MAT237 Lecture Notes

Tyler Holden, ©2015-2016

Contents

1 The Topology of $\mathbb{R}^n$

1.1 Sets and notation

As we start our adventure into the world of multivariate and vector calculus, we must first ensure that everybody is on the same page in terms of notation and basic set theory. While it is entirely possible that the reader may already be passingly familiar with all of the following topics, one could dedicate an entire course to exploring this subject, so it is worth meditating over, even if only superficially. We will begin by reviewing sets and the fundamental operations on sets, then follow this with functions between such sets.

1.1.1 Basic Set Theory

A *set* is any collection of distinct objects.[1] Some examples of sets might include

$$\text{the alphabet} = \{a, b, c, \ldots, x, y, z\}, \qquad \begin{matrix}\text{Universities in} \\ \text{Toronto}\end{matrix} = \{\text{UofT}, \text{Ryerson}, \text{York}\},$$

$$\text{The Kardashian Sisters} = \{\text{Kim}, \text{Khloe}, \text{Kourtney}\}.$$

We use the symbol '$\in$' (read as 'in') to talk about when an element is in a set; for example, $1 \in \{1, 2, 3\}$ but $\ddot{\frown} \notin \{\text{dog}, \text{cat}\}$.

Each of the previous examples were *finite* sets, as they consisted of only a finite number of elements. A set can also have infinitely many elements. In such instances, it is inconvenient to write out every element of the set so we use *set builder notation.* Herein, if P is a proposition on the set S, such that for each $x \in S$, $P(x)$ is either true or false, then one can define the set

$$\{x \in S : P(x)\}$$

which consists of all the elements in S which make P true. For example, if M is the set of months in the year, then

$$\{m \in M : m \text{ has 31 days}\} = \{\text{January}, \text{March}, \text{May}, \text{July}, \text{August}, \text{October}, \text{December}\}.$$

This was an example where the resulting set was still finite, but it still demonstrates the compactness of setbuilder notation.

The following are some important infinite sets that we will see throughout the course:

- The **naturals**[2] $\mathbb{N} = \{0, 1, 2, 3, \ldots\}$,
- The **integers** $\mathbb{Z} = \{..., -2, -1, 0, 1, 2, ...\}$,
- The **rationals** $\mathbb{Q} = \{p/q : p, q \in \mathbb{Z}, q \neq 0, \gcd(p, q) = 1\}$,

[1] This is not true, since it is possible to define objects called *classes*, but we will not worry about this too much in this context

[2] Some mathematicians do not believe that 0 is a natural number. I am personally undecided, and always just choose which version is more convenient.

- The **reals** $\mathbb{R}$ (the set of all infinite decimal expansion).

We can also talk about *subsets*, which are collections of items in a set and indicated with a '$\subseteq$' sign. For example, if P is the set of prime numbers, then $P \subseteq \mathbb{Z}$, since every element on the left (a prime number) is also an element of the right (an integer). Alternatively, one has $\mathbb{N} \subseteq \mathbb{Z} \subseteq \mathbb{Q} \subseteq \mathbb{R}$. There is a particular distinguished set, known as the *empty set* and denoted by $\emptyset$, which contains no elements. Recalling the definition of a vacuous truth, it is not too hard to convince oneself that empty set is a subset of every set!

Exercise: Determine the subset relations for the following sets:

1. $S = \{x \in \mathbb{R} : x = 2n, n \in \mathbb{Z}\}$,
2. $T = \left\{x \in \mathbb{R} : x = a - \frac{1}{2}, \forall a \in \mathbb{N}\right\}$,
3. $U = \left\{x \in \mathbb{Q} : x = \frac{p}{2^n}, \gcd(p,k) = 1\right\}$,
4. $V = \{x \in \mathbb{Z} : x = 3^n, n \in \mathbb{N}\}$.

1.1.2 Operations on Sets

Union and Intersection Let S be a set and choose two sets $A, B \subseteq S$. We define the *union* of A and B to be

$$A \cup B = \{x \in S : x \in A \text{ or } x \in B\}$$

and the *intersection* of A and B to be

$$A \cap B = \{x \in S : x \in A \text{ and } x \in B\}.$$

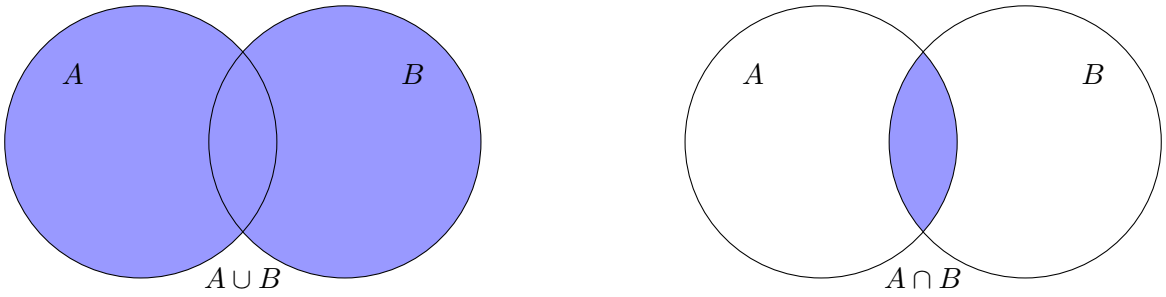


Figure 1: Left: The union of two sets is the collection of all elements which are in both (though remember that elements of sets are distinct, so we do not permit duplicates). Right: The intersection of two sets consists of all elements which are common to both sets.

Example 1.1

Determine the union and intersection of the following two sets:

$$A = \{x \in \mathbb{R} : x > 1\}, \qquad B = \{x \in \mathbb{R} : -1 < x < 2\}.$$

Solution. By definition, one has

$$\begin{aligned} A \cup B &= \{x \in \mathbb{R} : x \in A \text{ or } x \in B\} = \{x \in \mathbb{R} : x > 1 \text{ or } -1 < x < 2\} \\ &= \{x \in \mathbb{R} : x > -1\}, \\ A \cap B &= \{x \in \mathbb{R} : x \in A \text{ and } x \in B\} = \{x \in \mathbb{R} : x > 1 \text{ and } -1 < x < 2\} \\ &= \{x \in \mathbb{R} : 1 < x < 2\}. \end{aligned}$$

■

Let $I \subseteq \mathbb{N}$ be an indexing set: Given a collection of sets $\{A_i\}_{i \in I}$ in S, one can take the intersection or union over the entire collection, and this is often written as

$$\bigcup_{i \in I} A_i = \{x \in S : \exists i \in I, x \in A_i\}, \qquad \bigcap_{i \in I} A_i = \{x \in S : \forall i \in I, x \in A_i\}.$$

Example 1.2

Consider the set $\{x \in \mathbb{R} : \sin(x) > 0\}$. Write this set as as an infinite union of intervals.

Solution. We are well familiar with the fact that $\sin(x) > 0$ on $(0, \pi)$, $(2\pi, 3\pi)$, $(4\pi, 5\pi)$, etc. If we let the interval $I_n = (2n\pi, (2n+1)\pi)$ then the aforementioned intervals are I_0, I_1, and I_2. We can convince ourselves that that $\sin(x) > 0$ on any of the I_n, and hence

$$\{x \in \mathbb{R} : \sin(x) > 0\} = \bigcup_{n \in \mathbb{Z}} I_n = \bigcup_{n \in \mathbb{Z}} (2n\pi, (2n+1)\pi).$$ ■

Example 1.3

Define $I_n = \left(0, \frac{1}{n}\right) \subseteq \mathbb{R}$. Determine $I = \bigcap_{n \in \mathbb{N}} I_n$.

Solution. By definition, I consists of the elements which are in I_n for every $n \in \mathbb{N}$. We claim that I cannot consist of any positive real number. Indeed, if $p > 0$ then there exists $n \in \mathbb{N}$ such that $\frac{1}{n} < p$, which means that $p \notin I_k$ for all $k \geq n$, and hence cannot be in I. Since I has no positive real numbers, and certainly cannot contain any non-positive real numbers, we conclude that $I = \emptyset$. ■

Exercise: Let $I_n = (-n, n) \subseteq \mathbb{R}$ for $n \in \mathbb{N}$. Determine both $\bigcup_n I_n$ and $\bigcap_n I_n$.

Complement If $A \subseteq S$ then the *complement* of A with respect to S is all elements which are not in A; that is,

$$A^c = \{x \in S : x \notin A\}.$$

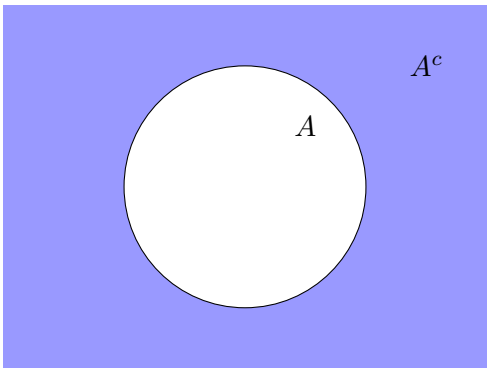


Figure 2: The complement of a set A with respect to S is the set of all elements which are in S but not in A.

Example 1.4

Determine the complement of $I = \bigcup_{n\in\mathbb{Z}}(2n\pi, (2n+1)\pi)$ from Example 1.2, with respect to $\mathbb{R}$.

Solution. Since I contains all the open intervals of the form $(2n\pi, (2n+1)\pi)$ we expect its complement to contain everything else. Namely,

$$I^c = \bigcup_{n\in\mathbb{Z}} [(2n-1)\pi, 2n\pi].$$ ■

Exercise:

1. Show that $(A \cup B)^c = A^c \cap B^c$,
2. Show that $(A \cap B)^c = A^c \cup B^c$,
3. Verify that $I^c = \bigcap_{n\in\mathbb{Z}} (2n\pi, (2n+1)\pi)^c$ is an equivalent solution for Example 1.1.2.

Cartesian Product The Cartesian product of two sets A and B is the collection of ordered pairs, one from A and one from B; namely,

$$A \times B = \{(a, b) : a \in A, b \in B\}.$$

A geometric way (which does not generalize well) is to visualize the Cartesian product as sticking a copy of B onto each element of A, or vice-versa. For our purposes, the main example of the product will be to define higher dimensional spaces. For example, we know that we can represent the plane $\mathbb{R}^2$ as an ordered pair of points $\mathbb{R}^2 = \{(x, y) : x, y \in \mathbb{R}\}$, while three dimensional space is an ordered triple $\mathbb{R}^3 = \{(x, y, z) : x, y, z \in \mathbb{R}\}$. In this sense, we see that $\mathbb{R}^2 = \mathbb{R} \times \mathbb{R}$, $\mathbb{R}^3 = \mathbb{R} \times \mathbb{R} \times \mathbb{R}$, and motivates the more general definition of $\mathbb{R}^n$ as an ordered n-tuple

$$\mathbb{R}^n = \underbrace{\mathbb{R} \times \cdots \times \mathbb{R}}_{n\text{-times}}.$$

Exercise: We have swept some things under the rug in defining $\mathbb{R}^n$, largely because the true nature is technical and boring. There is no immediate reason to suspect that $\mathbb{R} \times \mathbb{R} \times \mathbb{R}$ should be well defined: we first need to check that the Cartesian product is associative; that is, $(\mathbb{R} \times \mathbb{R}) \times \mathbb{R} = \mathbb{R} \times (\mathbb{R} \times \mathbb{R})$. By definition, the left-hand-side is

$$(\mathbb{R} \times \mathbb{R}) \times \mathbb{R} = \{((a,b),c) : (a,b) \in \mathbb{R} \times \mathbb{R}, c \in \mathbb{R}\}$$

while the right-hand-side is

$$\mathbb{R} \times (\mathbb{R} \times \mathbb{R}) = \{(a,(b,c)) : a \in \mathbb{R}, (b,c) \in \mathbb{R} \times \mathbb{R}\}.$$

Syntactically, neither of these looks the same as $\mathbb{R}^3 = \{(a,b,c) : a,b,c \in \mathbb{R}\}$, but nonetheless they all define the same data.

Exercise: Let $S^1 = \left\{(x,y) : x^2 + y^2 = 1\right\} \subseteq \mathbb{R}^2$ be the unit circle. What familiar shape is $S^1 \times S^1$?

1.1.3 Functions Between Sets

Given two sets A, B, a function $f : A \to B$ is a map which assigns to every point in A a *unique* point of B. If $a \in A$, we usually denote the corresponding element of B by $f(a)$. When specifying the function, one may write $a \mapsto f(a)$. The set A is termed the *domain*, while B is termed the *codomain*.

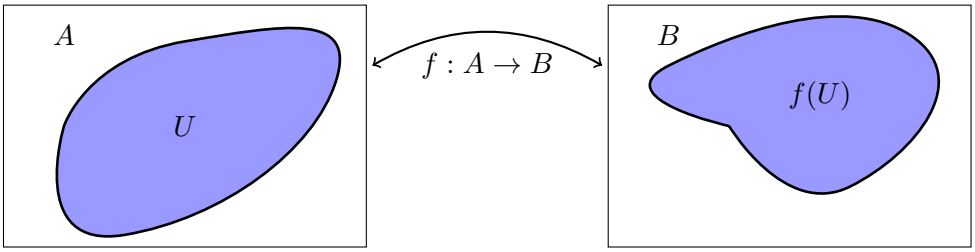
It is important to note that not every element of B needs to be hit by f; that is, B is not necessarily the range of f. Rather, B represents the ambient space to which f maps. Also, if either of the domain or codomain changes the function itself changes. This is because the data of the domain and codomain are intrinsic to the definition of a function. For example, $f : \mathbb{R} \to \mathbb{R}$ given by $f(x) = x^2$ is a different function than $g : \mathbb{R} \to [0,\infty)$, $g(x) = x^2$.

Definition 1.5

Let $f : A \to B$ be a function.

1. If $U \subseteq A$, then we define the *image* of U to be
$$f(U) = \{y \in B : \exists x \in U, f(x) = y\} = \{f(x) : x \in U\}.$$
2. If $V \subseteq B$, we define the *pre-image* of V to be
$$f^{-1}(V) = \{x \in A : f(x) \in V\}.$$

Note that despite being written as $f^{-1}(V)$, the preimage of a set does not say anything about the existence of an inverse function.



Example 1.6

Let $f : \mathbb{R} \to \mathbb{R}$ be specified by $f(x) = x^2$. Determine $f([0,1])$ and $f^{-1}(f([0,1])$.

Solution. By definition, one has

$$f([0,1]) = \{f(x) : x \in [0,1]\} = [0,1].$$

On the other hand, since $f([0,1]) = [0,1]$ we know that $f^{-1}(f([0,1])) = f^{-1}([0,1])$ for which

$$f^{-1}([0,1]) = \{x \in \mathbb{R} : f(x) \in [0,1]\} = [-1,1].$$ ■

Example 1.7

Let $f : \mathbb{R}^3 \to \mathbb{R}^2$ be given by $f(x,y,z) = (x,y)$. If

$$S^2 = \left\{(x,y,z) \in \mathbb{R}^3 : x^2 + y^2 + z^2 = 1\right\},$$

determine $f(S^2)$.

Solution. Let $(a,b,c) \in S^2$ so that $a^2 + b^2 + c^2 = 1$. The image of this point under f is $f(a,b,c) = (a,b)$. It must be the case that $a^2 + b^2 \leq 1$, and so $f(S^2) \subseteq D^2 = \left\{(x,y) \in \mathbb{R}^2 : x^2 + y^2 \leq 1\right\}$. We claim that this is actually an equality; that is, $f(S^2) = D^2$. In general, to show that two sets A and B are equal, we need to show $A \subseteq B$ and $B \subseteq A$. As we have already shown that $f(S^2) \subseteq D^2$, we must now show that $D^2 \subseteq f(S^2)$.

Let $(a,b) \in D^2$ so that $a^2 + b^2 \leq 1$. Let $c = \sqrt{1 - a^2 - b^2}$, which is well-defined by hypothesis. Then $a^2 + b^2 + c^2 = 1$ so that $(a,b,c) \in S^2$, and $f(a,b,c) = (a,b)$. Thus $f(S^2) = D^2$. ■

Exercise: Let $f : \mathbb{R}^3 \to \mathbb{R}^2$ be the function given in Example 1.7. Determine $f^{-1}(D^2)$.

We conclude this section by mentioning three important classes of function:

Definition 1.8

Let $f : A \to B$ be a function. We say that

1. f is *injective* if whenever $f(x) = f(y)$ then $x = y$,
2. f is *surjective* if for every $y \in B$ there exists an $x \in A$ such that $f(x) = y$,
3. f is *bijective* if f is both injective and surjective.

Notice that the choice of domain and codomain are exceptionally important determining whether a function is injective or surjective. For example, the function $f : \mathbb{R} \to \mathbb{R}$ given by $f(x) = x^2$ is not surjective (it misses the negative real numbers), while the function $f : \mathbb{R} \to [0, \infty)$ is surjective (there are no negative real numbers to miss).

Example 1.9

Determine whether the following functions are injective, surjective, or bijective.

1. $f : \mathbb{R}^3 \to \mathbb{R}^2$, $f(x, y, z) = (x, y)$,
2. $g : \mathbb{R}^2 \to \mathbb{R}^2$, $g(x, y) = (e^x, (x^2 + 1)y)$,
3. $h : \mathbb{R}^2 \to \mathbb{R}^2$, $h(x, y) = (y, x)$.

Solution.

1. The function f is certainly not injective, since $f(x, y, a) = (x, y) = f(x, y, b)$ for any a and b. On the other hand, it is surjective, since if $(x_0, y_0) \in \mathbb{R}^2$ then $f(x_0, y_0, 0) = (x_0, y_0)$.
2. The function g is injective: to see this, note that if $g(a_1, b_1) = g(a_2, b_2)$ then $(e^{a_1}, (a_1^2+1)b_1) = (e^{a_2}, (a_2^2+1)b_2)$ which can only happen if $e^{a_1} = e^{a_2}$. Since the exponential function is injective, $a_1 = a_2$. This in turn implies that $(a_1^2 + 1) = (a_2^2 + 1)$ and neither can be zero, so dividing the second component we get $b_1 = b_2$ as required. On the other hand, g is not surjective. For example, there is no point which maps to $(0, 0)$. codomain too large, some y does not have corresponding x in domain
3. This function is both injective and surjective. Both are left as simple exercises. We conclude that h is bijective.

■

1.2 Structures on $\mathbb{R}^n$

1.2.1 The Vector Space Structure

Any student familiar with linear algebra knows that $\mathbb{R}^n$ admits a vector space structure: one can add vectors in $\mathbb{R}^n$ and multiply by scalars. For those uninitiated, we briefly review the subject here.

Very roughly, a real vector space is any set in which two elements may be added to get another element of the set, as well as multiplied by a real number. Additionally, there must be an element $\mathbf{0}$ such that summing against zero does nothing. The full collection of axioms that define a vector space are too many to write down and are the topic of a linear algebra course, so refer the student to their favourite textbook.

The elements of the set are called *vectors*, while the real number multiples are called *scalars*. For notation sake, we will denote vectors by bold font $\mathbf{x}$ and scalars by non-bold font.

Recall that elements $\mathbf{x} \in \mathbb{R}^n$ just look like n-tuples of real numbers. If $\mathbf{x} = (x_1, \ldots, x_n), \mathbf{y} = (y_1, \ldots, y_n)$ are elements of $\mathbb{R}^n$ we can add them together and multiply by a scalar $c \in \mathbb{R}$ in a pointwise fashion

$$\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} + \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} x_1 + y_1 \\ x_2 + y_2 \\ \vdots \\ x_n + y_n \end{pmatrix}, \qquad c \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} cx_1 \\ cx_2 \\ \vdots \\ cx_n \end{pmatrix}.$$

The zero vector is, unsurprisingly, the n-tuple consisting entirely of zeroes: $\mathbf{0} = (0, \ldots, 0)$. See Figure 3 for a visualization of vector addition and scalar multiplication.

The diligent student might notice that I have been sloppy in writing vectors: there is a technical but subtle difference between vectors written horizontally and those written vertically. Once again, we will not be terribly concerned with the distinction in this course, so we will use whichever convention is simplest. In the event that the distinction is necessary, we will mention that point explicitly at the time.

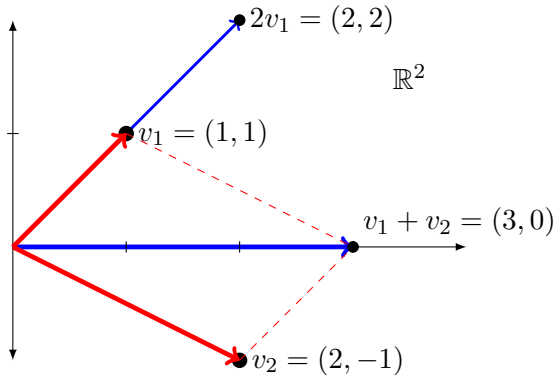


Figure 3: One may think of a vector as either representing a point in the plane (represented by the black dots) or as direction with magnitude (represented by the red arrows). The blue arrows correspond to the sum $v_1 + v_2$ and the scalar multiple $2v_1$. Notice that both are simply computed pointwise.

1.2.2 Of Lengths and Such

There are three intimately related structures which we will now impose on $\mathbb{R}$, which are the notion of an inner product, a norm, and a metric. The first is that of an inner product. While there

are many different kinds of inner products, the one with which we will be most concerned is the *Euclidean inner product*, also known as simply the *dot product*. Given two vectors $\mathbf{x} = (x_1, \ldots, x_n)$ and $\mathbf{y} = (y_1, \ldots, y_n)$ in $\mathbb{R}^n$, we write

$$\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{x} \cdot \mathbf{y} := \sum_{i=1}^{n} x_i y_i = x_1 y_1 + x_2 y_2 + \cdots + x_n y_n.$$

Geometrically, the dot product $\mathbf{x} \cdot \mathbf{y}$ is the length of the projection of $\mathbf{x}$ onto the unit vector in the $\mathbf{y}$ direction, or vice versa. More precisely, if $\mathbf{y} \in \mathbb{R}^n$ then $\hat{\mathbf{y}} = \frac{\mathbf{y}}{\|\mathbf{y}\|}$ is a unit vector that points in the same direction as $\mathbf{y}$, and

$$\frac{\langle \mathbf{x}, \mathbf{y} \rangle}{\|\mathbf{y}\|} \hat{\mathbf{y}} = \frac{\langle \mathbf{x}, \mathbf{y} \rangle}{\|\mathbf{y}\|^2} \mathbf{y}$$

is the projection of $\mathbf{x}$ into $\hat{\mathbf{y}}$. If $\langle \mathbf{v}, \mathbf{w} \rangle = 0$, we say that $\mathbf{v}$ and $\mathbf{w}$ are *orthogonal*, which we recognize will happen precisely when $\mathbf{x}$ and $\mathbf{y}$ are perpendicular.

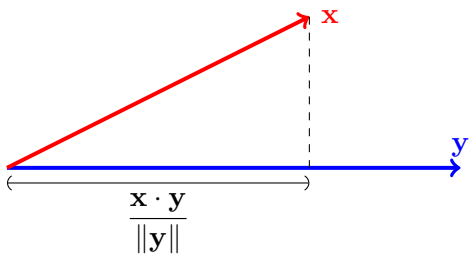


Figure 4: The inner product of $\mathbf{x}$ and $\mathbf{y}$, written $\mathbf{x} \cdot \mathbf{y}$ is the length of the projection of the vector $\mathbf{x}$ onto $\mathbf{y}$.

Example 1.10

If $\mathbf{v} = (1, -1, 2)$ and $\mathbf{w} = (2, 0, 4)$ are vectors in $\mathbb{R}^3$, determine $\langle \mathbf{v}, \mathbf{w} \rangle$.

Solution. We need only apply the definition to find that

$$\langle \mathbf{v}, \mathbf{w} \rangle = v_1 w_1 + v_2 w_2 + v_3 w_3 = (1 \times 2) + (-1 \times 0) + (2 \times 4) = 10.$$ ■

Proposition 1.11

The inner product satisfies the following properties: Let $\mathbf{x}, \mathbf{y}, \mathbf{z} \in \mathbb{R}^n$ and $c \in \mathbb{R}$,

1. **Symmetry:** $\langle \mathbf{x}, \mathbf{y} \rangle = \langle \mathbf{y}, \mathbf{x} \rangle$,
2. **Non-negative:** $\langle \mathbf{x}, \mathbf{x} \rangle \geq 0$ and is equal to zero if and only if $\mathbf{x} = 0$,
3. **Linearity:** $\langle c\mathbf{x} + \mathbf{y}, \mathbf{z} \rangle = c \langle \mathbf{x}, \mathbf{z} \rangle + \langle \mathbf{y}, \mathbf{z} \rangle$.

These properties are straightforward to verify and are left as an exercise for the student.

The next structure is called a *norm*, and prescribes a way of measuring the length of a vector. Our motivation comes from the one-dimensional case, where we know that the absolute value $|\cdot|$ is used to measure distance. As such, we define $\|\cdot\| : \mathbb{R}^n \to \mathbb{R}$ as the function

$$\|\mathbf{x}\| := \sqrt{\langle \mathbf{x}, \mathbf{x} \rangle} = \left(\sum_{i=1}^{n} x_i^2 \right)^{1/2} = \sqrt{x_1^2 + x_2^2 + \cdots + x_n^2}.$$

First, we recognize that this generalizes the Pythagorean Theorem in $\mathbb{R}^2$, since if $\mathbf{x} = (x, y)$ then the vector $\mathbf{x}$ looks like the hypotenuse of a triangle with side lengths x and y. The length of the hypotenuse is just $\sqrt{x^2 + y^2} = \|\mathbf{x}\|$ (See Figure 5).

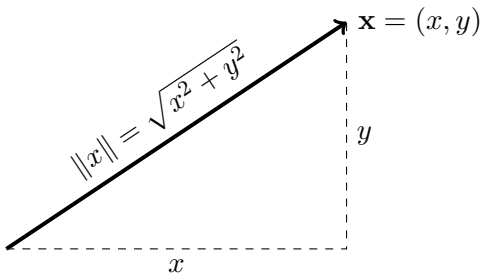


Figure 5: In $\mathbb{R}^2$, the length of a vector can be derived from the Pythagorean theorem. The norm $\|\cdot\|$ generalizes this notion to multiple dimensions.

Exercise: Let $\mathbf{x} = (x, y, z) \in \mathbb{R}^3$. Determine the length of this vector using the Pythagorean theorem and confirm that one gets the same value as $\|\mathbf{x}\|$.

Example 1.12

If $\mathbf{v} = (6, 8, 5)$, compute $\|\mathbf{v}\|$.

Solution. We again just apply the definition to get

$$\|\mathbf{v}\| = \sqrt{6^2 + 8^2 + 5^2} = \sqrt{125} = 5\sqrt{5}.$$ ■

A very important relationship between the inner product and the norm of $\mathbf{v}$ is the *Cauchy-Schwarz inequality*:

Proposition 1.13

If $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ then

$$|\langle \mathbf{x}, \mathbf{y} \rangle| \leq \|\mathbf{x}\| \|\mathbf{y}\|.$$

This proof is not terribly enlightening, nor is it very intuitive. The student may refer to the textbook for a proof.

Example 1.14

Let $\mathbf{v} = (1, -1, 2)$ and $\mathbf{w} = (2, 0, 4)$ as in Example 1.10. Compute $\|\mathbf{v}\|$ and $\|\mathbf{w}\|$ and confirm that the Cauchy-Schwarz inequality holds.

Solution. We already saw that $\langle \mathbf{v}, \mathbf{w} \rangle = 10$. Computing the norms one gets

$$\|\mathbf{v}\| = \sqrt{6}, \qquad \|\mathbf{w}\| = \sqrt{20}$$

so that $\|\mathbf{v}\|\|\mathbf{w}\| = \sqrt{120}$ which is greater than $10 = \sqrt{100}$. ■

Proposition 1.15

Let $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ and $c \in \mathbb{R}$. The norm $\|\cdot\|$ satisfies the following properties:

1. **Non-degeneracy:** $\|\mathbf{x}\| \geq 0$ with equality if and only if $\mathbf{x} = 0$,
2. **Normality:** $\|c\mathbf{x}\| = |c|\|\mathbf{x}\|$,
3. **Triangle Inequality:** $\|\mathbf{x} + \mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|$.

Proof. The first two properties follow immediately from properties of the inner product and are left as an exercise for the student. We resolve thus to prove the Triangle Inequality. Here one has

$$\begin{aligned} \langle \mathbf{x} + \mathbf{y}, \mathbf{x} + \mathbf{y} \rangle &= \langle \mathbf{x}, \mathbf{x} \rangle + 2\langle \mathbf{x}, \mathbf{y} \rangle + \langle \mathbf{y}, \mathbf{y} \rangle \\ &\leq \langle \mathbf{x}, \mathbf{x} \rangle + 2\|\mathbf{x}\|\|\mathbf{y}\| + \langle \mathbf{y}, \mathbf{y} \rangle \qquad \text{by Cauchy-Schwarz} \\ &= (\|\mathbf{x}\| + \|\mathbf{y}\|)^2 . \end{aligned}$$

By taking the square root of both sides, we get the desired result. □

The triangle inequality is so named because it relates the sides of a triangle. Indeed, if $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ and we form the triangle whose vertices are the point $\mathbf{x}, \mathbf{y}$ and $(\mathbf{x} + \mathbf{y})$, then the length of $\mathbf{x} + \mathbf{y}$ will be less than the sum of the other two side angles. Equality will occur precisely when $\mathbf{x} = c\mathbf{y}$ for some $c \in \mathbb{R}$.

Finally, one has a *metric*, which is a method for determining the distance between two vectors. If $\mathbf{x} = (x_1, \ldots, x_n)$ and $\mathbf{y} = (y_1, \ldots, y_n)$ then the *Euclidean metric* is

$$d(\mathbf{x}, \mathbf{y}) = \|\mathbf{x} - \mathbf{y}\| = \left(\sum_{i=1}^{n} (x_i - y_i)^2 \right) = \sqrt{(x_1 - y_1)^2 + \cdots + (x_n - y_n)^2}.$$

In $\mathbb{R}^2$, this exactly agrees with the usual distance formula.

Proposition 1.16

Let $\mathbf{x}, \mathbf{y}, \mathbf{z} \in \mathbb{R}^3$.

1. **Symmetry:** $d(\mathbf{x}, \mathbf{y}) = d(\mathbf{y}, \mathbf{x})$
2. **Non-degeneracy:** $d(\mathbf{x}, \mathbf{y}) \geq 0$ with equality if and only if $\mathbf{x} = \mathbf{y}$,
3. **Triangle Inequality:** $d(\mathbf{x}, \mathbf{z}) \leq d(\mathbf{x}, \mathbf{y}) + d(\mathbf{y}, \mathbf{z})$.

All of these properties follow immediately from the properties of the norm and are left as an exercise for the student. We will often omit the $d(\mathbf{x}, \mathbf{y})$ notation as some students may find it confusion, though this is typically how metrics are denoted in more abstract courses.

1.2.3 Cross product

In $\mathbb{R}^3$, the cross product of two vectors is a way of determining a third vector which is orthogonal to the original two. It is defined as follows: If $\mathbf{v} = (v_1, v_2, v_3)$ and $\mathbf{w} = (w_1, w_2, w_3)$ then

$$\mathbf{v} \times \mathbf{w} = (v_2 w_3 - w_2 v_3, w_1 v_3 - v_1 w_3, v_1 w_2 - w_1 v_2).$$

This is rather terrible to remember though, so if the student is familiar with determinants, it can be written as

$$\mathbf{v} \times \mathbf{w} = \det \begin{pmatrix} \hat{\imath} & \hat{\jmath} & \hat{k} \\ v_1 & v_2 & v_3 \\ w_1 & w_2 & w_3 \end{pmatrix}.$$

Here $\hat{\imath}, \hat{\jmath}, \hat{k}$ represent the standard unit vectors in $\mathbb{R}^3$, so that $(a, b, c) = a\hat{\imath} + b\hat{\jmath} + c\hat{k}$.

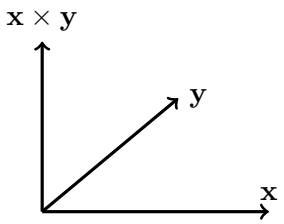


Figure 6: The cross product of two vectors $\mathbf{x} \times \mathbf{y}$.

Example 1.17

If $\mathbf{v} = (1, 0, 1)$ and $\mathbf{w} = (1, 2, 3)$, determine $\mathbf{v} \times \mathbf{w}$.

Solution. Using our definition, one has

$$(1, 0, 1) \times (1, 2, 3) = \det \begin{pmatrix} \hat{\imath} & \hat{\jmath} & \hat{k} \\ 1 & 0 & 1 \\ 1 & 2 & 3 \end{pmatrix} = (-2, -2, 2)$$

As we mentioned, this new vector should be orthogonal to the other two. Computing the dot products, we have

$$\langle \mathbf{v}, \mathbf{v} \times \mathbf{w} \rangle = (1,0,1) \cdot (-2,-2,2) = -2 + 2 = 0$$
$$\langle \mathbf{w}, \mathbf{v} \times \mathbf{w} \rangle = (1,2,3) \cdot (-2,-2,2) = -2 - 4 + 6 = 0$$ ■

Exercise:

1. Show that $\mathbf{v} \times \mathbf{w} = -\mathbf{w} \times \mathbf{v}$.
2. Show that $\langle \mathbf{v}, \mathbf{v} \times \mathbf{w} \rangle = 0$ in general.
3. Show that if $\mathbf{w} = \lambda \mathbf{v}$ for some $\lambda \in \mathbb{R}$, then $\mathbf{v} \times \mathbf{w} = 0$. Conclude that the cross product of two vectors in $\mathbb{R}^3$ is non-zero if and only if the vectors are linearly independent.

1.3 Open, Closed, and Everything in Between

The goal of the next several sections is to discuss the notion of *topology*, which is the coarse grained geometry and structure of a space. In single variable calculus, one was exposed to the notions of open intervals (a,b), closed intervals $[a,b]$, and the knowledge that some intervals are neither open nor closed. What motivates the nomenclature for these sets? Intuitively, the idea seems to be that the set (a,b) does not contain its endpoints a and b: it contains points which are arbitrarily close, but not those two specific points. A closed interval does contain its endpoints, it is closed off. Our goal is to extend this notion to $\mathbb{R}^n$, where the addition of dimensions significantly complicates our picture. However, we can at least start somewhere nice, by defining the generalization of an interval:

Definition 1.18

Let $\mathbf{x} \in \mathbb{R}^n$ and $r > 0$ a real number. We define the *open ball of radius r at the point $\mathbf{x}$* as

$$B_r(\mathbf{x}) := \{\mathbf{y} \in \mathbb{R}^n : \|\mathbf{x} - \mathbf{y}\| < r\}.$$

Recalling that $\|\mathbf{x} - \mathbf{y}\|$ is equivalent to the distance between $\mathbf{x}$ and $\mathbf{y}$, the open ball $B_r(\mathbf{x})$ is nothing more than the collection of points which are a distance at most r from $\mathbf{x}$. This indeed generalizes the interval, since in $\mathbb{R}^1$ we have

$$B_r(x) = \{y \in \mathbb{R} : |x - y| < r\} = (x - r, x + r),$$

or if we centre around 0, $B_r(0) = (-r, r)$. In $\mathbb{R}^2$ we get a disk of radius r,

$$B_r(\mathbf{0}) = \left\{(x,y) \in \mathbb{R}^2 : \sqrt{x^2 + y^2} \leq r\right\},$$

which we recognize as being the same as $x^2 + y^2 \leq r^2$.

Definition 1.19

A set $S \subseteq \mathbb{R}^n$ is bounded if there exists an $r > 0$ such that $S \subseteq B_r(\mathbf{0})$.

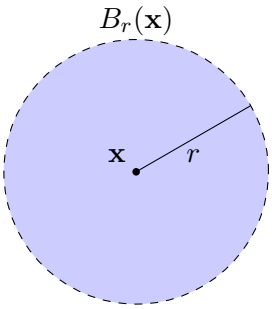


Figure 7: The open ball of radius r centred at $\mathbf{x}$ consists of all points which are a distance r from $\mathbf{x}$.

One hopes that this is fairly intuitive: A set is bounded if we can put a ball around it. If we can place a ball around the set, it cannot grow arbitrarily large. For example, the set $S = \{(x, y) \in \mathbb{R}^2 : xy > 0\}$ consists of the first and third quadrants of the plane. Since both x and y can become arbitrarily large is absolute value, no ball centred at the origin entirely contains S. On the other hand, $C = \{(x, y) \in \mathbb{R}^2 : (x - a)^2 + (y - b)^2 \leq r^2\}$ is bounded *for any* choice of $a, b, c \in \mathbb{R}$.

Exercise: For an arbitrary choice of $a, b, r \in \mathbb{R}$, determine the open ball that bounds C as defined above.

These balls will be our way of "looking around" a point; namely, if we know something $B_r(\mathbf{x})$ then we know what is happening within a distance r of the point $\mathbf{x}$. We can use these open balls to define different types of points of interest:

Definition 1.20

Let $S \subseteq \mathbb{R}^n$ be an arbitrary set.

1. We say that $\mathbf{x} \in \mathbb{R}^n$ is an *interior point of* S if there exists an $r > 0$ such that $B_r(\mathbf{x}) \subseteq S$; that is, $\mathbf{x}$ is an interior point if we can enclose it in an open ball which is entirely contained in S.

2. We say that $\mathbf{x} \in \mathbb{R}^n$ is a *boundary point of* S if for every $r > 0$, $B_r(\mathbf{x}) \cap S \neq \emptyset$ and $B_r(\mathbf{x}) \cap S^c \neq \emptyset$; that is, $\mathbf{x}$ is a boundary point if no matter what ball we place around $\mathbf{x}$, that ball lives both inside and outside of S.

If S is a set, we define the *interior of* S, denoted S^{int} to be the collection of interior points of S. We define the *boundary of* S, denoted ∂S, to be the collection of boundary points of S.

We should take a moment and think about these definitions, and why they make sense. Let us start with a boundary point. Intuitively, a boundary point is any point which occurs at the very fringe of the set; that is, if I push a little further I will leave the set. An interior point should be a point inside of S, such that if I move in any direction a sufficiently small distance, I stay within the set. This is exactly what Definition 1.20 is conveying. By definition, if $\mathbf{x}$ is an interior point then we must have that $\mathbf{x} \in S$; however, boundary points *do not* need to be in the set. We start

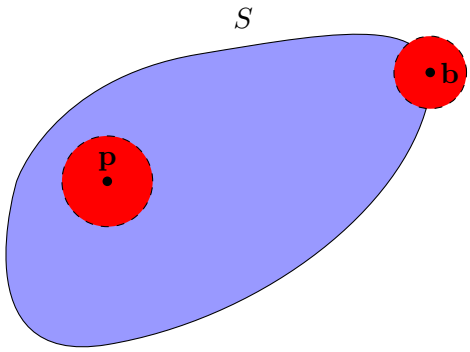


Figure 8: The point **b** is a boundary point. No matter what size ball we place around **b**, that ball will intersect both S and S^c. On the other hand, **p** is an interior point, since we can place a ball around it which lives entirely within S.

with a silly example:

Example 1.21

Let $S = (-1, 1]$. What are the interior points and the boundary points of S?

Solution. We claim that any point in $(-1, 1)$ is an interior point. To see that this is the case, let $p \in (-1, 1)$ be an arbitrary point. We need to place a ball around p which lives entirely within $(-1, 1)$. To do this, assume without loss of generality that $p \geq 0$. If $p = 0$ then we can set $r = \frac{1}{2}$ and $B_{1/2}(0) = (-1/2, 1/2) \subseteq (-1, 1)$. Thus assume that $p \neq 0$ and let $r = \frac{1-p}{2}$, which represents half the distance from p to 1.

We claim that $B_r(p) \subseteq (-1, 1)$. Indeed, let $x \in B_r(p)$ be any point, so that $|x - p| < r$ by definition. Then

$$\begin{aligned} |x| &= |x - p + p| \leq |x - p| + p \\ &\leq r + p = \frac{1-p}{2} + p \\ &= \frac{1+p}{2} < 1 \end{aligned}$$

where in the last inequality we have used the fact that $p < 1$ so $1 + p < 2$. Thus $x \in (-1, 1)$, and since x was arbitrary, $B_r(p) \subseteq (-1, 1)$.

The boundary points are ± 1, where we note that even though $-1 \notin (-1, 1]$, it is still a boundary point. To see that $+1$ is a boundary point, let $r > 0$ be arbitrary, so that $B_r(p) = (1 - r, 1 + r)$. We then have

$$B_r(p) \cap (-1, 1] = (1 - r, 1] \neq \emptyset, \qquad B_r(p) \cap (-1, 1)^c = (1, 1 + r) \neq \emptyset,$$

as required. The proof for -1 is analogous and left as an exercise. ■

Example 1.22

What is the boundary of $\mathbb{Q}$ in $\mathbb{R}$?

Solution. We claim that $\partial\mathbb{Q} = \mathbb{R}$. Since both the irrationals and rationals are dense in the real numbers, we know that every non-empty open interval in $\mathbb{R}$ contains both a rational and irrational number. Thus let $x \in \mathbb{R}$ be any real number, and $r > 0$ be arbitrary. The set $B_r(x)$ is an open interval around x, and contains a rational number, showing that $B_r(x) \cap \mathbb{Q} \neq \emptyset$. Similarly, $B_r(x)$ contains an irrational number, showing that $B_r(x) \cap \mathbb{Q}^c \neq \emptyset$, so $x \in \partial\mathbb{Q}$. Since x was arbitrary, we conclude that $\partial\mathbb{Q} = \mathbb{R}$. ■

Exercise: Show that it is impossible for a boundary point to also be an interior point, and vice-versa.

Definition 1.23

A set $S \subseteq \mathbb{R}^n$ is said to be *open* if every point of S is an interior point; that is, S is open if for every $\mathbf{x} \in S$ there exists an $r > 0$ such that $B_r(\mathbf{x}) \subseteq S$. The set S is *closed* if S^c is open.

Example 1.24

The set $S = \{(x, y) \in \mathbb{R}^2 : y > 0\}$ is open.

Solution. We need to show that around every point in S we can place an open ball that remains entirely within S. Choose a point $\mathbf{p} = (p_x, p_y) \in S$, so that $p_y > 0$, and let $r = p_y/2$. Consider the ball $B_r(\mathbf{p})$, which we claim lives entirely within S. To see that this is the case, choose any other point $\mathbf{q} = (q_x, q_y) \in B_r(\mathbf{p})$. Now used the fact that norm of a vector's component is less then the vector itself

$$|q_y - p_y| \leq \|\mathbf{q} - \mathbf{p}\| < r = \frac{p_y}{2}$$

which implies that $q_y > p_y - \frac{p_y}{2} = \frac{p_y}{2} > 0$. Since $q_y > 0$ this shows that $\mathbf{q} \in S$, and since $\mathbf{q}$ was arbitrary, $B_r(\mathbf{p}) \subseteq S$ as required. ■

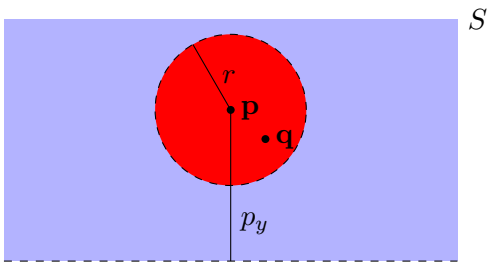


Figure 9: The upper half plane is open. For any point, look at its y-coordinate p_y and use the ball of radius $p_y/2$.

Example 1.25

Every open ball is open.

Solution. One certainly hopes that this result is true, otherwise the name would by quite the misnomer. To show this, let $\mathbf{x} \in \mathbb{R}^n$ and $r > 0$ both be arbitrary, and consider $S = B_r(\mathbf{x})$. We need to show that every point in S can in turn be enclosed with a smaller ball which lives entirely within S. Choose some $\mathbf{p} \in S$ and let $d = \|\mathbf{x} - \mathbf{p}\|$ so that $d < r$ be definition.

We claim that the ball of radius $r' = (r-d)/2 > 0$ will work. To see this, choose an arbitrary $\mathbf{y} \in B_{r'}(\mathbf{p})$ so that $\|\mathbf{p} - \mathbf{y}\| < r'$. One has that

$$\begin{aligned} \|\mathbf{x} - \mathbf{y}\| &\leq \|\mathbf{x} - \mathbf{p}\| + \|\mathbf{p} - \mathbf{y}\| && \text{triangle inequality} \\ &\leq d + r' = d + \frac{r-d}{2} = \frac{r+d}{2} \\ &\leq \frac{2r}{2} < r && \text{since } d < r \quad . \end{aligned}$$

Since $\mathbf{y}$ was arbitrary, $B_{r'}(\mathbf{p}) \subseteq B_r(\mathbf{x})$ as required. ■

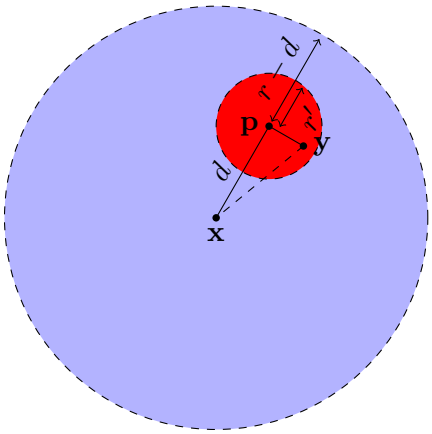


Figure 10: A visualization of the solution to Example 1.25.

Proposition 1.26

A set $S \subseteq \mathbb{R}^n$ is closed if and only if $\partial S \subseteq S$.

Proof. [$\Rightarrow$] Assume that S is closed, and for the sake of contradiction assume that $\partial S \nsubseteq S$. Choose an element $\mathbf{x} \in \partial S$ which is not in S, so that $\mathbf{x} \in S^c$. Now since S is closed, S^c is open, so we can find an $\epsilon > 0$ such that $B_\epsilon(\mathbf{x}) \subseteq S^c$. However, this is a contradiction: since $\mathbf{x} \in \partial S$ then every open ball must intersect both S and S^c, and this shows that $B_\epsilon(\mathbf{x})$ is an open ball around $\mathbf{x}$ which fails to intersect S. We thus conclude $\partial S \subseteq S$ as required.

[$\Leftarrow$] We will proceed by contrapositive, and show that if S is not closed, then $\partial S \nsubseteq S$. If S is not closed, then S^c is not open, and hence there is some point $\mathbf{x} \in S^c$ such that for every $r > 0$,

$B_r(\mathbf{x}) \cap S \neq \emptyset$. Certainly $B_r(\mathbf{x}) \cap S^c \neq \emptyset$ (since both sets contain $\mathbf{x}$) and hence $\mathbf{x} \in \partial S$. Thus $\mathbf{x}$ is a point in $\partial S \cap S^c$, and so $\partial S \not\subseteq S$. □

Just as in the case of intervals in $\mathbb{R}$, it is possible for a set to be neither open nor closed. A previous exercise showed that a point in a set cannot be both a boundary point and an interior point, so failing to be open somehow amounts to containing some of your boundary points. If you have all of your boundary points, Proposition 1.26 shows that you are actually closed. Thus sets which fail to be both open or closed contain some of their boundary points, but not all of them. By adding all the boundary points, we can "close off" a set.

Definition 1.27

If $S \subseteq \mathbb{R}^n$ then the closure of S is the set $\overline{S} = S \cup \partial S$.

Exercise:

1. Show that $\overline{S}$ is always a closed set.
2. Show that S is closed if and only if $S = \overline{S}$.
3. Show that $\overline{S}$ is the *smallest* closed set containing S.

The closure of an interval (a, b) is the closed interval $[a, b]$. The student should check that the closure of the open ball in $\mathbb{R}^n$ is

$$\overline{B_r(\mathbf{x})} = \{\mathbf{y} \in \mathbb{R}^n : \|\mathbf{x} - \mathbf{y}\| \leq r\}$$

where we note that the inequality need no longer be strict. Similarly, the closure of the open half plane in Example 1.24 is the closed half plane

$$\overline{\{(x, y) \in \mathbb{R}^2 : y > 0\}} = \{(x, y) \in \mathbb{R}^2 : y \geq 0\}.$$

1.4 Sequences and Completeness

The student is already passingly familiar with the notion of sequences in $\mathbb{R}$. We will quickly review the pertinent points before introducing sequences in $\mathbb{R}^n$.

1.4.1 Sequences in $\mathbb{R}$

The rough idea of a sequence is that it resembles an ordered collection of real numbers. We can make this more formal by writing a sequence as a map $x : \mathbb{N} \to \mathbb{R}$, so that $x(n)$ is a real number. For example, the sequence $x(n) = n^2$ is such that

$$x(1) = 1, \quad x(2) = 4, \quad x(3) = 9, \quad x(4) = 16, \quad x(5) = 25, \ldots.$$

For brevity of notation, one often writes x_n instead of $x(n)$, so that for example, $x_4 = 16$. In addition to this, we often choose to conflate the function x itself with its (ordered) image in $\mathbb{R}$, in

which case we write the sequence as $(x_n)_{n=0}^{\infty}$. When we are feeling particularly lazy, we will even drop the indexing and just write (x_n).

A subsequence is a method for extracting elements from a sequence, while keeping them in the same order defined by the original sequence. In particular, given a sequence $x : \mathbb{N} \to \mathbb{R}$ and an *increasing* map $n : \mathbb{N} \to \mathbb{N}$, a subsequence is the sequence formed by the composition $x(n(k))$, which we often write as x_{n_k}. For example, if $x_n = n^2$ as above and $n(k) = 2k$, then

$$x_{n_1} = x_2 = 4, \quad x_{n_2} = x_4 = 16, \quad x_{n_3} = x_6 = 36, \quad x_{n_4} = x_8 = 64, \quad \dots$$

This is the subsequence which picks out every other member of the original sequence:

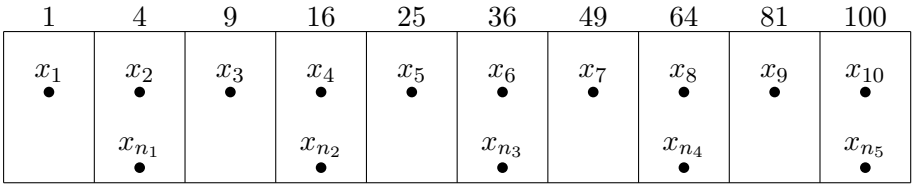


Figure 11: The subsequence x_{n_k} picks out every other number from the sequence, and defines a new sequence.

Definition 1.28

Let $(x_n)_{n=1}^{\infty}$ be a sequence in $\mathbb{R}$. We say that

1. (x_n) is *bounded* if there exists an $M > 0$ such that $|x_n| \leq M$ for every $n \in \mathbb{N}$,
2. (x_n) is *increasing* if $x_{n+1} > x_n$ for every $n \in \mathbb{N}$. Similarly, (x_n) is *decreasing* if $x_{n+1} < x_n$ for every $n \in \mathbb{N}$.

Example 1.29

Determine whether the sequence is increasing/decreasing or bounded:

1. $x_n = \frac{1}{n}$,
2. $y_n = (-1)^n$
3. $z_n = 2^n$.

Solution.

1. The sequence $x_n = \frac{1}{n}$ is bounded and decreasing. Setting $M = 1$ we have $|x_n| = \left|\frac{1}{n}\right| \leq 1$ since $n \geq 1$. In addition, it is well known that $x_{n+1} = \frac{1}{n+1} < \frac{1}{n} = x_n$, so the sequence is decreasing as required.
2. This sequence just oscillates between the numbers ± 1, so it is certainly bounded (choose $M = 1$). On the other hand, it is neither increasing nor decreasing, since $y_{2n} > y_{2n+1}$ but also $y_{2n+2} > y_{2n+1}$.
3. This sequence is increasing, since $z_{n+1} = 2^{n+1} = 2z_n > z_n$. However, the sequence is unbounded. ■

We can talk about when such sequences converge, in a manner similar to horizontal asymptotes of functions:

Definition 1.30

If $(x_n)_{n=1}^{\infty}$ is a sequence in $\mathbb{R}$, we say that (x_n) *converges with limit* L, written as $(x_n) \xrightarrow{n\to\infty} L$, if for every $\epsilon > 0$ there exists an $N \in \mathbb{N}$ such that whenever $n > N$, $|x_n - L| < \epsilon$.

This definition says that by progressing sufficiently far into the sequence, we can ensure that the x_n get arbitrarily close to L.

Example 1.31

If $x_n = \frac{1}{n}$, show that $(x_n) \to 0$.

Solution. Let $\epsilon > 0$ be given, and choose an $N \in \mathbb{N}$ such that $\frac{1}{N} < \epsilon$. If $n > N$ then

$$|x_n - 0| = \frac{1}{n} < \frac{1}{N} < \epsilon$$

which is what we wanted to show. ■

It seems intuitive to expect that the limit of a sequence should be unique, so that we may talk about *the* limit of the sequence. We demonstrate this with the following proposition:

Proposition 1.32

If $(x_n)_{n=1}^{\infty}$ is a sequence in $\mathbb{R}$ such that $(x_n) \to x$ and $(x_n) \to y$ then $x = y$; that is, limits of convergent sequences are unique.

Proof. It is sufficient to show that for every $\epsilon > 0$ we have $|x - y| < \epsilon$. Indeed, this will show that $|x - y|$ cannot be positive, and so must necessarily be zero, from which it will follow that $x = y$.

Let $\epsilon > 0$ be arbitrary. As $(x_n) \to x$ there exists $N_1 \in \mathbb{N}$ such that $|x_n - x| < \frac{\epsilon}{2}$. By the same token, there exists $N_2 \in \mathbb{N}$ such that $|x_n - y| < \frac{\epsilon}{2}$. Let $N = \max\{N_1, N_2\}$ and fix any $n > N$, so that

$$|x - y| \leq |x - x_n| + |x_n - y| \leq \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon.$$ □

since epsilon is arbitrary, we can take epsilon arbitrarily small, together with |x-y| < epsilon. we prove that x-y cannot be positive

Our profound laziness as mathematicians means that if we can avoid doing more work, we will. We can use the following proposition, akin to the limit laws for functions, to infer convergence of sequences and their limits.

Theorem 1.33: Limit Laws for Sequences

Let $(a_n) \to L$ and $(b_n) \to M$ be convergent sequences.

1. The sequence $(a_n + b_n)$ is convergent and $(a_n + b_n) \to L + M$,
2. The sequence $(a_n b_n)$ is convergent and $(a_n b_n) \to LM$,
3. For any $\alpha \in \mathbb{R}$ the sequence (αa_n) converges and $(\alpha a_n) \to \alpha L$,
4. If $M \neq 0$ then the sequence (a_n / b_n) converges and $(a_n / b_n) \to L/M$.

Proof. The proof of these are almost identical to those of the limit laws for functions. We will prove (1) and leave the remainder as a (non-trivial) exercise:

Assume that $(a_n) \to L$ and $(b_n) \to M$. Let $\epsilon > 0$ be given and choose $M_1, M_2 \in \mathbb{N}$ such that if $k \geq M_1$ then $|a_n - L| < \frac{\epsilon}{2}$ and if $k \geq M_2$ then $|b_n - M| < \frac{\epsilon}{2}$. Let $M = \max\{M_1, M_2\}$, so that if $k \geq M$ then

$$|a_n + b_n - (L + M)| \leq |a_n - L| + |b_n - M| < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon./qedhere$$

triangular inequality again!!

□

There is also a version of the Squeeze Theorem for sequences, again proven in almost an identical fashion:

Theorem 1.34: Squeeze Theorem for Sequences

Let $(a_n), (b_n)$, and (c_n) be sequences in $\mathbb{R}$, and assume that for sufficiently large k we have $a_n \leq b_n \leq c_n$. If $(a_n) \to L$ and $(c_n) \to L$, then (b_n) is also convergent with limit L.

We know from our previous experience that every convergent sequence is bounded (prove it yourself!). A partial converse is the following:

Theorem 1.35: Monotone Convergence Theorem

If (a_n) is bounded from above and non-decreasing, then (a_n) is convergent with its limit given by $\sup\{a_n : n \in \mathbb{N}\}$.

Proof. Let $L = \sup_n a_n$, which we know exists by the completeness axiom. Let $\epsilon > 0$ be given. By definition of the supremum, we know that there exists some $M \in \mathbb{N}$ such that

$$L - \epsilon < a_M \leq L.$$

Since (a_n) is non-decreasing, we have that for all $k \geq M$

$$L - \epsilon < a_M < a_k \leq L < L + \epsilon;$$

that is, $|a_n - L| < \epsilon$. Hence $(a_n) \to L$ as required. □

A similar argument shows that the theorem also holds if (a_n) is bounded and non-increasing.

Example 1.36

Determine whether the sequence $a_n = \frac{2^n}{n!}$ is convergent.

Solution. A quick computation shows that

$$\frac{a_{n+1}}{a_n} = \frac{2^{n+1}}{(n+1)!}\frac{n!}{2^n} = \frac{2}{n+1},$$

so that if $n \geq 2$ we have $a_{n+1} < a_n$ and the sequence is decreasing. It is easy to see that a_n is always positive, hence bounded below by 0. By the Monotone Convergence Theorem we know that (a_n) converges. ■

Proposition 1.37

If $(x_n)_{n=1}^{\infty} \to x$ is a convergent sequence, then every subsequence (x_{n_k}) is also convergent with the same limit.

Proof. Let (x_{n_k}) be any subsequence and note that $n_k \geq k$ for all k (prove this). Let $\epsilon > 0$ be given. Since $(x_n) \to x$ there exists $N \in \mathbb{N}$ such that if $n > N$ then $|x_n - x| < \epsilon$. This N will also work for (x_{n_k}), since if $k > N$ then $n_k > n_N > N$, implying that $|x_{n_k} - x| < \epsilon$. □

1.4.2 Sequences in $\mathbb{R}^m$

With our rapid review of sequences in $\mathbb{R}$, we can now begin considering sequences in $\mathbb{R}^m$. A *sequence* in $\mathbb{R}^m$ is any function $\mathbf{x} : \mathbb{N} \to \mathbb{R}^m$. Just like before, we will often write such as sequences as $\mathbf{x}_n := \mathbf{x}(n)$. For example, the map $\mathbf{x}(n) = (n, n^2 - 1)$ is a sequence in $\mathbb{R}^2$ whose first few elements are given by

$$\mathbf{x}_1 = (1, 0), \quad \mathbf{x}_2 = (2, 3), \quad \mathbf{x}_3 = (3, 8), \quad \mathbf{x}_4 = (4, 15), \quad \mathbf{x}_5 = (5, 24), \quad \ldots$$

If $n : \mathbb{N} \to \mathbb{N}$ then one can define a subsequence by $\mathbf{x}(k(n)) = \mathbf{x}_{k_n}$. For example, if $n(k) = 3k - 1$, then

$$\mathbf{x}_{k_1} = (2, 3), \quad \mathbf{x}_{k_2} = (5, 24), \quad \mathbf{x}_{k_3} = (8, 63), \quad \mathbf{x}_{k_4} = (11, 120), \quad \ldots$$

Remark 1.38 Let $(\mathbf{x}_n)$ be a sequence in $\mathbb{R}^m$, and write $\mathbf{x}_n = (x_n^1, \ldots, x_n^m) \in \mathbb{R}^m$. By picking out the components, we can define m sequences in $\mathbb{R}$ by $\left(x_n^k\right)_{n=1}^{\infty}$. If $(\mathbf{x}_{n_\ell})_{\ell=1}^{\infty}$ is a subsequence, this defines subsequences $(\mathbf{x}_{n_\ell}^k)_{\ell=1}^{\infty}$ (we are running out of letters!). However, notice that the converse is certainly not true: One cannot take subsequences of each $(\mathbf{x}_n^k)$ and stitch them back together to get a subsequence of $(\mathbf{x}_n)$. For example, consider the sequence $\mathbf{x}_n = (n, -n)$ in $\mathbb{R}^2$. This defines two sequences in $\mathbb{R}$, one by $x_n = n$ and $y_n = -n$. Let's take the subsequence of (x_n) consisting of even indices, so that $x_{n_r} = 2n$, and the subsequence of y_n consisting of odd indices $y_{n_s} = -(2n-1)$.

$$x_{n_1} = 2, \quad x_{n_2} = 4, \quad x_{n_3} = 6, \quad x_{n_4} = 8, \quad \ldots$$
$$y_{n_1} = -1, \quad y_{n_2} = -3, \quad y_{n_3} = -5, \quad y_{n_4} = -7, \quad \ldots$$

There is no way of combining these individual subsequences to arrive at a subsequence of $(\mathbf{x}_n)$.

Our interest lies principally with sequences which converge, for which the definition is almost identical to the one for sequences in $\mathbb{R}$:

Definition 1.39

Let $(\mathbf{x}_n)_{n=1}^{\infty}$ be a sequence in $\mathbb{R}^n$. We say that $(\mathbf{x}_n)$ converges with limit $\mathbf{x} \in \mathbb{R}^n$, written $(\mathbf{x}_n) \to \mathbf{x}$, if for every $\epsilon > 0$ there exists an $N \in \mathbb{N}$ such that whenever $n > N$ then $\|\mathbf{x}_n - \mathbf{x}\| < \epsilon$.

Exercise: Let $(\mathbf{x}_n)_{n=1}^{\infty}$ be a sequence in $\mathbb{R}^m$ and write $\mathbf{x}_n = (x_n^1, \ldots, x_n^m)$. Show that $(\mathbf{x}_n)$ converges if and only if (x_n^i) converges for $i = 1, \ldots, m$.

The corresponding theorems about uniqueness of limits, the limit laws, and the Squeeze Theorem all hold in $\mathbb{R}^n$ as well as $\mathbb{R}$, with the only effective change being that the absolute value $|\cdot|$ becomes the norm $\|\cdot\|$. The student is encouraged to prove these in multiple dimensions to check that they also work.

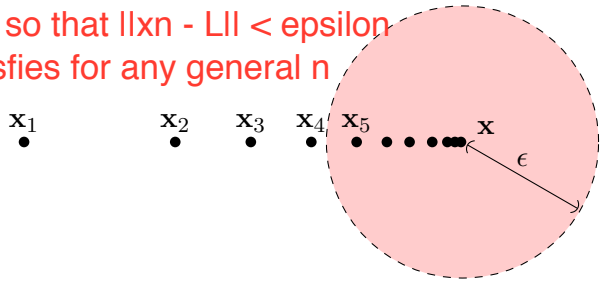
We may feel comfortable with the definition of convergent sequences, since it is only a slight modification of what we have seen repetitively in both this course and its prequel. However, with our discussion of balls, we now have an opportunity to associate a strong geometric interpretation to the idea of convergence. The condition $\|\mathbf{x}_n - \mathbf{x}\| < \epsilon$ says that $\mathbf{x}_n$ is in $B_\epsilon(\mathbf{x})$, so the definition of convergence can equivalently be colloquialized by saying that $(\mathbf{x}_n) \to \mathbf{x}$ if

"Every ball around $\mathbf{x}$ contains all but finitely many points of the sequence."

Example 1.40

Show that the sequence $\mathbf{x}_n = (x_n, y_n) = \left(\frac{1}{n}, \frac{1}{n^2}\right) \to (0, 0)$.

1. let epsilon be arbitrary
2. choose N in terms of epsilon so that ||xn - L|| < epsilon
3. prove that the inequality satisfies for any general n



Solution. Let $\epsilon > 0$ be given, and choose N such that $\frac{1}{N} < \frac{\epsilon}{\sqrt{2}}$. If $n > N$ then $\frac{\sqrt{2}}{n} < \frac{\sqrt{2}}{N}$ and

$$\|(x_n, y_n) - (0,0)\| = \sqrt{\frac{1}{n^2} + \frac{1}{n^4}} = \frac{1}{n}\sqrt{1 + \frac{1}{n^2}} \leq \frac{\sqrt{2}}{n} \qquad 1 + \frac{1}{n^2} \leq 2$$
$$< \frac{\sqrt{2}}{n} < \epsilon.$$

■

another way to determine if set is closed

One can use sequences to characterize the closure of a set, and hence determine whether or not a set is closed. If $S \subseteq \mathbb{R}^n$, we say that $(\mathbf{x}_n)$ is a *sequence in* S if $\mathbf{x}_n \in S$ for every $n \in \mathbb{N}$. The closure of S is the collection of all limit points of convergent sequences in S:

Proposition 1.41

meaning every element in sequence is in S

If $S \subseteq \mathbb{R}^n$, then $x \in \overline{S}$ if and only if there exists a convergent sequence $(\mathbf{x}_n)$ in S such that $(\mathbf{x}_n) \to \mathbf{x}$.

construction of xn based on a ball centered at x, which is in closure of S

Proof. [⇒] Assume that $\mathbf{x} \in \overline{S}$ so that every ball around around $\mathbf{x}$ intersects S. We need to construct a sequence in S which converges to $\mathbf{x}$. For each $n \in \mathbb{N}$, choose an element $\mathbf{x}_n \in B_{1/n}(\mathbf{x}) \cap S$, which is non-empty by assumption (See Figure 12). By construction, the sequence $(\mathbf{x}_n)_{n=1}^{\infty}$ is a sequence in S, so we need only show that $(\mathbf{x}_n) \to \mathbf{x}$. Let $\epsilon > 0$ be given and choose N such that $\frac{1}{N} < \epsilon$. When $n > N$ we have $\frac{1}{n} < \frac{1}{N} < \epsilon$, and by construction $\mathbf{x}_n \in B_{1/n}(\mathbf{x}) \subseteq B_\epsilon(\mathbf{x})$, or equivalently

essentially, every point in closure of S has a ball containing a sequence; the hard part is choosing the N, but its also intuitive to set sequence to 1/n

$$\|\mathbf{x}_n - \mathbf{x}\| < \frac{1}{n} < \epsilon.$$

[⇐] Let $(\mathbf{x}_n)_{n=1}^{\infty}$ be a convergent sequence in S with limit point $\mathbf{x}$. If $\epsilon > 0$ is any arbitrary real number, then there exists an $N \in \mathbb{N}$ such that for all $n > N$ we have $\mathbf{x}_n \in B_\epsilon(\mathbf{x})$. Since $\mathbf{x}_n \in S$, this implies that $B_\epsilon(\mathbf{x}) \cap S \neq \emptyset$. Since ϵ was arbitrary, $\mathbf{x} \in S$ or $\mathbf{x} \in \partial S$. In either case, $\mathbf{x} \in \overline{S}$. □

closure implies every ball around any point in S intersects S

Since we know that a set S is closed if and only if $S = \overline{S}$, one immediately gets the following Corollary.

Corollary 1.42

= x in \overline{S}

A set $S \subseteq \mathbb{R}^n$ is closed if and only if whenever $(\mathbf{x}_n)_{n=1}^{\infty}$ is a convergent sequence in S with $(\mathbf{x}_n) \to \mathbf{x}$, then $\mathbf{x} \in S$.

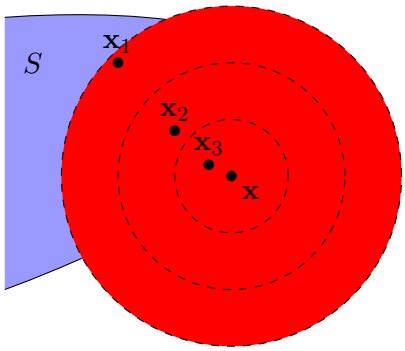


Figure 12: In the proof of Proposition 1.41, we need to construct a sequence which converges to $\mathbf{x}$. This is done by choosing points in successively smaller balls around $\mathbf{x}$.

1.4.3 Completeness

Our goal in this section is to extract convergent subsequences from bounded sequences, in an effort to facilitate our future discussion of compactness. We have already reviewed the Monotone Convergence Theorem, from which the following (equivalent) Theorem follows:

Theorem 1.43: Nested Interval Theorem

For each $k \in \mathbb{N}$, let $I_k = [a_k, b_k]$ be a closed interval such that

$$I_1 \supseteq I_2 \supseteq I_3 \supseteq I_4 \supseteq \cdots \supseteq I_k \supseteq \cdots$$

is a nested collection of intervals, and $(b_k - a_k) \xrightarrow{k\to\infty} 0$; that is, the length of the intervals is getting smaller. Then the intersection of these intervals is non-empty, and in particular consists of a single element, say p. Notationally,

$$\bigcap_{k=1}^{\infty} I_k = \{p\}\,.$$

skipped…

Proof. Consider the sequences $(a_k)_{k=1}^{\infty}$ and $(b_k)_{k=1}^{\infty}$ defined by the endpoints of the intervals. Since the intervals are contained within one another, (a_k) is monotone increasing, while (b_k) is monotone decreasing. By the Monotone convergence theorem, both sequences converge. Moreover, since the length of the subintervals approach zero, the sequences converge to the same point (prove this more rigorously if you do not see it). Let this limit point be p, for which $a_k \leq p \leq b_k$ for every k, showing that

$$p \in \bigcap_{k=1}^{\infty} I_k.$$

Since the lengths of the intervals tend to zero, this is the only possible point in the intersection (once again, provide a more rigorous proof of this fact). □

Theorem 1.44

Every bounded sequence in $\mathbb{R}$ has a convergent subsequence.

Proof. The idea of the proof will be to exploit Theorem 1.43 by successively bisecting the sequence into two halves. This will lead to a chain of nested intervals, which must have a single point in common. We will then construct a sequence which converges to this point.

More formally, let $(a_n)_{n=1}^{\infty}$ be a bounded sequence, and $M > 0$ be such that $|a_n| \leq M$ for all $n \in \mathbb{N}$. In particular, $a_n \in [-M, M]$. Consider the two halves $[-M, 0]$ and $[0, M]$, one of which must contain infinitely many elements of the sequences. Call this interval I_1. We inductively construct the closed interval I_n as follows: Assume that I_{n-1} has been given, and split I_n into two halves. At least one of these halves must contain infinitely many elements of the set, so choose one and call it I_n.

By construction,

$$I_1 \supseteq I_2 \supseteq I_3 \supseteq \cdots,$$

and the length of the interval I_k is $M/2^{k-1}$. Clearly, as $k \to \infty$ the length of the subintervals tends to 0, and as such the Nested Interval Theorem implies there exists a point p which is contained in every such interval.

We now construct a sequence which converges to p. Let x_{k_1} be any element of (x_k) which lives in I_1. We construct x_{k_n} inductively as follows: Assume that $x_{k_{n-1}}$ has been specified. Since I_n contains infinitely many elements, there exists an element in I_n which is further along the sequence than $x_{k_{n-1}}$. Call this element x_{k_n}.

Finally, we show that $(x_{k_n}) \to p$. Let $\epsilon > 0$ be given and choose $N \in \mathbb{N}$ such that $\frac{M}{2^{N-1}} < \epsilon$. If $n > N$ then

$$|x_{k_n} - x| < (\text{length of } I_n) < \frac{M}{2^{n-1}} < \frac{M}{2^{N-1}} < \epsilon$$

as required. □

We wish to extend this to discuss sequences in $\mathbb{R}^n$. Though it no longer makes sense to talk about increasing or decreasing sequences (there is no natural way of ordering n-tuples), we can still talk about when a sequence is bounded.

or exists a ball such that xn in ball center at origin

Definition 1.45

A sequence $(\mathbf{x}_n)_{n=1}^{\infty}$ in $\mathbb{R}^m$ is bounded if there exists $M > 0$ such that $\|\mathbf{x}_n\| \leq M$ for every $n \in \mathbb{N}$.

Proposition 1.46

Every bounded sequence in $\mathbb{R}^n$ has a convergent subsequence.

Proof. We will give the explicit proof for $n = 2$, which contains all the important ideas, and comment on how to generalize it afterwards. Let $\mathbf{x}_n = (x_n, y_n)$ be a bounded sequence in $\mathbb{R}^2$. Note that $|x_n| \leq \|\mathbf{x}_n\|$ and so the sequences (x_n) and (y_n) are each bounded in $\mathbb{R}$. It is very tempting to

simply take a convergent subsequence of each, but the problem is that we cannot stitch them back together (See Remark 1.38).

Instead, let (x_{n_k}) be a convergent subsequence of (x_n), with limit say $\mathbf{x}$. *Using the same indices*, consider the subsequence (y_{n_k}). This sequence does not necessarily converge, but it is bounded, so it in turn has a convergent subsequence $(y_{n_{k_\ell}}) \to y$. We claim that the (sub)subsequence $(x_{n_{k_\ell}}, y_{n_{k_\ell}})$ converges. We already know that $(y_{n_{k_\ell}}) \to y$. Furthermore, since $(x_{n_{k_\ell}})$ is a subsequence of (x_{n_k}), which we know is convergent, Proposition 1.37 implies that $(x_{n_{k_\ell}}) \to x$. By Exercise 1.4.2, since each component converges, $(x_{n_{k_\ell}}, y_{n_{k_\ell}})$ converges, as required. □

A very closely related notion of convergence is the notion of a Cauchy sequence.

Definition 1.47

A sequence $(\mathbf{x}_n)_{n=1}^{\infty}$ is said to be *Cauchy* if for every $\epsilon > 0$ there exists a $N \in \mathbb{N}$ such that if $n, k > N$ then $\|x_n - x_k\| < \epsilon$.

Cauchy sequences are precisely those sequences whose elements get closer together the further we travel into the sequence. Indeed, if we translate the definition of a Cauchy sequence, it says

> "By going far enough into a Cauchy sequence, we can ensure that any two elements are as close together as we want."

benefit of a Cauchy sequence

The benefit of Cauchy sequences is that they seem to encapsulate the basic behaviour of a convergent sequence, but one does not need to know *a priori* the limit itself. The following proposition confirms this suspicion.

Proposition 1.48

If $(\mathbf{x}_n)_{n=1}^{\infty}$ is a sequence in $\mathbb{R}^n$, then $(\mathbf{x}_n)$ is Cauchy if and only if $(\mathbf{x}_n)$ is convergent.

triangular inequality again

Solution. [⇐] This is the easier of the two directions. Assume that $(\mathbf{x}_n)$ is convergent with limit point $\mathbf{x}$ and let $\epsilon > 0$ be given. Choose $N \in \mathbb{N}$ such that if $n > N$ then $\|\mathbf{x}_n - \mathbf{x}\| < \frac{\epsilon}{2}$. We claim that this N works for the definition of a Cauchy sequence. Indeed, let $k, n > N$ so that

$$\|\mathbf{x}_k - \mathbf{x}_n\| \leq \|\mathbf{x}_k - \mathbf{x}\| + \|\mathbf{x}_n - \mathbf{x}\| < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon$$

as required.

[⇒] Conversely, let us now assume that $(\mathbf{x}_n)$ is Cauchy. We will first show that $(\mathbf{x}_n)$ is bounded. Setting $\epsilon = 1$ there exists an $N \in \mathbb{N}$ such that whenever $n > N$ then $\|\mathbf{x}_n - \mathbf{x}_N\| < 1$, from which it follows that

$$\|\mathbf{x}_n\| < \|\mathbf{x}_n - \mathbf{x}_N\| + \|\mathbf{x}_N\| = 1 + \|\mathbf{x}_N\|.$$

By setting $M = \max\{\|\mathbf{x}_1\|, \ldots, \|\mathbf{x}_N\|, 1 + \|\mathbf{x}_N\|\}$ then $\|\mathbf{x}_n\| \leq M$ for all $k \in \mathbb{N}$.

By Proposition 1.46, $(\mathbf{x}_n)$ thus has a convergent subsequence $(\mathbf{x}_{n_k})$, say with limit point $\mathbf{x}$. We now claim that the original sequence actually converges $\mathbf{x}$ as well. Indeed, let $\epsilon > 0$ be chosen, and $N_1 \in \mathbb{N}$ be such that for all $k, \ell > N_1$ we have $\|\mathbf{x}_k - \mathbf{x}_\ell\| < \frac{\epsilon}{2}$. Similarly, choose $K \in \mathbb{N}$ such that

1. xn cauchy
2. xn bounded
3. xn has a convergent subsequence
4. xn converges to the same limist

for all $k > K$ we have $\|\mathbf{x}_{n_k} - \mathbf{x}\| < \frac{\epsilon}{2}$. Fix an integer $k > N_1$ such that $n_k > K$ so that if $n > K$ we have

$$\|\mathbf{x}_n - \mathbf{x}\| < \|\mathbf{x}_n - \mathbf{x}_{n_k}\| + \|\mathbf{x}_{n_k} - \mathbf{x}\| < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon. \qquad \blacksquare$$

Definition 1.49

We say that $S \subseteq \mathbb{R}^n$ is *complete* if every Cauchy sequence converges.

Proposition 1.48 implies that $\mathbb{R}^n$ is complete. We leave it as an exercise for the student to show that $S \subseteq \mathbb{R}^n$ is complete if and only if S is closed.

1.5 Continuity

We pause our discussion of sequence for the moment (to be resumed quite shortly) to discuss the notion of limits and continuity for functions of several variables. Let us briefly recall the definitions in a single variable, upon which we will generalize our discussion to multiple dimensions.

Definition 1.50

Let $f : \mathbb{R} \to \mathbb{R}$ with $c, L \in \mathbb{R}$. We say that $\lim_{x \to c} f(x) = L$ if for every $\epsilon > 0$ there exists $\delta > 0$ such that whenever $0 < |x - c| < \delta$ then $|f(x) - L| < \epsilon$.
We say that f is *continuous at* c if $\lim_{x \to c} f(x) = f(c)$. If f is continuous at every point in its domain, we simply say that f is continuous.

Continuity is a way of saying that the function behaves well under limits, or equivalently that limits can be taken "inside" the function, since

$$\lim_{x \to c} f(x) = f\left(\lim_{x \to c} x\right) = f(c).$$

This idea that continuous functions permit one to interchange the function evaluation with the limit will become more evident in a second. We presume that the student is still familiar with these notions (albeit perhaps a bit rusty), so we will not explore them further at this time and instead pass to multivariable functions.

Of particular interest will be functions of the form $f : \mathbb{R}^n \to \mathbb{R}$ (though a similar conversation holds for functions $f : \mathbb{R}^n \to \mathbb{R}^m$). For functions of a single variable, the idea of a limit is that as x gets arbitrarily close to c, the function value $f(x)$ becomes arbitrarily close to L. These notions were made formal by way of the absolute value, which measured distance: $|x - c|$ is the distance between x and c, while $|f(x) - L|$ is the distance between f and L. In $\mathbb{R}^n$ we have adapted to use the norm to measure distance, so it seems natural to replace $|x - c|$ with $\|\mathbf{x} - \mathbf{c}\|$.

Definition 1.51

Let $\mathbf{f} : \mathbb{R}^n \to \mathbb{R}^m$ with $\mathbf{c} \in \mathbb{R}^n$ and $\mathbf{L} \in \mathbb{R}^m$. We say that

$$\lim_{\mathbf{x}\to\mathbf{c}} \mathbf{f}(\mathbf{x}) = \mathbf{L}$$

if for every $\epsilon > 0$ there exists a $\delta > 0$ such that whenever $0 < \|\mathbf{x} - \mathbf{c}\| < \delta$ then $\|\mathbf{f}(\mathbf{x}) - \mathbf{L}\| < \epsilon$.

Note that these are *different* norms; the norm for $\|\mathbf{x} - \mathbf{c}\|$ is the $\mathbb{R}^n$ norm, while the norm for $\|\mathbf{f}(\mathbf{x}) - \mathbf{L}\|$ is in $\mathbb{R}^m$. The student is likely familiar with the unwieldy ϵ-δ approach to limits, and we assure the reader that this situation is significantly more exacerbated in multiple dimensions.

Example 1.52

Show that $\lim_{(x,y)\to(1,1)} (x + y) = 2$.

Solution. Recall that in general, for any arbitrary $(a, b) \in \mathbb{R}^2$ one has

$$|a| \le \sqrt{a^2 + b^2}, \qquad |b| \le \sqrt{a^2 + b^2} \tag{1.1}$$

Let $\epsilon > 0$ be given and choose $\delta = \frac{\epsilon}{2}$. Assume that $(x, y) \in \mathbb{R}^2$ satisfy $\|(x, y) - (1, 1)\| < \delta$ so that

$$\begin{aligned} |(x + y) - 2| &= |(x - 1) + (y - 1)| \le |x - 1| + |y - 1| \\ &\le \sqrt{(x-1)^2 + (y-1)^2} + \sqrt{(x-1)^2 + (y-1)^2} \qquad \text{by (1.1)} \\ &= 2\|(x, y) - (1, 1)\| < \epsilon. \end{aligned}$$

■

Example 1.53

Show that $\lim_{(x,y)\to(0,0)} \dfrac{xy}{\sqrt{x^2 + y^2}} = 0$.

Solution. Let $\epsilon > 0$ be given and choose $\delta = \epsilon$. If $(x, y) \in \mathbb{R}^2$ satisfy $\|(x, y)\| < \delta$ then

$$\begin{aligned} \left| \frac{xy}{\sqrt{x^2 + y^2}} - 0 \right| = \frac{|x||y|}{\sqrt{x^2 + y^2}} &\le \frac{\sqrt{x^2 + y^2}\sqrt{x^2 + y^2}}{\sqrt{x^2 + y^2}} \\ &= \sqrt{x^2 + y^2} = \|(x, y)\| < \epsilon. \end{aligned}$$

■

Example 1.54

Let $\mathbf{f} : \mathbb{R}^2 \to \mathbb{R}^3$ be given by $(x, y) \mapsto (x, x + y, x - y)$. Show that

$$\lim_{(x,y)\to(1,0)} \mathbf{f}(x, y) = (1, 1, 1).$$

Solution. Let $\epsilon > 0$ be given and choose $\delta = \dfrac{\epsilon}{\sqrt{3}}$. Notice that

$$\begin{aligned}\|(x-1, x+y-1, x-y-1)\|^2 &= (x-1)^2 + (x+y-1)^2 + (x-y-1)^2 \\ &= (x-1)^2 + \left[(x-1)^2 + 2(x-1)y + y^2\right] \\ &\qquad + \left[(x-1)^2 - 2(x-1)y + y^2\right] \\ &= 3(x-1)^2 + 2y^2 \leq 3\left[(x-1)^2 + y^2\right] \\ &= 3\|(x-1, y)\|^2\end{aligned}$$

and as such

$$\|\mathbf{f}(\mathbf{x}) - \mathbf{L}\| \leq \sqrt{3}\|\mathbf{x} - (1,0)\| < \epsilon. \qquad \blacksquare$$

Recall that

$$\lim_{x\to c} f(x) \text{ exists} \quad \Leftrightarrow \quad \lim_{x\to c^+} f(x) \text{ exists and } \lim_{x\to c^-} f(x) \text{ exists.}$$

This represents the fact that the limit exists if and only if the limit is the same regardless of which path I take to get to c. The problem in $\mathbb{R}^n$ is much more difficult, since even in $\mathbb{R}^2$ the number of ways in which a limit can be approached is infinite. In Example 1.53 we took the limit as $(x, y) \to (0, 0)$. We can approach the origin $(0, 0)$ along the x-axis, the y-axis, or along any line in $\mathbb{R}^2$ (see Figure 13). In fact, one need not even approach along lines, you can approach along any path in $\mathbb{R}^2$ that leads to the origin. For the limit to exist overall, the limit along every possible path to the origin must exists, and they must all be equal.

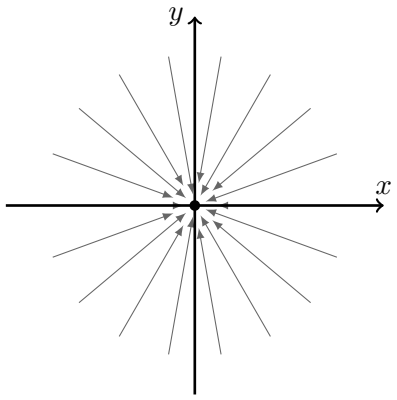


Figure 13: Even in $\mathbb{R}^2$, there are infinitely many ways of approaching a point. For a limit to exist, the limit along each path must exist and must be equal to that achieved from every other path.

Example 1.55

Show that the limit $\displaystyle\lim_{(x,y)\to(0,0)} \frac{x^2y^2}{x^4+y^4}$ does not exist.

Solution. Let us approach the origin along the straight lines $y = mx$, where $m \in \mathbb{R}$ is arbitrary. If the limit exists, it must be the same regardless of our choice of m. Let $f(x, y) = \frac{x^2y^2}{x^4+y^4}$, and note

that the path $y = mx$ can be written pairwise as (x, mx), and so

$$\begin{aligned}\lim_{x\to 0} f(x, mx) &= \lim_{x\to 0} \frac{x^2(mx)^2}{x^4 + (mx)^4} = \lim_{x\to 0} \frac{m^2x^4}{x^4 + m^4x^4} \\ &= \lim_{x\to 0} \frac{m^2x^4}{(m^4+1)x^4} = \lim_{x\to 0} \frac{m^2}{m^4+1} \\ &= \frac{m^2}{m^4+1}.\end{aligned}$$

This limit clearly depends upon the choice of m, and so we conclude that the limit does not exist. ■

The inquisitive reader might suspect that it is only straight lines that pose problems. For example, could it be the case that if the function exists along every line $y = mx$ then the limit can be guaranteed to exist? The following examples shows that this is note the case:

Example 1.56

Show that the function $f(x, y) = \frac{2xy^2}{x^2+y^4}$ admits a limit along every line $y = mx$, but fails along the parabola $x = my^2$.

Solution. Proceeding as suggested, we take the limit along the lines $y = mx$:

$$\lim_{x\to 0} f(x, mx) = \lim_{x\to 0} \frac{2x(mx)^2}{x^2 + (mx)^4} = \lim_{x\to 0} \frac{2m^2x^3}{x^2 + m^4x^4} = \lim_{x\to 0} \frac{2mx}{1 + m^4x^2} = 0$$

On the other hand, along the line $x = my^2$ we get

$$\lim_{y\to 0} f(my^2, y) = \lim_{y\to 0} \frac{2(my^2)y^2}{(my^2)^2 + y^4} = \lim_{y\to 0} \frac{2my^4}{m^2y^4 + y^4} = \lim_{y\to 0} \frac{2m}{m^2+1} = \frac{2m}{m^2+1}$$

and this clearly depends on m. We conclude that the limit does not exist. ■

Things seem rather hopeless: The ϵ-δ definition is tricky to work with, and the above examples show that we cannot even limits be evaluating along typical paths. What progress can we possibly make? Our salvation lies with the fact that the Squeeze Theorem also holds for functions $f : \mathbb{R}^n \to \mathbb{R}$.

Theorem 1.57: Multivariable Squeeze Theorem

Let $f, g, h : \mathbb{R}^n \to \mathbb{R}$ be functions and $\mathbf{c} \in \mathbb{R}^n$. Assume that in some neighbourhood of $\mathbf{c}$, such that $f(\mathbf{x}) \leq g(\mathbf{x}) \leq h(\mathbf{x})$ for all $\mathbf{x}$ in that neighbourhood. If

$$\lim_{\mathbf{x}\to\mathbf{c}} f(\mathbf{x}) = \lim_{\mathbf{x}\to\mathbf{c}} g(\mathbf{x}) = L, \qquad \text{then} \qquad \lim_{\mathbf{x}\to\mathbf{c}} g(\mathbf{x}) = L.$$

The proof is identical to that of the single variable squeeze theorem, so we leave it as an exercise.

Example 1.58

Show that $\lim_{(x,y)\to(0,0)} \dfrac{3x^2y^2}{x^2+y^2} = 0.$

Solution. Note that $y^2 \leq x^2 + y^2$, and so for $(x, y) \neq (0, 0)$,

$$0 \leq \frac{3x^2y^2}{x^2+y^2} \leq \frac{3x^2(x^2+y^2)}{x^2+y^2} = 3x^2.$$

In the limit as $(x, y) \to (0, 0)$ the bounding functions both tend to zero, so by the Squeeze Theorem we conclude

$$\lim_{(x,y)\to(0,0)} \frac{3x^2y^2}{x^2+y^2} = 0.$$ ■

Example 1.59

Determine the limit $\lim_{(x,y)\to(0,0)} \dfrac{y^4 \sin^2(xy)}{x^2+y^2}$.

Solution. Taking absolute values and using the fact that $|\sin(xy)| \leq 1$ and $y^2 \leq x^2 + y^2$ we get

$$0 \leq \left|\frac{y^4 \sin^2(xy)}{x^2+y^2}\right| \leq \frac{y^4}{x^2+y^2} \leq \frac{y^2(x^2+y^2)}{(x^2+y^2)} = y^2.$$

As both sides tend to zero as $(x, y) \to (0, 0)$ we conclude that

$$\lim_{(x,y)\to(0,0)} \left|\frac{y^2 \sin^2(xy)}{x^2+y^2}\right| = 0$$

from which the limit follows.[3] ■

Now that we have tools for discussing limits, we can move onto the notion of continuity, which in a multivariable context is nearly identical to the single variable definition.

Definition 1.60

A function $\mathbf{f} : \mathbb{R}^n \to \mathbb{R}^m$ is *continuous* at $\mathbf{c} \in \mathbb{R}^n$ if

$$\lim_{\mathbf{x}\to\mathbf{c}} \mathbf{f}(\mathbf{x}) = \mathbf{f}(\mathbf{c}).$$

If $\mathbf{f}$ is continuous at every point in its domain, we just say that $\mathbf{f}$ is continuous.

For example, the function $f(x, y) = \frac{y^4 \sin^2(xy)}{x^2+y^2}$ from Example 1.59 is undefined at $(0, 0)$, but if we define

$$g(x,y) = \begin{cases} \frac{y^4 \sin^2(xy)}{x^2+y^2}, & \text{if } x \neq 0 \\ 0, & \text{if } (x,y) = (0,0) \end{cases}$$

[3] Recall that $-|f(\mathbf{x})| \leq f(\mathbf{x}) \leq |f(\mathbf{x})|$, so if $|f(\mathbf{x})| \xrightarrow{\mathbf{x}\to c} 0$, the Squeeze Theorem implies that $f(\mathbf{x}) \xrightarrow{\mathbf{x}\to c} 0$.

then g is a continuous function.

It turns that there are (at least) two other equivalent notions of continuity, which will prove invaluable tools for studying topology.

Theorem 1.61

A function $\mathbf{f} : \mathbb{R}^n \to \mathbb{R}^m$ is continuous if and only if whenever $(\mathbf{a}_n)_{n=1}^{\infty} \to \mathbf{a}$ is a convergent sequence in $\mathbb{R}^n$, then $(\mathbf{f}(\mathbf{a}_n))_{n=1}^{\infty} \to \mathbf{f}(\mathbf{a})$ is a convergent sequence in $\mathbb{R}^m$.

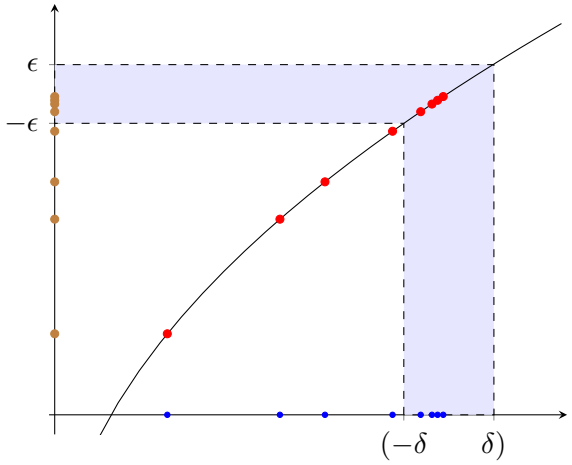


Figure 14: If $(a_n) \to a$, then by going far enough into our sequence (blue) we can guarantee that we will be in δ-neighbourhood of a. The image of these points are the $f(a_n)$ (brown), which live in the desired ϵ-neighbourhood because of the continuity of f.

Proof. [$\Rightarrow$] Assume that $\mathbf{f}$ is continuous, and let $(\mathbf{a}_n) \to \mathbf{a}$. We want to show that $(\mathbf{f}(\mathbf{a}_n)) \to \mathbf{f}(\mathbf{a})$. Let $\epsilon > 0$ be given. Since $\mathbf{f}$ is continuous, there exists a $\delta > 0$ such that for each $\mathbf{x}$ satisfying $\|\mathbf{x} - \mathbf{a}\| < \delta$ we have $\|\mathbf{f}(\mathbf{x}) - \mathbf{f}(\mathbf{a})\| < \epsilon$. Since $(\mathbf{a}_n)$ is convergent, there exists an $N \in \mathbb{N}$ such that for all $n \geq N$ we have $\|\mathbf{a}_n - \mathbf{a}\| < \delta$. Combining these, we see that whenever $n \geq N$ we have

$$\|\mathbf{a}_n - \mathbf{a}\| < \delta, \text{ and so } \|\mathbf{f}(\mathbf{a}_n) - \mathbf{f}(\mathbf{a})\| < \epsilon.$$

which is exactly what we want to show.

[$\Leftarrow$] Conversely, assume that $\mathbf{f}$ is not continuous, say at $\mathbf{c}$. Hence there exists an $\epsilon > 0$ such that for any $\delta > 0$ there is an $\mathbf{x}$ such that $\|\mathbf{x} - \mathbf{c}\| < \delta$ and $\|\mathbf{f}(\mathbf{x}) - \mathbf{f}(\mathbf{c})\| \geq \epsilon$. For each $\delta_n = \frac{1}{n}$, choose an element $\mathbf{x}_n$ satisfying $\|\mathbf{x}_n - \mathbf{c}\| < \delta_n$ and $\|\mathbf{f}(\mathbf{x}_n) - \mathbf{f}(\mathbf{c})\| \geq \epsilon$. Then $(\mathbf{x}_n) \to \mathbf{c}$ but $\mathbf{f}(\mathbf{x}_n)$ does not converge to $\mathbf{f}(\mathbf{c})$. □

Theorem 1.61 shows that a function is continuous if and only if it it maps convergent sequences to convergent sequences. This is precisely what we mean when we say that we can interchange a function with the limit, since if $(\mathbf{x}_n) \to \mathbf{a}$ then

$$\lim_{n\to\infty} \mathbf{f}(\mathbf{x}_n) = \mathbf{f}\left(\lim_{n\to\infty} \mathbf{x}_n\right) = \mathbf{f}(\mathbf{a}).$$

The other equivalent definition of continuity is more topological in nature:

Theorem 1.62

A function $\mathbf{f} : \mathbb{R}^n \to \mathbb{R}^m$ is continuous if and only if whenever $U \subseteq \mathbb{R}^m$ is an open set, then $\mathbf{f}^{-1}(U) \subseteq \mathbb{R}^n$ is also an open set.

Proof. [$\Rightarrow$] Assume that $\mathbf{f}$ is continuous and let $U \subseteq \mathbb{R}^m$ be an open set. Let $\mathbf{x} \in \mathbf{f}^{-1}(U)$ be arbitrary and consider $\mathbf{f}(\mathbf{x}) \in U$. Since U is open, there exists and $\epsilon > 0$ such that $B_\epsilon(\mathbf{f}(\mathbf{x})) \subseteq U$, and since $\mathbf{f}$ is continuous, let $\delta > 0$ be the choice of delta which corresponds to this epsilon. We claim that $B_\delta(\mathbf{x}) \subseteq \mathbf{f}^{-1}(U)$. Indeed, let $\mathbf{y} \in B_\delta(\mathbf{x})$ so that $\|\mathbf{x} - \mathbf{y}\| < \delta$. By continuity, $\|\mathbf{f}(\mathbf{x}) - \mathbf{f}(\mathbf{y})\| < \epsilon$ which shows that $\mathbf{f}(\mathbf{y}) \in B_\epsilon(\mathbf{f}(\mathbf{x})) \subseteq U$, thus $\mathbf{y} \in \mathbf{f}^{-1}(U)$ as required.

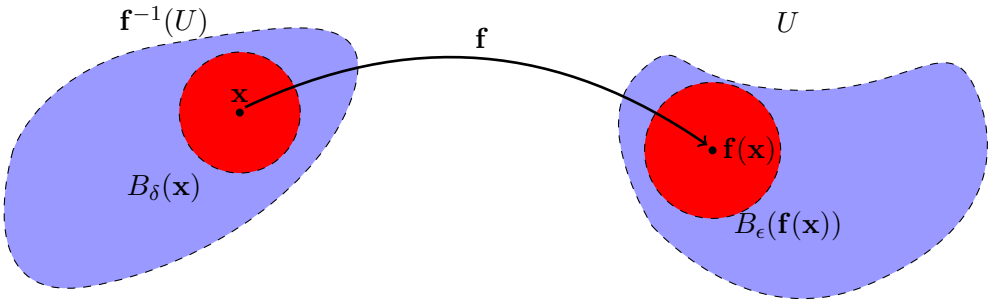


Figure 15: To show that the pre-image of open sets is open, we use the fact that the condition $\|\mathbf{f}(\mathbf{x}) - \mathbf{f}(\mathbf{y})\| < \epsilon$ is exactly the same thing as looking at an ϵ-ball around $\mathbf{f}(\mathbf{x})$.

[$\Leftarrow$] Assume that the preimage of open sets is open, for which we want to show that $\mathbf{f}$ is continuous, say at $\mathbf{x}$. Let $\epsilon > 0$ be given, and set $U = B_\epsilon(\mathbf{f}(x))$. Certainly we have $x \in \mathbf{f}^{-1}(U)$, and since this is an open set by assumption, there exists a $\delta > 0$ such that $B_\delta(\mathbf{x}) \subseteq \mathbf{f}^{-1}(U)$. We claim that this choice of delta will satisfy the continuity requirement. Indeed, let $\mathbf{y}$ be a point such that $\|\mathbf{x} - \mathbf{y}\| < \delta$; that is, $\mathbf{y} \in B_\delta(\mathbf{x})$. Since $B_\delta(\mathbf{x}) \subseteq \mathbf{f}^{-1}(U)$ we know that $f(\mathbf{y}) \in f(B_\delta(\mathbf{x})) \subseteq U = B_\epsilon(\mathbf{f}(\mathbf{x}))$; that is, $\|\mathbf{f}(\mathbf{y}) - \mathbf{f}(\mathbf{x})\| < \epsilon$, as required. □

Example 1.63

Show that the set $S = \{(x, y) : y > 0\} \subseteq \mathbb{R}^2$ is open.

Solution. This is the same set as in Example 1.24, wherein we showed that S was open by constructing an open ball around every point. Consider the function $f : \mathbb{R}^2 \to \mathbb{R}$ given by $f(x, y) = y$. The student can convince him/herself that this function is continuous, and moreover, that $S = f^{-1}((0, \infty))$. Since $(0, \infty)$ is open in $\mathbb{R}$ and f is continuous, it follows that S is open as well. ■

Exercise: Show that $\mathbf{f} : \mathbb{R}^n \to \mathbb{R}^m$ is continuous if and only if whenever $V \subseteq \mathbb{R}^m$ is closed then $\mathbf{f}^{-1}(V)$ is closed.

Many of the theorems about continuous functions in a single variable carry over to multiple dimensions, for example

Theorem 1.64

If $\mathbf{f} : \mathbb{R}^n \to \mathbb{R}^m$ is continuous at $\mathbf{c}$ and $\mathbf{g} : \mathbb{R}^m \to \mathbb{R}^k$ is continuous at $\mathbf{f}(\mathbf{c})$, then $\mathbf{g} \circ \mathbf{f} : \mathbb{R}^n \to \mathbb{R}^k$ is continuous at $\mathbf{c}$.

This is a simple theorem to prove, so it is left as an exercise for the student. It is straightforward to show that the functions $x \pm y$, and xy are continuous, which immediately imply that the sum and product of continuous functions is also continuous.

1.6 Compactness

In our study of calculus on $\mathbb{R}$, there is a very real sense in which the sets $[a, b]$ are the best behaved: They are closed, which means we need to not worry about the distinction between infimum/supremum and minimum/maximum, and they are bounded so need not worry about wandering off to infinity. In fact, one might recall that the Extreme Value Theorem was stated for an interval of this type.

We have since explored the notions of closed and boundedness in multiple dimensions, and once again it seems as though the same benefits afforded in the single variable case also extend to $\mathbb{R}^n$. We give such sets a very special name:

Definition 1.65

A set $S \subseteq \mathbb{R}^n$ is *compact* if it is both closed and bounded.

Example 1.66

1. As mentioned, the interval $[a, b] \subseteq \mathbb{R}$ is compact. More generally, any closed ball $\overline{B_r(\mathbf{x})} \subseteq \mathbb{R}^n$ is compact.
2. As finite unions of closed and bounded sets are closed and bounded, the finite union of compacts sets is compact.
3. The set consisting of a single point is compact. By the previous example, every finite set is also compact.
4. $\mathbb{R}^n$ is *not* compact. While it is closed, it is certainly not bounded.
5. The rationals $\mathbb{Q} \subseteq \mathbb{R}$ are neither closed nor bounded, and hence are not compact.

Exercise: Prove property (2); that is, show that a finite union of compact sets is still compact. Give an example to show that the result is not true if we allow infinite unions.

It turns out that this definition, while convenient conceptually, does not lend itself to proving results about compact sets. Nor does it generalize to more abstract spaces. As such, we have the following two equivalent definitions of compactness:

Theorem 1.67

If $S \subseteq \mathbb{R}^n$ then the following are equivalent:

1. S is compact,
2. [**Bolzano-Weierstrass**] Every sequence in S has a convergent subsequence; that is, if $(\mathbf{x}_n)_{n=1}^{\infty} \subseteq S$, then there exists a subsequence $(\mathbf{x}_{k_n})$ and a point $\mathbf{x} \in S$ such that $(\mathbf{x}_{k_n}) \to \mathbf{x}$.
3. [**Heine-Borel**] Every open cover of S admits a finite subcover; that is, if $\{U_i\}_{i \in I}$ is a collection of open sets such that $S \subseteq \bigcup_{i \in I} U_i$, then there exists a finite subset $J \subseteq I$ such that $S \subseteq \bigcup_{i \in J} U_i$.

Proof. This is typically stated as two separate theorems: The Heine-Borel Theorem and the Bolzano-Weierstrass Theorem, in which one shows that each of the corresponding alternate definitions are equivalent to closed and bounded. We will only prove Bolzano-Weierstrass, as Heine-Borel is more complicated.

[(1) $\Rightarrow$ (2)] Suppose that S is closed and bounded, and let $(\mathbf{x}_n)_{n=1}^{\infty} \subseteq S$. Since S is bounded, so too is $(\mathbf{x}_n)$, in which case Theorem 1.44 implies there exists a convergent subsequence $(\mathbf{x}_{n_k}) \to \mathbf{x}$. *A priori*, we only know that $\mathbf{x} \in \mathbb{R}^n$, but since S is closed, by Corollary 1.42 we know that $\mathbf{x} \in S$. Thus $(\mathbf{x}_{n_k})$ is a convergent subsequence.

[(2) $\Rightarrow$ (1)] We will proceed by contrapositive. Assume therefore that S is either not closed or not bounded.

If S is not closed, there exists $\mathbf{x} \in \overline{S} \setminus S$. By Corollary 1.42 there exists a sequence $(\mathbf{x}_n)_{n=1}^{\infty} \subseteq S$ such that $(\mathbf{x}_n) \to \mathbf{x}$. Since $(\mathbf{x}_n)$ converges, by Proposition 1.37 every subsequence also converges, and to the same limit point. Thus $(\mathbf{x}_n)$ is a sequence in S with no convergent subsequence in S.

Now assume that S is not bounded. One can easily construct a sequence $(\mathbf{x}_n)$ such that $\|\mathbf{x}_n\| \xrightarrow{n \to \infty} \infty$. Necessarily, any subsequence of $\mathbf{x}_n$ also satisfies this property, and so $(\mathbf{x}_n)$ has no convergent subsequence. □

Remark 1.68 There are many more equivalent definitions of compactness, some of which are equivalent, depending on the more general topological context. In general, none of these definitions are actually equivalent. The statement corresponding to Heine-Borel is the "true" definition of compactness, though it is sometimes known as quasi-compactness, while the Bolzano-Weierstrass definition is referred to as *sequential compactness*.

One of the more potent results about compact sets is the following theorem

Theorem 1.69

Let $\mathbf{f} : \mathbb{R}^n \to \mathbb{R}^m$ be a continuous function. If $K \subseteq \mathbb{R}^n$ is compact, then $\mathbf{f}(K)$ is also compact. More concisely, the continuous image of compact sets is compact.

Proof. We will proceed via the Bolzano-Weierstrass theorem. Consider a sequence $(\mathbf{y}_n)_{n=1}^{\infty}$ in $\mathbf{f}(K)$, for which our intent is to find a convergent subsequence. By definition of $\mathbf{f}(K)$, for each $\mathbf{y}_n$ there exists an $\mathbf{x}_n \in K$ such that $\mathbf{y}_n = \mathbf{f}(\mathbf{x}_n)$, and hence we can define a sequence $(\mathbf{x}_n)$ in K. Since K is compact, there exists a convergent subsequence $(\mathbf{x}_{n_k}) \to \mathbf{x}$, with $\mathbf{x} \in K$.

We claim that the corresponding subsequence $(\mathbf{y}_{n_k})$ converges to $\mathbf{f}(\mathbf{x})$. Indeed, since $\mathbf{f}$ is continuous, we know that

$$\lim_{k\to\infty} \mathbf{f}(\mathbf{x}_{n_k}) = \mathbf{f}\left(\lim_{k\to\infty} \mathbf{x}_{n_k}\right) = \mathbf{f}(\mathbf{x})$$

and since $\mathbf{x} \in K$, we know $\mathbf{f}(\mathbf{x}) \in \mathbf{f}(K)$. Thus $(\mathbf{y}_{n_k})$ is a convergent subsequence in $\mathbf{f}(K)$, and we conclude that $\mathbf{f}(K)$ is compact. □

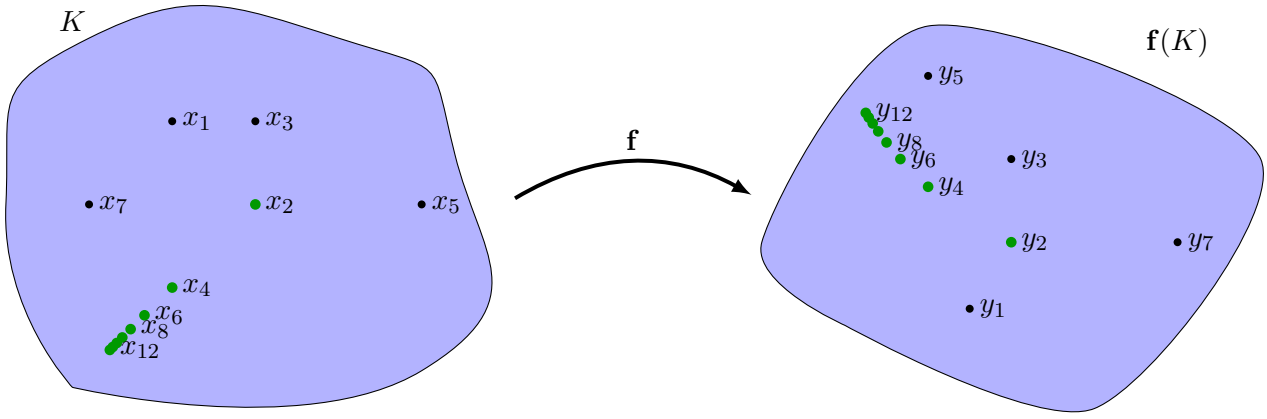


Figure 16: Proof of Theorem 1.69. We start with a random sequence (y_n) in $\mathbf{f}(K)$ and look at the points (x_n) in K which map to (y_n). We choose a convergent subsequence (x_{n_k}) (green), and use that subsequence to define a convergent sequence (y_{n_k}) (green) in $\mathbf{f}(K)$.

Exercise: Prove Theorem 1.69 using the Heine-Borel Theorem.

This now immediately implies a familiar theorem from single variable calculus:

Corollary 1.70: Extreme Value Theorem

If $f : \mathbb{R}^n \to \mathbb{R}$ is a continuous function and $K \subseteq \mathbb{R}^n$ is a compact set, then there exists $\mathbf{x}_{\min}, \mathbf{x}_{\max} \in K$ such that for every $\mathbf{x} \in K, f(\mathbf{x}_{\min}) \leq f(\mathbf{x}) \leq f(\mathbf{x}_{\max})$; that is, f achieves both its extreme values on K.

Proof. Since f is continuous and K is compact, by Theorem 1.69 we know that $f(K)$ is compact, and as such is both closed and bounded. Since $f(K) \subseteq \mathbb{R}$, the completeness axiom implies that $\sup f(K)$ and $\inf f(K)$ both exist. Since $f(K)$ is closed, the supremum and infimum are actually in $f(K)$, so there exist $\mathbf{x}_{\min}, \mathbf{x}_{\max} \in K$ such that

$$f(\mathbf{x}_{\min}) = \inf f(K), \qquad f(\mathbf{x}_{\max}) = \sup f(K),$$

and by definition of inf and sup, for every $\mathbf{x} \in K$

$$f(\mathbf{x}_{\min}) = \inf f(K) \leq f(\mathbf{x}) \leq \sup f(K) = f(\mathbf{x}_{\max})$$

as required. □

1.7 Connectedness

Connectedness is an expansive and important topic, but one which is also quite subtle. The "true definition" embodies pathological cases which we will not be of concern in the majority of our work, and so it is more intuitive to introduce a weaker notion known as path connectedness.

Intuitively, we would like something to be connected if it cannot be decomposed into two separate pieces. Hence we might say that a set S is not connected if there exist S_1, S_2 such that $S = S_1 \cup S_2$ and $S_1 \cap S_2 = \emptyset$. This latter condition is important to guarantee that the two sets do not overlap. Unfortunately, this condition does not actually capture the idea we are trying convey.

For example, one expects that the interval $S = (0, 2)$ should be connected: it looks like all one piece. Nonetheless, we can write $(0, 2) = (0, 1) \cup [1, 2)$, so that if $S_1 = (0, 1)$ and $S_2 = [1, 2)$ then $S = S_1 \cup S_2$ and $S_1 \cap S_2 = \emptyset$.

The remedy is to enforce a condition on the closure of each set; namely, that $\overline{S}_1 \cap S_2 = \emptyset$ and $S_1 \cap \overline{S}_2 = \emptyset$. Ensuring that these intersections are empty ensures that our sets are far enough apart.

Definition 1.71

A set $S \subseteq \mathbb{R}^n$ is said to be *disconnected* if there exist non-empty $S_1, S_2 \subseteq S$ such that

1. $S = S_1 \cup S_2$,
2. $\overline{S}_1 \cap S_2 = \emptyset$ and $S_1 \cap \overline{S_2} = \emptyset$.

We refer to (S_1, S_2) as a *disconnection* of S. If S admits no disconnection, we say that S is *connected*.

This definition is such that it is much easier to show that a set is disconnected rather than connected, since to show that a set is connected we must then show that there is no disconnection amongst all possible candidates.

Example 1.72

Show that the following sets are not connected:

1. $S = [0,1] \cup [2,3] \subseteq \mathbb{R}$,
2. $\mathbb{Q} \subseteq \mathbb{R}$,
3. $T = \{(x,y) \in \mathbb{R}^2 : y \neq x\}$.

Solution.

1. The disconnection for this case is evident: by setting $S_1 = [0,1]$ and $S_2 = [2,3]$, all conditions are satisfied. Hence (S_1, S_2) is a disconnection of S.
2. This example requires us to think more carefully. We know that $\pi \in \mathbb{Q}$ is irrational, so consider $S_1 = \mathbb{Q} \cap (-\infty, \pi)$ and $S_2 = \mathbb{Q} \cap (\pi, \infty)$. Clearly $S_1 \cup S_2 = \mathbb{Q} \cap (\mathbb{R} \setminus \{\pi\}) = \mathbb{Q}$, while
$$\overline{S_1} \cap S_2 = (-\infty, \pi] \cap (\pi, \infty) = \emptyset.$$
Similarly, $S_1 \cap \overline{S_2} = \emptyset$. Thus (S_1, S_2) does indeed form a disconnection of $\mathbb{Q}$. ■
3. Our set T looks like the plane with the line $y = x$ removed. Since the line $y = x$ somehow splits the space, one might be unsurprised that this set is disconnected. Let $S_1 = \{(x,y) : y > x\}$ and $S_2 = \{(x,y) : y < x\}$, so that $T = S_1 \cup S_2$. Furthermore,
$$\overline{S_1} \cap S_2 = \{(x,y) : y \geq x\} \cap \{(x,y) : y < x\} = \emptyset.$$
Similarly, $S_1 \cap \overline{S_2} = \emptyset$, and so (S_1, S_2) is a disconnection of T.

Remark 1.73 Examples (2) and (3) above show that the elements of the disconnection can be arbitrarily close to one another yet still form a disconnection.

Proposition 1.74

A set $S \subseteq \mathbb{R}$ is connected if and only if S is an interval.

Despite the simplicity of the statement, the proof of this result is non-trivial. It can be found in the textbook, so we leave it as an exercise for the interested student.

So in general, it seems that proving that a set is connected can prove quite bothersome. An excellent tool for proving connectedness will be the following weaker definition:

Definition 1.75

If $S \subseteq \mathbb{R}^n$ then a *path in* S is any continuous map $\gamma : [0,1] \to S$. We say that S is *path-connected* if for every two points $\mathbf{a}, \mathbf{b} \in S$ there exists a path $\gamma : [0,1] \to S$ such that $\gamma(0) = \mathbf{a}$ and $\gamma(1) = \mathbf{b}$.

Intuitively, a set is path connected if between any two points in our set, we can draw a line between those two points which never leaves the set.

Example 1.76

Show that every interval $[a, b] \subseteq \mathbb{R}$ is path connected.

Solution. Let $c, d \in [a, b]$ be arbitrary, and define the map $\gamma : [0, 1] \to [a, b]$ by $\gamma(t) = td + (1 - t)c$. One can easily check that γ is continuous, and $\gamma(0) = c$, $\gamma(1) = d$. We conclude that $[a, b]$ is path connected. ■

Naturally, the solution above would also work for (half) open intervals. We invite the student to compare this proof to the one for Proposition 1.74, to see the difference in complexity required to show that a set is connected as compared to path connected.

Example 1.77

Show that the set $S = \{(x, y) \in \mathbb{R}^2 : x \neq 0\} \cup \{(0, 0)\}$ is path-connected.

Solution. Consider Figure 17 which suggests how we might proceed. If the two components lie in the same half of the plane, we can connected them with a straight line. If they lie in separate halves of the plane, we can connected them with lines that must first go through the origin.

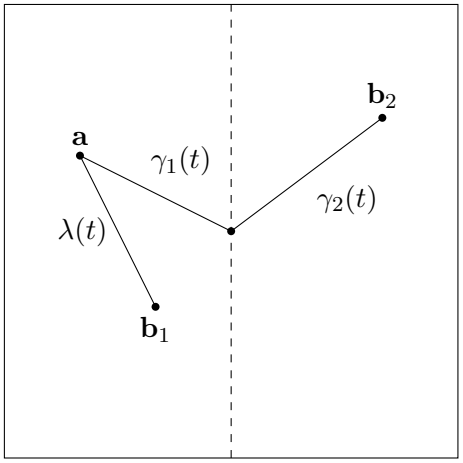


Figure 17: If $\mathbf{a}$ and $\mathbf{b}_1$ lie in the same plane, we can connect them with a straight line. If $\mathbf{a}$ and $\mathbf{b}_2$ lie in separate planes, we can connect them with a line through the origin.

Choose two points $\mathbf{a} = (a_1, a_2), \mathbf{b} = (b_1, b_2) \in S$. Our first case will be to assume that both $\mathbf{a}$ and $\mathbf{b}$ lie in the same half of the plane. Without loss of generality, assume that $a_1, b_1 > 0$. Define the path

$$\lambda(t) = \mathbf{a}t + (1 - t)\mathbf{b} = (a_1t + (1 - t)b_1, a_2t + (1 - t)b_2).$$

Since a_1 and b_1 are both positive, the x-coordinate of the path $a_2t + (1-t)b_2$ is also always positive. Thus λ is a path entirely in S.

For our other case, assume then that $a_1 < 0$ and $b_1 > 0$. Consider the two paths $\gamma_1(t) = \mathbf{a}(1-t)$ and $\gamma_2(t) = \mathbf{b}t$, both of which are paths from their respective points to the origin, which remain entirely within S. By concatenating these paths, we can define a new path

$$\gamma(t) = \begin{cases} \gamma_1(2t) & t \in [0, 1/2] \\ \gamma_2(2t-1) & t \in [1/2, 1] \end{cases}.$$

It is easy to check that γ is continuous, $\gamma(0) = \mathbf{a}$ and $\gamma(1) = \mathbf{b}$. As each constituent path lies entirely within S, so too does the concatenated path, as required. We conclude that S is path connected. ■

Theorem 1.78

The continuous image of a (path) connected set is (path) connected. More precisely, if $\mathbf{f} : \mathbb{R}^n \to \mathbb{R}^m$ is continuous and $S \subseteq \mathbb{R}^n$ is (path) connected, then $\mathbf{f}(S)$ is (path) connected.

Proof. We will show the (much simpler) proof when S is path connected, and leave the connected case as an exercise.

Assume then that S is path connected, and consider $\mathbf{f}(S)$. Let $\mathbf{a}, \mathbf{b} \in \mathbf{f}(S)$ be any two points, and choose $\mathbf{x}, \mathbf{y} \in S$ such that $\mathbf{f}(\mathbf{x}) = \mathbf{a}$ and $\mathbf{f}(\mathbf{y}) = \mathbf{b}$. Since S is path connected, there is a path $\gamma : [0, 1] \to S$ such that $\gamma(0) = \mathbf{x}$ and $\gamma(1) = \mathbf{y}$. We claim that $\mathbf{f} \circ \gamma : [0, 1] \to \mathbf{f}(S)$ is a path in $\mathbf{f}(S)$ connecting $\mathbf{a}$ and $\mathbf{b}$. Indeed, since γ and $\mathbf{f}$ are both continuous their composition $\mathbf{f} \circ \gamma$ is also continuous. Evaluating the endpoints, we have

$$(\mathbf{f} \circ \gamma)(0) = \mathbf{f}(\gamma(0)) = \mathbf{f}(\mathbf{x}) = \mathbf{a}, \qquad (\mathbf{f} \circ \gamma)(1) = \mathbf{f}(\gamma(1)) = \mathbf{f}(\mathbf{y}) = \mathbf{b}$$

so $(\mathbf{f} \circ \gamma)$ is a path connecting $\mathbf{a}$ and $\mathbf{b}$ as required. □

Corollary 1.79: Intermediate Value Theorem

Let $V \subseteq \mathbb{R}^n$ be a (path) connected set and $\mathbf{f} : \mathbb{R}^n \to \mathbb{R}$ be a continuous function. Let $\mathbf{a}, \mathbf{b} \in V$ and assume that $f(\mathbf{a}) < f(\mathbf{b})$. Then for every c such that $\mathbf{f}(a) < c < \mathbf{f}(b)$ there exists an $\mathbf{x} \in V$ such that $f(\mathbf{x}) = c$.

Proof. Regardless of whether we allow V to be connected or path connected, we know that the image $f(V)$ is an interval. Since $f(\mathbf{a}), f(\mathbf{b}) \in f(V)$ then $[f(\mathbf{a}), f(\mathbf{b})] \subseteq f(V)$, and the result follows. □

I have mentioned that path connectedness is a strictly weaker notion of connectedness; that is, any path connected set is necessarily connected, but the converse need not be true. This is demonstrated by the following proposition and the example thereafter.

Proposition 1.80

Any set which is path connected is also connected.

Proof. We will proceed by contradiction. Assume then that $S \subseteq \mathbb{R}^n$ is path connected but not connected, so there exists a disconnection (S_1, S_2). Choose $\mathbf{a} \in S_1$ and $\mathbf{b} \in S_2$ and let $\gamma : [0,1] \to S$ be a path from $\mathbf{a}$ to $\mathbf{b}$. Since γ is continuous, $P = \gamma([0,1])$ is necessarily connected. On the other hand, let $P_1 = P \cap S_1$ and $P_2 = P \cap S_2$, so that $P_1 \cup P_2 = (S_1 \cup S_2) \cap P = P$, while

$$\overline{P_1} \cap P_2 = (\overline{P} \cap \overline{S}_1) \cap S_2 \subseteq \overline{S}_1 \cap S_2 = \emptyset,$$

implies that $\overline{P_1} \cap P_2 = \emptyset$. Similarly, $P_1 \cap \overline{P_2} = \emptyset$, showing that (P_1, P_2) is a disconnection of P, which is a contradiction. We conclude that S is connected. □

To see that connected does not imply path connected, consider the following set, known as the *Topologist's Sine Curve*:

$$\left\{ \left(x, \sin\left(\frac{1}{x}\right) \right) : x \in \mathbb{R} \setminus \{0\} \right\} \cup (0,0).$$

It is possible to show that this set is connected (convince yourself of this) but not path connected (also convince yourself of this). Thus path connectedness is not equivalent to connectedness. A partial converse is given by the following:

Proposition 1.81

If $S \subseteq \mathbb{R}^n$ is connected and open, then S is path-connected.

1.8 Uniform Continuity

Stronger than continuity, there is a notion of uniform continuity, which plays nicer with Cauchy sequences than a simple continuous function. The idea is as follows: If we write out the ϵ-δ definition of continuity, in full quantifiers, we get

$$\forall \epsilon > 0, \forall \mathbf{x} \in D, \exists \delta > 0, \forall y \in D, |\mathbf{x} - \mathbf{y}| < \delta \Rightarrow |\mathbf{f}(\mathbf{x}) - \mathbf{f}(\mathbf{y})| < \epsilon.$$

The fact that the delta is specified *after* both the ϵ and the point $\mathbf{x}$ means that $\delta(\epsilon, \mathbf{x})$ is a function of both these terms; that is, changing either ϵ or the point $\mathbf{x}$ will change the necessary value of δ. This is perhaps unsurprising, since the choice of δ really corresponds to how quickly the function is growing at a point (See Figure 18).

The idea of uniform continuity is that given a fixed $\epsilon > 0$, one can find a δ which works *for every* point $\mathbf{x}$.

Definition 1.82

Let $D \subseteq \mathbb{R}^n$ and $\mathbf{f} : D \to \mathbb{R}^m$. We say that f is *uniformly continuous* if for every $\epsilon > 0$, there exists a $\delta > 0$ such that **for every** $x, y \in D$ satisfying $\|\mathbf{x} - \mathbf{y}\| < \delta$ then $\|\mathbf{f}(\mathbf{x}) - \mathbf{f}(\mathbf{y})\| < \epsilon$.

As stated, the definition of uniform continuity implies that δ only depends upon the choice of ϵ, not on the particular point that we choose. Intuitively, uniformly continuous function are in some sense bounded in how quickly they are permitted to grow.

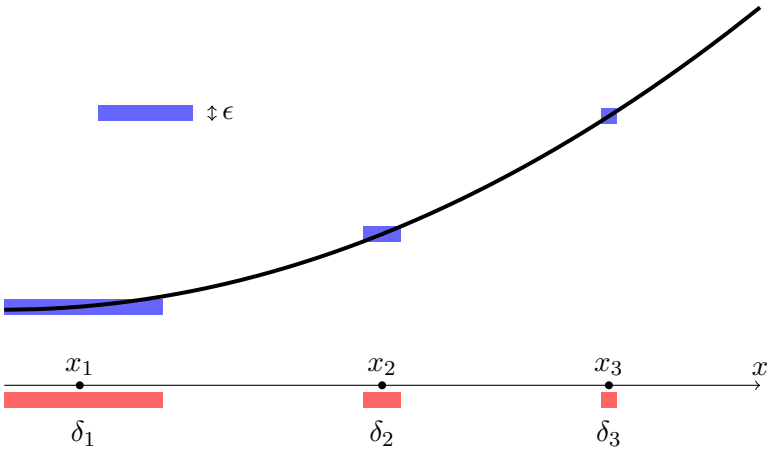


Figure 18: For a fixed $\epsilon > 0$, the value of δ depends on the choice of the point x. In fact, the faster a function grows at a point, the smaller the corresponding δ will be.

Example 1.83

The function $f : \mathbb{R} \to \mathbb{R}$ given by $f(x) = 2x + 5$ is uniformly continuous.

Solution. Let $\epsilon > 0$ and choose $\delta = \frac{\epsilon}{2}$. Let $x, y \in \mathbb{R}$ be any points such that $|x - y| < \delta$, and notice that

$$|f(x) - f(y)| = |(2x + 5) - (2y + 5)| = 2|x - y| < 2\delta = \epsilon.$$ ■

The domain is an exceptionally important piece of information when determining uniform continuity, as the following example shows.

Example 1.84

Let $f : \mathbb{R} \to \mathbb{R}, x \mapsto x^2$ and $g : [-2, 2] \to \mathbb{R}, x \mapsto x^2$. Show that g is uniformly continuous but f is not uniformly continuous.

Solution. Let $\epsilon > 0$ and choose $\delta = \frac{\epsilon}{4}$. Let $x, y \in \mathbb{R}$ be such that $|x - y| < \delta$. Since $x, y \in [-2, 2]$ we know that $-2 < x, y < 2$ so

$$|x + y| < |x| + |y| < 2 + 2 = 4.$$

and moreover

$$|f(x) - f(y)| = |x^2 - y^2| = |x + y||x - y| < 4|x - y| < 4\delta = \epsilon$$

as required. On the other hand, suppose for the sake of contradiction that f is uniformly continuous. Let $\epsilon = 1$ and choose the $\delta > 0$ guaranteed by uniform continuity. Choose $x \in \mathbb{R}$ such that $|x| > 1/\delta$, and set $y = x + \delta/2$. Clearly $|x - y| < \delta$, but

$$\left| x^2 - \left(x + \frac{\delta}{2} \right)^2 \right| = \left| \delta x + \delta^2 \right| \geq \delta |x| > 1 = \epsilon$$

which is a contradiction. ■

Notice that the proof for why f fails to be uniformly continuous cannot be applied to g, precisely because g is only defined on the interval $[-2, 2]$ and as such, we cannot guarantee there exists an x such that $|x| > 1/\delta$.

So far, our examples have been limited to those function $\mathbb{R} \to \mathbb{R}$, and naturally the situation becomes more complicated in higher dimensions. Luckily, with the use of compactness, we can prove the following theorem:

Theorem 1.85

If $D \subseteq \mathbb{R}^n$ is a compact set and $\mathbf{f} : D \to \mathbb{R}^m$ is continuous, then $\mathbf{f}$ is uniformly continuous. More concisely, continuous functions with compact domain are uniformly continuous.

Proof. The proof of this theorem is particularly slick using the Heine-Borel theorem, but as our characterization of compactness has been principally the Bolzano-Weierstrass theorem, we will proceed with that.

Assume, for the sake of contradiction, that $\mathbf{f}$ is not uniformly continuous; that is, there exists an $\epsilon > 0$ such that for all $\delta > 0$ we can find a pair $\mathbf{x}, \mathbf{y} \in D$ such that $\|\mathbf{x} - \mathbf{y}\| < \delta$ and $\|\mathbf{f}(\mathbf{x}) - \mathbf{f}(\mathbf{y})\| \geq \epsilon$. Let $\epsilon > 0$ be as given, and for each $n \in \mathbb{N}$ let $\delta_n = \frac{1}{n}$. Define $\mathbf{x}_n, \mathbf{y}_n$ to be the pair such that $\|\mathbf{x}_n - \mathbf{y}_n\| < \delta_n$ and $\|\mathbf{f}(\mathbf{x}_n) - \mathbf{f}(\mathbf{y}_n)\| \geq \epsilon$.

The sequence $(\mathbf{x}_n)_{n=1}^{\infty}$ is a sequence in the compact set D, and so by Bolzano-Weierstrass, it has a convergent subsequence $(\mathbf{x}_{n_k}) \to \mathbf{x}$. Since $\|\mathbf{x}_n - \mathbf{y}_n\| \xrightarrow{n\to\infty} 0$, one can show that $(\mathbf{y}_{n_k}) \to \mathbf{x}$ as well. Since f is continuous,

$$\lim_{k\to\infty} [\mathbf{f}(\mathbf{x}_{n_k}) - \mathbf{f}(\mathbf{y}_{n_k})] = \mathbf{f}(\mathbf{x}) - \mathbf{f}(\mathbf{x}) = 0$$

which contradicts the fact that $\|\mathbf{f}(\mathbf{x}_n) - \mathbf{f}(\mathbf{y}_n)\| \geq \epsilon$. □

Exercise: Prove the above theorem using the open covering version of compactness.

This allows us to immediately deduce that some functions are uniformly continuous, without having to go through the trouble of proving the ϵ-δ version. For example, the function $f(x, y) = \sqrt{\sin(x) + \cos^2(y)}$, defined on $\overline{B_1(0)}$ is uniformly continuous by virtue of the fact that f is continuous and $\overline{B_1(0)}$ is compact.

2 Differential Calculus

2.1 Derivatives

2.1.1 Single Variable: $\mathbb{R} \to \mathbb{R}$

Recall that if $f : \mathbb{R} \to \mathbb{R}$ we say that f is differentiable at a point $a \in \mathbb{R}$ if

$$\lim_{h \to 0} \frac{f(a+h) - f(a)}{h} \text{ exists,}$$

and moreover, when the limit exists we denote it by $f'(a)$.

It turns out that this paradigm may not be the most pragmatic when dealing with functions from $\mathbb{R}^n \to \mathbb{R}^m$, so we take a moment to re-introduce the topic with a view which is more amenable to our future discussion.

The idea is that a function f is differentiable at a if it can be well-approximated by a linear function sufficiently close to a. In particular, if h is sufficiently small, one would hope that there exists an m such that

$$f(a+h) = f(a) + mh + \text{error}(h) \tag{2.1}$$

where $\text{error}(h)$ is the corresponding error in the linear approximation. For the approximation to be good, the error should go to zero faster than linearly in h; that is,

$$\lim_{h \to 0} \frac{\text{error}(h)}{h} = 0.$$

This leads us to the following equivalent definition of differentiability:

Definition 2.1

A function $f : \mathbb{R} \to \mathbb{R}$ is differentiable at $a \in \mathbb{R}$ if there exists an $m \in \mathbb{R}$ such that

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

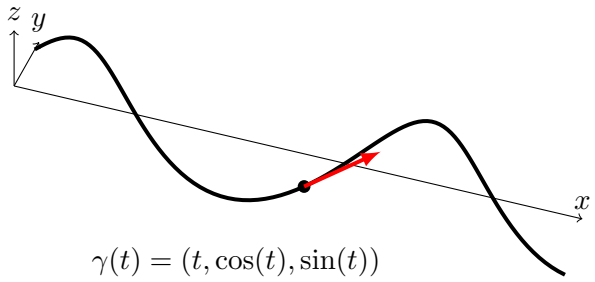
One can manipulate (2.1) to show that $m = f'(a)$ under the usual definition. Of course, everything we know about single variable calculus is still true: The product rule, the chain rule, our theorems regarding differentiability. We will not replicate the list here, for it is too large and the student should be well familiar with it.

2.1.2 Vector Valued: $\mathbb{R} \to \mathbb{R}^n$

The first and simplest generalization of the derivative comes from looking at vector valued functions $\gamma : \mathbb{R} \to \mathbb{R}^n$. Such function are often visualized as parameterized paths in $\mathbb{R}^n$.

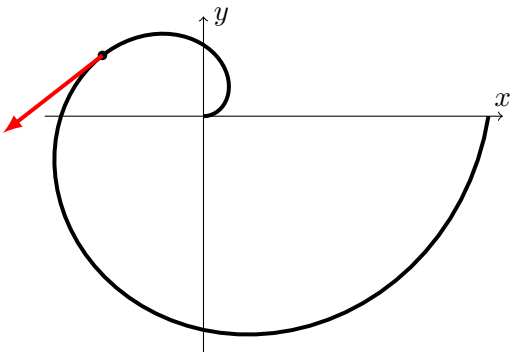
Example 2.2

1. Consider the function $\gamma_1 : [0, 2\pi) \to \mathbb{R}^2, t \mapsto (\cos(t), \sin(t))$. By plotting the values of the function for $t \in [0, 2\pi)$, we see that γ_1 traces out the unit circle in $\mathbb{R}^2$.
2. The map $\gamma_2 : (0, \infty) \to \mathbb{R}^2$ given by $\gamma_2(t) = (t\cos(t), t\sin(t))$ is a spiral (see Figure 19b).
3. The function $\gamma_3 : \mathbb{R} \to \mathbb{R}^3$ given by $\gamma_3(t) = (t, \cos(t), \sin(t))$ is a Helix (see Figure 19a).



$\gamma(t) = (t, \cos(t), \sin(t))$

(a)



$\gamma(t) = (t\cos(t), t\sin(t))$

(b)

Figure 19: Examples of parameterized curves and their derivatives. Left: A helix in $\mathbb{R}^3$. Right: A spiral in $\mathbb{R}^2$.

Definition 2.3

We say that a function $\gamma : \mathbb{R} \to \mathbb{R}^n$ is *differentiable* if at t_0 if

$$\begin{aligned}\gamma'(t_0) &= \lim_{h\to 0} \frac{\gamma(t_0+h) - \gamma(t_0)}{h} \\ &= \left(\lim_{h\to 0} \frac{\gamma_1(t_0+h) - \gamma_1(t_0)}{h}, \ldots, \lim_{h\to 0} \frac{\gamma_n(t_0+h) - \gamma_n(t_0)}{h}\right)\end{aligned}$$

exists. If γ is differentiable at every point in its domain, we will say that γ is differentiable.

Thus a vector-value function of a single variable is differentiable precisely when each of its component functions is differentiable, and the derivative may be computed by differentiating each component separately. For example, we can immediately deduce that the curve $\gamma(t) = (e^t, \cos(t^2), (t^2+1)^{-1})$ is differentiable everywhere, since each of its component functions are differentiable everywhere, and moreover its derivative is given by

$$\gamma'(t) = \left(e^t, -2t\sin(t^2), \frac{-2t}{(1+t^2)^2}\right).$$

Similarly, every curve given in Example 2.2 is differentiable.

Example 2.4

Determine the derivatives of each curve given in Example 2.2.

Solution. In every case we need only read off the derivatives of each component:

$$\begin{aligned}\gamma_1'(t) &= (-\sin(t), \cos(t))\\ \gamma_2'(t) &= (\cos(t) - t\sin(t), \sin(t) + t\cos(t))\\ \gamma_3'(t) &= (1, -\sin(t), \cos(t)).\end{aligned}$$

■

In the context of $\gamma : \mathbb{R} \to \mathbb{R}^n$ parameterizing a curve in $\mathbb{R}^n$, its derivative $\gamma'(t_0)$ represents the instantaneous velocity of the curve at that point (both the speed at the direction). The corresponding vector is tangent to the curve. For example, see Figure-19.

Proposition 2.5

Let $\mathbf{f}, \mathbf{g} : \mathbb{R} \to \mathbb{R}^n$ and $\varphi : \mathbb{R} \to \mathbb{R}$ be differentiable functions.

1. $(\varphi\mathbf{f})' = \varphi'\mathbf{f} + \varphi\mathbf{f}'$,
2. $(\mathbf{f} \cdot \mathbf{g})' = \mathbf{f}' \cdot \mathbf{g} + \mathbf{f} \cdot \mathbf{g}'$,
3. $(\mathbf{f} \times \mathbf{g})' = \mathbf{f}' \times \mathbf{g} + \mathbf{f} \times \mathbf{g}'$ (if $n = 3$).

In particular, since the cross-product is not-commutative, the order of $\mathbf{f}$ and $\mathbf{g}$ matters.

Proof. We will do the proof for (2) and leave the others as an exercise for the students. Let $\mathbf{f}(t) = (f_1(t), \ldots, f_n(t))$ and $\mathbf{g}(t) = (g_1(t), \ldots, g_n(t)$. Differentiating their dot product yields

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}(\mathbf{f}(t) \cdot \mathbf{g}(t)) &= \frac{\mathrm{d}}{\mathrm{d}t}(f_1(t)g_1(t) + \cdots + f_n(t)g_n(t))\\ &= \left[f_1'(t)g_1(t) + f_1(t)g_1'(t)\right] + \cdots + \left[f_n'(t)g_n(t) + f_n(t)g_n'(t)\right]\\ &= \left[f_1'(t)g_1(t) + f_2'(t)g_2(t) + \cdots + f_n'(t)g_n(t)\right]\\ &\qquad + \left[f_1(t)g_1'(t) + f_2(t)g_2'(t) + \cdots + f_n(t)g_n'(t)\right]\\ &= \mathbf{f}'(t) \cdot \mathbf{g}(t) + \mathbf{f}(t) \cdot \mathbf{g}'(t).\end{aligned}$$

□

2.1.3 Multivariable $\mathbb{R}^n \to \mathbb{R}$

The previous section represented the simplest generalization of the derivative to multiple dimensions. In this section, we will now examine what happens when our function takes in multiple parameters. This situation is *significantly* more complicated: In the previous two sections, only a single variable existed to be differentiated. Now that our functions have multiple parameters, making sense of how to meaningfully define a derivative becomes its own challenge.

Let $f : \mathbb{R}^n \to \mathbb{R}$ be a function. We can visualize this function by thinking about its graph,

$$\Gamma(f) = \{(\mathbf{x}, f(\mathbf{x})) : \mathbf{x} \in \mathbb{R}^n\} \subseteq \mathbb{R}^{n+1}.$$

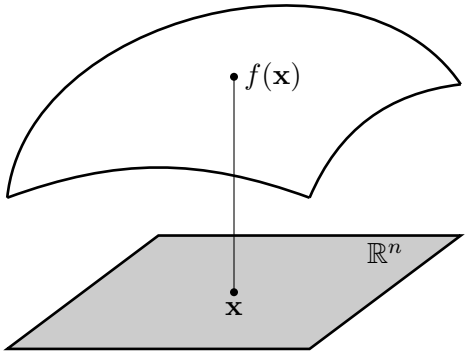


Figure 20: A function $f : \mathbb{R}^n \to \mathbb{R}$ can be visualized in terms of its graph.

as illustrated in figure 20. In this case, what does it mean to behave linearly? The correct notion of a linear object in $\mathbb{R}^{n+1}$ is that of an n-plane. An n-plane through the origin has the equation

$$c_1x_1 + c_2x_2 + \cdots + c_nx_n + c_{n+1}x_{n+1} = \mathbf{c} \cdot \mathbf{x} = 0.$$

In the case where $n = 1$ then this reduces to $c_1x_1 + c_2x_2 = 0$, which we recognize as a straight line through the origin. If we instead would like this plane to pass through a point $\mathbf{a} \in \mathbb{R}^{n+1}$, we can change this to

$$0 = \mathbf{c} \cdot (\mathbf{x} - \mathbf{a}) = \mathbf{c} \cdot \mathbf{x} - \mathbf{c} \cdot \mathbf{a}$$

or equivalently, $\mathbf{c} \cdot \mathbf{x} = d$ for some constant $d = \mathbf{c} \cdot \mathbf{a}$. The difference between writing an n-plane as $\mathbf{c} \cdot \mathbf{x} = d$ and $\mathbf{c} \cdot (\mathbf{x} - \mathbf{a}) = 0$ is equivalent to the difference between writing a line as $y = mx + b$ or in point-slope format $(y - y_0) = m(x - x_0)$.

For a multivariate function $f : \mathbb{R}^n \to \mathbb{R}$, our generalization of being differentiable at a point $\mathbf{a}$ should then be that f behaves like an n-plane near $\mathbf{a}$.

Definition 2.6

We say that a function $f : \mathbb{R}^n \to \mathbb{R}$ is differentiable at a point $\mathbf{a} \in \mathbb{R}^n$ if there exists $\mathbf{c} \in \mathbb{R}^n$ such that

$$\lim_{\mathbf{h} \to \mathbf{0}} \frac{f(\mathbf{a} + \mathbf{h}) - f(\mathbf{a}) - \mathbf{c} \cdot \mathbf{h}}{\|\mathbf{h}\|} = 0.$$

The constant $\mathbf{c}$, if it exists, is called the *gradient of* f, and is often denoted $\nabla f(\mathbf{a})$.

Recall that the limit $\mathbf{h} \to 0$ means that we approach $\mathbf{0}$ from every possible direction, which is why we had to use an n-plane to capture the idea of approaching the point $\mathbf{a}$ from every conceivable direction. One can show that the equation of the tangent n-plane at $f(\mathbf{a})$ is given by $x_{n+1} = f(\mathbf{a}) + \nabla f(\mathbf{a}) \cdot \mathbf{x}$ so that $f(\mathbf{a}+\mathbf{h}) - \mathbf{f}(\mathbf{a}) - \nabla f(\mathbf{a}) \cdot \mathbf{h}$ represents the error between between the value of the function and that of the tangent plane. Once again, the condition on differentiability means that this error goes to zero faster than linearly.

Example 2.7

Show that the function $f(x,y) = x^2 + y^2$ is differentiable at the point $\mathbf{a} = (1,0)$ with $\nabla f(1,0) = (2,0)$. Determine more generally what $\nabla f(\mathbf{a})$ should be for general $\mathbf{a}$.

Solution. Let $\mathbf{h} = (h_1, h_2)$. Checking the definition of differentiability, we have

$$\begin{aligned}\lim_{\mathbf{h}\to mb0} \frac{f(\mathbf{a}+\mathbf{h}) - f(\mathbf{a}) - \nabla f(\mathbf{a})\cdot\mathbf{h}}{\|\mathbf{h}\|} &= \lim_{\mathbf{h}\to\mathbf{0}} \frac{f(1+h_1,h_2) - f(1,0) - (2,0)\cdot(h_1,h_2)}{\sqrt{h_1^2+h_2^2}} \\ &= \lim_{\mathbf{h}\to\mathbf{0}} \frac{(1+h_1)^2 + h_2^2 - 1 - 2h_1}{\sqrt{h_1^2+h_2^2}} \\ &= \lim_{\mathbf{h}\to\mathbf{0}} \frac{1 + 2h_1 + h_1^2 + h_2^2 - 1 - 2h_1}{\sqrt{h_1^2+h_2^2}} \\ &= \lim_{\mathbf{h}\to\mathbf{0}} \sqrt{h_1^2+h_2^2} = 0\end{aligned}$$

which is precisely what we wanted to show. More generally, let $\mathbf{a} = (x,y)$ and $\nabla f(\mathbf{a}) = (c_1, c_2)$, so that differentiability becomes

$$\begin{aligned}&\lim_{\mathbf{h}\to\mathbf{0}} \frac{f(x+h_1, y+h_2) - f(x,y) - (c_1,c_2)\cdot(h_1,h_2)}{\|\mathbf{h}\|} \\ &\quad= \lim_{\mathbf{h}\to\mathbf{0}} \frac{(x^2 + 2xh_1 + h_1^2) + (y + 2yh_2 + h_1)^2 - x^2 - y^2 - c_1h_1 - c_2h_2}{\sqrt{h_1^2+h_2^2}} \\ &\quad= \lim_{\mathbf{h}\to\mathbf{0}} \frac{h_1(2x - c_1) + h_2(2y - c_2) + h_1^2 + h_2^2}{\sqrt{h_1^2+h_2^2}}.\end{aligned}$$

If either $2x - c_1, 2y - c_2 \neq 0$ then this limit does not exist, which implies that $c_1 = 2x$ and $c_2 = 2y$; that is, $\nabla f(x,y) = (2x, 2y)$. ■

Remark 2.8

1. It was very necessary that we considered the entire $f(\mathbf{a}+h) - \mathbf{f}(\mathbf{a}) - \nabla f(\mathbf{a})\cdot\mathbf{h}$ term, since the cancellations were necessary to ensure that the limit exists. As such, we cannot just drop the $\nabla f(\mathbf{a})\cdot\mathbf{h}$ like we were able to do when our functions were maps from $\mathbb{R}$ to $\mathbb{R}$.

2. Notice that our gradient $\nabla f(\mathbf{x}) = (2x, 2y)$ contains the terms $2x$ and $2y$ which are the derivatives of x^2 and y^2 respectively. So it seems like the gradient might still be related to the one-dimensional derivatives.

Theorem 2.9

If $f : \mathbb{R}^n \to \mathbb{R}$ is differentiable at $\mathbf{a}$ then f is continuous at $\mathbf{a}$.

Proof. Since f is differentiable at $\mathbf{a}$ we have that

$$\begin{aligned}\lim_{\mathbf{h}\to\mathbf{0}}\left[f(\mathbf{a}+\mathbf{h})-f(\mathbf{a})-\nabla f(\mathbf{a})\cdot\mathbf{h}\right] &= \lim_{\mathbf{h}\to\mathbf{0}}\frac{f(\mathbf{a}+h)-f(\mathbf{a})-\nabla f(\mathbf{a})\cdot\mathbf{h}}{\|\mathbf{h}\|}\|\mathbf{h}\| \\ &= \left[\lim_{\mathbf{h}\to\mathbf{0}}\frac{f(\mathbf{a}+h)-f(\mathbf{a})-\nabla f(\mathbf{a})\cdot\mathbf{h}}{\|\mathbf{h}\|}\right]\left[\lim_{\mathbf{h}\to\mathbf{0}}\|\mathbf{h}\|\right] \\ &= 0.\end{aligned}$$

Since this limit exists, we can conclude continuity as follows:

$$\begin{aligned}0 &= \lim_{\mathbf{h}\to\mathbf{0}}\Big[\left[f(\mathbf{a}+\mathbf{h})-f(\mathbf{a})-\nabla f(\mathbf{a})\cdot\mathbf{h}\right]+\nabla f(\mathbf{a})\cdot\mathbf{h}\Big] \\ &= \lim_{\mathbf{h}\to\mathbf{0}}\left[f(\mathbf{a}+\mathbf{h})-f(\mathbf{a})\right].\end{aligned}$$

□

Partial Derivatives: We take a small detour to develop some machinery before returning to differentiability. We have seen that it can be very difficult to capture how a limit approaches a point in multiple dimensions, precisely because there are infinitely many possible ways to approach a point. The same argument works for differentiability: There is no obvious way of writing down the rate of change of a function in infinitely many directions simultaneously.

However, we know from linear algebra that we do not have to be able to describe every vector in $\mathbb{R}^n$, only a finite subset of basis vectors, from which every other vector can be built through a linear combination. We will apply this idea here, and determine the rate of change of the function f in each of standard unit vectors.

Definition 2.10

Write $(x_1,\ldots,x_n)$ to denote the coordinates of $\mathbb{R}^n$. If $f:\mathbb{R}^n\to\mathbb{R}$, we define the partial derivative of f with respect to x_i at $\mathbf{a}=(a_1,\ldots,a_n)\in\mathbb{R}^n$ as

$$\frac{\partial f}{\partial x_i}(\mathbf{a})=\lim_{h\to 0}\frac{f(a_1,\ldots,a_i+h,\ldots,a_n)-f(a_1,\ldots,a_n)}{h}.$$

That is, $\frac{\partial f}{\partial x_i}$ is the one-variable derivative of $f(x_1,\ldots,x_n)$ with respect to x_i, where all other variables are held constant.

Example 2.11

Determine the partial derivatives of the function $f(x,y,z)=xy+sin(x^2z)+z^{-2}e^y$.

Solution. Remember that when computing the partial derivative with respect to x_i, we treat all other variables as constants. Hence

$$\begin{aligned}\frac{\partial f}{\partial x} &= y+2xz\cos(x^2z)\\ \frac{\partial f}{\partial y} &= x+\frac{e^y}{z^2}\\ \frac{\partial f}{\partial z} &= x^2\cos(x^2z)-\frac{2e^y}{z^3}.\end{aligned}$$

■

It can be quite cumbersome to write $\frac{\partial f}{\partial x_i}$, so we will often interchange it with any of the following when it is unambiguous:

$$\frac{\partial f}{\partial x_i}, \quad \partial_{x_i} f, \quad \partial_i f, \quad f_{x_i}, \quad f_i.$$

This will be particularly convenient when we start taking higher order partial derivatives.

Recall in Example 2.7 we showed that if $f(x,y) = x^2+y^2$ then $\nabla f(x,y) = (2x, 2y) = (\partial_x f, \partial_y f)$. Is this just a coincidence, or does it hold more generally?

Theorem 2.12

If $f : \mathbb{R}^n \to \mathbb{R}$ is differentiable at $\mathbf{a}$ then the partials of f exist at $\mathbf{a}$ and

$$\nabla f(\mathbf{a}) = \left(\frac{\partial f}{\partial x_1}(\mathbf{a}), \ldots, \frac{\partial f}{\partial x_n}(\mathbf{a}) \right).$$

Proof. This is actually a fairly natural result. If differentiability means that the limit exists from every direction and partial derivatives are only capturing information about a single direction, it seems natural that one would imply the other.

More directly, let $e_i = (\underbrace{0, \ldots, 1}_{i\text{-times}}, 0, \ldots, 0)$ be the standard unit vector in the i-th direction. We are going to use our knowledge that the function is differentiable and approach along the i-th coordinate axis. Indeed, let $\mathbf{h} = he_i$ and $\nabla f(\mathbf{a}) = (c_1, \ldots, c_n)$ so that

$$\begin{aligned} 0 &= \lim_{h \to 0} \frac{f(\mathbf{a}+\mathbf{h}) - f(\mathbf{a}) - \nabla f(\mathbf{a}) \cdot \mathbf{h}}{\|h\|} \\ &= \lim_{h \to 0} \frac{f(a_1, \ldots, a_i + h, \ldots, a_n) - f(a_1, \ldots, a_n)}{h} - c_i \end{aligned}$$

[Note: The final equation above is always true, but how we arrive at it depends on the sign of h. Convince yourself of this!] Re-arranging gives $\frac{\partial f}{\partial x_i} = c_i$. Since this holds for arbitrary i, we conclude that x

$$\nabla f(\mathbf{a}) = (\partial_{x_1} f, \ldots, \partial_{x_n} f).$$ □

It is important to note however that the converse of this theorem is not true; that is, it is possible for the partial derivatives to exist but for the function to not be differentiable. Indeed, it is precisely because the partials only measure the differentiability in finitely many directions that the converse direction does not hold. Consider

$$f(x,y) = \begin{cases} \frac{xy}{x^2+y^2}, & \text{if } (x,y) \neq (0,0) \\ 0 & \text{if } (x,y) = (0,0). \end{cases} \tag{2.2}$$

We know that this function is not continuous at $(0,0)$ (for example, approach along the line $y = mx$) and so has no chance of being differentiable at $(0,0)$. Nonetheless, the partial derivatives exist at $(0,0)$ since

$$\frac{\partial f}{\partial x}(0,0) = \lim_{h \to 0} \frac{f(h,0) - f(0,0)}{h} = 0 = \lim_{h \to 0} \frac{f(0,h) - f(0,0)}{h} = \frac{\partial f}{\partial y}(0,0).$$

To arrive at a meaningful converse, we need to add an extra regularity condition:

Theorem 2.13

Let $f : \mathbb{R}^n \to \mathbb{R}$ and $\mathbf{a} \in \mathbb{R}^n$. If $\partial_i f(\mathbf{x})$ all exist and are continuous in an open neighbourhood of $\mathbf{a}$, then f is differentiable at $\mathbf{a}$.

The proof of this theorem is very slick, but is not terribly enlightening. As such, we omit its proof and leave it as an exercise for the student (the proof may be found in Folland). Once again let f be the function in (2.2). Notice that its partial derivatives are given by

$$\frac{\partial f}{\partial x} = \frac{y^3 - x^2 y}{(x^2+y^2)^2}, \qquad \frac{\partial f}{\partial y} = \frac{x^3 - y^2 x}{(x^2+y^2)^2}$$

and these functions do not have limits as $(x, y) \to (0, 0)$ (try the line $y = -x$). Hence the partial derivatives are not continuous, and Theorem 2.13 does not apply.

Definition 2.14

We define the collection of C^1 *functions on* U to be

$$C^1(\mathbb{R}^n, \mathbb{R}) = \left\{ f : \mathbb{R}^n \to \mathbb{R} : \begin{array}{c} \partial_i f \text{ exists and is continuous} \\ i = 1, \ldots, n \end{array} \right\}.$$

That is, a function f is C^1 if all of its partial exist and are continuous.

All C^1 functions are automatically differentiable by Theorem 2.13; however, there are differentiable functions which are not C^1. For example, the function

$$f(x,y) = \begin{cases} (x^2+y^2)\sin\left(\frac{1}{\sqrt{x^2+y^2}}\right), & \text{if } (x,y) \neq (0,0) \\ 0 & \text{if } (x,y) = (0,0) \end{cases} \tag{2.3}$$

is everywhere differentiable, but its partial derivatives are not continuous at $(0, 0)$.

Exercise: Show that (2.3) is differentiable but its partial derivatives are not continuous at $(0, 0)$.

We have presented a lot of theorems and counter-examples, so let's take a moment to summarize what we have said:

Differentiable	$\Rightarrow$	Partials Exist	Theorem 2.12
Differentiable	$\not\Leftarrow$	Partials Exist	Function (2.2)
Partials exist and continuous	$\Rightarrow$	Differentiable	Theorem 2.13
Partials exist and continuous	$\not\Leftarrow$	Differentiable	Function (2.3)

Directional Derivatives: Partial derivatives gave us the ability to determine how a function was changing along the coordinate axes, but what if we want to see how the derivative is changing along other vectors? This is done via directional derivatives:

Definition 2.15

Let $f : \mathbb{R}^n \to \mathbb{R}$ and $\mathbf{a} \in \mathbb{R}^n$. If $\mathbf{u} \in \mathbb{R}^n$ is a unit vector ($\|\mathbf{u}\| = 1$) then the *directional derivative of f in the direction $\mathbf{u}$ at $\mathbf{a}$* is

$$\partial_{\mathbf{u}} f(\mathbf{a}) = \lim_{t\to 0} \frac{f(\mathbf{a}+t\mathbf{u}) - f(\mathbf{a})}{t} = \left.\frac{\mathrm{d}}{\mathrm{d}t}\right|_{t=0} f(\mathbf{a}+t\mathbf{u}).$$

This represents an idea that is prevalent throughout mathematics, and especially the field of *differential geometry*. First of all, notice that $\gamma : \mathbb{R} \to \mathbb{R}^n$ given by $\gamma(t) = \mathbf{a} + t\mathbf{u}$ is the straight line through $\mathbf{a}$ in the direction of $\mathbf{u}$, and hence is a curve. By composing with f, we get $g = f \circ \gamma : \mathbb{R} \to \mathbb{R}^n \to \mathbb{R}$ which is just a normal, one-variable function, and hence can be differentiated as normal. We know that $\gamma'(t) = \mathbf{u}$ is the velocity vector of the curve, so to see how the function behaves in the direction $\mathbf{u}$ we look at how the function f behaves in a neighbourhood of our point $\mathbf{a}$, and differentiate at $t = 0$ to get the behaviour in this direction.

Example 2.16

Determine the directional derivative of $f(x, y) = \sin(xy) + e^x$ in the direction $\mathbf{u} = \frac{1}{\sqrt{5}}(1, 2)$ at the point $\mathbf{a} = (0, 0)$.

Solution. We can proceed by direct computation:

$$\begin{aligned}
\left.\frac{\mathrm{d}}{\mathrm{d}t}\right|_{t=0} f(\mathbf{a}+t\mathbf{u}) &= \left.\frac{\mathrm{d}}{\mathrm{d}t}\right|_{t=0} f\left(\frac{t}{\sqrt{5}}, \frac{2t}{\sqrt{5}}\right) \\
&= \left.\frac{\mathrm{d}}{\mathrm{d}t}\right|_{t=0} \left(\sin\left(\frac{2}{5}t^2\right) + e^{t/\sqrt{5}}\right) \\
&= \left[\frac{4}{5}t\cos\left(\frac{2}{5}t^2\right) + \frac{1}{\sqrt{5}}e^{t/\sqrt{5}}\right]_{t=0} \\
&= \frac{1}{\sqrt{5}}
\end{aligned}$$

■

Theorem 2.17

If $f : \mathbb{R}^n \to \mathbb{R}$ is differentiable at $\mathbf{a}$, then for any unit vector $\mathbf{u}$, $\partial_{\mathbf{u}} f$ exists. Moreover, $\partial_{\mathbf{u}} f(\mathbf{a}) = \nabla f(\mathbf{a}) \cdot \mathbf{u}$.

Proof. The idea is almost exactly the same as Theorem 2.12. We will approach along the line $\mathbf{a}+t\mathbf{u}$ and use differentiability to conclude that the limit exists. As such, let $\mathbf{h} = t\mathbf{u}$ for $t \in \mathbb{R}$, so that

$$\begin{aligned}
0 &= \lim_{t\to 0} \frac{f(\mathbf{a}+\mathbf{h}) - f(\mathbf{a}) - \nabla f(\mathbf{a}) \cdot \mathbf{h}}{\|h\|} \\
&= \lim_{t\to 0} \frac{f(\mathbf{a}+t\mathbf{u}) - f(\mathbf{a}) - t\nabla f(\mathbf{a}) \cdot \mathbf{u}}{t} \\
&= \left(\lim_{t\to 0} \frac{f(\mathbf{a}+t\mathbf{u}) - f(\mathbf{a})}{t}\right) - \nabla f(\mathbf{a}) \cdot \mathbf{u}.
\end{aligned}$$

Re-arranging, we get $\partial_{\mathbf{u}} f(\mathbf{a}) = \nabla f(\mathbf{a}) \cdot \mathbf{u}$ as required. □

Example 2.18

Verify the result from Example 2.16 by using the above theorem.

Solution. Our function $f(x, y) = \sin(xy) + e^x$ is clearly differentiable, as its partial derivatives exist and are continuous:

$$\frac{\partial f}{\partial x} = y\cos(xy) + e^x, \quad \frac{\partial f}{\partial y} = x\cos(xy).$$

At the point $\mathbf{a} = (0, 0)$ the gradient is $\nabla f(0, 0) = (1, 0)$, and so

$$\partial_{\mathbf{u}} f(0,0) = \nabla f(0,0) \cdot \mathbf{u} = (1,0) \cdot \left(\frac{1}{\sqrt{5}}, \frac{2}{\sqrt{5}}\right) = \frac{1}{\sqrt{5}}.$$ ■

Exercise: Show that the converse of Theorem 2.17 is false; that is, there is a function in which every directional derivative exists at a point, but the function is not differentiable.

2.1.4 Functions $\mathbb{R}^n \to \mathbb{R}^m$

Our motivation for defining the derivative thus far has been that a function is differentiable if it can be approximated by a linear function, with an error that tends to zero faster than linear. In such instances, that linear approximation is what we call the derivative. The same story will hold for functions $\mathbf{f} : \mathbb{R}^n \to \mathbb{R}^m$.

So what does it mean for a function $\mathbf{L} : \mathbb{R}^n \to \mathbb{R}^m$ to be linear? Here, the word linear has the same interpretation as it does in *linear* algebra; that is, for every $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$

$$\mathbf{L}(a\mathbf{x} + b\mathbf{y}) = a\mathbf{L}(\mathbf{x}) + b\mathbf{L}(\mathbf{y}).$$

The student is hopefully familiar with the fact that such maps can be represented by matrices, with respect to some basis. In particular, if $\mathbf{L} : \mathbb{R}^n \to \mathbb{R}^m$ then $\mathbf{L}$ must take an n-vector to an m-vector, and thus must be an $m \times n$-matrix, say A. In this basis, we can write $\mathbf{L}(\mathbf{x}) = A\mathbf{x}$. Thus we would like to say something along the lines of "A function $\mathbf{f} : \mathbb{R}^n \to \mathbb{R}^m$ is differentiable at $\mathbf{a} \in \mathbb{R}^n$ if there exists a matrix A such that

$$\mathbf{f}(\mathbf{a} + \mathbf{h}) = \mathbf{f}(\mathbf{a}) + A\mathbf{h} + \text{error}(\mathbf{h}).$$

Solving for the error we get

$$\text{error}(\mathbf{h}) = \mathbf{f}(\mathbf{a} + \mathbf{h}) - \mathbf{f}(\mathbf{a}) - A\mathbf{h}.$$

For this approximation to do a good job, the error should tend to zero faster than linearly, leading us to the following definition:

Definition 2.19

A function $\mathbf{f} : \mathbb{R}^n \to \mathbb{R}^m$ is differentiable at the point $\mathbf{a} \in \mathbb{R}^n$ if there exists an $m \times n$ matrix A such that

$$\lim_{\mathbf{h}\to 0} \frac{\|\mathbf{f}(\mathbf{a}+\mathbf{h}) - \mathbf{f}(\mathbf{a}) - A\mathbf{h}\|_{\mathbb{R}^m}}{\|\mathbf{h}\|_{\mathbb{R}^n}} = 0. \tag{2.4}$$

We often denote the quantity A by $D\mathbf{f}(\mathbf{a})$, referred to as the *Jacobian matrix*.

Example 2.20

Let $f : \mathbb{R}^3 \to \mathbb{R}^2$ be given by $f(x, y, z) = (x^2, xz + y)^T$. Show that f is differentiable at the point $\mathbf{a} = (-1, 1, 0)$ with

$$D\mathbf{f}(\mathbf{a}) = \begin{pmatrix} -2 & 0 & 0 \\ 0 & 1 & -1 \end{pmatrix}.$$

Solution. Let $\mathbf{h} = (h_1, h_2, h_3)$ so that $\mathbf{a} + \mathbf{h} = (-1 + h_1, 1 + h_2, h_3)$. Computing the numerator, we get

$$\begin{aligned} \|\mathbf{f}(\mathbf{a}+\mathbf{h}) - \mathbf{f}(\mathbf{a}) - D\mathbf{f}(\mathbf{a})\mathbf{h}\| &= \left\| \begin{pmatrix} h_1^2 - 2h_1 + 1 \\ -h_3 + h_1h_3 + h_2 + 1 \end{pmatrix} - \begin{pmatrix} 1 \\ 1 \end{pmatrix} - \begin{pmatrix} -2 & 0 & 0 \\ 0 & 1 & -1 \end{pmatrix} \begin{pmatrix} h_1 \\ h_2 \\ h_3 \end{pmatrix} \right\| \\ &= \left\| \begin{pmatrix} h_1^2 \\ h_1h_3 \end{pmatrix} \right\| = \sqrt{h_1^4 + h_1^2h_3^2} \\ &= h_1\sqrt{h_1^2 + h_3^2}. \end{aligned}$$

We will have to proceed by the Squeeze Theorem. Taking the entire difference quotient into consideration, we have

$$0 \le \frac{\|\mathbf{f}(\mathbf{a}+\mathbf{h}) - \mathbf{f}(\mathbf{a}) - A\mathbf{h}\|_{\mathbb{R}^m}}{\|\mathbf{h}\|_{\mathbb{R}^n}} = \frac{h_1\sqrt{h_1^2 + h_3^2}}{\sqrt{h_1^2 + h_2^2 + h_3^2}} \le \frac{h_1\sqrt{h_1^2 + h_3^2}}{\sqrt{h_1^2 + h_3^2}} = h_1.$$

As both the upper and lower bounds limit to 0, we conclude that f is differentiable as required. ■

We would like to find a much better way of determining $D\mathbf{f}(\mathbf{a})$ than using the limit definition. If $\mathbf{f}$ is differentiable, then

$$\lim_{\mathbf{h}\to\mathbf{0}} \|\mathbf{f}(\mathbf{a}+\mathbf{h}) - \mathbf{f}(\mathbf{a}) - Df(\mathbf{a})\mathbf{h}\| = 0.$$

Furthermore, the norm of a vector tends to zero if and only if each of its terms also tends to zero. Write $[D\mathbf{f}(a)]_i$ for the i-th row of $D\mathbf{f}(a)$ and let $\mathbf{f}(\mathbf{x}) = (f_1(\mathbf{x}), \ldots, f_m(\mathbf{x}))$. Then (2.4) is equivalent to the statement that for each $i = 1, \ldots, m$ we have

$$\lim_{\mathbf{h}\to 0} \frac{|f_i(\mathbf{a}+\mathbf{h}) - f_i(\mathbf{a}) - [D\mathbf{f}(\mathbf{a})]_i \cdot \mathbf{h}|}{\|\mathbf{h}\|} = 0.$$

Notice that this is exactly the definition of the gradient, and so $[D\mathbf{f}(\mathbf{a})]_i = \nabla f_i(\mathbf{a})$. We thus get the following result for free:

Proposition 2.21

If $\mathbf{f} : \mathbb{R}^n \to \mathbb{R}^m$ is given by $\mathbf{f}(\mathbf{x}) = (f_1(\mathbf{x}), \ldots, f_m(\mathbf{x}))$, then $\mathbf{f}$ is differentiable if and only if each of the $f_i : \mathbb{R}^n \to \mathbb{R}$ is differentiable, and in this case

$$D\mathbf{f}(\mathbf{a}) = \begin{pmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} & \cdots & \frac{\partial f_2}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1} & \frac{\partial f_m}{\partial x_2} & \cdots & \frac{\partial f_m}{\partial x_n} \end{pmatrix}.$$

Example 2.22

Determine the derivative of the function $\mathbf{f}(r, \theta) = (r\sin(\theta), r\cos(\theta))$.

Solution. By definition, the derivative is the matrix of partial derivatives, so we can compute this to be

$$d\mathbf{f}(r, \theta) = \begin{pmatrix} \sin(\theta) & r\cos(\theta) \\ \cos(\theta) & -r\sin(\theta) \end{pmatrix}. \qquad \blacksquare$$

Example 2.23

Determine the derivative of the function $\mathbf{f}(x, y, z) = (xy, z\sin(xy), e^{xz})$.

Solution. Once again, we compute the matrix of partial derivatives:

$$d\mathbf{f}(x, y, z) = \begin{pmatrix} y & x & 0 \\ zy\cos(xy) & xz\cos(xy) & \sin(xy) \\ ze^{xz} & 0 & xe^{xz} \end{pmatrix}. \qquad \blacksquare$$

2.2 The Chain Rule

Given two functions $\mathbf{g} : \mathbb{R}^k \to \mathbb{R}^n$ and $\mathbf{f} : \mathbb{R}^n \to \mathbb{R}^m$, their composition is given by $\mathbf{f} \circ \mathbf{g} : \mathbb{R}^k \to \mathbb{R}^n \to \mathbb{R}^m$. Just as was the case in one-variable, we would like to determine when this new function is differentiable, and how to write its derivative in terms of $D\mathbf{f}$ and $D\mathbf{g}$.

Let's start by looking at what happens in one dimension. If $k = n = m = 1$ then the derivative of $f \circ g$ is given by $(f \circ g)'(a) = f'(g(a))g'(a)$. For more general k, n, and m, we know that $D\mathbf{f}$ is an $m \times n$ matrix, $D\mathbf{g}$ is an $n \times k$ matrix, and $D(\mathbf{f} \circ \mathbf{g})$ needs to be an $m \times k$ matrix. There is only one way to combine these matrices:

Theorem 2.24: Chain Rule

Let $\mathbf{g} : \mathbb{R}^k \to \mathbb{R}^n$ and $\mathbf{f} : \mathbb{R}^n \to \mathbb{R}^m$. If $\mathbf{g}$ is differentiable at $\mathbf{a} \in \mathbb{R}^k$ and $\mathbf{f}$ is differentiable at $\mathbf{g}(\mathbf{a}) \in \mathbb{R}^n$, then $\mathbf{f} \circ \mathbf{g}$ is differentiable at $\mathbf{a}$, and moreover its derivative can be written as

$$D(\mathbf{f} \circ \mathbf{g})(\mathbf{a}) = D\mathbf{f}(\mathbf{g}(\mathbf{a}))D\mathbf{g}(a).$$

The proof of the Chain Rule in even a single dimension is tricky. The addition of multiple dimensions only serves to make the proof messy, so we will omit it. Here now it is important to make the distinction between which objects are treated as rows and which are treated as columns. If $f : \mathbb{R}^n \to \mathbb{R}^m$ then $D\mathbf{f}(\mathbf{a})$ should reduce a gradient when $m = 1$, and should be curve derivative when $n = 1$. In particular, this implies that the gradient of a function $\mathbb{R}^n \to \mathbb{R}$ is a *row vector*, while the derivative of a function $\mathbb{R} \to \mathbb{R}^n$ is a *column vector*.

There are a few notable cases that we should take into account. Let $\mathbf{g} : \mathbb{R} \to \mathbb{R}^n$ and $f : \mathbb{R}^n \to \mathbb{R}$ so that $f \circ \mathbf{g} : \mathbb{R} \to \mathbb{R}$. By the Chain Rule, we must then have

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}(f \circ \mathbf{g})(t) &= \nabla f(\mathbf{g}(t)) \cdot \mathbf{g}'(t) \\ &= \left.\frac{\partial f}{\partial x_1}\right|_{\mathbf{g}(t)} g_1'(t) + \cdot + \left.\frac{\partial f}{\partial x_n}\right|_{\mathbf{g}(t)} g_n'(t).\end{aligned}$$

Using Leibniz notation, let $y = f(\mathbf{x})$ and set $(x_1, \ldots, x_n) = \mathbf{g}(t) = (g_1(t), \ldots, g_n(t))$ so that $g_i'(t) = \frac{\mathrm{d}x_i}{\mathrm{d}t}$. Our derivative now becomes

$$\frac{\mathrm{d}}{\mathrm{d}t}(f \circ \mathbf{g}) = \frac{\partial y}{\partial x_1}\frac{\partial x_1}{\partial t} + \cdots + \frac{\partial y}{\partial x_n}\frac{\partial x_n}{\partial t}.$$

Once again, it seems as though the derivatives are 'cancelling,' though this is not the case.

Example 2.25

Let $\mathbf{g}(t) = (\sin(t), \cos(t), t^2)$ and $f(x, y, z) = x^2 + y^2 + xyz$. Determine the derivative of $f \circ \mathbf{g}$.

Solution. One does not need to use the chain rule here, since we can explicitly write

$$f(\mathbf{g}(t)) = f(\sin(t), \cos(t), t^2) = \sin^2(t) + \cos^2(t) + t^2 \sin(t)\cos(t) = 1 + t^2 \sin(t)\cos(t),$$

and differentiating yields

$$\frac{\mathrm{d}}{\mathrm{d}t} f(\mathbf{g}(t)) = 2t\sin(t)\cos(t) + t^2\cos^2(t) - t^2\sin^2(t).$$

Let's see that we get the same answer with the chain rule. We know that $\mathbf{g}'(t) = (\cos(t), -\sin(t), 2t)$ and $\nabla f(x, y, z) = (2x + yz, 2y + xz, xy)$ so that

$$\begin{aligned}\nabla f(\mathbf{g}(t)) \cdot \mathbf{g}'(t) &= (2\sin(t) + t^2\cos(t), 2\cos(t) + t^2\sin(t), \sin(t)\cos(t)) \cdot (\cos(t), -\sin(t), 2t) \\ &= 2\sin(t)\cos^2(t) + t^2\cos(t) - 2\cos(t)\sin 9t) - t^2\sin^2(t) + 2t\cos(t)\sin(t) \\ &= 2t\sin(t)\cos(t) + t^2\cos^2(t) - t^2\sin^2(t).\end{aligned}$$

■

Now let $\mathbf{g}:\mathbb{R}^n\to\mathbb{R}^m$ and $f:\mathbb{R}^m\to\mathbb{R}$ so that $f\circ\mathbf{g}:\mathbb{R}^n\to\mathbb{R}$. The Chain Rule tells us that

$$\begin{aligned}\nabla(f\circ\mathbf{g})(\mathbf{x}) &= \nabla f(\mathbf{g}(\mathbf{x}))D\mathbf{g}(\mathbf{x})\\ &= \left(\frac{\partial f}{\partial x_1},\dots,\frac{\partial f}{\partial x_n}\right)\begin{pmatrix}\frac{\partial g_1}{\partial t_1} & \frac{\partial g_1}{\partial t_2} & \cdots & \frac{\partial g_1}{\partial t_n}\\ \frac{\partial g_2}{\partial t_1} & \frac{\partial g_2}{\partial t_2} & \cdots & \frac{\partial g_2}{\partial t_n}\\ \vdots & \vdots & \ddots & \vdots\\ \frac{\partial g_m}{\partial t_1} & \frac{\partial g_m}{\partial t_2} & \cdots & \frac{\partial g_m}{\partial t_n}\end{pmatrix}.\end{aligned}$$

Thus if we set $y=f(\mathbf{x})$ and $\mathbf{x}=\mathbf{g}(\mathbf{t})$ then

$$\frac{\partial}{\partial t_i}(f\circ\mathbf{g})(\mathbf{x}) = \frac{\partial y}{\partial x_1}\frac{\partial x_1}{\partial t_i}+\frac{\partial y}{\partial x_2}\frac{\partial x_2}{\partial t_i}+\cdots+\frac{\partial y}{\partial x_n}\frac{\partial x_n}{\partial t_i}.$$

Example 2.26

Let $f(x,y,z)=xz+e^{yz}$ and $\mathbf{g}(t_1,t_2)=(t_1,t_2,t_1t_2)$. Determine $\nabla(f\circ\mathbf{g})$.

Solution. This can again be computed by hand. Notice that

$$(f\circ\mathbf{g})(t_1,t_2)=f(t_1,t_2,t_1t_2)=t_1^2t_2+e^{t_1t_2^2},$$

and so

$$\nabla(f\circ\mathbf{g})(t_1,t_2)=\left(2t_1t_2+t_2^2e^{t_1t_2^2},t_1^2+2t_1t_2e^{t_1t_2^2}\right).$$

On the other hand, $\nabla f=(z,ze^{yz},x+ye^{yz})$ so by the Chain Rule

$$\begin{aligned}\nabla(f\circ\mathbf{g})(t_1,t_2) &= (t_1t_2,t_1t_2e^{t_1t_2^2},t_1+t_2e^{t_1t_2^2})\begin{pmatrix}1&0\\0&1\\t_2&t_1\end{pmatrix}\\ &= \left(t_1t_2+t_1t_2+t_2^2e^{t_1t_2^2},t_1t_2e^{t_1t_2^2}+t_1^2+t_1t_2e^{t_1t_2^2}\right)\\ &= \left(2t_1t_2+t_2^2e^{t_1t_2^2},t_1^2+2t_1t_2e^{t_1t_2^2}\right).\end{aligned}$$

■

The next example is if $\mathbf{g}:\mathbb{R}\to\mathbb{R}^n$ and $\mathbf{f}:\mathbb{R}^n\to\mathbb{R}^m$. The composition is a map $\mathbf{f}\circ\mathbf{g}:\mathbb{R}\to\mathbb{R}^m$ and so in this case the Chain Rule tells us that

$$\frac{\mathrm{d}}{\mathrm{d}t}(\mathbf{f}\circ\mathbf{g})(t)=D\mathbf{f}(\mathbf{g}(t))\cdot\mathbf{g}'(t).$$

Example 2.27

Let $\mathbf{f}(x,y)=(xy,x+y,x-y)$ and $\mathbf{g}(t)=(t,t^2)$. Compute $\frac{\mathrm{d}}{\mathrm{d}t}\left(\mathbf{f}\circ\mathbf{g}\right)(t)$.

Solution. Explicitly computing the map, we have

$$(\mathbf{f} \circ \mathbf{g})(t) = (t^3, t + t^2, t - t^2)^T$$

and so

$$\frac{\mathrm{d}}{\mathrm{d}t}(\mathbf{f} \circ \mathbf{g})(t) = (3t^2, 1 + 2t, 1 - 2t)^T.$$

On the other hand,

$$D\mathbf{f}(x, y) = \begin{pmatrix} y & x \\ 1 & 1 \\ 1 & -1 \end{pmatrix}, \qquad g'(t) = (1, 2t)^T$$

so by the Chain Rule

$$\frac{\mathrm{d}}{\mathrm{d}t}(\mathbf{f} \circ \mathbf{g})(t) = \begin{pmatrix} t^2 & t \\ 1 & 1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} 1 \\ 2t \end{pmatrix} = \begin{pmatrix} 3t^2 \\ 1 + 2t \\ 1 - 2t \end{pmatrix}. \qquad \blacksquare$$

Finally, we do an example using the full Chain Rule:

Example 2.28

Let $\mathbf{g}(r, s) = (r + rs, r^2, s^2)$ and $\mathbf{f}(x, y, z) = (y^2 + z^2, xy)$. Determine $D(\mathbf{f} \circ \mathbf{g})$.

Solution. One can check that

$$D\mathbf{g}(r, s) = \begin{pmatrix} 1 + s & r \\ 2r & 0 \\ 0 & 2s \end{pmatrix}, \qquad D\mathbf{f}(x, y, z) = \begin{pmatrix} 0 & 2y & 2z \\ y & x & 0 \end{pmatrix},$$

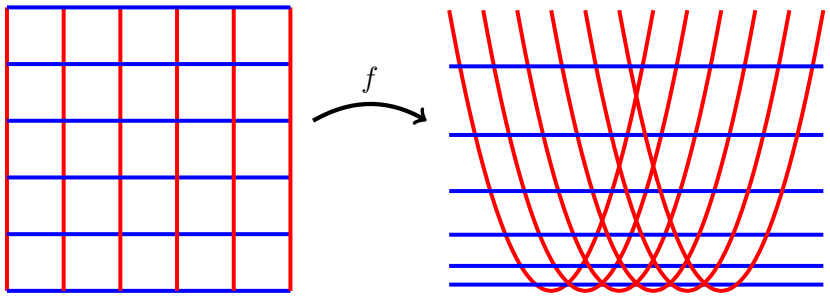
so that by the Chain Rule we have

$$D(\mathbf{f} \circ \mathbf{g})(r, s) = \begin{pmatrix} 0 & 2r^2 & 2s^2 \\ r^2 & r + rs & 0 \end{pmatrix} \begin{pmatrix} 1 + s & r \\ 2r & 0 \\ 0 & 2s \end{pmatrix} = \begin{pmatrix} 4r^3 & 4s^3 \\ 3r^2 + 3r^2 s & r^3 \end{pmatrix}. \qquad \blacksquare$$

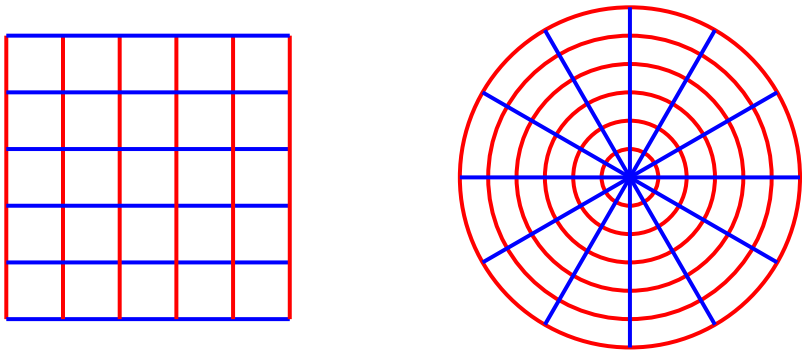
Exercise: In Example 2.28 we used the Chain Rule without explicitly computing the map $\mathbf{f} \circ \mathbf{g}$. Write down the map $\mathbf{f} \circ \mathbf{g}$, compute its derivative, and verify the result of Example 2.28.

An intuition for the derivative: For functions $f : \mathbb{R} \to \mathbb{R}$ and $g : \mathbb{R}^n \to \mathbb{R}$ we had a way of visualizing the derivative: in the former case $f'(a)$ described the slope of the tangent line through a, while in the latter case $\nabla g(\mathbf{a})$ defined a tangent plane. In the case of functions $\mathbb{R}^n \to \mathbb{R}^m$, the visual picture becomes somewhat more complicated.

It is important that we get away from the idea of thinking of such maps as curves or graphs, since neither of these fits into this context of multivariable vector valued maps. Instead, we must truly think of a function as a black-box, which takes an input (elements of $\mathbb{R}^n$) and delivers an



(a) The function $f(x, y) = (x + y, y^2)$ acts on an orthogonal grid in the way pictured.



(b) The function $f(x, y) = (x\cos(y), x\sin(y))$ acts on an orthogonal grid in the way pictured.

Figure 21: One can visualize maps $\mathbb{R}^n \to \mathbb{R}^m$ by how they map orthogonal grids.

output (elements of $\mathbb{R}^m$). If we are lucky and $m = n$, we can try to visualize how such functions work by looking at how orthogonal grids transform (see Figure 21).

So what should derivatives do in this regime? The idea is roughly as follow: Given a point $\mathbf{a} \in \mathbb{R}^n$ and an infinitesimal change in a direction $\mathbf{u}$, we want to characterize how our function transforms that infinitesimal change. Alternatively, pretend that we are driving a car in $\mathbb{R}^n$ and our path is described by the curve $\gamma : \mathbb{R} \to \mathbb{R}^n$ and satisfies

$$\gamma(0) = \mathbf{a}, \qquad \gamma'(0) = \mathbf{u};$$

that is, we pass through the point $\mathbf{a}$ at the time $t = 0$ and here we have a certainly velocity vector $\mathbf{u}$. Now let $\mathbf{f} : \mathbb{R}^n \to \mathbb{R}^m$ be a differentiable (hence continuous) function. The composition $\mathbf{f} \circ \gamma : \mathbb{R} \to \mathbb{R}^m$ is a path in $\mathbb{R}^m$, and so $(\mathbf{f} \circ \gamma)'(0) = \mathbf{v}$ describes the velocity vector at the point $(\mathbf{f} \circ \gamma)(0) = \mathbf{f}(\mathbf{a})$. By the chain rule, we know that

$$\mathbf{v} = (\mathbf{f} \circ \gamma)'(0) = D\mathbf{f}(\mathbf{a})\gamma'(0) = D\mathbf{f}(\mathbf{a})\mathbf{u};$$

namely, $D\mathbf{f}(\mathbf{a})$ describes how our velocity vector $\mathbf{u}$ transforms into the velocity vector $\mathbf{v}$. In fact, this holds regardless of the choice of curve through $\mathbf{a}$, and so

> "$D\mathbf{f}(\mathbf{a})$ describes how velocity vectors through $\mathbf{a}$ transform into velocity vectors through $\mathbf{f}(\mathbf{a})$."

This is illustrated in Figure-**??**.

Change in scale: The quantity $D\mathbf{f}(\mathbf{a})$ represents how velocity vectors transform at the point $\mathbf{a}$. If $f : \mathbb{R}^n \to \mathbb{R}^n$ then $D\mathbf{f}(\mathbf{a})$ is actually a square matrix. A result that the student may be familiar with is that given a linear transformation $A : \mathbb{R}^n \to \mathbb{R}^n$ and a set S, then

$$\text{Area}(A(S)) = \det(A)\text{Area}(S).$$

Of course, we have not been very careful by what the word area means, but this is something we will fix in a later section. Thus $D\mathbf{f}(\mathbf{a})$ can tell us information about how infinitesimal volumes change near $\mathbf{a}$, and leads to the following:

Definition 2.29

If $\mathbf{f} : \mathbb{R}^n \to \mathbb{R}^n$ is differentiable at $\mathbf{a}$, then we define the *Jacobian* (determinant) of $\mathbf{f}$ to be $\det d\mathbf{f}(\mathbf{a})$.

The Jacobian will appear a great deal in later sections, but we will not have too much occasion to use it now. The idea is that the Jacobian describes infinitesimally how areas change under the map f.

Example 2.30

Determine the Jacobian determinant of the maps $\mathbf{f}(r, \theta) = (r\cos(\theta), r\sin(\theta))$ and $\mathbf{g}(x, y) = (x + y, y^2)$.

Solution. These are the maps plotted in Figure-21, and it is a straightforward exercise to compute the Jacobian matrices to be

$$D\mathbf{f}(r, \theta) = \begin{pmatrix} \cos(\theta) & -r\sin(\theta) \\ \sin(\theta) & r\cos(\theta) \end{pmatrix} \qquad D\mathbf{g}(x, y) = \begin{pmatrix} 1 & 1 \\ 0 & 2y \end{pmatrix}.$$

Thus taking determinants, we get the Jacobian determinants

$$\det D\mathbf{f}(r, \theta) = r, \qquad \det D\mathbf{g}(x, y) = 2y.$$ ■

2.3 The Mean Value Theorem

The Mean Value Theorem is one of the most interesting theorem of mathematics. It appears relatively innocuous at first sight, but leads to a plethora of powerful results. In this section we will take a brief moment to examine whether the MVT generalizes to multiple dimensions, and if so how that generalization takes hold.

To begin with, we recall the statement of the Mean Value Theorem:

Theorem 2.31: Mean Value Theorem

If $f : [a, b] \to \mathbb{R}$ is continuous on $[a, b]$ and differentiable on (a, b), then there exists $c \in (a, b)$ such that

$$f(b) - f(a) = f'(c)(b - a). \tag{2.5}$$

One can apply the MVT to prove several important results, such as the following:

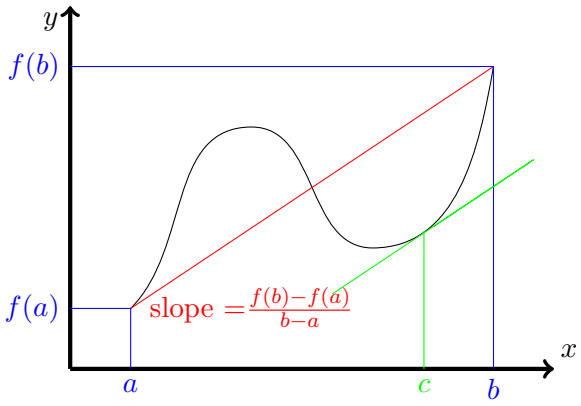


Figure 22: The Mean Value Theorem says that there is a point on this graph such that the tangent line has the same slope as the secant between $(a, f(a))$ and $(b, f(b))$.

1. If $f : [a, b] \to \mathbb{R}$ is differentiable with bounded derivative, say $|f'(x)| \leq M$ for all $x, y \in [a, b]$, then $|f(y) - f(x)| \leq M|y - x|$.

2. If $f'(x) \equiv 0$ for all $x \in [a, b]$ then f is the constant function on $[a, b]$.

3. If $f'(x) > 0$ for all $x \in [a, b]$ then f is an increasing (and hence injective) function.

This is but a short collection of useful theorems; naturally, there are many more.

As a first look at whether or not the MVT generalizes, we should consider functions of the type $\mathbf{f} : \mathbb{R} \to \mathbb{R}^n$. If one were to guess as to what kind of statement a mean value theorem here might have, it would probably be something of the form:

> "If $\mathbf{f} : [a, b] \to \mathbb{R}^n$ is continuous on $[a, b]$ and differentiable on (a, b) then there exists a $c \in [a, b]$ such that
> $$\mathbf{f}(b) - \mathbf{f}(a) = \mathbf{f}'(c)\,(b - a)\,."$$

One should check that the equality sign above even makes sense. The left-hand-side consists of a vector in $\mathbb{R}^n$, while the right-hand-side consists of multiplying a scalar $(b - a)$ with a vector $\mathbf{f}'(c)$. Okay, so the result does make sense. However, applying this to even simple functions immediately results in nonsense.

For example, consider the function $f : [0, 2\pi] \to \mathbb{R}^3$ given by $f(t) = (\cos(t), \sin(t))$. This certainly satisfies our hypotheses, as it is every continuous and everywhere differentiable. On the other hand, $f(0) = (1, 0)$ and $f(2\pi) = (1, 0)$ so that $f(1) - f(0) = (0, 0)$. However, this would then imply that there exists a c such that

$$(0, 0) = (-2\pi \sin(c), 2\pi \cos(c))$$

and this is impossible, since there is no point at which both $\sin(t)$ and $\cos(t)$ are zero.

There is a way to fix this, but we are not interested in how to do this at the moment.

So vector-valued functions fail to admit a generalization of the MVT. Do real-valued multivariate functions have a version of the Mean Value Theorem? The answer is affirmative, and the key lies with the Chain Rule.

Theorem 2.32: Mean Value Theorem for Multivariate Functions

Let $U \subseteq \mathbb{R}^n$ and let $\mathbf{a}, \mathbf{b} \in U$ be such that the straight line connecting them lives entirely within U. More precisely, the curve $\gamma : [0,1] \to \mathbb{R}^n$ given by $\gamma(t) = (1-t)\mathbf{a} + t\mathbf{b}$ satisfies $\gamma(t) \in U$ for all $t \in [0,1]$. If $f : U \to \mathbb{R}$ is a function such that $f \circ \gamma$ is continuous on $[0,1]$ and differentiable on $(0,1)$, then there exists a $t_0 \in (0,1)$ such that $\mathbf{c} = \gamma(t_0)$ and

$$f(\mathbf{b}) - f(\mathbf{a}) = \nabla f(\mathbf{c}) \cdot (\mathbf{b} - \mathbf{a}).$$

Proof. The idea is that we have used the chain rule to reduce this multivariate function to a real-valued function of a single variable. Thinking of the line $\gamma(t) = \mathbf{a}(1-t) + t\mathbf{b}$ as a copy of the interval $[0,1]$ inside of U, restricting f to this line gives a function $f \circ \gamma : [0,1] \to \mathbb{R}$ to which we can apply the original MVT.

More formally, we know that $f \circ \gamma : [0,1] \to \mathbb{R}$ is continuous on $[0,1]$ and differentiable on $(0,1)$, so by the Mean Value Theorem there exists $t_0 \in (0,1)$ such that

$$(f \circ \gamma)(1) - (f \circ \gamma)(0) = (f \circ \gamma)'(t_0)(1-0).$$

Now $(f \circ \gamma)(1) = f(\gamma(1)) = f(\mathbf{b})$ and $(f \circ \gamma)(0) = f(\gamma(0)) = f(\mathbf{a})$. In addition, the Chain Rule tells us that

$$(f \circ \gamma)'(t_0) = \nabla f(\gamma(t_0)) \cdot \gamma'(t_0) = \nabla f(\mathbf{c}) \cdot (\mathbf{b} - \mathbf{a}).$$

Combining everything together, we get

$$f(\mathbf{b}) - f(\mathbf{a}) = \nabla f(\mathbf{c}) \cdot (\mathbf{b} - \mathbf{a}),$$

as required. □

Important to the statement of the Mean Value Theorem is the fact that the line segment connecting $\mathbf{a}$ and $\mathbf{b}$ lives entirely within U. Conveniently, we have already seen that convex sets satisfy this property for any pair of points within the set.

Corollary 2.33

If $U \subseteq \mathbb{R}^n$ is convex and $f : U \to \mathbb{R}$ is a differentiable function such that $|\nabla f(\mathbf{x})| \leq M$ for all $\mathbf{x} \in U$, then for every $\mathbf{a}, \mathbf{b} \in U$ we have

$$|f(\mathbf{b}) - f(\mathbf{a})| \leq M|\mathbf{b} - \mathbf{a}|.$$

Corollary 2.34

If $U \subseteq \mathbb{R}^n$ is convex and $f : U \to \mathbb{R}$ is a differentiable function such that $\nabla f(\mathbf{x}) = 0$ for all $\mathbf{x} \in U$, then f is a constant function on U.

Exercise: The proofs of Corollaries 2.33 and 2.34 are almost identical to their single variable equivalents. Prove these theorems.

2.4 Multi-indices and higher order partials

2.4.1 Second-Order Partial Derivatives

For differentiable functions of the type $f : \mathbb{R} \to \mathbb{R}$, a lot of information about f could be derived not only from its first derivative f', but from its higher order derivatives $f^{(n)}$. For example, if f represents some physical quantity such as position as a function of time, we know that f' is its velocity, f'' is its acceleration, and $f^{(3)}$ is its jerk. This means that the higher-order derivatives are essential when modelling differential equations. We used an infinite number of derivatives when computing Taylor series, and we exploited the second derivative test to determine optimality of critical points. All of these applications and more will extend to functions $f : \mathbb{R}^n \to \mathbb{R}$.

The first step is second-order derivatives; that is, to differentiate a function twice. Interestingly though, we now have many different ways of computing a second derivative. For example, if $f : \mathbb{R}^2 \to \mathbb{R}$ then there are four possible second derivatives:

$$\partial_{xx} f = \frac{\partial}{\partial x}\left[\frac{\partial f}{\partial x}\right], \qquad \partial_{xy} f = \frac{\partial}{\partial x}\left[\frac{\partial f}{\partial y}\right], \qquad \partial_{yx} f = \frac{\partial}{\partial y}\left[\frac{\partial f}{\partial x}\right], \qquad \partial_{yy} f = \frac{\partial}{\partial y}\left[\frac{\partial f}{\partial y}\right].$$

The terms $\partial_{xx} f, \partial_{yy} f$ are called *pure partial derivatives*, while $\partial_{xy} f, \partial_{yx} f$ are called *mixed partial derivatives*. In general, given a function $f : \mathbb{R}^n \to \mathbb{R}$, there are n^2 different second-order partial derivatives.

Example 2.35

Determine the second-order partial derivatives of the function $f(x, y) = e^{xy} + x^2 \sin(y)$.

Solution. This is a matter of straightforward computation. The first order partial derivatives are given by

$$\frac{\partial f}{\partial x} = ye^{xy} + 2x\sin(y), \qquad \frac{\partial f}{\partial y} = xe^{xy} + x^2\cos(y).$$

To compute the second order partials, we treat each of the first order partials as functions of x and y and repeat the process:

$$\begin{aligned} \partial_{xx} f &= y^2 e^{xy} + 2\sin(y) & \partial_{xy} f &= e^{xy} + xye^{xy} + 2x\cos(y) \\ \partial_{yx} f &= e^{xy} + xye^{xy} + 2x\cos(y) & \partial_{yy} f &= x^2 e^{xy} - x^2\sin(y). \end{aligned}$$

Interestingly, note that $\partial_{yx} f = \partial_{xy} f$. ■

Example 2.36

Determine the second-order partial derivatives of the function $f(x, y) = e^{\cos(xy)}$.

Solution. The first order partial derivatives are given by

$$\partial_x f = -y\sin(xy)e^{\cos(xy)}, \qquad \partial_y f = -x\sin(xy)e^{\cos(xy)}.$$

The second order derivatives are given by

$$\begin{aligned}
\partial_{xx} f &= e^{\cos(xy)}\left(y^2\sin^2(xy) - y^2\cos(xy)\right)\\
\partial_{xy} f &= e^{\cos(xy)}\left(xy\sin^2(xy) - xy\cos(xy) - \sin(xy)\right)\\
\partial_{yx} f &= e^{\cos(xy)}\left(xy\sin^2(xy) - xy\cos(xy) - \sin(xy)\right)\\
\partial_{yy} f &= e^{\cos(xy)}\left(x^2\sin^2(xy) - x^2\cos(xy)\right).
\end{aligned}$$

Here we still have $\partial_{xy} f = \partial_{yx} f$. ■

The fact that $\partial_{xx} f = \partial_{yy} f$ in Example 2.36 is a consequence of the symmetry of the function $f(x, y) = e^{\cos(xy)}$. However, somewhat more surprising is that in both of the previous two examples our mixed partial derivatives were the same. It turns out that this is a fairly common occurrence.

Theorem 2.37: Clairut's Theorem

Let $f : \mathbb{R}^n \to \mathbb{R}$ be a function and $\mathbf{a} \in \mathbb{R}^n$ a point. Let $i, j \in \{1, \ldots, n\}$ with $i \neq j$. If $\partial_{ij} f(\mathbf{a})$ and $\partial_{ji} f(\mathbf{a})$ both exist and are continuous in a neighbourhood of $\mathbf{a}$, then $\partial_{ij} f(\mathbf{a}) = \partial_{ji} f(\mathbf{a})$.

This is a technical theorem, and to present a readable version of this proof will require some sort of sophistry (either making an argument about the ability to interchange limits, or an argument about the existence of points in the Mean Value Theorem). In either case, we encourage the student to think hard about this theorem, but to not worry about the proof. To make our lives a little bit easier, we introduce the following class of functions:

Definition 2.38

Let $U \subseteq \mathbb{R}^n$ be an open set. We define $C^2(U, \mathbb{R})$ to be the collection of $f : \mathbb{R}^n \to \mathbb{R}$ whose second partial derivatives exist and are continuous at every point in U.

If f is a C^2 function, Clairut's theorem immediately imply that it's mixed partial derivatives exist, are continuous, and hence are equal.

2.4.2 The Chain Rule

Despite having constantly and consistently cautioned against treating differentials as fractions, there have not been too many instances to date where ignoring this advice could have caused any damage. Here at last our efforts will be vindicated, as we show the student some of the deeper subtleties in using higher-order partial derivatives in conjunction with the chain rule.

Let's start with a simple but general example. To make a point, we will write all partial derivatives using Leibniz notation. Let $u = f(x, y)$ and suppose that both x, y are functions of (s, t); that is, $x(s, t)$ and $y(s, t)$. Let's say that we wish to compute $\frac{\partial^2 u}{\partial s^2}$. Using the chain rule, we can find the first order partial as

$$\frac{\partial u}{\partial s} = \frac{\partial u}{\partial x}\frac{\partial x}{\partial s} + \frac{\partial u}{\partial y}\frac{\partial y}{\partial s}.$$

Next, we again take a partial derivative with respect to s, to get

$$\frac{\partial^2 u}{\partial s^2} = \frac{\partial}{\partial s}\left[\frac{\partial u}{\partial s}\right] = \frac{\partial}{\partial s}\left[\frac{\partial u}{\partial x}\frac{\partial x}{\partial s}\right] + \frac{\partial}{\partial s}\left[\frac{\partial u}{\partial y}\frac{\partial y}{\partial s}\right].$$

Now realize that since $u = f(x, y)$ is a function of x and y, $\frac{\partial u}{\partial x}$ is also a function of (x, y). Thus to differentiate this function with respect to s, we must once again use the chain rule. Thus looking at only the first summand, we have

$$\begin{aligned}
\frac{\partial}{\partial s}\left[\frac{\partial u}{\partial x}\frac{\partial x}{\partial s}\right] &= \left[\frac{\partial}{\partial s}\frac{\partial u}{\partial x}\right]\frac{\partial x}{\partial s} + \frac{\partial u}{\partial x}\frac{\partial^2 x}{\partial s^2} && \text{product rule} \\
&= \left[\frac{\partial^2 u}{\partial x^2}\frac{\partial x}{\partial s} + \frac{\partial^2 u}{\partial x \partial y}\frac{\partial y}{\partial s}\right]\frac{\partial x}{\partial s} + \frac{\partial u}{\partial x}\frac{\partial^2 x}{\partial s^2} && \text{chain rule} \\
&= \frac{\partial^2 u}{\partial x^2}\left[\frac{\partial x}{\partial s}\right]^2 + \frac{\partial^2 u}{\partial x \partial y}\frac{\partial y}{\partial s}\frac{\partial x}{\partial s} + \frac{\partial u}{\partial x}\frac{\partial^2 x}{\partial s^2}.
\end{aligned}$$

What a mess! A similar computation on the second summand yields

$$\frac{\partial}{\partial s}\left[\frac{\partial u}{\partial y}\frac{\partial y}{\partial s}\right] = \frac{\partial^2 u}{\partial y^2}\left[\frac{\partial y}{\partial s}\right]^2 + \frac{\partial^2 u}{\partial x \partial y}\frac{\partial y}{\partial s}\frac{\partial x}{\partial s} + \frac{\partial u}{\partial y}\frac{\partial^2 y}{\partial s^2}.$$

Putting everything together:

$$\frac{\partial^2 u}{\partial s^2} = \frac{\partial^2 u}{\partial x^2}\left[\frac{\partial x}{\partial s}\right]^2 + \frac{\partial^2 u}{\partial y^2}\left[\frac{\partial y}{\partial s}\right]^2 + 2\frac{\partial^2 u}{\partial x \partial y}\frac{\partial y}{\partial s}\frac{\partial x}{\partial s} + \frac{\partial u}{\partial x}\frac{\partial^2 x}{\partial s^2} + \frac{\partial u}{\partial y}\frac{\partial^2 y}{\partial s^2}. \tag{2.6}$$

This is only a single partial derivative. The same procedure must also be used to compute $\partial_{xy} u$ and $\partial_{yy} u$. These are left as exercises for the student.

Exercise: Hurt your brain a little bit more! Let $u = f(x, y, s)$ and $x(s, t)$ and $y(s, t)$. Now determine $\partial_{ss} u$.

2.4.3 Higher-Order Partials

We have limited our discussion to just second-order partial derivatives, in hopes that this simplest of cases would serve as a gentle introduction. Even in this case though, Equation (2.6) shows that things can get unpleasant very quickly. We begin by generalizing Clairut's theorem to higher dimensions.

Definition 2.39

If $U \subseteq \mathbb{R}^n$ is an open set, then for $k \in \mathbb{N}$ we define $C^k(U, \mathbb{R})$ to be the collection of functions $f : \mathbb{R}^n \to \mathbb{R}$ such that the k-th order partial derivatives of f all exist and are continuous on U. If the partials exist and are continuous for all k, we say that f is of type $C^\infty(U, \mathbb{R})$.

Theorem 2.40: Generalized Clairuit's Theorem

If $f : U \subseteq \mathbb{R}^n \to \mathbb{R}$ is of type C^k, then

$$\partial_{i_1,\ldots,i_k} f = \partial_{j_1,\ldots,j_k} f$$

whenever $(i_1, \ldots, i_k)$ and $(j_1, \ldots, j_k)$ are re-orderings of one another.

Notice that

$$C^k(U, \mathbb{R}) \subseteq C^{k-1}(U, \mathbb{R}) \subseteq C^{k-2}(U, \mathbb{R}) \subseteq \cdots \subseteq C^1(U, \mathbb{R}).$$

So in particular, if f is of type C^k, then we know that the mixed partials all agree up to and including order k.

Now let's make sure that we understand what Clairut's theorem is saying. For example, if $f : \mathbb{R}^3 \to \mathbb{R}$ is of type C^4, then the theorem does *not* say that all the fourth order derivatives are the same (there are 81 fourth order derivatives). Rather, the theorem says the partial derivatives of the same 'type' are equivalent:

$$\partial_{xxyz} f, \quad \partial_{xyxz} f, \quad \partial_{xyzx} f, \quad \partial_{yxxz} f, \quad \partial_{yxzx} f, \quad \partial_{yzxx} f,$$

$$\partial_{xxzy} f, \quad \partial_{xzxy} f, \quad \partial_{xzyx} f, \quad \partial_{zxxy} f, \quad \partial_{zxyx} f, \quad \partial_{zyxx} f.$$

The point being that every partial derivative above consists of exactly two x-derivatives, one y-derivative, and one z-derivative.

2.4.4 Multi-indices

When a function is of type C^k, then we know that in computing a k-th order derivative the order of the derivatives does not matter, only the total number of derivatives we take with respect to each variable. This suggests a very convenient notation. In the above example, we can write $(2, 1, 1)$ to capture the fact that we are differentiating the first variable twice, the second variable one, and the third variable once. This leads us to the notion of a multi-index.

A *multi-index* α is a tuple of non-negative integers $\alpha = (\alpha_1, \ldots, \alpha_n)$. The *order* of α is the sum of its components

$$|\alpha| = \alpha_1 + \alpha_2 + \cdots + \alpha_n.$$

We define the *multi-index factorial* to be

$$\alpha! = \alpha_1! \alpha_2! \cdots \alpha_n!.$$

If $\mathbf{x} = (x_1, \ldots, x_n) \in \mathbb{R}^n$ then the *multi-index exponential* is

$$\mathbf{x}^\alpha = x_1^{\alpha_1} x_2^{\alpha_2} \cdots x_n^{\alpha_n}$$

and if $f : \mathbb{R}^n \to \mathbb{R}$ we write

$$\partial^\alpha = \frac{\partial^{|\alpha|} f}{\partial x_1^{\alpha_1} \partial x_2^{\alpha_2} \cdots \partial x_n^{\alpha_n}}.$$

The multi-index factorial and exponential will be crucial pieces of notation in the following section. For now, we would like to capitalize on partial derivative notation. So for example, if $f : \mathbb{R}^4 \to \mathbb{R}$ and we endow $\mathbb{R}^4$ with the coordinates (x, y, z, w), then

$$\partial^{(0,0,0,1)} f = \partial_w f, \quad \partial^{(0,1,1,0)} f = \partial_{yz} f, \quad \partial^{(2,0,1,0)} f = \partial_{xxz} f, \quad \partial^{(0,1,2,1)} f = \partial_{yzzw} f.$$

et cetera.

2.5 Taylor Series

2.5.1 A Quick Review

Before talking about how multivariate Taylor series work, let's review what we learned in the single variable case. We have seen that the derivative can be used as a tool for linearly approximating a function. If f is differentiable at a point a, then for x near a we have the approximation

$$f(x) \approx f(a) + f'(a)(x - a).$$

Note that this is also sometimes written in terms of the distance $h = x - a$ from a, so that

$$f(a + h) \approx f(a) + f'(a)h.$$

Again, the top equation is a function of the absolute position x, while the bottom equation is a function of the relative distance h. The relationship between these two representations of Taylor series are akin to the two equivalent definitions for the derivative at a:

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

Now one can extend the conversation beyond just linear approximations, and introduce quadratic, cubic, and quartic approximations. More generally, given some $n \in \mathbb{N}$ we can set $p_{n,a}(x) = c_n x^n + c_{n-1} x^{n-1} + \cdots + c_1 x + c_0$ and ask what conditions on the c_k guarantee that $f^{(k)}(a) = p^{(k)}_{n,a}(a)$. This is a fairly straightforward exercise, and the student will find that

$$c_k = \frac{f^{(k)}(a)}{k!}, \quad \text{so that} \quad p_{n,a}(x) = \sum_{k=0}^{n} \frac{f^{(k)}(a)}{k!} (x - a)^k.$$

In order to ensure that this is a good approximation, we need to look at the error term $r_{n,a}(x) = f(x) - p_{n,a}(x)$. In particular, for $p_{n,a}(x)$ to represent a good k-th order approximation to f, we should require that the remainder tends to zero faster than k-th order; that is,

$$\lim_{x \to a} \frac{r_{n,a}(x)}{(x - a)^k} = 0.$$

There are many different approximations to $r_{n,a}(x)$, which vary depending on the regularity of the function (is f of type C^n or C^{n+1}?), or on the technique used to approximate the error. In general we will only be working with C^∞ functions, so we are not going to concern ourselves too much with regularity. It is quite a mess to introduce all of the technical approximations, so we content ourselves with only deriving a single one, called *Lagrange's form of the remainder*.

Lemma 2.41: Higher Order Rolle's Theorem

Assume that $f : \mathbb{R} \to \mathbb{R}$ is continuous on $[a, b]$ and $n + 1$ times differentiable on $[a, b]$. If $f(a) = f(b)$ and $f^{(k)}(a) = 0$ for all $k \in \{1, \ldots, n\}$ then there exists a $c \in (a, b)$ such that $f^{(n+1)}(c) = 0$.

Proof. All of the conditions of Rolle's theorem apply with $f(a) = f(b)$, so there exists a $\theta_1 \in (a, b)$ such that $f'(\theta_1) = 0$. Similarly, we know that f' is continuous on $[a, b]$ and differentiable on (a, b), and $f'(a) = f'(\theta_1) = 0$, so there exists $\theta_2 \in (a, \theta_1)$ such that $f''(\theta_2) = 0$. We can continue inductively in this fashion, until $f^{(n)}(a) = f^{(n)}(\theta_k)$, so that there exists $c := \theta_{n+1} \in (a, \theta_n) \subseteq (a, b)$ such that $f^{(n+1)}(c) = 0$, as required. □

Theorem 2.42: Taylor's Theorem with Lagrange Remainder

Suppose that f is $n+1$ times differentiable on an interval I with $a \in I$. For each $x \in I$ there is a point c between a and x such that

$$r_{n,a}(x) = \frac{f^{(n+1)}(c)}{(n+1)!}(x-a)^{n+1}. \tag{2.7}$$

Proof. Assume for the moment that $x > a$ and define the function

$$g(t) = r_{n,a}(t) - r_{n,a}(x)\frac{(t-a)^{n+1}}{(x-a)^{n+1}}$$

so that $g(a) = g(x) = 0$. Writing $r_{n,a}(t) = f(t) - p_{n,a}(t)$ we have

$$g(t) = f(t) - f'(a)(t-a) - \frac{f''(a)}{2}(t-a)^2 - \cdots - \frac{f^{(n)}(a)}{n!}(t-a)^n - r_{n,a}(x)\frac{(t-a)^{n+1}}{(x-a)^{n+1}}.$$

It is straightforward to check that

$$g^{(k)}(t) = f^{(k)}(t) - f^{(k)}(a) - f^{(k+1)}(a)(x-a) - \cdots - \frac{f^{(n)}(a)}{(n-k)!}(t-a)^{n-k} - r_{n,a}(x)\frac{(n+1)!}{(n+1-k)!}\frac{(t-a)^{n+1-k}}{(x-a)^{n+1}}$$

so that $g^{(k)}(a) = 0$ for all $k = 1, \ldots, n$. By the Higher Order Rolle's Theorem, we know there exists a $c \in (a, x)$ such that $g^{(n+1)}(c) = 0$, but this is precisely equivalent to

$$0 = g^{(n+1)}(c) = f^{(n+1)}(c) - r_{n,a}(x)\frac{(n+1)!}{(x-a)^{n+1}}$$

we we can re-arrange to get (2.7). □

Corollary 2.43

If f is of type C^{n+1} on an open interval I with $a \in I$, then

$$\lim_{x \to a} \frac{r_{n,a}(x)}{|x-a|^n} = 0.$$

Proof. Since f is of type C^{n+1} we know that $f^{(n+1)}$ is continuous on I. Since I is open and $a \in I$, we can find a closed interval J such that $a \in J \subseteq I$. By the Extreme Value Theorem, there exists $M > 0$ such that that $|f^{(n+1)}(x)| \leq M$ for all $x \in J$. Since f is $n+1$ times differentiable in a neighbourhood of a, Theorem 2.42 implies that

$$\begin{aligned}\lim_{x\to a} \frac{|r_{n,a}(x)|}{|x-a|^n} &= \lim_{x\to a} \frac{|f^{(n+1)}(c)|}{(n+1)!} \frac{|x-a|^{n+1}}{|x-a|^n} && c \text{ depends on } x \\ &= \lim_{x\to a} \frac{M}{(n+1)!}|x-a| \\ &= 0.\end{aligned}$$

The result then follows by applying the Squeeze Theorem to

$$-\frac{|r_{n,a}(x)|}{|x-a|^n} \leq \frac{r_{n,a}(x)}{|x-a|^n} \leq \frac{|r_{n,a}(x)|}{|x-a|^n}.$$

□

This corollary implies that the Taylor remainder is a good approximation, since the error vanishes faster than order n. Moreover, in the proof we found that we could bound $r_{n,a}(x)$ as

$$|r_{n,a}(x)| \leq \frac{M}{(n+1)!}|x-a|^{n+1} \tag{2.8}$$

for some $M > 0$. This allows us to determine error bounds on Taylor series.

Example 2.44

Let $f(x) = \sin(x)$ and $g(x) = e^x$. Determine the number of terms needed in the Taylor series to ensure that the Taylor polynomials at $a = 0$ are accurate to within 8 decimal places on $[-1, 1]$.

Solution. This is a problem you might have if you worked for a classical calculator company. If your calculator is only capable of holding eight significant digits then you need only ensure accuracy to eight digits, so you need to determine how many terms of the Taylor polynomial you need to program.

For $f(x) = \sin(x)$ we know that regardless of how many derivatives we take, $|f^{(k)}(x)| \leq 1$ for all x, and since we are looking at the interval $[-1, 1]$, we know that $|x - a| = |x| \leq 1$. Substituting this information into (2.8) we get that $|r_{n,a}(x)| \leq [(k+1)!]^{-1}$. We need to find a value of k such that $1/(k+1)! < 10^{-8}$. The student can check that this first happens when $k = 11$.

Similarly, for $g(x) = e^x$ we know that $g^{(k)}(x) = e^x$, and on the interval $[-1, 1]$ we can bound this as $|g^{(k)}(x)| < 3$. We still have $|x - a| < 1$, so (2.8) gives us $|r_{n,a}(x)| \leq 3[(k+1)!]^{-1}$, which also becomes smaller than 10^{-8} when $k = 11$. ■

2.5.2 Multivariate Taylor Series

Just like with the Multivariate Mean Value Theorem, we will introduce the multivariate Taylor Series by examining what happens when we restrict our function to a line. For simplicity, assume

that $S \subseteq \mathbb{R}^n$ is a convex set and choose some point $\mathbf{a} = (a^1, \ldots, a^n) \in S$ around which we will compute our Taylor series for $f : S \to \mathbb{R}$. Let $\mathbf{x}_0 = (x_0^1, \ldots, x_0^n) \in S$ be some point at which we want to compute $f(\mathbf{x}_0)$ and consider the line

$$\gamma(t) = (1-t)\mathbf{a} + t\mathbf{x}_0 = a + t(\mathbf{x}_0 - \mathbf{a}).$$

Pre-composing f by γ we get the function $g : \mathbb{R} \to \mathbb{R}, g(t) = f(\gamma(t))$. Notice that $g(0) = f(\mathbf{a})$ and $g(1) = f(\mathbf{x})$. Furthermore, since g is a real-valued function of a single variable, it admits a Taylor polynomial centred at 0, which can be evaluated at $t = 1$:

$$g(1) = \sum_{k=0}^{n} \frac{g^{(k)}(0)}{k!} + \text{remainder}. \tag{2.9}$$

Let's look at the derivatives of g. The first derivative is easily computed via the chain rule, and we get

$$g'(t) = (\mathbf{x}_0 - \mathbf{a}) \cdot \nabla f(\mathbf{a} + t(\mathbf{x}_0 - \mathbf{a})).$$

If we think of $\nabla = \left(\frac{\partial}{\partial x_1}, \ldots, \frac{\partial}{\partial x_n}\right)$ then we can define a new operator

$$(\mathbf{x}_0 - \mathbf{a}) \cdot \nabla = (x_0^1 - a^1)\frac{\partial}{\partial x_1} + \cdots + (x_0^n - a^n)\frac{\partial}{\partial x_n},$$

and $g'(t) = [(\mathbf{x}_0 - \mathbf{a}) \cdot \nabla] f(\mathbf{a} + t(\mathbf{x}_0 - \mathbf{a}))$. Differentiating k-times in general will give us

$$g^{(k)}(t) = [(\mathbf{x}_0 - \mathbf{a}) \cdot \nabla]^k f(\mathbf{a} + t(\mathbf{x}_0 - \mathbf{a})).$$

Substituting this into (2.9) and evaluating at $t = 0$ we have

$$f(\mathbf{x}) = \sum_{k=0}^{n} \frac{[(\mathbf{x}_0 - \mathbf{a}) \cdot \nabla]^k f(\mathbf{a})}{k!}.$$

This is theoretically complete, but computationally quite messy. Let's see if we can get a better grip on what these operators $[(\mathbf{x}_0 - \mathbf{a}) \cdot \nabla]^k$ look like. For the sake determining what this looks like, let $n = 2$ and $\mathbf{a} = (0, 0)$, so that

$$\begin{aligned}
[(\mathbf{x}_0 - \mathbf{a}) \cdot \nabla]^2 f &= [\mathbf{x}_0 \cdot \nabla]\,[x_0 f_x + y_0 f_y] \\
&= x_0\,[f_{xx} + y_0 f_{xy}] + y_0\,[f_{yx} + y_0 f_{yy}] \\
&= x_0^2 f_{xx} + x_0 y_0 f_{xy} + y_0 x_0 f_{yx} + y_0^2 f_{yy} \\
&= \mathbf{x}_0^{(2,0)} \partial^{(2,0)} f + 2\mathbf{x}_0^{(1,1)} \partial^{(1,1)} f + \mathbf{x}_0^{(0,2)} \partial^{(0,2)} f.
\end{aligned}$$

Notice that we get a perfect correspondence between the coefficient and the derivatives. For example, the coefficient of f_{yx} is $y_0 x_0$. The last line is written in multi-index notation, where the order of every multi-index in 2. One can imagine this also works for general n and general $\mathbf{a}$, so that

$$[(\mathbf{x}_0 - \mathbf{a}) \cdot \nabla]^k f = \sum_{|\alpha| = k} \frac{k!}{\alpha!} \left(\partial^\alpha f\right)(\mathbf{a}) \left(\mathbf{x}_0 - \mathbf{a}\right)^\alpha.$$

In conclusion, the equation for our multivariate Taylor polynomial is given by

Multivariate Taylor Polynomial

$$f(\mathbf{x}) = \sum_{|\alpha| \leq n} \frac{(\partial^\alpha f)(\mathbf{a})}{\alpha!} (\mathbf{x} - \mathbf{a})^\alpha + r_{n,\mathbf{a}}(\mathbf{x})$$

Example 2.45

Determine the 2nd order Taylor polynomial for $f(x, y) = \sin(x^2 + y^2)$ about $\mathbf{a} = (0, 0)$.

Solution. We have collected the data in a handy table below:

$\lvert\alpha\rvert$	α	$\alpha!$	$(\mathbf{x} - \mathbf{a})^\alpha$	$\partial^\alpha f$	$\partial^\alpha f(\mathbf{a})$
0	$(0,0)$	1	1	$\sin(x^2 + y^2)$	0
1	$(1,0)$	1	x	$2x\cos(x^2 + y^2)$	0
1	$(0,1)$	1	y	$2y\cos(x^2 + y^2)$	0
2	$(2,0)$	2	x^2	$2\cos(x^2 + y^2) - 4x^2\sin(x^2 + y^2)$	2
2	$(0,2)$	2	y^2	$2\cos(x^2 + y^2) - 4y^2\sin(x^2 + y^2)$	2
2	$(1,1)$	1	xy	$-4xy\sin(x^2 + y^2)$	0

Putting this information together, we get the relatively simple Taylor polynomial $\sin(x^2 + y^2) \approx x^2 + y^2$. ■

Example 2.46

Determine the 2nd order Taylor polynomial for $f(x, y) = xe^y$ at $\mathbf{a} = (0, 0)$.

Solution. Once again, we collate the data in the following table:

$\lvert\alpha\rvert$	α	$\alpha!$	$(\mathbf{x} - \mathbf{a})^\alpha$	$\partial^\alpha f$	$\partial^\alpha f(\mathbf{a})$
0	$(0,0)$	1	1	xe^y	0
1	$(1,0)$	1	x	e^y	1
1	$(0,1)$	1	y	xe^y	0
2	$(2,0)$	2	x^2	0	0
2	$(0,2)$	2	x^2	xe^y	0
2	$(1,1)$	1	xy	e^y	1

which gives us the Taylor polynomial $xe^y \approx x + xy$. ■

Something interesting is happening here: We know that the Taylor series for e^x and $\sin(x)$ are

$$e^x = \sum_{k=0}^{\infty} \frac{x^k}{k!}, \qquad \sin(x) = \sum_{k=0}^{\infty} \frac{(-1)^k x^{2k+1}}{(2k+1)!}.$$

It is tempting to substitute the appropriate polynomials in (x, y) into these expressions:

$$\begin{aligned} xe^y &= x\left[\sum_{k=0}^{\infty}\frac{y^k}{k!}\right] = \sum_{k=0}^{\infty}\frac{xy^k}{k!} \\ &= \left[x + xy + \frac{xy^2}{2} + \frac{xy^3}{3!} + \cdots\right] \\ \sin(x^2+y^2) &= \sum_{k=0}^{\infty}\frac{(-1)^k(x^2+y^2)^{2k+1}}{(2k+1)!} \\ &= \left[(x^2+y^2) - \frac{(x^2+y^2)^3}{3!} + \cdots\right]. \end{aligned}$$

Notice that to second order, these series both agree with what we computed above. Indeed, these are the correct Taylor series. This follows from the fact that Taylor polynomials are *unique*; that is, if we have an order k polynomial approximation to a function whose error vanishes in order $k+1$, then that polynomial is necessarily the Taylor polynomial. This also immediately implies that the Taylor series of any polynomial is that polynomial itself.

2.5.3 The Hessian Matrix

We know that if $f : \mathbb{R}^n \to \mathbb{R}$ is at least class C^2, then there are n^2 second-order partial derivatives $\partial_{ij}f, i, j \in \{1, \ldots, n\}$. Moreover, the mixed partial derivatives commute, so that $\partial_{ij}f = \partial_{ji}f$. This information can all be conveniently written in terms of a matrix:

Definition 2.47

If $f : \mathbb{R}^n \to \mathbb{R}$ is of class C^2 then the *Hessian matrix* of f at $\mathbf{a} \in \mathbb{R}^n$ is the symmetric $n \times n$-matrix of second order partial derivatives:

$$H(\mathbf{a}) = \begin{pmatrix} \partial_{11}f(\mathbf{a}) & \partial_{12}f(\mathbf{a}) & \cdots & \partial_{1n}f(\mathbf{a}) \\ \partial_{21}f(\mathbf{a}) & \partial_{22}f(\mathbf{a}) & \cdots & \partial_{2n}f(\mathbf{a}) \\ \vdots & \vdots & \ddots & \vdots \\ \partial_{n1}f(\mathbf{a}) & \partial_{n2}f(\mathbf{a}) & \cdots & \partial_{nn}f(\mathbf{a}) \end{pmatrix}.$$

The Hessian matrix makes writing down the Taylor series of a function very compact. Notice that the first order terms of the Taylor expansion are given by

$$\sum_{|\alpha|=1}\frac{1}{\alpha!}\left(\partial^\alpha f\right)(\mathbf{a})(\mathbf{x}_0 - \mathbf{a})^\alpha = \nabla f(\mathbf{a}) \cdot (\mathbf{x}_0 - \mathbf{a}).$$

Similarly, the second order terms involve the second-order partials and can be written as

$$\sum_{|\alpha|=2}\frac{2}{\alpha!}\left(\partial^\alpha f\right)(\mathbf{a})(\mathbf{x}_0 - \mathbf{a})^\alpha = (\mathbf{x}_0 - \mathbf{a})^T H(\mathbf{a})(\mathbf{x}_0 - \mathbf{a}),$$

so that the second-order Taylor polynomial is just

$$f(\mathbf{x}) = f(\mathbf{a}) + \nabla f(\mathbf{a})(\mathbf{x} - \mathbf{a}) + \frac{1}{2}(\mathbf{x} - \mathbf{a})^T H(\mathbf{a})(\mathbf{x} - \mathbf{a}) + O(\|\mathbf{x}^3\|).$$

Example 2.48

Determine the Hessian of the function $f(x,y,z) = x^2y + e^{yz}$ at the point $(1,1,0)$.

Solution. We start be computing the gradient $\nabla f = (2xy, x^2 + ze^{yz}, ye^{yz})$. The Hessian is now the matrix of second order partial derivatives, and may be computed as

$$H(x,y,z) = \begin{pmatrix} 2y & 2x & 0 \\ 2x & z^2e^{yz} & e^{yz}(1+zy) \\ 0 & e^{yz}(1+zy) & y^2e^{yz} \end{pmatrix}.$$

Evaluating at the point $(x,y,z) = (1,1,0)$ we get

$$H(1,1,0) = \begin{pmatrix} 2 & 2 & 0 \\ 2 & 0 & 1 \\ 0 & 1 & 1 \end{pmatrix}.$$

We can take one extra step and evaluate the gradient at this point $\nabla f(1,1,0) = (2,1,1)$, and write down the Taylor series:

$$\begin{aligned} f(\mathbf{x}) &= f(1,1,0) + \nabla f(1,1,0) \cdot \begin{pmatrix} x-1 \\ y-1 \\ z \end{pmatrix} + \frac{1}{2}(x-1, y-1, z) H(1,1,0) \begin{pmatrix} x-1 \\ y-1 \\ z \end{pmatrix} + O(\|\mathbf{x}\|^3) \\ &= 2 + (2,1,1) \begin{pmatrix} x-1 \\ y-1 \\ z \end{pmatrix} + \frac{1}{2}(x-1, y-1, z) \begin{pmatrix} 2 & 2 & 0 \\ 2 & 0 & 1 \\ 0 & 1 & 1 \end{pmatrix} \begin{pmatrix} x-1 \\ y-1 \\ z \end{pmatrix} + O(\|\mathbf{x}\|^3) \\ &= 2 + 2(x-1) + (y-1) + z + (x-1)^2 + 2(x-1)(y-1) + (y-1)z + \frac{1}{2}z^2 + O(\|\mathbf{x}\|^3). \blacksquare \end{aligned}$$

We can make even further strides if we allow ourselves to import a powerful theorem from linear algebra:

Theorem 2.49: Spectral Theorem

If $A : \mathbb{R}^n \to \mathbb{R}^n$ is a symmetric matrix then there exists an orthonormal basis consisting of eigenvectors of A.

Writing A in the basis guaranteed by the Spectral Theorem is called the *eigendecomposition* of A. In the eigendecomposition, the matrix A is a diagonal matrix with the eigenvalues on the diagonal. We will make use of the spectral theorem in the following section.

2.6 Optimization

When dealing with differentiable real-valued functions of a single variable $f : [a,b] \to \mathbb{R}$ we had a standard procedure for determining maxima and minima. This amounted to checking critical points on the interior (a,b) and then checking the boundary points. The necessity of checking the boundary separately arose from the non-differentiability of the function at the boundary. In the

multiple dimension regime, we will now be looking at functions $f : S \subseteq \mathbb{R}^n \to \mathbb{R}$. Once again, we will use differentiability to establish a necessary condition for extrema to occur on the interior, and check the boundary separately. However, unlike the former example where the boundary consisted of two points $\{a, b\}$, in multiple dimensions our boundaries become much larger. This will necessitate and entirely different approach to determining maxima on the boundary.

For now, we recall the definition of what it means to be a local maximum and minimum.

Definition 2.50

Let $f : \mathbb{R}^n \to \mathbb{R}$.

1. We say that $\mathbf{a} \in \mathbb{R}^n$ is a *local maximum* of f if there exists a neighbourhood $U \subseteq \mathbb{R}^n$ containing $\mathbf{a}$ such that $f(\mathbf{x}) \leq f(\mathbf{a})$ for all $\mathbf{x} \in U$.
2. We say that $\mathbf{a} \in \mathbb{R}^n$ is a *local minimum* of f if there exists a neighbourhood $U \subseteq \mathbb{R}^n$ containing $\mathbf{a}$ such that $f(\mathbf{x}) \geq f(\mathbf{a})$ for all $\mathbf{x} \in U$.

When $n = 1$ this is exactly our usual definition of a maximum/minimum point.

2.6.1 Critical Points

Definition 2.51

If $f : \mathbb{R}^n \to \mathbb{R}$ is differentiable, we say that $\mathbf{c} \in \mathbb{R}^n$ is a *critical point* of f if $\nabla f(\mathbf{c}) = 0$.[a] If $\mathbf{c}$ is a critical point, we say that $f(\mathbf{c})$ is a *critical value*. All points which are not critical are termed *regular points*.

[a] More generally, if $\mathbf{f} : \mathbb{R}^n \to \mathbb{R}^k$ then we say that $\mathbf{c} \in \mathbb{R}^n$ is a critical point if $D\mathbf{f}(\mathbf{c})$ does not have maximal rank.

We see that the above definition of a critical point agrees with the our usual definition when $n = 1$; namely, that $f'(c) = 0$.

Example 2.52

Determine the critical points of the following functions:

$$f(x, y) = x^3 + y^3, \qquad g(x, y, z) = xy + xz + x$$

Solution. The gradient of f is easily determined to be $\nabla f(x, y) = (3x^2, 3y^2)$. Setting this to be $(0, 0)$ implies that $3x^2 = 0 = 3y^2$ so that the only critical point is $(x, y) = 0$. For the function g we compute $\nabla g(x, y, z) = (y + z + 1, x, x)$. Setting this equal to zero implies that $x = 0$ while $y + z + 1 = 0$. Thus there is an entire line worth of critical points. ■

Notice that critical points do not need to be isolated: one can have entire curves or planes represent critical points. The important property of critical points is that they give a schema for determining when a point is a maximum or minimum, through the following theorem:

Proposition 2.53

If $f : [a, b] \to \mathbb{R}$ is differentiable and c is interior point which is either a local maximum or local minimum, then necessarily $f'(c) = 0$.

Proof. We shall do the proof for the case when c corresponds to a local maximum and leave the proof of the other case to the student. Since c is a local maximum, we know there is some neighbourhood $I \subseteq D$ of c such that for all $x \in I, f(x) \leq f(c)$.

Since c corresponds to a maximum of f, for all $h > 0$ sufficiently small so that $c + h \in I$, we have that $f(c+h) \leq f(c)$. Hence $f(c+h) - f(c) \leq 0$, and since h is positive, the difference quotient satisfies $\frac{f(c+h)-f(c)}{h} \leq 0$. In the limit as $h \to 0^+$ we thus have

$$\lim_{h\to 0^+} \frac{f(c+h) - f(c)}{h} \leq 0 \tag{2.10}$$

Similarly, if $h < 0$ we still have $f(c + h) - f(c) \leq 0$ but now with a negative denominator our difference quotient is non-negative and

$$\lim_{h\to 0^-} \frac{f(c+h) - f(c)}{h} \geq 0. \tag{2.11}$$

Combining (2.10) and (2.11) and using the fact that f is differentiable at c, we have

$$0 \leq \lim_{h\to 0^-} \frac{f(c+h) - f(c)}{h} = f'(c) = \lim_{h\to 0^+} \frac{f(c+h) - f(c)}{h} \leq 0$$

which implies that $f'(c) = 0$. □

Of course, we know that this proposition is only necessary, not sufficient; that is, there are critical points which do not yield extrema. The quintessential example is the function $f(x) = x^3$, which has a critical point at $x = 0$, despite this point being neither a maximum nor minimum. A more interesting example, which we leave for the student, is the function $f(x) = x \sin(x)$, which has infinitely many critical points but no local maxima or minima.

Our theme for the last several sections has been to adapt our single-variable theorems to multivariate theorems by examining the behaviour of functions through a line. This part will be no different.

Corollary 2.54

Let $U \subseteq \mathbb{R}^n$. If $f : U \to \mathbb{R}$ is differentiable and $\mathbf{c} \in U$ is either a local maximum or minimum of f, then $\nabla f(\mathbf{c}) = \mathbf{0}$.

Proof. We do the case where $\mathbf{c}$ is a maximum and leave the other case as an exercise. Since $\mathbf{c}$ is a maximum, we know there is a neighbourhood $U \subseteq \mathbb{R}^n$ containing $\mathbf{c}$ such that $f(\mathbf{x}) \leq f(\mathbf{c})$ for all $\mathbf{x} \in U$. Since this holds in general, it certainly holds locally along any line through $\mathbf{c}$; that is, for any unit vector $\mathbf{u} \in \mathbb{R}^n$ there exists $\epsilon > 0$ such that for all $t \in (-\epsilon, \epsilon)$, we have

$$g(t) := f(\mathbf{c} + t\mathbf{u}) \leq f(\mathbf{c}).$$

Since g attains its maximum at $t = 0$ (an interior point), Proposition 2.53 implies that $g'(0) = 0$. Using the chain rule, this implies that $\nabla f(\mathbf{c}) \cdot \mathbf{u} = 0$. This holds for all unit vectors $\mathbf{u}$, so in particular if we let $\mathbf{u} = \mathbf{e}_i = (0, \dots, 1, \dots, 0)$ be one of the standard unit normal vectors, then

$$0 = \nabla f(\mathbf{c}) \cdot \mathbf{e}_i = \partial_{x_i} f(\mathbf{c}).$$

This holds for every standard unit vector, so $\nabla f(\mathbf{c}) = 0$. □

Once again, this theorem will be necessary, but not sufficient. For example, consider the functions $f_1(x, y) = x^2 + y^2$ and $f_2(x, y) = y^2 - x^2$. Both function have critical points at $(x, y) = (0, 0)$, however the former is a minimum while the later is not. In particular, the latter function gives an example of a *saddle point*. Graphing functions is a terrible way to determine maxima and minima though, so we need to develop another criteria for determining extrema. This comes in the form of the second derivative test.

Proposition 2.55

Let $f : \mathbb{R}^n \to \mathbb{R}$ be class C^2 in a neighbourhood of a critical point $\mathbf{c}$.

1. If $H(\mathbf{c})$ has all positive eigenvalues, then $\mathbf{c}$ is a local minimum,
2. If $H(\mathbf{c})$ has all negative eigenvalues, then $\mathbf{c}$ is a local maximum,
3. If $H(\mathbf{c})$ has a mix of positive and negative eigenvalues, then $\mathbf{c}$ is a saddle point.

We will not give the proof of this proposition, but instead present a heuristic which essentially captures the idea of the proof. Recall from our discussion at the end of 2.5.3 that $H(\mathbf{c})$ admits an eigendecomposition with eigenvectors $\{\lambda_i\}_{i=1}^n$. As $\nabla f(\mathbf{c}) = 0$, the second-order Taylor polynomial tells us that in this basis, our function is approximately

$$f(\mathbf{x}) = f(\mathbf{c}) + (\mathbf{x} - \mathbf{c})^T H(\mathbf{c})(\mathbf{x} - \mathbf{c}) = f(\mathbf{c}) + \sum_{i=1}^{n} \lambda_i (x_i - c_i)^2.$$

If all of the eigenvalues are positive, this function is approximately an upward facing paraboloid centered at $\mathbf{c}$, meaning that it has a minimum. Similarly, if all the eigenvalues are negative, it is a downward facing paraboloid and hence $\mathbf{c}$ is a maximum. In the case where the λ_i are of mixed sign, we have that the function looks like a maximum in the direction of the eigenvectors corresponding to positive eigenvalues, and a minimum in the direction of eigenvectors corresponding to negative eigenvalues, and hence is a saddle point.

Example 2.56

Determine the critical points of the function $f(x, y) = x^4 - 2x^2 + y^3 - 6y$ and classify each as a maxima, minima, or saddle point.

Solution. The gradient can be quickly computed to be $\nabla f(x, y) = (4x(x^2 - 1), 3(y^2 - 2))$. The first component is zero when $x = 0, \pm 1$ and the second component is zero when $y = \pm\sqrt{2}$, giving six

critical points: $(0, \pm\sqrt{2}), (-1, \pm\sqrt{2})$, and $(1, \pm\sqrt{2})$. The Hessian is easily computed to be

$$H(x, y) = \begin{pmatrix} 12x^2 - 4 & 0 \\ 0 & 6y \end{pmatrix}.$$

Since the matrix is diagonal, its eigenvalues are exactly the $12x^2 - 4$ and $6y$. Thus the maximum is $(0, -\sqrt{2})$, the minima are $(\pm 1, \sqrt{2})$, and the other three points are saddles. ■

There is one additional kind of critical point which can appear. The above discussion of maxima, minima, and saddle points amounted to the function looking as though it had either a maximum or a minimum in every direction, and whether or not those directions all agreed with one another. This has not yet captured the idea of an inflection point.

Definition 2.57

If $f : \mathbb{R}^n \to \mathbb{R}$ is C^2 and $\mathbf{c}$ is a critical point of f, then we say that $\mathbf{c}$ is a *degenerate critical point* if $\mathbf{f}$ is $\operatorname{rank} H(\mathbf{c}) < n$.

Example 2.58

Show that the function $f(x, y) = y^2 - x^3$ has a degenerate critical point at $(x, y) = (0, 0)$.

Solution. The gradient is $\nabla f(x, y) = (-3x^2, 2y)$ which indeed has a critical point at $(0, 0)$. Furthermore, the Hessian is

$$H(x, y) = \begin{pmatrix} -6x & 0 \\ 0 & 2 \end{pmatrix}, \quad \Rightarrow \quad H(0, 0) = \begin{pmatrix} 0 & 0 \\ 0 & 2 \end{pmatrix}$$

so $H(0, 0)$ has rank 1, and we conclude that $(0, 0)$ is a degenerate critical point. ■

In the special case of function $f : \mathbb{R}^n \to \mathbb{R}$, one can use the determinant of the Hessian to quickly ascertain whether critical points are maxima or minima.

Proposition 2.59

Let $f : \mathbb{R}^2 \to \mathbb{R}$ and $\mathbf{c}$ be a critical point.

1. If $\det H(\mathbf{c}) < 0$ then $\mathbf{c}$ is a saddle point
2. If $\det H(\mathbf{c}) > 0$:
   (a) If $\partial_{11} f > 0$ then $\mathbf{c}$ is a minimum,
   (b) If $\partial_{11} f < 0$ then $\mathbf{c}$ is a maximum.

If $\det H(\mathbf{c}) = 0$ then this is inconclusive.

Proof. For any matrix, the determinant is the product of the eigenvalues (this follows immediately from the spectral theorem and the fact that the determinant is basis independent). Since $f : \mathbb{R}^2 \to \mathbb{R}$

the Hessian has two eigenvalues. If the determinant is negative, this means that the two eigenvalues have different signs and hence the critical point is a saddle point. If the determinant is positive, both eigenvalues have the same sign, and we need only determine if both are positive or negative. This last check can be done by looking at $\partial_{11} f$. □

2.6.2 Constrained Optimization

The previous section introduced the notion of critical points, which can be used to determine maxima/minima on the interior of a set. However, what happens when we are given a set with empty interior? Similarly, if one is told to optimize over a compact set, it is not sufficient to only optimize over the interior, one must also check the boundary.

We have seen problems of constrained optimization before. A typical example might consist of something along the lines of

> "You are building a fenced rectangular pasture, with one edge located along a river. Given that you have 200m of fencing, find the dimensions which maximize the volume of the pasture."

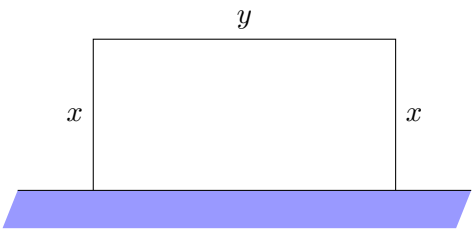


Figure 23: A visualization of simple optimization problem.

Translating this problem into mathematics, we let x be the length and y be the width of the pasture. We must then maximize the function $f(x, y) = xy$ subject to the constraint $2x + y = 200$. The equation $2x + y = 200$ is a line in $\mathbb{R}^2$, so we are being asked to determine the maximum value of the function f along this line. The way that this was typically handled in first year was to use the constraint to rewrite one variable in terms of another, and use this to reduce our function to a single variable. For example, if we write $y = 200 - 2x$ then

$$f(x, y) = x(200 - 2x) = 200x - 2x^2.$$

The lone critical point of this function occurs at $x = 50$, which gives a value of $y = 100$, and one can quickly check that this is the max.

Another technique that one could employ is the following: Recognizing that $2x + y = 200$ is just a line in $\mathbb{R}^2$, we can parameterize that line by a function $\gamma(t) = (t, 200 - 2t)$. The composition $f \circ \gamma$ is now a function in terms of the independent parameter t, yielding $f(\gamma(t)) = 200t - 2t^2$ which of course gives the same answer.

The fact that our constraint was just a simple line made this problem exceptionally simple. What if we wanted to optimize over a more difficult one-dimensional space, or even a two dimensional surface? Once again we can try to emulate the procedures above, and we may even meet

with some success. However, there is a more novel way of approaching such problems, using the method of *Lagrange multipliers*.

Theorem 2.60

Let $f, G : \mathbb{R}^n \to \mathbb{R}$ be C^1 functions, and set $S = G^{-1}(0)$. If the restriction $f : S \to \mathbb{R}$ has a maximum or minimum at a point $\mathbf{c} \in S$ and $\nabla G(\mathbf{c}) \neq 0$ then there exists $\lambda \in \mathbb{R}$ such that

$$\nabla f(\mathbf{c}) = \lambda \nabla G(\mathbf{c}).$$

Proof. Let $\gamma : (-\epsilon, \epsilon) \to S$ be any path such that $\gamma(0) = \mathbf{c}$, so that $\gamma'(0)$ is a vector which is tangent to S at $\mathbf{c}$. Since $\gamma(t) \in S$ for all $t \in (-\epsilon, \epsilon)$, by the definition of S we must have $G(\gamma(t)) = 0$. Differentiating at $t = 0$ yields the identity

$$0 = \nabla G(\mathbf{c}) \cdot \gamma'(0).$$

On the other hand, since $\mathbf{c}$ is a local maximum/minimum of f we have that $t = 0$ is a local maximum/minimum for $f(\gamma(t))$ and hence is a critical point. Using the chain rule, this implies that

$$0 = \frac{\mathrm{d}}{\mathrm{d}t}\bigg|_{t=0} f(\gamma(t)) = \nabla f(\mathbf{c}) \cdot \gamma'(0).$$

Since $\gamma'(0)$ can be chosen arbitrarily, this implies that both $\nabla G(\mathbf{c})$ and $\nabla f(\mathbf{c})$ are perpendicular to tangent plane at $\mathbf{c}$, and thus they must be proportional[4]; that is, there exists some $\lambda \in \mathbb{R}$ such that $\nabla f(\mathbf{c}) = \lambda \nabla G(\mathbf{c})$ as required. □

Example 2.61

Use the method of Lagrange multipliers to solve the problem given in Figure 23.

Solution. The constraint in our fencing problem is given by the function $G(x, y) = 2x + y - 200 = 0$. We can easily compute $\nabla f(x, y) = (y, x)$ and $\nabla G(x, y) = (2, 1)$, so by the method of Lagrange multipliers, there exists $\lambda \in \mathbb{R}$ such that $\nabla f(x, y) = \lambda \nabla G(x, y)$; that is,

$$\begin{pmatrix} y \\ x \end{pmatrix} = \lambda \begin{pmatrix} 2 \\ 1 \end{pmatrix}.$$

We thus know that $y = 2\lambda, x = \lambda$, and substituting this into $2x - y = 200$ gives $4\lambda = 200$. Thus $\lambda = 50$, from which we conclude that $y = 2\lambda = 100$ and $x = \lambda = 50$ as required. ■

Example 2.62

Maximize the function $f(x, y, z) = xyz$ on the ellipsoid $x^2 + 2y^2 + 3z^2 = 1$.

[4] Here we are sweeping some stuff under the rug. In particular, one must believe us that since G is C^1 then $S = G^{-1}(0)$ is a 'smooth' surface, so that its tangent plane has dimension $n - 1$.

Solution. The constraint equation is given by $G(x,y,z) = x^2+2y^2+3z^2-1 = 0$. When we compute our gradients, the method of Lagrange multipliers gives the following system of equations:

$$\begin{aligned} yz &= 2\lambda x \\ xz &= 4\lambda y \\ xy &= 6\lambda z \end{aligned}$$

If we combine this with the constraint $x^2+2y^2+3z^2 = 1$ we have four equations in four unknowns, though all the equations are certainly non-linear! Herein we must be clever, and start manipulating our equations to try and solve for (x,y,z). Notice that if we play with the term xyz then depending on how we use the associativity of multiplication, we can get an additional set of conditions. For example

$$\begin{aligned} x(yz) &= x(2\lambda x) = 2\lambda x^2 \\ y(xz) &= y(4\lambda y) = 4\lambda y^2 \\ z(xy) &= z(6\lambda z) = 6\lambda z^2 \end{aligned}$$

and all of these must be equal. We can make a small simplification by removing a factor of 2 to get

$$\lambda x^2 = 2\lambda y^2 = 3\lambda z^2. \tag{2.12}$$

Case 1 ($\lambda = 0$): If $\lambda = 0$ then $yz = xz = xy = 0$. This immediately implies that two of x, y, or z must be zero, so $f(x,y,z) = xyz = 0$. If $x = y = 0$ then the constraint equation gives $\left(0,0,\pm\frac{1}{\sqrt{3}}\right)$. If $x = z = 0$ then $\left(0,\pm\frac{1}{\sqrt{2}},0\right)$ and if $y = z = 0$ then $(\pm 1,0,0)$. So all of these points give a result of $f(x,y,z) = 0$ and are candidates for maxima/minima.

Case 2 ($\lambda \neq 0$): If $\lambda \neq 0$ then we can divide (2.12) by λ to get that $x^2 = 2y^2 = 3z^2$. Substituting this into the constraint equation we get $1 = x^2+x^2+x^2 = 3x^2$ so that $x = \pm\frac{1}{\sqrt{3}}$, which we can use to find y and z. This gives us eight possible critical points corresponding to the following choice of signs:

$$x = \pm\frac{1}{\sqrt{3}}, \quad y = \pm\frac{1}{\sqrt{6}}, \quad z = \pm\frac{1}{3}.$$

There are only two possible values of f for these points, namely $f(x,y,z) = \pm\frac{1}{9\sqrt{2}}$. Since these are both either bigger than 0 or smaller than 0, these are the corresponding global maxima/minima of the function. ■

Example 2.63

Determine the maximum and minimum of the function $f(x,y) = x^2 + 2y^2$ on the disk $x^2+y^2 \leq 4$.

Solution. We begin by determining critical points on the interior. Here we have $\nabla f(x,y) = (2x, 4y)$ which can only be $(0,0)$ if $x = y = 0$. Here we have $f(0,0) = 0$.

Next we determine the extreme points on the boundary $x^2+y^2 = 4$, for which we set up the constraint function $G(x,y) = x^2+y^2-4$ with gradient $\nabla G(x,y) = (2x, 2y)$. Using the method of

Lagrange multipliers, we thus have

$$2x = 2\lambda x$$
$$4y = 2\lambda y$$

Case 1 ($x \neq 0$): If $x \neq 0$ then we can solve $2x = 2\lambda x$ to find that $\lambda = 1$. This implies that $y = 2y$ which is only possible if $y = 0$. Plugging this into the constraint gives $x^2 = 4$ so that $x = \pm 2$, so our candidate points are $(\pm 2, 0)$, which give values $f(\pm 2, 0) = 4$.

Case 2 ($y \neq 0$): If $y \neq 0$ then we can solve $4y = 2\lambda y$ to find that $\lambda = 2$. This implies that $2x = 4x$ which is only possible if $x = 0$. Solving the constraint equation thus gives the candidates $(0, \pm 2)$, which gives values $f(0, \pm 2) = 8$.

The case where $\lambda = 0$ gives no additional information. Hence we conclude that the minimum occurs at $(0, 0)$ with a value of $f(0, 0) = 0$, while the maximum occurs at the two points $(0, \pm 2)$ with a value of $f(0, \pm 2) = 8$. ■

If multiple constraints are given, the procedure is similar, except that we now need additional multipliers. More precisely, if $\mathbf{G} : \mathbb{R}^n \to \mathbb{R}^m$ is given by $\mathbf{G}(\mathbf{x}) = (G_1(\mathbf{x}), \ldots, G_m(\mathbf{x}))$, we set $S = G^{-1}(\mathbf{0})$, and we are tasked with optimizing $f : S \to \mathbb{R}$, then if $\mathbf{c} \in S$ is a maximum or minimum there exist $\lambda_1, \ldots \lambda_m \in \mathbb{R}$ such that

$$\nabla f(\mathbf{c}) = \sum_{i=1}^{m} \lambda_i \nabla G_i(\mathbf{c}).$$

3 Local Invertibility

Given the plethora of ways of defining functions, curves, or surfaces over $\mathbb{R}^n$, a natural question which arises is whether such characterizations are (locally) invertible. For example, given a function $f \in C^1(\mathbb{R}^2, \mathbb{R}^2)$, $(x, y) \mapsto (xy, xe^y)$, is there a differentiable function $f^{-1} : \mathbb{R}^2 \to \mathbb{R}^2$ which inverts it everywhere? If not, can we find a function which at least inverts it locally, or perhaps conditions which tell us which points are troublesome for inverting?

Alternatively, what if one is given the zero locus of a C^1 function $F(x, y) = 0$ and is asked to determine y as a function of x? What conditions guarantee that this is possible? This section is dedicated to elucidating this information.

3.1 Implicit Function Theorem

We begin by analyzing the latter case first; namely, given a C^1 function $\mathbf{F} : \mathbb{R}^{n+k} \to \mathbb{R}^k$, when can we solve the equation

$$\mathbf{F}(x_1, \ldots, x_n, y_1, \ldots, y_k) = \mathbf{0}$$

for the y_i as functions of the x_i? More precisely, do there exist C^1 functions $f_i : \mathbb{R}^n \to \mathbb{R}$ such that $y_i = f(x_1, \ldots, x_n)$. In this great of generality, it can be difficult to see the forest for the trees, so we treat the $k = 1$ case as a special example to glean some insight into the problem.

3.1.1 Scalar Valued Functions

Consider a function $F \in C^1(\mathbb{R}^n, \mathbb{R})$ and endow $\mathbb{R}^{n+1}$ with the coordinates $(x_1, \ldots, x_n, y)$, whose purpose is to make our exposition clear with regards to which variable is solved in terms of the other variables. Can we solve the equation $F(\mathbf{x}, y) = 0$ for y as a function of x? Alternatively, can we realize $F(\mathbf{x}, y) = 0$ as the graph of a function $y = f(\mathbf{x})$? Some simple examples suggest that the answer could be yes.

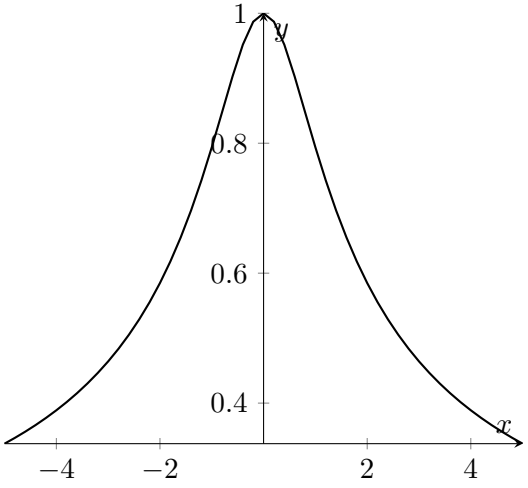
Example 3.1 Let $F : \mathbb{R}^2 \to \mathbb{R}$ be given by $F(x, y) = (x^2 + 1)y^3 - 1$. The zero-locus $F(x, y) = 0$ can be solved in terms of y to yield

$$y = \sqrt[3]{\frac{1}{x^2 + 1}},$$

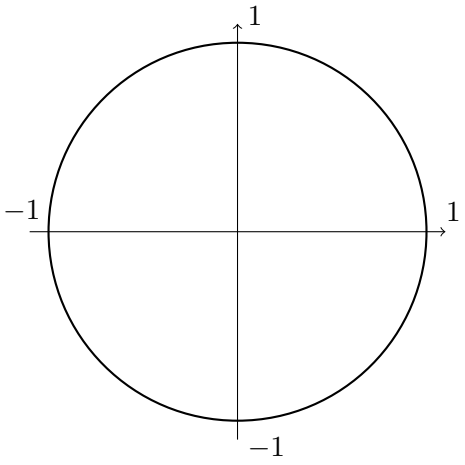
and this holds for all $x, y \in \mathbb{R}^2$. ▲

Unfortunately, it turns out that such examples are exceptionally rare and in general the answer is no:

Example 3.2 Let $F(x, y) = x^2 + y^2 - 1$. The zero-locus $F(x, y) = 0$ is equivalent to the circle $x^2 + y^2 = 1$. If one tries to solve y as a function of x, we get $y = \pm\sqrt{1 - x^2}$. In particular, for each x-value there are two possible y-values. Since functions must map a single input to a single output, this means that y *cannot* be written as a function of x. ▲



(a) A plot of the graph $(x^2+1)y^3 = 1$. It is easily to believe that this curve can be written as the graph of a function.



(b) The circle $x^2 + y^2 = 1$ cannot be written as the graph of a function: It fails the vertical line test.

The primary difference between Examples 3.1 and 3.2 is that the former was in a sense "injective" with respect to y (since y^3 is one-to-one) while the latter was not (y^2 is two-to-one). This example in hand, the situation seems rather dire: even such simple examples preclude the hope of solving one variable in terms of another. Nonetheless, one could argue that there are parts of the

circle $x^2+y^2=1$ that look like the graphs, one being $y=\sqrt{1-x^2}$ while the other is $y=-\sqrt{1-x^2}$. If it was our lofty goal of solving y as a function of x everywhere that presented a problem, perhaps by restricting ourselves to local solutions we might make more progress.

Since calculus is, in many ways, the study of functions by looking at their linear approximations, let's see what happens in the simplest non-trivial case where $F(\mathbf{x}, y)$ is linear:

$$F(x,y) = \alpha_1 x_1 + \cdots + \alpha_n x_n + \beta y_n + c = \sum_{i=1}^{n} \alpha_i x_i + \beta y + c.$$

In this case, it is easy to see that we can solve for y as a function of $\mathbf{x}$ so long as $\beta \neq 0$. Now recall that if $F(\mathbf{x}, y)$ is a (not necessarily linear) C^1 function, and $(\mathbf{a}, b)$ satisfies $F(\mathbf{a}, b) = 0$, then the equation of the tangent hyperplane at $(\mathbf{a}, b)$ is given by

$$\begin{aligned}&\frac{\partial F}{\partial x_1}(\mathbf{a},b)x_1 + \cdots + \frac{\partial F}{\partial x_n}(\mathbf{a},b)x_n + \frac{\partial F}{\partial y}(\mathbf{a},b)y + d \\ &= \nabla_x F(\mathbf{a},b)\cdot \mathbf{x} + \frac{\partial F}{\partial y}(\mathbf{a},b)y + d = 0\end{aligned}$$

for some constant d. By analogy, $\frac{\partial F}{\partial y}(\mathbf{a}, b)$ plays the role of β, which suggests that so long as $\frac{\partial F}{\partial y}(\mathbf{a}, b) \neq 0$, y should be solvable as a function of $\mathbf{x}$ in a neighbourhood of $(\mathbf{a}, b)$.

Aside: Recall that in single variable calculus, a continuously differentiable function $f : \mathbb{R} \to \mathbb{R}$ which satisfies $f'(x) \neq 0$ for all points x in a neighbourhood of a point p is injective on that neighbourhood. This is certainly compatible with our notion of "injectivity" above.

Theorem 3.3

If $F(\mathbf{x}, y)$ is C^1 on some neighbourhood $U \subseteq \mathbb{R}^{n+1}$ of the point $(\mathbf{a}, b) \in \mathbb{R}^{n+1}$, $F(\mathbf{a}, b) = 0$, and $\frac{\partial F}{\partial y}(\mathbf{a}, b) \neq 0$, then there exists an $r \in \mathbb{R}_{\geq 0}$ together with a unique C^1 function $f : B_{\mathbf{a}}(r) \to \mathbb{R}$ such that $F(\mathbf{x}, f(\mathbf{x})) = 0$ for all $x \in B_{\mathbf{a}}(r)$.

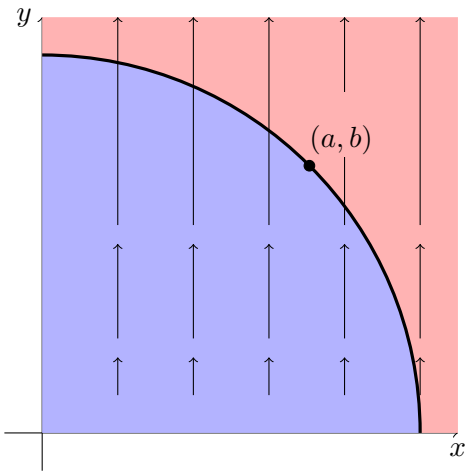
Proof. We break our proof into several steps: we begin by showing that there is an $r > 0$ such that for each $x_0 \in B_{\mathbf{a}}(r)$ there exists a unique y_0 such that $F(x_0, y_0) = 0$. We call the mapping which takes $x_0 \mapsto y_0$ the function $f(\mathbf{x}, y)$. After this, we show that this function is actually differentiable.

Existence and Uniqueness: This spirit of this part of the proof is actually akin to the proof of injectivity mentioned in the previous aside. Without loss of generality, assume that $\frac{\partial F}{\partial y}(\mathbf{a}, b) > 0$, so that there exists $r_1 > 0$ such that $\partial_y F > 0$ on the neighbourhood $B_{\mathbf{a},b}(r_1) \subseteq \mathbb{R}^{n+1}$. By taking smaller r_1 if necessary, we can ensure that $B_{\mathbf{a},b}(r) \subseteq U$.

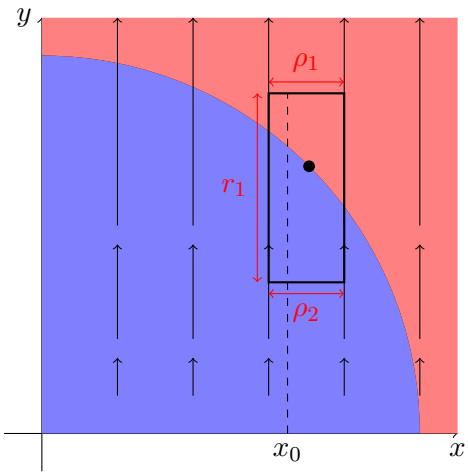
Now the positivity of $\partial_y F$ on $B_{\mathbf{a},b}(r)$ ensures that

$$F(\mathbf{a}, b - r_1) < 0, \qquad F(\mathbf{a}, b + r_1) > 0.$$

Once again, by continuity there exists $\rho_1, \rho_2 > 0$ such that $F(\mathbf{x}, b - r_1) < 0$ for all $\mathbf{x} \in B_{\mathbf{a}}(\rho_1)$ and $F(x, b + r_1) > 0$ for all $\mathbf{x} \in B_{\mathbf{a}}(\rho_2)$. Let $r = \min\{r_1, \rho_1, \rho_2\}$, so that for any fixed $\mathbf{x}_0 \in B_{\mathbf{a}}(r)$ we have $F(\mathbf{x}_0, b - r_1) < 0$ and $F(\mathbf{x}_0, b + r_1) > 0$. By the single variable Intermediate Value Theorem, there is at least one $y_0 \in B_b(r)$ such that $F(\mathbf{x}_0, y_0) = 0$. Furthermore, because $F(x_0, y)$ is strictly increasing as a function of y, this y is unique by the Mean Value Theorem.



(a) The graph of $F(x,y) = x^2 + y^2 - 1$, wherein the blue represents where $F(x,y) < 0$ and the red where $F(x,y) > 0$. The arrows are the values of $\frac{\partial F}{\partial y}$.



(b) Notice how the bottom of the rectangle lies entirely within the blue, and the top lies entirely within the red.

Differentiability: Fix some $x_0 \in B_{\mathbf{a}}(r)$, set $y_0 = f(x_0)$, and choose $h \in \mathbb{R}$ sufficiently small so that $\mathbf{h}_i = h\mathbf{e}_i$ satisfies $\mathbf{x}_0 + \mathbf{h}_i \in B_{\mathbf{a}}(r)$. Define $k = f(\mathbf{x}_0 + \mathbf{h}_i) - f(\mathbf{x}_0)$ to be the i^{th} difference quotient, so that $y_0 + k = f(\mathbf{x}_0 + \mathbf{h}_i)$. Now $F(\mathbf{x}_0 + \mathbf{h}_i, y_0 + k) = F(\mathbf{x}_0, y_0) = 0$ since both points lie in $B_{\mathbf{a}}(r)$, so by the Mean Value Theorem there exists some $t \in (0,1)$ such that

$$0 = F(\mathbf{x}_0 + \mathbf{h}_i, y_0 + k) - F(\mathbf{x}, y_0) = h\frac{\partial F}{\partial x_i}(\mathbf{x}_0 + t\mathbf{h}_i, y_0 + tk) + k\frac{\partial F}{\partial y}(\mathbf{x}_0 + t\mathbf{h}_i, y_0 + tk).$$

Re-arranging we can write

$$\frac{f(\mathbf{x}_0 + \mathbf{h}_i) - f(\mathbf{x}_0)}{h} = \frac{k}{h} = -\frac{\frac{\partial F}{\partial x_i}(\mathbf{x}_0 + t\mathbf{h}_i, y + tk)}{\frac{\partial F}{\partial y}(\mathbf{x}_0 + t\mathbf{h}_i, y_0 + tk)}.$$

As the quotient on the right-hand-side consists of continuous functions and $\frac{\partial F}{\partial y} \neq 0$ in $B_{\mathbf{a},b}(r)$, taking the $h \to 0$ limit yields

$$\frac{\partial f}{\partial x_i} = -\frac{\frac{\partial F}{\partial x_i}(\mathbf{x}_0, y_0)}{\frac{\partial F}{\partial y}(\mathbf{x}_0, y_0)}, \tag{3.1}$$

which is a continuous function. □

A useful consequence of the proof of Theorem 3.3 is equation (3.1) which gives a formula for the partial derivatives of $f(\mathbf{x})$. This is not surprising though, since if $y = f(\mathbf{x})$ satisfies $F(\mathbf{x}, f(\mathbf{x})) = 0$ then we may differentiate with respect to x_j to find that

$$0 = \partial_j F + \partial_{n+1} F \partial_j f, \qquad \Rightarrow \qquad \partial_j f = -\frac{\partial_j F}{\partial_{n+1} F}$$

which agrees with what we found in the course of the proof.

Recall that when implicit differentiation is used in first year calculus, we wave our hands and tell the student to assume that what we are doing is kosher. The Implicit Function Theorem is the

theorem which justifies the fact that this can be done in general (though naturally only at the places where the theorem actually applies). Furthermore, while equation (3.1) is useful theoretically, it effectively amounts to implicit differentiation, which is what we will often use to actually compute these derivatives.

Corollary 3.4

If $F \in C^1(\mathbb{R}^{n+1}, \mathbb{R})$ satisfies $\nabla F \neq 0$, then for every $\mathbf{x}_0 \in S = \{\mathbf{x} : F(\mathbf{x}) = 0\}$ there is a neighbourhood N containing $\mathbf{x}_0$ such that $S \cap N$ is the graph of a C^1 function.

This corollary is of course immediate. The fact that $\nabla F \neq 0$ means that at every point, one of the components $\partial_j F \neq 0$. We may then apply Theorem 3.3 to solve for x_j in terms of the remaining variables, and the result follows.

Example 3.5 Recall that the circle defined by the zero locus of $F(x, y) = x^2 + y^2 - 1$ cannot globally be solved for either x or y. However, $\nabla F(x, y) = (2x, 2y)$, which means that whenever $y \neq 0$ we may determine y in terms of x, and vice versa. Indeed, this is what we expect, since any neighbourhood about the points $(0, \pm 1)$ is necessarily two-to-one in terms of y. Furthermore, if $y \neq 0$, let $y = f(x)$ be the local solution for y in terms of x. From equation (3.1) the derivative $\frac{\mathrm{d}f}{\mathrm{d}x}$ is then

$$\frac{\mathrm{d}f}{\mathrm{d}x} = -\frac{\partial_1 F}{\partial_2 F} = -\frac{2x}{2y} = -\frac{x}{y}.$$

This agrees with both implicit differentiation as well as explicit differentiation of $y = \pm\sqrt{1 - x^2}$, and is left as an exercise for the student. ▲

Example 3.6

Consider the function $F(x, y, z) = (2x + y^3 - z^2)^{1/2} - \cos(z)$. If $S = f^{-1}(0)$, determine which variables may be determined by the others in a neighbourhood of $(1, -1, 0)$ and compute the corresponding partial derivatives.

Solution. First notice that $F(1, -1, 0) = 0$ so that this point is in S. We need to determine which partial derivatives are non-zero at $(1, -1, 0)$, so we compute to find

$$\nabla F = \frac{1}{\sqrt{2x + y^3 - z^2}} \left(1, \tfrac{3}{2}y^2, -z + \sqrt{2x + y^3 - z^2}\sin(z)\right).$$

At the point $(1, -1, 0)$ this reduces to $\nabla F(1, -1, 0) = (1, 3/2, 0)$, so we may find C^1 functions f and g such that $x = f(y, z)$ and $y = g(x, z)$, but it is not possible to solve for z in a neighbourhood of $(1, -1, 0)$.

For the partial derivatives, we start with $x = f(y,z)$.

$$\frac{\partial f}{\partial y} = -\frac{\partial_y F}{\partial_x F} = -\frac{3}{2}y^2$$
$$\frac{\partial f}{\partial z} = -\frac{\partial_z F}{\partial_x F} = z - \sin(z)\sqrt{2x+y^3-z^2}.$$

Similarly, for $y = g(x,z)$ we have

$$\frac{\partial g}{\partial x} = -\frac{\partial_x F}{\partial_y F} = -\frac{2}{3y^2}$$
$$\frac{\partial g}{\partial z} = -\frac{\partial_z F}{\partial_y F} = \frac{2z - 2\sin(z)\sqrt{2x+y^3-z^2}}{3y^2}.$$

Again, the student may check that this is consistent with implicit differentiation. ■

3.1.2 The General Case

The general case of a C^1 function $F : \mathbb{R}^{n+k} \to \mathbb{R}^k$ is not much more difficult: The major change will be evaluating what the analogous condition to $\frac{\partial F}{\partial y} \neq 0$ should be. Let $\mathbf{x} = (x_1, \ldots, x_n)$ and $\mathbf{y} = (y_1, \ldots, y_k)$. We once again return to the case where $F(\mathbf{x}, \mathbf{y})$ is a linear function. In this case, let $A \in M_{k\times n}(\mathbb{R})$ and $B \in M_{k\times k}(\mathbb{R})$ be real matrices, and define $F(\mathbf{x},\mathbf{y}) = A\mathbf{x} + B\mathbf{y} + \mathbf{c}$ for some $\mathbf{c} \in \mathbb{R}^k$. If $(\mathbf{x}_0, \mathbf{y}_0)$ is some point where $F(\mathbf{x}_0, \mathbf{y}_0) = 0$, then we can express $\mathbf{y}$ as a function of $\mathbf{x}$ if and only if the matrix B is invertible.

If $F(\mathbf{x},\mathbf{y})$ is now a general function, the set $F(\mathbf{x},\mathbf{y}) = 0$ defines at surface of dimension at most n in $\mathbb{R}^{n+k}$. Let $F(\mathbf{x},\mathbf{y}) = (F_1(\mathbf{x},\mathbf{y}), \ldots, F_k(\mathbf{x},\mathbf{y}))$ for C^1 functions $F_i : \mathbb{R}^{n+k} \to \mathbb{R}$. The Jacobian of $F(\mathbf{x},\mathbf{y})$ is given by

$$dF = \underbrace{\begin{bmatrix} \frac{\partial F_1}{\partial x_1} & \cdots & \frac{\partial F_1}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial F_k}{\partial x_1} & \cdots & \frac{\partial F_k}{\partial x_n} \end{bmatrix}}_{A(\mathbf{x},\mathbf{y}) \in M_{k\times n}(C^1(\mathbb{R}^{n+k}))} \quad \underbrace{\begin{bmatrix} \frac{\partial F_1}{\partial y_1} & \cdots & \frac{\partial F_1}{\partial y_k} \\ \vdots & \ddots & \vdots \\ \frac{\partial F_k}{\partial y_1} & \cdots & \frac{\partial F_k}{\partial y_k} \end{bmatrix}}_{B(\mathbf{x},\mathbf{y}) \in M_{k\times k}(C^1(\mathbb{R}^{n+k}))}$$

so that the tangent plane to $F^{-1}(\mathbf{0})$ at $(\mathbf{x}_0, \mathbf{y}_0)$ is given by $A(\mathbf{x}_0,\mathbf{y}_0)\mathbf{x} + B(\mathbf{x}_0,\mathbf{y}_0)\mathbf{y} + \mathbf{d} = 0$. This tells us that our analogy to $\frac{\partial F}{\partial y} \neq 0$ from the single-variable case should now be changed to $\left(\frac{\partial F_i}{\partial y_j}\right)_{ij}$ should be an invertible function.

Theorem 3.7: General Implicit Function Theorem

Let $F : \mathbb{R}^{n+k} \to \mathbb{R}^n$ be a C^1 function, and write $(x_1, \ldots, x_n, y_1, \ldots, y_k)$ for the coordinates in $\mathbb{R}^{n+k}$. If $(\mathbf{a},\mathbf{b})$ satisfies $F(\mathbf{a},\mathbf{b}) = 0$ and $\left(\frac{\partial F_i}{\partial y_j}\right)_{ij}$ is invertible, there exists $r > 0$ and a unique C^1 function $\mathbf{f} : B_r(\mathbf{a}) \to \mathbb{R}^k$ such that for all $\mathbf{x} \in B_r(a)$, $F(\mathbf{x}, \mathbf{f}(\mathbf{x})) = 0$.

The proof of this theorem is done via induction on k, but it is quite messy and not particularly enlightening so we omit the proof. Of more immediate interest is whether we can determine an

equation for the partial derivatives of the $f_k(\mathbf{x})$. The boring answer to this is that we simply differentiate the equation $F(\mathbf{x}, \mathbf{f}(\mathbf{x})) = 0$ with respect to x_j to determine the result, but this does not clear things up as much as a simple example.

Example 3.8

Consider the function

$$\mathbf{F}(x, y, u, v) = \begin{pmatrix} x^2 - y^2 - u^3 + v^2 + 4 \\ 2xy + y^2 - 2u^2 + 3v^4 + 8 \end{pmatrix}.$$

If $S = \mathbf{F}^{-1}(\mathbf{0})$, show that (u, v) may be expressed as functions of (x, y) in a neighbourhood of $(2, -1, 2, 1)$ and compute the derivatives of those functions.

Solution. The (u, v) derivatives of $\mathbf{F}$ are given by

$$d_{(u,v)}\mathbf{F} = \begin{pmatrix} -3u^2 & 2v \\ -4u & 12v^3 \end{pmatrix}, \qquad \Rightarrow \qquad d_{(u,v)}\mathbf{F}(2, -1, 2, 1) = \begin{pmatrix} -12 & 2 \\ -8 & 12 \end{pmatrix}$$

which has determinant $-128 \neq 0$ and so is invertible. By Theorem 3.7 we know that (u, v) may thus be determined as functions of (x, y) in a neighbourhood of this point; say $u = g_1(x, y)$ and $v = g_2(x, y)$.

Now in order to determine the derivatives, we differentiate the function $\mathbf{F}(x, y, u, v)$ implicitly with respect to x and y, keeping in mind that $u = g_1(x, y)$ and $v = g_2(x, y)$. We thus have

$$\begin{pmatrix} 2x - 3u^2 \frac{\partial u}{\partial x} + 2v \frac{\partial v}{\partial x} \\ 2y - 4u \frac{\partial u}{\partial x} + 12v^3 \frac{\partial v}{\partial x} \end{pmatrix} = 0$$

$$\Leftrightarrow \begin{pmatrix} -3u^2 & 2v \\ -4u & 12v^3 \end{pmatrix} \begin{pmatrix} \frac{\partial u}{\partial x} \\ \frac{\partial v}{\partial x} \end{pmatrix} = \begin{pmatrix} -2x \\ 2y \end{pmatrix}$$

$$\Leftrightarrow \begin{pmatrix} \frac{\partial u}{\partial x} \\ \frac{\partial v}{\partial x} \end{pmatrix} = \begin{pmatrix} -3u^2 & 2v \\ -4u & 12v^3 \end{pmatrix}^{-1} \begin{pmatrix} -2x \\ -2y \end{pmatrix}$$

$$\Leftrightarrow \begin{pmatrix} \frac{\partial u}{\partial x} \\ \frac{\partial v}{\partial x} \end{pmatrix} = \frac{1}{8uv - 36u^2v^3} \begin{pmatrix} -24xv^3 + 4vy \\ -8ux + 6u^2y \end{pmatrix}.$$

Note that this solution makes sense in spite of the fact that the u and v appear in the solution, since $u = g_1(x, y)$ and $v = g_2(x, y)$ implies that these are functions of x, y alone.

■

3.1.3 The Inverse Function Theorem

If we are clever, we can use the Implicit Function Theorem to say something about invertibilty. Consider for example a function $F : \mathbb{R}^2 \to \mathbb{R}$ and its zero locus $S = F^{-1}(0)$. If both $\partial_x F$ and $\partial_y F$ are non-zero at a point (a, b), the Implicit Function Theorem implies that we can write y in

terms of x and vice-versa, in a neighbourhood of (a, b). More precisely, there exists f, g such that $y = f(x)$ and $x = g(y)$ locally.

By taking a sufficiently small neighbourhood around (a, b), we can guarantee that both f and g are injective (convince yourself this is true), and so by single variable results, both f and g have inverses. For example, this means that $f^{-1}(y) = x$. But the Implicit Function Theorem also told us that the function g satisfying $g(y) = x$ was unique, so necessarily $g = f^{-1}$.

We conclude that the Implicit Function Theorem might be able to say something about determining when a function is invertible. This culminates in the following theorem:

Theorem 3.9: The Inverse Function Theorem

Let $U, V \subseteq \mathbb{R}^n$ and fix some point $\mathbf{a} \in U$. If $\mathbf{f} : U \to V$ is of class C^1 and $D\mathbf{f}(\mathbf{a})$ is invertible, then there exists neighbourhoods $\tilde{U} \subseteq U$ of $\mathbf{a}$ and $\tilde{V} \subseteq V$ of $f(\mathbf{a})$ such that $\mathbf{f}|_{\tilde{U}} : \tilde{U} \to \tilde{V}$ is bijective with C^1 inverse $(\mathbf{f}|_U)^{-1} : \tilde{V} \to \tilde{U}$. Moreover, if $\mathbf{b} = \mathbf{f}(\mathbf{a})$ then the derivative of the inverse map is given by

$$[D\mathbf{f}^{-1}](\mathbf{b}) = [D\mathbf{f}(\mathbf{a})]^{-1}. \tag{3.2}$$

It turns out that the Inverse Function Theorem and the Implicit Function Theorem are actually equivalent; that is, the Implicit Function Theorem can be proven from "scratch" then used to prove the Inverse Function Theorem, or vice versa. We already have the Implicit Function Theorem, so we might as well use it, not to mention that the "scratch" proof of the Inverse Function Theorem is rather lengthy and uses the Contraction Mapping Theorem.

Proof. Define the function $\mathbf{F} : U \times V \subseteq \mathbb{R}^{2n} \to \mathbb{R}^n$ by $\mathbf{F}(\mathbf{x}, \mathbf{y}) = \mathbf{y} - \mathbf{f}(\mathbf{x})$ so that $\mathbf{F}(\mathbf{x}, \mathbf{y}) = 0$ is equivalent to $\mathbf{y} = \mathbf{f}(\mathbf{x})$. We want to determine if we can solve for $\mathbf{x}$ locally in terms of $\mathbf{y}$, so naturally we will use the Implicit Function Theorem. But this immediately follows, since the invertibility condition on $D\mathbf{f}(\mathbf{x})$ is precisely the requirement for the Implicit Function Theorem.

To derive Equation (3.2) we note that $\mathbf{f}^{-1}(\mathbf{f}(\mathbf{x})) = \mathbf{x}$, so differentiating and applying the chain rule yields

$$[D\mathbf{f}^{-1}](\mathbf{f}(\mathbf{x})) \cdot D\mathbf{f}(\mathbf{x}) = I,$$

and the result then follows. □

Example 3.10

Determine whether the function

$$\mathbf{f}(x, y) = (e^x \sin(y), e^x \cos(y))$$

is invertible in a neighbourhood of $(0, 0)$. More generally, show that $\mathbf{f}$ is invertible in a neighbourhood of any point.

Solution. Computing the derivative of $\mathbf{f}$, we get

$$D\mathbf{f}_{(x,y)} = \begin{pmatrix} e^x \sin(y) & e^x \cos(y) \\ e^x \cos(y) & -e^x \sin(y) \end{pmatrix}.$$

Evaluating at $(0,0)$ we get

$$D\mathbf{f}_{(0,0)} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

which is certainly invertible (in fact, it is its own inverse). More generally, we want to determine whether $D\mathbf{f}_{(x,y)}$ is invertible, so we compute the determinant to be

$$\det D\mathbf{f}_{(x,y)} = -e^{2x}\sin^2(y) - e^{2x}\cos^2(y) = -e^{2x}.$$

Since e^{2x} is never zero, $\mathbf{D}f_{(x,y)}$ will be invertible for any choice of (x,y), so the Inverse Function Theorem can be applied everywhere. ■

3.2 Curves, Surfaces, and Manifolds

The Implicit Function Theorem is the key to determining the appropriate definition of what it means for something to be smooth. Intuitively, an object is smooth if it contains no corners, such as a sphere. On the other hand, something like a cube will not be smooth, as each vertex and edge of the cube are sharp. It turns out that this is not the best of way of thinking about smooth: For example, what happens when we are given a surface, such as the *lemniscate* (see Figure 26)? When we draw the lemniscate, we can do it is a smooth fashion so that no sharp edges ever appear; nonetheless there does seem to be something odd about what happens at the overlap point.

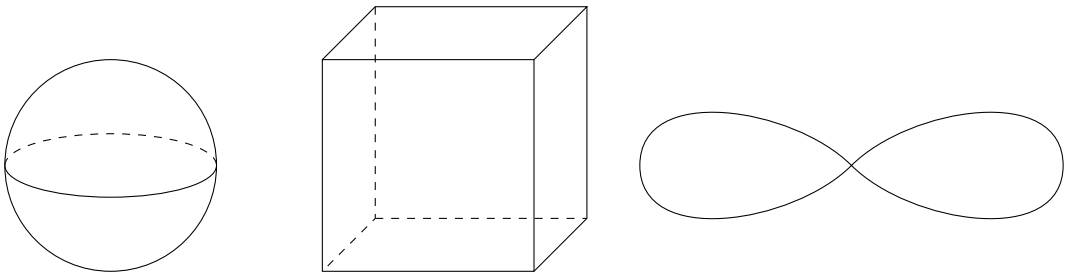


Figure 26: The sphere should be smooth, the cube should not be, but who knows about the lemniscate?

The criteria that we will see will look mysterious in the absence of geometric intuition, but the key to testing smoothness is to look at the tangent space. At each point on a curve, surface, or higher dimensional object, there is the notion of a line, plane, or hyperplane which is tangent to that point. Each of these tangent spaces just looks like a vector space, and thus has an associated dimension. A space is smooth if its tangent space does not change dimension.

For example, we will see that every point on the sphere has a two dimensional tangent space. For the cube, the interior of the faces will have two dimensional tangent spaces, while the edges will have 1-dimensional tangent spaces, and the vertices will have a 0-dimensional tangent space.

There are several ways of actually defining these spaces, such as via the graph of a function, through a parameterization, or as the zero locus of a function. In each of these cases, the dimensions of the domain and codomain will play an important role. In this section, we take an introductory look at the relationship between 1-dimensional curves, 2-dimensional surfaces, and n-dimensional manifolds.

3.2.1 Curves in $\mathbb{R}^2$

We have thus far seen three different ways of defining one-dimensional objects. Here we pay particular attention to the case of one-dimensional objects in $\mathbb{R}^2$. A curve can be written as

1. The graph of a function: Let $f : \mathbb{R} \to \mathbb{R}$, and define the graph of f to be
$$\Gamma(f) = \{(x, f(x)) : x \in \mathbb{R}\}.$$
2. The zero locus of a function: Let $F : \mathbb{R}^2 \to \mathbb{R}$, and let $C = F^{-1}(0)$. This object will generally be one-dimensional, as $F(x, y) = 0$ yields one equation with two-variables, meaning we can (locally) solve one in terms of the other.
3. As the image of a parametric function: Let $s : (a, b) \to \mathbb{R}^2$ by given by $t \mapsto (s_1(t), s_2(t))$, and define the curve to be $p((a, b))$.

For example, the curve which is the graph of $f(x) = \sqrt[3]{x^2}$, is the same as the zero locus of $F(x, y) = y^3 - x^2$ and the parametric curve $p(t) = (t^3, t^2)$.

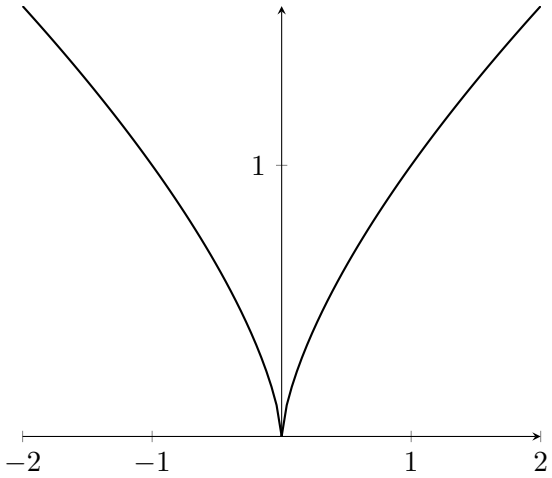


Figure 27: The curve defined by the graph of $f(x) = \sqrt[3]{x^2}$, the zero locus of $F(x, y) = y^3 - x^2$ and the parametric function $p(t) = (t^3, t^2)$.

Unfortunately, the set of all curves defined by one method need not be the same as those defined by another.

Proposition 3.11

Every curve that can be expressed as the graph of a function $f : \mathbb{R} \to \mathbb{R}$ may also be written as the zero locus of a function $F : \mathbb{R}^2 \to \mathbb{R}$ and parametrically as the image of $p : \mathbb{R} \to \mathbb{R}^2$.

Proof. Fix some $f : \mathbb{R} \to \mathbb{R}$ with graph $\Gamma(f) = \{(x, f(x)) : x \in \mathbb{R}\}$. Define the function $F(x, y) = y - f(x)$ and notice that

$$F^{-1}(0) = \{(x, y) : F(x, y) = 0\} = \{(x, y) : y = f(x)\} = \{(x, f(x)) : x \in \mathbb{R}\} = \Gamma(f).$$

In the parametric case, define the parametric function $p : \mathbb{R} \to \mathbb{R}^2$ by $t \mapsto (t, f(t))$. Once again we have

$$p(\mathbb{R}) = \{(t, f(t)) : t \in \mathbb{R}\} = \Gamma(f),$$

as required. □

The converse of this proposition is not true. For example, the circle is impossible to write as the graph of a function, since it fails the vertical line test. However, the circle is the zero locus of $F(x, y) = x^2 + y^2 - 1$, or the image of the function $p : [0, 2\pi) \to \mathbb{R}^2$ given by $p(t) = (\cos(t), \sin(t))$. Since graphs cannot describe even simple shapes like a circle, they are rarely used to define manifolds.

Now we are more interested in assessing the smoothness properties of a curve C. Our time studying calculus has told us that if the function $f : \mathbb{R} \to \mathbb{R}$ is C^1 then its graph defines a curve which is smooth. However, since not all curves of interest can be written as the graphs of functions, this will fail to be a good definition. Instead, we will just require that the curve *locally* look like the graph of a smooth function:

Definition 3.12

A *connected* set $C \subseteq \mathbb{R}^2$ is said to be a *smooth curve* if every point $\mathbf{a} \in C$ has a neighbourhood N on which $C \cap N$ is the graph of a C^1 function. If C is not connected, then we shall say that C is smooth if each of its connected components is a smooth curve.

Unfortunately, if the curve is defined as a zero locus or parametrically, we will not be able to "read off" the smoothness of the curve just by looking at the defining function. To see what kind of things can go wrong, consider the function $F(x, y) = y^3 - x^2$ (Figure 27). This is certainly a C^1 function (and is in fact infinitely differentiable), but the zero locus it defines is the curve $y^3 = x^2$. The student can easily check that this curve is not differentiable when $x = 0$, and so cannot be written as the graph of a C^1 function.

Thus the best we can hope to do is locally. Our work with the Implicit Function Theorem tells us that $\nabla F \neq 0$ on $F^{-1}(0)$ will guarantee that our curve looks locally like the graph of a C^1 function, but what is the condition we should impose on parametric definition? The following is a first step in the right direction:

Theorem 3.13

1. Let $F : \mathbb{R}^2 \to \mathbb{R}$ be a C^1 function and $S = F^{-1}(0)$. If $\mathbf{p} \in S$ and $\nabla F(p) \neq 0$ then there exists an open neighbourhood N of $\mathbf{p}$ such that $N \cap S$ is the graph of a C^1 function.
2. Let $p : (a, b) \to \mathbb{R}^2$ be a C^1 function and let $S = p(a, b)$. If $t_0 \in (a, b)$ satisfies $p'(t) \neq 0$ then there is an open subinterval $I \subseteq (a, b)$ such that $p(I)$ is the graph of a C^1-curve.

Proof. 1. This is immediate by the Implicit Function Theorem.

2. The image of p is $p(a, b) = \{(p_1(t), p_2(t)) : t \in (a, b)\}$. Morally, we would like to "change variables" by setting $x = p_1(t)$, inverting to write $t = p_1^{-1}(x)$ and substitute to write C as $C = \left\{(x, p_2(p_1^{-1}(x)))\right\}$. However, there is no reason to suspect that p_1 should be invertible.

The solution to this is effectively given by the Implicit Function Theorem, which in a sense says that we can locally invert.

More rigorously, since $p'(t_0) = (p_1'(t_0), p_2'(t_0)) \neq 0$ then one of the $p_i'(t_0) \neq 0$. Without loss of generality, assume that $p_1'(t_0) \neq 0$. Define the C^1 function $F(x,t) = x - p_1(t)$ with $x_0 = p_1(t_0)$ so that $F(x_0,t_0) = 0$ and $\partial_t F(x_0,t_0) = -p_1'(t_0) \neq 0$. By the Implicit Function Theorem, we may solve for t in terms of x; that is, there exists a C^1-function g such that $t = g(x)$ in a neighbourhood of (x_0,t_0). Thus

$$p(t) = (p_1(t), p_2(t)) = (x, p_2(g(x))).$$

Since p_2 and g are both C^1, so too is their composition $p_2 \circ g$, and this shows that the image of $p(t)$ is the graph of the C^1-curve $p_2 \circ g$ as required.

□

Example 3.14

Determine whether the curve defined by the image of $p : (0, 2\pi) \to \mathbb{R}^2$ given by $t \mapsto (t\cos(t), t\sin(t))$ is smooth.

Solution. If we differentiate $p'(t) = (\cos(t) - t\sin(t), \sin(t) + t\cos(t))$ and it is not clear whether or not this function is ever zero. Instead, let's try to rewrite this curve as the zero locus of a function $F : \mathbb{R}^2 \to \mathbb{R}$. Set $x = t\cos(t)$ and $y = t\sin(t)$ and notice that $x^2 + y^2 = t^2$. We can isolate t by recognizing that $t = \arctan(y/x)$, and so the curve as defined in the same thing as the zero locus of the function $F(x,y) = x^2 + y^2 - \arctan^2(y/x)$. It's gradient is given by

$$\nabla F(x,y) = \left(2x + \frac{2\arctan(y/x)}{x^2+y^2}, 2y - \frac{2x\arctan(y/x)}{x^2+y^2}\right).$$

The only time that this can be zero is when $(x,y) = (0,0)$; however, the restriction of $t \in (0, 2\pi)$ makes this impossible. ■

An important point to note is that part (2) of the theorem indicates there is an open *interval* $I \subseteq (a,b)$ on which the function is the graph of a C^1-curve, but this does not mean that there is an open neighbourhood of $\mathbb{R}^2$ on which this is a smooth curve. What could possibly go wrong? Well, the map could fail to be injective:

Example 3.15 Under Definition 3.12 we know that the lemniscate cannot be a smooth curve, since at the point of overlap there is no neighbourhood on which the function looks like the graph of a smooth curve. In parametric equations, one has (Figure 28)

$$p : \mathbb{R} \to \mathbb{R}^2, \qquad t \mapsto \frac{1}{1+\sin^2 t}\left(\cos(t), \sin(t)\cos(t)\right). \tag{3.3}$$

Notice that

$$p'(t) = \frac{1}{(1+\sin^2(t))^2}\left(-\sin(t)[2+\cos^2(t)], \cos(2t)[1+\sin^2(t)] - \sin(t)\cos(t)\sin(2t)\right).$$

This is never zero, since the first argument is only zero at $t = n\pi$ whereas the second argument is identically 1 at $n\pi$. But the function is certainly not injective since it is periodic. In fact, even restricting our attention to $(0, 2\pi)$ we see that $p(\pi/2) = p(3\pi/2) = (0, 0)$. Thus there is no neighbourhood of $(0, 0)$ where $p(0, 2\pi)$ looks like the graph of a C^1 function. Even worse, there are two different values of the derivative at $(0, 0)$ depending on whether we take $t = \pi/2$ or $3\pi/2$.

$$p'(\pi/2) = \tfrac{1}{2}(-1, 1), \qquad p'(3\pi/2) = \tfrac{1}{2}(1, -1).$$

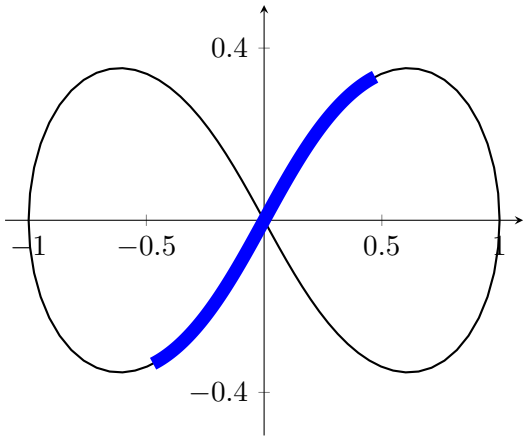


Figure 28: The lemniscate drawn with the parametric equation given by (3.3). The thick blue line is the set $p((\pi/4, 3\pi/4))$, which is the graph of a C^1-function. However, the whole curve fails to be smooth.

So what exactly did Theorem 3.13 tell us? It told us that since $p'(\pi/2) \neq 0$ there is a neighbourhood around $\pi/2$ whose image looked like the graph of a C^1-function. However, this did not take into consideration the more global nature of the curve. The way to remedy this situation is as follows:

Definition 3.16

If $I \subseteq \mathbb{R}$ is an interval, a C^1 map $\gamma : I \to \mathbb{R}^2$ is said to be

1. A *regular curve* if $\gamma'(t) \neq 0$ for all $t \in I$,
2. A *simple curve* if γ is injective on the interior of I.

Hence if γ is a regular, Theorem 3.13 implies that there is a neighbourhood of each point whose image looks like the graph of a C^1-function. Simplicity guarantees that no funny overlaps can happen, and this is what is needed for the curve to be smooth.

Example 3.17

Determine whether the curve defined by the parametric equation $p : \mathbb{R} \to \mathbb{R}^2$ given by $t \mapsto (t^3, e^{2t})$ is a smooth.

Solution. Differentiating, we get $p'(t) = (3t^2, 2e^{2t})$ and since the second component is never zero, p is certainly regular. On the other hand, both component functions t^3 and e^{2t} are injective, implying that p is also injective, so p is simple. We conclude that the image of p is smooth. ■

Summary: There are three ways to define curves: as the graph of a function $\mathbb{R} \to \mathbb{R}$, as the zero locus of a function $\mathbb{R}^2 \to \mathbb{R}$, or as the image of a parameterization $\mathbb{R} \to \mathbb{R}^2$. Smoothness is determined as follows:

1. If $C = \Gamma(f) = \{(x, f(x)) : x \in [a, b]\}$ is the graph of a C^1 function, then C is smooth.
2. If $C = F^{-1}(0)$ is the zero locus of $F : \mathbb{R}^2 \to \mathbb{R}$, then C is smooth if $\nabla F \neq 0$ for every point in C.
3. If $C = p(a, b)$ where $p : \mathbb{R} \to \mathbb{R}^2$ is the image of a parameterization, then C is smooth only if p is regular ($p'(t) \neq 0$) and simple (p is injective).

Curves in $\mathbb{R}^n$: Generalizing our discussion so far, a curve in $\mathbb{R}^n$ may be described in one of three ways:

1. The graph of a function $\mathbf{f} : \mathbb{R} \to \mathbb{R}^{n-1}$,
2. The zero locus of a function $\mathbf{F} : \mathbb{R}^n \to \mathbb{R}^{n-1}$,
3. The image of a function $\mathbf{p} : \mathbb{R} \to \mathbb{R}^n$.

Again, we define $C \subseteq \mathbb{R}^n$ to be a smooth curve if it is connected and locally the graph of a C^1 function. If $C = \mathbf{F}^{-1}(0)$ then the linear independence of $\{\nabla F_i\}$ guarantees that C is a smooth curve. The notions of regular and simple curves generalize in an obvious way to $\mathbb{R}^n$ and so the image of a regular, simple, C^1-map $\mathbb{R} \to \mathbb{R}^n$ is also a smooth curve.

Dimension of the Tangent Space: It was mentioned previously that the idea is to examine the dimension of the tangent space, and see whether or not it changes as we move along our curve. Notice that when $p : \mathbb{R} \to \mathbb{R}^n$ that $p'(t_0)$ is a vector which is tangent to the curve. Hence so long as $p'(t_0) \neq 0$, the tangent space is always one dimensional. Conversely, if the curve is given by the zero locus of $F : \mathbb{R}^n \to \mathbb{R}^{n-1}$ then there are $(n-1)$-vectors $\{\nabla F_i\}$. If these are linearly dependent, they define an $(n-1)$-hyper plane in $\mathbb{R}^n$. The perpendicular to this hyperplane is the tangent space, which will again be one dimensional.

3.2.2 Surfaces in $\mathbb{R}^3$

We have looked at 1-dimensional spaces of $\mathbb{R}^2$ and how to generalize them to $\mathbb{R}^n$. Now we increase the dimension of the space itself. The simplest case will be to look at 2-dimensional spaces (surfaces) or $\mathbb{R}^3$, afterwhich we will have a sufficient idea of the general procedure to be able to discuss k-dimensional spaces of $\mathbb{R}^n$.

Much akin to our treatment of curves, there are three ways to discuss surfaces:

1. The graph of a function $f : \mathbb{R}^2 \to \mathbb{R}$,
2. The zero-locus of a function $F : \mathbb{R}^3 \to \mathbb{R}$,
3. The image of a function $p : \mathbb{R}^2 \to \mathbb{R}^3$.

Naturally, given the graph of a function $f : \mathbb{R}^2 \to \mathbb{R}$ we may write it as the zero-locus of $F(x, y, z) = z - f(x, y)$, or parametrically as $(s, t) \mapsto (s, t, f(s, t))$.

Example 3.18 Fix $r, R > 0$ and consider the parameterization $g : [0, 2\pi) \times [0, 2\pi) \to \mathbb{R}^3$ given by

$$g(\phi, \theta) = [(R + r\cos(\phi))\cos(\theta), (R + r\cos(\phi))\sin(\theta), r\sin(\phi)]$$

or equivalently, the zero locus of the function

$$F(x, y, z) = (x^2 + y^2 + z^2 + R^2 - r^2)^2 - 4R^2(x^2 + y^2).$$

These define a torus, where r is the radius of the sweeping circle, and R is the radius to the center of that circle

Definition 3.19

A smooth surface of $\mathbb{R}^3$ is a connected subset $S \subseteq \mathbb{R}^3$ such that, for every $\mathbf{p} \in S$ there exists a neighbourhood N of $\mathbf{p}$ such that $S \cap N$ is the graph of a C^1 function $\mathbf{f} : \mathbb{R}^2 \to \mathbb{R}$.

The Implicit Function Theorem again tells us that $F^{-1}(0)$ will be a smooth surface so long as $\nabla F(\mathbf{x}) \neq 0$ for all $\mathbf{x} \in F^{-1}(0)$, but the case of the parametric definition is slightly more subtle. Consider a linear parametric function

$$\mathbf{f}(s, t) = \mathbf{u}s + \mathbf{v}t + \mathbf{w}.$$

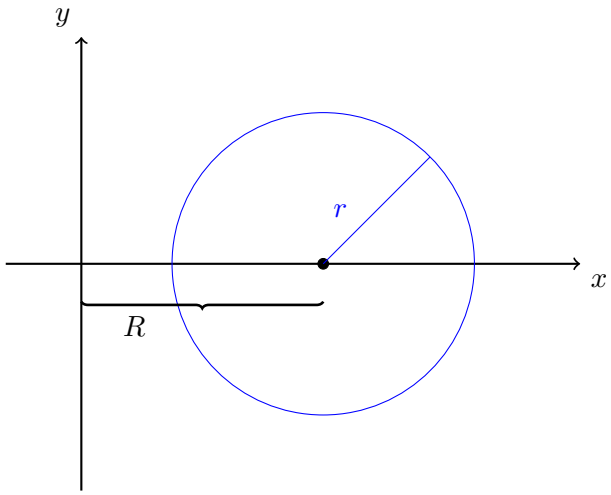


Figure 29: A slice of the torus.

If $\mathbf{u}$ and $\mathbf{v}$ are linearly dependent, the image of $\mathbf{p}$ will be a line, while if they are linearly independent the image will define a plane. Since we are only interested in bona-fide surfaces, we need to preclude the possibility of a line. When $\mathbf{p}(s,t)$ is not just a linear function, the usual generalization tells us that we need to tangent-vectors to be linearly independent; that is, $\frac{\partial \mathbf{p}}{\partial s}$ and $\frac{\partial \mathbf{p}}{\partial t}$ must be linearly independent. This can equivalently be phrased as saying that the matrix

$$\left[\begin{array}{c|c} \dfrac{\partial \mathbf{p}}{\partial s} & \dfrac{\partial \mathbf{p}}{\partial t} \end{array}\right]$$

must be full-rank.

Theorem 3.20

1. Let $F : \mathbb{R}^3 \to \mathbb{R}$ be a C^1 function and $S = F^{-1}(0)$. If $\mathbf{p} \in S$ and $\nabla F(p) \neq 0$ then there exists an open neighbourhood N of $\mathbf{p}$ such that $N \cap S$ is the graph of a C^1-curve.
2. Let $p : U \subseteq \mathbb{R}^2 \to \mathbb{R}^3$ be a C^1 function and let $S = p(U)$. If the point $(s_0, t_0) \in U$ causes the matrix
$$\left[\begin{array}{c|c} \dfrac{\partial \mathbf{p}}{\partial s}(s_0,t_0) & \dfrac{\partial \mathbf{p}}{\partial t}(s_0,t_0) \end{array}\right]$$
to have full rank, then there is an open subset $V \subseteq U$ of (s_0, t_0) such that $p(V) \cap S$ is the graph of a C^1-curve.

The proof of this theorem is almost identical to that of Theorem 3.13 and is left as an exercise for the student. It should be evident that a connected level set of a C^1 map $F : \mathbb{R}^3 \to \mathbb{R}$ with nowhere vanishing gradient defines a smooth surface. For the parametric definition, we once again require both regularity of the surface (linearly independent tangents) and simplicity (global injectivity).

Example 3.21

Consider the surface defined by the function

$$\mathbf{p}(s,t) = (s\cos(t), s\sin(t), s^2).$$

Find a zero-locus description of the surface, and find (using both the parametric and zero locus pictures) the points where the surface is singular.

Solution. Setting $x = s\cos(t), y = s\sin(t)$, and $z = s^2$, notice that

$$x^2 + y^2 = s^2(\cos^2(t) + \sin^2(t)) = s^2 = z,$$

so that the corresponding zero-locus is given by $F(x,y,z) = z - x^2 - y^2$. Our intuition tells us that this is a paraboloid and so should not admit any singularities. Differentiating the parametric definition, we have

$$\frac{\partial \mathbf{p}}{\partial s} = (\cos(t), \sin(t), 2s), \qquad \frac{\partial \mathbf{p}}{\partial t} = (-s\sin(t), s\cos(t), 0).$$

Now we could examine the rank of the matrix whose columns are made from the above matrices, but in the case of surfaces in $\mathbb{R}^3$ it is easier to check linear-independence by computing the cross product. We find that

$$\frac{\partial \mathbf{p}}{\partial s} \times \frac{\partial \mathbf{p}}{\partial t} = (-2s^2\cos(t), -2s^2\sin(t), s).$$

The only place where this could be zero is at $s = 0$ which corresponds to the origin $(0,0)$, but remember that this is not necessary for a singularity. Let's take a look at the zero-locus definition. The gradient of $F(x,y,z)$ is given by

$$\nabla F(x,y,z) = (-2x, -2y, 1)$$

and this is certainly never zero. This implies that the surface does not have any singularities. ■

Surfaces in General: More generally, a surface in $\mathbb{R}^n$ may be defined by the zero locus of a function $F : \mathbb{R}^n \to \mathbb{R}^{n-2}$, or the image of a parametric function $p : \mathbb{R}^2 \to \mathbb{R}^n$. To see what the general conditions for smoothness should be, we again think about the tangent space. For the zero locus picture, there are $(n-2)$ elements in the set $\{\nabla F_i\}$. If they are linearly independent, then they span an $(n-2)$-dimensional hyperplane of $\mathbb{R}^n$, whose orthogonal complement is the tangent space of the surface. In the parametric picture, $\partial_1 p$ and $\partial_2 p$ form a basis for the tangent space, so for this to be two dimensional, we require that they are everywhere linearly independent.

3.2.3 Dimension k-manifolds in $\mathbb{R}^n$

We now discuss (one last time!) how to form k-dimensional subspaces of $\mathbb{R}^n$. There are two methods we will consider:

1. The zero-locus of a function $\mathbf{F} : \mathbb{R}^n \to \mathbb{R}^{n-k}$,

2. The image of a function $\mathbf{p} : \mathbb{R}^k \to \mathbb{R}^n$.

If our space is $M = \mathbf{F}^{-1}(\mathbf{0})$, the conditions which guarantee that the defined object is a *smooth k-manifold* is that

$$\operatorname{rank} D\mathbf{F}(\mathbf{x}) = n - k, \qquad \forall x \in \mathbf{F}^{-1}(\mathbf{0}).$$

While if $M = \mathbf{p}(U \subseteq \mathbb{R}^k)$ then we must have $\mathbf{p}(\mathbf{t})$ is injective on U and

$$\operatorname{rank} \left[\begin{array}{c|c|c} \partial_1 \mathbf{p}(\mathbf{t}) & \cdots & \partial_k \mathbf{p}(\mathbf{t}) \end{array} \right] = k, \qquad \forall \mathbf{t} \in U.$$

In fact, rather than remembering which dimension corresponds to which, it is sufficient to state that either $D\mathbf{F}(\mathbf{x})$ or $D\mathbf{p}(\mathbf{t})$ must have *maximal rank* at every point on the surface. Rather than rehash our tangent space argument in this case, the student should try to convince his/herself that $\operatorname{rank} D\mathbf{F}(\mathbf{x}_0)$ is the rank of the normal plane at $\mathbf{x}_0$, while $\operatorname{rank} [\partial_i \mathbf{p}(\mathbf{t}_0)]$ is the rank of the tangent plane at $\mathbf{t}_0$.

4 Integration

Having effectively completed our study of differential calculus, we now move on to integral calculus. Students often find integral calculus more difficult than differential calculus, typically because computations are not nearly as straightforward as the "recipe book" style offered by differentiation. Nonetheless, it turns out that integration is actually a far more sound theory: it is easier to make rigorous in general contexts.

We will begin with a "review" of integration on the line (I say "review" because it will almost certainly be new material), before moving onto the general theory for integrating variables in several dimensions.

4.1 Integration on $\mathbb{R}$

Given a sufficiently nice function $f : \mathbb{R} \to \mathbb{R}$, the idea of integration on the interval $[a, b]$ is to estimate the signed[5] area between the graph of the function and the x-axis. The heuristic idea of how to proceed is to divide $[a, b]$ into subintervals and approximate the height of the function by rectangles. We then take a limit as the length of the subintervals goes to zero, and if we get a well-defined number, we call that the integral.

Unfortunately, there is no canonical choice for either how to divide $[a, b]$, nor for how high to make the rectangles. Typical choices for height often include left/right endpoints, or inf/sup values of the function on each subinterval, but of course these are not the only choices.

Aside: It turns out that Riemann integration, or integrating by partitions of the domain, is an inferior choice as there are many functions which are not integrable. A much more prudent choice is to actually break up the range of the function and integrate that way, in a manner known as *Lebesgue integration*. Unfortunately, Lebesgue integration is beyond the scope of the course.

4.1.1 Riemann Sums

For the remainder of this section, we fix an interval $[a, b] \subseteq \mathbb{R}$.

Definition 4.1

A *finite partition* P of $[a, b]$ is an ordered collection of points $P = \{a = x_0 < x_1 < x_2 < \cdots < x_n = b\}$. Define the *order* of P to be $|P| = n$ and the *length* of P to be

$$\ell(P) = \max_{i=1,\ldots,|P|} [x_i - x_{i-1}];$$

that is, the length of P is the length of the longest interval whose endpoints are in P.

One should think of partitions as a way of dividing the interval $[a, b]$ into subintervals. For example, on $[0, 1]$ we think of the partition $P = \left\{0 < \frac{1}{3} < \frac{2}{3} < 1\right\}$ as breaking $[0, 1]$ into $[0, 1/3] \cup [1/3, 2/3] \cup [2/3, 1]$. If $\mathcal{P}_{[a,b]}$ is the set of all finite partitions of $[a, b]$ then $\ell : \mathcal{P}_{[a,b]} \to \mathbb{R}_+$ gives us a "worst-case scenario" for the length of the subintervals, in much the same way as the sup-norm.

[5] Signed area simply means that area above the x-axis will be positive, while area below the x-axis will be negative

The idea is that when we do integration, we are going to want to take partitions whose length between endpoints gets smaller, corresponding to letting the width of our approximating rectangles get smaller. The number $\ell(P)$ then describes the widest width, which in a sense is our "worst" rectangle.

Definition 4.2

If P and Q are two partitions of $[a,b]$, then Q is a *refinement* of P if $P \subseteq Q$.

Example 4.3 Consider the interval $[0,1]$ and the partitions

$$P = \left\{0 < \tfrac{1}{2} < 1\right\}, \quad Q = \left\{0 < \tfrac{1}{3} < \tfrac{2}{3} < 1\right\}, \qquad R = \left\{0 < \tfrac{1}{4} < \tfrac{1}{3} < \tfrac{1}{2} < \tfrac{2}{3} < \tfrac{3}{4} < 1\right\}.$$

Note that P and Q cannot be compared, since one is not a subset of the other. However, $P \leq R$ and $Q \leq R$, so R is a common refinement of both P and Q.

It is not too hard to see that any two sets in $\mathcal{P}_{[a,b]}$ admit a common refinement: Given two partitions $P, Q \in \mathcal{P}_{[a,b]}$, define $R = P \cup Q$ so that $P \subseteq R$ and $Q \subseteq R$.

Definition 4.4

Given a function $f : [a,b] \to \mathbb{R}$, a *Riemann sum* of f with respect to the partition $P = \{x_0 < x_1 < \cdots < x_{n-1} < x_n\}$ is any sum of the form

$$S(f,P) = \sum_{i=1}^{n} f(t_i)(x_i - x_{i-1}), \qquad t_i \in [x_{i-1}, x_i].$$

Note that while the Riemann sum $S(f,P)$ certainly depends on how we choose the sampling t_i, we will often choose to ignore this fact. Some typical choices of Riemann sum that the student has likely seen amount to particular choices of the t_i. In the first case, we have the left- and right-endpoint Riemann sums

$$L(f,P) = \sum_{i=1}^{n} f(x_{i-1})(x_i - x_{i-1}), \qquad R(f,P) = \sum_{i=1}^{n} f(x_i)(x_i - x_{i-1}).$$

Of far greater use are the *lower* and *upper* Riemann sums, defined as follows. Fix a partition $P \in \mathcal{P}_{[a,b]}$ and $f : [a,b] \to \mathbb{R}$. Define

$$m_i = \inf_{x \in [x_{i-1}, x_i]} f(x), \qquad M_i = \sup_{x \in [x_{i-1}, x_i]} f(x),$$

so that m_i is the "smallest" value that f takes on $[x_{i-1}, x_i]$ while M_i is the largest. Now set

$$U(f,P) = \sum_{i=1}^{n} M_i(x_i - x_{i-1}), \qquad u(f,P) = \sum_{i=1}^{N} m_i(x_i - x_{i-1}).$$

The idea of the integral is that regardless of what partition we choose or how we choose to sample the partition, we should always arrive at the same answer. This leads us to the formal definition of Riemann integrability.

Definition 4.5

We say that a function $f : [a, b] \to \mathbb{R}$ is *Riemann integrable* on $[a, b]$ with integral I if for every $\epsilon > 0$ there exists a $\delta > 0$ such that whenever $P \in \mathcal{P}_{[a,b]}$ satisfies $\ell(P) < \delta$ then

$$|S(f, P) - I| < \epsilon.$$

The element I is often denoted $I = \int_a^b f(x)\,\mathrm{d}x$.

We know that the student abhors the idea of ϵ-δ proofs, so it would not surprise us if headaches are currently abound. Let's take a moment and read into what the definition of integrability really means: Roughly speaking, a function is Riemann integrable with integral I if we can approximate I arbitrarily well by taking a sufficiently fine partition P.

There are many different ways of defining Riemann integrability depending on how one chooses to set up the problem. We give here a statement of some equivalent definitions:

Theorem 4.6

If $f : [a, b] \to \mathbb{R}$ is a function, then the following are equivalent:

1. f is Riemann integrable,
2. $\sup_{P \in \mathcal{P}_{[a,b]}} u(f, P) = \inf_{P \in \mathcal{P}_{[a,b]}} U(f, P)$,
3. For every $\epsilon > 0$ there exists a partition $P \in \mathcal{P}_{[a,b]}$ such that $U(f, P) - u(f, P) < \epsilon$,
4. For every $\epsilon > 0$ there exists a $\delta > 0$ such whenever $P, Q \in \mathcal{P}_{[a,b]}$ satisfy $\ell(P) < \delta$ and $\ell(Q) < \delta$ then $|S(f, P) - S(f, Q)| < \epsilon$.

Students coming from Math 137 will recognize definition (2) as the statement that the lower and upper integrals are equal. Indeed, the supremum over the lower Riemann sums is the lower integral, and vice-versa for the upper integrals.

Each of these definitions offers its own advantage. For example, (1) and (2) are useful for theoretical reasons but are highly intractable for determining which functions are actually integrable. On the other hand, (3) and (4) are exceptionally useful as they do not require one to actually know the integral. In particular, (3) is useful because the upper and lower Riemann sums are nicely behaved, while (4) is useful because it offers the flexibility to choose samplings.

Example 4.7

Show that the function $f(x) = cx$ is integrable on $[a, b]$.

Solution. If $c = 0$ then there is nothing to do. Let us use definition (3) to proceed, and assume without loss of generality that $c > 0$. The advantage of using definition (3) is that we get to choose the partition, which gives us a great deal of power. Let n be any positive integer such that $\frac{c(b-a)^2}{n} < \epsilon$ (more on how to choose this later). Since our function is increasing, minima will occur at left endpoints, and maxima will occur at right endpoints. Choose a uniform partition of $[a, b]$

into $n+1$-subintervals $P = \{a = x_0, x_1, \ldots, x_n = b\}$, where $x_i = a + \frac{b-a}{n} i$, so that

$$u(f,P) = \sum_{k=0}^{n-1} f(x_k)(x_{k+1} - x_k) = \frac{c(b-a)}{n} \sum_{k=0}^{n-1} x_k$$
$$U(f,P) = \sum_{k=0}^{n-1} f(x_{k+1})(x_{k+1} - x_k) = \frac{c(b-a)}{n} \sum_{k=0}^{n-1} x_{k+1}.$$

Hence their difference yields

$$\begin{aligned} U(f,P) - u(f,P) &= \frac{c(b-a)}{n} \sum_{k=0}^{n-1} (x_{k+1} - x_k) \\ &= \frac{c(b-a)}{n}(b-a) \\ &< \epsilon. \end{aligned}$$

which is what we wanted to show. ■

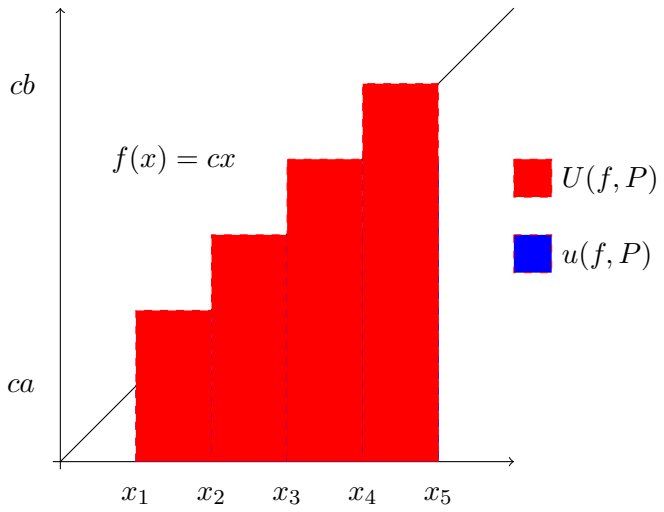


Figure 30: One can visually see why the difference between $U(f,P)$ and $u(f,P)$ results in a telescoping sum. For example, the red rectangle on $[x_1, x_2]$ is the same area as the blue rectangle on $[x_2, x_3]$, so they cancel in the difference.

Example 4.8

Show that the characteristic function of the rationals on $[0,1]$:

$$\chi_{\mathbb{Q}}(x) = \begin{cases} 1 & x \in \mathbb{Q} \cap [0,1] \\ 0 & \text{otherwise} \end{cases}$$

is not Riemann integrable.

Solution. Let $P = \{0 = x_0 < x_1 < \cdots < x_n = 1\}$ be an arbitrary partition of $\mathbb{Q} \cap [0,1]$, and recall that $\mathbb{Q}$ is dense in $[0,1]$ while the irrationals $\mathbb{R} \setminus \mathbb{Q}$ are dense in $[0,1]$. Hence on each subinterval $[x_{i-1}, x_i]$ we have

$$M_i = \sup_{x \in [x_{i-1}, x_i]} \chi_{\mathbb{Q}}(x) = 1, \qquad m_i = \inf_{x \in [x_{i-1}, x_i]} \chi_{\mathbb{Q}}(x) = 0$$

so in particular

$$\begin{aligned} U(f,P) &= \sum_{i=1}^{n} M_i(x_i - x_{i-1}) = \sum_{i=1}^{n} (x_i - x_{i-1}) \\ &= x_1 - x_0 = 1 \\ u(f,P) &= \sum_{i=1}^{n} m_i(x_i - x_{i-1}) = 0 \end{aligned}$$

so that $U(f,P) - u(f,P) = 1$. Since this holds for arbitrary partitions, any $\epsilon < 1$ will fail the definition of integrability, so $\chi_{\mathbb{Q}}$ is not integrable. ■

4.1.2 Properties of the Integral

Theorem 4.9

1. **Additivity of Domain:** If f is integrable on $[a,b]$ and $[b,c]$ then f is integrable on $[a,c]$ and
$$\int_a^c f(x)\,\mathrm{d}x = \int_a^b f(x)\,\mathrm{d}x + \int_b^c f(x)\,\mathrm{d}x.$$
2. **Additivity of Integral:** If f,g are integral on $[a,b]$ then $f+g$ is integrable on $[a,b]$ and
$$\int_a^b [f(x)+g(x)]\,\mathrm{d}x = \int_a^b f(x)\,\mathrm{d}x + \int_a^b g(x)\,\mathrm{d}x.$$
3. **Scalar Multiplication:** If f is integrable on $[a,b]$ and $c\in\mathbb{R}$, then cf is integrable on $[a,b]$ and
$$\int_a^b cf(x)\,\mathrm{d}x = c\int_a^b f(x)\,\mathrm{d}x.$$
4. **Inherited Integrability:** If f is integrable on $[a,b]$ then f is integrable on any subinterval $[c,d]\subseteq[a,b]$.
5. **Monotonicity of Integral:** If f,g are integrable on $[a,b]$ and $f(x)\leq g(x)$ for all $x\in[a,b]$ then
$$\int_a^b f(x)\,\mathrm{d}x \leq \int_a^b g(x)\,\mathrm{d}x.$$
6. **Subnormality:** If f is integrable on $[a,b]$ then $|f|$ is integrable on $[a,b]$ and satisfies
$$\left|\int_a^b f(x)\,\mathrm{d}x\right| \leq \int_a^b |f(x)|\,\mathrm{d}x.$$

These proofs are standard and fairly fundamental results. We will not go into them at this time, but encourage the student to give them a try.

Of course, we also have the following important theorem which guarantees that integral calculus is actually computable:

Theorem 4.10: The Fundamental Theorem of Calculus

1. If f is integrable on $[a,b]$ and $x\in[a,b]$ define $F(x)=\int_a^x f(t)dt$. The function F is continuous on $[a,b]$ and moreover, $F'(x)$ exists and equals $f(x)$ at every point x at which f is continuous.
2. Let F be a continuous function on $[a,b]$ that is differentiable except possibly at finitely many points in $[a,b]$, and take $f=F'$ at all such points. If f is integrable on $[a,b]$, then $\int_a^b f(x)\,\mathrm{d}x = F(b)-F(a)$.

The fundamental theorem say that, up to functions being "almost the same" and additive constants, the processes of integration and differentiation are mutually inverting. The proof is a standard exercise and so we omit it.

4.1.3 Sufficient Conditions for Integrability

Theorem 4.6 gave multiple equivalent definitions for integrability, each with its own strengths depending on context. Of great use was parts (3) and (4) which gave conditions on integrability without needing to know the limiting integral. Unfortunately, these criteria fail to really expound upon which of our everyday functions are integrable.

There are a great deal of functions, absent of any regularity conditions such as continuity or differentiability, which prove to be integrable. Example 4.8 shows that there are also functions which fail to integrable. We will develop several sufficient conditions for integrability, one which looks similar to "Bolzano-Weierstrauss" and one which amounts to being "almost continuous," which is certainly the case with most functions we have seen and will see.

Theorem 4.11

If f is bounded and monotone on $[a, b]$ then f is integrable.

Proof. The idea of the proof is the upper and lower Riemann sums are very easy to write down for monotone functions, and the fact that f is additionally bounded means that we can make the difference between the upper and lower Riemann sums arbitrarily small (which is one of our integrability conditions). In fact, the proof is effectively identical to the one given in Example 4.7 (see Figure 30).

More formally, assume without loss of generality that f is increasing on $[a, b]$ (just replace f with $-f$ if it is decreasing and apply Theorem 4.9 (3)). For any partition $P = \{a = x_0 < x_1 < \cdots < x_n = b\}$ we then have that the lower and upper Riemann sums are determined by the left- and right-endpoints on each interval:

$$u(f, P) = \sum_{i=1}^{n} f(x_{i-1})(x_i - x_{i-1}), \qquad U(f, P) = \sum_{i=1}^{n} f(x_i)(x_i - x_{i-1}).$$

Let $\epsilon > 0$ be given and choose $\delta < \epsilon[f(b) - f(a)]^{-1}$. Let P be any partition of $[a, b]$ such that $\ell(P) < \delta$, so that

$$\begin{aligned} U(f, P) - u(f, P) &= \sum_{i=1}^{n} [f(x_i) - f(x_{i-1})] (x_i - x_{i-1}) \\ &\leq \delta \sum_{i=1}^{n} [f(x_i) - f(x_{i-1})] \\ &\leq \delta(f(b) - f(a)) \\ &\leq \frac{\epsilon}{f(b) - f(a)}(f(b) - f(a)) < \epsilon. \end{aligned}$$

Since ϵ was arbitrary, Theorem 4.6 part (3) implies that f is integrable.

Note: We could have used uniform partitions here, which would have removed the need to take $\delta < \epsilon[f(b) - f(a)]^{-1}$. Try repeating the proof using uniform partitions to test whether you actually understand the proof. □

Theorem 4.12

Every continuous function on $[a, b]$ is integrable.

It is tempting to use Theorem 4.11, since f is certainly bounded and we should be able to restrict f to intervals on which it is monotone. Applying Theorem 4.9 part (1) we would then be done. However, this does not work, since it can be shown that there are continuous functions on $[a, b]$ which are not monotone on any interval! (Think about the function $\sin(1/x)$ and consider yourself this is not monotone in any interval around 0. Such functions are similar.) Luckily, we can actually just prove the theorem directly:

Proof. The idea of the theorem is as follows: Continuous function on compact sets are necessarily uniformly continuous: in effect, this means that we can control how quickly our function grows by choosing neighbourhoods of identical but sufficiently small size. By choosing a partition to have length smaller than these neighbourhoods, we can thus control the distance between the maximum and minimum of a function on each subinterval, and force the upper and lower Riemann sums to converge.

More formally: Let $\epsilon > 0$ be given. Since any continuous function on a compact set is uniformly continuous, we can find a $\delta > 0$ such that whenever $|x - y| < \delta$ then $|f(x) - f(y)| < \frac{\epsilon}{b-a}$. Now let $P = \{x_0 < \cdots < x_n\}$ be a partition such that $\ell(P) < \delta$. The restriction of f to each subinterval $[x_{i-1}, x_i]$ is still continuous, and so by the Extreme Value Theorem, f must attain its maximum and minimum on $[x_{i-1}, x_i]$. Let ξ_M correspond to the max and ξ_m correspond to the min so that $M_i = f(\xi_M)$ and $m_i = f(\xi_m)$. Since $\xi_M, \xi_m \in [x_{i-1}, x_i]$ we have $|\xi_M - \xi_m| \leq |x_i - x_{i-1}| < \delta$ so that

$$M_i - m_i = |M_i - m_i| = |f(\xi_M) - f(\xi_m)| < \frac{\epsilon}{b-a}.$$

Hence the difference in Riemann sums becomes

$$\begin{aligned} U(f, P) - u(f, P) &= \sum_{i=1}^{n} (M_i - m_i)(x_i - x_{i-1}) \leq \sum_{i=1}^{n} \left[\frac{\epsilon}{b-a}\right] (x_i - x_{i-1}) \\ &\leq \frac{\epsilon}{b-a} \sum_{i=1}^{n} (x_i - x_{i-1}) = \frac{\epsilon}{b-a}(b-a) = \epsilon. \end{aligned}$$

Applying Theorem 4.6 part (3), this shows that f is integrable. □

With any luck, your previous courses have taught you that integration over a single point yields an integral of 0, regardless of the function. In essence, this occurs because a single point has no "width," and so any Riemann sum over it is zero. We should be able to readily extend this to any *finite* number of points, so that an integral over a finite set is still zero, but what happens when we want to talk infinitely many points? What does it mean to have zero width in this case?

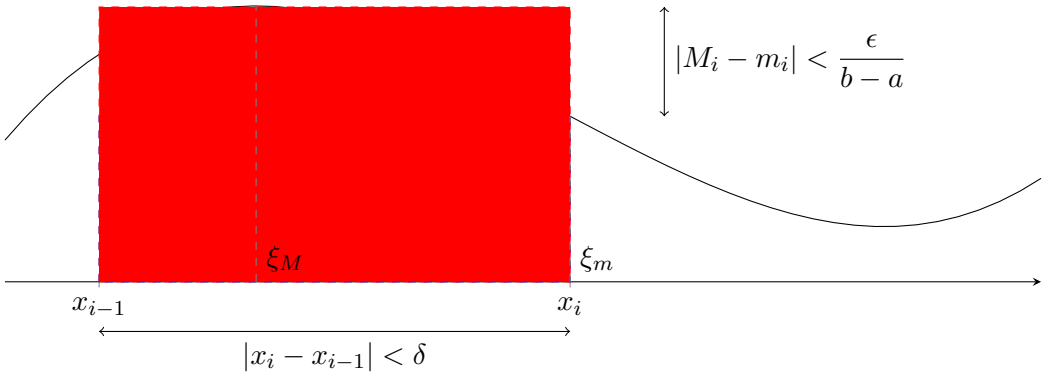


Figure 31: Since our function is uniformly continuous, whenever $|x-y| < \delta$ then $|f(x)-f(y)| < \frac{\epsilon}{b-a}$. By choosing a partition for which the maximal length of a subinterval is less than δ, we can ensure that the difference between the upper and lower integrals on each region is bounded.

Definition 4.13

If $I = [a, b]$ let the length of I be $\ell(I) = b - a$. If $\mathcal{P}(\mathbb{R})$ is the power-set of $\mathbb{R}$, we define the *Jordan outer measure* [a] as the function $m : \mathcal{P}(\mathbb{R}) \to \mathbb{R}_{\geq 0}$ given by

$$m(S) = \inf \left\{ \sum_{k=1}^{n} \ell(I_k) : \begin{array}{c} I_k \text{ is an interval} \\ S \subseteq \bigcup_{k=1}^{n} I_k \end{array} \right\}.$$

If $m(S)$ exists and $m(\partial S) = 0$, we say that S is *Jordan measurable*. If $m(S) = 0$ we say that S has *Jordan measure zero*.

[a] There is a much more useful notion called the *Lebesgue measure*, which is essentially the same as the Jordan measure except that we no longer consider a finite covering by intervals, and instead take a countable collection of intervals.

Most well behaved sets that we can think of are Jordan measurable. An example of a set which is not Jordan measurable is $\mathbb{Q} \cap [0, 1]$. Notice that $m(\partial(\mathbb{Q} \cap [0, 1])) = m([0, 1]) = 1$, so that its boundary does not have zero measure.

Example 4.14

Let S be a set containing a single point. Show that S has zero Jordan measure.

Solution. Let $S = \{x\}$ so that the point has a name. It suffices to show that for every $\epsilon > 0$, $m(S) < \epsilon$ (why?). Notice that $I = \left(x - \frac{\epsilon}{2}, x + \frac{\epsilon}{2}\right)$ covers S, and $\ell(I) = \epsilon$. Since $m(S)$ is the infimum over all such covers, we have $m(S) < \ell(I) = \epsilon$ as required. ■

Exercise: Show that the measure of any finite set is also 0.

Since integration does not seem to recognize individual points, we suspect that changing a

function at a finite collection of points should not affect the integral.

Example 4.15 Let $f(x) = x$ on $[0, 2]$ and define

$$g(x) = \begin{cases} f(x) & x \neq 1 \\ 10^6 & x = 1 \end{cases}.$$

It seems likely that f and g have the same integral on $[0, 2]$. In order to show that this is true, we apply a tried-and-tested analysis technique, which essentially involves ignoring the point which is different and taking a limit. More rigorously, for sufficiently small $\epsilon > 0$, let $U_\epsilon = (1 - \epsilon, 1 + \epsilon)$. On $V_\epsilon = [0, 2] \setminus U_\epsilon = [0, 1 - \epsilon] \cup [1 + \epsilon, 2]$ we have that $f(x) = g(x)$, and these are integrable since they are continuous on V_ϵ. Furthermore, by Theorem 4.9 we have

$$\begin{aligned} \int_0^2 g(x)\,\mathrm{d}x &= \int_{V_\epsilon} g(x)\,\mathrm{d}x + \int_{U_\epsilon} g(x)\,\mathrm{d}x \\ &= \int_{V_\epsilon} f(x)\,\mathrm{d}x + \int_{U_\epsilon} g(x)\,\mathrm{d}x. \end{aligned}$$

We want to show that in the limit $\epsilon \to 0$ we get $\int_{U_\epsilon} g(x)\,\mathrm{d}x \to 0$, so that $\int_0^2 f(x)\,\mathrm{d}x = \int_0^2 g(x)\,\mathrm{d}x$. While the approximation is rather terrible, notice that $g(x) \geq 0$ for all $x \in U_\epsilon$ and

$$\max_{x \in U_\epsilon} g(x) = 10^6,$$

so that $0 \leq \int_{U_\epsilon} g(x)\,\mathrm{d}x \leq 2\epsilon 10^6$. By the Squeeze Theorem, it then follows that

$$\int_{U_\epsilon} g(x)\,\mathrm{d}x \xrightarrow{\epsilon \to 0} 0.$$

▲

Theorem 4.16

If $S \subseteq [a, b]$ is a Jordan measure zero set, and $f : [a, b] \to \mathbb{R}$ is bounded and continuous everywhere except possibly at S, then f is integrable.

Proof. Let M and m be the supremum and infimum of f on $[a, b]$ and let $\epsilon > 0$ be given. Since S has Jordan measure zero, we can find a finite collection of intervals $(I_j)_{j=1}^k$ such that $S \subseteq \cup_j I_j \subseteq [a, b]$ and $\sum_j \ell(I_j) < \frac{\epsilon}{2(M-m)}$. Set $W = \cup_j I_j$ and $V = [a, b] \setminus W$. Since f is continuous on V, it is integrable on V and hence there exists some partition P such that $U(f|_V, P) - u(f|_V, P) < \frac{\epsilon}{2}$. If necessary, refine P so that it contains the endpoints of the intervals I_j. Writing the upper and lower Riemann sums over $[a, b]$ we get

$$U(f, P) = U(f|_W, P) + U(f|_V, P), \qquad u(f, P) = u(f|_W, P) + u(f|_V, P).$$

Since we already know how to bound the V contribution, we need now only look at the W contri-

bution. Notice on W we have

$$U(f|_W, P) - u(f|_W, P) < \sum_{j=1}^{k} (M-m)\ell(I_j) \leq (M-m)\frac{\epsilon}{2(M-m)} = \frac{\epsilon}{2},$$

thus

$$\begin{aligned} U(f,P) - u(f,P) &= [U(f|_W, P) - u(f|_W, P)] + [U(f|_V, P) - u(f|_V, P)] \\ &\leq \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon. \end{aligned}$$

□

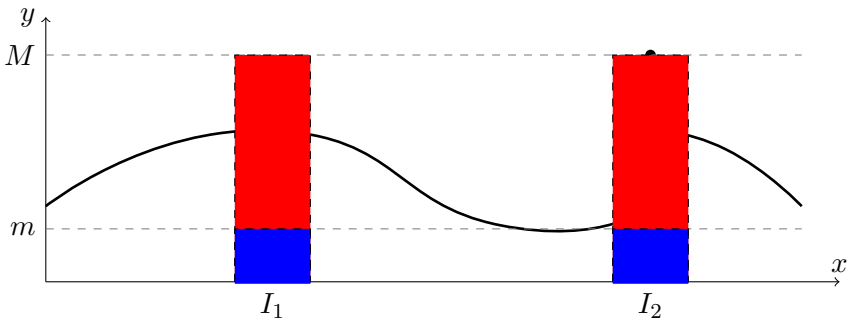


Figure 32: The set $W = I_1 \cup I_2$ contains the discontinuities of our function. Since our function is continuous away from W, we can make the difference between the upper and lower sums as small as we want, hence we need only bound the function on W. The difference in height will always be at worst $M - m$, but we can make the length of the intervals I_1 and I_2 as small as we want, making the W contribution arbitrarily small.

Corollary 4.17

If f, g are integrable on $[a, b]$ and $f = g$ up to a set of Jordan measure zero, then $\int_a^b f(x)\,\mathrm{d}x = \int_a^b g(x)\,\mathrm{d}x$.

This is an easy corollary, whose proof effectively emulates that of Remark 4.15, so we leave it as an exercise for the student.

4.2 Integration in $\mathbb{R}^n$

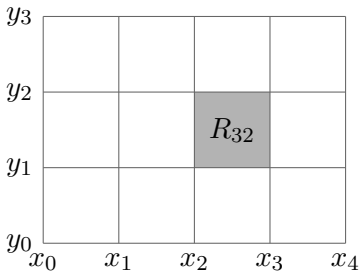
The process of integration for $\mathbb{R}^n$ is effectively identical to that of $\mathbb{R}$, except now we must use rectangles instead of intervals, rectangles being a possible analog for higher-dimensional intervals. We start by focusing on $\mathbb{R}^2$ to gain a familiarity with the concepts before moving to general $\mathbb{R}^n$.

Note: It could be argued that the generalization of a closed interval $[a, b]$ is a closed ball. One can develop the following theory with balls, but taking the area/volume of balls usually involves a nasty factor of π hanging around. We want to avoid this, so let us just use rectangles.

4.2.1 Integration in the Plane

cp By realizing (non-canonically) $\mathbb{R}^2 = \mathbb{R}\times\mathbb{R}$, we can define a *rectangle* R in $\mathbb{R}^2$ as any set which can be written as $R = [a,b]\times[c,d]$: this truly looks like a rectangle if drawn in the plane. A partition of R may then be given by a partition of $[a,b]$ and $[c,d]$; namely, if $P_x = \{a = x_0 < \cdots < x_n = b\}$ and $P_y = \{c = y_0 < \cdots < y_m = d\}$ are partitions of their respective intervals, then $P = P_x \times P_y$ is a partition of R, with subrectangles

$$R_{ij} = [x_{i-1}, x_i]\times[y_{j-1}, y_j], \qquad \substack{i=1,\dots,n\\ j=1,\dots,m}.$$



It should be intuitively clear that the area of R_{ij} will be given by $A(R_{ij}) = (x_i - x_{i-1})(y_j - y_{j-1})$, in which case a *Riemann sum* for $f : \mathbb{R}^2 \to \mathbb{R}$ over the partition P is given by

$$S(f,P) = \sum_{\substack{i=1,\dots,n\\ j=1,\dots,m}} f(t_{ij})A(R_{ij}), \qquad t_{ij} \in R_{ij}.$$

The notion of left- and right-Riemann sums no longer make sense, but certainly the upper and lower Riemann sums are still well-defined:

$$U(f,P) = \sum_{\substack{i=1,\dots,n\\ j=1,\dots,m}} \left[\sup_{\mathbf{x}\in R_{ij}} f(\mathbf{x})\right] A(R_{ij}), \quad u(f,P) = \sum_{\substack{i=1,\dots,n\\ j=1,\dots,m}} \left[\inf_{\mathbf{x}\in R_{ij}} f(\mathbf{x})\right] A(R_{ij}).$$

The usual definitions of Riemann integrability then carry over directly from Definition 4.5. Restricting ourselves to just one definition for the moment, we will then say that $f : R \to \mathbb{R}$ is Riemann integrable if for any $\epsilon > 0$ we can find a partition P such that $U(f,P) - u(f,P) < \epsilon$, and we will write the integral as

$$\iint_R f \,\mathrm{d}A, \qquad \text{or} \qquad \iint f(x,y)\,\mathrm{d}x\,\mathrm{d}y.$$

The usual theorems of integration apply:

Theorem 4.18

1. **Linearity of the Integral:** If f_1, f_2 are integrable on R and $c_1, c_2 \in \mathbb{R}$ then $c_1 f_1 + c_2 f_2$ is integrable on S and
$$\iint_R [c_1 f_1 + c_2 f_2]\,\mathrm{d}A = c_1 \iint_R f_1\,\mathrm{d}A + c_2 \iint_R f_2\,\mathrm{d}A.$$
2. **Additivity of Domain:** If f is integrable on disjoint rectangles R_1 and R_2 then f is integrable on $R_1 \cup R_2$ and
$$\iint_{R_1 \cup R_2} f\,\mathrm{d}A = \iint_{R_1} f\,\mathrm{d}A + \iint_{R_2} f\,\mathrm{d}A.$$
3. **Monotonicity:** If $f_1 \leq f_2$ are integrable functions on R then
$$\iint_R f_1\,\mathrm{d}A \leq \iint_R f_2\,\mathrm{d}A.$$
4. **Subnormality:** If f is integrable on R and $|f|$ is integrable on R and $\left|\iint f\,\mathrm{d}A\right| \leq \iint |f|\,\mathrm{d}A$.
5. If f is continuous, then f is integrable.

Up to zero measure: As before, we will only be interested in functions up to sets of zero measure. The notion of zero measurable sets immediately generalizes from the 1-dimensional case as follows:

Definition 4.19

The *Jordan outer measure* of a set $S \in \mathbb{R}^2$ is defined to be
$$m(S) = \inf\left\{ \sum_{i,j} A(R_{ij}) : \begin{array}{c} R_{ij} \text{ is a rectangle} \\ S \subseteq \bigcup_{ij} R_{ij} \end{array} \right\}.$$
If $m(S)$ is defined, and $m(\partial S) = 0$ we say that S is *Jordan measurable*. Additionally, if S is Jordan measurable and $m(S) = 0$, we say S has Jordan measure zero.

For any reasonably nice set, one can think of the Jordan measure as the area. For example, if $B^2 = \left\{(x, y) \in \mathbb{R}^2 : x^2 + y^2 \leq 1\right\}$ is the unit disk, then $m(B^2) = \pi$ (though this is extremely tough to show by hand!). Intuitively, zero-measure sets of $\mathbb{R}^2$ are those which do not have any area, and one would suspect that "one-dimensional" objects should have no area.

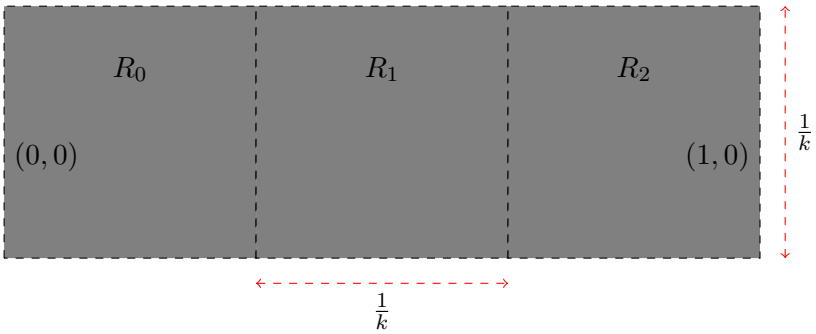
Example 4.20

Show that the set $S = [0, 1] \times \{0\} \subseteq \mathbb{R}^2$ has zero Jordan measure.

Solution. Fix a positive integer $k \in \mathbb{Z}_+$ and consider the k squares R_i defined as

$$R_i = \left[\frac{i}{k}, \frac{i+1}{k}\right] \times \left[-\frac{1}{2k}, \frac{1}{2k}\right], \qquad i = 0, \dots, k-1$$

each of which has an area of $\frac{1}{k^2}$. The student can check that $S \subseteq \cup_{i=0}^{k-1} R_i$ so that $\{R_i\}$ cover S. Moreover, there are exactly k such squares, so their total area is $k \times \frac{1}{k^2} = \frac{1}{k}$. Since the Jordan measure is the infimum over all possible measures, we have that $\mu(S) \leq \frac{1}{k}$. Since we chose k arbitrarily, we can make $\mu(S)$ as small as we want, showing that $\mu(S) = 0$. ■



This likely seemed like an unnecessarily difficult way of doing the problem: certainly we could have just placed an rectangle of length 1 and height $\frac{1}{k}$ around the interval and let k shrink to zero. The important point here is that as we let k grow, the number of rectangles increased proportional to k, while the area decreased proportional to k^2.

Theorem 4.21

If $\mathbf{f} : \mathbb{R} \to \mathbb{R}^2$ is of class C^1, then for every interval $I \subseteq \mathbb{R}$ we have that $\mathbf{f}(I)$ has zero content.

Proof. As mentioned before, the idea of the proof is that the image of $\mathbb{R}$ under a C^1 function has no width, but how do we show this? By thinking of $\mathbf{f}(t) = (f_1(t), f_2(t))$ as a curve, its derivatives $\mathbf{f}'(t) = (f_1'(t), f_2'(t))$ represent the velocity of the curve. If we take the maximum horizontal speed $C = \max f_1'(t)$, then by restricting to an interval $[a, b]$, we see that the maximum horizontal distance that the function can travel is bounded above by $C \times (b - a)$; that is, distance = speed $\times$ time. Proceeding similarly with the vertical direction means that we can put $\mathbf{f}([a, b])$ into a box whose area is proportional to $C(b - a)^2$, and since we have control over how to partition our curve, we can always force this number as small as we want.

More formally, let I be a fixed interval and $\epsilon > 0$ be given. Since $\mathbf{f}$ is of class C^1, we know that $|\mathbf{f}'(t)|$ is continuous and hence attains its max and min on I. Let $S = \max_{t \in I} |\mathbf{f}'(t)|$ and choose an integer k such that $k > \frac{\ell(I)^2 S^2}{\epsilon}$. Let P be a uniform partition of I into k sub-intervals and notice then that

$$\ell(P) = \frac{\ell(I)}{k} < \frac{1}{S}\sqrt{\frac{\epsilon}{k}}.$$

Fix a sub-interval $[x_i, x_{i+1}]$ and apply the Mean Value Theorem to the component functions

$f_1(t)$ and $f_2(t)$ on this interval to find $t_i \in [x_i, x_{i+1}]$ satisfying

$$|f_1(x_i) - f_1(x_{i-1})| \leq |f_1'(t_i)||x_i - x_{i-1}| \leq S|x_i - x_{i-1}| \leq S\frac{1}{S}\sqrt{\frac{\epsilon}{k}} = \sqrt{\frac{\epsilon}{k}},$$

and similarly $|f_2(x_i) - f_2(x_{i-1})| < \sqrt{\frac{\epsilon}{k}}$. Hence $\mathbf{f}([x_i, x_{i-1}])$ is contained is a box whose with area at most $\left(\sqrt{\frac{\epsilon}{k}}\right)^2 = \frac{\epsilon}{k}$. Since there are k such partitions, this means that $\mathbf{f}(I)$ can covered by k-rectangles whose total width at is at most $k \times \frac{\epsilon}{k} = \epsilon$. Since ϵ was arbitrary, this completes the proof. □

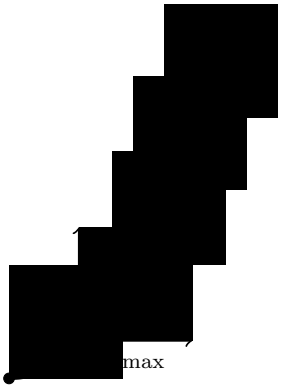


Figure 33: By looking at the maximum speed that the function attains, one can find the worst case box the each subinterval (marked by black dots) fits into. As we increase the number of subintervals, the number of necessary boxes increases linearly, while the area of each box decreases proportional to the $-n$-th power.

Definition 4.22

A curve $f : [a, b] \to \mathbb{R}^n$ is said to be *piecewise* C^1 if it is C^1 at all but a finite number of points.

Corollary 4.23

Any set $S \subseteq \mathbb{R}^2$ such that ∂S is defined by a piecewise C^1 curve is Jordan measurable.

Proof. The proof of this corollary is immediate. If S has a boundary defined by a piecewise smooth curve, then its boundary has zero Jordan measure by Theorem 4.21. This is precisely the definition for S to be Jordan measurable. □

Theorem 4.24

If R is a rectangle and f is continuous on R up to a set of Jordan measure 0, then f is integrable.

Proof. This proof is effectively the same as Theorem 4.16. □

Integrability over non-Rectangles: Of course, we would like to be able to integrate functions over other (bounded) sets that aren't just rectangles! If $S \subseteq \mathbb{R}^2$ is a bounded set, we can always find a sufficiently large rectangle R containing S. We thus need only extend $f : S \to \mathbb{R}^2$ in a way that should not affect which rectangle we take. The way to do this is to define the *characteristic function of* S:

$$\chi_S(\mathbf{x}) = \begin{cases} 1 & \mathbf{x} \in S \\ 0 & \text{otherwise} \end{cases}.$$

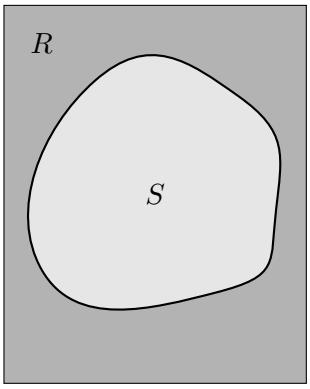


Figure 34: Every bounded set can be placed inside a rectangle.

Thus the function $f\chi_S : R \to \mathbb{R}$ is just $f(\mathbf{x})$ on S and is identically 0 everywhere else. Note that the choice of enveloping rectangle really doesn't affect $f\chi_S$ since we have extended f by zero outside of S. We would now like to check that $f\chi_S$ is integrable on R so that it makes sense to write down $\iint_S f \, \mathrm{d}A$.

Theorem 4.25

If S is Jordan measurable and the set of discontinuities of $f : S \to \mathbb{R}^2$ has zero measure, then f is Riemann integrable on S.

Proof. It is easy to convince ourselves that the discontinuities of the characteristic function χ_S occur exactly at the boundary ∂S. If S is Jordan measurable, then $m(\partial S) = 0$. The discontinuities of f are also Jordan measure zero, hence the total discontinuities of $f\chi_S$ has zero measure, so this function is integrable.

More rigorously, fix a rectangle R such that $S \subseteq R$. Let D be the set of discontinuities of f and note that the set of discontinuities of χ_S is given by ∂S. It then follows that the set of discontinuities of $f\chi_S$ on R is $D \cup \partial S$. Since the union of zero measure sets has zero measure, $f\chi_S$ has zero-measure discontinuities on R and hence is Riemann Integrable by Theorem 4.24. □

In particular, we have the following Corollary:

Corollary 4.26

If $S \subseteq \mathbb{R}^2$ is Jordan measurable then $m(S) = \int_S \chi_S$.

4.2.2 Integration Beyond 2-dimensions

Now we generalize things for (hopefully!) the last time. A rectangle in $\mathbb{R}^n$ is any set of the form

$$R = [a_1, b_1] \times \cdots \times [a_n, b_n],$$

and has volume $V(R) = (b_1 - a_1) \times \cdots \times (b_n - a_n)$. A partition of R may be specified by an n-partitions of $\mathbb{R}$, each one decomposing $[a_i, b_i]$. For $(i_1, \ldots, i_n)$ a collection of positive integers, let $R_{(i_1,\ldots,i_n)}$ be the sub-rectangle corresponding to the $(i_1, \ldots, i_n)$ element. A tagged Riemann sum over R is any sum of the form

$$S(f, P) = \sum_{(i_1,\ldots,i_n)} f\left(t_{(i_1,\ldots,i_n)}\right) V(R_{(i_1,\ldots,i_n)}), \qquad t \in R_{(i_1,\ldots,i_n)}.$$

As usual, one can define the upper $U(f, P)$ and lower $u(f, P)$ Riemann sums using the supremum and infimum, in which case we say that $f : R \subseteq \mathbb{R}^n \to \mathbb{R}$ is integrable precisely when for every $\epsilon > 0$ there exists a partition P such that

$$U(f, P) - u(f, p) < \epsilon.$$

To extend the definition of the integral beyond rectangles, we once again introduce the Jordan measure. The Jordan measure of a set S is defined as the infimum of the volumes of all covering rectangles, and S is Jordan measurable if its boundary has measure zero. If $k < n$ then the image of a C^1 map $f : \mathbb{R}^k \to \mathbb{R}^n$ has Jordan measure zero. A function $f : S \to \mathbb{R}$ is then integrable if S is Jordan measurable and if the set of discontinuities of f on S has Jordan measure zero. We denote the integral of such a function as:

$$\int \cdots \int_S f dV = \int \cdots \int f(\mathbf{x}) d^n x = \int \cdots \int f(x_1, \ldots, x_n)\, \mathrm{d}x_1 \cdots \mathrm{d}x_n.$$

The only thing left to mention is the Mean Value Theorem:

Theorem 4.27: Mean Value Theorem

Let $S \subseteq \mathbb{R}^n$ be a compact, connected, and Jordan measurable set, with continuous functions $f, g : S \to \mathbb{R}$ satisfying $g \geq 0$. Then there exists a point $\mathbf{a} \in S$ such that

$$\int \cdots \int_S f(\mathbf{x}) g(\mathbf{x}) d^n \mathbf{x} = f(\mathbf{a}) \int \cdots \int_S g(\mathbf{x}) d^n \mathbf{x}.$$

Proof. Since S is compact and f is continuous on S, it attains its max and min on S, say M and m respectively. Since $g \geq 0$ we have

$$m \int \cdots \int_S g(\mathbf{x}) d^n \mathbf{x} \leq \int \cdots \int_S f(\mathbf{x}) g(\mathbf{x}) d^n \mathbf{x} \leq M \int \cdots \int_S g(\mathbf{x}) d^n \mathbf{x}.$$

or equivalently

$$m \leq \frac{\int \cdots \int_S f(\mathbf{x})g(\mathbf{x})d^n\mathbf{x}}{\int \cdots \int_S g(\mathbf{x})d^n\mathbf{x}} \leq M.$$

Since f is continuous and S is connected, f is surjective on $[m, M]$ and hence the Intermediate Value Theorem implies the middle term is $f(\mathbf{a})$ for some $\mathbf{a} \in S$, as required. □

The student has likely noticed that this section is filled with theory, and zero computation. The reason for this is that computing integrals in multiple dimensions is an incredibly difficult thing to do. The reason is that for any partitioning subrectangle, we are looking at the supremum/infimum of our function restricted to that n-dimensional rectangle. In a sense, we have to integrate in all n-dimension simultaneously. This is not easy to do, so our next section will introduce a method by which we integrate our function in 'slices.'

4.3 Iterated Integrals

In developing the theory of integration in the plane and higher, it was necessary to consider partitions of rectangles and hence, in essence, to consider the area of function with respect to an infinitesimal area $\mathrm{d}A$. Of importance is that this area term encapsulates information about every dimension simultaneously, but simultaneity is a computational obstacle. For example, when learning to differentiate a multivariate function, we needed to invest a great deal of energy into simply analyzing the change of the function *in a single, specific direction* (ie the partial derivatives). If we want to know how the function is changing in an arbitrary direction, we then have the directional derivative $d_u f = \nabla f \cdot u$, so that the gradient ∇f somehow represents that simultaneous derivative of f at any point.

Consider now the problem of computing the upper sum $U(f, P)$ for a function f on a partition P. For *each* subrectangle R_{ij}, one would need to determine the supremum of f on R_{ij}. If our function is C^1, even this involves solving for critical points on the interior, then using the method of Lagrange multipliers on the boundary. What a nightmare!

From our single variable calculus days, we know that integration is often more difficult that the formulaic recipe-following nature of differentiation. The fact that "simultaneous" differentiation required so much work does not bode well for the idea of simultaneous integration. So as mathematicians, we won't bother trying to figure it out. Instead, we will apply the mathematicians favourite tool: We will reduce simultaneous integration to a problem we have solved before: one dimensional integration.

As always, we start out with a rectangle $R = [a, b] \times [c, d]$ in the plane, partitioned into $P = P_x \times P_y = \{x_0 < \cdots x_n\} \times \{y_0 < \cdots y_m\}$. The prototypical Riemann sum which corresponds to this partition is

$$S(f, P) = \sum_{\substack{i\in\{1,\ldots,n\}\\ j\in\{1,\ldots,m\}}} f(\widetilde{\mathbf{x}_{ij}})A(R_{ij}) = \sum_{\substack{i\in\{1,\ldots,n\}\\ j\in\{1,\ldots,m\}}} f(\tilde{x}_i, \tilde{y}_j)\Delta x_i \Delta y_j$$

where $(\tilde{x}_i, \tilde{y}_j) \in [x_{i-1}, x_i] \times [y_{j-1}, y_j]$ and $\Delta x_i = (x_i - x_{i-1}), \Delta y_j = (y_j - y_{j-1})$. Now if we look at

this sum, we can decompose it as

$$\sum_{\substack{i\in\{1,\dots,n\}\\ j\in\{1,\dots,m\}}} f(\tilde{x}_i,\tilde{y}_j)\Delta x_i\Delta y_j = \sum_{j=1}^{m}\Bigg[\underbrace{\sum_{i=1}^{n} f(\tilde{x}_i,\tilde{y}_j)\Delta x_i}_{\approx\int_a^b f(x,\tilde{y}_j)\,\mathrm{d}x}\Bigg]\Delta y_j. \tag{4.1}$$

The heuristic idea is as follows: if we define the function

$$g_k(\tilde{y}) = \lim_{\ell(P_x)\to 0} S(f, P_x\times P_y) = \int_a^b f(x,\tilde{y})\,\mathrm{d}x$$

then (4.1) gives

$$\begin{aligned}\int_R f(x,y)\,\mathrm{d}x &= \lim_{\ell(P)\to 0}\sum_{j=1}^{m}\left[\sum_{i=1}^{n} f(\tilde{x}_i,\tilde{y}_j)\Delta x_i\right]\Delta y_j\\ &= \lim_{\ell(P_y)\to 0}\sum_{j=1}^{m} g_k(\tilde{y}_k)\Delta y_k\\ &= \int_c^d\left[\int_a^b f(x,y)\,\mathrm{d}x\right]\,\mathrm{d}y\end{aligned}$$

Now strictly speaking, what we have done here is not kosher, since in particular we had to assume two things:

1. The limit $\ell(P)\to 0$ is equivalent to first doing $\ell(P_x)\to 0$ then $\ell(P_y)\to 0$, and
2. Each of the "slices" $f(x,\tilde{y}_k)$ is integrable.

If we make these assumptions and add a pinch of rigour (which we will not do here), we get

Theorem 4.28: Fubini's Theorem

Let $R=[a,b]\times[c,d]$ be a rectangle and $f: R\to\mathbb{R}$ an integrable function on R. If for each $y_0\in[c,d]$ the function $f_{y_0}:[a,b]\to\mathbb{R}$ given by $x\mapsto f(x,y_0)$ is integrable on $[a,b]$, and $g(y)=\int_a^b f(x,y)\,\mathrm{d}x$ is integrable on $[c,d]$, then

$$\int_R f\,\mathrm{d}A = \int_c^d\left[\int_a^b f(x,y)\,\mathrm{d}x\right]\,\mathrm{d}y.$$

Of course, the theorem also holds with the roles of x and y reversed.

Example 4.29

Determine the volume under the function $f(x,y)=xe^{x^2-y}$ on the rectangle $R=[0,1]\times[0,1]$.

Solution. Since f is a continuous function on R it is integrable, and so certainly each of the slices $f_y(x)$ or $f_x(y)$ are integrable as well. We will do the calculation both ways to show that the integral

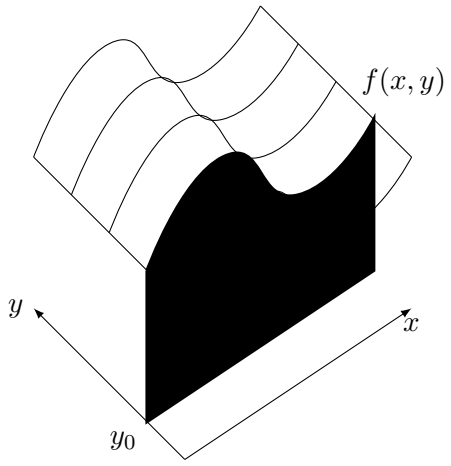


Figure 35: Fixing a y_0, we look at the function $f(x, y_0)$. If this function is integrable for each y_0, then the value of $g(y_0)$ is precisely $\int_a^b f(x, y_0)\,\mathrm{d}x$, the shaded region. If g is also integrable, then we can compute the integral of f by these slices.

yields the same results. If we integrate first with respect to x then y, we have

$$\begin{aligned}\int_0^1 \left[\int_0^1 xe^{x^2-y}\,\mathrm{d}x\right]\,\mathrm{d}y &= \int_0^1 \left[\frac{1}{2}e^{x^2-y}\right]_{x=0}^1\,\mathrm{d}y \\ &= \frac{1}{2}(e-1)\int_0^1 e^{-y}\,\mathrm{d}y \\ &= \frac{1}{2}(e-1)\left[-e^{-y}\right]_0^1 = -\frac{1}{2}(e-1)(e^{-1}-1) \\ &= 1-\cosh(1).\end{aligned}$$

Conversely, let us instead integrate with respect to y first. We have

$$\begin{aligned}\int_0^1 \left[\int_0^1 xe^{x^2-y}\,\mathrm{d}y\right]\,\mathrm{d}x &= -(e^{-1}-1)\int_0^1 xe^{x^2}\,\mathrm{d}x \\ &= -(e^{-1}-1)(e-1) = 1-\cosh(1).\end{aligned}$$

As expected, the result was the same either way. ■

Of course, the above example was very simple since we could decompose our function $f(x, y) = f_1(x)f_2(y)$, but the result still holds even when such a decomposition is not possible.

Now rectangles are rather boring objects about which to integrate, so we again look at Jordan measurable sets $S \subseteq \mathbb{R}^2$. In particular, we will suppose that S has its boundary defined by piecewise C^1 curves; say

$$S = \{(x, y) : a \le x \le b, \alpha(x) \le y \le \beta(x)\}.$$

In this case, our integration becomes

$$\int_S f\,\mathrm{d}A = \int_a^b \left[\int_{\alpha(x)}^{\beta(x)} f(x, y)\,\mathrm{d}y\right]\,\mathrm{d}x.$$

Often times, the most difficult part of solving an iterated integral question comes from determining the bounding functions, though sometimes we are fortunate and they are already prescribed.

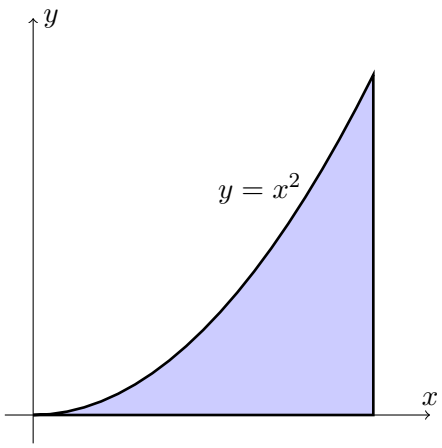
Example 4.30

Find the integral of the function $f(x,y) = \frac{y}{x^5+1}$ on the intersection of

$$\{y \geq 0\} \cap \{x \leq 1\} \cap \left\{y \leq x^2\right\}.$$

Solution. In any situation of performing iterated integrals, it is best to draw a diagram of the region over which we are integrating. In our case, we can see that the region may be summarily described as

$$S = \left\{(x,y) : 0 \leq x \leq 1, 0 \leq y \leq x^2\right\}.$$



Certainly our function is continuous on S (since $x^5 + 1 \neq 0$ on this set) and so is integrable, along with any of the slices. This means we may apply Fubini's theorem:

$$\begin{aligned}\iint_S f \, dA &= \int_0^1 \left[\int_0^{x^2} \frac{y}{x^5+1} \, dy\right] dx \\ &= \frac{1}{2}\int_0^1 \left[\frac{y^2}{x^5+1}\right]_0^{x^2} dx = \frac{1}{2}\int_0^1 \frac{x^4}{x^5+1} \, dx \\ &= \frac{1}{10}\ln|x^5+1|_0^1 = \frac{\ln(2)}{10}.\end{aligned}$$

■

Note that the region in Example 4.30 also could have been described by

$$S = \{0 \leq y \leq 1, \ \sqrt{y} \leq x \leq 1\},$$

so we also could have (attempted to) compute the integral as

$$\iint_S f \, dA = \int_0^1 \left[\int_{\sqrt{y}}^1 \frac{y}{x^5+1} \, dx\right] dy.$$

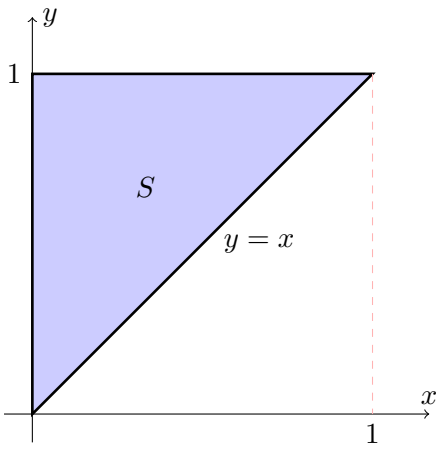


Figure 36

This probably would not have worked as nicely though, since $\frac{1}{x^5+1}$ is not easy to integrate. This suggests that being able to rewrite our domain is a useful skill, since sometimes we are given the boundary, but the problem is not amenable to the given description.

Example 4.31

Determine the integral of the function $f(x,y) = e^{y^2}$ on the region bounded by the lines $y = 1$, $x = 0$ and $y = x$.

Solution. The region is a simple triangle, given in Figure 36, which can be written as either of the following two sets

$$\begin{aligned} S &= \{0 \le x \le 1,\ x \le y \le 1\} \\ &= \{0 \le y \le 1,\ 0 \le x \le y\}. \end{aligned}$$

If we try to use the first description, we get

$$\int_S f \,\mathrm{d}A = \int_0^1 \left[\int_x^1 e^{y^2} \,\mathrm{d}y\right] \mathrm{d}x$$

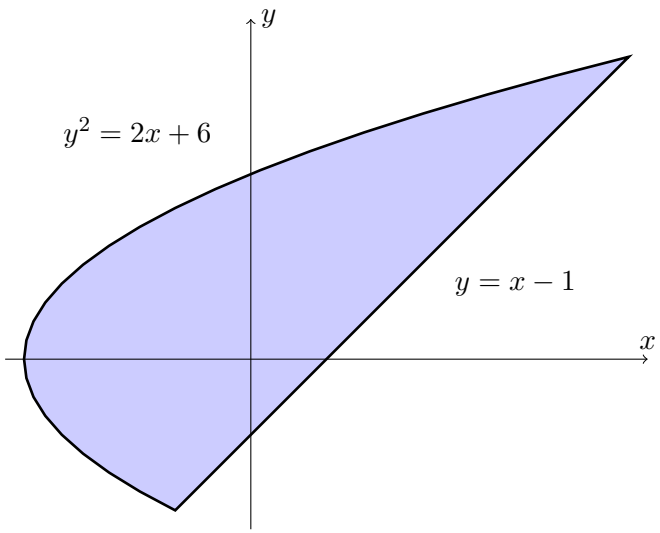
but the function e^{y^2} has no elementary anti-derivative, and we are stuck. On the other hand, using the second description gives

$$\begin{aligned} \int_S f \,\mathrm{d}A &= \int_0^1 \left[\int_0^y e^{y^2} \,\mathrm{d}x\right] \mathrm{d}y \\ &= \int_0^1 \left[x e^{y^2}\right]_{x=0}^{x=y} \mathrm{d}y = \int_0^1 y e^{y^2} \,\mathrm{d}y \\ &= \left[\frac{1}{2} e^{y^2}\right]_{y=0}^1 = \frac{1}{2}(e-1). \end{aligned}$$

■

Example 4.32

Determine $\iint_S xy \, \mathrm{d}A$ where S is the region bounded by $y = x - 1$ and $y^2 = 2x + 6$.



Solution. We begin by drawing a rough picture of what the boundary looks like. Notice that the intersection of these two lines occurs when

$$(x-1)^2 = 2x + 6, \qquad \Leftrightarrow \qquad x^2 - 4x - 5 = 0, \qquad \Leftrightarrow \qquad x = 5, -1,$$

which corresponds to the pairs $(-1, -2)$ and $(5, 4)$. Now our figure shows that it will be very hard to write this as $\{a \le x \le b, \alpha(x) \le y \le \beta(x)\}$, so instead we try to switch the variables. In that case, notice that we can write S as

$$S = \left\{ -2 \le y \le 4, \ \frac{1}{2}y^2 - 3 \le x \le y + 1 \right\}.$$

Now integrating, we get

$$\begin{aligned}
\int_S xy \, \mathrm{d}A &= \int_{-2}^{4} \left[\int_{\frac{1}{2}y^2 - 3}^{y+1} xy \, \mathrm{d}x \right] \mathrm{d}y \\
&= \frac{1}{2} \int_{-2}^{4} \left[x^2 y \right]_{\frac{1}{2}y^2 - 3}^{y+1} \mathrm{d}y \\
&= \frac{1}{2} \int_{-2}^{4} y \left[(y+1)^2 - \left(\frac{1}{2}y^2 - 3 \right)^2 \right] \mathrm{d}y \\
&= \frac{1}{2} \int_{-2}^{4} \left[-\frac{y^5}{4} + 4y^3 + 2y^2 - 8y \right] \mathrm{d}y \\
&= \frac{1}{2} \left[-\frac{y^6}{24} + y^4 + \frac{2y^3}{3} - 4y^2 \right]_{-2}^{4} = 36.
\end{aligned}$$

■

Thus far we have been fortunate: most of our examples are clearly C^1 on the region on which they are defined, and all the hypotheses of Fubini's theorem become easily verified. However, there are instances where Fubini will not hold, as the following example demonstrates.

Example 4.33

Consider the function $f(x,y) = \dfrac{xy(x^2-y^2)}{(x^2+y^2)^3}$ on the rectangle $R = [0,1]\times[0,1]$.

Solution. Let us naïvely assume that Fubini's theorem applies. Notice that f is symmetric in x and y with the exception of a negative sign in the numerator. Hence

$$\begin{aligned}
\int_0^1 \frac{xy(x^2-y^2)}{(x^2+y^2)^3}\,\mathrm{d}x &= \frac{1}{2}\int_{y^2}^{1+y^2} \frac{y(u-2y^2)}{u^3}\,\mathrm{d}u && \text{substitution with } u = x^2+y^2\\
&= \frac{y}{2}\int_{y^2}^{1+y^2} \frac{1}{u^2}\,\mathrm{d}u - y^3\int_{y^2}^{1+y^2}\frac{1}{u^3}\,\mathrm{d}u\\
&= \left[-\frac{y}{2u} + \frac{y^3}{2u^2}\right]_{y^2}^{1+y^2}\\
&= -\frac{y}{2(1+y^2)} + \frac{y^3}{2(1+y^2)^2}\\
&= -\frac{y}{2(1+y^2)^2}.
\end{aligned}$$

This in turn is easily integrated with respect to y, to yield

$$\begin{aligned}
\int_0^1 \left[-\frac{y}{2(1+y^2)^2}\right]\mathrm{d}y &= -\frac{1}{4}\int_1^2 \frac{1}{u^2}\,\mathrm{d}u && u = 1+y^2\\
&= -\frac{1}{4}\left[\frac{1}{u}\right]_1^2 = \frac{1}{8}.
\end{aligned}$$

The computation in the other order is exactly the same, except one gets an extra negative sign coming from the original substitution $u = x^2+y^2$. Thus

$$\int_0^1\left[\int_0^1 \frac{xy(x^2-y^2)}{(x^2+y^2)^3}\,\mathrm{d}y\right]\mathrm{d}x = -\int_0^1\left[\int_0^1 \frac{xy(x^2-y^2)}{(x^2+y^2)^3}\,\mathrm{d}x\right]\mathrm{d}y$$

and the integrals are *not* equal. The reason why Fubini's theorem fails is that f is not integrable on R. Indeed, f is not even bounded on R and so certainly cannot be integrable.

One might wonder if the only way the solutions will disagree is a minus-sign. The answer is no, as can be checked by using a non-symmetric rectangle. As an exercise, the student should check that if the rectangle $R = [0,2]\times[0,1]$ is used instead, the resulting integrals will differ in value as well as sign.

■

Triple! Integrals: Of course we have limited our discussion thus far to functions of two variables, but there was no reason to (other than to keep ourselves from headaches). Naturally, we can extend to three dimensions and beyond, and so perform integration in n-variables. However, because drawing diagrams is so critical for doing iterated integrals, we typically tend to avoid doing them in 4-dimensions or greater. In this course, we will not see integrals in more than 3-variables.

This being said, what happens when we want to integrate a function in three variables? The solution is to proceed just as before, except that now we write our domain as

$$S = \{(x, y, z) : a \leq x \leq b,\ \alpha(x) \leq y \leq \beta(x),\ \varphi(x, y) \leq z \leq \psi(x, y)\},$$

and the corresponding integral becomes

$$\iiint_S f(x, y, z)\, \mathrm{d}A = \int_a^b \left[\int_{\alpha(x)}^{\beta(x)} \left[\int_{\varphi(x,y)}^{\psi(x,y)} f(x, y, z)\, \mathrm{d}z \right] \mathrm{d}y \right] \mathrm{d}x.$$

Example 4.34

Determine $\iiint_S z\ \mathrm{d}A$ if S is the set bounded by the planes $x = 0, y = 0, z = 0$ and $x+y+z = 1$.

Solution. This shape is a tetrahedron whose boundaries are the three standard unit normals $\{e_i\}_{i=1,2,3}$ and the origin $(0, 0, 0)$. Now $0 \leq x \leq 1$ is evident, and projecting into the xy-plane we see that $0 \leq y \leq 1 - x$. Finally, we clearly have that $0 \leq z \leq 1 - x - y$ so that

$$\begin{aligned}
\iiint_S z\ \mathrm{d}A &= \int_0^1 \left[\int_0^{1-x} \left[\int_0^{1-x-y} z\, \mathrm{d}z \right] \mathrm{d}y \right] \mathrm{d}x \\
&= \int_0^1 \left[\int_0^{1-x} \left[\frac{z^2}{2} \right]_0^{1-x-y} \mathrm{d}y \right] \mathrm{d}x \\
&= \frac{1}{2} \int_0^1 \left[\int_0^{1-x} (1 - x - y)^2\, \mathrm{d}y \right] \mathrm{d}x = \frac{1}{2} \int_0^1 \left[-\frac{(1 - x - y)^3}{3} \right]_0^{1-x} \mathrm{d}x \\
&= \frac{1}{6} \int_0^1 (1 - x)^3\, \mathrm{d}x = \frac{1}{6} \left[-\frac{(1 - x)^4}{4} \right]_0^1 = \frac{1}{24}
\end{aligned}$$

■

Example 4.35

Determine $\iiint_S (2x + 4z) dV$ where S is the region bounded by the planes $y = x$, $z = x$, $z = 0$, and $y = x^2$.

Solution. The student should stare at these equations for some time and try to visualize the space. In particular, a nice parameterization of the space can be given as

$$S = \left\{0 \leq x \leq 1,\ x^2 \leq y \leq x,\ 0 \leq z \leq x\right\}.$$

Our function is clearly C^1 on this set, so we can apply Fubini to get

$$\begin{aligned}\iiint_S f dV &= \int_0^1 \left[\int_{x^2}^x \left[\int_0^x (2x+4z)\,\mathrm{d}z\right]\,\mathrm{d}y\right]\,\mathrm{d}x \\ &= \int_0^1 \left[\int_{x^2}^x 2x^2+2x^2\,\mathrm{d}y\right]\,\mathrm{d}x \\ &= 2\int_0^1 \left(4x^3-4x^4\right)\,\mathrm{d}x \\ &= 4\left[\frac{1}{4}x^4-\frac{1}{5}x^5\right]_0^1 = \frac{1}{5}.\end{aligned}$$

■

4.4 Change of Variables

There is a great idea amongst physicists that the properties of a physical system should be invariant of how you choose to look at that system. Consider for example, a driver racing around a circular track. We should be able to determine fundamental physical facts about the driver regardless of whether we are looking at the driver from the stands, from the center of the track, or even from the backseat of the car. However, each point of view offers its own advantages and disadvantages. For example, the observer at the center of the track only sees a change in the angle of the car relative to the observer, with the distance remaining constant. On the other hand, the backseat observer will see the driver experience the fictitious centrifugal force, while the external observers will simply see inertia.

Another exceptionally important example is the theory of special relativity. Effectively, if one starts with the simple (but unintuitive) assumption that the speed of light is constant in every frame of reference, then much of theory of special relativity (such as time/length dilation, breaking simultaneity) can be derived simply by analyzing what happens from different view points. This section is dedicated to analyzing how this is done mathematically, and how we can use this to make headway on difficult integrals.

4.4.1 Coordinates

It is difficult to describe what we mean by a set of coordinates without using more technical language. The effective idea is that a coordinate system should be a way of (uniquely) and continuously describing a point in your space. Cartesian coordinates are those with which we are most familiar, and are given by (x, y), describing the horizontal and vertical displacement of a point from the origin. However, the origin itself corresponds to an arbitrary choice: choose some other point in the plane and call that the origin, and notice that fundamentally, our space has not changed. For example, a circle $x^2+y^2=1$ is in many ways the same as the circle $(x-a)^2+(y-b)^2=1$ for any choice of (a, b), we have simply "moved it." Such a transformation is called a *translation* and are described as functions $f(x, y) = (x-a, y-b)$.

Similarly, one might choose to change how we want to measure distances, resulting in a *scaling* of the from $f(x, y) = (\alpha x, \beta y)$ for $\alpha, \beta \neq 0$ (when $\alpha < 0$ this corresponds to reflecting about the y-axis, and similarly $\beta < 0$ is reflection about the x-axis). We could even rotate our coordinate

system by an angle θ via the map $f(x,y) = (\cos(\theta)x + \sin(\theta)y, \cos(\theta)y - \sin(\theta)x)$. Combining scaling, rotations, and translations, one gets *affine transformations* $f(x,y) = (c_1x + c_2y + c_3, d_1x + d_2y + d_3)$.

But of course we have seen many other types of coordinate systems. For example, *polar coordinates* are described by the function $(x,y) = f(r,\theta) = (r\cos(\theta), r\sin(\theta))$. In $\mathbb{R}^3$ we have cylindrical and spherical coordinates:

$$(x,y,z) = f(r,\theta,z) = (r\cos(\theta), r\sin(\theta), z)$$
$$(x,y,z) = g(\rho,\theta,\phi) = (\rho\sin(\theta)\cos(\phi), \rho\sin(\theta)\sin(\phi), \rho\cos(\theta))$$

Though one problem faced with these set of coordinates is that without restrictions on r, ρ, θ, ϕ, the coordinate system may not be unique! For example, the following all represent the same set of points

$$(\cos(\pi), \sin(\pi)) = (\cos(-\pi), \sin(-\pi)) = (-\cos(0), -\sin(0)).$$

For polar coordinates we thus demand that $r \in (0,\infty)$ and $\theta \in [0, 2\pi)$. For spherical coordinates, one takes $\rho \in (0,\infty), \theta \in [0,\pi]$, and $\phi \in [0, 2\pi)$. Unfortunately, this means that we must make a sacrfice in the collection of points we are able to represent, for example, the origin $(0,0)$ cannot be written in polar coordinates. Hence our function is a map $f : (0,\infty) \times [0,2\pi) \to \mathbb{R}^2 \setminus \{(0,0)\}$.

There are countless other types of coordinate systems one might want to use, for example $(x,y) = f(\xi,\zeta) = (e^\xi, \zeta^2)$, though again we run into uniqueness issues and need to restrict our sets in order to have a "good" coordinate system. For example, in this case our good coordinate system is between the sets $f : \mathbb{R} \times [0,\infty) \to (0,\infty) \times [0,\infty)$.

So what restriction should we place on f to ensure that we have a good coordinate system between sets $U, V \subseteq \mathbb{R}^n$? Just for things to be unique we should certainly require that $f : U \to V$ is bijective (so that $f^{-1} : V \to U$ exists) and for things to play well with calculus, we should also require that f and f^{-1} be differentiable.

Definition 4.36

If $U, V \subseteq \mathbb{R}^n$ and $\mathbf{f} : U \to V$ is a C^1 bijection with C^1 inverse $\mathbf{f}^{-1} : V \to U$, then we say that $\mathbf{f}$ is a *diffeomorphism*.

Once we have a diffeomorphism $\mathbf{f} : U \to V$ we know that our spaces U, V are, in a sense, identical with respect to differentiation. Importantly however, the notion of lengths/volume may have changed. As our end goal will be to apply diffeomorphisms to integrals, we want to examine infinitesimal changes.

Consider for example the diffeomorphism $\mathbf{f} : (0,\infty) \times [0,2\pi) \to \mathbb{R}^2 \setminus \{(0,0)\}$ given by polar coordinates

$$(x,y) = \mathbf{f}(r,\theta) = (r\cos(\theta), r\sin(\theta)).$$

It is easy to see that $\mathbf{f}$ is C^1, and moreover it has an inverse

$$(r,\theta) = \mathbf{f}^{-1}(x,y) = \left(\sqrt{x^2+y^2}, \arctan\left(\frac{y}{x}\right)\right).$$

It may not be immediately obvious that $\mathbf{f}^{-1}$ is differentiable, but the absence of the origin $(0,0)$ ensures that this is the case.

Let's see how areas change under this transformation. Consider a rectangle $[a, b] \times [\alpha, \beta]$ in (r, θ)-space; that is, $a \le r \le b$ and $\alpha \le \theta \le \beta$. Applying $\mathbf{f}$, we get an arc-segment, as illustrated in Figure 37.

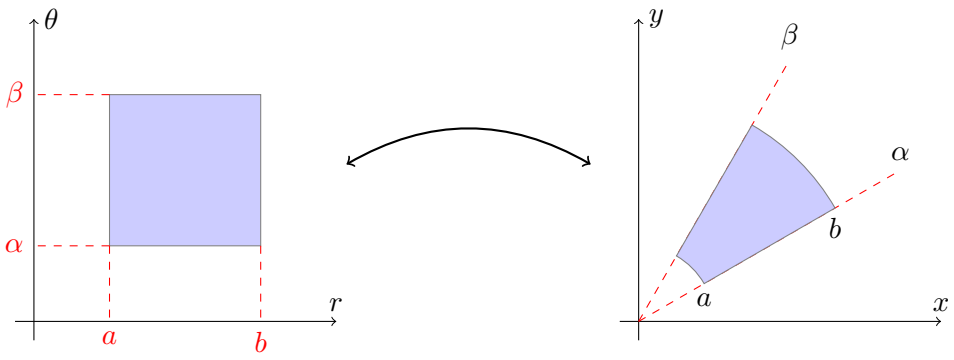


Figure 37: How a simple square in polar coordinates (left) changes under the map $\mathbf{f}(r, \theta) = (r\cos(\theta), r\sin(\theta))$.

If we think of these as describing infinitesimal elements, then $D\mathbf{f} : \mathbb{R}^n \to \mathbb{R}^n$ is a linear transformation between the $\{\,\mathrm{d}r,\, \mathrm{d}\theta\}$ basis and the $\{\,\mathrm{d}x,\, \mathrm{d}y\}$ basis,

$$D\mathbf{f}_{r,\theta} = \begin{pmatrix} \cos(\theta) & -r\sin(\theta) \\ \sin(\theta) & r\cos(\theta) \end{pmatrix},$$

and we can apply the following theorem:

Theorem 4.37

If $T : V \to W$ is a linear transformation between vector spaces of the same dimension, and $S \subseteq V$ is measurable with measure $m(S)$, then

$$m(TS) = |\det T| m(S).$$

One can easily check that $|\det D\mathbf{f}| = r$, and so $\mathrm{d}x\,\mathrm{d}y = r\,\mathrm{d}r\,\mathrm{d}\theta$.

Exterior Algebra: There is another technique for deriving this relationship, although the rigours of the theory would take us very far afield. Instead, one can summarize the basic rules of how to manipulate infinitesimal terms. Let $\mathrm{d}x$, $\mathrm{d}y$, and $\mathrm{d}z$ represent three infinitesimal terms (though this generalizes to a higher number of terms, and not just Cartesian coordinates).

1. The order of multiplication matters: $\mathrm{d}x\,\mathrm{d}y \neq \mathrm{d}y\,\mathrm{d}x$, so pay careful attention to the ordering,
2. Otherwise, multiplication can be done as normal:

$$(f\,\mathrm{d}x + g\,\mathrm{d}y)(h\,\mathrm{d}z) = fh\;\mathrm{d}x\;\mathrm{d}z + gh\;\mathrm{d}y\;\mathrm{d}z,$$

3. To interchange two infinitesimals which are adjacent, introduce a $-$ sign:

$$\mathrm{d}x\;\mathrm{d}y\;\mathrm{d}z = -\,\mathrm{d}y\;\mathrm{d}x\;\mathrm{d}z = -\,\mathrm{d}x\;\mathrm{d}z\;\mathrm{d}y,$$

4. If two of the same infinitesimal appear, the term becomes zero:

$$\mathrm{d}x\ \mathrm{d}x = 0, \quad \mathrm{d}x\ \mathrm{d}y\ \mathrm{d}y = 0, \quad \mathrm{d}x\ \mathrm{d}z\ \mathrm{d}x = 0.$$

For example, if $(x, y) = (r\cos(\theta), r\sin(\theta))$, then

$$\begin{aligned}
\mathrm{d}x &= d(r\cos(\theta)) = \cos(\theta)\,\mathrm{d}r - r\sin(\theta)\,\mathrm{d}\theta \\
\mathrm{d}y &= d(r\sin(\theta)) = \sin(\theta)\,\mathrm{d}r + r\cos(\theta)d\theta \\
\mathrm{d}x\ \mathrm{d}y &= (\cos(\theta)\,\mathrm{d}r - r\sin(\theta)\,\mathrm{d}\theta)(\sin(\theta)\,\mathrm{d}r + r\cos(\theta)\,\mathrm{d}\theta) \\
&= \underbrace{\cos(\theta)\sin(\theta)\,\mathrm{d}r\ \mathrm{d}r}_{=0} - r\sin^2(\theta)\,\mathrm{d}\theta\ \mathrm{d}r + r\cos^2(\theta)\,\mathrm{d}r\ \mathrm{d}\theta - \underbrace{r^2\sin(\theta)\cos(\theta)\,\mathrm{d}\theta\ \mathrm{d}\theta}_{=0} \\
&= r(\sin^2(\theta) + \cos^2(\theta))\,\mathrm{d}r\ \mathrm{d}\theta \\
&= r\ \mathrm{d}r\ \mathrm{d}\theta.
\end{aligned}$$

In fact, notice that rules 3 and 4 are very similar to determinants: Interchanging two columns corresponds to introducing a minus sign, and if two rows are linearly dependent, the determinant is zero. This is not a coincidence, as it turns out the exterior algebra for infinitesimals is intimately related to determinants.

4.4.2 Integration

The content of this section is extraordinarily useful but the proofs are cumbersome and not particularly enlightening. Consequently, we will motivate the situation by analyzing what happens in the one-dimensional case, before stating the major theorem (without proof).

In the one-dimensional case, there is not much in the way of variable changing that can be done! Nonetheless, the student has already seen a plethora of examples which greatly emulate coordinate changing: The method of substitution. For example, when integrating the equation

$$\int_2^3 \frac{x}{x^2 - 1}\,\mathrm{d}x,$$

the student should (hopefully) realize that the appropriate substitution here is $u = x^2 - 1$ so that $\mathrm{d}u = 2x\,\mathrm{d}x$, and the integral becomes

$$\int_2^3 \frac{x}{1 - x^2}\,\mathrm{d}x = \frac{1}{2}\int_3^8 \frac{1}{u}\,\mathrm{d}u = [\ln|u|]_3^8 = \ln(8) - \ln(3).$$

In effect, the theory behind why this works is that we have realized that working in the x-coordinate system is rather silly since it makes our integral look complicated. By changing to the $u = 1 + x^2$ coordinate system, the integral reduces to something which we can easily solve.

The theory is as follows (though our presentation might seem a bit backwards compared to how such integrals are usually computed): The fundamental theorem of calculus easily tells us that

$$\int_a^b f(g(x))g'(x)\,\mathrm{d}x = \int_{g(a)}^{g(b)} f(u)\,\mathrm{d}u \tag{4.2}$$

where $u = g(x)$ so that $\mathrm{d}u = g'(x)\,\mathrm{d}x$. The idea is that by introducing the auxiliary function $u = g(x)$ we were able to greatly reduce the problem to something more elementary, and that is the goal of changing variables.

Unfortunately, there is never a single way to change variables, and it can make our notation a bit of a headache. For example, what if we had instead chosen the substitution $u = 1 - x^2$ in the previous example, so that the integral became

$$\int_2^3 \frac{x}{x^2-1}\,\mathrm{d}x = \frac{1}{2}\int_{-3}^{-8} \frac{1}{u}\,\mathrm{d}u.$$

Notice that the bounds of integration are in the wrong order, since certainly $-3 > -8$. We of course fix this by introducing a negative sign and interchanging the bounds and arrive at the same answer, but the point is that we do not want to have to worry about whether we have changed the *orientation*[6] of the interval (since this will become a grand nightmare in multiple dimensions!). Hence if $I = [a, b]$, we will write (4.2) as

$$\int_I f'(g(x))|g'(x)|\,\mathrm{d}x = \int_{g(I)} f(u)\,\mathrm{d}u.$$

What is bothersome about this equation is that g appears on both sides of the equation. If g is a change of coordinates (so that it is a diffeomorphism on I), then there is no harm in replacing g with g^{-1}. Let $J = g(I)$ so that we get

$$\int_{g^{-1}(J)} f'(g(x))|g'(x)|\,\mathrm{d}x = \int_J f(u)\,\mathrm{d}u.$$

So what do we do in higher dimension?

Theorem 4.38: Change of Variables

If $S, T \subseteq \mathbb{R}^n$ are measurable and $\mathbf{G} : S \to T$ is a diffeomorphism, then for any integrable function $f : T \to \mathbb{R}$ we have

$$\int_T f(\mathbf{u})\,\mathrm{d}\mathbf{u} = \int_{\mathbf{G}^{-1}(T)} f(\mathbf{G}(x))|\det D\mathbf{G}(\mathbf{x})|\,\mathrm{d}\mathbf{x}.$$

The term $|\det D\mathbf{G}|$ is known as the *Jacobian* of the change of variables.

Again, this proof is laborious and of no great value, so we omit it here. Note that this effectively say that the element $|\det D\mathbf{G}(\mathbf{x})|$ represent the scaling of the volume element we had discussed before. Indeed, previously we saw that the change in area resulting from using polar coordinates was to multiply by r. If $(x, y) = G(r, \theta) = (r\cos(\theta), r\sin(\theta))$ then

$$\begin{aligned} |\det DG(r,\theta)| &= \left|\det \begin{pmatrix} \cos(\theta) & -r\sin(\theta) \\ \sin(\theta) & r\cos(\theta) \end{pmatrix}\right| \\ &= \left|r\cos^2(\theta) + r\sin^2(\theta)\right| \\ &= r \end{aligned}$$

[6] This is a remarkably subtle but important point that does not manifest in 1-dimension but proves to be truly inconvenient in higher dimensions. There is an entire theory of *orientability* of surfaces and higher dimensional spaces, and if your space is not orientable then it is difficult to do integration.

which exactly agrees with our previous assessment that $\mathrm{d}x\ \mathrm{d}y = r\ \mathrm{d}r\ \mathrm{d}\theta$. The following are the two most often used coordinate changes in three dimensions:

Example 4.39

1. **Cylindrical Coordinates:** Recall that cylindrical coordinates are related to Cartesian coordinates by $(x, y, z) = g(r, \theta, z) = (r\cos(\theta), r\sin(\theta), z)$. Hence

$$\begin{aligned} |\det Dg(r,\theta,z)| &= \left|\det \begin{pmatrix} \cos(\theta) & -r\sin(\theta) & 0 \\ \sin(\theta) & r\cos(\theta) & 0 \\ 0 & 0 & 1 \end{pmatrix}\right| \\ &= r. \end{aligned}$$

   This is not terribly surprising: cylindrical coordinates are polar coordinates with the z-direction unaffected. Hence we only expect the scaling to occur in the xy-dimensions, and this is indeed what we see.

2. **Spherical Coordinates:** Cartesian and Spherical coordinates are related by $(x, y, z) = g(\rho, \theta, \phi) = (\rho\sin\theta\cos\phi, \rho\sin\theta\sin\phi, \rho\cos\theta)$, and

$$\begin{aligned} |\det Dg(\rho,\theta,\phi)| &= \left|\det \begin{pmatrix} \sin\theta\cos\phi & \rho\cos\theta\cos\phi & -\rho\sin\theta\sin\phi \\ \sin\theta\sin\phi & \rho\cos\theta\sin\phi & \rho\sin\theta\cos\phi \\ \cos\theta & -\rho\sin\theta & 0 \end{pmatrix}\right| \\ &= \cos\theta\left[\rho^2\cos\theta\sin\theta\cos^2\phi + \rho^2\cos\theta\sin\theta\sin^2\phi\right] \\ &\quad + \rho\sin\theta\left[\rho\sin^2\theta\cos^2\phi + \rho\sin^2\theta\sin^2\phi\right] \\ &= \rho^2\cos^2\theta\sin\theta + \rho^2\sin^2\theta\sin\theta \\ &= \rho^2\sin\theta \end{aligned}$$

Example 4.40

Let $(u, v) = (e^r\cos(\theta), e^r\sin(\theta))$. Determine $\mathrm{d}u\,\mathrm{d}v$ as a function of $\mathrm{d}r\,\mathrm{d}\theta$ and vice versa.

Solution. Computing the Jacobian of the transformation one gets

$$\det \begin{pmatrix} e^r\cos(\theta) & e^r\sin(\theta) \\ -e^r\sin(\theta) & e^r\cos(\theta) \end{pmatrix} = e^{2r},$$

and so $\mathrm{d}u\ \mathrm{d}v = e^{2r}\ \mathrm{d}r\ \mathrm{d}\theta$.

To compute $\mathrm{d}r\ \mathrm{d}\theta$ in terms of $\mathrm{d}u$, $\mathrm{d}v$ one could try to find the inverse of the coordinate transformation, but that would prove exceptionally difficult. Instead, recognize that $u^2 + v^2 = e^{2r}$ and hence

$$\mathrm{d}r\ \mathrm{d}\theta = \frac{1}{e^{2r}}\,\mathrm{d}u\ \mathrm{d}v = \frac{\mathrm{d}u\ \mathrm{d}v}{u^2+v^2}.$$

■

We can now exploit change of variables to make integration much easier:

Example 4.41

Let $R = \{(x, y) \in \mathbb{R}^2 : 1 \leq x^2 + y^2 \leq 3\}$. Evaluate $\iint_R e^{x^2+y^2}\, \mathrm{d}A$.

Solution. Our region R is simply the area between the circles of radius 1 and $\sqrt{3}$, so we use polar coordinates. Let $(x, y) = (r\cos(\theta), r\sin(\theta))$ so that $S = [1, \sqrt{3}) \times [0, 2\pi)$ is just a rectangle in (r, θ) space, and $\mathbf{G} : S \to T$ is a diffeomorphism. Integrating using change of variables gives

$$\begin{aligned}\iint_R e^{x^2+y^2}\, \mathrm{d}A &= \int_1^{\sqrt{3}} \int_0^{2\pi} e^{r^2} r\, \mathrm{d}r\, \mathrm{d}\theta = 2\pi \int_1^{\sqrt{3}} r e^{r^2}\, \mathrm{d}r\ \mathrm{d}\theta \\ &= \pi \left[e^{r^2}\right]_{r=1}^{\sqrt{3}} = \pi \left[e^3 - e\right]\end{aligned}$$

■

Example 4.42

Let S be the region bounded by the curves $xy = 1$, $xy = 3$, $x^2 - y^2 = 1$ and $x^2 - y^2 = 4$. Compute $\iint_T (x^2 + y^2)\, \mathrm{d}A$.

Solution. The region suggests that we should take a change of variables of the form $u = xy$ and $v = x^2 - y^2$, so that setting

$$T = \{1 \leq u \leq 3,\ 1 \leq v \leq 4\}$$

implies that $G : S \to T$ given by $(u, v) = G(x, y) = (xy, x^2 - y^2)$ is a diffeomorphism. Now

$$\begin{aligned}|\det DG(x, y)| &= \left|\det \begin{pmatrix} y & x \\ 2x & -2y \end{pmatrix}\right| \\ &= 2(x^2 + y^2).\end{aligned}$$

Thus $\mathrm{d}u\, \mathrm{d}v = 2(x^2 + y^2)\, \mathrm{d}x\, \mathrm{d}y$ and our integral becomes

$$\iint_S (x^2 + y^2)\, \mathrm{d}x\, \mathrm{d}y = \frac{1}{2} \int_T \mathrm{d}u\ \mathrm{d}v = 3.$$

■

Example 4.43

Find the area bounded between the sphere $x^2 + y^2 + z^2 = 4$ and the cylinder $x^2 + y^2 = 1$.

Solution. Let B be the region bounded. Let's use cylindrical coordinates, so that $\mathrm{d}x\, \mathrm{d}y\, \mathrm{d}z = r\, \mathrm{d}r\, \mathrm{d}\theta\, \mathrm{d}z$. Now by drawing a picture, it is clear that we are symmetric about reflection in the xy-plane, so we need only find the volume bounded by the upper-half hemisphere and the cylinder, B_+. The total area will be governed by $r \in (0, 1)$ and $\theta \in (0, 2\pi)$, but our z coordinate will by

represented by $z = \sqrt{4 - x^2 - y^2} = \sqrt{4 - r^2}$. Our integral is thus

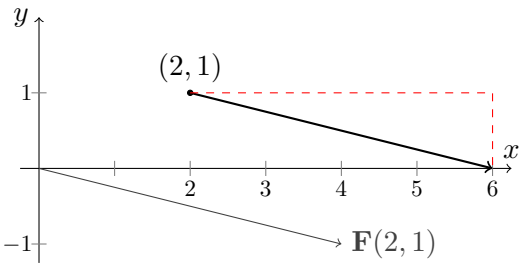
$$\begin{aligned}\int_{B_+} \mathrm{d}x\ \mathrm{d}y\ \mathrm{d}z &= \int_0^{2\pi} \int_0^1 \int_0^{\sqrt{4-r^2}} r\ \mathrm{d}z\ \mathrm{d}r\ \mathrm{d}\theta \\ &= \int_0^{2\pi} \int_0^1 r\sqrt{4 - r^2}\,\mathrm{d}r\ \mathrm{d}\theta \\ &= \int_0^{2\pi} \left[-\frac{1}{3}(4 - r^2)^{3/2}\right]_{r=0}^1 \mathrm{d}\theta \\ &= \frac{2\pi}{3}\left[8 - 3\sqrt{3}\right]\end{aligned}$$

Hence the fully bounded area is $2B_+ = \frac{4\pi}{3}\left[8 - 3\sqrt{3}\right]$. ■

5 Vector Fields

Section 4 was principally concerned with integrating functions $f : \mathbb{R}^n \to \mathbb{R}$, whose geometric interpretation was to find the area under the graph of f on some domain. In contrast to this, we now turn our focus to the more general case of functions $\mathbf{F} : \mathbb{R}^n \to \mathbb{R}^n$. However, the geometric interpretation of what an integral is will change dramatically. It no longer makes sense to ask about things like upper and lower Riemann sums since $\mathbf{F}(\mathbf{x}) \in \mathbb{R}^n$, there is measure of what is "largest" or "smallest," so we are no longer thinking about areas under graphs.

Instead, the function $\mathbf{F}$ describes what is called a *vector field*. A vector field is function which prescribes to every point $\mathbf{x} \in \mathbb{R}^n$ an arrow, $\mathbf{F}(\mathbf{x})$. For example, consider the vector field $\mathbf{F}(x, y) = (x^2, -y)$. To determine what arrow to place at $\mathbf{x} = (1, 2)$ we compute $\mathbf{F}(2, 1) = (4, -1)$.



We can visualize vector fields by choosing multiple points and drawing the vectors which correspond to them, as in Figure 38

Vector fields can be used to describe physical fields and forces. For example, the force exhibited at a single point by an electromagnetic field or gravity may be conveyed as a vector field. On the other hand, a vector field might describe the flow of a liquid, such as water in a stream or air over wing. Our goal in this section is to see how we can use vector fields to compute useful quantities, which often have physical interpretations such as flux or work.

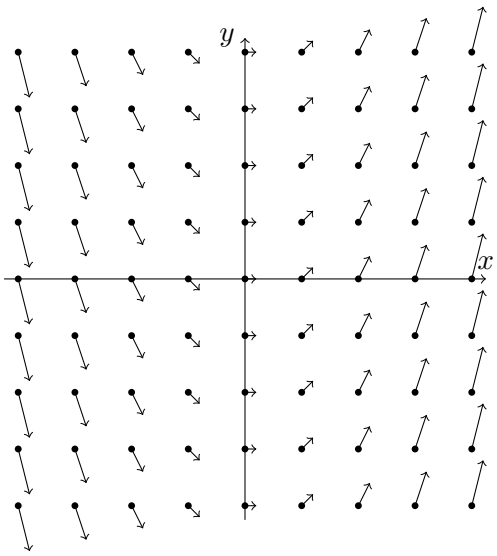
5.1 Vector Derivatives

Depending on whether a given function is vector-valued or not, there are multiple different kinds of derivatives that we can take. In this section, we are going to look at four such operators: the gradient (which you have already seen), divergence, curl, and the Laplacian. In turns out that the first three of these are all actually the same operator in disguise, but that is a rather advanced topic which we (probably) won't cover in this class. The trick in all of these cases is to think of the nabla operator ∇ as a vector whose components are the partial derivative operators. Hence in $\mathbb{R}^n$, the nabla operator is

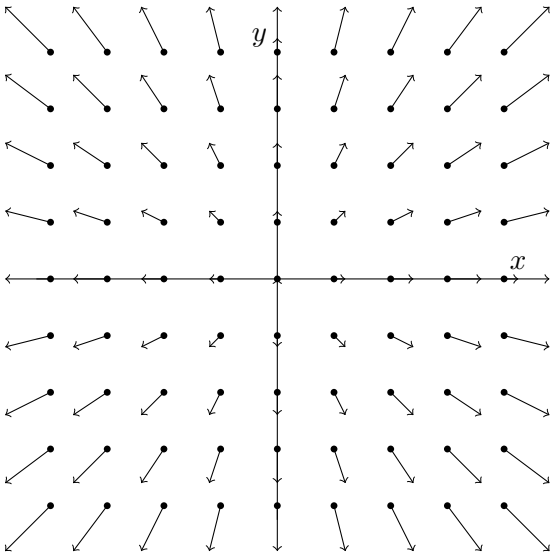
$$\nabla = \left(\frac{\partial}{\partial x_1}, \frac{\partial}{\partial x_2}, \ldots, \frac{\partial}{\partial x_n} \right).$$

1. **Gradient:** Let $f : \mathbb{R}^n \to \mathbb{R}$ be a C^1 function. The *gradient* of f is

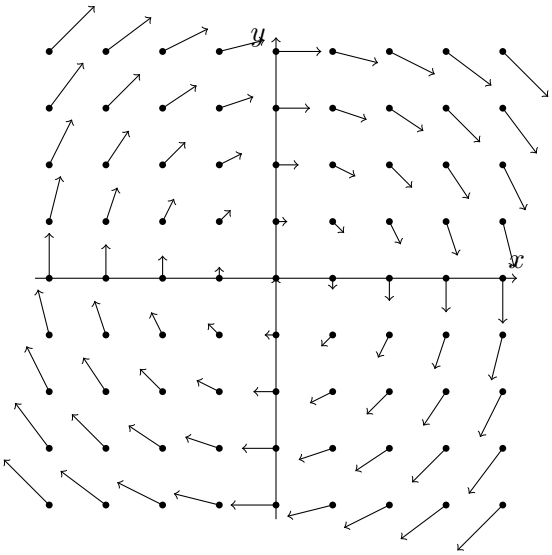
$$\operatorname{grad} f = \nabla f = \left(\frac{\partial f}{\partial x_1}, \ldots, \frac{\partial f}{\partial x_n} \right).$$



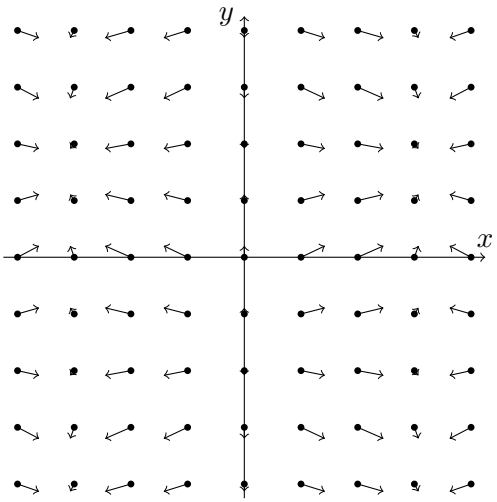
$\mathbf{F}(x, y) = (1, x)$



$\mathbf{F}(x, y) = (x, y)$



$\mathbf{F}(x, y) = (y, -x)$



$\mathbf{F}(x, y) = (\sin(x), \cos(y))$

Figure 38: A visualization of several vector fields. Using many points gives us an intuitive idea for how these vectors "flow."

The gradient measures how quickly the function f is changing in each of the given coordinate axes, and ∇f in its totality gives the direction of steepest ascent. As an example computation, if $f(x,y,z) = z\sin(xy)$ then

$$\begin{aligned}\nabla f(x,y,z) &= \left(\frac{\partial}{\partial x}\left[z\sin(xy)\right], \frac{\partial}{\partial y}\left[z\sin(xy)\right], \frac{\partial}{\partial z}\left[z\sin(xy)\right]\right) \\ &= \left(zy\cos(xy), zx\cos(xy), \sin(xy)\right).\end{aligned}$$

2. **Divergence:** If $\mathbf{F} : \mathbb{R}^n \to \mathbb{R}^n$ is a C^1-vector field, then the *divergence* of $\mathbf{F}$ is

$$\operatorname{div}\mathbf{F} = \nabla \cdot F = \frac{\partial F_1}{\partial x_1} + \cdots + \frac{\partial F_n}{\partial x_n}.$$

The divergence is a measure of the *infinitesimal flux* of the vector field; that is, the amount of the field which is passing through an infinitesimal surface area. As an example, if $\mathbf{F}(x,y,z) = (x^2, y^2, z^2)$ then

$$\begin{aligned}\operatorname{div}\mathbf{F}(x,y,z) &= \left[\frac{\partial}{\partial x}x^2\right] + \left[\frac{\partial}{\partial y}y^2\right] + \left[\frac{\partial}{\partial z}z^2\right] \\ &= 2x + 2y + 2z.\end{aligned}$$

3. **Curl:** If $\mathbf{F} : \mathbb{R}^3 \to \mathbb{R}^3$ is a C^1 vector field in $\mathbb{R}^3$, then the *curl* of $\mathbf{F}$ is

$$\operatorname{curl}\mathbf{F} = \nabla \times F = \left(\frac{\partial F_3}{\partial x_2} - \frac{\partial F_2}{\partial x_3}, \frac{\partial F_1}{\partial x_3} - \frac{\partial F_3}{\partial x_1}, \frac{\partial F_2}{\partial x_1} - \frac{\partial F_1}{\partial x_2}\right).$$

The curl measures the *infinitesimal circulation* of the vector field. Furthermore, notice that the curl only makes sense in $\mathbb{R}^3$. There are higher dimensional analogs of the curl, but they are very messy to write down. For example, if $\mathbf{F}(x,y,z) = (x^2y, xyz, -x^2y^2)$ then

$$\begin{aligned}\operatorname{curl}\mathbf{F}(x,y,z) &= \left(-2x^2y - xy, 0 - (-2xy^2), yz - x^2\right) \\ &= \left(-xy(2x+1), 2xy^2, yz - x^2\right).\end{aligned}$$

4. **Laplacian:** If $f : \mathbb{R}^n \to \mathbb{R}$ is a C^1 function, then the *Laplacian* of f is

$$\nabla^2 f = \Delta f = \frac{\partial^2 f}{\partial x_1^2} + \cdots + \frac{\partial^2 f}{\partial x_n^2}.$$

Notice that one can also write $\nabla^2 = \nabla \cdot \nabla$ so that the Laplacian is the divergence of the gradient. In essence, the Laplacian measures the infinitesimal rate of change of the function f in outward rays along spheres. If $f(x,y,z) = x^2y + z^3$, then an example of computing the Laplacian is given by

$$\begin{aligned}\nabla^2 f(x,y,z) &= \left[\frac{\partial^2}{\partial x^2}\left(x^2y + z^3\right)\right] + \left[\frac{\partial^2}{\partial y^2}\left(x^2y + z^3\right)\right] + \left[\frac{\partial^2}{\partial z^2}\left(x^2y + z^3\right)\right] \\ &= 2y + 6z.\end{aligned}$$

All of these vector derivatives are exceptionally important in physics and mathematics. Of particular note is the Laplacian, which is central to the study of partial differential equations and harmonic analysis.

We have already seen the gradient: it physically represents the direction of steepest ascent. The names associated to divergence and curl are also done with a purpose. We do not yet have the tools, but one can show that the curl of a vector field in $\mathbb{R}^3$ corresponds to infinitesimal circulation of the vector field (how quickly the field is spinning around), while the divergence is the infinitesimal flux of the vector field (how quickly the field is spreading out). For this reason, if $\mathbf{F}$ is a vector field such that $\operatorname{curl}\mathbf{F} = 0$, we say that $\mathbf{F}$ is *irrotational*. Similarly, if $\operatorname{div} F = 0$ we say that $\mathbf{F}$ is *incompressible*.

Proposition 5.1

Let $f, g : \mathbb{R}^n \to \mathbb{R}$ and $\mathbf{F}, \mathbf{G} : \mathbb{R}^n \to \mathbb{R}^n$ all be C^1 (taking $n = 3$ when appropriate). Then the gradient, divergence, and curl all satisfy the following properties:

$$\begin{aligned}
\operatorname{grad}(fg) &= f\operatorname{grad} g + g\operatorname{grad} f \\
\operatorname{grad}(\mathbf{F}\cdot\mathbf{G}) &= (\mathbf{F}\cdot\nabla)\mathbf{G} + \mathbf{F}\times(\operatorname{curl}\mathbf{G}) + (\mathbf{G}\cdot\nabla)\mathbf{F} + \mathbf{G}\times(\operatorname{curl}\mathbf{F}) \\
\operatorname{curl}(f\mathbf{G}) &= f\operatorname{curl}\mathbf{G} + (\operatorname{grad} f)\times\mathbf{G} \\
\operatorname{curl}(\mathbf{F}\times\mathbf{G}) &= (\mathbf{G}\cdot\nabla)\mathbf{F} + (\operatorname{div}\mathbf{G})\mathbf{F} - (\mathbf{F}\cdot\nabla)\mathbf{G} - (\operatorname{div}\mathbf{F})\mathbf{G} \\
\operatorname{div}(f\mathbf{G}) &= f\operatorname{div}\mathbf{G} + (\operatorname{grad} f)\mathbf{G} \\
\operatorname{div}(\mathbf{F}\times\mathbf{G}) &= \mathbf{G}\cdot(\operatorname{curl} F) - \mathbf{F}\cdot(\operatorname{curl}\mathbf{G})
\end{aligned}$$

Proof. The majority of these are straightforward if laborious, so we will only do one as an example. Let's show that

$$\operatorname{curl}(f\mathbf{G}) = f\operatorname{curl}\mathbf{G} + (\operatorname{grad} f)\times\mathbf{G}.$$

Let $\mathbf{G} = (G_1, G_2, G_2)$ so that $f\mathbf{G} = (fG_1, fG_2, fG_3)$. The x-component of $\operatorname{curl}(f\mathbf{G})$ is

$$\begin{aligned}
\operatorname{curl}(f\mathbf{G})_1 &= \frac{\partial}{\partial y}(fG_3) - \frac{\partial}{\partial z}(fG_2) \\
&= \frac{\partial f}{\partial y}G_3 + f\frac{\partial G_3}{\partial y} - \frac{\partial f}{\partial z}G_2 - f\frac{\partial G_2}{\partial z} \\
&= f\left(\frac{\partial G_3}{\partial y} - \frac{\partial G_2}{\partial z}\right) + \left(\frac{\partial f}{\partial y}G_3 - \frac{\partial f}{\partial z}G_2\right) \\
&= f(\operatorname{curl} G)_1 + [\operatorname{grad} f\times\mathbf{G})]_1 \\
&= [f(\operatorname{curl} G)_1 + \operatorname{grad} f\times\mathbf{G})]_1 .
\end{aligned}$$

Hence the x-coordinates of both vectors agree. Since all other components follow precisely the same reasoning (just replace y and z with z and x respectively) the result follows. □

The next two identities worth pointing out are the following:

$$\operatorname{curl}(\operatorname{grad} f) = 0, \qquad \operatorname{div}(\operatorname{curl}\mathbf{F}) = 0.$$

In fact, in higher level mathematics this is effectively contained in definition of divergence and curl. A very nice diagram is the following:

$$\begin{matrix}\text{scalar}\\ \text{function}\end{matrix} \xrightarrow{\text{grad}} \begin{matrix}\text{vector}\\ \text{fields}\end{matrix} \xrightarrow{\text{curl}} \begin{matrix}\text{vector}\\ \text{fields}\end{matrix} \xrightarrow{\text{div}} \begin{matrix}\text{scalar}\\ \text{functions}\end{matrix},$$

which is (up to renaming some things) called the *de Rham complex*.

5.2 Arc Length

While arc length is a subject that can be discussed in introductory calculus, its generalization to curves in $\mathbb{R}^n$ will prove important for this section so we re-iterate its treatment here. Given a curve, one can naively attempt calculate its arc length by approximating it with successively finer and finer piecewise linear curves. The non-calculus way of doing this requires suprema and partitions and involves introducing a notion of *rectifiability* (read as: ability to break into piecewise functions). This offers the advantage that it allows us to compute the arc length of many curves which cannot be described as C^1 functions, but will not be useful for our purposes.

Instead, let's use our formidable calculus experience to formulate an expression for arc length. Recall that given two points $x = (x_1, \dots, x_n)$ and $y = (y_1, \dots, y_n)$ in $\mathbb{R}^n$ their distance is described as

$$|x - y| = \sqrt{(x_1 - y_1)^2 + \cdots + (x_n - y_n)^2} = \left[\sum_{i=1}^{n} (x_i - y_i)^2\right]^{\frac{1}{2}}.$$

Assume we were to approximate our curve C by infinitesimal straight line components, and fix a point x so that the straight line extends to $x + \mathrm{d}x$. The distance between these two points is then

$$|x - (x + \mathrm{d}x)| = |\,\mathrm{d}x| = \sqrt{\mathrm{d}x_1^2 + \cdots + \mathrm{d}x_n^2}. \tag{5.1}$$

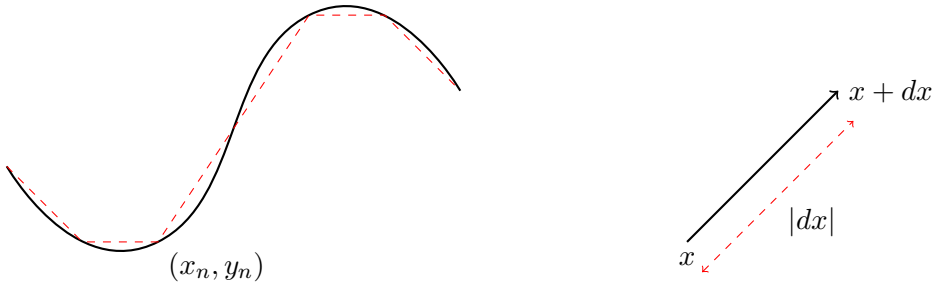


Figure 39: Left: We can approximate a C^1 curve by a piecewise linear curve. Right: In the infinitesimal limit, the length of each piecewise linear segment is $|\,\mathrm{d}x$—.

We often write $ds = |\,\mathrm{d}x|$, which we call an *element of arc*. The total arc length will then be given by integrating ds over the length of the curve. As it stands, this is not a particularly attractive element to work with, so to facilitate our computations we introduce a parameterization of our curve. Assume that C is given by the equation $x = g(t)$, where $g : [a, b] \to \mathbb{R}^n$, so that

$$\mathrm{d}x = g'(t)dt = \left(\frac{\mathrm{d}g_1}{\mathrm{d}t}, \dots, \frac{\mathrm{d}g_n}{\mathrm{d}t}\right) dt$$

which in turn means that we can rewrite (5.1) as

$$ds = |\,\mathrm{d}x| = \sqrt{\left(\frac{\mathrm{d}g_1}{\mathrm{d}t}\right)^2 + \cdots + \left(\frac{\mathrm{d}g_n}{\mathrm{d}t}\right)^2}\,dt.$$

By integrating from a to b we then have

$$\text{Arclength}(C) = \int_C ds = \int_a^b |g'(t)|dt = \int_a^b \sqrt{\left(\frac{\mathrm{d}g_1}{\mathrm{d}t}\right)^2 + \cdots + \left(\frac{\mathrm{d}g_n}{\mathrm{d}t}\right)^2}\,dt.$$

This has a particularly appealing physical interpretation: If $g(t)$ describes the position of a particle with respect to time, then $g'(t)$ is its velocity and $|g'(t)|$ is its speed. By integrating the speed over all time, we then get the distance travelled which is precisely the arc length.

Example 5.2

Show that the circumference of a circle with radius r is precisely $2\pi r$.

Solution. This is a result with which we are all familiar, but that familiarity is because we were told that it is true, and not because we have ever derived the solution ourselves. Our curve in question is the circle of radius r, which we know admits a very simple parametric descriptions as

$$(x, y) = g(t) = (r\cos(t), r\sin(t)), \qquad 0 \le t \le 2\pi.$$

The velocity is then $g'(t) = (-r\sin(t), r\cos(t))$ and the speed is

$$|g'(t)| = \sqrt{r^2\sin^2(t) + r^2\cos^2(t)} = r$$

Our arc length formula then gives

$$\text{Arclength}(C) = \int_0^{2\pi} |g'(t)|dt = \int_0^{2\pi} r\,dt = 2\pi r$$

as required. ■

Notes:

1. Typically, the arc-element $ds = |g'(t)|dt$ is exceptionally difficult to integrate, since the square root of a sum is rarely amenable to standard tricks.
2. The arc-length formula computes *total* distance travelled, not necessarily the arc length of the graph of the curve. For example, in Example 5.2, letting $0 \le t \le 4\pi$ corresponds to traversing the circle twice. It is easy to see that our arc-length in this case is $4\pi r = 2 \times 2\pi r$. Thus even though the graph only shows a single circle, the parameterization travelled the circle twice.
3. Arc length should be independent of parameterization, as entailed by our physical intuition; that is, our distance travelled shouldn't depend on how quickly I moved. As an example, if you run a kilometre or crawl a kilometre, the distance you travelled is still just one kilometre! We prove this in more detail in the following proposition.

Proposition 5.3

Arc length is invariant under re-parameterization. More precisely, if $g : [a, b] \to \mathbb{R}^n$ is a C^1 function and $\phi : [a, b] \to [c, d]$ is a re-parameterization of $g(t)$ (so that $g \circ \phi : [c, d] \to \mathbb{R}^n$) then

$$\int_{[c,d]} \left| \frac{\mathrm{d}}{\mathrm{d}t}(g \circ \phi)(t) \right| dt = \int_{[a,b]} |g'(t)| dt.$$

Proof. The proof is immediate, since

$$\left| \frac{\mathrm{d}}{\mathrm{d}t} g(\phi(t)) \right| = |g'(\phi(t))||\phi'(t)|$$

and by Theorem 4.38 we have

$$\begin{aligned} \int_{[a,b]} |g'(t)| dt &= \int_{\phi^{-1}([a,b])} |g'(\phi(t))||\det D\phi(t)| dt \\ &= \int_{[c,d]} |g'(\phi(t))||\phi'(t)| dt \\ &= \int_{[c,d]} \left| \frac{\mathrm{d}}{\mathrm{d}t}(g \circ \phi)(t) \right| dt. \end{aligned}$$

□

5.3 Line Integrals

Scalar functions: The next type of integration we are going to look at is a generalization of our multivariable integration. Let $f : U \subseteq \mathbb{R}^n \to \mathbb{R}$ be a continuous function and $C \subseteq U$ a smooth curve. We already know how to integrate f over U to get the total volume between the graph of $f(\mathbf{x})$ and the $\mathbb{R}^n$ plane. What if instead we wanted to know the area which lies between the the curve C and the graph of $f(\mathbf{x})$? The answer is the *line integral of a scalar function*, defined as follows: Let $g : [a, b] \to \mathbb{R}^n$ be a parameterization of C, and take

$$\int_C f \, \mathrm{d}s = \int_a^b f(\mathbf{g}(t))|\mathbf{g}'(t)| \, \mathrm{d}t.$$

If we think about this formula, we will see that it gives the desired result. In particular, if $f \equiv 1$ then our formula just returns the formula for arc length. By including the $f(\mathbf{g}(t))$ term, we are *weighting* the value of the curve at each point by the value that the curve takes on $f(\mathbf{x})$. Integrating over all of $[a, b]$ then gives the area under $f(\mathbf{x})$ which lies on the curve C.

Example 5.4

Compute $\int_C \frac{1}{1+z/2}$ where C is the curve $g(t) = (-2\sin(t), 2\cos(t), 2t^2)$ from $[0, 1]$.

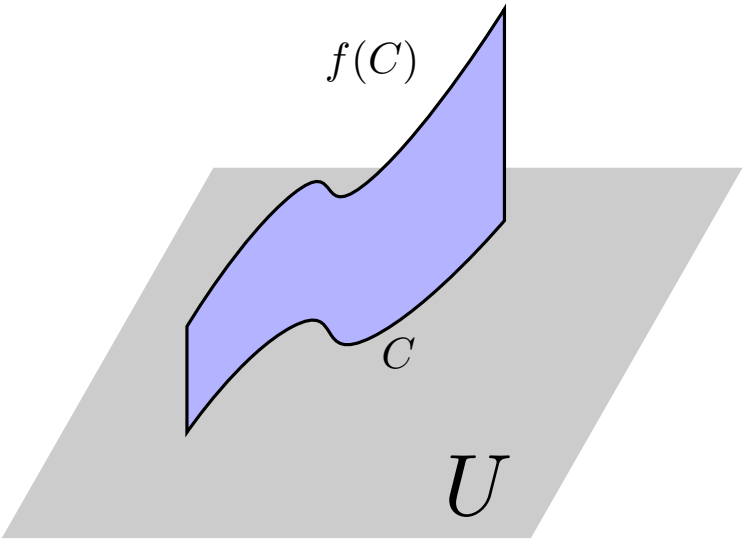


Figure 40: Integrating a scalar-valued function over U yields the area under the graph of f, but only along the curve C.

Solution. We begin by setting $f(x,y,z) = \frac{1}{1+z/2}$. One can easily compute $g'(t) = (-2\sin(t), 2\cos(t), 2t)$ so that $|g'(t)| = 2\sqrt{1+t^2}$. Furthermore,

$$f(g(t)) = f(-2\sin(t), 2\cos(t), 2t^2) = \frac{1}{1+t^2}$$

so our line integral becomes

$$\begin{aligned}
\int_C \frac{1}{1+z^2}\,ds &= \int_0^1 \underbrace{\frac{1}{1+t^2}}_{f(x,y,z)} \underbrace{2\sqrt{1+t^2}}_{|g'(t)|}\,dt \\
&= 2\int_0^1 \frac{1}{\sqrt{1+t^2}} \\
&= 2\int_0^{\pi/4} \frac{\sec^2(\theta)}{\sec(\theta)}\,\mathrm{d}\theta \qquad\qquad \text{let } t = \tan(\theta) \\
&= 2\ln\left|\sec(\theta)+\tan(\theta)\right|_0^{\pi/4} \\
&= 2\ln\left|t+\sqrt{1+t^2}\right|_0^1 \\
&= \ln\left|\sqrt{2}+1\right|.
\end{aligned}$$

Alternatively, one could realize that $\frac{\mathrm{d}}{\mathrm{d}t}\sinh^{-1}(t) = \frac{1}{\sqrt{1+t^2}}$ and that $\sinh^{-1}(t) = \ln(t+\sqrt{t^2+1})$. ■

Vector Fields: In my experience, line integrals of scalar functions are not particularly interesting and do not often manifest naturally (either in mathematics or otherwise). However, of far greater utility is computing line integrals through vector fields. The set up is as follows: Let $\mathbf{F} : \mathbb{R}^n \to \mathbb{R}^n$ be a vector field, and $C \subseteq \mathbb{R}^n$ be some smooth curve. Parameterize this curve by $\mathbf{g} : [a,b] \to \mathbb{R}^n$,

and think of the vector field acting on the curve at each point $t_0 \in [a, b]$. Our goal is to compute the "total amount of work" done by the vector field on the curve.

To put this into a more physical setting, imagine a fish that swims along a curve C, and let $\mathbf{F} : \mathbb{R}^3 \to \mathbb{R}^3$ be a vector field describing the current of the water at each point. As the fish swims through the water, the current acts on the fish by pushing it in the direction indicated by the vector field. We want to compute the amount of work done by vector field on the fish.

If we are travelling in the direction $\mathrm{d}\mathbf{x}$ then the force experienced is given by $\mathbf{F} \cdot \mathrm{d}\mathbf{x}$. In our original set up (in $\mathbb{R}^n$) our line integral is thus

$$\int_C \mathbf{F} \cdot \mathrm{d}x = \int_C (F_1 \,\mathrm{d}x_1 + \cdots + F_n \,\mathrm{d}x_n) = \int_a^b \mathbf{F}(\mathbf{g}(t)) \cdot \mathbf{g}'(t) dt.$$

The change of variable formula will quickly convince you that this formula's *magnitude* is invariant under change of parameterization, but notice that it can change sign. In our fish-analogy, imagine the fish swimming with the current in a river, versus swimming exactly the same path but against the current of the river. In both cases, the total magnitude of work is the same, but in one instance the fish had to exert energy to work against the current, and in the other the fish was moved by the current with no energy required. Hence orientation can change the sign of the line integral. Furthemore, notice that multiplying and dividing by the speed function $|g'(t)|$ and recalling that the element of arc satisfies $\mathrm{d}s = |g'(t)| \,\mathrm{d}t$ we have

$$\int_a^b \mathbf{F}(\mathbf{g}(t)) \cdot \frac{\mathbf{g}'(t)}{|\mathbf{g}'(t)|} |\mathbf{g}'(t)| dt = \int_a^b \mathbf{F}(\mathbf{g}(t)) \cdot \hat{\mathbf{T}}(t) ds$$

where $\hat{\mathbf{T}}(t) = \frac{\mathbf{g}'(t)}{|\mathbf{g}'(t)|}$ is the unit speed vector. The component $\mathbf{F}(\mathbf{g}(t)) \cdot \hat{\mathbf{T}}(t)$ is the projection of $\mathbf{F}$ in the direction of $\hat{\mathbf{T}}$, and is precisely the component of the field $\mathbf{F}$ doing work in the direction of $\hat{\mathbf{T}}$.

Example 5.5

Evaluate the line integral $\int_C \mathbf{F} \cdot \mathrm{d}\mathbf{x}$ if C is the curve $g(t) = (t, t^2, t^2)$ for $0 \leq t \leq 1$, and $\mathbf{F}(x, y, z) = (xyz, y^2, z)$.

Solution. Clearly $g'(t) = (1, 2t, 2t)$ and $\mathbf{F}(g(t)) = \mathbf{F}(t, t^2, t^2) = (t^5, t^4, t^2)$ so their dot product yields

$$\mathbf{F}(g(t)) \cdot g'(t) = (t^5, t^4, t^2) \cdot (1, 2t, 2t) = t^5 + 2t^5 + 2t^3 = 3t^5 + 2t^3.$$

Integrating gives

$$\begin{aligned} \int_0^1 \mathbf{F}(g(t)) \cdot g'(t) dt &= \int_0^1 3t^5 + 2t^3 \\ &= \frac{1}{2} \left[t^6 + t^4 \right]_0^1 \\ &= 1. \end{aligned}$$

■

Example 5.6

Evaluate the line integral $\int_C \mathbf{F} \cdot \mathrm{d}\mathbf{x}$ if $\mathbf{F}$ is the same vector field in Example 5.5 but C is the curve

$$C = \left\{(x, y, z) : x^2 + y^2 = 1, z = 1\right\}.$$

Solution. We can parameterize C via the function $g(t, z) = (\cos(t), \sin(t), 1)$ where $0 \leq t \leq 2\pi$. One can compute that

$$\mathbf{F}(g(t)) = \left(\cos(t)\sin(t), \sin^2(t), 1\right), \quad g'(t) = (-\sin(t), \cos(t), 0),$$

$$\mathbf{F}(g(t)) \cdot g'(t) = -\cos(t)\sin^2(t) + \cos(t)\sin^2(t) + 0 = 0$$

and hence $\displaystyle\int_C \mathbf{F} \cdot \mathrm{d}\mathbf{x} = 0$. ■

These examples are not typical in that they were actually easily solved. Example 5.5 was simple because everything was written as polynomials, while Example 5.6 magically became zero before having to integrate. In general, line integrals will yield exceptionally nasty integrands, necessitating that we expand our line integral toolbox.

5.4 Green's Theorem

Line integrals can be quite tricky to compute and so we would like to develop tools to facilitate their computation. Before proceeding, we will have to make a few preliminary definitions:

Definition 5.7

1. A *simple closed curve* of $\mathbb{R}^n$ is a simple curve (see Definition 3.16) whose endpoints coincide.
2. A *regular region* in $\mathbb{R}^n$ is a compact subset of $\mathbb{R}^n$ which is the closure of its interior.
3. A regular region $S \subseteq \mathbb{R}^n$ has a *piecewise smooth boundary* if its boundary ∂S is a finite union of piecewise, simple closed curves.

Example 5.8

1. The circle S^1 is a simple closed curve. Indeed, choose $g(t) = (\cos(t), \sin(t))$, $0 \leq t \leq 2\pi$ to parameterize the circle. It is clearly injective on $(0, 2\pi)$, and the endpoints coincide since $g(0) = g(2\pi)$.
2. The set $[0, 1] \cup \{2\}$ is certainly compact, but is not a regular region. Indeed, its interior is the set $(0, 1)$ whose closure is $[0, 1]$.

Given a regular region $S \subseteq \mathbb{R}^2$ with a piecewise smooth boundary ∂S, we define the Stokes' orientation to be the orientation of the boundary such that the interior of the set is "on the left."

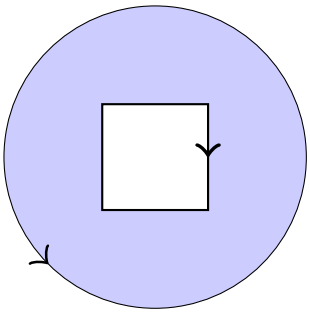


Figure 41: A regular region with a piecewise smooth boundary and the Stokes orientation. Notice that the orientation on the internal boundary is the opposite of the external boundary.

Notice in particular that this will mean that a space with a hole in it will have its external boundary oriented counter-clockwise, while the interior boundary will be oriented clockwise.

Theorem 5.9: Green's Theorem

If $S \subseteq \mathbb{R}^2$ is a regular region with piecewise smooth boundary ∂S, endowed with the Stokes orientation, and $\mathbf{F} : \mathbb{R}^2 \to \mathbb{R}^2$ is a C^1-vector field, then

$$\int_{\partial S} \mathbf{F} \cdot \mathrm{d}\mathbf{x} = \iint_S \left(\frac{\partial F_2}{\partial x_1} - \frac{\partial F_1}{\partial x_2} \right) \mathrm{d}A.$$

The full proof requires an argument that every space S in the hypothesis admits a decomposition into "simple" sets. Rather than harp on this point, we will choose examine how the proof would ideally work given such a simple set.

Before doing that however, let us take a second to think about what this theorem is saying: Depending on how we choose to look at it, we can determine what is happening on the interior of S just by looking at something on its boundary ∂S; or vice-versa, we can determine something about what's happening on the boundary of S, just by looking at what's happening on the interior. *A priori*, this is quite a surprising result: why should information about the interior and boundary be in any way related?

On the other hand, an argument can be made that the Fundamental Theorem of Calculus only cares about information on the boundary (and this will manifest in the proof), or alternatively that if we know how our vector field is leaving/entering the set, then we can infer a lot of information about the interior. Either way, Green's theorem is powerful and deserves some time contemplating.

Proof. Recall from our time doing iterated integrals that it is often convenient to be able to write a region as being bounded either as a function of x or a function of y. We will say that S is *x-simple* if it can be written as

$$S = \{a \leq x \leq b, \quad \phi_1(x) \leq y \leq \phi_2(x)\}$$

and *y-simple* if

$$S = \{c \leq y \leq d, \quad \psi_1(y) \leq x \leq \psi_2(y)\}.$$

Assume then that S is both x-simple and y-simple, and consider for now only the x-simple

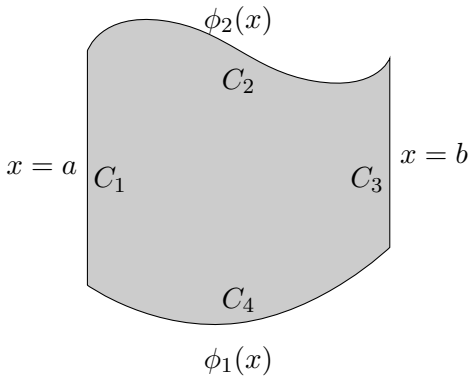


Figure 42: An x-simple description of our set S.

description. Label the edges $\partial S = C_1 + C_2 + C_3 + C_4$ as illustrated in Figure 42. For this part of the proof, we are going to compute $\int_C F_1 \, \mathrm{d}x$. On the straight line components (corresponding to $x = a$ and $x = b$) we have $\mathrm{d}x = 0$, and hence

$$\int_C F_1 \, \mathrm{d}x = \int_{C_2} F_1 \, \mathrm{d}x + \int_{C_4} F_1 \, \mathrm{d}x.$$

Notice that $\phi_2(x)$ runs from b to a in the Stokes orientation, so we must introduce a negative sign to get

$$\int_C F_1 \, \mathrm{d}x = \int_a^b F_1(x, \phi_1(x)) \, \mathrm{d}x - \int_a^b F_1(x, \phi_2(x)) \, \mathrm{d}x. \tag{5.2}$$

On the other hand, applying the Fundamental Theorem of Calculus to the following iterated integral:

$$\iint_S \frac{\partial F_1}{\partial y} \, \mathrm{d}A = \int_a^b \int_{\phi_1(x)}^{\phi_2(x)} \frac{\partial F_1}{\partial y} \, \mathrm{d}x = \int_a^b \left[F_1(x, \phi_2(x)) - F_1(x, \phi_1(x)) \right] \, \mathrm{d}x. \tag{5.3}$$

Comparing (5.2) and (5.3) yields

$$\int_{\partial S} F_1 \, \mathrm{d}x = - \iint_S \frac{\partial F_1}{\partial y} \, \mathrm{d}A.$$

Proceeding in precisely the same manner but using y-simple description of S results in

$$\int_{\partial S} F_2 \, \mathrm{d}y = \iint_S \frac{\partial F_2}{\partial x} \, \mathrm{d}A.$$

Naturally, combining these two results tells us that

$$\int_{\partial S} \mathbf{F} \cdot \, \mathrm{d}x = \iint_S \left[\frac{\partial F_2}{\partial x} - \frac{\partial F_1}{\partial y} \right] \, \mathrm{d}A.$$

More generally, the remainder of the proof hinges upon the ability to decompose S into subsets which are both x- and y-simple; namely $S = S_1 \cup \cdots \cup S_n$ where the S_n have disjoint interior and are

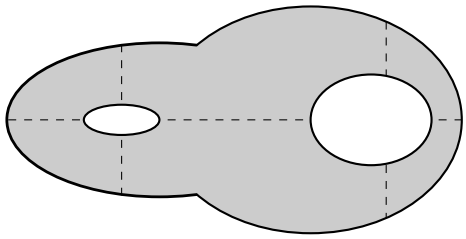


Figure 43: To prove Green's Theorem on more general regions, we decompose the region into subregions which are both x- and y-simple.

xy-simple. We will omit the fact that any regular region with piecewise smooth boundary admits such a decomposition.

Notice that interior boundaries (those that make up part of the boundary of ∂S_i but not of ∂S) have orientations which "cancel" each other out. By the additivity of line integrals and iterated integrals, the result then follows. □

Example 5.10

Compute the line integral

$$\oint_C \left[\left(2y + \sqrt{1+x^5}\right)\,\mathrm{d}x + \left(5x - e^{y^2}\right)\,\mathrm{d}y\right],$$

when C is the curve $x^2 + y^2 = R^2$.

Solution. This would be an exceptionally difficult integral to calculate if one was not permitted to use Green's Theorem; however, it becomes almost trivial after applying Green's Theorem. Let D be the interior of the radius R-circle, which we know has area πR^2. Green's Theorem gives

$$\begin{aligned}\oint_C \Big[\underbrace{\left(2y + \sqrt{1+x^5}\right)}_{F_1}\,\mathrm{d}x + \underbrace{\left(5x - e^{y^2}\right)}_{F_2}\,\mathrm{d}y\Big] &= \iint_D \left(\frac{\partial F_2}{\partial x} - \frac{\partial F_1}{\partial y}\right)\mathrm{d}A \\ &= \iint_D (5-2)\,\mathrm{d}A \\ &= 3\pi R^2.\end{aligned}$$

We did not even have to compute the iterated integral since we know the area of D! ■

Example 5.11

Determine the line integral $\int_{\partial S} \mathbf{F}\cdot\,\mathrm{d}\mathbf{x}$ where $\mathbf{F}(x,y) = (1, xy)$ and S is the triangle whose vertices are $(0,0), (1,0)$ and $(1,1)$, oriented counter clockwise.

Solution. We can write the triangle as an x-simple set

$$S = \{0 \le x \le 1,\ 0 \le y \le x\}$$

so that by Green's Theorem

$$\begin{aligned}\int_{\partial_S} \mathbf{F} \cdot \mathrm{d}\mathbf{x} &= \iint_S \left[\frac{\partial F_2}{\partial x} - \frac{\partial F_1}{\partial y}\right] \mathrm{d}A = \iint_S y \,\mathrm{d}A \\ &= \int_0^1 \int_0^x y \,\mathrm{d}y \,\mathrm{d}x = \frac{1}{6}.\end{aligned}$$

Let's compute the line integral and verify that we get the same result. Parameterize the line C_1 by $g_1(t) = (t, 0)$ for $0 \leq t \leq 1$, yielding

$$\int_{C_1} \mathbf{F} \cdot \mathrm{d}\mathbf{x} = \int_0^1 \mathbf{F}(g_1(t)) \cdot g_1'(t) dt = \int_0^1 (1, 0) \cdot (1, 0) dt = 1$$

To parameterize C_2 we use $g_2(t) = (1, t)$ for $0 \leq t \leq 1$ yielding

$$\int_{C_2} \mathbf{F} \cdot \mathrm{d}\mathbf{x} = \int_0^1 (1, t) \cdot (0, 1) dt = \int_0^1 t dt = \frac{1}{2}.$$

Finally, we parameterize C_2 and $g_3(t) = (1-t, 1-t)$ for $0 \leq t \leq 1$ (we choose this over $g_3(t) = (t, t)$ to keep the correct orientation).

$$\begin{aligned}\int_{C_3} \mathbf{F} \cdot \mathrm{d}\mathbf{x} &= \int_0^1 (1, (1-t)^2) \cdot (-1, -1) dt \\ &= \int_0^1 -1 - (1-t)^2 dt = -\frac{4}{3}.\end{aligned}$$

Combining everything together we get

$$\int_C \mathbf{F} \cdot \mathrm{d}\mathbf{x} = \int_{C_1} \mathbf{F} \cdot \mathrm{d}\mathbf{x} + \int_{C_2} \mathbf{F} \cdot \mathrm{d}\mathbf{x} + \int_{C_3} \mathbf{F} \cdot \mathrm{d}\mathbf{x} = 1 + \frac{1}{2} - \frac{4}{3} = \frac{1}{6}$$

exactly as we expected. ■

5.5 Exact and Closed Vector Fields

Line integrals can have some pretty surprising properties. In fact, it turns out that one can actually use line integrals to tell you something about the geometry of a surface (though this is a rather advanced topic for this course). Nonetheless, let's set up the framework for utilizing/understanding how this works.

5.5.1 Exact Vector Fields

Our first result is a version of the Fundamental Theorem of Calculus:

Theorem 5.12: Fundamental Theorem of Calculus for Line Integrals

If $C \subseteq \mathbb{R}^n$ is a C^1 curve given by a parameterization $g : [a, b] \to \mathbb{R}^n$ and $\mathbf{F} : \mathbb{R}^n \to \mathbb{R}^n$ is a vector field such that there exists a C^1 function $f : \mathbb{R}^n \to \mathbb{R}$ satisfying $\mathbf{F} = \nabla f$ then

$$\int_C \mathbf{F} \cdot d\mathbf{x} = f(\gamma(b)) - f(\gamma(a)).$$

In particular, the integral only depends on the endpoints $\gamma(a)$ and $\gamma(b)$ of the curve C.

Proof. Assume that $\mathbf{F} = \nabla f$ and let C be some oriented curve with parameterization $\gamma : [a, b] \to \mathbb{R}^n$. Straightforward computation then reveals that

$$\begin{aligned}
\int_C \mathbf{F} \cdot \mathrm{d}x &= \int_a^b \mathbf{F}(\gamma(t)) \cdot \gamma'(t)\,\mathrm{d}t \\
&= \int_a^b \nabla f(\gamma(t)) \cdot \gamma'(t)\,\mathrm{d}t && \text{by the chain rule} \\
&= \int_a^b \frac{\mathrm{d}}{\mathrm{d}t} f(\gamma(t))\,\mathrm{d}t \\
&= f(\gamma(b)) - f(\gamma(a))
\end{aligned}$$

by the Fundamental Theorem of Calculus. □

In general we know that the choice of curve makes a significant difference to the value of the line integral, so this theorem tells us that there is a particular class of vector fields on which the line integral does not seem to care about the path we choose. These vector fields are so important that we give them a special name.

Definition 5.13

Any vector field $\mathbf{F} : \mathbb{R}^n \to \mathbb{R}^n$ satisfying $\mathbf{F} = \nabla f$ for some C^1-function $f : \mathbb{R}^n \to \mathbb{R}$ is called an *exact vector field*. The function f is sometimes referred to as a *scalar potential*.

Example 5.14

Determine which of the following vector fields are exact:

1. $\mathbf{F}(x, y, z) = (yze^{xyz}, xze^{xyz}, xye^{xyz})$,
2. $\mathbf{G}(x, y, z) = (2xy, x^2 + \cos(z), -y\sin(z))$,
3. $\mathbf{H}(x, y, z) = (x + y, x + z, y + z)$.

Solution. Our strategy will be to work as follows: If $\mathbf{F} = \nabla f$ then we know $F_1 = \frac{\partial f}{\partial x}$. We thus integrate the 1st component with respect to x, to get $f(x, y, z) = \hat{f}(x, y, z) + g(y, z)$, where $\hat{f}(x, y, z)$ is what we compute from the integral, and $g(y, z)$ is the "constant" (with respect to x) of integration.

We can then differentiate f with respect to y to get

$$\frac{\partial f}{\partial y} = \frac{\partial \hat{f}}{\partial y} + \frac{\partial g}{\partial y}$$

and compare this to F_2. With any luck, we will be able to solve $g(y,z) = \hat{g}(y,z) + h(z)$, and perform a similar technique to compute h. Of course, at the end of the day we can only evaluate f up to a constant, but this constant will not affect the value of the integral.

1. The student can quickly check that $f(x,y,z) = e^{xyz}$ works.

2. This example requires a bit more work. We integrate the first term with respect to x to get $f(x,y,z) = x^2y + g(y,z)$ for some function $g(y,z)$. Differentiating with respect to y and setting $\frac{\partial f}{\partial y} = G_2$ we get

$$\frac{\partial f}{\partial y} = x^2 + \frac{\partial g}{\partial y} = x^2 + \cos(z), \qquad \frac{\partial g}{\partial y} = \cos(z).$$

   We integrate to find that $g(y,z) = y\cos(z) + h(z)$ for some yet to be determined function $h(z)$, giving $f(x,y,z) = x^2y + y\cos(z) + h(z)$. Differentiating with respect to z we get $\frac{\partial f}{\partial z} = -y\sin(z)$ which is exactly G_3. Hence $h(z)$ is a constant, which we might as well set to 0, and we conclude that $f(x,y,z) = x^2y + y\cos(z)$.

3. We integrate F_1 with respect to x to get $f(x,y,z) = \frac{1}{2}x^2 + yx + g(y,z)$. Differentiating with respect to y gives $\frac{\partial f}{\partial y} = x + \frac{\partial g}{\partial y}$. Equating to H_2 tells us that $\frac{\partial g}{\partial y} = z$, so that $f(x,y,z) = \frac{1}{2}x^2 + yx + yz + h(z)$. Finally, $\frac{\partial f}{\partial z} = y + \frac{\partial h}{\partial z} = H_3 = y + z$ so it must be the case that $\frac{\partial h}{\partial z} = z$, and we conclude that

$$f(x,y,z) = \frac{1}{2}x^2 + yx + yz + \frac{1}{2}z^2.$$ ■

Example 5.15

Determine the line integral $\int_C \mathbf{F} \cdot \mathrm{d}\mathbf{x}$ where $\mathbf{F}(x,y,z) = (2xy, x^2 + \cos(z), -y\sin(z))$ and C is the curve

$$C = \left\{(x,y,z) : x^2 + y^2 + z^2 = 1, y = -z, y \leq 0\right\},$$

oriented to start at $(-1,0,0)$

Solution. The curve C lies on the intersection of the unit sphere S^2 and the plane $z = -y$. This would normally be a full circle, except for the fact that the condition $y \leq 0$ ensures that we only pass through one hemisphere. One could parameterize this and try to compute the line integral by hand, except that the result is truly terrifying. Instead, all one needs to realize is that the endpoints of this curve are $(\pm 1, 0, 0)$. Furthermore, in Example 5.14 we showed that $\mathbf{F} = \nabla f$ where $f(x,y,z) = x^2y + y\cos(z)$. Consequently, the line integral is just

$$\int_C \mathbf{F} \cdot \mathrm{d}x = f(1,0,0) - f(-1,0,0) = 0.$$ ■

5.5.2 Conservative Vector Fields

We would like to explore whether there are other vector fields for which line integrals only depend upon endpoints. To this end, we have the following lemma:

Lemma 5.16

If $\mathbf{F}$ is a continuous vector field on an open set $U \subseteq \mathbb{R}^n$ then the following are equivalent:

1. If C_1 and C_2 are any two oriented curves in U with the same endpoints, then
$$\int_{C_1} \mathbf{F} \cdot \mathrm{d}\mathbf{x} = \int_{C_2} \mathbf{F} \cdot \mathrm{d}\mathbf{x}.$$
2. If C is a closed curve, then
$$\int_C \mathbf{F} \cdot \mathrm{d}\mathbf{x} = 0.$$

Proof.

(1)⇒(2) Pick a point $\mathbf{a}$ on C and declare that C has both endpoints equal to $\mathbf{a}$. Clearly, these are the same endpoints as the constant curve at $\mathbf{a}$, which we call $\hat{C}$, and so by (1) we have
$$\int_C \mathbf{F} \cdot \mathrm{d}\mathbf{x} = \int_{\hat{C}} \mathbf{F} \cdot \mathrm{d}\mathbf{x} = 0$$
where we note that integrating over the constant curve will certainly give a result of zero.

(2)⇒(1) Let the endpoints of C_1 be called $\mathbf{a}$ and $\mathbf{b}$. Since C_2 has the same endpoints, we may define a closed curve C as the one which traverses C_1 from $\mathbf{a}$ to $\mathbf{b}$, and then traverses C_2 from $\mathbf{b}$ to $\mathbf{a}$. Now C_2 has the opposite orientation of C_1, so applying (2) we get
$$0 = \int_C \mathbf{F} \cdot \mathrm{d}\mathbf{x} = \int_{C_1} \mathbf{F} \cdot \mathrm{d}\mathbf{x} - \int_{C_2} \mathbf{F} \cdot d\mathbf{x},$$
from which the result follows. □

Any vector field which satisfies either of the above (equivalent) conditions is called an *conservative vector field*. The name is derived from physics: In a system in which energy is conserved, only the initial and terminal configurations of the state determine energy transfer and the system ignores all other things which happen in between.

The FTC for Line Integrals tells us that exact vector fields are conservative. It turns out that that this exhausts the list of all conservative vector fields.

Theorem 5.17

If $S \subseteq \mathbb{R}^n$ is an open set, then a continuous vector field $F : S \to \mathbb{R}^n$ is conservative if and only if there is a C^1 function $f : S \to \mathbb{R}$ such that $F = \nabla f$.

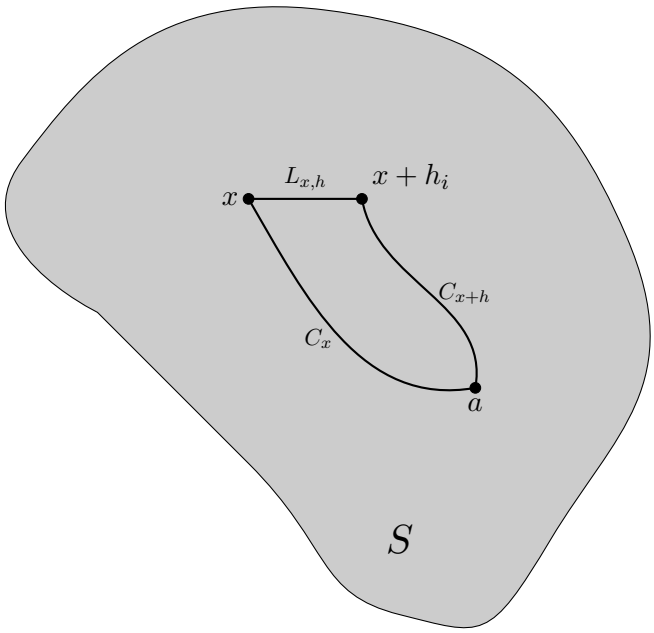


Figure 44

Proof. ($\Leftarrow$) This follows from the Fundamental Theorem of Calculus for Line Integrals.

($\Rightarrow$) Conversely, assume that $F : \mathbb{R}^n \to \mathbb{R}^n$ is a conservative vector field. Without loss of generality, we shall assume that S is connected (otherwise just do the following for each connected component). Fix some point $a \in S$, and define for each $\mathbf{x} \in S$ let $C_{\mathbf{x}}$ be a curve from $\mathbf{a}$ to $\mathbf{x}$ (which always exists since open connected sets are path-connected) and define the function

$$f(\mathbf{x}) = \int_{C_{\mathbf{x}}} F \cdot \mathrm{d}\mathbf{x}.$$

This is well defined since, by assumption, the definition is invariant of our choice of curve $C_{\mathbf{x}}$. Now we claim that $\nabla f = F$ and is C^1, and both claims will be obvious if we show that $\partial_i f = F_i$ for each $i = 1, \ldots, n$.

To see that this is the case, we will show that the ith partial of f is precisely F_i. Fix $\mathbf{x} \in S$ and choose $\mathbf{h} = he_i$ sufficiently small so that the line $L_{\mathbf{x},\mathbf{h}}$ between $\mathbf{x}$ and $\mathbf{x} + \mathbf{h}$ remains in S (see Figure 44). Let $\hat{C}_{\mathbf{x}+\mathbf{h}}$ be $C_{\mathbf{x}}$ followed by $L_{\mathbf{x},\mathbf{h}}$ so that

$$\begin{aligned} f(\mathbf{x}+\mathbf{h}) &= \int_{C_{\mathbf{x}+\mathbf{h}}} \mathbf{F} \cdot \mathrm{d}\mathbf{x} = \int_{C_{\mathbf{x}}} \mathbf{F} \cdot \mathrm{d}\mathbf{x} + \int_{L_{\mathbf{x},\mathbf{h}}} \mathbf{F} \cdot \mathrm{d}\mathbf{x} \\ &= f(\mathbf{x}) + \int_{L_{\mathbf{x},\mathbf{h}}} \mathbf{F} \cdot \mathrm{d}\mathbf{x}. \end{aligned}$$

Parameterize the line $L_{\mathbf{x},\mathbf{h}}$ be $g(t) = \mathbf{x} + te_i$ for $0 \leq t \leq h$ so that $g'(t) = e_i$ and

$$\begin{aligned}\int_{L_{\mathbf{x},\mathbf{h}}} \mathbf{F} \cdot \mathrm{d}\mathbf{x} &= \int_0^h \mathbf{F}(x_1, \ldots, x_{i-1}, x_i + t, x_{i+1}, \ldots, x_n) \cdot (0, \ldots, 0, 1, 0, \ldots, 0) dt \\ &= \int_0^h F_i(x_1, \ldots, x_i + t, \ldots, x_n) dt.\end{aligned}$$

Computing $\frac{\partial f}{\partial x_i}$ we have

$$\begin{aligned}\frac{\partial f}{\partial x_i} &= \lim_{h\to 0} \frac{f(\mathbf{x}+\mathbf{h}) - f(\mathbf{x})}{h} = \lim_{h\to 0} \frac{1}{h} \int_{L_{\mathbf{x},\mathbf{h}}} \mathbf{F} \cdot \mathrm{d}\mathbf{x} \\ &= \lim_{h\to 0} \frac{1}{h} \int_0^h F_i(x_1, \ldots, x_i + t, \ldots, x_n) dt \\ &= F_i(\mathbf{x}) \qquad \text{L'Hôpital} \qquad \square\end{aligned}$$

5.5.3 Closed Vector Fields

Theorem 5.17 is a very nice condition, but it is fairly intractable to compute all possible line integrals, and it can often be difficult to ascertain whether you are the gradient of a function.

There is a relatively simple *necessary* condition to determine whether a vector field is conservative. If $F = \nabla f$ then $F_i = \partial_i f$. Since mixed partials commute, we then have

$$\partial_i F_j = \partial_i \partial_j f = \partial_j \partial_i f = \partial_j F_i,$$

or alternatively

$$\frac{\partial F_i}{\partial x_j} - \frac{\partial F_j}{\partial x_i} = 0, \qquad i \neq j. \tag{5.4}$$

Vector fields which satisfy (5.4) are called *closed vector fields.* However, not all closed vector fields are exact. Also, notice that if we are working in $\mathbb{R}^3$ then the components given in (5.4) correspond to the components of the curl. Hence closed vector fields of $\mathbb{R}^3$ are irrotational.

Example 5.18 Consider the vector field $F(x,y) = \frac{1}{x^2+y^2}(-y,x)$ defined on the open set $\mathbb{R}^2\backslash_{\{(0,0)\}}$. It is easy to see that

$$\frac{\partial F_2}{\partial x} = \frac{\partial}{\partial x}\frac{x}{x^2+y^2} = \frac{y^2-x^2}{(x^2+y^2)^2}$$
$$\frac{\partial F_1}{\partial y} = \frac{\partial}{\partial y}\frac{-y}{x^2+y^2} = \frac{y^2-x^2}{(x^2+y^2)^2},$$

so $\mathbf{F}$ is a closed vector field.
On the other hand, let C be any circle, parameterize by $\gamma(\theta) = (r\cos(\theta), r\sin(\theta))$. Then $\gamma'(t) = (-r\sin(\theta), r\cos(\theta))$ and our line integral becomes:

$$\begin{aligned}\int_{C_r} \mathbf{F}\cdot \mathrm{d}\mathbf{x} &= \frac{1}{r^2}\int_0^{2\pi} (-r\sin(\theta), r\cos(\theta))\cdot(-r\sin(\theta), r\cos(\theta))\,\mathrm{d}\theta \\ &= \frac{1}{r^2}\int_0^{2\pi} \left[r^2\sin^2(\theta) + r^2\cos^2(\theta)\right]\,\mathrm{d}\theta \\ &= 2\pi.\end{aligned}$$

If F were conservative, this would have to be zero; hence F is an example of a closed vector field which is not exact.

So what went wrong with the above example? Essentially, because our vector field $\mathbf{F}$ is not C^1 unless our space has a hole at the origin $(0,0)$, our vector field/line integral was able to detect that hole. In fact, try computing the above line integral around any closed curve which does not contain the origin, and you will see that the result is zero.

It turns out that closed vector fields are locally exact. In order to describe what we mean, we must introduce a new definition:

Definition 5.19

A set $U \subseteq \mathbb{R}^n$ is said to be *star-shaped* if there exists a point $a \in U$ such that for every point $x \in U$ the straight line connected x to a is contained in U.

Notice that every convex set is star shaped, though the converse need not be true. For example, Figure 45 gives an example of a star shaped set in $\mathbb{R}^2$ that is not convex.

Theorem 5.20: Poincaré Lemma

If $U \subseteq \mathbb{R}^n$ is star-shaped and $\mathbf{F}$ is a closed vector field on U, then $\mathbf{F}$ is exact on U.

Proof. Without loss of generality, lets assume that U is star shaped about the origin. For any $x \in U$ let $\gamma_x(t) = tx$ be the straight line connecting the origin to x, and define the function

$$f(x) = \int_{\gamma_x} \mathbf{F}\cdot\mathrm{d}\mathbf{x} = \int_0^1 F_1(tx)x_1 + \cdots + F_n(tx)x_n\,\mathrm{d}t.$$

Notice that this is well defined since $\gamma_x(t) \in U$ for all t, and there is a unique straight line connecting

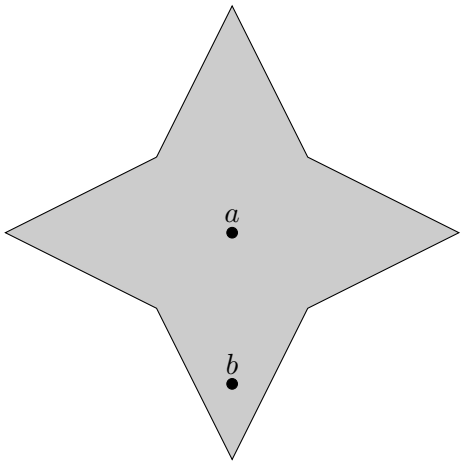


Figure 45: An example of a star shaped set which is not convex. The point a satisfies the required definition, while the point b does not.

0 to x. Now we claim that $\mathbf{F} = \nabla f$ on U. Inspecting one component at a time, we have

$$\begin{aligned}
\frac{\partial f}{\partial x_k} &= \int_0^1 \left[\sum_{i=1}^n \frac{\partial F_i}{\partial x_k}(tx) t x_i + F_k(tx) \right] \mathrm{d}t \\
&= \int_0^1 \left[\sum_{i=1}^n \frac{\partial F_k}{\partial x_i}(tx) t x_i + F_k(tx) \right] \mathrm{d}t && \text{since } F \text{ is closed} \\
&= \int_0^1 \frac{\mathrm{d}}{\mathrm{d}t} \left(t F_k(tx) \right) \mathrm{d}t \\
&= F_k(x).
\end{aligned}$$

Hence $\nabla f = \mathbf{F}$ as required. □

5.6 Surface Integrals

Line integrals captured the idea of a vector field doing work on a particle as it travelled a particular path. A similar idea is the surface integral, which calculates the amount of *flux* of a vector field passing through a surface.

5.6.1 Surface Area

Just as when we calculated the arc length of a curve, in order to compute surface area we are going to heuristically examine what an area element might look like. To do this, recall that given two linearly independent vectors $v, w \in \mathbb{R}^3$, the area of the parallelogram with vertices $0, v, w, v + w$ is given by $|v \times w|$ (Figure 46). The idea is to use this, but apply it to infinitesimal elements to get the corresponding area of a surface, so that the infinitesimal area-element is $\mathrm{d}A = |\,\mathrm{d}x \times \mathrm{d}y|$.

If we set $\mathrm{d}A = |\,\mathrm{d}x \times \mathrm{d}y|$, it should not come as a surprise that the surface area of a surface S

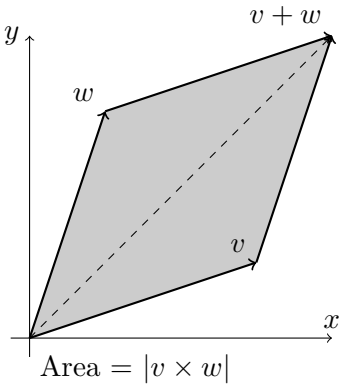


Figure 46: The area of a parallelogram in $\mathbb{R}^2$ is given by the determinant of the defining edges.

will be given by just integrating the area element

$$A(S) = \iint_S \mathrm{d}A$$

but just as in the case of arc-length, this is a infeasible equation if we don't have a parameterization of the surface.

Thus let $G : R \subseteq \mathbb{R}^2 \to \mathbb{R}^3$ be a parameterization of a surface S in $\mathbb{R}^3$, and fix some $(u_0, v_0) \in \mathbb{R}^2$. Applying infinitesimal translations $\mathrm{d}u$ and $\mathrm{d}v$ to (u_0, v_0), we get the corresponding vectors

$$G(u, v + \mathrm{d}v) - G(u, v) = \frac{\partial G}{\partial v}\,\mathrm{d}v, \qquad G(u + \mathrm{d}u, v) - G(u, v) = \frac{\partial G}{\partial u}\,\mathrm{d}u.$$

These are our two "vectors" with which we will use to compute the area element. Just as in the $\mathbb{R}^3$ case, we will take their cross product to get

$$\mathrm{d}A = \left|\frac{\partial G}{\partial u} \times \frac{\partial G}{\partial v}\right| \mathrm{d}u\,\mathrm{d}v,$$

and integrating over the whole surface thus gives us our surface area

$$A(S) = \iint_R \left|\frac{\partial G}{\partial u} \times \frac{\partial G}{\partial v}\right| \mathrm{d}u\,\mathrm{d}v.$$

If we use coordinates, this expression will even look a little like our arc-length formula. Set $(x, y, z) = G(u, v)$ so that $\frac{\partial G}{\partial u} = (x_u, y_u, z_u)$ and $\frac{\partial G}{\partial v} = (x_v, y_v, z_v)$. Notice that

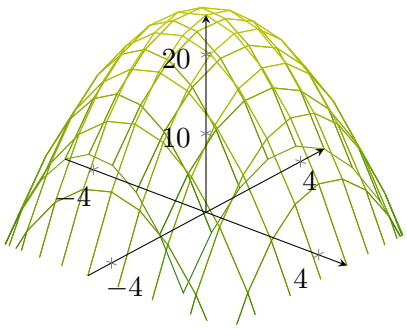
$$\begin{aligned}
\left|\frac{\partial G}{\partial u} \times \frac{\partial G}{\partial v}\right| &= \left|\det \begin{pmatrix} \mathbf{e}_1 & \mathbf{e}_2 & \mathbf{e}_3 \\ x_u & y_u & z_u \\ x_v & y_v & z_v \end{pmatrix}\right| \\
&= |(y_u z_v - z_u y_v, z_u x_v - x_u z_v, x_u y_v - y_u x_v)| \\
&= \sqrt{\left[\frac{\partial(y,z)}{\partial(u,v)}\right]^2 + \left[\frac{\partial(z,x)}{\partial(u,v)}\right]^2 + \left[\frac{\partial(x,y)}{\partial(u,v)}\right]^2}
\end{aligned}$$

This works if we have a parameterization of our surface, but what if we are given S as the graph of a C^1-function? Recall that if $z = f(x, y)$ then we can write this parametrically as $G(u, v) = (u, v, f(u, v))$ in which case

$$\left|\frac{\partial G}{\partial u} \times \frac{\partial G}{\partial v}\right| = \sqrt{1 + \left[\frac{\partial f}{\partial x}\right]^2 + \left[\frac{\partial f}{\partial y}\right]^2}.$$

Example 5.21

Find the surface area of surface defined by $x^2 + y^2 + z = 25$, lying above the xy-plane.



Solution. We can write our surface as the graph of the function $z = 25 - x^2 - y^2$, so that

$$\frac{\partial z}{\partial x} = -2x, \qquad \frac{\partial z}{\partial y} = -2y,$$

and the surface element is $\mathrm{d}A = \sqrt{1 + 4x^2 + 4y^2}$. The easiest way to integrate this is going to be through polar coordinates. Notice that in this case we have $z = 25 - r^2$, and since $z > 0$ this implies that $0 \le r \le 5$, and $0 \le \theta \le 2\pi$. Thus our integral becomes

$$\begin{aligned} A(S) &= \iint_S \sqrt{1 + 4x^2 + 4y^2}\,\mathrm{d}x\,\mathrm{d}y \\ &= \int_0^{2\pi}\int_0^5 r\sqrt{1 + 4r^2}\,\mathrm{d}r\,\mathrm{d}\theta \\ &= \frac{\pi}{4}\int_1^{101} \sqrt{u}\,\mathrm{d}u \qquad\qquad u = 1 + 4r^2 \\ &= \frac{\pi}{6}u^{3/2}\Big|_1^{101} = \frac{\pi}{6}(101^{3/2} - 1) \qquad\qquad \blacksquare \end{aligned}$$

5.6.2 Surface Integrals over Vector Fields

As with line integrals, surface integrals are going to depend on a choice of orientation, so what does it mean to orient a surface? An *orientation* of a surface S is a consistent (read: continuous) choice of normal vector to the surface. One can think of this as saying "I wish to everywhere define what it means to be right-handed," and an orientation does that.

Of particular use is that if S is parameterized $G(u,v)$, then $\frac{\partial G}{\partial u} \times \frac{\partial G}{\partial v}$ not only encodes information about the infinitesimal area, but is already a normal vector! Thus *a parameterization determines an orientation of the surface*, and we often write $\frac{\partial G}{\partial u} \times \frac{\partial G}{\partial v} = \hat{\mathbf{n}}\, \mathrm{d}A$, where $\hat{\mathbf{n}}$ is a unit normal vector to the surface. In particular, we can reverse the orientation simply by interchanging the roles of u and v! If S bounds a 3-manifold, we will say that it has *Stokes' orientation* if the normal vector of S points outwards with respect to the space it bounds.

It is worth mentioning at this point that not all surfaces are orientable. The simplest example of a non-orientable surface is a Möbius strip, formed by twisting a rectangle and gluing to ends together. The student can convince him/herself that if we start with a normal vector and traverse a circle around the Möbius band, our normal vector will have flipped when we arrive back at our original starting point. Hence one can find the surface area of non-orientable manifolds, but surface integrals will not make sense.

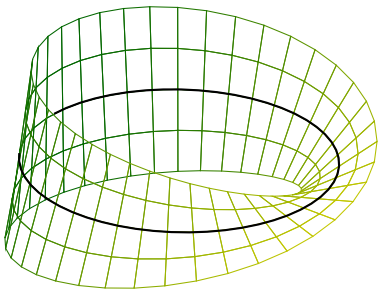


Figure 47: The Mobius Band is a non-orientable surface.

The idea of a surface integral is thus the following: Given a vector field $\mathbf{F} : \mathbb{R}^3 \to \mathbb{R}^3$ and a surface S, we want to compute the *flux* of the vector field through the surface. Effectively, if we again think of a vector field as representing forces or the flow of a fluid, the flux represents the amount of force/fluid passing through S. The vector field travelling in the direction $\hat{\mathbf{n}}$ is given by $\mathbf{F} \cdot \hat{\mathbf{n}}$ and so the surface integral is given by integrating each of these components:

$$\iint_S \mathbf{F} \cdot \hat{\mathbf{n}}\ \mathrm{d}A.$$

Notice that $\mathbf{F} \cdot \hat{\mathbf{n}}$ is precisely the vector field projected onto the normal of the surface, and so at the surface represents the vector field passing through the surface. Of course, this is not easily computed without a parameterization. If $G : R \subseteq \mathbb{R}^2 \to S \subseteq \mathbb{R}^3$ is such a parameterization, our flux becomes

$$\iint_S \mathbf{F} \cdot \hat{\mathbf{n}}\ \mathrm{d}A = \iint_R \mathbf{F}(G(u,v)) \cdot \left[\frac{\partial G}{\partial u} \times \frac{\partial G}{\partial v}\right] \mathrm{d}u\, \mathrm{d}v.$$

Example 5.22

Evaluate the flux of $\mathbf{F}(x,y,z) = (x^2+y, y^2z, x^2y)$ through the surface $S = [0,1] \times [0,1] \times \{0\}$, oriented pointing in the positive z-direction.

Solution. It is easy to parameterize this surface as

$$G(s,t) = (s,t,0), \quad 0 \le s \le 1, 0 \le t \le 1,$$

from which we get

$$\frac{\partial G}{\partial s} \times \frac{\partial G}{\partial t} = \begin{vmatrix} i & j & k \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{vmatrix} = (0, 0, 1).$$

This is oriented in the correct direction, so we proceed with the surface integral to get

$$\begin{aligned} \iint_S \mathbf{F} \cdot \hat{\mathbf{n}} \, dA &= \int_0^1 \int_0^1 (s^2 + t, t^2, s^2 t^2) \cdot (0, 0, 1) ds dt \\ &= \int_0^1 \int_0^1 s^2 t^2 ds dt = \frac{1}{9}. \end{aligned}$$

■

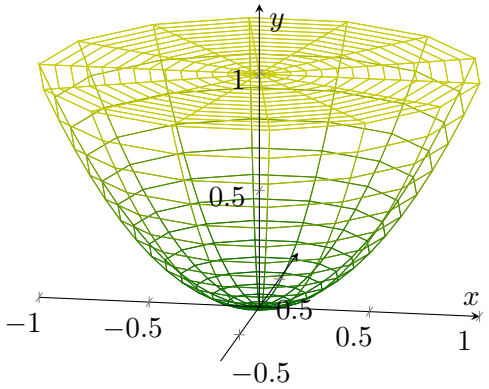
Sometimes it is necessary to break our surfaces into several pieces in order to determine the integral, as the next example demonstrates.

Example 5.23

Evaluate $\iint_S \mathbf{F} \cdot \hat{\mathbf{n}} \, dA$ where $F(x, y, z) = (0, y, -z)$ and S is the surface defined by

$$S = \left\{ y = x^2 + z^2 : 0 \leq y \leq 1 \right\} \cup \left\{ x^2 + z^2 \leq 1 : y = 1 \right\},$$

endowed with the Stokes orientation.



Solution. This space looks like the paraboloid, capped by the unit disk. Rather than trying to handle both parts of S at the same time, we break it into the paraboloid S_1 and the disk D separately.

We can parameterize the paraboloid as $(x, y, z) = (r\cos(\theta), r^2, r\sin(\theta))$ with $0 \leq r \leq 1$ and $0 \leq \theta \leq 2\pi$. Then

$$\frac{\partial G}{\partial r} = (\cos(\theta), 2r, \sin(\theta)), \quad \frac{\partial G}{\partial \theta} = (-r\sin(\theta), 0, r\cos(\theta)).$$

The cross product is then easily computed and we get

$$\frac{\partial G}{\partial r} \times \frac{\partial G}{\partial \theta} = \left(2r^2\cos(\theta), -r, 2r^2\sin(\theta)\right).$$

Hence

$$\begin{aligned}\mathbf{F}(G(r,\theta)) \cdot \left[\frac{\partial G}{\partial r} \times \frac{\partial G}{\partial \theta}\right] &= (0, r^2, -r\sin(\theta)) \cdot (2r^2\cos(\theta), -r, 2r^2\sin(\theta)) \\ &= -r^3\left(1 + 2\sin^2(\theta)\right).\end{aligned}$$

Integrating thus gives us

$$\begin{aligned}\iint_S \mathbf{F} \cdot \hat{\mathbf{n}} \ \mathrm{d}A &= -\left[\int_0^1 r^3 \,\mathrm{d}r\right]\left[\int_0^{2\pi} 1 + 2\sin^2(\theta)\,\mathrm{d}\theta\,\mathrm{d}\theta\right] \\ &= -\pi.\end{aligned}$$

Now the tricky part of doing the cap is making sure that we choose a parameterization of the cap which gives the Stokes orientation (that is, always points in the positive y-direction). The student can verify that

$$G(r,\theta) = (r\cos(\theta), 1, r\sin(\theta)), \quad 0 \le r \le 1, 0 \le \theta \le 2\pi$$

satisfies

$$\frac{\partial G}{\partial r} \times \frac{\partial G}{\partial \theta} = (0, -r, 0),$$

so that this is actually oriented the wrong way! This is fine, and we can continue to work with this parameterization, so long as we remember to re-introduce a negative sign at the end of our computation. Now

$$\iint_D \mathbf{F} \cdot \hat{\mathbf{n}} \ \mathrm{d}A = \int_0^1 \int_0^{2\pi} -r \,\mathrm{d}r\,\mathrm{d}\theta = -\pi$$

so properly orienting gives the result π. Adding both factors we get $\pi + (-\pi) = 0$, so we conclude that the flux is 0. ■

5.7 The Divergence Theorem

The Divergence Theorem (or Gauss' Theorem) is the analog of Green's theorem for surface integrals.

Theorem 5.24: Divergence Theorem

Let $R \subseteq \mathbb{R}^3$ be a regular region with piecewise smooth boundary ∂R. If $\mathbf{F} : \mathbb{R}^3 \to \mathbb{R}^3$ is a C^1 vector field and ∂R is positively oriented with respect to R then

$$\iint_{\partial R} \mathbf{F} \cdot \hat{\mathbf{n}} \ \mathrm{d}A = \iiint_R \operatorname{div} \mathbf{F} dV.$$

Proof. As with Green's Theorem, we will only provide a very simplified proof which nonetheless captures the idea of the Divergence Theorem.

Assume that R is an xy-simple set, so that we can write it as

$$R = \{(x,y) \in W, \psi_1(x,y) \le z \le \psi_2(x,y)\}$$

where $W \in \mathbb{R}^2$ is some regular region. Breaking the statement of the Divergence Theorem into its vector components, our aim is to show that

$$\iint_{\partial R} F_3 \cdot \hat{\mathbf{e}_3} \cdot \hat{\mathbf{n}} \; \mathrm{d}A = \iiint_R \frac{\partial F_3}{\partial x_3} \; dV$$

where $\hat{\mathbf{e}_3} = (0, 0, 1)$ is the standard unit normal in the positive z-direction. Note that $\hat{\mathbf{e}_3} \cdot \hat{\mathbf{n}} = 0$ along the vertical line segments occurring along the boundary of W, while $\hat{\mathbf{e}_3}$ is consistent with the orientation of the top surface $(x, y, \phi_2(x, y))$ and is the opposite orientation of the bottom surface $(x, y, \phi_1(x, y))$. Thus,

$$\begin{aligned}
\iint_{\partial R} F_3 \hat{\mathbf{e}_3} \cdot \hat{\mathbf{n}} \; \mathrm{d}A &= \iint_W \left[F_3(x, y, \phi_2(x, y)) - F_3(x, y, \phi_1(x, y)) \right] \mathrm{d}x \, \mathrm{d}y \\
&= \iint_W \int_{\phi_1(x,y)}^{\phi_2(x,y)} \frac{\partial F_3}{\partial x_3}(x, y, z) \, \mathrm{d}z \\
&= \iiint_R \frac{\partial F_3}{\partial x_3} \; dV
\end{aligned}$$

which is exactly what we wanted to show. □

Example 5.25

Evaluate $\iint_S \mathbf{F} \cdot \hat{\mathbf{n}} \; \mathrm{d}A$ where $\mathbf{F}(x, y, z) = (y^2 z, y^3, xz)$ and S is the boundary of the cube defined by

$$C = \{-1 \leq x \leq 1, -1 \leq y \leq 1, 0 \leq z \leq 2\},$$

oriented so that the normal points outwards.

Solution. This would normally be a rather tedious exercise: The reason is that the vector field provides no obvious symmetry, requiring that we compute the surface integral through each of the six faces of the separately and then add them all up (try it yourself!). With the Divergence Theorem however, this becomes much more simple. It is not too hard to see that the cube is a regular region and $\mathbf{F}$ is a C^1 vector field, hence the Divergence Theorem applies and

$$\begin{aligned}
\iint_S \mathbf{F} \cdot \hat{\mathbf{n}} \; \mathrm{d}A &= \iiint_C \operatorname{div} \mathbf{F} \; \mathrm{d}A \\
&= \int_{-1}^{1} \int_{-1}^{1} \int_{0}^{2} \left[3y^2 + x \right] \mathrm{d}z \, \mathrm{d}y \, \mathrm{d}x \\
&= 2 \int_{-1}^{1} \left[y^3 + xy \right]_{y=-1}^{y=1} \\
&= 4 \int_{-1}^{1} [1 + x] \, \mathrm{d}x = 8.
\end{aligned}$$

■

Example 5.26

Determine the flux of

$$\mathbf{F}(x,y,z) = (xz\sin(yz) + x^3, \cos(yz), 3zy^2 - e^{x^2+y^2}),$$

through the capped paraboloid

$$S = \left\{x^2+y^2+z=4\right\} \cup \left\{x^2+y^2 \le 4, z=0\right\}.$$

Solution. This is an almost impossible exercise to approach from the definition, so instead we use the Divergence Theorem. One can easily compute that

$$\operatorname{div}\mathbf{F} = \left(z\sin(yz) + 3x^3\right) + (-z\sin(yz)) + (3y^2) = 3x^3 + 3y^2.$$

Hence if V is the filled paraboloid so that $\partial V = S$ then our surface integral becomes

$$\iint_S \mathbf{F}\cdot\hat{\mathbf{n}}\,\mathrm{d}A = \iiint_V (3x^2+3y^2)dV.$$

To determine this integral, notice that we can write

$$\iiint_V (3x^2+3y^2)dV = \iint_D \int_0^{4-x^2-y^2} (3x^2+3y^2)\,\mathrm{d}z\,\mathrm{d}A$$

where D is the unit disk. Changing to polar coordinates (or cylindrical if we skip the previous step) gives

$$\begin{aligned}\int_0^2\int_0^{2\pi}\int_0^{4-r^2} 3r^3\,\mathrm{d}z\,\mathrm{d}\theta\,\mathrm{d}r &= 6\pi\int_0^2 r^3(4-r^2)\,\mathrm{d}r\,\mathrm{d}\theta \\ &= 6\pi\left(16-\frac{64}{6}\right) = 32\pi.\end{aligned}$$

■

5.8 Stokes' Theorem

Stokes' Theorem, in another form, is the ultimate theorem from which Green's Theorem and the Divergence Theorem are derivative; albeit we will not likely see this form in this class. Hence we present to you, the "baby Stokes'" theorem. The idea of Stokes theorem is that we take a step back, and examine how one computes line integrals in $\mathbb{R}^3$ in general.

Unsurprisingly, the theorem says something about looking at the boundary of a set. Since we know that such integrations are dependent upon orientation, we need to define our final notion of positive/Stokes' orientation. Let S be a smooth surface in $\mathbb{R}^3$ with piecewise smooth boundary ∂S. We say that ∂S is *positively oriented* or *endowed with the Stokes' orientation* with respect to S if, whenever $\mathbf{t}$ is the tangent vector of a parameterization of ∂S and $\mathbf{n}$ is the orientation of S, then $\mathbf{n}\times\mathbf{t}$ points into S. More informally, if we walk around ∂S, our body aligned with $\mathbf{n}$, then S will be to the left.

Theorem 5.27: Stokes' Theorem

Let S be a smooth surface with piecewise smooth boundary ∂S, endowed with the Stokes' orientation. If $\mathbf{F} : \mathbb{R}^3 \to \mathbb{R}^3$ is a C^1- vector field in a neighbourhood of S, then

$$\int_{\partial S} \mathbf{F} \cdot \, \mathrm{d}\mathbf{x} = \iint_S (\operatorname{curl} \mathbf{F}) \cdot \hat{\mathbf{n}} \, \, \mathrm{d}A.$$

Proof. First note that if S is just a region in the xy-plane, then $\hat{\mathbf{n}} = (0, 0, 1)$ and so

$$(\operatorname{curl} \mathbf{F}) \cdot \hat{\mathbf{n}} = \frac{\partial F_2}{\partial x_1} - \frac{\partial F_1}{\partial x_2}$$

and hence Green's Theorem gives

$$\int_{\partial S} \mathbf{F} \cdot \, \mathrm{d}\mathbf{x} = \iint_S \left[\frac{\partial F_2}{\partial x_1} - \frac{\partial F_1}{\partial x_2} \right] \mathrm{d}A.$$

Thus we see that Stokes' theorem in the xy-plane is just Green's theorem.

Now assume that S is a surface which does not live in one of the coordinate planes and let $G : W \to S$ be a parameterization of S, where the region W lives in the uv-plane. Furthermore, assume that G preserves boundaries and gives an orientation which coincides with the orientation of S (if $G(u, v)$ gives the opposite orientation, just switch the roles of u and v).

The idea is that since the boundaries are preserved under G and since Stokes' theorem is just Green's theorem, we will "pullback" the calculation to the uv-plane and apply Green's Theorem. As always, we shall do this component by component; in particular, we shall just look at the F_1 component. In effect, take $\mathbf{F} = (F_1, 0, 0)$ so that this amounts to showing

$$\int_{\partial S} F_1 \, \mathrm{d}x_1 = \iint_S \left(0, \frac{\partial F_1}{\partial x_3}, -\frac{\partial F_1}{\partial x_2} \right) \cdot \hat{\mathbf{n}} \, \, \mathrm{d}A.$$

Applying our parameterization, the right-hand side becomes

$$\begin{aligned} \iint_S \left(0, \frac{\partial F_1}{\partial x_3}, -\frac{\partial F_1}{\partial x_2} \right) \cdot \hat{\mathbf{n}} \, \, \mathrm{d}A &= \iint_W \left(0, \frac{\partial F_1}{\partial x_3}, -\frac{\partial F_1}{\partial x_2} \right) \cdot \left(\frac{\partial G}{\partial u} \times \frac{\partial G}{\partial v} \right) \mathrm{d}u \, \mathrm{d}v \\ &= \iint_W \left(\frac{\partial F_1}{\partial x_3} \frac{\partial (z, x)}{\partial (u, v)} - \frac{\partial F_1}{\partial x_2} \frac{\partial (x, y)}{\partial (u, v)} \right) \mathrm{d}u \, \mathrm{d}v. \end{aligned}$$

On the other hand, using the Chain rule and Green's Theorem, the left-hand-side yields

$$\begin{aligned} \int_{\partial W} F_1 \left(\frac{\partial x}{\partial u} \, \mathrm{d}u + \frac{\partial x}{\partial v} \, \mathrm{d}v \right) &= \iint_W \left(\frac{\partial}{\partial u} \left[F_1 \frac{\partial x}{\partial v} \right] - \frac{\partial}{\partial v} \left[F_1 \frac{\partial x}{\partial u} \right] \right) \mathrm{d}u \, \mathrm{d}v \\ &= \iint_W \left(\frac{\partial F_1}{\partial x_3} \frac{\partial (z, x)}{\partial (u, v)} - \frac{\partial F_1}{\partial x_2} \frac{\partial (x, y)}{\partial (u, v)} \right) \mathrm{d}u \, \mathrm{d}v. \end{aligned}$$

which is exactly what we had above, giving the desired equality. □

Example 5.28

Let C be the curve given by the intersection of $z = x$ and $x^2 + y^2 = 1$, oriented counter clockwise when examined from $(0, 0, 1)$, with S such that $\partial S = C$. Let $\mathbf{F}(x, y, z) = (x, z, 2y)$. Compute both

$$\oint_C \mathbf{F} \cdot \mathrm{d}\mathbf{x}, \qquad \iint_S (\operatorname{curl} \mathbf{F}) \cdot \hat{\mathbf{n}} \ \mathrm{d}A.$$

Solution. We can parameterize C as

$$\gamma(\theta) = (\cos(\theta), \sin(\theta), \cos(\theta)), \qquad 0 \leq \theta \leq 2\pi$$

so that

$$\begin{aligned}\oint_C \mathbf{F} \cdot \mathrm{d}\mathbf{r} &= \int_0^{2\pi} (\cos(\theta), \cos(\theta), 2\sin(\theta)) \cdot (-\sin(\theta), \cos(\theta), -\sin(\theta)) \, \mathrm{d}\theta \\ &= \int_0^{2\pi} -\cos(\theta)\sin(\theta) + \cos^2(\theta) - 2\sin^2(\theta) \, \mathrm{d}\theta \\ &= 0 + \pi - 2\pi = -2\pi.\end{aligned}$$

On the other hand, the curl of $\mathbf{F}$ is easily computed to be

$$\operatorname{curl} F = \det \begin{pmatrix} \mathbf{e}_1 & \mathbf{e}_2 & \mathbf{e}_3 \\ \partial_x & \partial_y & \partial_z \\ x & z & 2y \end{pmatrix} = (1, 0, 0).$$

We can parameterize our surface is almost exactly the same way as the curve (though now we let our radius vary) as

$$g(r, \theta) = (r\cos(\theta), r\sin(\theta), r\cos(\theta)), \qquad 0 \leq r \leq 1, 0 \leq \theta \leq 2\pi.$$

Hence

$$\frac{\partial g}{\partial r} = (\cos(\theta), \sin(\theta), \cos(\theta)), \qquad \frac{\partial g}{\partial \theta} = (-r\sin(\theta), r\cos(\theta), -r\sin(\theta 0)$$

giving an area element of

$$\frac{\partial g}{\partial r} \times \frac{\partial g}{\partial \theta} = (-r, 0, -r).$$

Integrating gives

$$\begin{aligned}\int_0^{2\pi} \int_0^{\sqrt{2}} (1, 0, 0) \cdot (-r, 0, r) \, \mathrm{d}r \, \mathrm{d}\theta &= \int_0^{2\pi} \int_0^{\sqrt{2}} -r \, \mathrm{d}r \, \mathrm{d}\theta \\ &= 2\pi \left[-\frac{1}{2} r^2 \right]_{r=0}^{\sqrt{2}} \\ &= -\pi.\end{aligned}$$

■

Example 5.29

Let $S = \{(x, y, z) : x^2 + y^2 + z^2 = 1, z \geq 0\}$. If ∂S is oriented counter clockwise when viewed from $(0, 0, 1)$, and

$$\mathbf{F}(x, y, z) = \left(xy + xe^z, \frac{1}{6}\left(2x^3 + 3x^2 + y^2 z\right), \sqrt{1 + x^2 + zy}\right),$$

compute $\oint_C \mathbf{F} \cdot \mathrm{d}\mathbf{x}$.

Solution. It is clear that ∂S is just the unit circle in the xy-plane, and so we can parameterize it as $g(t) = (\cos(t), \sin(t))$ for $t \in [0, 2\pi]$; however, it makes this integral almost impossible to compute directly. Our goal is thus to use Stokes theorem, so we compute the curl to be

$$\begin{aligned}\nabla \times \mathbf{F} &= \left(\frac{\partial F_3}{\partial y} - \frac{\partial F_2}{\partial z}, \frac{\partial F_1}{\partial z} - \frac{\partial F_3}{\partial x}, \frac{\partial F_2}{\partial x} - \frac{\partial F_1}{\partial y}\right) \\ &= \left(\frac{z}{2\sqrt{1 + x^2 + zy}} - \frac{zy}{3}, xe^z - \frac{x}{\sqrt{1 + x^2 + zy}}, x^2\right).\end{aligned}$$

Unfortunately, the unit normal on S is constantly changing and the integral is still rather horrific. However, one of the beautiful things about Stokes theorem is that it tells us is that the line integral over C can be computed in terms of an integral over S, but it does not say *which* S that has to be. In particular, if there is a more convenient S to choose, we should take it!

We notice then that C is just the boundary of the unit disk $S' = \{(x, y, 0) : x^2 + y^2 = 1\}$, and the corresponding orientation on S which yields the counterclockwise orientation on C is $\hat{\mathbf{n}} = (0, 0, 1)$. Hence our integral simply becomes

$$\oint_C \mathbf{F} \cdot \mathrm{d}\mathbf{x} = \iint_{S'} (\operatorname{curl} \mathbf{F}) \cdot \hat{\mathbf{n}} dS = \iint_{S'} x^2 dS$$

This integral is much easier done. Converting to polar coordinates, we get

$$\int_{S'} x^2 dS = \left[\int_0^1 r^3 \, \mathrm{d}r\right]\left[\int_0^{2\pi} \sin^2(\theta) \, \mathrm{d}\theta\right] = \frac{\pi}{4}.$$

■

5.9 Tensor Products

5.9.1 The Definition

We know that, given two vectors in the same space, there isn't a very meaningful way of multiplying them together to get a vector back. Sure, one can perform pointwise multiplication on the components, but the object that is returned is not useful for studying vector spaces. Furthermore, what happens if we want to multiply two vectors which are from different vector spaces?

We are faced with the following challenge: Given two $\mathbb{R}$ vector spaces V and W, is there a meaningful way to 'multiply' them together? What is meaningful? Our motivation is the following two examples:

1. To devise a method for describing product states. This is especially useful in statistic and quantum mechanics, as will be described later.

2. To approximate or describe multilinear objects via linear objects, in the most efficient way possible. This is the reason of greatest interest to mathematicians, and will be our principal motivation.

Again, the important property here is the idea of multilinearity, which we define below:

Definition 5.30

Let $V_1, \dots, V_n$ and W be vector spaces and $T : V_1 \times \cdots \times V_n \to W$ be a map. We say that T is *multilinear* if for each $i \in \{1, \dots, n\}$, the map T is linear in the i-th component when all other components are held constant.

Remark 5.31 Consider the map $f : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ given by $f(x, y) = xy$. This map is multilinear since

$$f(x_1 + x_2, y) = (x_1 + x_2)y = x_1y + x_2y = f(x_1, y) + f(x_2, y),$$

and similarly $f(x, y_1 + y_2) = f(x, y_1) + f(x, y_2)$. However, f is not a linear map of vector spaces, since

$$\begin{aligned} f\Big((x_1, y_1) + (x_2, y_2)\Big) &= f(x_1 + x_2, y_1 + y_2) \\ &= (x_1 + x_2)(y_1 + y_2) = x_1y_1 + x_1y_2 + y_1x_2 + y_2x_2 \\ &= [f(x_1, y_1) + f(x_2, y_2)] + [f(x_1, y_2) + f(x_2, y_1)]\,. \end{aligned}$$

▲

There are lots of interesting multilinear maps that appear in linear algebra, but the failure of their linearity means they cannot be properly studied within the realm of linear algebra (where only linear maps are permitted). For example, the determinant map is multilinear: If $\dim V = n$ then $\det : \underbrace{V \times \cdots \times V}_{n\text{-times}} \to \mathbb{R}$ is a multilinear map. The student can check that the following are also multilinear (but not linear) maps:

- The cross product $\cdot \times \cdot : \mathbb{R}^3 \times \mathbb{R}^3 \to \mathbb{R}^3$ is also a multilinear map,
- The dot product $\langle \cdot, \cdot \rangle : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$.

The properties that we would like our product to satisfy should be natural, in the sense that it should perform very much like a product, and in particular if we temporarily write the product of $v \in V$ and $w \in W$ as (v, w), then it should satisfy

1. $(v_1, w) + (v_2, w) = (v_1 + v_2, w)$,
2. $(v, w_1) + (v, w_2) = (v, w_1 + w_2)$,

3. $c(v, w) = (cv, w) = (v, cw), c \in \mathbb{R}$.

In order to ensure that these things happen, we will effectively force them to happen.

Definition 5.32

Given a set S, we define the *free vector space on* S to be a vector space $F(S)$ such that S is a basis for $F(S)$.

It turns out that free vector spaces are unique, up to an invertible linear map (an isomorphism), and this is easily determined since vector spaces are uniquely defined by the cardinal of their dimension.

Example 5.33 f $S = \{v_1, v_2, v_3\}$ then $F(S)$ is the (real) vector space with S as a basis. In particular, the elements of S are linearly independent and span $F(S)$, so every vector $v \in F(S)$ can be written uniquely as

$$v = c_1 v_1 + c_2 v_2 + c_3 v_3$$

for some $c_i \in \mathbb{R}$. If $e_i \in \mathbb{R}^3$ is the standard basis vectors for $\mathbb{R}^3$, then $T : F(S) \to \mathbb{R}^3$ given by $T(v_i) = e_i$ is an invertible linear map, so that $F(S) \cong \mathbb{R}^3$.

To ensure that our desired properties happen, we will construct a vector space with these properties. Consider the space $F(V \times W)$, which is the free vector space whose basis is given by all the elements of $V \times W$. Note that this is a *very* large vector space: if one of V and W is not the trivial vector space, then $F(V \times W)$ is an infinite dimensional vector space.

Next, we will consider the subspace $S \subseteq F(V \times W)$ generated by the following elements:

$$(v_1, w) + (v_2, w) - (v_1 + v_2, w), \qquad (v, w_1) + (v, w_2) - (v, w_1 + w_2),$$

$$c(v, w) - (cv, w), \qquad c(v, w) - (v, cw).$$

We define the *tensor product* of V and W, denoted $V \otimes W$, to be $F(V \times W)/S$.

Proposition 5.34

If V, W are real vector spaces and $\dim V = n$, $\dim W = m$ then the following facts hold true in the vector space $V \otimes W$:

1. Properties (1)-(3) all hold,
2. $V \otimes W$ is a finite dimensional vector space with $\dim(V \otimes W) = mn$,
3. If $\{v_1, \ldots, v_n\}$ is a basis of V and $\{w_1, \ldots, w_m\}$ is a basis for W then
$$\{v_i \otimes w_j : i \in \{1, \ldots, n\}, j \in \{1, \ldots, m\}\}$$
is a basis for $V \otimes W$.
4. If $f : V_1 \times \cdots \times V_n \to W$ is a multilinear map, then there exists a unique *linear* map $F : V_1 \otimes \cdots \otimes V_n \to W$ such that $F(v_1 \otimes \cdots \otimes v_n) = f(v_1, \ldots, v_n)$.

This proposition is fairly involved, so we will omit its proof. However, note that property (4) in particular tells us that we can use the tensor product to turn multilinear maps into linear maps over a different vector space, and hence study those maps using the tools of linear algebra. In fact, the correspondence in (4) is bijective, so we will sometimes not distinguish between f and F

Dual Spaces: If V is a vector space, we define its dual vector space, denoted V^* as
$$V^* = \{f : V \to \mathbb{R} : f \text{ is linear}\}.$$
The student can check that this is a vector space, and if $\dim V = n$ then $\dim V^* = n$. Furthermore, there is a canonical isomorphism $\phi : V \to (V^*)^*$ defined by $\phi(v)f = f(v)$.

Let $\{e_i\}$ be a basis for V. We define the *dual basis* $\{f_i\}$ of V^* to be the basis which satisfies the condition
$$f_i(e_j) = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}.$$
Now if $f : V^k \to \mathbb{R}$ is a multilinear map, then the corresponding linear function on the tensor product space (which we will also denote by f, is a linear map $f : V^{\otimes k} \to \mathbb{R}$; that is, $f \in (V^{\otimes k})^* \cong (V^*)^{\otimes k}$. Since $\{f_i\}$ is a basis for V^*, $\{f_{i_1} \otimes \cdots \otimes f_{i_k}\}$ is a basis for $(V^*)^{\otimes k}$ and hence we can write
$$f = \sum_{(i_1, \ldots, i_k)} c_{i_1, \ldots, i_k} f_{i_1} \otimes \cdots \otimes f_{i_k}.$$
For notation sake, this is rather clumsy. Recall that in the discussion of Taylor Series, we learned about multi-indices. If $I = (i_1, \ldots, i_k)$ we will denote by $f_I = f_{i_1} \otimes \cdots \otimes f_{i_k}$, and we can rewrite the above as
$$f = \sum_I c_I f_I$$

We can go one step further, and say that if $f : V_1 \to W_1$ and $g : V_2 \to W_2$ are both linear maps, there is an induced map $f \otimes g : V_1 \otimes V_2 \to W_1 \otimes W_2$ given by
$$(f \otimes g)(v_1 \otimes v_2) = f(v_1) \otimes g(v_2).$$

If $\{f_i\}$ and $\{h_i\}$ are dual bases for V_1^* and V_2^* respectively, and we write f and g in terms of their dual bases

$$f = \sum_i c_i f_i, \qquad g = \sum_j d_j h_j$$

then their product is $f \otimes j = \sum_{ij} c_i d_j f_i \otimes h_j$.

Exercise:

1. Check that $(V^*)^k \cong (V^k)^*$ and $(V^*)^{\otimes k} \cong (V^{\otimes k})^*$.
2. Let $\mathrm{Hom}(V, W) = \{f : V \to W, f \text{ is linear}\}$. Show that $\mathrm{Hom}(V, W) \cong V^* \otimes W$.

5.9.2 (Anti-)Symmetric Maps

So tensor products give us a means of studying multilinear maps using linear tools, so long as we are willing to modify our vector space. There are two very important types of multilinear maps in which one is typically interested: Let $f : V \times \cdots \times V \to W$ be a multilinear map.

1. We say that f is *symmetric* if for any $i < j$ we have

$$f(v_1, \ldots, v_j, \ldots, v_i, \ldots, v_n) = f(v_1, \ldots, v_i, \ldots, v_j, \ldots, v_n).$$

2. We say that f is *anti-symmetric* if for any $i < j$ we have

$$f(v_1, \ldots, v_j, \ldots, v_i, \ldots, v_n) = -f(v_1, \ldots, v_i, \ldots, v_j, \ldots, v_n).$$

Symmetric tensors often arise in the study of inner products or hermitian products, since those maps are symmetric multilinear. However, this is not the goal of our discussion, so we will not spend much time thinking symmetric maps. Instead, we will be more interested in anti-symmetric maps.

Proposition 5.35

Let V and W be a finite dimensional vector space with $\dim V = n$.

1. If $k \leq n$ and $f : V^k \to W$ is an anti-symmetric map, then if $\{v_1, \ldots, v_k\}$ is linearly dependent, necessarily
$$f(v_1, \ldots, v_k) = 0.$$
2. If $k > n$ then there are no anti-symmetric maps $f : V^k \to W$.

This proposition is not too difficult and its proof is left as an exercise for the student. It can be shown that the collection of k-multilinear alternating maps is a vector subspace of the space of k-multilinear maps, and as such we will denote this set by $\Lambda^k(V)$. To determine the dimension of this subspace, we need to introduce a basis:

Let V be a vector space with basis $\{e_i\}$ and let $\{f_i\}$ be a dual basis for V^*. If $I = (i_1, \ldots, i_k)$ is a multi-index, define the map $f^I : V^{\otimes k} \to \mathbb{R}$ by

$$f^I(v_1, \ldots, v_k) = \det \begin{pmatrix} f_1(v_1) & f_1(v_2) & \cdots & f_1(v_k) \\ f_2(v_1) & f_2(v_2) & \cdots & f_2(v_k) \\ \vdots & \vdots & \ddots & \vdots \\ f_k(v_1) & f_k(v_2) & \cdots & f_k(v_k) \end{pmatrix}.$$

Proposition 5.36

If V is an n-dimensional vector space with dual basis $\{f_i\}$ for V^*, the set

$$\left\{ f^{(i_1,\ldots,i_k)} : i_1 < i_2 < \ldots < i_k \right\}$$

is a basis $\Lambda^k(V)$. Consequently, $\dim \Lambda^k(V) = \binom{n}{k}$.

We define the *wedge product* as the following map on the f^I defined above

$$f^I \wedge f^J = f^{IJ},$$

and extend linearly.

Proposition 5.37

The wedge product satisfies the following properties:

1. **Anti-symmetry:** If $v \in \Lambda^k(V)$ and $w \in \Lambda^\ell(V)$ then $v \wedge w \in \Lambda^{k+\ell}(V)$ and
$$v \wedge w = (-1)^{k\ell} w \wedge v.$$
2. **Linearity:** $(cv_1 + v_2) \wedge w = c\,(v_1 \wedge w) + (v_2 \wedge w)$
3. **Associativity** $(u \wedge v) \wedge w = u \wedge (v \wedge w)$,
4. If $I = (i_1, \ldots, i_k)$ then $f^I = f^{i_1} \wedge \cdots \wedge f^{i_k}$.

5.9.3 Differential Forms

Okay, that is enough about tensors in general. It is now time to look at differential forms and how they are defined. Let S be a n-manifold, and for each $p \in S$ let V_p be the tangent space at p. Choose a basis $\{v_1^p, \ldots, v_n^p\}$ be a basis for this tangent space, and $\{dx_1^p, \ldots, dx_n^p\}$ be a basis of its dual space V_p^*. A differential k-form is a C^1 function $S \to \bigsqcup_{p \in S} \Lambda^k(V_p)$; that is, a function which assigns to every point $p \in S$ an element of the dual space of the tangent space at p. The collection of differential k-forms on S is denoted $\Omega^k(S)$.

Let us consider the case when S is a 3-manifold.

- The 0-forms on S are just the C^1-functions $S \to \mathbb{R}$.

- The 1-forms are functions which look $p \mapsto f(p)dx_1^p + g(p)dx_2^p + h(p)dx_3^p$. We will often drop the p dependence and just write $fdx_1 + gdx_2 + hdx_3$.
- The 2 forms look like $fdx_1 \wedge dx_2 + gdx_1 \wedge dx_3 + hdx_2 \wedge dx_3$.
- The 3 forms look like $fdx_1 \wedge dx_2 \wedge dx_3$.

Exterior Derivative: The exterior derivative is a map $d : \Omega^k(S) \to \Omega^{k+1}(S)$ defined as follows: If $f : S \to \mathbb{R}$ is a function, then

$$df = \sum_{k=1}^{n} \frac{\partial f}{\partial x_k} dx_k$$

is the usual differential of f. If $\omega = fdx_1 + gdx_2 + hdx_3$ is a 1-form, then we define

$$d\omega = df \wedge dx_1 + dg \wedge dx_2 + dh \wedge dx_3.$$

In general, if $\omega = \sum_I f_I dx_I$ is a differential k-form, then $d\omega = \sum_I df_I \wedge dx_I$.

Relation to Vector Fields: In $\mathbb{R}^3$ there are ways to realize differential forms as vector fields.

- In the case of 0-forms, there is nothing to do.
- Identify the 1-form $\omega = F_1dx + F_2dy + F_3dz$ with the vector field $\mathbf{F} = (F_1, F_2, F_3)$.
- Identify the 2-form $\omega = F_1dy \wedge dz + F_2dx \wedge dz + F_3dx \wedge dy$ with the vector field $\mathbf{F} = (F_1, F_2, F_3)$.
- Identify the 3-form $\omega = fdx \wedge dy \wedge dz$ with the function $f : S \to \mathbb{R}$.

These identifications allow us to realize the exterior derivative as our vector derivatives gradient, curl, and divergence. Indeed, if $f : S \to \mathbb{R}$ is a 0-form/function, then

$$df = \frac{\partial f}{\partial x}dx + \frac{\partial f}{\partial y}dy + \frac{\partial f}{\partial z}dz \sim \left(\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, \frac{\partial f}{\partial z}\right) = \nabla f.$$

If $\omega = F_1dx + F_2dy + F_3dz \sim \mathbf{F} = (F_1, F_2, F_3)$ then

$$\begin{aligned}
d\omega &= dF_1 \wedge dx + dF_2 \wedge dy + dF_3 \wedge dz \\
&= \left(\frac{\partial F_1}{\partial x}dx + \frac{\partial F_1}{\partial y}dy + \frac{\partial F_1}{\partial z}dz\right) \wedge dx + \left(\frac{\partial F_2}{\partial x}dx + \frac{\partial F_2}{\partial y}dy + \frac{\partial F_2}{\partial z}dz\right) \wedge dy \\
&\qquad + \left(\frac{\partial F_3}{\partial x}dx + \frac{\partial F_3}{\partial y}dy + \frac{\partial F_3}{\partial z}dz\right) \wedge dz \\
&= \left(\frac{\partial F_3}{\partial y} - \frac{\partial F_2}{\partial z}\right) dy \wedge dz + \left(\frac{\partial F_3}{\partial x} - \frac{\partial F_1}{\partial z}\right) dx \wedge dz + \left(\frac{\partial F_2}{\partial x} - \frac{\partial F_1}{\partial y}\right) dx \wedge dy \\
&\sim \left(\frac{\partial F_3}{\partial y} - \frac{\partial F_2}{\partial z}, \frac{\partial F_3}{\partial x} - \frac{\partial F_1}{\partial z}, \frac{\partial F_2}{\partial x} - \frac{\partial F_1}{\partial y}\right) = \operatorname{curl} \mathbf{F}.
\end{aligned}$$

Finally, if $\omega = F_1 dy \wedge dz + F_2 dx \wedge dy + F_3 dx \wedge dz \sim \mathbf{F} = (F_1, F_2, F_3)$ then

$$\begin{aligned} d\omega &= dF_1 \wedge dy \wedge dz + dF_2 \wedge dx \wedge dz + dF_3 \wedge dx \wedge dz \\ &= \left(\frac{\partial F_1}{\partial x}dx + \frac{\partial F_1}{\partial y}dy + \frac{\partial F_1}{\partial z}dz\right) \wedge dy \wedge dz + \left(\frac{\partial F_2}{\partial x}dx + \frac{\partial F_2}{\partial y}dy + \frac{\partial F_2}{\partial z}dz\right) \wedge dx \wedge dz \\ &\qquad + \left(\frac{\partial F_3}{\partial x}dx + \frac{\partial F_3}{\partial y}dy + \frac{\partial F_3}{\partial z}dz\right) \wedge dx \wedge dy \\ &= \left(\frac{\partial F_1}{\partial x} + \frac{\partial F_2}{\partial y} + \frac{\partial F_3}{\partial z}\right) dx \wedge dy \wedge dz \\ &\sim \frac{\partial F_1}{\partial x} + \frac{\partial F_2}{\partial y} + \frac{\partial F_3}{\partial z} = \operatorname{div} \mathbf{F} \end{aligned}$$

Interestingly, one can show that $d \circ d = 0$ regardless of the dimension of the manifold and the forms to which it is being applied.

Pullbacks: Let $F : S \to T$ be a function between manifolds, and let $\{dx_1, \ldots, dx_n\}$ be differential forms on T. One can define the *pullback* of a differential form on T to be the differential form on S given by

$$F^*(f dx_1 \wedge \cdots \wedge dx_n) = (f \circ F) d(x_1 \circ F) \wedge \cdots \wedge d(x_n \circ F)$$

where $x_i \circ F = F_i$ is the ith component of the function F. For example, let $S = \times[0,1] \times [0, 2\pi]$ and $T = D_1$ where D_1 is the unit disk. Define the map $F : S \to T$ by $F(r, \theta) = (r\cos(\theta), r\sin(\theta))$. The pullback of the form $dx \wedge dy$ is then given by

$$\begin{aligned} F^*(f(x,y)dx \wedge dy) &= d(x_1 \circ F) \wedge d(x_2 \circ F) = dF_1 \wedge dF_2 \\ &= f(F(r,\theta))(\cos(\theta)dr - r\sin(\theta)d\theta) \wedge (\sin(\theta)dr + r\cos(\theta)d\theta) \\ &= f(F(r,\theta)) - r\sin^2(\theta)d\theta \wedge dr + r\cos^2(\theta)dr \wedge d\theta - \\ &= f(F(r,\theta))r(\sin^2(\theta) + \cos^2(\theta))dr \wedge d\theta \\ &= f(F(r,\theta))r\, dr \wedge d\theta. \end{aligned}$$

In fact, if we think carefully about how differential forms are defined, then if $F : S \to T$ is a diffeomorphism with $\{x_1, \ldots, x_n\}$ the coordinates of S and $\{y_1, \ldots, y_n\}$, then

$$F^*(f dy_1 \wedge \cdots \wedge dy_n) = (f \circ F) \det\left(\frac{\partial F_i}{\partial x_j}\right) dx_1 \wedge \cdots \wedge dx_n.$$

Effectively, this allows us to write the Change of Variable Theorem as follows:

Theorem 5.38: Change of Variables

If $F : S \to T$ is a diffeomorphism between S and T, then

$$\int_S F^*(\omega) = \int_T \omega.$$

Note however that this version of the Change of Variables Theorem does keep track of orientation, so it is not quite identical to Theorem 4.38.

Stokes' Theorem: The power of differential forms is that it allows us to generalize Stokes' theorem to higher dimensions, and to see in fact that Green's Theorem, the Divergence Theorem, and Baby Stokes' Theorem, are all equivalent. As with all the aforementioned cases, one needs to talk about some suitable notion of the orientation of the boundary with respect to the thing which it bounds. As a general rule, we orient the boundary in a manner that points inwards.

Theorem 5.39: Stokes' Theorem

Let M be an n-dimensional C^1 manifold with boundary ∂M, oriented in a compatible way. If $\omega \in \Omega^n(M)$ then

$$\int_{\partial M} \omega = \int_M d\omega.$$

5.9.4 Closed and Exact Forms

Our notions of closed and exact conservative vector fields now extends to the context of differential forms.

Definition 5.40

Let ω be an n-form in[a] $\mathbb{R}^k$.

1. We say that ω is *exact* if there exists an $(n-1)$-form η such that $d\eta = \omega$. We denote the exact k-forms on S as $B^k(S)$.
2. We say that ω is *closed* if $d\omega = 0$. We denote the closed k-forms on S as $Z^k(S)$.

[a] The B stands for boundary, since there is a sense in which $d\eta$ is the boundary of η. The Z stands for Zyklen, which is the German word for cycle.

It was mentioned before that $d \circ d = 0$. This means that all exact forms are closed, since if $\omega = d\eta$ then $d\omega = d(d\eta) = 0$. In particular, this means that $B^k(S) \subseteq Z^k(S)$. In general, the converse is not true. For example, the one form

$$\frac{x}{x^2+y^2}dx - \frac{y}{x^2+y^2}dy \in \Omega^1(\mathbb{R}^2 \setminus \{0\})$$

is closed but not exact. As in the case of conservative vector fields, the problem is somehow captured by the presence of the hole at the origin. If our space does not have holes, then all closed vector fields will be exact, as exemplified by the following generalized version of Poincaré's Lemma:

Theorem 5.41: Poincaré's Lemma

If $S \subseteq \mathbb{R}^n$ is a star shaped set and $\omega \in \Omega^k(S)$ is an closed k-form, then ω is exact.

Finally, one can precisely measure the failure of closed forms from being exact, by computing the *de Rham cohomology* of a space. Let S be a smooth k-manifold, and recognize that we have the following "chain" of vector spaces

$$0 \xrightarrow{d} \Omega^0(S) \xrightarrow{d} \Omega^1(S) \xrightarrow{d} \cdots \xrightarrow{d} \Omega^{k-1}(S) \xrightarrow{d} \Omega^k(S) \xrightarrow{d} 0$$

The image of d is the closed forms, and the kernel of d is the exact forms, so one defines the k-th *de Rham cohomology* as

$$H^k(S) = \frac{\ker\left(d : \Omega^k(S) \to \Omega^{k+1}(S)\right)}{\operatorname{im}\left(d : \Omega^{k-1}(S) \to \Omega^k(S)\right)} = \frac{Z^k(S)}{B^k(S)}.$$