MAT257
Course Notes

Tyler Holden
Mathematics and Computational Sciences
University of Toronto Mississauga
tyler.holden@utoronto.ca

Contents

1 Review

We'll begin with a brief review of concepts which will be important in the following section. This is not intended to be a thorough review, so you should ensure that you fill the gaps in your knowledge by referencing an appropriate text. I will assume you are familiar with sets and set builder notation $S = \{x : P(x)\}$. The arithmetic hierarchy will be denoted as follows:

- The **naturals** $\mathbb{N} = \{0, 1, 2, 3, \ldots\}$,
- The **integers** $\mathbb{Z} = \{..., -2, -1, 0, 1, 2, ...\}$,
- The **rationals** $\mathbb{Q} = \{p/q : p, q \in \mathbb{Z}, q \neq 0, \gcd(p, q) = 1\}$,
- The **reals** $\mathbb{R}$ (choose your favourite construction).

1.1 Quantifiers

Quantifiers allow us to discuss the number of objects which satisfy a predicate. If we wish to discuss *every* element of a set, we use the *universal quantifier* $\forall$, read as "for all." To state that an element in a set *exists*, we use the *existential quantifier* $\exists$, read as "there exists."

When combined with a predicate P, we can assign truth values to quantified statements. For example, let S be an universe of discourse. The statement $\forall x \in S, P(x)$ will be true precisely when $P(x)$ is true for every element in S. On the other hand, $\exists x \in S, P(x)$ will be true as long as a single element of S makes $P(x)$ true.

The addition of quantifiers allows us to make statements such as the following:

- Every cow has a favourite radio station.
- There is a black horse.
- In every sport, someone breaks the rules.
- There is one textbook used in every class.

These last two examples have multiple quantifiers. Can you spot them?

Example 1.1

Determine whether each of the quantified statements is true or false.

1. $\forall x \in \mathbb{N}, x^2 \geq 0$
2. $\exists x \in \mathbb{R}, x = \sqrt{-1}$,
3. $\forall x \in \mathbb{Q}, \forall y \in \mathbb{Q}, x + y \in \mathbb{Q}$,
4. $\exists x \in \mathbb{N}, \exists y \in \mathbb{N}, x/y \in \mathbb{N}$.

Solution.

1. This statement is true, since squaring any number results in a non-negative number.

2. This statement is false. If such an x existed, it would also satisfy $x^2 = -1$. By our comment in part 1, the square of a non-zero number is always positive, leading us to a contradiction.

3. This statement is true. Write $x = a/b$ and $y = c/d$ so that
$$x + y = \frac{a}{b} + \frac{c}{d} = \frac{ad + bc}{bd}.$$
Since $ad + bc \in \mathbb{Z}$ and $bd \in \mathbb{N}$, this is also a rational number.

4. This statement is true. For example, by setting $x = 4$ and $y = 2$ we have $x/y = 4/2 = 2$, which is also a natural number. ■

Notice how the above solutions demonstrated the truth of quantifier statements. To show that $\exists x \in S, P(x)$, we find a single example of an $x \in S$ which makes $P(x)$ true. To show that $\forall x \in S, P(x)$ is more subtle. Rather than try to demonstrate $P(x)$ for every x, we choose an *arbitrary* $x \in S$. If $P(x)$ is true for an arbitrary x, then it must be true for every x.

Doubly quantified statements must be treated with caution. You may freely interchange two adjacent quantifiers of the *same* type, but not of different type. For example, the statements
$$\forall x \in \mathbb{Q}, \forall y \in \mathbb{Q}, x + y \in \mathbb{Q} \quad \text{is logically equivalent to} \quad \forall y \in \mathbb{Q}, \forall x \in \mathbb{Q}, x + y \in \mathbb{Q},$$
and
$$\exists x \in \mathbb{N}, \exists y \in \mathbb{N}, x/y \in \mathbb{N} \quad \text{is logically equivalent to} \quad \exists y \in \mathbb{N}, \exists x \in \mathbb{N}, x/y \in \mathbb{N}.$$
However, interchanging existential and universal quantifiers can lead to trouble.

Example 1.2

Consider the statements
$$\forall x \in \mathbb{R}, \exists y \in \mathbb{R},\ x + y = 0 \tag{1.1}$$
and
$$\exists x \in \mathbb{R}, \forall y \in \mathbb{R},\ x + y = 0. \tag{1.2}$$
Compare these expressions by translating them as follows:

1. Convert the mathematical notation into English.
2. Turn the sentence derived above into a simple sentence, which does not involve any variables.
3. Evaluate whether each statement is true or false.

Solution. We start with equation (1.1), for which a direct translation into English is

"For all x in the real numbers, there exists y in the real numbers, (such that) $x + y = 0$."

This is fine but not very enlightening. By recognizing that $x + y = 0$ is equivalent to $x = -y$, we could re-interpret this sentence as saying "For every real number there is another real number which is its negative." Dropping the superfluous words we arrive at the intuitive statement

"Every real number has a negative."

This statement is certainly true: Given an integer a, we can always construct its negative $-a$.

Looking at (1.2) we have

$$\exists y \in \mathbb{R}, \forall x \in \mathbb{R}, \; x + y = 0.$$

Using the same translation process as above, the corresponding simple sentence is given by

"There is an element which is the negative of every real number."

This says there is a number to which we can add any other number and always get zero. Certainly this is not true! If it were, then there would be a number n such that $n + a = 0$ and $n + b = 0$ for any real numbers a and b. Equating these expressions, we would find that $n + a = n + b$ which in turn implies that $a = b$. This would force all integers to be equal, which is nonsense. ■

Example 1.2 teaches us that changing the order of the quantifiers significantly changes the logical statement, and hence the truth of that statement. To borrow a term from the computer scientists, universal quantifiers admit a 'scope' to the existential quantifiers they precede. For example, the statement $\forall x, \exists y, P(x, y)$ means that the choice of y is allowed to depend upon x. The statement $\exists y, \forall x, P(x, y)$ does not confer this dependence: the choice of y must work for every x.

Example 1.3

Let S be the set of all students in a classroom, and $B(a, b)$ be the statement "student a has the same birthday as student b." Write the mathematical statements

$$\forall a \in S, \forall b \in S, B(a, b), \qquad \forall a \in S, \exists b \in S, B(a, b)$$

$$\exists a \in S, \forall b \in S, B(a, b), \qquad \exists a \in S, \exists b \in S, B(a, b)$$

in plain language.

Solution. We may interpret each of the statements as follows:

$\forall a \in S, \forall b \in S, B(a, b)$ Every student (a) in the classroom has the same birthday as every other student (b) in the classroom

$\forall a \in S, \exists b \in S, B(a, b)$ For each student (a) in the classroom, there exists some other student (b) in the classroom with the same birthday.

$\exists a \in S, \forall b \in S, B(a, b)$ There exists a student (a) who has the same birthday as every other student (b) in the classroom.

$\exists a \in S, \exists b \in S, B(a, b)$ There exists a student (a) in the classroom who has the same birthday as another student (b) in the classroom.

■

Quantifiers have an odd interaction with the empty set $\emptyset$. In particular, for any predicate P the statement $\forall x \in \emptyset, P(x)$ is true vacuously, while $\exists x \in \emptyset, P(x)$ is false.

1.1.1 Negating Quantifiers

To develop intuition for negating quantifiers, let's think about how we would disprove a statement involving a quantifier. For example, the universally quantified statement "every horse is black" may be disproved by showing that there exists a non-black horse. Mathematically, if $P(x)$ is "x is a black horse,

$$\text{the negation of} \quad (\forall x, P(x)) \quad \text{is} \quad (\exists x, \neg P(x)) .$$

The existentially quantified statement "there exists a pink horse" is disproved by showing that "every horse is not pink." Mathematically, if $P(x)$ is the statement "x is a pink horse," then

$$\text{the negation of} \quad (\exists x, P(x)) \quad \text{is} \quad (\forall x, \neg P(x)) .$$

By thinking about the case of a general predicate P, the negation rules above still apply.

Example 1.4

Consider the mathematical statement $\forall x \in \mathbb{R}, x < x^2$. Determine whether this sentence is true or false, and write the negation of this sentence.

Solution. This sentence is false. For example, if $x = 1/2$ then $x^2 = 1/4$, showing that $x > x^2$. The negation of this sentence is

$$\exists x \in \mathbb{R} : x \geq x^2.$$

Our counter-example satisfies the negation of our sentence, as one would expect. ■

Example 1.5

Negate the sentence "Every real number has a negative."

Solution. From Example 1.2 we know that the given sentence can be stated mathematically as

$$\forall x \in \mathbb{R}, \exists y \in \mathbb{R}, x + y = 0.$$

Applying our rules for negation, the negative of this sentence becomes

$$\exists x \in \mathbb{R}, \forall y \in \mathbb{R}, x + y \neq 0.$$

Translating this back into an English sentence, we have "There is a real number which has no negative." ■

1.2 More on Sets

To a set S we can discuss its *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). Though used less often, the notion of a *superset* reverses the inclusion, and is written $\mathbb{Z} \supseteq P$. The *power set* of a set S is the collection of all subsets of S, and will be denoted by $\mathcal{P}(S)$. For example, if $S = \{a, b, c\}$ then

$$\mathcal{P}(S) = \{\emptyset, \{a\}, \{b\}, \{c\}, \{a, b\}, \{a, c\}, \{b, c\}, \{a, b, c\}\}.$$

If S has finite cardinal $|S| = n$, then $|\mathcal{P}(S)| = 2^n$ (Exercise 1-1). For this reason the power set is sometimes denoted 2^S, regardless of whether S is finite or not.

1.2.1 Operations on Sets

Union and Intersection: Let S be a set and fix two subsets $A, B \subseteq S$. The *union* of A and B is the collection of elements common to one of A or B:

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

The *intersection* of A and B to be collection of element common to both A and B,

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

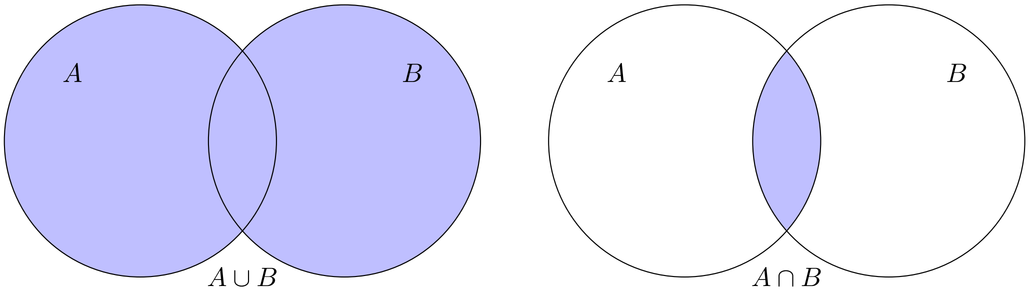


Figure 1.1: Left: The union of two sets is the collection of all elements which are in both, though remember that elements of sets are distinct, so we do not permit duplicates. Right: The intersection of two sets consists of all elements which are common to both sets.

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

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

Example 1.6

Suppose $A, B \subseteq C$. Define $A \setminus B = \{x \in A : x \notin B\}$. Show that $A \setminus B = A \cap B^c$.

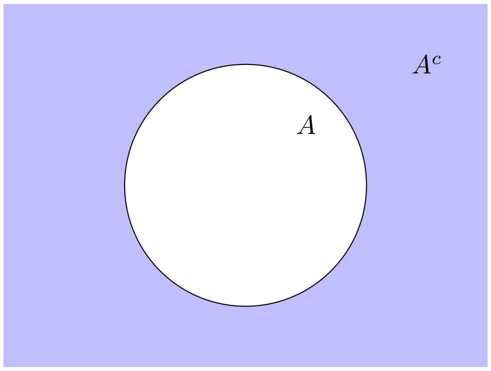


Figure 1.2: The complement of a set A with respect to S is the set of all elements which are in S but not in A.

Solution. We begin by showing that $A \setminus B \subseteq A \cap B^c$. Let $x \in A \setminus B$, so that $x \in A$ but $x \notin B$. Since $x \notin B$ we know that $x \in B^c$, and since $x \in A$ and $x \in B^c$ we know $x \in A \cap B^c$. This shows that $A \setminus B \subseteq A \cap B^c$.

The reverse direction is almost identical. Let $x \in A \cap B^c$ so that $x \in A$ and $x \in B^c$. The statement $x \in B^c$ is equivalent to saying that $x \notin B$, so $x \in A$ and $x \notin B^c$ implies that $x \in A \setminus B$. Both inclusions give the equality $A \setminus B = A \cap B^c$, as required. ■

Complements play nicely with intersections and unions through de Morgan's Laws.

Theorem 1.7: de Morgan's Laws

If S is a universe of discourse, with $A, B \subseteq S$, then

1. $(A \cup B)^c = A^c \cap B^c$
2. $(A \cap B)^c = A^c \cup B^c$.

The proof is a straightforward exercise is set theory, left to Exercise 1-3.

1.2.2 Quantifiers and Sets

Fix a universe of discourse S. We can use set-builder notation together with predicates to build sets. If P is a predicate, define

$$U_P = \{x \in S : P(x) \text{ is true}\}.$$

For example, if $S = \mathbb{Z}$ and $P(x)$ is the statement "x is even," then U_P contains the even numbers. Our three operations of AND, OR, and NOT then become familiar set operations.

Proposition 1.8

Let S be a universe of discourse and P, Q be predicates. If U_P and U_Q are those elements in S which satisfy P and Q respectively, then

1. $U_{P\wedge Q} = U_P \cap U_Q$,
2. $U_{P\vee Q} = U_P \cup U_Q$,
3. $U_{\neg P} = U_P^c$.

Proof. I'll prove (1) here and leave the rest as an exercise. By definition, we have

$$\begin{aligned} U_{P\wedge Q} &= \{x \in S : P(x) \wedge Q(x) \text{ is true}\} \\ &= \{x \in S : P(x) \text{ is true, and } Q(x) \text{ is true}\} \\ &= \{x \in S : P(x) \text{ is true}\} \cap \{x \in S : Q(x) \text{ is true}\} \\ &= U_P \cap U_Q. \end{aligned}$$

□

Remark 1.9 This is one of the few occasions where set equality followed without using double subset inclusion. In general, you will have no choice but to show two inclusions.

If P and Q are predicates, the statement $P \Rightarrow Q$ is equivalent to $U_P \subseteq U_Q$, as you will demonstrate in Exercise 1-5. Notice that if P is a predicate which always evaluates to false, then $U_P = \emptyset$. Since $\emptyset \subseteq U_Q$ for any Q, this corroborates why vacuous truths behave the way they do.

The universal and existential quantifiers work as follows: Let S be a universe, $A \subseteq S$, and P be a predicate,

- $\forall x \in A, P(x)$ is equivalent to $A \subseteq U_P$,
- $\exists x \in A, P(x)$ is equivalent to $A \cap U_P \neq \emptyset$.

This leads to interesting statements about the empty set. For example, if $A = \emptyset$ then for any predicate P we have

$$\forall x \in \emptyset, P(x) \quad \text{is equivalent to} \quad \emptyset \subseteq U_P,$$

which is always true. On the other hand

$$\exists x \in \emptyset, P(x) \quad \text{is equivalent to} \quad \emptyset \cap U_P \neq \emptyset,$$

which is always false.

1.2.3 Arbitrary Unions and Intersections

Taking a union or intersection of two sets can be combined to form longer chains, such as

$$A_1 \cup A_1 \cup A_3 \cup \cdots \cap A_n \quad \text{or} \quad A_1 \cap A_2 \cap A_3 \cap \cdots \cap A_n.$$

Quantifiers give the ability to form arbitrary unions and intersections, using infinitely many sets. Let $\{U_i : i \in I\}$ be a collection of sets, for some indexing set I. Here I could be finite or infinite of any size. We define

$$\bigcup_{i\in I} A_i = \{x : \exists k \in I, x \in A_k\} \quad \text{and} \quad \bigcap_{i\in I} A_i = \{x : \forall k \in I, x \in A_k\}.$$

These are generalizations of the definitions we saw before; namely, $x \in \bigcup_{i\in I} A_i$ if it is a member of some A_i, and $x \in \bigcap_{i\in I} A_i$ if it is a member of every A_i. When $I = \mathbb{N}$, we will write these as

$$\bigcup_{i=1}^{\infty} A_i \quad \text{and} \quad \bigcap_{i=1}^{\infty} A_i.$$

Example 1.10

Consider the sets $A_n = (-1/n, 1/n) \subseteq \mathbb{R}$. Show that

$$\bigcap_{i=1}^{\infty} A_n = \bigcap_{i=1}^{\infty} \left(-\frac{1}{n}, \frac{1}{n}\right) = \{0\}.$$

Solution. We need to show two inclusions. The $(\supseteq)$ direction is easiest. Since $0 \in (-1/n, 1/n)$ for all $n \in \mathbb{N}$, it is in the intersection. To show the $(\subseteq)$ inclusion, we will show that no non-zero number can be in the intersection. Fix an $x > 0$ in the real numbers, and choose some $N \in \mathbb{N}$ with $N > 1/x$. Since $x > 1/N$, x cannot be in $A_N = (-1/N, 1/N)$, and so x is not in the intersection. The proof for $x < 0$ is exactly the same, choosing N larger than $-1/x$. ■

On occasion we will discuss collections of sets $\mathcal{C} = \{U_i\}$, and need to take unions and intersections over all the elements in $\mathcal{C}$. As shorthand, we will let

$$\bigcup \mathcal{C} = \bigcup_{U_i \in \mathcal{C}} U_i \quad \text{and} \quad \bigcap \mathcal{C} = \bigcap_{U_i \in \mathcal{C}} U_i.$$

1.2.4 Cartesian Products

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[1]

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

[1] I'm lying to you here and hiding a technical and subtle problem. Look at Exercise 1-11 for more details.

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*.

Definition 1.11

Let A, B be two sets and $f : A \to B$ a function between them.

1. If $U \subseteq A$, the *image of U under f* is

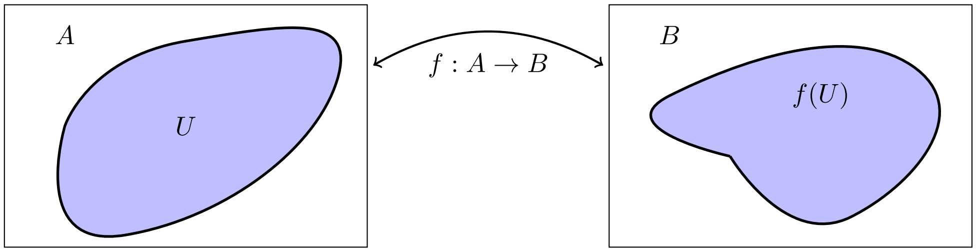
$$f(U) = \{y \in B : \exists x \in U, f(x) = y\} = \{f(x) : x \in U\}.$$

When $U = A$, we say that $f(A)$ is the *range of f*.

2. If $V \subseteq B$, the *pre-image of V under f* is

$$f^{-1}(V) = \{x \in A : f(x) \in V\}.$$

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 $g : \mathbb{R} \to [0, \infty), g(x) = x^2$.



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 1.12

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. I claim that $f([0, 1]) = [0, 1]$. Indeed, suppose $y \in f([0, 1])$ so that $y = f(x) = x^2$ for some $x \in [0, 1]$. Since $0 \leq x \leq 1$ we in turn know that $0 \leq x^2 = y \leq 1$ so that $y \in [0, 1]$ as required. Conversely, choose a $y \in [0, 1]$ and let $x = \sqrt{y}$, which exists and is also in $[0, 1]$. Then $f(x) = (\sqrt{y})^2 = y$ shows that $y \in f([0, 1])$. Both inclusions give the necessary equality.

On the other hand, since $f([0, 1]) = [0, 1]$ we know that $f^{-1}(f([0, 1])) = f^{-1}([0, 1])$, which I claim is the interval $[-1, 1]$. Indeed, if $x \in [-1, 1]$ then $f(x) = x^2 \in [0, 1]$ so that $[-1, 1] \subseteq f^{-1}([0, 1])$.

Conversely, if $x \in f^{-1}([0,1])$ then $f(x) = x^2 \in [0,1]$ by definition. Solving $0 \leq x^2 \leq 1$ gives $-1 \leq x \leq 1$ so that $x \in [-1,1]$. Hence $f^{-1}([0,1]) \subseteq [-1,1]$ and both inclusions give equality. ■

Example 1.13

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, $f(S^2) = D^2$. 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 $f(S^2) = D^2$. ■

1.3.1 Injective and Surjective Functions

Definition 1.14

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 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 as it misses the negative real numbers, while the function $f : \mathbb{R} \to [0,\infty)$ is surjective as there are no negative real numbers to miss. In addition, a function is surjective if $f(A) = B$.

Example 1.15

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 $(0, 0)$.

3. This function is both injective and surjective. Both are left as simple exercises. We conclude that h is bijective. ■

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

Here I'll quickly review some important structures we'll be using on $\mathbb{R}^n$. The portion concerning linear algebra should be review, but is mentioned here for completeness.

1.4.1 The Vector Space Structure

If F is a field, an *F-vector space* is a set V together with two operations; vector addition $V \times V \to V$, $(\mathbf{x}, \mathbf{y}) \to \mathbf{x} + \mathbf{y}$; and scalar multiplication $F \times V \to V$, $(\alpha, \mathbf{x}) \to \alpha\mathbf{x}$; such that for any $\mathbf{x}, \mathbf{y}, \mathbf{z} \in V$ and $\alpha, \beta \in F$

1. $\mathbf{x} + \mathbf{y} = \mathbf{y} + \mathbf{x}$
2. $\mathbf{x} + (\mathbf{y} + \mathbf{z}) = (\mathbf{x} + \mathbf{y}) + \mathbf{z}$
3. There is a distinguished element $\mathbf{0}$ such that $\mathbf{x} + \mathbf{0} = \mathbf{x}$.
4. $1_F\mathbf{x} = \mathbf{x}$
5. $\mathbf{x} + (-1_F)\mathbf{x} = \mathbf{0}$
6. $\alpha(\beta\mathbf{x}) = (\alpha\beta)\mathbf{x}$
7. $\alpha(\mathbf{x} + \mathbf{y}) = \alpha\mathbf{x} + \alpha\mathbf{y}$
8. $(\alpha + \beta)\mathbf{x} = \alpha\mathbf{x} + \beta\mathbf{x}$

A disadvantage of learning vector spaces so early in your mathematical career is that it's probably not clear why these are the axioms for a vector space. Those of you who take more abstract algebra will learn that vector spaces are example of "free modules," where a module is a group with a ring-action. The above axioms are then simply the group axioms and some compatibility conditions.

The vector space of greatest interest to us will be the $\mathbb{R}$ vector space $\mathbb{R}^n$, $n \in \mathbb{N}$. Elements $\mathbf{x} \in \mathbb{R}^n$ look like n-tuples of real numbers. If $\mathbf{x} = (x_1, \ldots, x_n), \mathbf{y} = (y_1, \ldots, y_n) \in \mathbb{R}^n$ and $\alpha \in \mathbb{R}$ then addition and scalar multiplication are done in a pointwise fashion

$$\mathbf{x} + \mathbf{y} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} + \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} x_1 + y_1 \\ x_2 + y_2 \\ \vdots \\ x_n + y_n \end{bmatrix}, \qquad \alpha\mathbf{x} = \alpha \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} \alpha x_1 \\ \alpha x_2 \\ \vdots \\ \alpha x_n \end{bmatrix}.$$

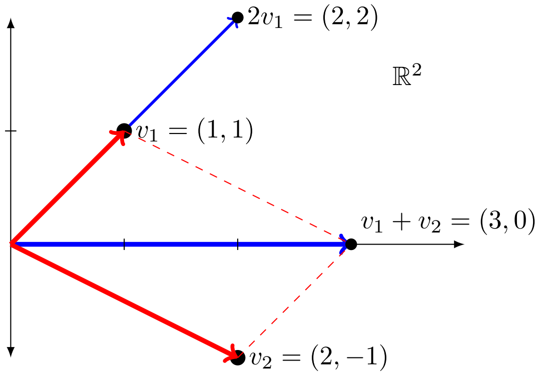


Figure 1.3: One may think of a vector in $\mathbb{R}^n$ as either representing a point in the plane (represented by the black dots) or as direction with magnitude (represented by the red arrows). The blue arrows correspond to the sum $v_1 + v_2$ and the scalar multiple $2v_1$. Notice that both are simply computed pointwise.

There are however, other examples of vector spaces that appear frequently or are of interest. For example, the set of continuous functions on the interval $[0,1]$,

$$C([0,1]) = \{f : [0,1] \to \mathbb{R} : f \text{ continuous}\},$$

is a real vector space.

Recall that a set of vectors $\mathcal{B} \subseteq V$ is *linearly independent* in V if for every finite subset $\{v_1, \ldots, v_n\} \subseteq \mathcal{B}$, $\sum_i c_i \mathbf{v}_i = 0$ implies that $c_i = 0$ for all $i = 1, \ldots, n$. The same set is said to *span* V if for every $\mathbf{v} \in V$ there exists a finite collection of vectors $\{v_1, \ldots, v_n\} \subseteq \mathcal{B}$ and constants $c_i \in F$ such that $\mathbf{v} = \sum_i c_i \mathbf{v}_i$. A *basis* for V is any linearly independent spanning set.

Theorem 1.16

If V is a real vector space, there exists a cardinal m – called the *dimension of* V and written $\dim(V)$ – such that

1. If $\mathcal{B} \subseteq V$ is linearly independent, $|\mathcal{B}| \leq m$,
2. If $\mathcal{B} \subseteq V$ spans V, then $|\mathcal{B}| \geq m$,
3. If $\mathcal{B} \subseteq V$ is a basis for V, then $|\mathcal{B}| = m$.

The proof for this theorem in finite dimensions is not too terrible, though in infinite dimensions the situation becomes more complex. The fact that *every* vector space admits a basis in the first place is in fact equivalent to the Axiom of Choice. I am a pro-choice mathematician, and the course will be taught as such.

A vector space is *finite-dimensional* if its dimension is a natural number, and *infinite dimensional* otherwise. For example, $\mathbb{R}^n$ has dimension n, as realized by its standard basis $\{\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n\}$ where $\mathbf{e}_i$ is 1 in its ith coordinate and zero everywhere else. The real vector space $C([0,1])$ is infinite dimensional, though this is harder to prove (Exercise 1-19).

A *linear transformation* between two F-vector spaces V and W is a function $T : V \to W$ such that $T(\alpha\mathbf{x}+\beta\mathbf{y}) = \alpha T(\mathbf{x}) + \beta T(\mathbf{y})$ for all $\mathbf{x},\mathbf{y} \in V$ and $\alpha,\beta \in F$. If $\dim(V) = n, \dim(W) = m$, then every linear transformation $T : V \to W$ is equivalent to

$$T(\mathbf{x}) = A\mathbf{x}$$

where A is an $m \times n$ matrix. The two vector spaces are said to be *isomorphic* – written $V \cong W$ – if there is a bijective linear transformation between them. All finite dimensional F-vector spaces are isomorphic (Exercise 1-22).

A *subspace* of an F-vector space V is a set $S \subseteq V$ satisfying

1. $\mathbf{x}+\mathbf{y} \in S$ whenever $\mathbf{x},\mathbf{y} \in S$,
2. $\alpha\mathbf{x} \in S$ whenever $\mathbf{x} \in S$ and $\alpha \in F$,
3. $\mathbf{0} \in S$.

In effect, subspaces are themselves vector spaces which live in a larger ambient vector space. For example, in $\mathbb{R}^3$ we can classify all the subspaces according to their dimension:

- The zero dimensional subspace $\{\mathbf{0}\}$,
- The one dimensional subspaces, corresponding to lines through the origin.
- The two dimensional subspaces, corresponding to planes through the origin.
- The three dimensional subspace $\mathbb{R}^3$.

The *codimension* of a subspace $S \subseteq V$ is $\operatorname{codim} S = \dim V - \dim S$.

Suppose $T : V \to W$ is a linear transformation. The *kernel* of T is the set

$$\ker T = \{\mathbf{v} \in V : T(\mathbf{v}) = \mathbf{0}_W\}$$

while the *image* of T is

$$\operatorname{image} T = \{\mathbf{w} : T(\mathbf{v}) = \mathbf{w} \text{ for some } \mathbf{v} \in V\}.$$

These are sometimes referred to as the null-space and column-space respectively, and $\ker T$ is a subspace of V while $\operatorname{image} T$ is a subspace of W. The *rank* of T is the dimension of its image, so $\operatorname{rank} T = \dim(\operatorname{image} T)$

The notion of matrix multiplication is done so as to agree with function composition. Suppose then that $T : U \to V$ has a matrix representation as $T(\mathbf{x}) = B\mathbf{x}$, and $S : V \to W$ has matrix representation $S(\mathbf{x}) = A\mathbf{x}$. Their composition is a map $(S \circ T) : U \to W$ and should be given by $ST(\mathbf{x}) = (AB)\mathbf{x}$, for an appropriate definition of matrix multiplication AB. If A is an $n \times k$ matrix and B is $k \times m$, the product AB is an $n \times m$ matrix, whose ith column is $A\mathbf{b}_i$ (where $\mathbf{b}_i$ is the ith column of B), or whose (i,j)th element is

$$(AB)_{ij} = \sum_{r=1}^{k} A_{ik}B_{kj}.$$

Explicitly multiplying two 2×2 matrices $A = [A_{ij}], B = [B_{ij}]$, we get the 2×2 matrix

$$\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix} = \begin{bmatrix} a_{11}b_{11} + a_{12}b_{21} & a_{11}b_{12} + a_{12}b_{22} \\ a_{21}b_{11} + a_{22}b_{21} & a_{21}b_{12} + a_{22}b_{22} \end{bmatrix}.$$

If $T : V \to W$, $x \mapsto A\mathbf{x}$ is an invertible transformation with inverse $T^{-1} : W \to V$, $x \mapsto B\mathbf{x}$, we must then have $T \circ T^{-1} = \mathrm{id}$, or $\mathbf{x} \mapsto (AB)\mathbf{x} = \mathbf{x}$ showing that $AB = I_n$, the $n \times n$ identity matrix. Such a matrix B is called the *inverse matrix* to A and is denoted $B = A^{-1}$. It is not hard to check that T is invertible if and only if $n = \dim V = \dim W$, in which case A and A^{-1} are both $n \times n$ matrices. There are multiple ways of computing A^{-1} from A, so choose your favourite.

Finally, the determinant is a map $\det : M_n(\mathbb{R}) \to \mathbb{R}$, where $M_n(\mathbb{R})$ are the collection of $n \times n$ square matrices. When thought of as a function on the columns of the input matrix $\det : (\mathbb{R}^n)^n \to \mathbb{R}$, it is the unique map which satisfies the following three properties:

1. [n-linear] det is linear in each of its arguments,
2. [Anti-symmetric] Interchanging two arguments results in a negative sign,
3. [Normal] $\det(I_n) = 1$.

Geometrically, the determinant measures the multiplicative volume change under the transformation $T(\mathbf{x}) = A\mathbf{x}$; for example, the parallelepiped which arises as the image of the unit n-cube will have volume $\det(A)$. As such, it can be used to determine whether a matrix is invertible: A non-invertible matrix will have a non-trivial kernel (Exercise 1-20), meaning the image of any n-dimensional shape will have smaller dimension than n, thus always admitting zero volume and giving $\det(A) = 0$.

The most straightforward way of computing the determinant is via cofactor expansion. Once again, I refer you to your favourite linear algebra text for more information.

1.4.2 Of Lengths and Such

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

Definition 1.17

Given a real vector space V, an *inner product* on V is a map $\langle \cdot, \cdot \rangle : V \times V \to \mathbb{R}$ satisfying

1. [Symmetric] $\langle \mathbf{x}, \mathbf{y} \rangle = \langle \mathbf{y}, \mathbf{x} \rangle$ for every $\mathbf{x}, \mathbf{y} \in V$,
2. [Linear] $\langle a\mathbf{x} + b\mathbf{y}, \mathbf{z} \rangle = a \langle \mathbf{x}, \mathbf{z} \rangle + b \langle \mathbf{y}, \mathbf{z} \rangle$ for all $\mathbf{x}, \mathbf{y}, \mathbf{z} \in V$ and $a, b \in \mathbb{R}$
3. [Positive Definite] $\langle \mathbf{x}, \mathbf{x} \rangle \geq 0$ and $\langle \mathbf{x}, \mathbf{x} \rangle = 0$ if and only if $\mathbf{x} = 0$.

Combining the symmetry and linear properties of an inner product tell us that an inner product is actually *bilinear*, or linear in each of its components:

$$\langle \mathbf{z}, a\mathbf{x} + b\mathbf{y} \rangle = \langle a\mathbf{x} + b\mathbf{y}, \mathbf{z} \rangle = a \langle \mathbf{x}, \mathbf{z} \rangle + b \langle \mathbf{y}, \mathbf{z} \rangle = a \langle \mathbf{z}, \mathbf{x} \rangle + b \langle \mathbf{z}, \mathbf{y} \rangle.$$

For this reason, you might see an inner product defined as a positive definite, symmetric, bilinear mapping.

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

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

I will leave it to you to show this satisfies the axioms of an inner product. Geometrically, the dot product $\mathbf{x} \cdot \mathbf{y}$ is the magnitude of the projection of $\mathbf{x}$ onto the unit vector in the $\mathbf{y}$ direction, or vice versa.

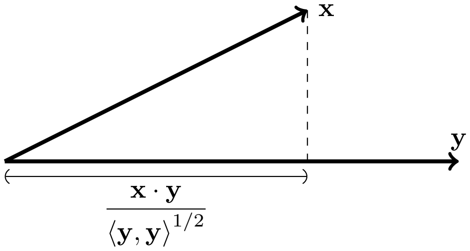

$\mathbf{x}$
$\mathbf{y}$
$\frac{\mathbf{x} \cdot \mathbf{y}}{\langle \mathbf{y}, \mathbf{y} \rangle^{1/2}}$

Figure 1.4: 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 $\mathbf{y}$.

We can use an inner product to define orthogonal complements of subspaces. So if V is a vector space endowed with an inner product $\langle \cdot, \cdot \rangle$ and S is a subspace, we define the orthogonal complement to S as

$$S^\perp = \{ \mathbf{v} \in V : \langle \mathbf{v}, \mathbf{w} \rangle = 0, \forall \mathbf{w} \in S \}.$$

You will show in Exercise 1-32 that $\dim(S^\perp) = \operatorname{codim}(S)$, and that if V is finite dimensional, $(S^\perp)^\perp = S$. This can be convenient for defining subspaces. For example, in $\mathbb{R}^3$ the two dimensional subspaces are planes through the origin. Rather than specify two basis elements which span the plane, it's sufficient to specify a single vector which is orthogonal to every element of the plane.

Example 1.18

Take $V = \mathbb{R}^3$ with the standard inner product, $\mathbf{v} = (1, 0, -1)^T$, and let $S = \operatorname{span}\{v\}$. Find $S^\perp$.

Solution. As $\dim(S) = 1$ we know $\dim(S^\perp) = 2$, and as such is a plane. Every element of S looks like $\alpha \mathbf{x} = (\alpha, 0, -\alpha)$ for some $\alpha \in \mathbb{R}$, so the orthogonal complement is

$$\begin{aligned} S^\perp &= \{ \mathbf{y} : \mathbf{y} \cdot \mathbf{x} = 0, \forall x \in \mathbf{v} \} \\ &= \{ (x, y, z) : (x, y, z) \cdot (\alpha, 0, -\alpha) = 0 \} \\ &= \{ (x, y, z) : x - z = 0 \}. \end{aligned}$$ ■

Cross Products: 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

$$\mathbf{v} \times \mathbf{w} = (v_2w_3 - w_2v_3, w_1v_3 - v_1w_3, v_1w_2 - w_1v_2).$$

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

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

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

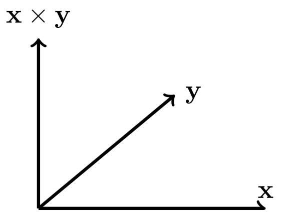


Figure 1.5: The cross product of two vectors $\mathbf{x} \times \mathbf{y}$.

Example 1.19

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

Solution. Using our definition, one has

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

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

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

One final property we'll need for the next section is the following inequality:

Proposition 1.20: Cauchy-Schwarz

If V is a real vector space with an inner product $\langle \cdot, \cdot \rangle$, then for every $\mathbf{x}, \mathbf{y} \in V$ we have

$$\langle \mathbf{x}, \mathbf{y} \rangle^2 \leq \langle \mathbf{x}, \mathbf{x} \rangle \langle \mathbf{y}, \mathbf{y} \rangle$$

with equality if and only if $\mathbf{x}$ and $\mathbf{y}$ are linearly independent.

Proof. If $\mathbf{x} = 0$ both sides are zero and the result is trivially true, so assume $\mathbf{x} \neq 0$. Let $p = \langle \mathbf{x}, \mathbf{y}\rangle / \langle \mathbf{x}, \mathbf{x}\rangle$, which you may recognize as the projection coefficient of $\mathbf{y}$ onto $\mathbf{x}$, so that $p\mathbf{x}$ is shown in Figure 1.4. Now

$$\begin{aligned}\langle \mathbf{y} - p\mathbf{x}, \mathbf{y} - p\mathbf{x}\rangle &= \langle \mathbf{y}, \mathbf{y}\rangle - p\langle \mathbf{y}, \mathbf{x}\rangle - p\langle \mathbf{x}, \mathbf{y}\rangle + p^2\langle \mathbf{x}, \mathbf{x}\rangle \\ &= \langle \mathbf{y}, \mathbf{y}\rangle - 2\frac{\langle \mathbf{x}, \mathbf{y}\rangle^2}{\langle \mathbf{x}, \mathbf{x}\rangle} + \frac{\langle \mathbf{x}, \mathbf{y}\rangle^2\langle \mathbf{x}, \mathbf{x}\rangle}{\langle \mathbf{x}, \mathbf{x}\rangle^2} \\ &= \langle \mathbf{y}, \mathbf{y}\rangle - \frac{\langle \mathbf{x}, \mathbf{y}\rangle^2}{\langle \mathbf{x}, \mathbf{x}\rangle}.\end{aligned}$$

This term is always non-negative by the positive-definite property of the inner product, hence

$$\langle \mathbf{y}, \mathbf{y}\rangle - \frac{\langle \mathbf{x}, \mathbf{y}\rangle^2}{\langle \mathbf{x}, \mathbf{x}\rangle} \geq 0 \quad \Rightarrow \quad \langle \mathbf{x}, \mathbf{y}\rangle^2 \leq \langle \mathbf{x}, \mathbf{x}\rangle\langle \mathbf{y}, \mathbf{y}\rangle$$

from which the inequality follows by taking square roots. For equality, note that the first line is zero if and only if $\mathbf{y} - p\mathbf{x} = 0$, or $\mathbf{y} = p\mathbf{x}$, showing that $\mathbf{x}$ and $\mathbf{y}$ are linearly dependent. □

Norms: The next structure is called a *norm*, and prescribes a way of measuring the length of a vector.

Definition 1.21

Let V be a real vector space. A norm on V is a map $\|\cdot\| : V \to V$ satisfying,

1. [Non-degenerate] $\|\mathbf{x}\| \geq 0$ for all $\mathbf{x} \in V$ and $\|\mathbf{x}\| = 0$ if and only if $\mathbf{x} = 0$,
2. [Homogeneous] $\|\alpha\mathbf{x}\| = |\alpha|\|\mathbf{x}\|$ for all $\mathbf{x} \in V$ and $\alpha \in \mathbb{R}$,
3. [Triangle Inequality] $\|\mathbf{x} + \mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|$.

The *Euclidean norm* is induced from the Euclidean inner product

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

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 1.6). I will leave it as an exercise to show that the Euclidean norm is indeed a norm.

In general, if $\langle \cdot, \cdot\rangle$ is an inner product on V, then $\|\cdot\| = \sqrt{\langle \cdot, \cdot\rangle}$ is a norm on V (Exercise 1-33). Indeed, the first two properties follow immediately from the definition of an inner product, leaving only the Triangle Inequality to be shown. Here one has

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

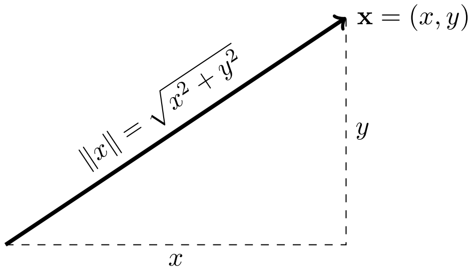


Figure 1.6: In $\mathbb{R}^2$, the length of a vector can be derived from the Pythagorean theorem. The norm $\|\cdot\|$ generalizes this notion to multiple dimensions.

By taking the square root of both sides gives the desired result.

On the other hand, there are plenty of norms which do not come from inner products. A popular example on $\mathbb{R}^n$ are the p-norms: If $p \geq 1$ is a real number, define

$$\|\mathbf{x}\|_p = \left(\sum_{i=1}^n |x_i|^p \right)^{1/p}.$$

The p-norm comes from an inner product if and only if $p = 2$, which is precisely the Euclidean norm.

Metrics: Finally, one has a *metric*. Metrics are the most flexible, least rigid structure we'll impose on a space, and as such will be our focus in the next chapter. Loosely speaking, metrics prescribe a method for determining the distance between two vectors.

Definition 1.22

A set X with a function $d : X \times X \to \mathbb{R}$ is said to be a *metric space* if

1. [Symmetry] $d(x, y) = d(y, x)$ for all $x, y \in X$,
2. [Non-degenerate] $d(x, y) \geq 0$ for all $x, y \in X$, with $d(x, x) = 0$ if and only if $x = y$.
3. [Triangle Inequality] $d(x, z) \leq d(x, y) + d(y, z)$ for all $x, y, z \in X$.

Note that a metric space does *not* need to be a vector space: The definition of a metric has no mention of addition, multiplication, or even of an $\mathbf{0}$ element. However, just as inner products induced norms, so too do norms induce metrics. If V is a real vector space with a norm $\|\cdot\|$, then the induced metric is

$$d(\mathbf{x}, \mathbf{y}) = \|\mathbf{x} - \mathbf{y}\|.$$

That the three properties of a metric are satisfied is almost immediate.

This means the from the Euclidean norm we have the *Euclidean metric*. If $\mathbf{x} = (x_1, \ldots, x_n)$ and $\mathbf{y} = (y_1, \ldots, y_n)$ then the Euclidean metric is

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

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

1.5 Exercises

1-1. Let X be a finite set. Show that $|\mathcal{P}(X)| = 2^{|X|}$.

1-2. For two sets A, B, we define $B^A = \{f : A \to B\}$; that is, the set of all functions from A to B. The reason for this curious notation will be illuminated below. Let $\hat{n} = \{1, 2, \ldots, n\}$.

(a) Show that there is a bijection between $\mathbb{R}^{\hat{n}}$ and $\mathbb{R}^n$.

(b) Show that for any set X, there is a bijection between $\hat{2}^X$ and $\mathcal{P}(X)$.

1-3. (a) Prove de Morgan's Laws (Theorem 1.7).

(b) Generalize de Morgan's Laws as follows: Suppose $\{A_i : i = 1, \ldots, n\}$ is any finite collection of sets. Show that

$$\left[\bigcup_{i=1}^n A_i\right]^c = \bigcap_{i=1}^n A_i^c, \qquad \left[\bigcap_{i=1}^n A_i\right]^c = \bigcup_{i=1}^n A_i^c$$

(c) Generalize the result even further as follows: Suppose I is some arbitrary indexing set, and $\{A_i : i \in I\}$ is an arbitrary collection of sets. Show that

$$\left[\bigcup_{i\in I} A_i\right]^c = \bigcap_{i\in I} A_i^c, \qquad \left[\bigcap_{i\in I} A_i\right]^c = \bigcup_{i\in I} A_i^c$$

1-4. Finish the proof of Proposition 1.8.

1-5. Suppose that P, Q are two predicates. Show that $P \Rightarrow Q$ is equivalent to the statement that $U_P \subseteq U_Q$.

1-6. Let $A \subseteq S$ and $B \subseteq S$. Prove each of the following statements

(a) $A \subseteq B$ if and only if $A \cup B = B$

(b) $A^c \subseteq B$ if and only if $A \cup B = S$

(c) $A \subseteq B$ if and only if $B^c \subseteq A^c$

(d) $A \subseteq B^c$ if and only if $A \cap B = \emptyset$

1-7. Let $A_1, A_2, \ldots, A_n$ be sets. If $A_1 \subseteq A_2 \subseteq A_3 \subseteq \cdots \subset A_n$ and $A_n \subseteq A_1$, then $A_1 = A_2 = A_3 = \cdots = A_n$.

1-8. Let I be an index for a collection of subsets $A_i \subseteq S$, $i \in I$. Show that for every $k \in I$, $\bigcap_{i\in I} A_i \subseteq A_k$

1-9. Let $f : A \to B$ be a function.

(a) For every $X \subseteq A$, $X \subseteq f^{-1}(f(X))$

(b) For every $Y \subseteq B$, $Y \supseteq f(f^{-1}(Y))$

(c) If $f : A \to B$ is injective, then for every $X \subseteq A$ we have $X = f^{-1}(f(X))$

(d) If $f : A \to B$ is surjective, then for every $Y \subseteq B$ we have $Y = f(f^{-1}(Y))$

1-10. Let $f : A \to B$ be a map of sets, and let $\{X_i\}_{i\in I}$ be an indexed collection of subsets of A.

(a) Prove that $f\left(\bigcup_{i\in I} X_i\right) = \bigcup_{i\in I} f(X_i)$

(b) Prove that $f\left(\bigcap_{i\in I} X_i\right) \subset \bigcap_{i\in I} f(X_i)$

(c) When does equality of sets hold in the above part?

1-11. Writing $\mathbb{R}^3 = \mathbb{R}\times\mathbb{R}\times\mathbb{R}$ is technically nonsense, since we have not defined what it means to take a threefold Cartesian product. Instead, we need to realize this as either

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

or

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

which are not technically the same. How do we get around this? Let X, Y, Z be three sets.

(a) Show there is a bijection between $(X \times Y) \times Z$ and $X \times (Y \times Z)$, so that up to bijection we can assume these are the same set. We say that "the Cartesian product is associative up to isomorphism in the category of sets."

(b) Show that there is a bijection between $X \times Y$ and $Y \times X$. We say that "the Cartesian product is commutative up to isomorphism in the category of sets."

1-12. Let $S^1 = \left\{(x,y) : x^2+y^2 = 1\right\}$ be the unit circle in $\mathbb{R}^2$. Normally we'd realize $S^1 \times S^1$ as a subset of $\mathbb{R}^4$; however, show that there is a bijection between $S^1 \times S^1$ and the torus in $\mathbb{R}^3$. Hint: The torus in $\mathbb{R}^3$ can be described as

$$\mathbb{T}_{r,R} = \left\{(x,y,z) : z^2 + (R - \sqrt{x^2+y^2})^2 = r^2\right\},$$

where R is the distance from the origin to the center of the tube, and r is the radius of the tube.

1-13. A *Boolean algebra* is a set S together with two binary operators $+, \times$, one unary operator $-$, and two elements 0 and 1 such that for any $x, y, z \in S$

- $x + (y + z) = (x + y) + z$
- $x \times (y \times z) = (x \times y) \times z$
- $x + y = y + x$ and $x \times y = y \times x$
- $x + (x \times b) = x$ and $x \times (x + b) = x$
- $x \times (y + z) = x \times y + x \times z$
- $x + (y \times z) = (x + y) \times (x + z)$
- $x + 0 = x$ and $x \times 1 = x$
- $x + (-x) = 0$ and $x \times -x = 1$.

Let S be any set. Show that taking $\cup, \cap$ as the binary operators, the complement ^{c} as the unary operator, and $\emptyset, S$ as 0 and 1 makes S into a Boolean algebra.

1-14. Let $f : \mathbb{R}^3 \to \mathbb{R}$ be given by $f(x,y,z) = (x,y)$. If $D^2 = \left\{(x,y) : x^2+y^2 \leq 1\right\} \subseteq \mathbb{R}^2$, determine $f^{-1}(D^2)$.

1-15. Suppose $S \subseteq \mathbb{R}$ and $f, g : S \to \mathbb{R}$ are two functions.

(a) If f, g are injective, is $F : S \to \mathbb{R}^2$, $(x,y) \mapsto (f(x), g(x))$ an injective function?

(b) If f, g are injective, is $F : S \times S \to \mathbb{R}^2$, $(x, y) \mapsto (f(x), g(y))$ an injective function?

(c) If $F : S \to \mathbb{R}^2$, $(x, y) \mapsto (f(x), g(x))$ is an injective function, need both f and g be injective?

1-16. Suppose $S \subseteq \mathbb{R}$ and $f, g : S \to \mathbb{R}$ are two functions.

(a) If f, g are surjective, is $F : S \to \mathbb{R}^2$, $(x, y) \mapsto (f(x), g(x))$ a surjective function?

(b) If f, g are surjective, is $F : S \times S \to \mathbb{R}^2$, $(x, y) \mapsto (f(x), g(y))$ a surjective function?

(c) If $F : S \to \mathbb{R}^2$, $(x, y) \mapsto (f(x), g(x))$ is a surjective function, need both f and g be surjective?

1-17. Prove that the following is true: The map $(\mathbf{x}, \mathbf{y}) \mapsto \langle \mathbf{x}, \mathbf{y} \rangle$ defines an inner product if and only if there exists a $n \times n$ symmetric, positive definite matrix A such that $\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{x}^T A\mathbf{y}$.

1-18. Show that $C([0, 1])$, the set of real-valued continuous functions on the interval $[0, 1]$, is a real vector space.

1-19. Show that the real vector space $C([0, 1])$ is infinite dimensional as follows:

(a) Convince yourself that it is sufficient to show that for every $n \in \mathbb{N}$, $C([0, 1])$ contains a linearly independent set of dimension n.

(b) Let $P_n(\mathbb{R})$ be the collection of real-coefficient polynomials of degree n. Show that $P_n(\mathbb{R})$ is a real vector space of dimension $n + 1$.

(c) Show that $P_n(\mathbb{R})$ is a subspace of $C([0, 1])$ for any $n \in \mathbb{N}$, and conclude that $C([0, 1])$ has infinite dimension.

1-20. Show that the transformation $T : V \to W$ is injective if and only if $\ker T = \{0\}$.

1-21. Fix a real vector space V and a set of vectors $\mathcal{B} = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\} \subseteq V$. Define a map $T : \mathbb{R}^n \to V, (c_1, \ldots, c_n) \mapsto c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots c_n\mathbf{v}_n$.

(a) Show that $\mathcal{B}$ is linearly independent if and only if T is injective.

(b) Show that $\mathcal{B}$ spans V if and only if T is surjective.

(c) Show that $\mathcal{B}$ is a basis for V if and only if T is bijective. Conclude that T is an isomorphism of vector spaces.

1-22. Suppose that V and W are two finite dimensional real vector spaces satisfying $\dim(V) = \dim(W)$. Show that $V \cong W$.

1-23. Let $V = \mathbb{R}^3$ with the Euclidean inner product $\langle \cdot, \cdot \rangle$.

(a) Show that $\mathbf{v} \times \mathbf{w} = -\mathbf{w} \times \mathbf{v}$.

(b) Show that $\langle \mathbf{v}, \mathbf{v} \times \mathbf{w} \rangle = 0$ in general.

(c) 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.

1-24. Let $T : V \to W$ be a linear transformation between F-vector spaces V and W. Show that $\ker T$ is a subspace of V, and $\operatorname{image} T$ is a subspace of W.

1-25. Suppose $T : V \to W$ is a rank k linear transformation between finite dimensional vector spaces.

(a) Show that T has k linearly indepenent columns,

(b) Show that T has k linearly independent rows,

(c) Show that T has a $k \times k$ submatrix which is invertible.

1-26. Show that the collection of infinite differentiable functions $C^\infty([0,1])$ is a subspace of $C([0,1])$.

1-27. Let V, W be real vector spaces, and let $L(V, W)$ be the collection of linear transformations between V and W.

(a) Show that $L(V, W)$ is itself a real vector space.

(b) If $\dim(V) = n$ and $\dim(W) = m$, show that $L(V, W) \cong \mathbb{R}^{nm}$.

1-28. Let $L(\mathbb{R}^n, \mathbb{R}^m)$ be the set of all linear transformations from $\mathbb{R}^n \to \mathbb{R}^m$, with $\|\cdot\|_{\text{Eu}}$ the Euclidean norm..

(a) Define a map $L(\mathbb{R}^n, \mathbb{R}^m) \to \mathbb{R}$ by $T \mapsto \|T\|_{\text{op}} = \sup \{\|T(\mathbf{x})\| : \|\mathbf{x}\|_{\text{Eu}} = 1\}$. This is called the *operator norm*. Show that $\|\cdot\|_{\text{op}}$ is indeed a norm on $L(\mathbb{R}^n, \mathbb{R}^m)$.

(b) Show that for any $\mathbf{v} \in \mathbb{R}^n$, $\|A\mathbf{v}\| \leq \|A\|_{\text{op}} \|\mathbf{v}\|$.

(c) Let $\{\mathbf{e}_i\}_{i=1}^n$ be the standard basis for $\mathbb{R}^n$, and define a map $L(\mathbb{R}^n, \mathbb{R}^m) \to \mathbb{R}$ by

$$T \mapsto \|T\|_* = \left[\sum_{k=1}^n \|T(\mathbf{e}_k)\|_{\text{Eu}}\right]^{1/2}.$$

Show that $\|T\|_*$ is a norm, and that under the isomorphism $L(\mathbb{R}^n, \mathbb{R}^m) \cong \mathbb{R}^{nm}$, $\|T\|_*$ agrees with the Euclidean norm.

(d) Show that $\|T\|_{\text{op}} \leq \|T\|_*$. *Hint:* Write out $T\mathbf{x}$ and use the Cauchy-Schwarz inequality.

1-29. Similar to Exercise 1-27, let V be a real vector space and define $V^* = L(V, \mathbb{R})$ as the linear maps from V to $\mathbb{R}$. This is known as the *dual space to V*.

(a) Define a map $P : V \times V^* \to \mathbb{R}$ by $(\mathbf{v}, f) \mapsto f(\mathbf{v})$. Show that P is a bilinear map.

(b) Suppose V is endowed with an inner product $\langle \cdot, \cdot \rangle$, and define the map $V \to V^*, v \mapsto v^* = \langle v, \cdot \rangle$. If V is a finite dimensional vector space with basis $\{v_1, \ldots, v_n\}$, show that $\{\mathbf{v}_1^*, \cdots, \mathbf{v}_n^*\}$ is a basis for V^*.

(c) If V is finite dimensional, show that $f \in V^*$ if and only if there exists a unique $v \in V$ such that $f = v^*$. Conclude that the inner product $\langle v, w \rangle$ is equivalent to the pairing $v^*(w)$.

Note: The result in part (c) generalizes to infinite dimensions if additional hypotheses are added to the vector space V.

1-30. If $T : V \to W$ is a linear transformation, its *adjoint* is the map $T^* : W^* \to V^*$, $f \mapsto f \circ T$.

(a) Convince yourself that for all $f \in W^*$ and $v \in V$, $T^*(f)(v) = f(T(v))$.

(b) Suppose V is endowed with an inner product $\langle \cdot, \cdot \rangle$. Using Exercise 1-29c and part (a), convince yourself that the adjoint satisfies $\langle f, T(v) \rangle = \langle T^*(f), v \rangle$.

(c) Let $\langle \cdot, \cdot \rangle$ be the Euclidean inner product, and $T(\mathbf{x}) = A\mathbf{x}$ be a matrix representation of the linear transformation. Show that $T^*(f) = A^T f$, where A^T is the transpose of A.

1-31. Show that the Euclidean inner product satisfies the axioms of Definition 1.17.

1-32. Let V be a finite dimensional vector space and $S \subseteq V$ a subspace.

(a) Show that $\dim(S^\perp) = \operatorname{codim}(S)$.

(b) Show that $(S^\perp)^\perp = S$.

(c) Let $\operatorname{Sub}(V)$ be the collection of all subspaces of V, and define the map $\operatorname{Sub}(V) \to \operatorname{Sub}(V)$ by $S \mapsto S^\perp$. Show this map is a bijection.

1-33. Suppose V is a vector space endowed with an inner product $\langle \cdot, \cdot \rangle$. Define $\|\cdot\| = \sqrt{\langle \cdot, \cdot \rangle}$. Show that $\|\cdot\|$ is a norm on V.

1-34. Suppose V is a vector space.

(a) Suppose that $\langle \cdot, \cdot \rangle$ is an inner product on this vector space, and $\|\cdot\|$ is the induced norm. Show that

$$2\|\mathbf{x}\|^2 + 2\|\mathbf{y}\|^2 = \|\mathbf{x} + \mathbf{y}\|^2 + \|\mathbf{x} - \mathbf{y}\|^2. \tag{1.3}$$

(b) Equation (1.3) is called the *Parallelogram Law* or the *polarization identity.* If V has a norm $\|\cdot\|$ which satisfies the polarization identity, then in fact it comes from an inner product (hence the norm comes from an inner product if and only if Equation (1.3) is true). This is difficult to prove, though you are free to try. Suppose then that the polarization identity is true and $\|\cdot\|$ is induced by an inner product. Find the corresponding inner product.

(c) Show that the p-norm satisfies the Parallelogram Law if and only if $p = 2$.

1-35. Show that $\|f\|_{\text{sup}} = \sup_{x \in [0,1]} \|f(x)\|$ is a norm on $C([0,1])$, where $\|\cdot\|$ is the Euclidean norm.

1-36. If V is a vector space, a metric $d : V \times V \to \mathbb{R}$ is said to be *homogeneous* if $d(\alpha\mathbf{x}, \alpha\mathbf{y}) = |\alpha| d(\mathbf{x}, \mathbf{y})$ for all $\mathbf{x}, \mathbf{y} \in V$ and $\alpha \in \mathbb{R}$. It is said to be *translation invariant* if $d(\mathbf{x} + \mathbf{z}, \mathbf{y} + \mathbf{z}) = d(\mathbf{x}, \mathbf{y})$ for all $\mathbf{x}, \mathbf{y}, \mathbf{z} \in V$. Show that a homogeneous, translation invariant metric is induced by a norm on V.

1-37. Let $\mathbf{x} = (x_1, \ldots, x_n)$ and $\mathbf{y} = (y_1, \ldots, y_n)$ be points in $\mathbb{R}^n$. Show that each of the following is a metric on $\mathbb{R}^n$.

(a) $d(\mathbf{x}, \mathbf{y}) = \sum_{i=1}^n |x_i - y_i|$

(b) $d(\mathbf{x}, \mathbf{y}) = \sup_{i=1,\ldots,n} |x_i - y_i|$

(c) $d(\mathbf{x}, \mathbf{y}) = \begin{cases} 1 & \mathbf{x} \neq \mathbf{y} \\ 0 & \text{otherwise} \end{cases}$

(d) $d(\mathbf{x},\mathbf{y}) = \begin{cases} \|\mathbf{x}-\mathbf{y}\| & \text{if } \mathbf{x} = \lambda\mathbf{y} \\ \|\mathbf{x}\| + \|\mathbf{y}\| & \text{otherwise} \end{cases}$ where $\|\cdot\|$ is the Euclidean norm.

1-38. Suppose V is a real vector space. If X, Y are two subspaces of V such that $X \cap Y = \{\mathbf{0}\}$, we define the *internal direct sum* $X \oplus Y = \{\mathbf{x}+\mathbf{y} : \mathbf{x} \in X, \mathbf{y} \in Y\}$.

(a) Show that $X \oplus Y$ is a subspace of V.

(b) Suppose X and Y are finite dimensional. Show that $\dim(X \oplus Y) = \dim X + \dim Y$.

(c) Assume V is finite dimensional, and comes equipped with an inner product $\langle \cdot, \cdot \rangle : V \times V \to \mathbb{R}$. If X is a subspace of V, recall the definition $X^\perp$ from Exercise 1-32. Show that $V = X \oplus X^\perp$.

1-39. If V is a real vector space, a map $f : V^n \to \mathbb{R}$ is said to be n-linear if f is linear in each its components:

$$f(v_1, \ldots, v_i + cw_i, \ldots, v_n) = f(v_1, \ldots, v_1, \ldots, v_n) + cf(v_1, \ldots, w_i, \ldots, v_n), \qquad i = 1, \ldots, n.$$

For example, $f : \mathbb{R} \times \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ given by $f(x,y,z) = xyz$ is trilinear, or $g : \mathbb{R}^2 \times \mathbb{R}^2 \to \mathbb{R}$ $g([a,b],[c,d]) = ad - bc$ is bilinear.

(a) Let V and W be real vector spaces, and let $L^n(V,W)$ be the set of n-linear maps between V and W. Show there is an isomorphism $L(V, L(V,W)) \cong L^2(V,W)$.

(b) Inductively show there is an isomorphism $L(V, L^{k-1}(V,W)) \cong L^k(V,W)$.

(c) Conclude that $\dim L^k(\mathbb{R}^n, \mathbb{R}^m) = n^k m$.

1-40. Let V be a n-dimensional real vector space. We say that a basis $\mathcal{B} = \{\mathbf{b}_1, \ldots, \mathbf{b}_n\}$ is *positively oriented* if $\det[\mathbf{b}_1, \ldots, \mathbf{b}_n] > 0$, and negatively oriented otherwise. If $\mathbf{D} = \{\mathbf{d}_1, \ldots, \mathbf{d}_n\}$ is another basis, let T be the change of basis transformation from $\mathcal{B}$ to $\mathcal{D}$. We say that T is *orientation preserving* if $\det T > 0$, in which case we say $\mathcal{B}$ and $\mathcal{D}$ are *compatibly oriented.* If $\det T < 0$, we say T is orientation reversing, and that $\mathcal{B}$ and $\mathcal{D}$ are *incompatibly oriented.*

(a) Describe – using non-mathematical terms if you like – the positive orientations in $\mathbb{R}$, $\mathbb{R}^2$, and $\mathbb{R}^3$.

(b) Show that a change of basis T is orientation preserving if and only if T^{-1} is orientation preserving.

(c) Define a relation on the collection of bases of V by saying that $\mathcal{B} \sim \mathcal{D}$ if the change of basis matrix between $\mathcal{B}$ and $\mathcal{D}$ is orientation preserving. Show that this defines an equivalence relation. How many equivalence classes are there?

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

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 analysis, you learned important theorems like the Extreme and Intermediate Value Theorems. While these two theorems normally begin with "Let f be a continuous function on $[a,b]$ ...," it is not necessary in either case to take a closed interval $[a,b]$. For example, the Intermediate Value Theorem applies to a continuous function on $(0,\infty)$, while the Extreme Value Theorem applies to a continuous function on $[0,1] \cup [5,6]$. Take a moment to think about how the conditions on $[a,b]$ can be loosened such that each theorem is still true.

2.1 Open, Closed, and Everything in Between

We begin by looking at the notion of open and closed sets. The idea is to generalize the behaviour of open and closed intervals in $\mathbb{R}$.

2.1.1 Interior and Boundary Points

Definition 2.1

Let (X, d) be a metric space. If $\mathbf{x} \in X$ and $r > 0$, we define the *open ball of radius* r *centred at* $\mathbf{x}$ as

$$B_r(\mathbf{x}) := \{\mathbf{y} \in X : d(\mathbf{x}, \mathbf{y}) < r\}.$$

In $\mathbb{R}^n$ with the Euclidean metric $d(\mathbf{x}, \mathbf{y}) = \|\mathbf{x} - \mathbf{y}\|$, 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 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,

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

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

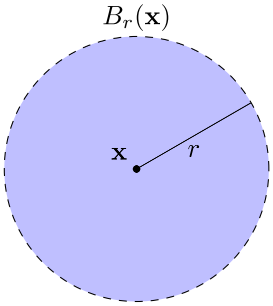


Figure 2.1: In $\mathbb{R}^2$, the open ball of radius r centred at $\mathbf{x}$ consists of all points which are a distance at most r from $\mathbf{x}$.

Definition 2.2

A subspace S of a metric space X is bounded if there exists an $r > 0$ and an $\mathbf{a} \in X$ such that $S \subseteq B_r(\mathbf{a})$.

This should not be surprising: A set is bounded if we can put a ball around it, prohibiting it from growing arbitrarily large. For example, the set $S = \left\{(x, y) \in \mathbb{R}^2 : xy > 0\right\}$ consists of the first and third quadrants of the plane. Since both x and y can become arbitrarily large in absolute value, no ball centred at the origin entirely contains S. On the other hand, $C = \left\{(x, y) \in \mathbb{R}^2 : (x - a)^2 + (y - b)^2 \leq r^2\right\}$ is bounded *for any* choice of $a, b, c \in \mathbb{R}$.

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 2.3

Let (X, d) be a metric space, and $S \subseteq \mathbb{R}^n$ be an arbitrary set.

1. We say that $\mathbf{x} \in S$ 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 $\mathbf{x} \in S$ 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.

The *interior of S* – denoted S^{int} – is the collection of interior points of S, while *boundary of S* – denoted ∂S – is the collection of boundary points of S.

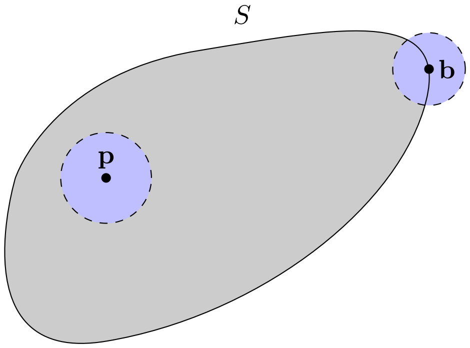


Figure 2.2: The point $\mathbf{b}$ is a boundary point. No matter what size ball we place around $\mathbf{b}$, that ball will intersect both S and S^c. On the other hand, $\mathbf{p}$ is an interior point, since we can place a ball around it which lives entirely within S.

We should take a moment and think about these definitions, and why they make sense. 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. Note that 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 with a simple example.

Example 2.4

Let $S = (-1, 1] \subseteq \mathbb{R}$, endowed with the Euclidean metric. What are the interior points and the boundary points of S?

Solution. I 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 $r = 1/2$ and $B_{1/2}(0) = (-1/2, 1/2) \subseteq (-1,1)$. Thus assume that $p \neq 0$ and let $r = (1-p)/2$, which represents half the distance from p to 1. I 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 2.5

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

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 $\partial\mathbb{Q} = \mathbb{R}$. ■

2.1.2 Open and Closed Sets

Definition 2.6

A set S in a metric space (X, d) 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. Given a point $\mathbf{x} \in X$, an *open neighbourhood* of $\mathbf{x}$ is some open set containing $\mathbf{x}$.

Example 2.7

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

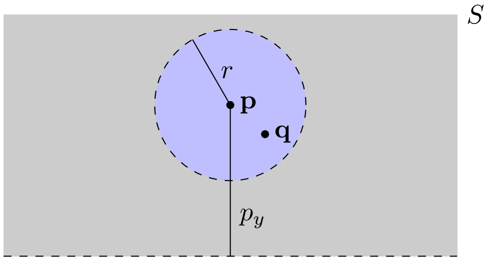


Figure 2.3: The upper half plane is open. For any point, look at its y-coordinate p_y and use the ball of radius $p_y/2$.

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

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

which implies that $q_y > p_y - p_y/2 = p_y/2 > 0$. 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. ■

Theorem 2.8

If (X, d) is a metric space, $\mathbf{x} \in X$, and $r > 0$, then $B_r(\mathbf{x})$ is open. More concisely, open balls are always open.

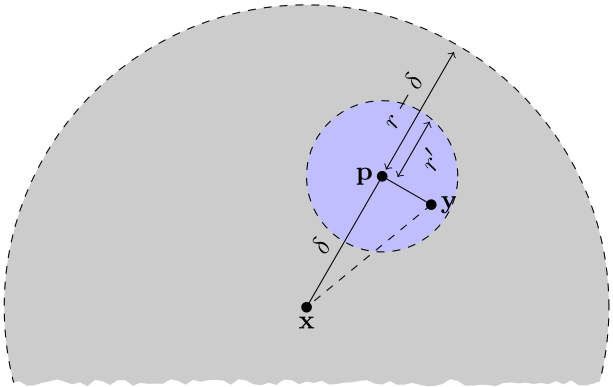


Figure 2.4: A visualization of the solution to Theorem 2.8.

Solution. The name 'open ball' certainly suggests this is the case. To show the result, let $\mathbf{x} \in X$ 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 $\delta = d(\mathbf{x}, \mathbf{p})$ so that $\delta < r$ be definition.

I claim that the open ball of radius $r' = (r-\delta)/2 > 0$ will live inside $B_r(\mathbf{x})$. To see this, choose an arbitrary $\mathbf{y} \in B_{r'}(\mathbf{p})$ so that $d(\mathbf{p},\mathbf{y}) < r'$. One has that

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

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

Theorem 2.9: The Metric Topology

Let (X,d) be a metric space. If $\mathcal{T}$ denotes the collection of open sets in X, then

1. Both X and $\emptyset$ are open; that is, $\emptyset, X \in \mathcal{T}$;
2. Arbitrary unions of open sets are open; that is, if
$$\{\mathcal{O}_i : i \in I\} \subseteq \mathcal{T}, \quad \text{then} \quad \bigcup_{i\in I} \mathcal{O}_i \in \mathcal{T};$$
3. Finite intersections of open sets are open; that is, if
$$\{\mathcal{O}_i : i = 1,\ldots,n\} \subseteq \mathcal{T}, \quad \text{then} \quad \bigcap_{i\in I} \mathcal{O}_i \in \mathcal{T}.$$

Proof. In each case, we choose an arbitrary point and show that it is an interior point.

1. That $\emptyset$ is open is vacuously true. To show that X is open, let $\mathbf{x} \in X$ and choose any $r > 0$. Then $B_r(\mathbf{x}) \subseteq X$, showing that $\mathbf{x}$ is an interior point.

2. Suppose $\{\mathcal{O}_i : i \in I\} \subseteq \mathcal{T}$ and let $\mathcal{O} = \bigcup_{i\in I} \mathcal{O}_i$. Fix an $\mathbf{x} \in \mathcal{O}$, in which case there exists an $i \in I$ such that $\mathbf{x} \in \mathcal{O}_i$. As $\mathcal{O}_i$ is open, there exists an $r > 0$ such that $B_r(\mathbf{x}) \subseteq \mathcal{O}_i$, in which case $B_r(\mathbf{x}) \subseteq \mathcal{O}$, showing that $\mathbf{x}$ is an interior point of $\mathcal{O}$.

3. Suppose $\{\mathcal{O}_i : i = 1,\ldots,n\} \subseteq \mathcal{T}$, and let $\mathcal{O} = \bigcap_{i=1}^n \mathcal{O}_i$. Fix an $\mathbf{x} \in \mathcal{O}$, in which case $\mathbf{x} \in \mathcal{O}_i$ for each $i \in \{1,\ldots,n\}$. As each $\mathcal{O}_i$ is open, there exist a collection of positive real numbers r_i such that $B_{r_i}(\mathbf{x}) \subseteq \mathcal{O}_i$. Let $r = \min\{r_1,\ldots,r_n\}$ so that $B_r(\mathbf{x}) \subseteq B_{r_i}(\mathbf{x}) \subseteq \mathcal{O}_i$. This shows that $B_r(\mathbf{x}) \subseteq \mathcal{O}$ and hence $\mathbf{x}$ is an interior point of $\mathcal{O}$ as required. □

Theorem 2.9 is representative of a much larger field within mathematics, called *Topology*. A topology $\mathcal{T}$ on a set X is a collection of open sets $\mathcal{T}$ which satisfy the listed properties: The empty set and X are both in $\mathcal{T}$, and $\mathcal{T}$ is closed under arbitrary unions and finite intersections. A metric on a space always defines a topology, but there are many topologies (even on $\mathbb{R}$) which cannot be defined by means of a metric.

In addition, the specification that $\mathcal{T}$ be closed under *finite* unions is essential. Note for example that for any $n \in \mathbb{N}$ the set $I_n = (-1/n, 1/n)$ is open in $\mathbb{R}$ with respect to the Euclidean metric, but

$$\bigcap_{n=1}^{\infty} I_n = \bigcap_{n=1}^{\infty} \left(-\frac{1}{n}, \frac{1}{n}\right) = \{0\}$$

which is a closed set.

Definition 2.10

Let (X, d) be a metric space, and $Y \subseteq X$. We say that a set $U \subseteq Y$ is *open in* Y if $U = V \cap Y$ for some open set $V \in Y$. The set U is *closed in* Y if its complement *in* Y, $U^c = \{\mathbf{x} \in Y : \mathbf{x} \notin U\}$ is open in Y.

Note that open sets relative to a subspace need not be open in the ambient space. For example, if $X = \mathbb{R}$ and d is the Euclidean metric, let $Y = [0, 2]$. Writing $[0, 1) = (-1, 1) \cap [0, 2]$ shows that $[0, 1)$ is open in Y, though it's not open in $\mathbb{R}$.

Defining relatively open and closed sets allows us to think of (Y, d) as a metric space in its own right. You might ask why we don't just restrict the metric to the subspace Y. It turns out that it doesn't matter if you do or not, they both define the same open sets (Exercise 2-12).

2.1.3 Closures

Just as in the case of intervals in $\mathbb{R}$, it is possible for a set to be neither open nor closed. In Exercise 2-1 you will show 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 the set S contains all of its boundary points, Proposition 2.11 shows that it's 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.

Proposition 2.11

A set S in (X, d) 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 $\mathbf{x} \in \partial S$ implies that 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 $\partial S \nsubseteq S$. □

Definition 2.12

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

The closure of an interval (a, b) in the Euclidean metric is the closed interval $\overline{(a, b)} = [a, b]$. Similarly, the closure of the open half plane in Example 2.7 is the closed half plane

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

You should check that the closure of the open ball in X is

$$\overline{B_r(\mathbf{x})} = \overline{\{\mathbf{y} \in \mathbb{R}^n : d(\mathbf{x}, \mathbf{y}) < r\}} = \{\mathbf{y} \in \mathbb{R}^n : d(\mathbf{x}, \mathbf{y}) \leq r\}.$$

Proposition 2.13

Suppose (X, d) is a metric space with $S \subseteq X$.

1. $\overline{S}$ is always a closed set,
2. S is closed if and only if $S = \overline{S}$,
3. $\overline{S} = \bigcap \{C : S \subseteq C, C \text{ closed}\}$,
4. $\overline{S}$ is the *smallest* closed set containing S.

Proof. There is a nicer characterization of closed sets that we'll learn in the next section, but for now we're forced to use the open ball definition.

1. By Exercise 2-16c, we know that $(\overline{S})^c = (S^c)^{\text{int}}$, and the right hand side, being a collection of interior points, is always open. Hence $\overline{S}$ is closed as required.
2. Suppose S is closed, so that Proposition 2.11 implies $\partial S \subseteq S$. But then $\overline{S} = S \cup \partial S = S$ as required. Conversely, if $S = \overline{S}$ then S is closed by part 1.
3. It suffices to show that $\overline{S}$ is contained in every closed set containing S. If C is a closed set such that $S \subseteq C$, then $\overline{S} \subseteq \overline{C} = C$ (part 2 and Exercise 2-15), which is what we wanted to show.
4. Let $\mathcal{K} = \{C : S \subseteq C, C \text{ closed}\}$, for which we want to show that $\overline{S} = \cap \mathcal{K}$. Since $S \subseteq \overline{S}$ and $\overline{S}$ is closed, $\overline{S} \in \mathcal{K}$ and hence $\cap \mathcal{K} \subseteq \overline{S}$. Conversely, by part 3 we know that $\overline{S} \subseteq C$ for all $C \in \mathcal{K}$, so $\overline{S} \subseteq \cap \mathcal{K}$. Both inclusions give the desired equality. □

2.2 Sequences and Completeness

You should be familiar with sequences in $\mathbb{R}$, and much of the terminology and ideas from that study will translate over to general metric spaces. However, sequences provide a useful tool for studying topology, and in particular play well with closed sets.

2.2.1 Sequences in Metric Space

The rough idea of a sequence is that it represents and ordered collection of elements in your space. More formally, a *sequence* in a space X is any function $\mathbf{x} : \mathbb{N} \to X$, so that $x(n) \in X$. The sequence then inherits the ordering of $\mathbb{N}$. We will often write sequence elements 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

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

We often choose to conflate the function $\mathbf{x}$ itself with its ordered image in X, in which case we write the sequence as $(\mathbf{x}_n)_{n=1}^{\infty}$. When we're feeling particularly lazy or it's of no consequence, we'll even omit the indexing and write $(\mathbf{x}_n)$.

A *subsequence* is a sequence derived from another sequence which preserves the original ordering. If $n : \mathbb{N} \to \mathbb{N}$ is a strictly increasing function and $\mathbf{x} : \mathbb{N} \to X$, then one can define a subsequence by $\mathbf{x}(k(n)) = \mathbf{x}_{k_n}$. For example, if $n(k) = 3k - 1$, then

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

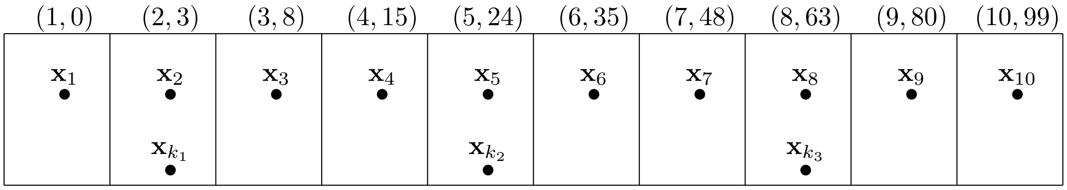


Figure 2.5: The sequence $\mathbf{x}_n = (n, n^2 - 1)$ and its subsequence defined by the strictly increasing map $n(k) = 3k - 1$; that is, $\mathbf{x}_{n_k} = (3k - 1, 9k^2 - 6k)$.

Remark 2.14 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 $\big(x^k{}_n\big)_{n=1}^{\infty}$. If $(\mathbf{x}_{n_\ell})_{\ell=1}^{\infty}$ is a subsequence, this defines subsequences $(x_{n_\ell}^k)_{\ell=1}^{\infty}$ (we are running out of letters!). However, notice that the converse is certainly not true: One cannot take subsequences of each (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 $y_{n_s} = -(2n - 1)$.

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

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

Definition 2.15

A sequence $(\mathbf{x}_n)$ in a metric space (X, d) is *bounded* if its image set $\{\mathbf{x}_n : n \in \mathbb{N}\}$ is bounded as a set.

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

Definition 2.16

Let $(\mathbf{x}_n)_{n=1}^\infty$ be a sequence in a metric space (X, d). We say that $(\mathbf{x}_n)$ converges with limit $\mathbf{x} \in X$, 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 $d(\mathbf{x}_n, \mathbf{x}) < \epsilon$.

The basic theorems from $\mathbb{R}$, such as uniqueness of limits and the limit laws, still hold true in general.

Theorem 2.17

If $(\mathbf{x}_n)$ is a sequence in a metric space (X, d) such that $(\mathbf{x}_n) \to \mathbf{x}$ and $(\mathbf{x}_n) \to \mathbf{y}$, then $\mathbf{x} = \mathbf{y}$.

Proof. By Exercise 2-3, it suffices to show that for every $\epsilon > 0$, $d(\mathbf{x}, \mathbf{y}) < \epsilon$. Let $\epsilon > 0$ be arbitrary, and fix $N_1, N_2 \in \mathbb{N}$ such that

$$d(\mathbf{x}_n, \mathbf{x}) < \frac{\epsilon}{2} \text{ whenever } n > N_1 \quad \text{and} \quad d(\mathbf{x}_n, \mathbf{y}) < \frac{\epsilon}{2} \text{ whenever } n > N_2.$$

Let $N = \max\{N_1, N_2\}$, so that if $n > N$ then

$$d(\mathbf{x}, \mathbf{y}) \leq d(\mathbf{x}, \mathbf{x}_n) + d(\mathbf{x}_n, \mathbf{y}) = \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon,$$

which is what we wanted to show. □

Theorem 2.18: Limit Laws

Suppose $(X, \|\cdot\|)$ is a normed real vector space with sequences $(\mathbf{x}_n)$ and $(\mathbf{y}_n)$. If $\mathbf{x}_n \to \mathbf{x}$ and $\mathbf{y}_n \to \mathbf{y}$ in the induced metric, then

1. $(\mathbf{x}_n + \mathbf{y}_n)$ converges and $(\mathbf{x}_n + \mathbf{y}_n) \to \mathbf{x} + \mathbf{y}$,
2. $(\alpha \mathbf{x}_n)$ converges for any $\alpha \in \mathbb{R}$, and $(\alpha \mathbf{x}_n) \to \alpha \mathbf{x}$.

Proof. I'll prove (1) and leave (2) to Exercise 2-19. Suppose $\mathbf{x}_n \to \mathbf{x}$ and $\mathbf{y}_n \to \mathbf{y}$. Fix an $\epsilon > 0$ and choose N_1, N_2 such that

$$\|\mathbf{x}_n - \mathbf{x}\| < \frac{\epsilon}{2} \text{ whenever } n > N_1 \quad \text{and} \quad \|\mathbf{y}_n - \mathbf{y}\| < \frac{\epsilon}{2} \text{ whenever } n > N_2.$$

Let $N = \max\{N_1, N_2\}$, so that if $n > N$ then

$$\|(\mathbf{x}_n + \mathbf{y}_n) - (\mathbf{x} + \mathbf{y})\| \leq \|\mathbf{x}_n - \mathbf{x}\| + \|\mathbf{y}_n - \mathbf{y}\| < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon,$$

showing that $(\mathbf{x}_n + \mathbf{y}_n) \to \mathbf{x} + \mathbf{y}$ as required. □

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 (Figure 2.6)."

Note that this is different than saying that every ball contains infinitely many points, as evidenced by Exercise 2-20.

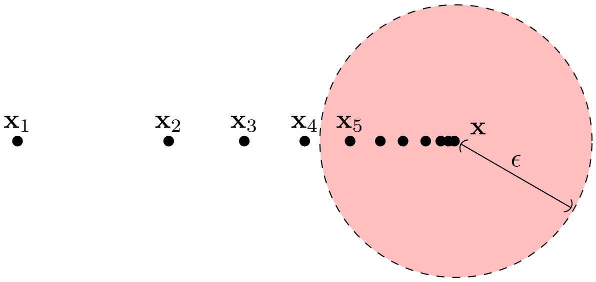


Figure 2.6: A sequence $\mathbf{x}_n$ converges to the point $\mathbf{x}$ if any ϵ ball around $\mathbf{x}$ contains all but finitely many points of the sequence.

Example 2.19

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

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

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

Proposition 2.20

If $(\mathbf{x}_n)$ is a sequence in the metric space (X, d), then $(\mathbf{x}_n) \to \mathbf{x}$ if and only if every subsequence $(\mathbf{x}_{n_k}) \to \mathbf{x}$ as well.

Proof. Suppose that $(\mathbf{x}_n) \to \mathbf{x}$ and fix a subsequence $\mathbf{x}_{n_k}$. In particular, the function $n : \mathbb{N} \to \mathbb{N}$ is strictly increasing, so if $k_1 < k_2$ then $n(k_1) < n(k_2)$. Let $\epsilon > 0$ and fix an $N \in \mathbb{N}$ such that $d(\mathbf{x}_n, \mathbf{x}) < \epsilon$ whenever $n > N$. If $k > N$ then $n(k) > n(N) > N$ and so $d(\mathbf{x}_{n_k}, \mathbf{x}) < \epsilon$ as well.

The other direction is trivial, as $(\mathbf{x}_n)$ is a subsequence of itself, which is assumed to converge. □

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 2.21

If S is subspace of a metric space (X, d), then $x \in \overline{S}$ if and only if there exists a convergent sequence $(\mathbf{x}_n)$ in S such that $(\mathbf{x}_n) \to \mathbf{x}$.

Proof. [$\Rightarrow$] 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 2.7). 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 $1/N < \epsilon$. When $n > N$ we have $1/n < 1/N < \epsilon$, and by construction $\mathbf{x}_n \in B_{1/n}(\mathbf{x}) \subseteq B_\epsilon(\mathbf{x})$, or equivalently

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

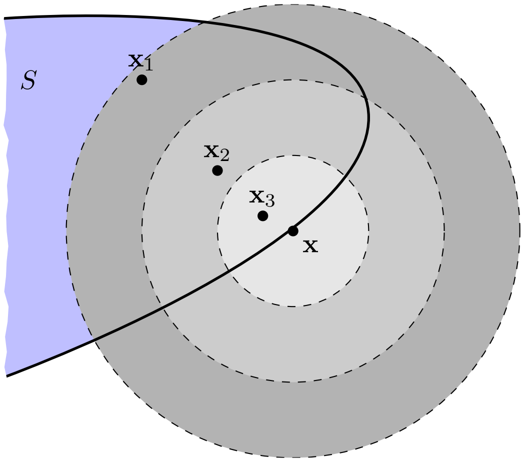


Figure 2.7: In the proof of Proposition 2.21, we need to construct a sequence which converges to $\mathbf{x}$. This is done by constructing the ball of radius $1/n$ around $\mathbf{x}$, which must intersect S. Choosing points in these successively smaller balls constructs a sequence $\mathbf{x}_n$ which converges to $\mathbf{x}$.

[$\Leftarrow$] 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, $\mathbf{x} \in \overline{S}$. □

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

Corollary 2.22

A subspace S of a metric space 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 $\mathbf{x} \in S$.

Hence we see that limits of sequences are convenient for determining whether a set is closed. This is in stark contrast to the open ball approach, whereby sets were closed if their complements were open. For our purposes, when we want to say something about closed sets we will generally use sequences, and when we want to say something about open sets, we'll generally use open balls.

2.2.2 Completeness

One of the problems with discussing convergence of a sequence is that you must have a candidate limit point in mind. To discuss convergent sequences without the need for a limiting value, we try to capture the idea of a convergent sequence with an alternate definition.

Definition 2.23

If (X, d) is a metric space, we say that a sequence $(\mathbf{x}_n)$ is a *Cauchy sequence* if for every $\epsilon > 0$, there exists an $N \in \mathbb{N}$ such that $d(\mathbf{x}_n, \mathbf{x}_m) < \epsilon$ whenever $n, m > N$.

Proposition 2.24

If $(\mathbf{x}_n)$ is a convergent sequence in (X, d), then $(\mathbf{x}_n)$ is Cauchy.

Proof. Suppose $(\mathbf{x}_n) \to \mathbf{x}$ and fix an $\epsilon > 0$. Choose an $N \in \mathbb{N}$ such that $d(\mathbf{x}_n, \mathbf{x}) < \epsilon/2$ if $n > N$. This same N shows that $(\mathbf{x}_n)$ is Cauchy, for if $m, n > N$ then

$$d(\mathbf{x}_n, \mathbf{x}_m) \leq d(\mathbf{x}_n, \mathbf{x}) + d(\mathbf{x}, \mathbf{x}_m) < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon. \qquad \square$$

The converse generally need not be true. For example, let $X = (0, \infty)$ be endowed with the Euclidean metric. The sequence $x_n = 1/n$ is certainly Cauchy, but does not converge to any point in X. For a more exotic example, let $X = \mathbb{R}$ and $d(x, y) = |e^x - e^y|$. Note that the sequence $x_n = -n$ is a Cauchy sequence (check this), but does not converge.

2.2.3 Completeness of Euclidean $\mathbb{R}^n$

Our goal in this section is to show that $\mathbb{R}^n$ with the Euclidean metric is in fact complete, and for the remainder of this section we will work exclusively with the Euclidean metric. Recall the Monotone Convergence Theorem:

Theorem 2.25: Monotone Convergence Theorem

If (a_n) is a sequence of real numbers, 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 $L = \sup_n \{a_n : N \in \mathbb{N}\}$, 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. □

Theorem 2.26: 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,

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

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 2.27

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

Proof. The idea of the proof will be to exploit Theorem 2.26 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.

Proposition 2.28

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 2.14).

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 2.20 implies that $(x_{n_{k_\ell}}) \to x$. By Exercise 2-18, since each component converges, $(x_{n_{k_\ell}}, y_{n_{k_\ell}})$ converges, as required. □

Proposition 2.29

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.

Solution. One direction has already been done, so it suffices to show that every Cauchy sequence converges. 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 2.28, $(\mathbf{x}_n)$ 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 $\|\mathbf{x}_k - \mathbf{x}_\ell\| < \epsilon/2$. Similarly, choose $K \in \mathbb{N}$ such that for all $k > K$ we have $\|\mathbf{x}_{n_k} - \mathbf{x}\| < \epsilon/2$. Fix an integer $k > N_1$ such that $n_k > K$ so that if $n > K$ we have

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

2.3 Continuity and Limits

Studying the functions on a space can reveal interesting information about the space itself. When we discussed functions between vector spaces, we demanded that those functions preserved the vector space structures, giving rise to the linear maps. In topology, we wish our functions to preserve the open sets. You may be surprised to learn that it is continuous functions which preserve the desired structure. Recall the definition of a continuous in a single variable:

Definition 2.30

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.

Let's break this down into the language of metrics and balls. The Euclidean metric on $\mathbb{R}$ is $d(x, y) = |x - y|$, and so the set $0 < |x - c| < \delta$ is $B_\delta^0(c) := B_\delta(c) \setminus \{c\}$, known as a *deleted neighbourhood of* c. Similarly, $|f(x) - L| < \epsilon$ corresponds to the ball $B_\epsilon(L)$ in the codomain. Recall that removing the condition of a deleted neighbourhood – that is, writing $|x - c| < \delta$ rather than $0 < |x - c| < \delta$ – is equivalent to continuity. Namely, $f : \mathbb{R} \to \mathbb{R}$ is continuous at c if for every $\epsilon > 0$ there exists a $\delta > 0$ such that $f(B_\delta(c)) \subseteq B_\epsilon(f(c))$.

We can generalize this to arbitrary metric spaces as follows:

Definition 2.31

Let (X, d_X) and (Y, d_Y) be metric spaces. If $f : X \to Y$, we say that *the limit of f as $\mathbf{x}$ approaches $\mathbf{c}$ is $\mathbf{L}$*, written as $\lim_{\mathbf{x} \to \mathbf{c}} f(\mathbf{x}) = \mathbf{L}$, if for every $\epsilon > 0$ there exists a $\delta > 0$ such that

$$0 < d_X(\mathbf{x}, \mathbf{c}) < \delta \quad \Rightarrow \quad d_Y(f(\mathbf{x}), \mathbf{L}) < \epsilon.$$

We say that f is *continuous* at $\mathbf{c}$ if for every $\epsilon > 0$ there exists a $\delta > 0$ such that

$$d_X(\mathbf{x}, \mathbf{c}) < \delta \quad \Rightarrow \quad d_Y(f(\mathbf{x}), f(\mathbf{c})) < \epsilon.$$

If f is continuous at every point in its domain, we will simply say that f is continuous.

In the language of balls, $f(\mathbf{x}) \to \mathbf{L}$ if for every $\epsilon > 0$ there exists a $\delta > 0$ such that $f\left(B_\delta^0(\mathbf{c})\right) \subseteq B_\epsilon(\mathbf{L})$, where it is understood that the balls are with respect to different metrics. Similarly, continuity at $\mathbf{c}$ is equivalent to the statement that for every $\epsilon > 0$ there is a $\delta > 0$ such that $f(B_\delta(\mathbf{c})) \subseteq B_\epsilon(f(\mathbf{c}))$.

Another way of thinking about continuity is 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 is best seen with sequences.

Theorem 2.32

Let (X, d_X) and (Y, d_Y) be two metric spaces. A function $f : X \to Y$ is continuous at $\mathbf{a}$ if and only if whenever $(\mathbf{a}_n)_{n=1}^\infty \to \mathbf{a}$ is a convergent sequence in X, then $(f(\mathbf{a}_n))_{n=1}^\infty \to f(\mathbf{a})$ is a convergent sequence in Y.

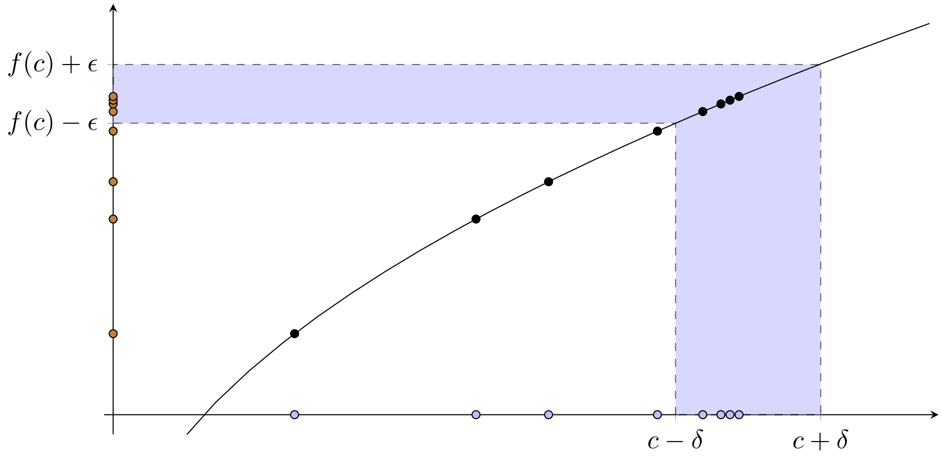


Figure 2.8: If $(a_n) \to a$, then by going far enough into our sequence (blue) we can guarantee that we will be in δ-neighbourhood of a. The image of these points are the $f(a_n)$ (brown), which live in the desired ϵ-neighbourhood because of the continuity of f.

Proof. [⇒] Assume that 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 f is continuous at $\mathbf{a}$, there exists a $\delta > 0$ such that $d_Y(f(\mathbf{x}), f(\mathbf{a})) < \epsilon$ whenever $d_X(\mathbf{x}, \mathbf{a}) < \delta$. Since $(\mathbf{a}_n)$ is convergent, there exists an $N \in \mathbb{N}$ such that $d_X(\mathbf{a}_n, \mathbf{a}) < \delta$ for all $n \geq N$. Combining these, we see that whenever $n \geq N$ we have

$$d_X(\mathbf{a}_n, \mathbf{a}) < \delta \quad \text{and so} \quad d_Y(f(\mathbf{a}_n), f(\mathbf{a})) < \epsilon.$$

which is exactly what we want to show.

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

Theorem 2.32 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

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

As alluded to earlier, continuous functions preserve the open sets between two metric spaces, as evidenced below:

Theorem 2.33

Let (X, d_X) and (Y, d_Y) be metric spaces. A function $f : X \to Y$ is continuous if and only if whenever $U \subseteq Y$ is an open set, then $f^{-1}(U) \subseteq X$ is an open set.

The above definition assumes f is continuous on all of X. There is a corresponding pointwise notion of continuity for topological spaces, but it's not as useful as the global continuity condition, so we omit it and leave it to the exercises.

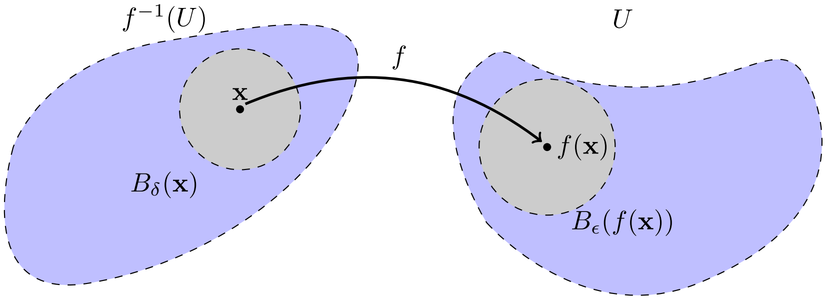


Figure 2.9: To show that the pre-image of open sets is open, we use the fact that the condition $d_Y(f(\mathbf{x}), f(\mathbf{y})) < \epsilon$ is exactly the same thing as looking at an ϵ-ball around $f(\mathbf{x})$.

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

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

Definition 2.34

If (X, d_X) and (Y, d_Y) are metric spaces, we say that X and Y are *homeomorphic* if there exist a continuous function $f : X \to Y$ which admits a continuous inverse $f^{-1} : Y \to X$. In this case, f is said to be a *homeomorphism.*

Homeomorphisms are 'isomorphisms' for topological spaces, and are sometimes called *bicontinuous* maps. Indeed, since $f : X \to Y$ admits an inverse, it is automatically bijective. Moreover, as both f and f^{-1} are continuous, open sets are perfectly preserved by f. This means that X and Y are identical as topological spaces.

With vector spaces, a bijective linear function was automatically an isomorphism, but the same is not true for topological spaces: There are continuous bijective maps whose inverses are not continuous. In particular, it is easy to create such functions by letting $f : X \to X$, and endowing the domain with a metric which has strictly more open sets.

2.3.1 Uniform Continuity

Note that continuous functions do not preserve Cauchy sequences. A straightforward example is to take $f : (0, \infty) \to \mathbb{R}$ and $x_n = 1/n$ with the Euclidean metrics in $\mathbb{R}$. The sequence (x_n) is certainly Cauchy, but $f(x_n) = n$ which is not Cauchy.

Stronger than continuity, there is a notion of uniform continuity which plays nicer with Cauchy sequences than simple continuous functions. The idea is as follows: Suppose (X, d_X) and (Y, d_Y) are metric spaces, with $f : X \to Y$ continuous everywhere on X. The ϵ-δ definition of continuity is

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

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 2.10). 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 2.35

Let (X, d_X) and (Y, d_Y) be metric spaces, and $f : X \to Y$. We say that f is *uniformly continuous* if for every $\epsilon > 0$, there exists a $\delta > 0$ such that $d_Y(f(\mathbf{x}), f(\mathbf{y})) < \epsilon$ **for every** $\mathbf{x}, \mathbf{y} \in X$ satisfying $d_X(\mathbf{x}, \mathbf{y}) < \delta$.

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. In fact, while continuity is defined at a point, uniform continuity is defined on a set. Intuitively, uniformly continuous function are in some sense bounded in how quickly they are permitted to grow. The following two examples both use the Euclidean metric in $\mathbb{R}^n$.

Example 2.36

The function $f : \mathbb{R}^2 \to \mathbb{R}$ given by $f(x) = 2x + 5y - 4$ is uniformly continuous.

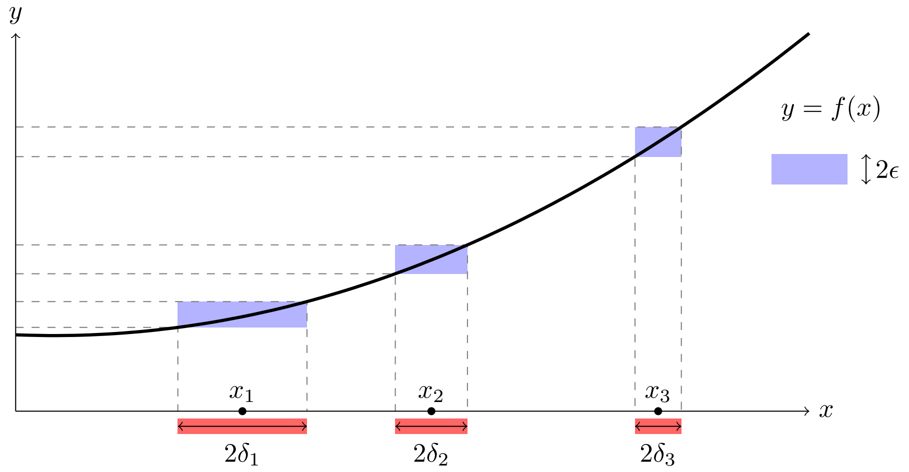


Figure 2.10: For a fixed $\epsilon > 0$, the value of δ depends on the choice of the point x. In fact, the faster a function grows at a point, the smaller the corresponding δ will be.

Solution. Let $\epsilon > 0$ and choose $\delta = \epsilon/10$. Let $(x_0, y_0), (x_1, y_1) \in \mathbb{R}^2$ be any points such that $\|(x_0, y_0) - (x_1, y_1)\| = \sqrt{(x_0 - x_1)^2 + (y_0 - y_1)^2} < \delta$, and notice that

$$\begin{aligned} |f(x_0, y_0) - f(x_1, y_1)| &\leq 2|x_0 - x_1| + 5|y_0 - y_1| \leq 5|x_0 - x_1| + 5|y_0 - y_1| \\ &\leq 10\sqrt{(x_0 - x_1)^2 + (y_0 - y_1)^2} < \epsilon, \end{aligned}$$

as required. ■

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

Example 2.37

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 $\delta = \epsilon/4$. 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

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

Moreover,

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

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 $|x| > 1/\delta$.

Proposition 2.38

If (X, d_X) and (Y, d_Y) are metric spaces, $(\mathbf{x}_n)$ is a Cauchy sequence in X, and $f : X \to Y$ is uniformly continuous, then $f(\mathbf{x}_n)$ defines a Cauchy sequence in Y.

Proof. Let $\epsilon > 0$ be given, and choose a $\delta > 0$ such that $d_Y(f(\mathbf{x}), f(\mathbf{y})) < \epsilon$ whenever $d_X(\mathbf{x}, \mathbf{y}) < \delta$. Since $(\mathbf{x}_n)$ is a Cauchy sequence, there exists an $M \in \mathbb{N}$ such that $d_X(\mathbf{x}_n, \mathbf{x}_m) < \delta$ whenever $m, n > N$. Hence if $m, n > M$ then

$$d_X(\mathbf{x}_n, \mathbf{x}_m) < \delta \quad \Rightarrow \quad d_Y(f(\mathbf{x}_n), f(\mathbf{x}_m)) < \epsilon,$$

showing that $(f(\mathbf{x}_n))$ is a Cauchy sequence in Y. □

2.3.2 Limits and Continuity in $\mathbb{R}^n$

So continuity is a topological notion, depending only upon open sets. However, we'll see that differentiability is more rigid than that, and really requires that we focus on normed vector spaces. To this end, we'll take a moment to see what happens when $f : \mathbb{R}^n \to \mathbb{R}^m$, and both spaces are endowed with the Euclidean metric/norm.

Naturally, translating the definition of the limit when our metric spaces are $\mathbb{R}^n$ gives

> "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 $\|\mathbf{f}(\mathbf{x}) - \mathbf{L}\| < \epsilon$."

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 $\mathbb{R}^m$.

Example 2.39

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

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

$$|a| \leq \sqrt{a^2 + b^2}, \qquad |b| \leq \sqrt{a^2 + b^2}. \tag{2.1}$$

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

$$\begin{aligned} |(x + y) - 2| &= |(x - 1) + (y - 1)| \leq |x - 1| + |y - 1| \\ &\leq \sqrt{(x - 1)^2 + (y - 1)^2} + \sqrt{(x - 1)^2 + (y - 1)^2} && \text{by (2.1)} \\ &= 2\|(x, y) - (1, 1)\| < \epsilon. \end{aligned}$$

■

Example 2.40

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

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

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

■

Example 2.41

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 $\delta = \epsilon/\sqrt{3}$. 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

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

■

Recall that

$$\lim_{x\to c} f(x) \text{ exists} \qquad \Leftrightarrow \qquad \lim_{x\to c^+} f(x) \text{ and } \lim_{x\to c^-} f(x) \text{ exist and are equal.}$$

This represents the fact that the limit exists if and only if the limit is the same regardless of which path we 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 2.40 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 2.11). 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 exist, and they must all be equal.

To see this more concretely, let $f : \mathbb{R}^2 \to \mathbb{R}$ be a function, whose limit we wish to evaluate at $\mathbf{0} = (0, 0)$. For any $a, b \in \mathbb{R}$ define $\gamma_{a,b} : [-\epsilon, \epsilon] \to \mathbb{R}^2$ by $\gamma(t) = (at, bt)$. The graph of γ corresponds to the curve $ay - bx = 0$, which is a line through the origin. You can check that this function is continuous on its domain. Hence if $\lim_{x\to\mathbf{c}} f(\mathbf{x})$ exists – with limit say L – then

$$\lim_{t\to 0} f(\gamma_{a,b}(t)) = L$$

holds for any $a, b \in \mathbb{R}$. By contrapositive, if there exists two pair (a_1, b_1) and (a_2, b_2) such that

$$\lim_{t\to 0} f(\gamma_{a_1,b_1}(t)) \neq \lim_{t\to 0} f(\gamma_{a_2,b_2}(t))$$

then the limit does not exist. You will generalize this result in Exercise 2-32.

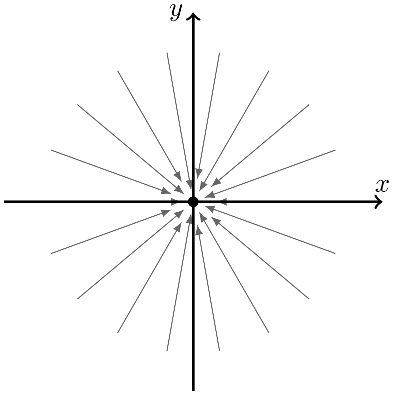


Figure 2.11: Even in $\mathbb{R}^2$, there are infinitely many ways of approaching a point. For a limit to exist, the limit along each path must exist and must be equal to that achieved from every other path.

Example 2.42

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 be the given function, and note that the path $y = mx$ can be written as $\gamma_m(t) = (t, mt)$. Composing the functions gives

$$\begin{aligned}\lim_{t\to 0} f(\gamma_m(t)) &= \lim_{t\to 0} \frac{t^2(mt)^2}{t^4+(mt)^4} = \lim_{t\to 0}\frac{m^2t^4}{t^4+m^4t^4} \\ &= \lim_{t\to 0}\frac{m^2t^4}{(m^4+1)t^4} = \lim_{t\to 0}\frac{m^2}{m^4+1} \\ &= \frac{m^2}{m^4+1}.\end{aligned}$$

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

You 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 $ax + by = 0$ then the limit can be guaranteed to exist? The following examples shows that this is not the case.

Example 2.43

Show that the function $f(x,y) = \dfrac{2xy^2}{x^2+y^4}$ admits a limit along every line $ax - by = 0$, but fails along the parabola $x = my^2$.

Solution. Proceeding as suggested, we take the limit along the lines $\gamma_{a,b}(t) = (at, bt)$:

$$\lim_{t\to 0} f(\gamma_{a,b}(t)) = \lim_{t\to 0}\frac{2(at)(bt)^2}{(at)^2+(bt)^4} = \lim_{t\to 0}\frac{2ab^2t^3}{a^2t^2+b^4t^4} = \lim_{t\to 0}\frac{2abt}{a^2+b^4t^2} = 0.$$

On the other hand, along the line $x = my^2$ we write $\psi_m(t) = (mt^2, t)$ to get

$$\lim_{t\to 0} f(\psi_m(t)) = \lim_{t\to 0}\frac{2(mt^2)t^2}{(mt^2)^2+t^4} = \lim_{t\to 0}\frac{2mt^4}{m^2t^4+t^4} = \lim_{t\to 0}\frac{2m}{m^2+1} = \frac{2m}{m^2+1}$$

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

There is a multivariate version of the Squeeze Theorem as well, which can be used to make short work of some functions.

Theorem 2.44: 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

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

The proof is nearly identical to that of the single variable Squeeze theorem, so I leave it as an exercise.

Example 2.45

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

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

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

Example 2.46

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

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.[2] $\blacksquare$

If $S \subseteq \mathbb{R}^n$, then continuity of functions $f : S \to \mathbb{R}^m$ is as simple as stating that

$$\lim_{x\to\mathbf{c}} f(\mathbf{x}) = f(\mathbf{c}) \quad \text{for all } \mathbf{c} \in S.$$

For example, the function $f(x,y) = (y^4\sin^2(xy))/(x^2+y^2)$ from Example 2.46 is undefined at $(0,0)$, but if we define

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

then g is a continuous function.

Example 2.47

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

[2] 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 $f(\mathbf{x}) \xrightarrow{\mathbf{x}\to c} 0$.

Solution. This is the same set as in Example 2.7, 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 $f(x, y) = y$. Convince yourself 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. ■

2.4 Compactness

In our study of analysis 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.

The idea of a compact set is one which generalizes these useful properties. The definition will seem odd at first, but we'll see later that the following definition is exactly what makes closed and bounded intervals so useful. We first begin with the notion of a cover.

Definition 2.48

Let (X, d_X) be a metric space, and $K \subseteq X$. An *open cover* for K is a collection of open sets $\mathcal{C} = \{U_i : U_i \text{ open}\}$ such that $K \subseteq \cup\mathcal{C}$. A *subcover* of a cover $\mathcal{C}$ is a collection of subsets $\mathcal{S} \subseteq \mathcal{C}$ such that $K \subseteq \cup\mathcal{S}$.

For example, let $\mathbb{D}^2 = \left\{(x, y) \in \mathbb{R}^2 : x^2 + y^2 < 1\right\}$ be the open disk in $\mathbb{R}^2$, and define $\mathcal{C}_1 = \left\{B_{1/n}(\mathbf{0}) : n \in \mathbb{N}\right\}$. This is an open cover of $\mathbb{D}^2$, and $\mathcal{C}_1{}' = \{B_1(\mathbf{0})\}$ is a subcover. Alternatively, for each $\mathbf{x} \in \mathbb{D}^2$, let $\rho(\mathbf{x}) = 1 - \|\mathbf{x}\|$, so that $\rho(\mathbf{x})$ is the distance from the point $\mathbf{x}$ to the boundary of the disk. Define $U_{\mathbf{x}} = B_{\rho(\mathbf{x})}(\mathbf{x})$. The set $\mathcal{C}_2 = \left\{U_{\mathbf{x}} : \mathbf{x} \in \mathbb{D}^2\right\}$ is certainly an open cover of $\mathbb{D}^2$, and $\mathcal{C}_2{}' = \left\{U_{\mathbf{x}} : \mathbf{x} \in \mathbb{D}^2 \cap (\mathbb{Q} \times \mathbb{Q})\right\}$ is a subcover.

Definition 2.49

A set K in a metric space (X, d_X) is *compact* if every open cover of K admits a finite subcover.

For example, the set $I = (0, 1)$ is not compact in $\mathbb{R}$ under the Euclidean metric. Indeed, $\mathcal{C} = \{(1/n, 1) \subseteq \mathbb{R} : n \in \mathbb{N}\}$ is an open cover of I, but no finite collection of subsets of $\mathcal{C}$ could possibly cover I.

Proposition 2.50

Intervals of the form $[a, b] \subseteq \mathbb{R}$ are compact under the Euclidean metric.

Proof. Let $\mathcal{C} = \{U_i : i \in I\}$ be an open cover of $[a, b]$, and let

$$V = \{x \in [a, b] : [a, x] \text{ is covered by a finite subcover of } \mathcal{C}\}.$$

Note that $a \in V$, since $\mathcal{C}$ is a covering for $[a, b]$ and hence there is at least one U_i such that $a \in U_i$, and this is a finite subcover of $[a, a] = \{a\}$. Hence V is non-empty. Moreover, this set is bounded from above by b, so by the Completeness Axiom, $r = \sup V$ exists. It is therefore enough to show that $r = b$.

For the sake of contradiction, assume this is not the case; namely, $r < b$. First, we note that $r \in S$. Indeed, since $\mathcal{C}$ covers $[a, b]$, there is some element $\hat{U} \in \mathcal{C}$ such that $r \in \hat{U}$. Since $\hat{U}$ is open, we can find an $\epsilon > 0$ such that $(r - \epsilon, r + \epsilon) \subseteq \hat{U}$. If necessary, make ϵ small enough to ensure that $r + \epsilon < b$. By definition of the least upper bound, there is some element $\xi \in (r - \epsilon, r]$ which is also in S, and so $[a, \xi]$ is covered by finitely many elements $\{U_{i_1}, \ldots, U_{i_m}\}$. But then $\left\{U_{i_1}, \ldots, U_{i_m}, \hat{U}\right\}$ is a cover for $[a, r]$, showing that $r \in S$.

This same argument also gives us our contradiction. Note that $[a, r + \epsilon/2]$ is covered by $\left\{U_{i_1}, \ldots, U_{i_m}, \hat{U}\right\}$, implying that $r + \epsilon/2 \in S$, which contradicts the fact that r was the least upper bound for S. Hence $r = b$, and we conclude that intervals $[a, b]$ are compact. □

Remark 2.51

1. What conditions on $[a, b]$ result in an invalid proof were we to replace this with (a, b) or $[a, \infty)$? The first was that $[a, b]$ is a closed interval – which we used to ensure that V was non-empty – and we used the fact that $[a, b]$ was bounded. Both were used to invoke the Completeness Axiom. If we used a non-closed or non-bounded interval, the proof would fail.

2. This proof may seem similar to one you've seen before; namely, the Extreme Value Theorem. Indeed, that is because compactness secretly underlies that theorem as well, as we'll see in Corollary 2.55.

Theorem 2.52

Let (X, d_X) and (Y, d_Y) be metric spaces and $f : X \to Y$ a continuous function. If $K \subseteq X$ is compact, then $f(K)$ is also compact. More concisely, the continuous image of compact sets is compact.

Proof. Let $\mathcal{D} = \{U_i : i \in I\}$ be an open cover for $f(K)$, and consider $\mathcal{C} = \left\{V_i = f^{-1}(U_i) : i \in I\right\}$. As f is continuous, each $f^{-1}(U_i)$ is open. Moreover, $K \subseteq f^{-1}(f(K))$, showing that $\cup\mathcal{C}$ covers K and hence $\mathcal{C}$ is an open cover for K. Since K is compact, there is some finite subcover $\mathcal{C}' = \{V_{i_1}, \ldots, V_{i_n}\}$ of $\mathcal{C}$.

I claim that $\mathcal{D}' = \{U_{i_1}, \ldots, U_{i_n}\}$ is a finite subcover of $\mathcal{D}$. Indeed, if $\mathbf{y} \in f(K)$ then we can choose some $\mathbf{x} \in K$ such that $f(\mathbf{x}) = \mathbf{y}$. Since $\mathcal{C}'$ covers K, there is some V_{i_m} which contains $\mathbf{x}$. But $V_{i_m} = f^{-1}(U_{i_m})$, so $f(\mathbf{x}) \in U_{i_m} \in \mathcal{D}'$. Hence $\mathcal{D}'$ is a finite subcover of $f(K)$, as required. □

Theorem 2.53

Let (X, d_X) and (Y, d_Y) be metric spaces and $K \subseteq X$ a compact set. If $f : K \to Y$ is a continuous function, then f is uniformly continuous; that is, continuous functions on compact domains are uniformly continuous.

Proof. Effectively, the idea is as follows: We'll fix an $\epsilon > 0$, and for each $\mathbf{x} \in K$ we'll choose a $\delta > 0$ that satisfies the definition of continuity. Using the balls of radius δ centred at $\mathbf{x}$ for every $\mathbf{x} \in K$ creates an open cover of K. Since K is compact, it admits a finite subcover; that is, there are finitely many different δ's which work everywhere in K. Taking their minimum will give the desired result.

However, there are technical points which will confuse the matter. For example, a cover could have δ balls which barely touch one-another, meaning that the distance between two points in the cover could be more than δ. We'll fix this by throwing in factors of $1/2$ everywhere.

Formally, let $\epsilon > 0$ be given. For each $\mathbf{x} \in K$, let $\delta_\mathbf{x} > 0$ be such that

$$d_X(\mathbf{x}, \mathbf{y}) < \delta_\mathbf{x} \quad \Rightarrow \quad d_Y(f(\mathbf{x}), f(\mathbf{y})) < \frac{\epsilon}{2}.$$

Let $U_\mathbf{x} = B_{\delta_\mathbf{x}/2}(\mathbf{x})$, and take $\mathcal{C} = \{U_\mathbf{x} : \mathbf{x} \in K\}$ – an open cover of K. Since K is compact, it admits a finite subcover, $\mathcal{C}' = \{U_{\mathbf{x}_i} : i = 1, \ldots, n\}$. Let $\delta = \frac{1}{2} \min \left\{\delta_{\mathbf{x}_1/2}, \ldots, \delta_{\mathbf{x}_n/2}\right\}$, which I claim will show that f is uniformly continuous for this ϵ.

Indeed, let $\mathbf{x}, \mathbf{y} \in K$ satisfying $d_X(\mathbf{x}, \mathbf{y}) < \delta$. As $\mathcal{C}'$ is a cover for K, $\mathbf{x} \in U_{\mathbf{x}_k}$ for some k, implying that $d_X(\mathbf{x}, \mathbf{x}_k) < \delta_\mathbf{x}/2$. Thus

$$d_X(\mathbf{y}, \mathbf{x}_k) \leq d_X(\mathbf{y}, \mathbf{x}) + d_X(\mathbf{x}, \mathbf{x}_k) < \delta + \frac{\delta_{\mathbf{x}_k}}{2} < \delta_\mathbf{x},$$

and so

$$d_Y(f(\mathbf{x}), f(\mathbf{y})) \leq d_Y(f(\mathbf{x}), f(\mathbf{x}_k)) + d_Y(f(\mathbf{x}_k), f(\mathbf{y})) < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon,$$

which is what we wanted to show. □

2.4.1 Compactness in Euclidean $\mathbb{R}^n$

Throughout this section, we will always assume the Euclidean metric. I mentioned in Remark 2.51 that it was the closed and bounded properties of the interval $[a, b]$ that made it compact. This applies generally to compact subspaces of $\mathbb{R}^n$.

Theorem 2.54: Heine-Borel

A set $K \subseteq \mathbb{R}^n$ is compact if and only if it is closed and bounded.

Proof. [$\Rightarrow$] Assume that K is compact. We first show that K is bounded. Take as an open cover of K the set $\{B_n(\mathbf{0}) : n \in \mathbb{N}\}$. As K is compact, this must have a finite subcover, say $\{B_{n_1}(\mathbf{0}), \ldots, B_{n_k}(\mathbf{0})\}$. Let $N = \max\{n_1, \ldots, n_k\}$, in which case $K \subseteq B_N(\mathbf{0})$, and so is bounded.

To show that K is closed, assume for the sake of contradiction that it is not. That is, there is some point $\mathbf{p} \in \overline{K} \setminus K$. For each $\mathbf{x} \in K$, let $\rho(\mathbf{x}) = \|\mathbf{x} - \mathbf{p}\|/2$ – half the distance between $\mathbf{p}$ and $\mathbf{x}$ – and let $U_\mathbf{x} = B_{\rho(\mathbf{x})}(\mathbf{x})$. The set $\mathcal{C}' = \{U_\mathbf{x} : \mathbf{x} \in K\}$ forms an open cover of K, and so must admit a finite subcover $\mathcal{C}' = \{U_{\mathbf{x}_1}, \ldots, U_{\mathbf{x}_n}\}$. However, let $\rho = \min\{\rho(\mathbf{x}_1), \ldots, \rho(\mathbf{x}_n)\}$. Since $\mathbf{p} \in \overline{K}$, $B_{\rho/2}(\mathbf{p}) \cap K \neq \emptyset$, but no element of $B_{\rho/2}(\mathbf{p}) \cap K$ is contained in any of the $U_{\mathbf{x}_i}$, a contradiction.

[$\Leftarrow$] Suppose that K is closed and bounded. As K is bounded, there is some rectangle $[-L, L]^n$ that covers K (Exercise 2-6). As closed subspaces of compact spaces are compact (Exercise 2-45), it suffices to show that $[-L, L]^n$ is compact.

To do this, suppose for the sake of a contradiction that $R_0 = [-L, L]^n$ is not compact; that is, there exists some open cover $\mathcal{C}$ of R_0 which admits no finite subcover. Bisect each edge to form 2^n closed subrectangles of R_0. At least one of these rectangles requires an infinite subcover, call this rectangle R_1. Suppose then that we have constructed n rectangles $R_0 \supseteq R_1 \supseteq R_2 \supseteq \cdots \supseteq R_n$. Divide R_n into 2^n subrectangles, one of which much require an infinite subcover of $\mathcal{C}$, and call this R_{n+1}.

Choose an $\mathbf{x}_k \in R_k$ for each k. Note that the side length of R_k is $L/2^{k-1}$, which tends to zero as $k \to \infty$. Thus $(\mathbf{x}_n)$ must be a Cauchy sequence, and hence converges with limit $\mathbf{x}$. As each R_k is closed, $\mathbf{x} \in R_k$ for all k. Since $\mathcal{C}$ covers R_0, there exist some element $U \in \mathcal{C}$ such that $\mathbf{x} \in U$. But since U is open, for sufficiently large k we must have $R_k \subseteq U$, contradicting the fact that R_k required an infinite subcover. □

Corollary 2.55: Extreme Value Theorem

Let (X, d_X) be a metric space, and endow $\mathbb{R}$ with the Euclidean metric. If $K \subseteq X$ is a compact set and $f : K \to \mathbb{R}$ is a continuous function, then there exists $\mathbf{x}_{\min}, \mathbf{x}_{\max} \in K$ such that $f(\mathbf{x}_{\min}) \leq f(\mathbf{x}) \leq f(\mathbf{x}_{\max})$ for every $\mathbf{x} \in K$; that is, f achieves both its extreme values on K.

Proof. Since f is continuous and K is compact, by Theorem 2.52 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. □

Another useful characterization of compactness in $\mathbb{R}^n$ is the following:

Theorem 2.56: Bolzano-Weierstrass

A set $K \subseteq \mathbb{R}^n$ is compact if and only if every sequence in K 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 $(\mathbf{x}_{k_n}) \to \mathbf{x}$.

Proof. We'll use the fact that compact subsets of $\mathbb{R}^n$ are closed and bounded to help us.

[$\Leftarrow$] 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 2.22 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 2.20 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.

[$\Rightarrow$] 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 2.27 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 2.22 we know that $\mathbf{x} \in S$. Thus $(\mathbf{x}_{n_k})$ is a convergent subsequence. □

The idea that every sequence has a convergent subsequence is known as *sequential compactness*. In turns out that sequential compactness and compactness are always equivalent in metric spaces, and can be shown directly (Exercise 2-49). On the other hand, compact sets are generally not closed and bounded (Exercise 2-51): This is a special property of $\mathbb{R}^n$. The correct way of generalizing "closed and bounded" to general metric spaces is given in Exercise 2-53.

2.5 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, 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 $S_1 \cap S_2 = \emptyset$. There are multiple possible remedies, discussed in Exercise 2-63, but we'll add the condition that both sets must be open.

Definition 2.57

If (X, d) is a metric space, a set $S \subseteq X$ is said to be *disconnected* if there exist disjoint non-empty *relatively* open sets $S_1, S_2 \subseteq S$ such that $S = S_1 \cup S_2$. We refer to (S_1, S_2) as a *disconnection* of S. If S admits no disconnection, we say that S is *connected.*

Note that S_1 and S_2 are also both relatively closed, since their complement in S is each other. A set that is both open and closed is said to be *clopen*, and hence disconnections are clopen partitions of a set.

Example 2.58

Show that the following sets are not connected in the Euclidean topology on $\mathbb{R}^n$:

1. $S = [0,1] \cup [3,4] \subseteq \mathbb{R}$,
2. $\mathbb{Q} \subseteq \mathbb{R}$,
3. $T = \left\{(x,y) \in \mathbb{R}^2 : y \neq x\right\}$.

Solution.

1. The disconnection for this case is evident, setting $S_1 = [0,1]$ and $S_2 = [3,4]$. Note that both sets are *relatively* open in S, as $S_1 = (-\infty, 2) \cap S$ and $S_2 = (2,\infty) \cap S$. Both sets are also disjoint, 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 $S_2 = \mathbb{Q} \cap (\pi, \infty)$. Both sets are relatively open in $\mathbb{Q}$ by construction, $S_1 \cup S_2 = \mathbb{Q} \cap (\mathbb{R} \setminus \{\pi\}) = \mathbb{Q}$, while $S_1 \cap S_2 = \emptyset$.
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 $T = S_1 \cup S_2$ and $S_1 \cap S_2 = \emptyset$. Define $f(x,y) = y - x$, a continuous function, so that $S_1 = f^{-1}(0,\infty)$ and $S_2 = f^{-1}(-\infty, 0)$ realizes S_1 and S_2 as the preimage of open sets. Thus (S_1, S_2) is a disconnection of T. ■

Remark 2.59 Examples (2) and (3) above show that the elements of the disconnection can be arbitrarily close to one another yet still form a disconnection.

Theorem 2.60

The continuous image of a connected set is connected. More precisely, if (X, d_X) and (Y, d_Y) are metric spaces, $f : X \to Y$ is continuous, and $S \subseteq X$ is connected, then $f(S)$ is connected.

Proof. Suppose for the sake of contradiction that $f(S)$ is not connected; namely, it admits a disconnection (T_1, T_2). Consider $S_i = f^{-1}(T_i) \cap S$ for $i = 1, 2$, which we claim forms a disconnection of S. Indeed, S_1 and S_2 are both relatively open as they are the preimage of open sets. They cover S, for if $\mathbf{s} \in S$ then $f(\mathbf{s}) \in f(S)$ means that either $f(\mathbf{s}) \in T_1$ or $f(\mathbf{s}) \in T_2$, which in turn implies that either $\mathbf{s} \in S_1$ or $\mathbf{s} \in S_2$. Finally, $S_1 \cap S_2 = \emptyset$, since if not then $\mathbf{s} \in S_1 \cap S_2$ then $f(\mathbf{s}) \in f(S_1) \cap f(S_2) = T_1 \cap T_2 = \emptyset$, which cannot happen. Thus (S_1, S_2) is a disconnection for S, but this is a contradiction, since S was connected by assumption. □

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. Alternatively, the following definition is more rigid, but nicer for demonstrating that a set is connected.

Definition 2.61

Let (X, d) be a metric space, and $S \subseteq X$. 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 curve between those two points which never leaves the set.

Example 2.62

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

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. ■

Example 2.63

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

Solution. Consider Figure 2.12 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.

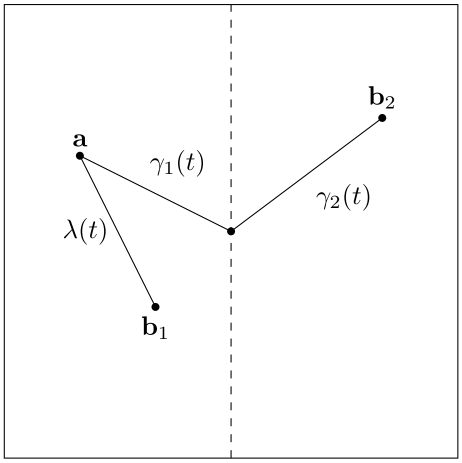


Figure 2.12: If $\mathbf{a}$ and $\mathbf{b}_1$ lie in the same plane, we can connect them with a straight line. If $\mathbf{a}$ and $\mathbf{b}_2$ lie in separate planes, we can connect them with a line through the origin.

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

$$\lambda(t) = \mathbf{a}t + (1-t)\mathbf{b} = (a_1 t + (1-t)b_1, a_2 t + (1-t)b_2).$$

Since a_1 and b_1 are both positive, the x-coordinate of the path $a_2 t + (1-t)b_2$ 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. ■

The following proposition has a relatively straightforward proof, which I leave to Exercise 2-62.

Proposition 2.64

Suppose (X, d_X) and (Y, d_Y) are metric spacse, with $f : X \to Y$ a continuous function. If $S \subseteq X$ is path connected, then $f(S)$ is path connected.

I mentioned that being path connected was a stronger condition than being connected. The below proposition demonstrates this fact.

Proposition 2.65

If (X, d) is a metric space and $S \subseteq X$ is path connected, then S is also connected.

Proof. We will proceed by contradiction. Assume then that $S \subseteq X$ is path connected but not connected, and fix 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 $P_2 = P \cap S_2$. Both sets are relatively open as subspaces of P, $P_1 \cup P_2 = (S_1 \cup S_2) \cap P = P$, but $P_1 \cap P_2 = P \cap (S_1 \cap S_2) = \emptyset$. Thus (P_1, P_2) is a disconnection of P, a contradiction. We conclude that S is connected. □

The converse of this proposition is not true, and is demonstrated by the *Topologist's Sine Curve*:

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

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.

2.5.1 Connectedness in $\mathbb{R}^n$

So far my examples have all be in Euclidean $\mathbb{R}^n$ anyway, as this is the easiest space to visualize. On the other hand, let's now focus exclusively on $\mathbb{R}^n$ with the Euclidean metric.

Proposition 2.66

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

Proof. For this proof we need a mathematical definition of an interval. We say that I is an interval if whenever $a, b \in I$ and x satisfies $a < x < b$, then $x \in I$ as well.

[$\Rightarrow$] We proceed by contrapositive. Suppose that S is not an interval, so there exist a, b, c such that $a, b \in S$ and $a < c < b$ but $c \notin S$. Let $S_1 = (-\infty, c) \cap S$ and $S_2 = (c, \infty) \cap S$. These sets are non-empty, since $a \in S_1$ and $b \in S_2$, are relatively open in S, and $S = S_1 \cup S_2$. Furthermore, $S_1 \cap S_2 = S \cap ((-\infty, c) \cap (c, \infty)) = \emptyset$, hence (S_1, S_2) is a disconnection of S.

[$\Leftarrow$] We again proceed by contrapositive. Suppose S is not connected and fix a disconnection (S_1, S_2). Since neither is empty, choose an $a \in S_1$ and $b \in S_2$, assuming without loss of generality that $a < b$. We will show there is a $c \notin S$ with $a < c < b$.

Let $p = \sup(S_1 \cap [a, b])$, so that $a \leq p \leq b$. Now $p \in \overline{S_1}$ so $p \notin S_2$ by Exercise 2-24. If $p \notin S_1$ then we're done, so assume that $p \in S_1$. Since $S_1 \subseteq \overline{S_2}^c$ which is open, there exists an $\epsilon > 0$ such that $B_\epsilon(p) \subseteq \overline{S_2}^c \subseteq S_2^c$. Now $b \in S_2$ means $b \notin B_\epsilon(p) = (p - \epsilon, p + \epsilon)$, so either $b < p - \epsilon$ or $b > p + \epsilon$. We also know $b \geq p$, so $b > p + \epsilon$. Fix some $c \in (p, p + \epsilon)$. By construction, we thus have

$$a \leq p < c < p + \epsilon < b,$$

and since $c > p$, the supremum of $S_1 \cap [a, b]$, we know $c \notin S_1$. Thus $c \notin S$ but $a < c < b$, as required. □

Corollary 2.67: Intermediate Value Theorem

Let $V \subseteq \mathbb{R}^n$ be a connected set and $f : V \to \mathbb{R}$ 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 $f(\mathbf{a}) < c < f(\mathbf{b})$ 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. □

Proposition 2.68

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

2.6 Exercises

2-1. Let (X, d) be a metric space and $S \subseteq X$. Show that $\partial S \cap S^{\text{int}} = \emptyset$.

2-2. Show that for an arbitrary choice of $a, b, r \in \mathbb{R}$, the closed disk $(x-a)^2 + (y-b)^2 \leq r^2$ is a bounded set in $\mathbb{R}^2$.

2-3. Let (X, d) be a metric space and fix $\mathbf{x}, \mathbf{y} \in X$. Show that if $d(\mathbf{x}, \mathbf{y}) < \epsilon$ for every $\epsilon > 0$, then $\mathbf{x} = \mathbf{y}$.

2-4. Let V be a vector space with norm $\|\cdot\|$.

(a) Show that for any $r > 0$ and $\mathbf{x} \in V$, $B_r(\mathbf{x}) \subseteq B_{r+\|\mathbf{x}\|}(\mathbf{0})$.

(b) We say that a set $S \subseteq V$ is bounded at $\mathbf{x} \in V$ if there exists some $r > 0$ such that $S \subseteq B_r(\mathbf{x})$. So that a set is bounded if and only if it's bounded at $\mathbf{x}$ for any $\mathbf{x} \in V$.

2-5. Let S be a subspace of the metric space (X, d). We say that S is *pointwise-bounded* if there exists a $P > 0$ such that $d(x, y) < C$ for all $x, y \in S$. Show that a set is pointwise bounded if and only if it is bounded.

2-6. We say that a set $S \subseteq \mathbb{R}^n$ is *rectangle-bounded* in the Euclidean metric if there exists some $L > 0$ such that $S \subseteq [-L, L]^n$. Show that a set S is rectangle bounded if and only if it is bounded.

2-7. Determine whether the following sets of Euclidean $\mathbb{R}^n$ are open, closed, or neither

(a) $\{(x, y) \in \mathbb{R}^2 : x > 1 \,\text{and}\, y > 0\}$

(b) $\{(x, y) \in \mathbb{R}^2 : x \notin \mathbb{Z}\}$

(c) $\{(x, y) \in \mathbb{R}^2 : x \in \mathbb{Z}\}$

(d) $\{(x, y) \in \mathbb{R}^2 : x^2 + y^2 = 1\}$

(e) $\{(x, y) \in \mathbb{R}^2 : x^2 + y^2 \leq 1\}$

(f) $\{(x, y) \in \mathbb{R}^2 : x^2 + y^2 \leq 1, (x, y) \neq (0, 0)\}$

(g) $\{(x, y) \in \mathbb{R}^2 : x > 0, y = \sin(1/x)\}$

(h) $\{(x, y) \in \mathbb{R}^2 : x, y \in \mathbb{Q} \cap [0, 1]\}$

(i) $\{(x, y) \in \mathbb{R}^2 : y > x^2\}$

2-8. Consider $C([0, 1])$ with $d(f, g) = \sup |f(x) - g(x)|$. Let $S \subseteq C([0, 1])$ be the collection of constant functions on $[0, 1]$. Find S^{int} and ∂S.

2-9. (a) Show that a finite set $F \subseteq \mathbb{R}^n$ is always closed in the Euclidean topology.

(b) Is this statement true in any topology? Namely, can you find a metric on $\mathbb{R}^n$ in which finite sets are not necessarily closed?

2-10. Consider $\mathbb{R}$ with the Euclidean metric. Construct an 'arbitrarily small' open set which contains all of $\mathbb{Q}$.

2-11. Let (X, d) be a metric space, and $\mathcal{T}'$ denote the collection of all *closed* subsets of X. Show that

(a) X and $\emptyset$ in $\mathcal{T}'$,

(b) $\mathcal{T}'$ is closed under finite unions,

(c) $\mathcal{T}'$ is closed under arbitrary intersections.

2-12. Suppose (X, d) is a metric space and $Y \subseteq X$ is a subspace. Let d' be the metric on X restricted to Y. Show that a set $U \subseteq Y$ is open in Y if and only if every point in U is an interior point relative to d'.

2-13. Akin to Example 2.7, determine if the set $S = \left\{(x, y) \in \mathbb{R}^2 : y > 0\right\}$ is open with respect to each of the metrics given in Exercise 1-37.

2-14. Two metrics d_1 and d_2 on a space X are said to be *equivalent* if there exists positive constants α, β such that

$$\alpha d_1(\mathbf{x}, \mathbf{y}) < d_2(\mathbf{x}, \mathbf{y}) < \beta d_1(\mathbf{x}, \mathbf{y})$$

for all $\mathbf{x}, \mathbf{y} \in X$.

(a) Show that equivalency of metrics is an equivalence relation on the set of all metrics.

(b) Show that two equivalent metrics define precisely the same open sets in X; namely, if d_1 and d_2 are equivalent metrics, a set $\mathcal{O}$ is open with respect to d_1 if and only if it is open with respect to d_2.

2-15. Suppose (X, d) is a metric space and $A \subseteq B \subseteq X$. Show that $A^{\text{int}} \subseteq B^{\text{int}}$ and $\overline{A} \subseteq \overline{B}$.

2-16. Let (X, d) be a metric space and $S \subseteq X$.

(a) Show that $(S^c)^c = S$.

(b) Show that $\partial S = \partial(S^c)$.

(c) Show that $(\overline{S})^c = (S^c)^{\text{int}}$.

(d) Show that $(S^{\text{int}})^c = \overline{S^c}$.

2-17. Determine if each metric in Exercise 1-37 is equivalent to the Euclidean metric.

2-18. Let $(\mathbf{x}_n)_{n=1}^\infty$ be a sequence in $\mathbb{R}^m$ with the Euclidean metric, 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 $i = 1, \ldots, m$.

2-19. Finish the proof of Theorem 2.18 by showing that sequence convergence is preserved under scalar multiplication.

2-20. Given a sequence $(\mathbf{x}_n)$ in a metric space (X, d), we say that $\mathbf{a} \in X$ is a *cluster point* of the sequence if for every $\epsilon > 0$, and every $n_0 \in \mathbb{N}$, there exists an $N \in \mathbb{N}$ with $N > n_0$ such that $d(\mathbf{x}_n, \mathbf{a}) < \epsilon$. You should convince yourself that this statement is equivalent to saying that "**a** is a cluster point of $(\mathbf{x}_n)$ if every open ball around **a** contains infinitely many elements of the sequence."

(a) Show that if $(\mathbf{x}_n) \to \mathbf{a}$, then $\mathbf{a}$ is a cluster point of $(\mathbf{x}_n)$.

(b) Give an example of sequence with two distinct cluster points, showing that not all cluster

2-21. Suppose (X, d) is a metric space. Let $(\mathbf{x}_n)$ be a Cauchy sequence, and suppose $\mathbf{x}$ is a cluster point for the sequence. Show that the $(\mathbf{x}_n) \to \mathbf{x}$.

2-22. Give an example of a sequence $(\mathbf{x}_n)$ such that $d(\mathbf{x}_n, \mathbf{x}_{n+1}) \to 0$ but $(\mathbf{x}_n)$ is not Cauchy.

2-23. Let S be a subspace of a metric space (X, d). We say that $\mathbf{x} \in S$ is a *limit point* of S if for every $\epsilon > 0$, $S \cap (B_\epsilon(\mathbf{x}) \setminus \{\mathbf{x}\}) \neq \emptyset$; that is, every ball around $\mathbf{x}$ contains points of S that are not $\mathbf{x}$ itself. The *derived set of* S, denoted S', is the collection of all limit points of S.

(a) Show that if $x \in S'$, there exists a sequence $(\mathbf{x}_n)$ in S such that $(\mathbf{x}_n) \to \mathbf{x}$.

(b) Conclude that $\overline{S} = S \cup S'$.

Both of these exercises can be done by effectively emulating the proof of Proposition 2.21.

2-24. Suppose (X, d) is a metric space and A, B are disjoint, non-empty, open subsets of X. Show that $\overline{A} \cap B = \emptyset$.

2-25. Let $X = \mathbb{R}$ be endowed with the exponential metric $d(x, y) = |e^x - e^y|$.

(a) Confirm that d is in fact a metric.

(b) Show that $x_n = -n$ is a Cauchy sequences which does not converge.

2-26. Find the following limits in Euclidean $\mathbb{R}^n$.

(a) $\displaystyle\lim_{(x,y)\to(0,0)} \left(5x^3 - x^2y^2\right)$

(b) $\displaystyle\lim_{(x,y)\to(0,0)} \frac{xy}{\sqrt{x^2+y^2}}$

(c) $\displaystyle\lim_{(x,y)\to(0,0)} \frac{x^4-4y^2}{x^2+2y^2}$

(d) $\displaystyle\lim_{(x,y)\to(0,0)} \frac{x^2+y}{\sqrt{x^2+y^2}}$

(e) $\displaystyle\lim_{(x,y)\to(0,0)} \frac{x^2\sin^2(y)}{x^2+2y^2}$

2-27. Let (X, d_X) and (Y, d_Y) be metric spaces. We say that f is continuous at $\mathbf{c} \in X$ if for every open set $U \subseteq Y$ containing $f(\mathbf{c})$, $f^{-1}(U)$ is an open set containing $\mathbf{c}$. Show that this definition of continuity at a point coincides with that of Definition 2.31.

2-28. Let (X, d_X) and (Y, d_Y) be metric spaces. Show that $f : X \to Y$ is continuous if and only if whenever $V \subseteq Y$ is a closed set, $f^{-1}(V) \subseteq X$ is a closed set.

2-29. Let $(X, d_X), (Y, d_Y)$, and (Z, d_Z) be metric spaces. Suppose that $g : X \to Y$ is continuous (at $\mathbf{c}$) and $f : Y \to Z$ is continuous (at $f(\mathbf{c})$). Show that $f \circ g$ is continuous (at $\mathbf{c}$) using

(a) The epsilon-delta definition of continuity,

(b) The sequential definition of continuity,

(c) The topological definition of continuity.

2-30. Show that $f : X \to Y$ is continuous if and only if for every set $A \subseteq X$, $f\left(\overline{A}\right) \subseteq \overline{f(A)}$.

2-31. Let $(X, d_X), (Y, d_Y)$, and (Z, d_Z) be metric spaces. Suppose that $g : X \to Y$ and $f : Y \to Z$, with $\mathbf{c} \in X, \mathbf{L} \in Y$, and $\mathbf{M} \in Z$ satisfying

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

(a) Show that if f is continuous at $\mathbf{L}$, or there is an open set $U \subseteq X$ containing $\mathbf{c}$ on which $g(\mathbf{x}) \neq \mathbf{L}$, then

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

(b) Show that the hypothesis that f be continuous at $\mathbf{L}$ or $g(\mathbf{x}) \neq \mathbf{L}$ on an open set containing $\mathbf{c}$ are necessary by explicitly constructing a counterexample to the result if this hypotheses is removed.

2-32. Suppose $\mathbf{p} \in \mathbb{R}^n$ and $U \subseteq \mathbb{R}^n$ is some open set containing $\mathbf{p}$. Let $f : U \to \mathbb{R}^m$ be a function, and define $C_{\mathbf{p}} = \{\gamma : (-1,1) \to U \colon \gamma(0) = \mathbf{p}\}$. Show that if

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

then for every $\gamma \in C_{\mathbf{p}}$,

$$\lim_{t\to 0} f(\gamma(t)) = \mathbf{L}.$$

2-33. Reread Examples 2.42 and 2.43. Fix an $n \in \mathbb{N}$. Construct a function $f : \mathbb{R}^2 \to \mathbb{R}$ whose limit exists for all paths $\gamma_{a,b}(t) = (at, bt^n)$ but fails for $\psi_m(t) = (t, mt^{n+1})$.

2-34. Show that the functions $\mathbb{R}^2 \to \mathbb{R}, (x,y) \mapsto x + y$ and $\mathbb{R}^2 \to \mathbb{R}, (x,y) \mapsto xy$ are continuous in the Euclidean topologies on $\mathbb{R}^2$ and $\mathbb{R}$.

2-35. Find an example of a continuous function $f : \mathbb{R}^n \to \mathbb{R}^m$ (with the Euclidean topologies) and an open set $U \subseteq \mathbb{R}^n$ such that $f(U)$ is not open. Still assuming that f is continuous, suppose there exists a continuous $g : \mathbb{R}^m \to \mathbb{R}^n$ such that $f(g(x)) = x$ and $g(f(x)) = x$. If U is open, must $f(U)$ be open? Must $g(U)$ be open?

2-36. Define a function $f : \mathbb{R}^2 \setminus \{(x,y) : x = 0\} \to \mathbb{R}$ as follows:

$$f(x,y) = \frac{\sin(xy)}{x}.$$

How should you define $f(x,y)$ at $x = 0$ so that $f(x,y)$ extends to a continuous function on all of $\mathbb{R}^2$?

2-37. Suppose $\mathbf{f}, \mathbf{g} : \mathbb{R}^n \to \mathbb{R}^m$ are continuous functions and $D \subseteq \mathbb{R}^n$ is a dense set. Suppose $\mathbf{f}(\mathbf{x}) = \mathbf{g}(\mathbf{x})$ for all $\mathbf{x} \in D$. Show that $\mathbf{f}(\mathbf{x}) = \mathbf{g}(\mathbf{x})$ for all $\mathbf{x} \in \mathbb{R}^n$.

2-38. Let $M_n(\mathbb{R})$ denote the set of $n \times n$ matrices with real coefficients.

(a) Show that $M_n(\mathbb{R})$ is a real vector space.

(b) Show that $\langle A, B\rangle = \operatorname{Tr}(A^T B)$ is an inner product on $M_n(\mathbb{R})$, where $\operatorname{Tr}(A) = \sum_{k=1}^n A_{k,k}$.

(c) Define $GL_n(\mathbb{R})$ as the set of invertible matrices.

i. Is $GL_n(\mathbb{R})$ a subspace of $M_n(\mathbb{R})$?

ii. Show that $GL_n(\mathbb{R})$ is an open subset of $M_n(\mathbb{R})$.

iii. Is $GL_n(\mathbb{R})$ connected?

(d) Let $SL_n(\mathbb{R})$ be the set of invertible matrices with determinant 1.

i. Show that $SL_n(\mathbb{R})$ is a closed subset of $M_n(\mathbb{R})$.

ii. Is $SL_n(\mathbb{R})$ connected?

(e) Let $O(n) = \left\{A \in M_n(\mathbb{R}) : A^T A = I_n\right\}$.

i. Show that $O(n)$ is a compact subset of $M_n(\mathbb{R})$.

ii. Is $O(n)$ connected?

iii. Let $\mathcal{B}_n$ denote the collection of ordered orthonormal bases of $\mathbb{R}^n$; that is, $\mathcal{B}_n \subseteq (\mathbb{R}^n)^n$, and $\mathfrak{b} \in \mathbf{B}_n$ is of the form $\mathfrak{b} = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ and[3] $\mathbf{v}_i \cdot \mathbf{v}_j = \delta_{ij}$. Show there is a bijection between $\mathcal{B}_n$ and $O(n)$.

[3] $\delta_{ij} = 1$ if $i = j$ and 0 otherwise. This is called the Kronecker delta.

(f) Let $SO(n) = \{A \in O(n) : \det(A) = 1\}$.

i. Show that $SO(n)$ is compact.

ii. Show that $SO(n)$ is connected (hard).

iii. Show that $SO(2)$ is homeomorphic to S^1, each with their subspace topologies.

2-39. Let $(V, \|\cdot\|)$ be a normed vector space. Show that the map $f : V \to \mathbb{R}, \mathbf{v} \mapsto \|\mathbf{v}\|$ is a continuous map.

2-40. Let (X, d_X) and (Y, d_Y) be metric spaces. Give an example of a continuous bijective function $f : X \to Y$ such that f^{-1} is not bijective.

2-41. A metric space (X, d) is said to be *bounded* if its metric is bounded; that is, there exists an $M > 0$ such that $d(\mathbf{x}, \mathbf{y}) < M$ for all $\mathbf{x}, \mathbf{y} \in X$. Show that every metric space is homeomorphic to a bounded metric space.

2-42. Show that any finite set is compact.

2-43. The distance between two sets $U, V \subseteq \mathbb{R}^n$ is defined by:

$$d(U, V) = \inf\{|x - y| : x \in U, y \in V\}$$

(a) Show that $d(U, V) = 0$ if either $\exists\, x \in \overline{U} \cap V$ or $\exists\, x \in \overline{V} \cap U$

(b) Show that if U is compact, V is closed, and $U \cap V = \emptyset$, then $d(U, V) > 0$.

(c) Show that the compactness of U in the previous part was necesssary by giving an example of two closed sets U and V in $\mathbb{R}^2$ which share no point in common, but satisfy $d(U, V) = 0$.

2-44. (a) Is the intersection of finitely many compact sets is always compact.

(b) Is the arbitrary intersection of compact sets compact?

(c) Is the union of finitely many compact sets is always compact.

(d) Is the arbitrary union of compact sets compact?

2-45. Let (X, d_X) be a compact metric space, and $C \subseteq X$ be a closed subspace. Show that C is also compact.

2-46. Let (X, d) be a metric space, and $K \subseteq X$ is compact. Using the open-cover definition of compactness, show that X is complete.

2-47. If (X, d) is a metric space, we say $K \subseteq X$ is *sequentially compact* if every sequence $(\mathbf{x}_n)$ in K has a convergent subsequence.

(a) Prove that the continuous image of sequentially compact sets is sequentially compact.

(b) Prove that sequentially compact spaces are complete; that is, every Cauchy sequence converges.

2-48. Let (X, d) be a metric space, and K be a sequentially compact subspace of X. Show that if $\mathcal{C}$ is an open cover of K, there exists some $r > 0$ such that for every $\mathbf{x} \in X$, $B_r(\mathbf{x}) \subseteq U$ for some $U \in \mathcal{C}$. *Hint:* Proceed by contradiction, and use this to construct a Cauchy sequence.

2-49. Let (X, d) be a metric space.

(a) Prove that if $K \subseteq X$ is sequentially compact then K is compact. *Hint:* Exercise 2-48.

(b) Prove that if $K \subseteq X$ is compact, then it is sequentially compact. *Hint:* Prove the contrapositive. Choose a sequence with no convergent subsequence. Argue that for every $\mathbf{x} \in K$, there is some open ball about $\mathbf{x}$ which contains only finitely many elements of the sequence.

2-50. Suppose (X, d_X) and (Y, d_Y) are compact metric spaces. Show that $X \times Y$ is also a compact metric space. *Hint:* This is very hard to do with the open-cover definition of compact.

2-51. Come up with an example of a metric space (X, d) and a set $K \subseteq X$ which is closed and bounded but not compact.

2-52. Suppose $f : X \to Y$ is bijective and continuous, with X compact. Show that f is a homeomorphism.

2-53. Let (X, d) be a metric space. A subspace $T \subseteq X$ is said to be *totally bounded* if for every $\epsilon > 0$ there exists a finite collection of points $\{\mathbf{a}_1, \ldots, \mathbf{a}_n\} \subseteq T$ (called an ϵ-net), such that $T \subseteq \bigcup_{i=1}^n B_\epsilon(\mathbf{a}_i)$.

(a) Show that a totally bounded set is necessarily bounded.

(b) Give an example of a bounded set which is not totally bounded.

(c) Show that every compact set is totally bounded.

(d) Show that every totally bounded and complete set is compact. *Hint:* Emulate the proof the Heine-Borel Theorem: Fix an open cover $\mathcal{C}$ and cover T with sets of radius $1/n$, arguing that one such set requires an infinite subcover. Construct a Cauchy sequence, and arrive at a contradiction.

We have now shown that K is compact if and only if K is totally bounded and complete. This is the correct way of generalizing "closed and bounded" to compactness outside of $\mathbb{R}^n$.

2-54. Suppose $f : (X, d_X) \to (Y, d_Y)$ is a homeomorphism, and $A \subseteq X$.

(a) Show that $f(\overline{A}) = \overline{f(A)}$.

(b) Show that $\partial f(A) = f(\partial A)$.

(c) Show that $f(A)^{\text{int}} = f(A^{\text{int}})$.

(d) By means of counterexample, show that each of these statements is false if f is merely continuous.

2-55. Suppose (X, d_X) is a metric space, with $U \subseteq X$ open and $K \subseteq U$ compact. Show that there is a compact space V such that $K \subseteq V^{\text{int}} \subseteq V \subseteq U$.

2-56. Let (X, d) be a metric space, with $U, V \subseteq X$. Define a function $\hat{d} : \mathcal{P}(X) \times \mathcal{P}(X) \to \mathbb{R}$ by $(U, V) \mapsto \hat{\inf} \{d(\mathbf{x}, \mathbf{y}) : \mathbf{x} \in U, \mathbf{y} \in V\}$.

(a) Show that $d(U, V) = 0$ if either $\exists x \in \overline{U} \cap V$ or $\exists x \in \overline{V} \cap U$.

(b) Show that if U is compact, V is closed, and $U \cap V = \emptyset$, then $d(U, V) > 0$.

(c) Show that the compactness of U in the previous part was necesssary by giving an example of two closed sets U and V in $\mathbb{R}^2$ which share no point in common, but satisfy $d(U, V) = 0$.

2-57. Let (X, d) be a metric space. Show that if $K_1 \supset K_2 \supset K_3 \supset K_4 \supset \ldots$ is a chain of proper containments with each K_i compact, then $\bigcap_{i=1}^\infty K_i \neq \emptyset$

2-58. Let $U \subseteq \mathbb{R}^n$ be an open set. We say that $f : U \to \mathbb{R}$ is *proper* if and only if f is continuous, and for every compact set $K \subseteq \mathbb{R}$, $f^{-1}(K) \subseteq U$ is compact. Suppose that we have a finite set $A \subseteq \mathbb{R}^n$ and a function $f : \mathbb{R}^n \setminus A \to \mathbb{R}$ such that for every $a \in A$,

$$\lim_{x\to a} |f(x)| = \lim_{x\to\infty} |f(x)| = \infty$$

Show that f is proper.

2-59. Determine if the following subsets of Euclidean $\mathbb{R}^n$ are connected or disconnected:

(a) The hyperbola $\left\{(x, y) \in \mathbb{R}^2 : x^2 - y^2 = 1\right\}$

(b) Any finite set of points in $\mathbb{R}^n$ with more than two elements

(c) $\left\{(x, y, z) \in \mathbb{R}^3 : xyz > 0\right\}$

(d) $\left\{(x, y, z) : x^2 + y^2 + z^2 = 1\right\}$

2-60. Determine whether the following statements are true or false. Give a proof if true, and a counter example if false.

(a) Every subset of a connected set is connected.

(b) The union of connected sets is connected.

(c) The intersection of connected sets is connected.

2-61. Let (X, d) be a metric space, and $S_i \subseteq X, i \in \mathbb{N}$ a collection of connected subsets with the property $S_i \cap S_{i+1} \neq \emptyset$ for all n. Show that $\bigcup_{i=1}^\infty S_i$ is connected.

2-62. Prove Proposition 2.64.

2-63. Let (X, d) be a metric space. Show that the following are equivalent:

(a) X is disconnected,

(b) There is a proper, non-trivial clopen subset of X,

(c) X contains a proper, non-trival subset S such that $\partial S = \emptyset$,

(d) X can be written as the union of two separated sets[4]

(e) There is a continuous surjective function $f : X \to \{0, 1\}$, where $\{0, 1\}$ has the discrete metric.

2-64. Endow $\mathbb{R}^n$ with the Euclidean topology. Suppose that S is an open connected subset of $\mathbb{R}^n$.

(a) Fix an $\mathbf{a} \in S$ and define S_1 as the collection of all points which are path connected to $\mathbf{a}$, and $S_2 = S_1^c$. Show that for every $\mathbf{b} \in S_1$, there exists an $\epsilon > 0$ such that $B_\epsilon(\mathbf{b}) \subseteq S_1$. Hence S_1 is open.

(b) Show that if $x \in \overline{S_1}$ then $x \in S_1$.

(c) Conclude that S is path connected.

[4] Two sets A, B are said to be *separated* if $\overline{A} \cap B = \emptyset = A \cap \overline{B}$.

2-65. Let (X, d) be a compact metric space, and take $C(X) = \{f : X \to \mathbb{R} : f \text{ continuous}\}$, where $\mathbb{R}$ has the Euclidean metric. Define a norm on $C(X)$ by $\|f\|_{\text{sup}} = \sup_{\mathbf{x}\in K} |f(\mathbf{x})|$. Show that $(f_k) \to f$ in $\|\cdot\|_{\text{sup}}$ is equivalent to uniform convergence.

2-66. Let $(X, \|\cdot\|)$ be a normed real vector space. A set $S \subseteq X$ is said to be *convex* if for every $\mathbf{x}, \mathbf{y} \in S$ and $t \in [0, 1]$, the line $(1 - t)\mathbf{x} + t\mathbf{y} \in S$. The set S is said to be *star-convex* if there exists a point $\mathbf{a} \in S$ such that for every $\mathbf{x} \in S$ and $t \in [0, 1]$, the line $(1 - t)\mathbf{a} + t\mathbf{x} \in S$.

(a) Show that every convex set is star-convex.

(b) Show that every star-convex set is connected.

(c) Give an example of a star-convex set which is not convex.

2-67. Two norms $\|\cdot\|_1, \|\cdot\|_2$ in a real vector space V are said to be *equivalent* if there are positive constants α, β such that $\alpha \|\mathbf{x}\|_1 \leq \|\mathbf{x}\|_2 \leq \beta \|\mathbf{x}\|_1$ for all $\mathbf{x} \in V$.

(a) Let d_1 and d_2 be the metrics defined by $\|\cdot\|_1, \|\cdot\|_2$ respectively. Show that equivalent norms induce equivalent metrics (Exercise 2-14).

(b) Suppose V is a finite dimensional real vector space. Show that any two norms on V are equivalent. *Hint:* Fix a basis $\{\mathbf{b}_1, \ldots, \mathbf{b}_n\}$ to identify $V \cong \mathbb{R}^n$, then use the fact that the unit sphere $\mathbb{S}^{n-1} \subseteq \mathbb{R}^n$ is compact.

2-68. Let $A \in L(\mathbb{R}^n, \mathbb{R}^m)$, where $\mathbb{R}^n$ and $\mathbb{R}^m$ are endowed with any norm, but say the Euclidean norm for this exercise. Recall the definition of the operator norm from Exercise **??**.

(a) An operator $A \in L(\mathbb{R}^n, \mathbb{R}^m)$ is said to be *bounded* if there exists an $M > 0$ such that $\|A\mathbf{x}\| \leq M\|\mathbf{x}\|$. Show that a A is bounded if and only if $\|A\|_{\text{op}}$ exists.

(b) Show that A is continuous if and only if A is *bounded.*

(c) Show that every $A \in L(\mathbb{R}^n, \mathbb{R}^m)$ is bounded. (Note: This is not true if the vector spaces in question are infinite dimensional.)

2-69. Let $U \subseteq \mathbb{R}^n$ be a bounded open set. An *exhaustion of U by compact sets* is a countable collection $\mathcal{K}$ of compact sets such that $U = \bigcup \mathcal{K}$ and $K_i \subseteq K_{i+1}^{\text{int}}$ for all $i \in \mathbb{N}$. In this exercise we'll demonstrate the existence of a compact exhaustion.

For each $n \in \mathbb{N}$, define $\mathcal{C}_n = \left\{B_{1/n}(\mathbf{x}) : \mathbf{x} \in \partial U\right\}$. Define $K_n = U \setminus \bigcup \mathcal{C}_n$. Show that the K_n form a compact exhaustion of U.

2-70. Generalize Exercise 2-69 to unbounded open sets U. *Hint:* For each $k \in \mathbb{N}$ let $U_k = U \cap B_k(\mathbf{0})$.

2-71. Let $\mathcal{O}$ be a collection of open sets in $\mathbb{R}^n$, and $U = \bigcup \mathcal{C}$. Show there is a countable collection of closed balls $\mathcal{B} = \left\{B_n = \overline{B_{r_n}(\mathbf{x}_n)} : n \in \mathbb{N}\right\}$ such that

i. $U = \bigcup_n B_n^{\text{int}}$,

ii. For each $n \in \mathbb{N}$, there exists $O \in \mathcal{O}$ such that $B_n \subseteq O$.

iii. For every point $\mathbf{x} \in U$, there exists a neighbourhod N of $\mathbf{x}$ such that only finitely many elements of $\mathcal{C}$ intersect N non-trivially.

To do this, proceed as follows:

(a) Fix a compact exhaustion $\{K_n : n \in \mathbb{N}\}$ of U (Exercise 2-70), and set $C_n = K_n - K_{n-1}^{\text{int}}$ (with the understanding that $C_1 = K_1$). Argue that the C_n are all compact.

(b) For every $\mathbf{x} \in C_i$, choose an $r_{\mathbf{x}} > 0$ sufficiently small so that $B_{\mathbf{x}} = \overline{B_{r_{\mathbf{x}}}(\mathbf{x})} \cap C_{i-2} = \emptyset$ and $B_{\mathbf{x}} \subseteq O$ for some $O \in \mathcal{O}$. Make sure to justify why you can do this. Use compactness to pick out a finite number of the $B_{\mathbf{x}}$, and let $\mathcal{B}_i$ denote the union of those finite number of balls.

(c) Let $\mathcal{B} = \bigcup_{i\in\mathbb{N}} \mathcal{B}_i$. Argue that this is the collection of closed balls which satisfies the required properties.

3 Differentiation

In the previous section, we considered metrics on arbitrary spaces – though we paid particular attention to $\mathbb{R}^n$. To define the derivative of a function, we'll have to add a bit more structure; namely, we're going to work on normed vector spaces. We'll write everything in terms of the Euclidean norm, but from Exercise 2-67 we know that any two norms on a finite dimensional vector space are equivalent, so this is not strictly necessary. In fact, a convenient norm to use other than the Euclidean norm is often the supremum norm.

3.1 A Brief Review: $\mathbb{R} \to \mathbb{R}$

Recall that if $a \in \mathbb{R}$ and U is an open neighbourhood of a, the function $f : U \to \mathbb{R}$ is said to be *differentiable at* a if

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

in which case we denote the value of the limit by $f'(a)$.

That we think of the derivative as a number – often representing the instantaneous rate of change of f at a – does not generalize when dealing with functions from $\mathbb{R}^n \to \mathbb{R}^m$. Hence let me redefine the derivative with a view towards something which does generalize to multivariate, vector-valued functions.

Recall that the set of linear operators $L(\mathbb{R}, \mathbb{R})$ is isomorphic to $\mathbb{R}$; more precisely, if $T \in L(\mathbb{R}, \mathbb{R})$ then there exists an $m \in \mathbb{R}$ such that $T(x) = [m]x$. Here we identify the 1×1 matrix $[m]$ with the real number m. The idea of a derivative is that a function f is differentiable at a if it can be well-approximated by a linear operator sufficiently close to a. In particular, if h is sufficiently small, one would hope that there exists an m such that

$$f(a+h) = f(a) + mh + \text{error}(h) \tag{3.1}$$

where 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 3.1

Let $a \in \mathbb{R}$ and U be an open neighbourhood of a. A function $f : U \to \mathbb{R}$ 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 (3.1) 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. I will not replicate the list here, for it is too large and the student should be well familiar with it.

3.2 Multivariate Vector-Valued Functions

A multivariate function is one whose domain is $\mathbb{R}^n$ for $n > 1$, while a vector-valued function in one whose codomain is $\mathbb{R}^m$ for $m > 1$. Our goal then is to define the correct notion of a derivative for functions $\mathbf{f} : \mathbb{R}^n \to \mathbb{R}^m$. Just as our discussion above suggested that we could think of the derivative of f as a linear operator on $\mathbb{R}$, the derivative of $\mathbf{f}$ will be a linear operator in $L(\mathbb{R}^n, \mathbb{R}^m)$. 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

$$\mathbf{f}(\mathbf{a}+\mathbf{h}) = \mathbf{f}(\mathbf{a}) + A(\mathbf{h}) + \operatorname{error}(\mathbf{h}),$$

for all $\mathbf{h}$ is a small neighbourhood of $\mathbf{0}$. Solving for the error we get

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

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

Definition 3.2

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 a linear operator $A \in L(\mathbb{R}^n, \mathbb{R}^m)$ such that

$$\lim_{\mathbf{h} \to 0} \frac{\|\mathbf{f}(\mathbf{a}+\mathbf{h}) - \mathbf{f}(\mathbf{a}) - A(\mathbf{h})\|}{\|\mathbf{h}\|} = 0. \tag{3.2}$$

We denote the linear operator A by $D\mathbf{f}(\mathbf{a})$. If $\mathbf{f}$ is differentiable for every point in its domain, we simply say that $\mathbf{f}$ is *differentiable*

Note that the two norms in (3.2) are different: The norm in the numerator is the Euclidean norm on $\mathbb{R}^m$, while the norm in the denominator is the Euclidean norm in $\mathbb{R}^n$.

Example 3.3

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

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

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

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

$$0 \leq \frac{\|\mathbf{f}(\mathbf{a}+\mathbf{h}) - \mathbf{f}(\mathbf{a}) - A\mathbf{h}\|}{\|\mathbf{h}\|} = \frac{h_1\sqrt{h_1^2 + h_3^2}}{\sqrt{h_1^2 + h_2^2 + h_3^2}} \leq \frac{h_1\sqrt{h_1^2 + h_3^2}}{\sqrt{h_1^2 + h_3^2}} = h_1.$$

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

We will work almost exclusively in the standard basis for $\mathbb{R}^n$. When the linear transformation $D\mathbf{f}(\mathbf{a})$ is written in the standard bases, it is called the *Jacobian matrix for* $\mathbf{f}$ *at* $\mathbf{a}$. Hence the Jacobian matrix for $D\mathbf{f}(\mathbf{a})$ in Example 3.3 is

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

We'll see how to explicitly compute the Jacobian matrix a while later.

Example 3.4

Show that the function $f(x, y) = x^2 + y^2$ is differentiable at the point $\mathbf{a} = (1, 0)$ with $Df(1, 0) = (2, 0)$. Determine more generally what $Df(\mathbf{a})$ 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 $Df(\mathbf{a}) = (c_1, c_2)$, 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, $\nabla f(x,y) = (2x, 2y)$. ■

Proposition 3.5

Suppose $\mathbf{a} \in \mathbb{R}^n$ and U is an open neighbourhood of $\mathbf{a}$. If $f : U \to \mathbb{R}^m$ is differentiable at $\mathbf{a}$, then its derivative $D\mathbf{f}(\mathbf{a})$ is unique.

Proof. Suppose $S, T \in L(\mathbb{R}^n, \mathbb{R}^m)$ satisfy (3.2). Let $\mathbf{d}(\mathbf{h}) = \mathbf{f}(\mathbf{a}+\mathbf{h}) - \mathbf{f}(\mathbf{a})$, and notice we can write

$$\begin{aligned} 0 \leq \frac{\|(S-T)(\mathbf{h})\|}{\|\mathbf{h}\|} &= \frac{\|S(\mathbf{h}) - \mathbf{d}(\mathbf{h}) + \mathbf{d}(\mathbf{h}) - T(\mathbf{h})\|}{\|\mathbf{h}\|} \\ &\leq \frac{\|f(\mathbf{a}+\mathbf{h}) - f(\mathbf{a}) - S(\mathbf{h})\|}{\|\mathbf{h}\|} + \frac{\|f(\mathbf{a}+\mathbf{h}) - \mathbf{f}(\mathbf{a}) - T(\mathbf{h})\|}{\|\mathbf{h}\|}. \end{aligned}$$

The right hand side goes to zero in the limit as $\mathbf{h} \to \mathbf{0}$, so we conclude $\|(S-T)\mathbf{h}\|/\|\mathbf{h}\| \xrightarrow{\mathbf{h}\to\mathbf{0}} 0$. Let $\mathbf{h}$ be any fixed, non-zero vector, so that

$$0 = \lim_{t\to 0} \frac{\|(S-T)(t\mathbf{h})\|}{\|t\mathbf{h}\|} = \lim_{t\to 0} \frac{|t|\|(S-T)\mathbf{h}\|}{|t|\|\mathbf{h}\|} = \frac{\|(S-T)(\mathbf{h})\|}{\|\mathbf{h}\|}.$$

Hence $\|(S-T)(\mathbf{h})\| = 0$ which implies $(S-T)(\mathbf{h}) = \mathbf{0}$, from which we conclude that $S(\mathbf{h}) = T(\mathbf{h})$. As $\mathbf{h}$ was arbitrary, we conclude that $S = T$. □

If $\mathbf{f} : \mathbb{R}^n \to \mathbb{R}^m$ is differentiable, knowing that $D\mathbf{f}(\mathbf{a})$ is unique gives rise to another function $D\mathbf{f} : \mathbb{R}^n \to L(\mathbb{R}^n, \mathbb{R}^m)$, $\mathbf{a} \mapsto D\mathbf{f}(\mathbf{a})$, which we will call the *derivative of* $\mathbf{f}$.

Proposition 3.6

Suppose $\mathbf{a} \in \mathbb{R}^n$ and U is an open neighbourhood of $\mathbf{a}$. If $f : U \to \mathbb{R}^m$ is differentiable at $\mathbf{a}$, then it is continuous at $\mathbf{a}$.

Proof. We can write

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

This last expression tends to zero by hypothesis, so by the Squeeze Theorem $\|\mathbf{f}(\mathbf{a}+\mathbf{h}) - \mathbf{f}(\mathbf{a})\| \xrightarrow{\mathbf{h}\to\mathbf{0}} 0$, from which we conclude that

$$\lim_{\mathbf{h}\to\mathbf{0}} \mathbf{f}(\mathbf{a}+\mathbf{h}) = \mathbf{f}(\mathbf{a})$$

Showing that $\mathbf{f}$ is continuous at $\mathbf{a}$ as required. □

3.2.1 Partial Derivatives

For the moment, let's restrict our attention to real-valued multivariate functions $f : \mathbb{R}^n \to \mathbb{R}$. We would like to way of determining $Df(\mathbf{a})$ other than using the limit definition of the derivative. 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 the direction each of standard unit vectors.

Definition 3.7

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

$$D_i f(\mathbf{a}) = \lim_{t\to 0} \frac{f(a_1, \ldots, a_i + t, \ldots, a_n) - f(a_1, \ldots, a_n)}{t} = \lim_{t\to 0} \frac{f(\mathbf{a} + t\mathbf{e}_i) - f(\mathbf{a})}{t}.$$

That is, $D_i(f)(\mathbf{a})$ 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 3.8

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

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

$$\begin{aligned} D_1 f(x, y, z) &= y + 2xz\cos(x^2 z) \\ D_2 f(x, y, z) &= x + \frac{e^y}{z^2} \\ D_3 f(x, y, z) &= x^2\cos(x^2 z) - \frac{2e^y}{z^3}. \end{aligned}$$ ■

Other common notation for indicating the partial derivative is given by the following:

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

This will be particularly convenient when we start taking higher order partial derivatives. I will often interchange between notation when it is unambiguous.

Theorem 3.9

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

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

Proof. This is actually a fairly natural result. Differentiability implies that the difference quotient limit limit exists from every direction, and partial derivatives are only capturing information about a single direction. More formally, let $\mathbf{e}_i$ be the standard unit vector in the ith direction. Let $\mathbf{h} = h\mathbf{e}_i$ and write the Jacobian matrix of f as $Df(\mathbf{a}) = (c_1, \ldots, c_n)$ so that

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

Re-arranging gives $\partial_{x_i} f(\mathbf{a}) = c_i$. Since this holds for arbitrary i, we conclude that x

$$Df(\mathbf{a}) = (\partial_{x_1} f(\mathbf{a}), \ldots, \partial_{x_n} f(\mathbf{a})). \qquad \square$$

It is important to note 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

$$f(x,y) = \begin{cases} \dfrac{xy}{x^2+y^2} & \text{if } (x,y) \neq (0,0) \\ 0 & \text{if } (x,y) = (0,0). \end{cases} \tag{3.3}$$

We know 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 3.10

Let $\mathbf{a} \in \mathbb{R}^n$, U an open set containing $\mathbf{a}$, and $f : U \to \mathbb{R}$. If every $\partial_i f(\mathbf{x})$ exists and is continuous in an open neighbourhood $V \subseteq U$ of $\mathbf{a}$, then f is differentiable at $\mathbf{a}$.

Proof. I'll do the proof for a function of two variables f just to keep the notation from getting out of hand. The idea for n variables is identical, if somewhat more cumbersome.

Suppose the partials exist and are continuous in a neighbourhood U of $\mathbf{a}$. Setting $\mathbf{d} = (\partial_x f(\mathbf{a}), \partial_y f(\mathbf{a}))$, we want to show that

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

Since U is open, choose an open ball $B_r(\mathbf{a}) \subseteq U$, and fix some $\mathbf{h} = (h_1, h_2)$ such that $\mathbf{a}+\mathbf{h} \in B_r(\mathbf{a})$. We introduce a cross-term to the following expression:

$$\begin{aligned} f(\mathbf{a}+\mathbf{h}) - f(\mathbf{a}) &= f(a_1+h_1, a_2+h_2) - f(a_1, a_2) \\ &= [f(a_1+h_1, a_2+h_2) - f(a_1+h_1, a_2)] + [f(a_1+h_1, a_2) - f(a_1, a_2)]. \end{aligned}$$

Define a function $g(t) = f(a_1 + h_1, a_2 + t)$. Note that g is continuous on $[0, h_2]$ and differentiable on $(0, h)$, by the assumption that $\partial_y f$ exists on U. By the Mean Value Theorem, there exists $\alpha \in [0, h_2]$ such that

$$\begin{aligned} g(h_2) - g(0) &= g'(\alpha) h_2 \\ f(a_1 + h_1, a_2 + h_2) - f(a_1 + h_1, a_2) &= \partial_y f(a_1 + h_1, a_2 + \alpha) h_2. \end{aligned}$$

By precisely the same reasoning, there is a $\beta \in [0, h_1]$ such that

$$f(a_1 + h_1, a_2) - f(a_1, a_2) = \partial_x f(a_1 + \beta, a_2) h_1.$$

Combining this, we get

$$\begin{aligned} &\frac{|f(\mathbf{a}+\mathbf{h}) - f(\mathbf{a}) - \mathbf{d}\cdot\mathbf{h}|}{\|\mathbf{h}\|} \\ &\quad = \frac{1}{\|\mathbf{h}\|} \left| \frac{\partial f}{\partial x}(a_1 + \beta, a_2) h_1 + \frac{\partial f}{\partial y}(a_1 + h_1, a_2 + \alpha) h_2 - \left(\frac{\partial f}{\partial x}(\mathbf{a}) h_1 + \frac{\partial f}{\partial y}(\mathbf{a}) h_2 \right) \right| \\ &\quad = \left| \frac{\partial f}{\partial y}(a_1 + h_1, a_2 + \alpha) - \frac{\partial f}{\partial y}(a_1, a_2) \right| \frac{|h_2|}{\|\mathbf{h}\|} + \left| \frac{\partial f}{\partial x}(a_1 + \beta, a_2) - \frac{\partial f}{\partial x}(a_1, a_2) \right| \frac{|h_1|}{\|\mathbf{h}\|} \\ &\quad \leq \underbrace{\left| \frac{\partial f}{\partial y}(a_1 + h_1, a_2 + \alpha) - \frac{\partial f}{\partial y}(a_1, a_2) \right|}_{(\dagger)} + \underbrace{\left| \frac{\partial f}{\partial x}(a_1 + \beta, a_2) - \frac{\partial f}{\partial y}(a_1, a_2) \right|}_{(\dagger\dagger)} \end{aligned}$$

where in the last equality we've used the fact that $|h_i|/\|\mathbf{h}\| \leq 1$. Since $\partial_i f$ is continuous in U, both $(\dagger)$ and $(\dagger\dagger)$ tend to zero as $\mathbf{h} \to \mathbf{0}$, showing that f is differentiable as a result. □

Once again let f be the function in (3.3). Its partial derivatives are given by

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

so the limits fail to exist as $(x, y) \to (0, 0)$ (try the line $y = -x$). Hence the partial derivatives are not continuous, and Theorem 3.10 does not apply.

Definition 3.11

Let $U \subseteq \mathbb{R}^n$. We define the collection of C^1 *functions on* U to be

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

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

All C^1 functions are automatically differentiable by Theorem 3.10; however, there are differentiable functions which are not C^1. For example, the function

$$f(x, y) = \begin{cases} (x^2 + y^2) \sin\left(\frac{1}{\sqrt{x^2 + y^2}} \right), & \text{if } (x, y) \neq (0, 0) \\ 0 & \text{if } (x, y) = (0, 0) \end{cases} \tag{3.4}$$

is everywhere differentiable, but its partial derivatives are not continuous at $(0,0)$. In some sense, this is example of the classical one-variable example of a differentiable functions which is not C^1. Indeed, along any straight line through the origin, this function looks likes $x \mapsto x^2\sin(1/x)$ when $x \neq 0$.

Proposition 3.12

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 $[D\mathbf{f}(\mathbf{a})]_{i,j} = \partial_j f_i(\mathbf{a})$.

Proposition 3.12 is straightforward to prove, so I leave it to Exercise 3-5. In practice, we can determine if $\mathbf{f}$ is differentiable by computing its Jacobian and seeing if its components are continuous in a neighbourhood of the point of interest.

Example 3.13

Determine if $\mathbf{f} : (0,\infty] \times [0, 2\pi) \to \mathbb{R}^2$, $\mathbf{f}(r,\theta) = (r\sin(\theta), r\cos(\theta))$ is differentiable, and if so find its derivative.

Solution. The Jacobian matrix is computed to be

$$D\mathbf{f}(r,\theta) = \begin{bmatrix} \sin(\theta) & r\cos(\theta) \\ \cos(\theta) & -r\sin(\theta) \end{bmatrix}.$$

Certainly each component is continuous on the domain of $\mathbf{f}$, and hence $\mathbf{f}$ is differentiable. ■

Example 3.14

Determine if $\mathbf{f} : \mathbb{R}^3 \to \mathbb{R}^3, \mathbf{f}(x,y,z) = (xy, z\sin(xy), e^{xz})$ is differentiable, and if so determine its derivative.

Solution. Once again we compute the Jacobian matrix:

$$D\mathbf{f}(x,y,z) = \begin{bmatrix} y & x & 0 \\ zy\cos(xy) & xz\cos(xy) & \sin(xy) \\ ze^{xz} & 0 & xe^{xz} \end{bmatrix}.$$ ■

We have presented a lot of theorems and counter-examples, which are summarized in Table 1.

Differentiable	$\Rightarrow$	Partials Exist	Theorem 3.9
Differentiable	$\not\Leftarrow$	Partials Exist	Function (3.3)
Partials exist and continuous	$\Rightarrow$	Differentiable	Theorem 3.10
Partials exist and continuous	$\not\Leftarrow$	Differentiable	Function (3.4)

Table 1: Implications of differentiability and existence of partial derivatives.

Directional Derivatives: Partial derivatives give us the ability to determine how a function changes 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 3.15

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

$$D_{\mathbf{u}} f(\mathbf{a}) = \lim_{t \to 0} \frac{\mathbf{f}(\mathbf{a} + t\mathbf{u}) - \mathbf{f}(\mathbf{a})}{t}$$

In the special case where $f : \mathbb{R}^n \to \mathbb{R}$, the first thing we notice is that $D_{\mathbf{e}_i} f = \partial_i f$. Moreover, the directional derivative can be more quickly computed as

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

See page 80 for more on this.

Example 3.16

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 3.17

Suppose $\mathbf{a} \in \mathbb{R}^n$ and U is an open neighbourhood of $\mathbf{a}$. If $\mathbf{f} : \mathbb{R}^n \to \mathbb{R}^m$ is differentiable at $\mathbf{a}$, then for any unit vector $\mathbf{u}$, $D_{\mathbf{u}} f(\mathbf{a})$ exists. Moreover, $D_{\mathbf{u}} f(\mathbf{a}) = Df(\mathbf{a})(\mathbf{u})$.

Proof. The proof is similar to that of Theorem 3.9. We will approach along the line $\mathbf{a} + t\mathbf{u}$ and use differentiability to conclude that the limit exists. Let $\mathbf{h} = t\mathbf{u}$ for $t \in \mathbb{R}$, so that

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

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

Example 3.18

Verify the result from Example 3.16 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:

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

At the point $\mathbf{a} = (0,0)$ the Jacobian is $Df(0,0) = (1,0)$, and so

$$D_{\mathbf{u}} f(0,0) = Df(0,0)\mathbf{u} = (1,0)\cdot\left(\frac{1}{\sqrt{5}}, \frac{2}{\sqrt{5}}\right) = \frac{1}{\sqrt{5}}. \qquad \blacksquare$$

You may have wondered why we defined $\mathbf{f} : U \subseteq \mathbb{R}^n \to \mathbb{R}^m$ to be C^1 in terms of the partial derivatives of $\mathbf{f}$, versus the more traditional idea that $D\mathbf{f} : U \to L(\mathbb{R}^n, \mathbb{R}^m)$ should be a continuous function. There are two obstacles here: The first is that we need a norm on $L(\mathbb{R}^n, \mathbb{R}^m)$, for which the natural choice is the operator norm. Of course, Exercise 2-67 tells us it doesn't matter which norm we use, as all norms on a finite dimensional vector space are equivalent. The second was the framework connected the derivative $D\mathbf{f}$ with its directional derivatives and partial derivatives. Having established these facts, we can prove the following:

Theorem 3.19

Let $U \subseteq \mathbb{R}^n$ be an open set and $\mathbf{f} : \mathbb{R}^n \to \mathbb{R}^m$. The function $D\mathbf{f} : U \to L(\mathbb{R}^n, \mathbb{R}^m)$ is continuous if and only if the partial derivatives of $\mathbf{f}$ exist and are continuous on U.

Proof. We will assume continuity of $D\mathbf{f}$ means we're working with the operator norm.

[⇒] Suppose $D\mathbf{f} : U \to L(\mathbb{R}^n, \mathbb{R}^m)$ is continuous, so that for any $\epsilon > 0$ we can find a $\delta > 0$ such that $\|D\mathbf{f}(\mathbf{x}) - D\mathbf{f}(\mathbf{y})\| < \epsilon$ whenever $\mathbf{x}, \mathbf{y} \in U$ and $\|\mathbf{x} - \mathbf{y}\| < \delta$.

Fix some $i \in \{1, \ldots, n\}$ and $j \in \{1, \ldots, m\}$, for which we'll focus our attention on $D_i f_j$. Let $\epsilon > 0$ be given and choose the same δ used in continuity of $D\mathbf{f}$. If $\mathbf{x}, \mathbf{y} \in U$ and $\|\mathbf{x} - \mathbf{y}\| < \delta$ then

$$|\partial_i f_j(\mathbf{x}) - \partial_i f_j(\mathbf{y})| = |Df_j(\mathbf{x})\mathbf{e}_i - Df_j(\mathbf{y})\mathbf{e}_i| \leq \|Df_j(\mathbf{x}) - Df_j(\mathbf{y})\| < \epsilon,$$

showing that $\partial_i f_j$ is continuous, as required.

[⇐] Conversely, suppose $\partial_i f_j$ is continuous on U for every $i \in \{1, \ldots, n\}$ and $j \in \{1, \ldots, m\}$. By Exercise 1-28, we know that if A is an $m \times n$ matrix with components a_{ij}, then

$$\|A\|_{\mathrm{op}} \leq \|A\|_{\mathrm{Eu}} = \sqrt{\sum_{i,j} a_{ij}^2}.$$

As such,

$$\|D\mathbf{f}(\mathbf{x}) - D\mathbf{f}(\mathbf{y})\|_{\mathrm{op}} \leq \sqrt{\sum_{i,j} |\partial_i f_j(\mathbf{x}) - \partial_i f_j(\mathbf{x})|^2}.$$

For a fixed $\epsilon > 0$, we can find a $\delta_{ij} > 0$ such that $|\partial_i f_j(\mathbf{x}) - \partial_i f_j(\mathbf{y})| < \epsilon/(mn)$ whenever $\|\mathbf{x} - \mathbf{y}\| < \delta_{ij}$. Taking $\delta = \min\{\delta_{ij} : i \in \{1, \ldots, n\}, j \in \{1, \ldots m\}\}$ gives the desired result. □

3.3 Special Cases and Specific Interpretations

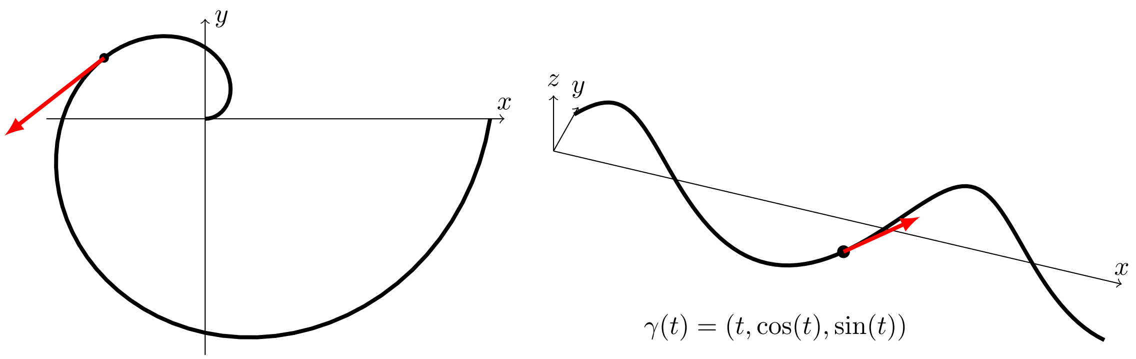
Here we'll look at two special cases of functions: Vector valued $\mathbb{R} \to \mathbb{R}^n$ and multivariate $\mathbb{R}^n \to \mathbb{R}$. We know how to differentiate these functions, but the specifics of the dimensions of domain and codomain give room for different interpretations. There is a third type of special case, the functions $\mathbf{F} : \mathbb{R}^n \to \mathbb{R}^n$. These can be interpreted as *vector-fields*, visualized by placing at $\mathbf{a} \in \mathbb{R}^n$ the vector $\mathbf{F}(\mathbf{a})$.

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

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

Example 3.20

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 3.1a).
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 3.1b).



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

(a)

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

(b)

Figure 3.1: Examples of parameterized curves and their derivatives. Left: A spiral in $\mathbb{R}^2$. Right: A helix in $\mathbb{R}^3$.

A vector-valued 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. In this case, we write $D\gamma = \gamma'$. 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 3.20 is differentiable.

Example 3.21

Determine the derivatives of each curve given in Example 3.20.

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 3.1.

Proposition 3.22

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. I will do the proof for (2) and leave the others as an exercise. 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

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

Paths are essential in analysis, and especially the field of *differential geometry*. For example, look at the definition of the direction derivative again. 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 $\mathbf{u}$. By composing with f, we get $g = f \circ \gamma : \mathbb{R} \to \mathbb{R}^n \to \mathbb{R}^m$ which is some curve in $\mathbb{R}^m$. 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. This will be made rigorous in Section 3.4.1.

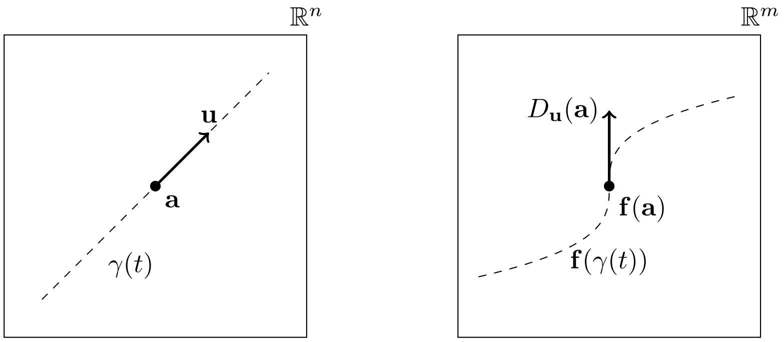


Figure 3.2: The directional derivative $D_{\mathbf{u}}f(\mathbf{a})$ can be thought of as the velocity of the curve $\mathbf{f}(\gamma(t))$ if $\gamma(t) = \mathbf{a} + t\mathbf{u}$.

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

Now we look at functions $f : \mathbb{R}^n \to \mathbb{R}$ that have multiple parameters. In this special case, the derivative of f is often denoted

$$Df(\mathbf{a}) = \nabla f(\mathbf{a}) = \left(\frac{\partial f}{\partial x_1}(\mathbf{a}), \frac{\partial f}{\partial x_2}(\mathbf{a}), \ldots, \frac{\partial f}{\partial x_n}(\mathbf{a}) \right),$$

where $\nabla f(\mathbf{a})$ is called the *gradient of f at $\mathbf{a}$*, and pronounced “nabla f” or “del f” or “grad f.” We can visualize this function by thinking about its graph,

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

as illustrated in Figure 3.3. 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. If $(\mathbf{x}, y)$ are the coordinates of $\mathbb{R}^{n+1}$, an n-plane through the origin has the equation

$$c_1x_1 + c_2x_2 + \cdots + c_nx_n + c_{n+1}y = \mathbf{c} \cdot \hat{\mathbf{x}} = 0,$$

where $\hat{\mathbf{x}} = (\mathbf{x}, y)$. In the case where $n = 1$ then this reduces to $c_1x_1 + c_2y = 0$, which we recognize as a straight line through the origin. If we instead would like this plane to pass through a point $\hat{\mathbf{a}} = (\mathbf{a}, f(\mathbf{a})) \in \mathbb{R}^{n+1}$, we can change this to

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

or equivalently, $\mathbf{c} \cdot \hat{\mathbf{x}} = d$ for some constant $d = \mathbf{c} \cdot \hat{\mathbf{a}}$. 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 $(y - y_0) = m(x - x_0)$.

Let’s look at the plane $\mathbf{c} \cdot (\hat{\mathbf{x}} - \hat{\mathbf{a}}) = 0$, which we want to be tangent to $y = f(\mathbf{x})$. If we fix every variable at $\mathbf{a}$ and only allow x_i to vary, we get the line $y - y_0 = c_i(x_i - a_i)$ – the equation of a line in the y-x_i plane. The slope of this line is precisely the partial derivative in the x_i direction, meaning $c_i = \partial_i f(\mathbf{a})$. Thus the *equation of the tangent plane to $z = f(\mathbf{x})$ at $\mathbf{a}$* is

$$y - f(\mathbf{a}) = \nabla f(\mathbf{a}) \cdot (\mathbf{x} - \hat{\mathbf{a}}).$$

We’ll see a more rigorous proof of this fact once we have learned the Implicit Function Theorem.

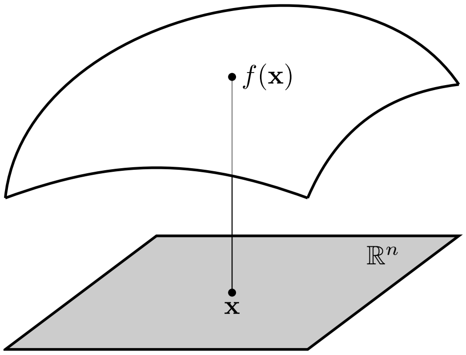


Figure 3.3: A function $f : \mathbb{R}^n \to \mathbb{R}$ can be visualized in terms of its graph.

3.4 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 3.23: 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

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

The proof is relatively long, though largely because we need to do some bookkeeping.

Proof. For notation sake, let $\mathbf{b} = \mathbf{g}(\mathbf{a})$ and set

$$\begin{aligned} \mathbf{u}(\mathbf{h}) &= \mathbf{g}(\mathbf{a} + \mathbf{h}) - \mathbf{g}(\mathbf{a}) - D\mathbf{g}(\mathbf{a})\mathbf{h} \\ \mathbf{v}(\mathbf{k}) &= \mathbf{f}(\mathbf{b} + \mathbf{k}) - \mathbf{f}(\mathbf{b}) - D\mathbf{f}(\mathbf{b})\mathbf{k}. \end{aligned}$$

Note that $\mathbf{u}(\mathbf{0}) = \mathbf{0}$ and $\mathbf{v}(\mathbf{0}) = \mathbf{0}$ since $\mathbf{g}$ and $\mathbf{f}$ are continuous at their respective points. Moreover, as $\mathbf{f}$ is differentiable at $\mathbf{b}$ and $\mathbf{g}$ is differentiable at $\mathbf{a}$,

$$\lim_{\mathbf{h} \to \mathbf{0}} \frac{\mathbf{u}(\mathbf{h})}{\|\mathbf{h}\|} = \mathbf{0} \quad \text{and} \quad \lim_{\mathbf{k} \to \mathbf{0}} \frac{\mathbf{v}(\mathbf{k})}{\|\mathbf{k}\|} = \mathbf{0}.$$

Let $\epsilon > 0$ be given and choose $\eta > 0$ so that $\|\mathbf{f}(\mathbf{y}) - f(\mathbf{b})\| < \epsilon$ whenever $\|\mathbf{y} - \mathbf{b}\| < \eta$. Let $\delta > 0$ be such that $\|\mathbf{g}(\mathbf{x}) - \mathbf{g}(\mathbf{a})\| < \eta$ whenever $\|\mathbf{x} - \mathbf{a}\| < \delta$. Hence $\mathbf{f} \circ \mathbf{g}$ is defined on $B_\delta(\mathbf{a})$. Choose $\mathbf{h}$

with $\|\mathbf{h}\| < \delta$ so $\mathbf{a}+\mathbf{h} \in B_\delta(\mathbf{a})$. Setting $\mathbf{k} = \mathbf{g}(\mathbf{a}+\mathbf{h}) - \mathbf{g}(\mathbf{a})$ we get

$$\|\mathbf{k}\| = \|D\mathbf{g}(\mathbf{a})\mathbf{h} + \mathbf{u}(\mathbf{h})\| \leq \|D\mathbf{g}(\mathbf{a})\|_{\text{op}} \|\mathbf{h}\| + \|\mathbf{u}(\mathbf{h})\|, \quad \text{(Exercise 1-28)}$$

which in turn implies that

$$\lim_{\mathbf{h}\to\mathbf{0}} \frac{\|\mathbf{v}(\mathbf{k})\|}{\|\mathbf{h}\|} \leq \lim_{\mathbf{h}\to\mathbf{0}} \frac{\|\mathbf{v}(\mathbf{k})\|}{\|\mathbf{k}\|} \left(\|D\mathbf{g}(\mathbf{a})\|_{\text{op}} + \frac{\|\mathbf{u}(\mathbf{h})\|}{\|\mathbf{h}\|} \right) = 0.$$

Now the numerator of the difference quotient can be manipulated as follows:

$$\begin{aligned}
\mathbf{f}(\mathbf{g}(\mathbf{a}+\mathbf{h})) - \mathbf{f}(\mathbf{g}(\mathbf{a})) - D\mathbf{f}(\mathbf{b})D\mathbf{g}(\mathbf{a})\mathbf{h} &= \mathbf{v}(\mathbf{k}) + D\mathbf{f}(\mathbf{b})\left[\mathbf{g}(\mathbf{a}+\mathbf{h}) - \mathbf{g}(\mathbf{a})\right] - D\mathbf{f}(\mathbf{b})D\mathbf{g}(\mathbf{a})\mathbf{h} \\
&= \mathbf{v}(\mathbf{k}) + D\mathbf{f}(\mathbf{b})\left[\mathbf{u}(\mathbf{h}) + D\mathbf{g}(\mathbf{a})\mathbf{h}\right] - D\mathbf{f}(\mathbf{b})D\mathbf{g}(\mathbf{a})\mathbf{h} \\
&= \mathbf{v}(\mathbf{k}) + D\mathbf{f}(\mathbf{b})\mathbf{u}(\mathbf{h}),
\end{aligned}$$

so that

$$0 \leq \lim_{\mathbf{h}\to\mathbf{0}} \frac{\|\mathbf{f}(\mathbf{g}(\mathbf{a}+\mathbf{h})) - \mathbf{f}(\mathbf{g}(\mathbf{a})) - D\mathbf{f}(\mathbf{b})D\mathbf{g}(\mathbf{a})\mathbf{h}\|}{\|\mathbf{b}\|} \leq \lim_{\mathbf{h}\to\mathbf{0}} \frac{\|\mathbf{v}(\mathbf{k})\|}{\|\mathbf{h}\|} + \|D\mathbf{f}(\mathbf{b})\|_{\text{op}} \lim_{\mathbf{h}\to\mathbf{0}} \frac{\|\mathbf{u}(\mathbf{h})\|}{\|\mathbf{h}\|} = 0.$$

This shows that $\mathbf{f}\circ\mathbf{g}$ is differentiable at $\mathbf{a}$, and moreover $D(\mathbf{f}\circ\mathbf{g})(\mathbf{a}) = D\mathbf{f}(\mathbf{g}(\mathbf{a}))D\mathbf{g}(\mathbf{a})$ as required.

□

Here now it is important to make the distinction between which objects are treated as rows and which are treated as columns. If $\mathbf{f} : \mathbb{R}^n \to \mathbb{R}^m$ then the Jacobian matrix $D\mathbf{f}(\mathbf{a})$ should reduce a gradient when $m = 1$, and should be curve derivative when $n = 1$. In particular, this implies that the Jacobian matrix of a function $\mathbb{R}^n \to \mathbb{R}$ is a *row vector*, while the Jacobian matrix of a function $\mathbb{R} \to \mathbb{R}^n$ is a *column vector*.

Here 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’ one another.

Example 3.24

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

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

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 3.25

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

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

and so

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

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

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

■

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 3.26

Let $\mathbf{f}(x, y) = (xy, x + y, x - y)$ and $\mathbf{g}(t) = (t, t^2)$. Compute $(\mathbf{f} \circ \mathbf{g})'(t)$.

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,

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

so by the Chain Rule

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

Finally, we'll do an example when both $\mathbf{f}$ and $\mathbf{g}$ have multiple dimensions in both domain and codomain.

Example 3.27

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

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

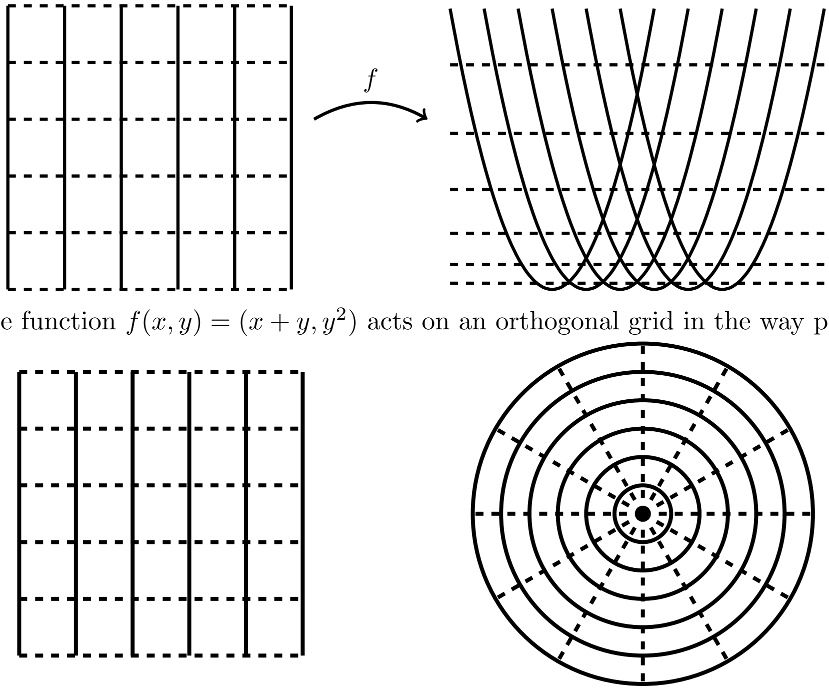
so that by the Chain Rule we have

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

3.4.1 More 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 think of a function as a black-box which takes an input elements of $\mathbb{R}^n$, and delivers an output element 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 3.4).



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

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 $(\mathbf{f} \circ \gamma)(0) = \mathbf{f}(\mathbf{a})$. See Figure 3.2 for a visual reminder. 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 $\mathbf{f}(\mathbf{a})$."

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

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

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

Definition 3.28

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 3.29

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-3.4, and it is a straightforward exercise to compute the Jacobian matrices to be

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

Thus taking determinants, we get the Jacobian determinants

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

3.5 The Mean Value Theorem

The Mean Value Theorem is powerful theorem. While it appears innocuous at first, it's ability to translate information between f and f' is invaluable. In this section we will examine how the MVT generalizes to multiple dimensions. To begin, recall the statement of the Mean Value Theorem in one dimension.

Theorem 3.30: Mean Value Theorem

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

$$f(b) - f(a) = f'(c)(b-a). \tag{3.5}$$

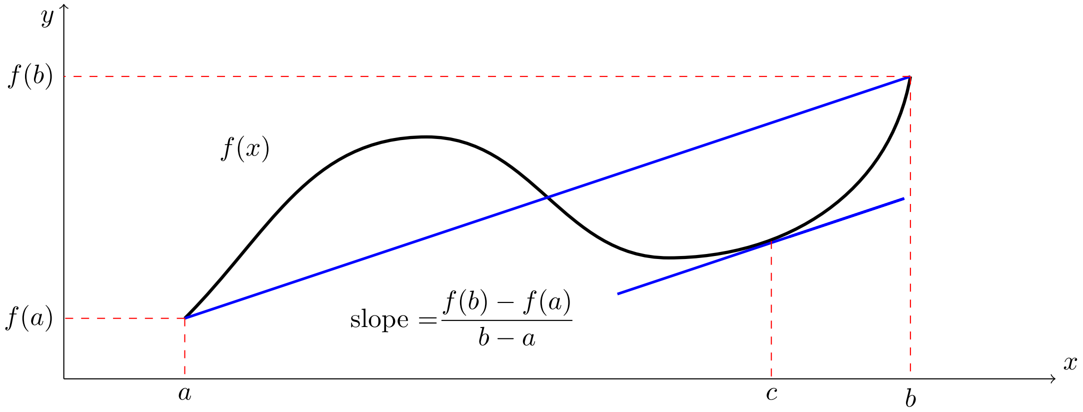


Figure 3.5: The Mean Value Theorem says that there is a point on this graph such that the tangent line has the same slope as the secant between $(a, f(a))$ and $(b, f(b))$.

The MVT can be used to prove several important results, such as the following:

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 we were to guess as to what the Mean Value Theorem might say, 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
> $$\mathbf{f}(b) - \mathbf{f}(a) = \mathbf{f}'(c)\,(b-a)\,.''$$

You 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 $f : [0, 2\pi] \to \mathbb{R}^2$ 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 $f(1) - f(0) = (0,0)$. 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. We conclude that 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 3.31: 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 the image of $\gamma(t) = \mathbf{a}(1-t) + t\mathbf{b}$ is a copy of the interval $[0,1]$ inside of U. Hence $f \circ \gamma : [0,1] \to \mathbb{R}$, to which we can apply the classical Mean Value Theorem.

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 gives

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

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 (Exercise 2-66) satisfy this property for any pair of points within the set. The following two corollaries are left to Exercises 3-12 and 3-15.

Corollary 3.32

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

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

Corollary 3.33

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

3.6 Higher Order Derivatives

Here is where our conversation starts to become abstract. To a differentiable function $\mathbf{f} : \mathbb{R}^n \to \mathbb{R}^m$ we associate another function $D\mathbf{f} : \mathbb{R}^n \to L(\mathbb{R}^n, \mathbb{R}^m)$. This is itself a map between two normed vector spaces – keeping in mind $L(\mathbb{R}^n, \mathbb{R}^m) \cong \mathbb{R}^{nm}$ if you prefer – and so has an associated derivative itself.

Definition 3.34

Let $\|\cdot\|_*$ be the Euclidean norm on $L(\mathbb{R}^n, \mathbb{R}^m)$. Let $\mathbf{a} \in \mathbb{R}^n$ and U an open neighbourhood of $\mathbf{a}$. We say that $\mathbf{f} : \mathbb{R}^n \to \mathbb{R}^m$ is *twice differentiable at* $\mathbf{a}$ if there exists a map $\Lambda \in L(\mathbb{R}^n, L(\mathbb{R}^n, \mathbb{R}^m))$ such that

$$\lim_{\mathbf{h} \to \mathbf{0}} \frac{\|D\mathbf{f}(\mathbf{a}+\mathbf{h}) - D\mathbf{f}(\mathbf{a}) - \Lambda(\mathbf{h})\|_*}{\|\mathbf{h}\|} = 0.$$

When this limit exists, we write $D(D\mathbf{f})(\mathbf{a}) = \Lambda$.

As any two norms on a finite dimensional vector space are equivalent (Exercise 2-67), we could have replaced the Euclidean norm $\|\cdot\|_*$ with the operator norm $\|\cdot\|_{\text{op}}$.

Example 3.35

Let $f : \mathbb{R}^2 \to \mathbb{R}, (x, y) \mapsto x^3 + x^2 y$. Show that $D(Df)(1, 2)(x, y) = \Lambda$, where

$$[\Lambda(x, y)](a, b) = 10xa + 2xb + 2ya \quad \text{or equivalently} \quad \Lambda(x, y) = \begin{bmatrix} 10x + 2y & 2x \end{bmatrix}.$$

Solution. Working with the Jacobian matrices, we have $D\mathbf{f}(x, y) = [3x^2 + 2xy, x^2]$, so if $\mathbf{a} = (1, 2)$ and $\mathbf{h} = (h, k)$ then $\mathbf{a} + \mathbf{h} = (1 + h, 2 + k)$ and

$$\begin{aligned} D\mathbf{f}(\mathbf{a}+\mathbf{h}) - D\mathbf{f}(\mathbf{a}) - \Lambda(\mathbf{h}) &= [3(1+h)^2 + 2(1+h)(2+k), (1+h)^2] - [9, 1] - [10h + 2k, 2h] \\ &= \left[3h^2 + 2hk, h^2\right] \end{aligned}$$

We can write the difference quotient as

$$\begin{aligned} \frac{\|D\mathbf{f}(\mathbf{a}+\mathbf{h}) - D\mathbf{f}(\mathbf{a}) - \Lambda(\mathbf{h})\|}{\|\mathbf{h}\|} &= \frac{\sqrt{10h^4 + 12h^3k^2 + 4h^2k^2}}{\sqrt{h^2+k^2}} \leq \frac{\sqrt{h^2+k^2}\sqrt{10h^2 + 12hk^2 + 4k^2}}{\sqrt{h^2+k^2}} \\ &= \sqrt{10h^2 + 12hk^2 + 4k^2}. \end{aligned}$$

This final expression is continuous in h and k, and hence goes to zero as $\mathbf{h} \to \mathbf{0}$, so by the Squeeze Theorem f is twice differentiable and $D(D\mathbf{f})(1, 2) = \Lambda$. ■

From Exercise 1-39 we know that $L(\mathbb{R}^n, L(\mathbb{R}^n, \mathbb{R}^m)) \cong L^2(\mathbb{R}^n, \mathbb{R}^m)$, where the right hand side is the collection of bilinear maps on $\mathbb{R}^n$. We can thus identify $D(D\mathbf{f})(\mathbf{a})$ with a bilinear map $D^2\mathbf{f}(\mathbf{a}) \in L^2(\mathbb{R}^n, \mathbb{R}^m)$. In the special case of $L^2(\mathbb{R}^n, \mathbb{R})$, the bilinear maps can be identified with matrices. If $B : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ is a bilinear map, then there exists a matrix A such that

$$f(\mathbf{u}, \mathbf{v}) = \mathbf{u}^T A\mathbf{v}.$$

This means the second derivatives of functions $f : \mathbb{R}^n \to \mathbb{R}$ can be written as matrices, though the situation becomes more complicated when the codomain is of higher dimension. We'll see how to write down this matrix in Definition 3.43.

Theorem 3.36

Suppose $\mathbf{a} \in \mathbb{R}^n$ and $U \subseteq \mathbb{R}^n$ is an open neighbourhood of $\mathbf{a}$. If $\mathbf{f} : U \to \mathbb{R}^m$ is C^2 on U, then for any vectors $\mathbf{v}, \mathbf{w} \in \mathbb{R}^n$, $D^2\mathbf{f}(\mathbf{a})(\mathbf{v}, \mathbf{w}) = D^2\mathbf{f}(\mathbf{a})(\mathbf{w}, \mathbf{v})$.

Proof. For this proof we'll be bootstrapping off our first derivative results. In particular, recall Theorem 3.17, which tells us that the directional derivative $D_\mathbf{v}\mathbf{f}(\mathbf{a}) = D\mathbf{f}(\mathbf{a})\mathbf{v}$. As $D\mathbf{f}$ is a differentiable function, so too is $D_\mathbf{v}\mathbf{f}$, so

$$D_\mathbf{w}(D_\mathbf{v}\mathbf{f})(\mathbf{a}) = D_\mathbf{w}(D\mathbf{f}(\mathbf{a})\mathbf{v}) = D(D\mathbf{f}(\mathbf{a}))(\mathbf{v})(\mathbf{w}) = D^2\mathbf{f}(\mathbf{a})(\mathbf{v}, \mathbf{w}).$$

Similarly, $D_\mathbf{v}(D_\mathbf{w}\mathbf{f})(\mathbf{a}) = D^2\mathbf{f}(\mathbf{w}, \mathbf{v})$.

Were we to write out the directional derivatives we would get

$$\begin{aligned} D_\mathbf{w}(D_\mathbf{v}\mathbf{f})(\mathbf{a}) &= \lim_{s\to 0} \frac{D_\mathbf{v}\mathbf{f}(\mathbf{a}+s\mathbf{w}) - D_\mathbf{v}\mathbf{f}(\mathbf{a})}{s} \\ &= \lim_{s\to 0} \frac{1}{s}\left[\lim_{t\to 0} \frac{\mathbf{f}(\mathbf{a}+s\mathbf{w}+t\mathbf{v}) - f(\mathbf{a}+s\mathbf{w})}{t} - \lim_{t\to 0} \frac{\mathbf{f}(\mathbf{a}+t\mathbf{v}) - \mathbf{f}(\mathbf{a})}{t}\right] \\ &= \lim_{s\to 0}\lim_{t\to 0} \frac{\mathbf{f}(\mathbf{a}+s\mathbf{w}+t\mathbf{v}) - \mathbf{f}(\mathbf{a}+s\mathbf{w}) - \mathbf{f}(\mathbf{a}+t\mathbf{v}) + \mathbf{f}(\mathbf{a})}{st}. \end{aligned}$$

In computing $D_\mathbf{v}(D_\mathbf{w}\mathbf{f})(\mathbf{a})$ we will get the same result, but with the roles of s and t reversed. Unfortunately, the process of interchanging two limits is tricky, so let's aim to attack this is a different way.

For the moment, let $f : U \to \mathbb{R}$ and fix s and t sufficiently small so that $\mathbf{a} + s\mathbf{w} + t\mathbf{v} \in U$. Let

$$\lambda(s,t) = f(\mathbf{a}+s\mathbf{w}+t\mathbf{v}) - f(\mathbf{a}+s\mathbf{w}) - f(\mathbf{a}+t\mathbf{v}) + f(\mathbf{a}),$$

which we recognize as the numerator of the previous limit. Furthermore, let's introduce $g_s : [0,t] \to \mathbb{R}$ by $g_s(u) = f(\mathbf{a}+s\mathbf{w}+u\mathbf{v}) - f(\mathbf{a}+u\mathbf{v})$ so that $\lambda(s,t) = g_s(t) - g_s(0)$. Since f is C^2 on U, we know g_s is continuous on $[0,t]$ and differentiable on $(0,t)$, from which the Mean Value Theorem gives a $t_0 \in [0,t]$ such that

$$\lambda(s,t) = g_s(t) - g_s(0) = g_s'(t)t = [D_\mathbf{v}f(\mathbf{a}+s\mathbf{w}+t_0\mathbf{v}) - D_\mathbf{v}f(\mathbf{a}+t_0\mathbf{v})]\, t.$$

Similarly, define $h : [0,s] \to \mathbb{R}$ by $h(u) = D_\mathbf{v}f(\mathbf{a}+s\mathbf{w}+t_0\mathbf{v})$. Again, our hypotheses on f ensure we can apply the Mean Value Theorem to h, so there exists some $s_0 \in [0,s]$ such that

$$D_\mathbf{v}f(\mathbf{a}+s\mathbf{w}+t_0\mathbf{v}) - D_\mathbf{v}f(\mathbf{a}+t_0\mathbf{v}) = h(s) - h(0) = h'(s_0)s = D_\mathbf{w}(D_\mathbf{v}f)(\mathbf{a}+s_0\mathbf{w}+t_0\mathbf{v}).$$

Thus

$$\begin{aligned}&\lim_{(s,t)\to(0,0)} \frac{\mathbf{f}(\mathbf{a}+s\mathbf{w}+t\mathbf{v})-\mathbf{f}(\mathbf{a}+s\mathbf{w})-\mathbf{f}(\mathbf{a}+t\mathbf{v})+\mathbf{f}(\mathbf{a})}{st}\\ &= \lim_{(s,t)\to(0,0)} \frac{\lambda(s,t)}{st} = \lim_{(s,t)\to(0,0)} \frac{D_{\mathbf{w}}(D_{\mathbf{v}}f)(\mathbf{a}+s_0\mathbf{w}+t_0\mathbf{v})st}{st}\\ &= D_{\mathbf{w}}(D_{\mathbf{v}}f)(\mathbf{a})\end{aligned}$$

But this limit is symmetric in $\mathbf{v}$ and $\mathbf{w}$ – or alternatively perform the same argument by interchange the role of s and t – to see that $D_{\mathbf{w}}(D_{\mathbf{v}}f)(\mathbf{a}) = D_{\mathbf{v}}(D_{\mathbf{w}}f)(\mathbf{a})$.

In the more general case where $\mathbf{f}: U \to \mathbb{R}^m$, $\mathbf{f}(\mathbf{x}) = (f_1(\mathbf{x}), \ldots, f_m(\mathbf{x}))$, each of the components f_i are symmetric, implying that the second derivative will also be symmetric. □

While the proofs become messier, we can inductively define the kth derivative as follows:

Definition 3.37

Suppose $\mathbf{a} \in \mathbb{R}^n$, $U \subseteq \mathbb{R}^n$ is an open neighbourhood of $\mathbf{a}$, and $k \in \mathbb{N}$. We say that $\mathbf{f}: U \to \mathbb{R}^m$ is k-times differentiable at $\mathbf{a}$ if $D^{k-1}\mathbf{f}$ exists on U, and there exists $\Lambda \in L^k(\mathbb{R}^n, \mathbb{R}^m)$ such that

$$\lim_{\mathbf{h}\to\mathbf{0}} \frac{\left\|D^{k-1}\mathbf{f}(\mathbf{a}+\mathbf{h}) - D^{k-1}\mathbf{f}(\mathbf{a}) - \Lambda(\mathbf{h})\right\|}{\|\mathbf{h}\|} = 0.$$

When this limit exists, we write $D(D^{k-1}\mathbf{f})(\mathbf{a}) = \Lambda$.

It is understood that the norm in the numerator is either the operator norm, or the Euclidean norm on $\mathbb{R}^{n^k m}$. As before, there is an isomorphism $L(\mathbb{R}^n, L^{k-1}(\mathbb{R}^n, \mathbb{R}^m)) \cong L^k(\mathbb{R}^n, \mathbb{R}^m)$ allowing us to identify the function $D(D^{k-1}\mathbf{f})$ with a k-linear map $D^k\mathbf{f}$.

Theorem 3.38

Suppose $\mathbf{a} \in \mathbb{R}^n$ and U is an open neighbourhood of $\mathbf{a}$. If $\mathbf{f}: U \to \mathbb{R}^m$ is k-times differentiable at $\mathbf{a}$, then

1. For any collection $\{\mathbf{v}_1, \ldots, \mathbf{v}_k\} \subseteq \mathbb{R}^n$,

$$D_{\mathbf{v_1}}(D_{\mathbf{v}_2}(\cdots D_{\mathbf{v_k}}\mathbf{f}(\mathbf{a}))\cdots) = D^k\mathbf{f}(\mathbf{a})(\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k).$$

2. If $\sigma: \{1, \ldots, k\} \to \{1, \ldots, k\}$ is any permutation of $\{1, \ldots, k\}$, then for any collection $\{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ we have

$$D^k\mathbf{f}(\mathbf{a})(\mathbf{v}_1, \ldots, \mathbf{v}_k) = D^k\mathbf{f}(\mathbf{a})(\mathbf{v}_{\sigma(1)}, \ldots, \mathbf{v}_{\sigma(k)}).$$

The second property above is known as *symmetry*, in that the order in which the vectors are substituted into $D^k\mathbf{f}(\mathbf{a})$ doesn't matter. Hence the kth derivative is a symmetric k-linear map.

3.6.1 Partial Derivatives

Partial derivatives form the computational framework for calculating derivatives. The situation becomes more complicated with higher order derivatives, since we generally don't have a nice way of writing down a k-linear function. Nonetheless, there is still value to computing these terms, so we see how to do so here, and how they are related to D^k in general.

Once again, let's restrict our attention to functions $f : \mathbb{R}^n \to \mathbb{R}$. The first step is second-order derivatives, of which there we now many different ways of computing a second derivative. For example, if $f : \mathbb{R}^2 \to \mathbb{R}$ then there are four possible second order partial 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.

Definition 3.39

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.

The definition of C^2 is equivalent to $D^2 f$ being a continuous function, but this is more readily realized in terms of the partial derivatives. The results summarized in Table 1 still hold.

Recall that $\partial_i f = D_{\mathbf{e}_i} f$, so that the partial derivatives are realized as the directional derivatives along the coordinate axes. Applying this to the second derivative means that

$$\frac{\partial}{\partial x_i}\left[\frac{\partial f}{\partial x_j}\right](\mathbf{a}) = \frac{\partial^2 f}{\partial x_i \partial x_j}(\mathbf{a}) = D_i(D_j f)(\mathbf{a}).$$

A natural consequence of Theorem 3.36 is therefore that $\partial_{x_i x_j} f = \partial_{x_j x_i} f$. However, this only applies if f is twice differentiable to begin with. It is possible to write down a function whose second order partials exist, is not twice differentiable, and whose mixed partials do not agree. All of this is summarized by the following theorem:

Theorem 3.40: 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 $\partial_{ij} f(\mathbf{a}) = \partial_{ji} f(\mathbf{a})$.

Example 3.41

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

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

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

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

■

Example 3.42

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

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

■

I mentioned earlier that in the special case $f : \mathbb{R}^n \to \mathbb{R}$ we could explicitly write down $D^2 f$. Indeed, $D^2 f(\mathbf{a}) : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ is a symmetric bilinear map, and in this special case these are all of the form

$$D^2 f(\mathbf{a})(\mathbf{v},\mathbf{w}) = \mathbf{v}^T A\mathbf{w}$$

for some $n \times n$ matrix A. Note that $A_{ij} = \mathbf{e}_i^T A\mathbf{e}_j$, so the matrix representing the second derivative has coordinates

$$A_{ij} = D^2 f(\mathbf{a})(\mathbf{e}_i, \mathbf{e}_j) = D_{\mathbf{e}_i}(D_{\mathbf{e}_j})f(\mathbf{a}) = \frac{\partial^2 f}{\partial x_i \partial x_j}(\mathbf{a}).$$

Definition 3.43

If $f : \mathbb{R}^n \to \mathbb{R}$ is twice differentiable at $\mathbf{a}$, then the *Hessian of f at* $\mathbf{a}$ is the matrix

$$\left[H_f(\mathbf{a})\right]_{ij} = \partial_{x_i x_j} f(\mathbf{a}).$$

In this case, $D^2 f(\mathbf{a})(\mathbf{v},\mathbf{w}) = \mathbf{v}^T H_f(\mathbf{a})\mathbf{w}$.

Example 3.44

Determine the Hessian of the function $f(x,y,z) = x^2 y + e^{yz}$ 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

$$H(x,y,z) = \begin{bmatrix} 2y & 2x & 0 \\ 2x & z^2 e^{yz} & e^{yz}(1+zy) \\ 0 & e^{yz}(1+zy) & y^2 e^{yz} \end{bmatrix}.$$

Evaluating at the point $(x, y, z) = (1, 1, 0)$ we get

$$H(1,1,0) = \begin{bmatrix} 2 & 2 & 0 \\ 2 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix}.$$

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:

$$\begin{aligned} f(\mathbf{x}) &= f(1,1,0) + \nabla f(1,1,0) \cdot \begin{bmatrix} x-1 \\ y-1 \\ z \end{bmatrix} + \frac{1}{2}(x-1, y-1, z) H(1,1,0) \begin{bmatrix} x-1 \\ y-1 \\ z \end{bmatrix} + O(\|\mathbf{x}\|^3) \\ &= 2 + (2,1,1) \begin{bmatrix} x-1 \\ y-1 \\ z \end{bmatrix} + \frac{1}{2}(x-1, y-1, z) \begin{bmatrix} 2 & 2 & 0 \\ 2 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} x-1 \\ y-1 \\ z \end{bmatrix} + O(\|\mathbf{x}\|^3) \\ &= 2 + 2(x-1) + (y-1) + z + (x-1)^2 + 2(x-1)(y-1) + (y-1)z + \frac{1}{2}z^2 + O(\|\mathbf{x}\|^3). \blacksquare \end{aligned}$$

We can make even further strides if we allow ourselves to import a powerful theorem from linear algebra:

Theorem 3.45: 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 Section 3.8.

I have limited our discussion so far to just second-order partial derivatives, in hopes that this simplest of cases would serve as a gentle introduction. To move further, we begin by defining C^k functions and generalizing Clairut's theorem to higher dimensions.

Definition 3.46

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 $C^\infty(U, \mathbb{R})$.

Notice that

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

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

Theorem 3.47: 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.

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.

Multi-indices When a function is of type C^k we know that the order of differentiation does not matter in computing a kth partial derivatives, 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 once, and the third variable once.

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 helpful pieces of notation in Section 3.7. 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.

3.6.2 Partial Derivatives with 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:

$$\frac{\partial^2 u}{\partial s^2} = \frac{\partial^2 u}{\partial x^2}\left[\frac{\partial x}{\partial s}\right]^2 + \frac{\partial^2 u}{\partial y^2}\left[\frac{\partial y}{\partial s}\right]^2 + 2\frac{\partial^2 u}{\partial x \partial y}\frac{\partial y}{\partial s}\frac{\partial x}{\partial s} + \frac{\partial u}{\partial x}\frac{\partial^2 x}{\partial s^2} + \frac{\partial u}{\partial y}\frac{\partial^2 y}{\partial s^2}. \tag{3.6}$$

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.

3.7 Taylor Series

3.7.1 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 $f^{(k)}(a) = p_{n,a}^{(k)}(a)$. This is a fairly straightforward exercise, and the student will find that

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

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 3.48: 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 3.49: 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

$$r_{n,a}(x) = \frac{f^{(n+1)}(c)}{(n+1)!}(x-a)^{n+1}. \tag{3.7}$$

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 (3.7). □

Corollary 3.50

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 3.49 implies that

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

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

$$|r_{n,a}(x)| \leq \frac{M}{(n+1)!}|x-a|^{n+1} \tag{3.8}$$

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

Example 3.51

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 (3.8) 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 (3.8) gives us $|r_{n,a}(x)| \leq 3[(k+1)!]^{-1}$, which also becomes smaller than 10^{-8} when $k = 11$. ■

3.7.2 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

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

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 $g(1) = f(\mathbf{x}_0)$. 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 $t = 1$:

$$g(1) = \sum_{k=0}^{n} \frac{g^{(k)}(0)}{k!} + r_{n,0}(1), \tag{3.9}$$

where $r_{n,0}(t)$ is the remainder. Let's look at the derivatives of g. The first derivative is easily computed via the chain rule, giving

$$g'(t) = Df(\gamma(t))\gamma'(t) = Df(\mathbf{a} + t(\mathbf{x}_0 - \mathbf{a}))(\mathbf{x}_0 - \mathbf{a}).$$

The second derivative gives

$$g''(t) = D(Df(\mathbf{a} + t(\mathbf{x}_0 - \mathbf{a})))(\mathbf{x}_0 - \mathbf{a})\gamma'(t) = D^2f(\mathbf{a} + t(\mathbf{x}_0 - \mathbf{a}))(\mathbf{x}_0 - \mathbf{a})^2,$$

where by $(\mathbf{x}_0 - \mathbf{a})^2$ I mean that the bilinear map takes $\mathbf{x}_0 - \mathbf{a}$ in both of its arguments. Inductively, we find that

$$g^{(k)}(t) = D^k f(\mathbf{a} + t(\mathbf{x}_0 - \mathbf{a}))(\mathbf{x}_0 - \mathbf{a})^k.$$

Similarly, the Taylor remainder becomes

$$r_{n,0}(1) = \frac{g^{(n+1)}(\theta)}{(n+1)!} = \frac{D^{n+1}f(\mathbf{c})}{(n+1)!}(\mathbf{x}_0 - \mathbf{a})^{n+1}$$

where $\theta \in (0,1)$, and $\mathbf{c} = \mathbf{a} + \theta(\mathbf{x}_0 + \mathbf{a})$ is some point on the line between $\mathbf{a}$ and $\mathbf{x}_0$. Plugging everything into (3.9) yields

$$f(\mathbf{x}_0) = \sum_{k=1}^{n} \frac{D^k(\mathbf{a})}{k!}(\mathbf{x}_0 - \mathbf{a})^k + \frac{D^{n+1}f(\mathbf{c})}{(n+1)!}(\mathbf{x}_0 - \mathbf{a})^{n+1},$$

which is remarkably similar to what we had for a single variable function.

While theoretically nice, from a computational standpoint this is not very useful. Let's repeat this calculation, but this time using coordinates. The first derivative yields

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

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 (3.9) and evaluating at $t = 0$ we have

$$f(\mathbf{x}) = \sum_{k=0}^{n} \frac{[(\mathbf{x}_0 - \mathbf{a}) \cdot \nabla]^k\, f(\mathbf{a})}{k!}.$$

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

$$[(\mathbf{x}_0 - \mathbf{a}) \cdot \nabla]^k\, f = \sum_{|\alpha| = k} \frac{k!}{\alpha!} \left(\partial^\alpha f\right)(\mathbf{a})\, (\mathbf{x}_0 - \mathbf{a})^\alpha.$$

In conclusion, the coordinate 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 3.52

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:

$\|\alpha\|$	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 3.53

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:

$\|\alpha\|$	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.

We know that if $f:\mathbb{R}^n\to\mathbb{R}$ is at least class C^2, its second derivative is a symmetric bilinear form and can be expressed as a matrix in terms of the Hessian. The Hessian matrix makes writing down the first few terms of the Taylor polynomial 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

$$f(\mathbf{x}) = f(\mathbf{a}) + \nabla f(\mathbf{a})(\mathbf{x}-\mathbf{a}) + \frac{1}{2}(\mathbf{x}-\mathbf{a})^T H(\mathbf{a})(\mathbf{x}-\mathbf{a}) + O(\|\mathbf{x}^3\|).$$

3.8 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 3.54

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.

3.8.1 Critical Points

Definition 3.55

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 3.56

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 3.57

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

$$\lim_{h\to 0^+} \frac{f(c+h)-f(c)}{h} \leq 0 \tag{3.10}$$

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

$$\lim_{h\to 0^-} \frac{f(c+h)-f(c)}{h} \geq 0. \tag{3.11}$$

Combining (3.10) and (3.11) 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 3.58

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 3.57 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 $\mathbf{u} = \mathbf{e}_i = (0, \ldots, 1, \ldots, 0)$ 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 $\nabla f(\mathbf{c}) = 0$. □

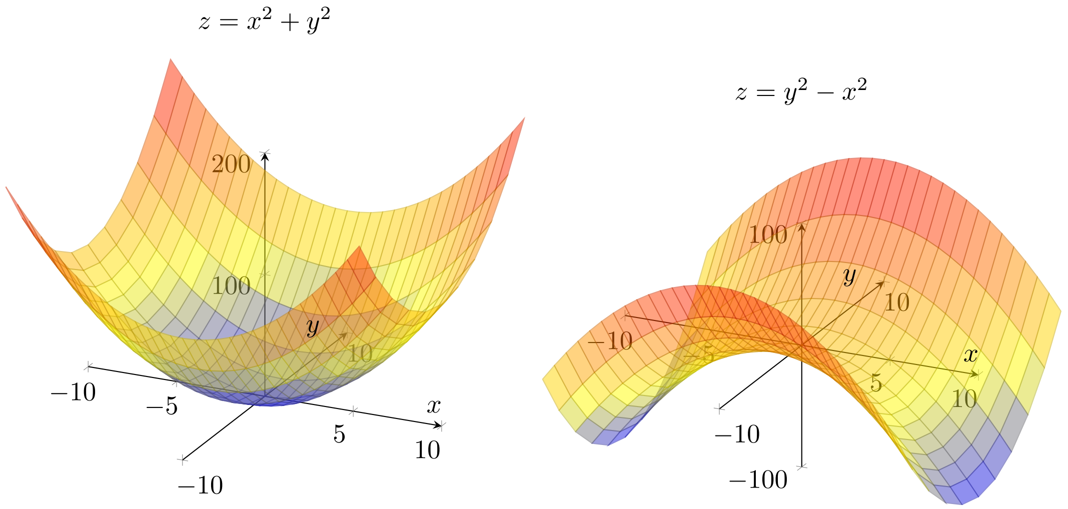


Figure 3.6: Plots of $z = x^2 + y^2$ (Left) and $z = y^2 - x^2$ (Right). Both surfaces have a critical point at $(x, y) = (0, 0)$, but only one of these corresponds to an extremum.

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* (See Figure 3.6) which we will define shortly. 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.

Recall that a bilinear map $B : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ is said to be *positive definite* if $B(\mathbf{u}, \mathbf{v}) \geq 0$ for all $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$, and *negative definite* if $B(\mathbf{u}, \mathbf{v}) \leq 0$ for all non-zero $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$. The following are equivalent: Let $B : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ be a bilinear map, with M_B its corresponding matrix representative.

- The map $B : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ is positive definite,
- The eigenvalues of M_B are all strictly positive,
- The leading principal minors of M_B are all strictly positive.

Similarly,

- The map $B : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ is negative definite,
- The eigenvalues of M_B are all strictly negative,
- The even leading principal minors of M_B are positive, while the negative leading principal minors are negative.

Proposition 3.59

Let $f : \mathbb{R}^n \to \mathbb{R}$ be class C^2 in a neighbourhood of a critical point $\mathbf{a}$.

1. If $D^2 f(\mathbf{a})$ is positive definite, then $\mathbf{a}$ is a local minimum,
2. If $D^2 f(\mathbf{a})$ has all negative eigenvalues, then $\mathbf{a}$ is a local maximum,
3. If $D^2 f(\mathbf{a})$ is neither positive nor negative definite , then $\mathbf{a}$ is neither a maximum nor a minimum.

Critical points where $D^2 f(\mathbf{a})$ is neither positive nor negative semi-definite are *saddle points*. Roughly speaking, saddle points look like both minima and maxima, depending on cross-sectional perspective. See Figure 3.6 for a visualization of a saddle point.

Proof. I'll do the proof for when $D^2 f(\mathbf{a})$ is positive definite, as the other cases follow similarly. The function $\mathbf{a} \mapsto D^2 f(\mathbf{a})$ is continuous by assumption. Identifying $D^2 f(\mathbf{a})$ with its Hessian representation, we therefore know that $\mathbf{a} \mapsto H_f(\mathbf{a})$ is continuous as well, and hence each of the components $\mathbf{a} \mapsto [H_f(\mathbf{a})]_{ij}$ are continuous. Let $\lambda_i : \mathbb{R}^n \to \mathbb{R}, \mathbf{a} \mapsto \lambda_i(\mathbf{a})$ for $i = 1, \ldots, n$ represent the eigenvalues of $H_f(\mathbf{a})$. As the λ_i are polynomials in the components of $H_f(\mathbf{a})$, themselves continuous, we we know the λ_i are continuous functions.

Since each λ_i is continuous and $\lambda_i(\mathbf{a}) > 0$, there is an open neighbourhood U_i of $\mathbf{a}$ on which $\lambda_i(\mathbf{x}) > 0$ for all $\mathbf{x} \in U_i$. Let $U = U_1 \cap \cdots \cap U_n$, and by taking an open ball in U if necessary, assume U is a convex open neighbourhood of $\mathbf{a}$. For each $\mathbf{x}_0 \in U$, $D^2 f(\mathbf{x}_0)$ is a positive definite matrix.

Fix some $\mathbf{x} \in U$, and write down the first order Taylor polynomial centred at $\mathbf{a}$. Note that $Df(\mathbf{a}) = 0$ and there exists some $\mathbf{c}$ on the line connecting $\mathbf{a}$ and $\mathbf{x}$ such that

$$f(\mathbf{x}) = f(\mathbf{a}) + \frac{D^2 f(\mathbf{c})}{2}(\mathbf{x} - \mathbf{a})^2.$$

Since U is convex, $\mathbf{c} \in U$ and hence $D^2 f(\mathbf{c})$ is positive semi-definite: $D^2 f(\mathbf{c})(\mathbf{x} - \mathbf{a})^2 \geq 0$. Putting this into our expression above,

$$f(\mathbf{x}) \geq f(\mathbf{a})$$

and this holds for all $\mathbf{x} \in U$. Hence $\mathbf{a}$ is a local minimum of f as required. □

Example 3.60

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

$$H(x, y) = \begin{bmatrix} 12x^2 - 4 & 0 \\ 0 & 6y \end{bmatrix}.$$

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 3.61

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* of $\mathbf{f}$ if $\operatorname{rank} H(\mathbf{c}) < n$.

Example 3.62

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

$$H(x, y) = \begin{bmatrix} -6x & 0 \\ 0 & 2 \end{bmatrix}, \qquad \Rightarrow \qquad H(0, 0) = \begin{bmatrix} 0 & 0 \\ 0 & 2 \end{bmatrix}$$

so $H(0, 0)$ has rank 1, and we conclude that $(0, 0)$ is a degenerate critical point. ■

3.8.2 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."

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

$$f(x, y) = x(200 - 2x) = 200x - 2x^2.$$

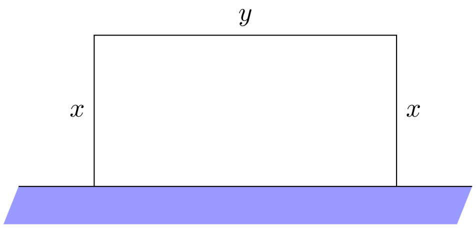


Figure 3.7: A visualization of simple optimization problem.

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 3.63

Let $f, G : \mathbb{R}^n \to \mathbb{R}$ be C^1 functions, and set $S = G^{-1}(0)$. If S is compact and 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

$$0 = \left.\frac{\mathrm{d}}{\mathrm{d}t}\right|_{t=0} f(\gamma(t)) = \nabla f(\mathbf{c}) \cdot \gamma'(0).$$

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[5]; that is, there exists some $\lambda \in \mathbb{R}$ such that $\nabla f(\mathbf{c}) = \lambda \nabla G(\mathbf{c})$ as required. □

[5] 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 $n - 1$.

Example 3.64

Use the method of Lagrange multipliers to solve the problem given in Figure 3.7.

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,

$$\begin{bmatrix} y \\ x \end{bmatrix} = \lambda \begin{bmatrix} 2 \\ 1 \end{bmatrix}.$$

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 3.65

Maximize the function $f(x, y, z) = xyz$ on the ellipsoid $x^2 + 2y^2 + 3z^2 = 1$.

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

$$\lambda x^2 = 2\lambda y^2 = 3\lambda z^2. \tag{3.12}$$

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 (3.12) 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 3.66

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

$$\begin{aligned} 2x &= 2\lambda x \\ 4y &= 2\lambda y \end{aligned}$$

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

$$\nabla f(\mathbf{c}) = \sum_{i=1}^{m} \lambda_i \nabla G_i(\mathbf{c}).$$

3.9 The Implicit and Inverse Function Theorems

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 exploring these questions.

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 $y_i = f(x_1, \ldots, x_n)$. This level of generality can make it difficult to see the forest for the trees, so let's treat the $k = 1$ first and develop insight.

3.9.1 Scalar Valued Functions

Consider a function $F \in C^1(U, \mathbb{R})$ for some open set $U \subseteq \mathbb{R}^{n+1}$. Let's endow $\mathbb{R}^{n+1}$ with the coordinates $(x_1, \ldots, x_n, y)$, whose purpose is to make it clear 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 $\mathbf{x}$? 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 3.67 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

$$y = \sqrt[3]{\frac{1}{x^2+1}},$$

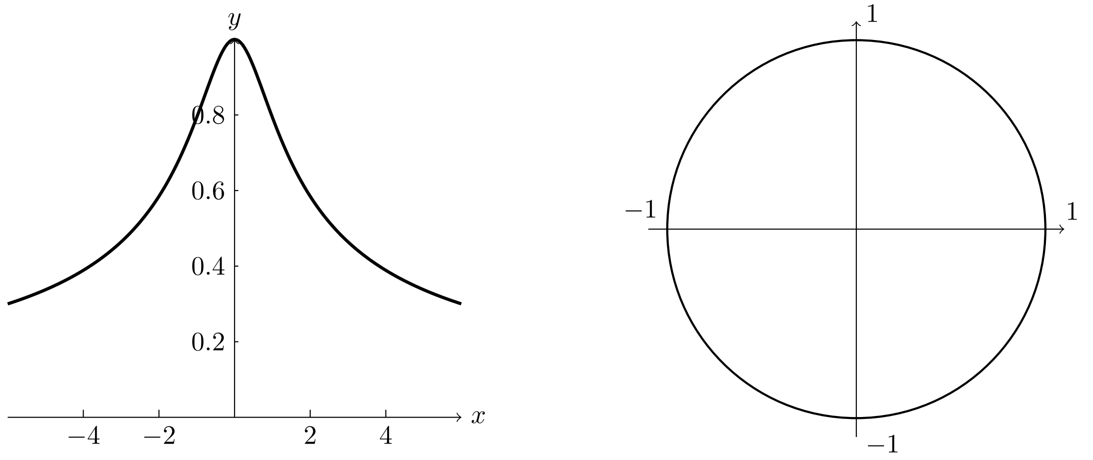
and this holds for all $x, y \in \mathbb{R}^2$. ▲

Unfortunately, it turns out that such examples are exceptionally rare and in general the answer is no:

Example 3.68 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. ▲

The primary difference between Examples 3.67 and 3.68 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,



(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.

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 plane at $(\mathbf{a}, b)$ is given by

$$\begin{aligned}&\partial_{x_1} F(\mathbf{a}, b) x_1 + \cdots + \partial_{x_n} F(\mathbf{a}, b) x_n + \partial_y F(\mathbf{a}, b) y + d \\ &= \nabla_x F(\mathbf{a}, b) \cdot \mathbf{x} + \partial_y F(\mathbf{a}, b) y + d = 0\end{aligned}$$

for some constant d. By analogy, $\partial_y F(\mathbf{a}, b)$ 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 $(\mathbf{a}, b)$.

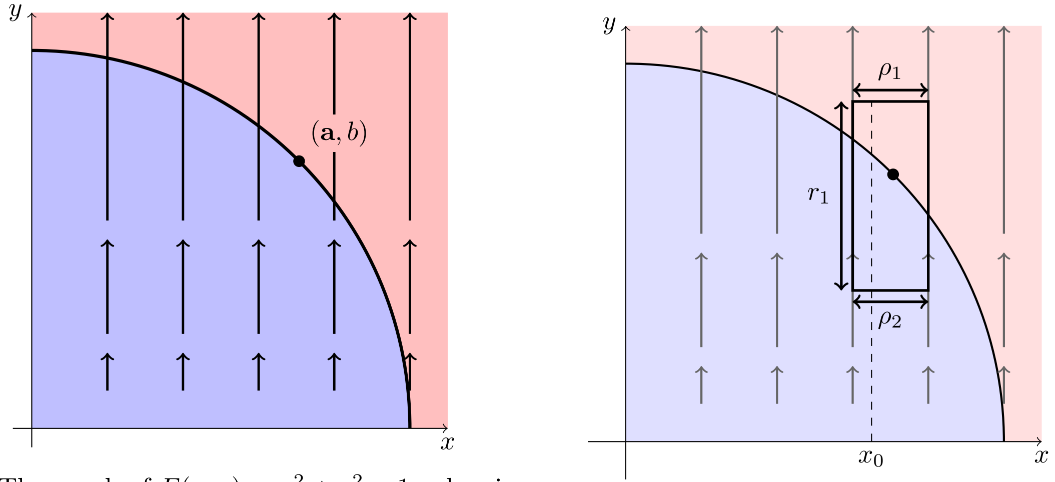
Theorem 3.69: Implicit Function Theorem

Let $U \subseteq \mathbb{R}^{n+1}$ be an open neighbourhood. Suppose $F(\mathbf{x}, y)$ is a C^1 function on U and $(\mathbf{a}, b) \in U$ satisfies $F(\mathbf{a}, b) = 0$. If $\partial_y F(\mathbf{a}, b) \neq 0$, then there exists an open set $V \subseteq \mathbb{R}^n$ together with a unique C^1 function $f : V \to \mathbb{R}$ such that $F(\mathbf{x}, f(\mathbf{x})) = 0$ for all $x \in V$. Moreover, the derivative of f is given by

$$\partial_{x_i} f(\mathbf{x}) = -\frac{\partial_{x_i} F(\mathbf{x}, y)}{\partial_y F(\mathbf{x}, y)}. \tag{3.13}$$

Proof. We break our proof into several steps: we begin by showing that there is an $r > 0$ such that for each $x_0 \in B_r(\mathbf{a})$ there exists a unique y_0 such that $F(\mathbf{x}_0, y_0) = 0$. We call the mapping which takes $\mathbf{x}_0 \mapsto y_0$ the function $f(\mathbf{x}, y)$. After this, we show that this function is actually differentiable, with the prescribed derivative.

Existence and Uniqueness: The spirit of this part of the proof is akin to showing that a single variable function with positive derivative is injective. Without loss of generality, assume that $\partial_y F(\mathbf{a}, b) > 0$, so that there exists $r_1 > 0$ such that $\partial_y F(\mathbf{x}, y) > 0$ for all $(\mathbf{x}, y)$ in $B_{r_1}(\mathbf{a}, b) \subseteq \mathbb{R}^{n+1}$. By taking smaller r_1 if necessary, we can ensure that $B_{\mathbf{a},b}(r_1) \subseteq U$.



(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.

Since $\partial_y F$ is positive on $B_{\mathbf{a},b}(r_1)$, it is increasing in the y-direction, so

$$F(\mathbf{a}, b - r_1) < 0 \qquad \text{and} \qquad F(\mathbf{a}, b + r_1) > 0.$$

Once again, by continuity there exists $\rho_1, \rho_2 > 0$ such that $F(\mathbf{x}, b - r_1) < 0$ for all $\mathbf{x} \in B_{\rho_1}(\mathbf{a})$ and $F(\mathbf{x}, b + r_1) > 0$ for all $\mathbf{x} \in B_{\rho_2}(\mathbf{a})$. 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 $F(\mathbf{x}_0, y)$ is strictly increasing as a function of y, this y is unique by the Mean Value Theorem.

Differentiability: Fix some $\mathbf{x}_0 \in B_r(\mathbf{a})$, set $y_0 = f(\mathbf{x}_0)$, and choose $h \in \mathbb{R}$ sufficiently small so that $\mathbf{h}_i = h\mathbf{e}_i$ satisfies $\mathbf{x}_0 + \mathbf{h}_i \in B_r(\mathbf{a})$. Define $k = f(\mathbf{x}_0 + \mathbf{h}_i) - f(\mathbf{x}_0)$ to be the ith 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 $B_r(\mathbf{a})$, so by the Mean Value Theorem there exists some $t \in (0, 1)$ such that

$$0 = F(\mathbf{x}_0 + \mathbf{h}_i, y_0 + k) - F(\mathbf{x}, y_0) = h\partial_{x_i} F(\mathbf{x}_0 + t\mathbf{h}_i, y_0 + tk) + k\partial_y F(\mathbf{x}_0 + t\mathbf{h}_i, y_0 + tk).$$

Re-arranging we can write

$$\frac{f(\mathbf{x}_0 + \mathbf{h}_i) - f(\mathbf{x}_0)}{h} = \frac{k}{h} = -\frac{\partial_{x_i} F(\mathbf{x}_0 + t\mathbf{h}_i, y + tk)}{\partial_y F(\mathbf{x}_0 + t\mathbf{h}_i, y_0 + tk)}.$$

As the quotient on the right-hand-side consists of continuous functions and $\partial_y F \neq 0$ in $B_r(\mathbf{a}, b)$, taking the $h \to 0$ limit yields

$$\frac{\partial f}{\partial x_i}(\mathbf{x}_0, y_0) = -\frac{\partial_{x_i} F(\mathbf{x}_0, y_0)}{\partial_y F(\mathbf{x}_0, y_0)},$$

which is a continuous function. □

A useful consequence of the proof of Theorem 3.69 is equation (3.13) 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

$$0 = \frac{\partial F}{\partial x_j}(\mathbf{x}) + \frac{\partial F}{\partial y}(\mathbf{x}, f(\mathbf{x}))\frac{\partial f}{\partial x_j}(\mathbf{x}) \qquad \Rightarrow \qquad \partial_{x_j} f(\mathbf{x}) = -\frac{\partial_{x_j} F(\mathbf{x}, f(\mathbf{x}))}{\partial_y F(\mathbf{x}, f(\mathbf{x}))}$$

which agrees with what we found in the course of the proof.

It's not worthwhile memorizing this formula. In fact, you already know how to compute $\partial_{x_j} f$, via implicit differentiation. When you learned the process of implicit differentiation in your first year analysis course, you assumed that $y = f(x)$ was something that made sense. It is precisely the Implicit Function Theorem above which justifies that procedure. Let's see this by recalling a straightforward example.

Example 3.70

Consider the set $\mathbb{S}^1 = \left\{(x, y) : x^2 + y^2 = 1\right\}$. Determine when y can be written as a function $y = f(x)$, and find the derivative $f'(x)$.

Solution. The circle $\mathbb{S}^1$ is defined by the zero locus of $F(x, y) = x^2 + y^2 - 1$, and cannot globally be solved for either x or y. However, $\nabla F(x, y) = (2x, 2y)$, which means that whenever $y \neq 0$ we solve for y in terms of xIndeed, 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 (3.13) 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 $y = \pm\sqrt{1 - x^2}$. ■

Example 3.71

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

$$\nabla F(x, y, z) = \frac{1}{\sqrt{2x + y^3 - z^2}}\left(1, \tfrac{3}{2}y^2, -z - \sqrt{2x + y^3 - z^2}\sin(z)\right).$$

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 $x = f(y,z)$.

$$\begin{aligned}\frac{\partial f}{\partial y} &= -\frac{\partial_y F}{\partial_x F} = -\frac{3}{2}y^2 \\ \frac{\partial f}{\partial z} &= -\frac{\partial_z F}{\partial_x F} = -z + \sin(z)\sqrt{2x+y^3-z^2}.\end{aligned}$$

Similarly, for $y = g(x,z)$ we have

$$\begin{aligned}\frac{\partial g}{\partial x} &= -\frac{\partial_x F}{\partial_y F} = \frac{2}{3y^2} \\ \frac{\partial g}{\partial z} &= -\frac{\partial_z F}{\partial_y F} = -\frac{2z + 2\sin(z)\sqrt{2x+y^3-z^2}}{3y^2}.\end{aligned}$$

Again, the student may check that this is consistent with implicit differentiation. ■

3.9.2 The General Case

Consider a C^1 function $\mathbf{F} : \mathbb{R}^{n+k} \to \mathbb{R}^k$. The hypotheses in this case are not too difficult. The major change will be evaluating what the analogous condition to $\partial_y \mathbf{F} \neq 0$ 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 $\mathbf{F}(\mathbf{x}, \mathbf{y})$ 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 $\mathbf{F}(\mathbf{x},\mathbf{y}) = A\mathbf{x} + B\mathbf{y} + \mathbf{c}$ for some $\mathbf{c} \in \mathbb{R}^k$. If $(\mathbf{x}_0, \mathbf{y}_0)$ is some point where $\mathbf{F}(\mathbf{x}_0,\mathbf{y}_0) = \mathbf{0}$, then we can express $\mathbf{y}$ as a function of $\mathbf{x}$ if and only if the matrix B is invertible.

If $\mathbf{F}(\mathbf{x},\mathbf{y})$ is now a general function, the set $\mathbf{F}(\mathbf{x},\mathbf{y}) = 0$ defines at surface of dimension at most n in $\mathbb{R}^{n+k}$. Let $\mathbf{F}(\mathbf{x},\mathbf{y}) = (F_1(\mathbf{x},\mathbf{y}), \ldots, F_k(\mathbf{x},\mathbf{y}))$ for C^1 functions $F_i : \mathbb{R}^{n+k} \to \mathbb{R}$. The Jacobian of $\mathbf{F}(\mathbf{x},\mathbf{y})$ is given by

$$D\mathbf{F}(\mathbf{x}) = \left[\underbrace{\begin{matrix}\partial_{x_1}F_1(\mathbf{x},\mathbf{y}) & \cdots & \partial_{x_n}F_1(\mathbf{x},\mathbf{y}) \\ \vdots & \ddots & \vdots \\ \partial_{x_1}F_k(\mathbf{x},\mathbf{y}) & \cdots & \partial_{x_n}F_k(\mathbf{x},\mathbf{y})\end{matrix}}_{A(\mathbf{x},\mathbf{y})\in M_{k\times n}(C^1(\mathbb{R}^{n+k}))} \middle| \underbrace{\begin{matrix}\partial_{y_1}F_1(\mathbf{x},\mathbf{y}) & \cdots & \partial_{y_k}F_1(\mathbf{x},\mathbf{y}) \\ \vdots & \ddots & \vdots \\ \partial_{y_1}F_k(\mathbf{x},\mathbf{y}) & \cdots & \partial_{y_k}F_k(\mathbf{x},\mathbf{y})\end{matrix}}_{B(\mathbf{x},\mathbf{y})\in M_{k\times k}(C^1(\mathbb{R}^{n+k}))}\right]$$

so 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 our analogy to $\partial_y F \neq 0$ from the single-variable case should be to require $\left(\partial_{y_j}F_i(\mathbf{a},\mathbf{b})\right)_{ij}$ to be an invertible matrix.

Theorem 3.72: General Implicit Function Theorem

Let U be an open subset of $\mathbb{R}^{n+k}$ and $\mathbf{F} : U \to \mathbb{R}^k$ be a C^1 function. 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 $\mathbf{F}(\mathbf{a},\mathbf{b}) = \mathbf{0}$ and $B_{ij} = \left(\partial_{y_j}F_i(\mathbf{a},\mathbf{b})\right)_{ij}$ is invertible, there exists an open set $V \subseteq \mathbb{R}^n$ and a unique C^1 function $\mathbf{f} : V \to \mathbb{R}^k$ such that $\mathbf{F}(\mathbf{x},\mathbf{f}(\mathbf{x})) = 0$ for all $\mathbf{x} \in V$.

Proof. I will give the proof when $k = 2$. The general proof is done via induction, but is a good exercise in symbol-pushing an linear algebra. I'll guide you through it in Exercise 3-49.

Suppose then that $\mathbf{F} : \mathbb{R}^{n+2} \to \mathbb{R}^2$ is of the form $F(\mathbf{x}, y_1, y_2) = (F_1(\mathbf{x}, y_1, y_2), F_2(\mathbf{x}, y_1, y_2))$ and $(\mathbf{a}, b_1, b_2)$ satisfies $\mathbf{F}(\mathbf{a}, b_1, b_2) = 0$. The matrix B has the form

$$B = \begin{bmatrix} \partial_{y_1} F_1(\mathbf{a}, b_1, b_2) & \partial_{y_2} F_1(\mathbf{a}, b_1, b_2) \\ \partial_{y_1} F_2(\mathbf{a}, b_1, b_2) & \partial_{y_2} F_2(\mathbf{a}, b_1, b_2) \end{bmatrix}.$$

As B is invertible, $\det(B) \neq 0$ and one of its entries B_{ij} must be non-zero. By re-arranging the entries if necessary, we can assume that $\partial_{y_1} F_1(\mathbf{a}, b_1, b_2) \neq 0$. By the Implicit Function Theorem, there exists a neighbourhood $\tilde{V} \subseteq \mathbb{R}^{n+1}$ and a C^1-function $f : V \to \mathbb{R}$ such that $y_1 = f(\mathbf{x}, y_2)$ satisfies $F_1(\mathbf{x}, f(\mathbf{x}, y_2), y_2) = 0$. Moreover,

$$\partial_{y_2} f(\mathbf{x}, y_2) = -\frac{\partial_{y_2} F_1(\mathbf{x}, y_2)}{\partial_{y_1} F_1(\mathbf{x}, y_2)}. \tag{3.14}$$

Define a function $G : \mathbb{R}^{n+1} \to \mathbb{R}$ by $G(\mathbf{x}, y_2) \mapsto F_2(\mathbf{x}, f(\mathbf{x}, y_2), y_2)$. If we can apply the Implicit Function Theorem to G with respect to y_2, then we'll have a neighbourhood $V \subseteq \mathbb{R}^n$ and a C^1-function $g : V \to \mathbb{R}$ such that $y_2 = g(\mathbf{x})$. Then $y_1 = f(\mathbf{x}, g(\mathbf{x}))$ is a function of the $\mathbf{x}$ alone, and we define $\mathbf{f} : V \to \mathbb{R}^2$ by $\mathbf{f}(\mathbf{x}) = (f(\mathbf{x}, g(\mathbf{x})), g(\mathbf{x}))$, completing the proof. Thus it suffices to show that we can apply the Implicit Function Theorem to G with respect to y_2, which amounts to showing that $\partial_{y_2} G(\mathbf{a}, b_2) \neq 0$.

To compute $\partial_{y_2} G$, we differentiate via the chain rule:

$$\begin{aligned} \partial_{y_2} G(\mathbf{a}, b_2) &= \partial_{y_1} F_2(\mathbf{a}, b_2) \partial_{y_2} f(\mathbf{a}, b_2) + \partial_{y_2} F_2(\mathbf{a}, b_2) \\ &= \partial_{y_1} F_2(\mathbf{a}, b_2) \left[-\frac{\partial_{y_2} F_1(\mathbf{a}, b_2)}{\partial_{y_1} F_1(\mathbf{a}, b_2)} \right] + \partial_{y_2} F_2(\mathbf{a}, b_2) && \text{from (3.14)} \\ &= \frac{\partial_{y_1} F_1(\mathbf{a}, b_2) \partial_{y_2} F_2(\mathbf{a}, b_2) - \partial_{y_1} F_2(\mathbf{a}, b_2) \partial_{y_2} F_1(\mathbf{a}, b_2)}{\partial_{y_1} F_1(\mathbf{a}, b_2)} \\ &= \frac{\det(B)}{\partial_{y_1} F_1(\mathbf{a}, b_2)} \end{aligned}$$

By assumption, both numerator and denominator are non-zero, hence $\partial_{y_2} G(\mathbf{a}, b_2) \neq 0$ as required. □

Example 3.73

Consider the function

$$\mathbf{F}(x, y, u, v) = \begin{bmatrix} x^2 - y^2 - u^3 + v^2 + 4 \\ 2xy + y^2 - 2u^2 + 3v^4 + 8 \end{bmatrix}.$$

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

$$D_{(u,v)} \mathbf{F} = \begin{bmatrix} -3u^2 & 2v \\ -4u & 12v^3 \end{bmatrix}, \qquad \Rightarrow \qquad D_{(u,v)} \mathbf{F}(2, -1, 2, 1) = \begin{bmatrix} -12 & 2 \\ -8 & 12 \end{bmatrix}$$

which has determinant $-128 \neq 0$ and so is invertible. By Theorem 3.72 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 $v = g_2(x, y)$.

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

$$\begin{aligned}
&\begin{bmatrix} 2x - 3u^2\partial_x u + 2v\partial_x v \\ 2y - 4u\partial_x u + 12v^3\partial_x v \end{bmatrix} = 0 \\
\Leftrightarrow &\begin{bmatrix} -3u^2 & 2v \\ -4u & 12v^3 \end{bmatrix} \begin{bmatrix} \partial_x u \\ \partial_x v \end{bmatrix} = \begin{bmatrix} -2x \\ 2y \end{bmatrix} \\
\Leftrightarrow &\begin{bmatrix} \partial_x u \\ \partial_x v \end{bmatrix} = \begin{bmatrix} -3u^2 & 2v \\ -4u & 12v^3 \end{bmatrix}^{-1} \begin{bmatrix} -2x \\ -2y \end{bmatrix} \\
\Leftrightarrow &\begin{bmatrix} \partial_x u \\ \partial_x v \end{bmatrix} = \frac{1}{8uv - 36u^2v^3} \begin{bmatrix} -24xv^3 + 4vy \\ -8ux + 6u^2y \end{bmatrix}.
\end{aligned}$$

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. ■

3.9.3 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 C^1-functions 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 3.74: 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

$$[D\mathbf{f}^{-1}](\mathbf{b}) = [D\mathbf{f}(\mathbf{a})]^{-1}. \tag{3.15}$$

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. Both theorems are special cases of a theorem called the *Constant Rank 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 (3.15) 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 3.75

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

$$D\mathbf{f}_{(x,y)} = \begin{bmatrix} e^x \sin(y) & e^x \cos(y) \\ e^x \cos(y) & -e^x \sin(y) \end{bmatrix}.$$

Evaluating at $(0, 0)$ we get

$$D\mathbf{f}_{(0,0)} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

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. ■

3.10 Partitions Of Unity

Partitions of unity are technical but important tools. We'll use them almost exclusively when for integrating, but we can state and prove their existence using only differentiability.

Definition 3.76

If $f : (X, d_X) \to \mathbb{R}$, we define the *support of* f as the closure of the collection of points where f is non-zero; that is,

$$\operatorname{supp}(f) = \overline{\{\mathbf{x} \in X : f(\mathbf{x}) \neq 0\}}.$$

We say that f has *compact support* if $\operatorname{supp}(f)$ is compact in X.

Theorem 3.77

If $\mathcal{O}$ be a collection of open sets, with $U = \bigcup \mathcal{O}$, then there exists a countable collection of non-negative, C^∞, compactly supported functions $\phi_i : U \to \mathbb{R}$ such that

1. [Subordinate] For every $S_i = \operatorname{supp}(\phi_i)$, there exists an $O \in \mathcal{O}$ such that $S_i \subseteq O$,
2. [Locally Finite] For every $\mathbf{x} \in U$, there is a neighbourhood N of $\mathbf{x}$ which intersects only finitely many of the S_i,
3. [Partition of Unity] For every $\mathbf{x} \in U$, $\sum_i \phi_i(\mathbf{x}) = 1$.

The set $\{\phi_i : i \in \mathbb{N}\}$ is said to be a *(smooth, compactly supported) partition of unity subordinate to* $\mathcal{O}$.

Proof. The proof is largely a collection of previously completed exercises.

From Exercise 2-71, we know we can find a countable collection of closed balls $\mathcal{B} = \{B_n : n \in I \subseteq \mathbb{N}\}$ such that:

1. $U = \bigcup_{i \in I} B_n^{\text{int}}$,
2. For every $i \in I$, there is an $O \in \mathcal{O}$ such that $B_i \subseteq O$,
3. Every point $\mathbf{x} \in U$ has a neighbourhood N which only intersects finitely many of the B_i non-trivially.

By Exercise 3-63, for each $i \in I$ we can find a bump function $\psi_i : U \to \mathbb{R}$ such that $\psi_i(\mathbf{x}) \neq 0$ for all $x \in B_n$; that is, $\operatorname{supp}(\psi_i) = B_n$, which is compact. At this point we're pretty much done. All that remains is to massage the ψ_i to ensure they always sum to 1.

Define $\psi(\mathbf{x}) = \sum_{i \in I} \psi_i(\mathbf{x})$. Since each $\mathbf{x} \in N$ only intersects finitely many of the B_n non-trivially, all but finitely many of the terms in the sum are non-zero, and hence $\psi(\mathbf{x})$ converges for all $\mathbf{x} \in U$. Moreover, since the interiors of the B_n cover U, we know ψ is never zero. It is also C^∞, as at every point it is a finite sum of C^∞ functions.

Define $\phi_i(\mathbf{x}) = \psi_i(\mathbf{x})/\psi(\mathbf{x})$. Note that $\operatorname{supp}(\phi_i) = \operatorname{supp}(\psi_i)$ and hence is compact. This is well defined as its denominator is never zero, it's a C^∞ function, and $\sum_i \phi_i(\mathbf{x}) = 1$ for all $\mathbf{x} \in U$. Thus the ϕ_i form a partition of unity as required. □

3.11 Exercises

3-1. Use the limit definition of the derivative to differentiate the following functions at the given point:

(a) $f(x,y) = x^2y^2, \mathbf{a} = (2,-1)$

(b) $f(x,y,z) = xy/z, \mathbf{a} = (1,2,3)$

(c) $f(x) = (2xy, \sqrt{xy}), \mathbf{a} = (2,2)$

(d) $f(x,y) = (y,x), \mathbf{a} = (a,b)$

(e) $f(x,y) = (x, x^2, x^3, \ldots, x^n), n \in \mathbb{N}, \mathbf{a} = (1, \ldots, 1)$

(f) $f(x,y,z) = (xy, xz, yz), \mathbf{a} = (-1,1,-1)$

3-2. Let $\mathbf{a} \in \mathbb{R}^n$ and U be an open neighbourhood of $\mathbf{a}$. One can define the derivative $f : U \to \mathbb{R}^m$ in several different ways:

(a) $\lim\limits_{\mathbf{h}\to 0} \dfrac{f(\mathbf{a}+\mathbf{h}) - f(\mathbf{a}) - A(\mathbf{h})}{\|\mathbf{h}\|} = 0,$

(b) $\lim\limits_{\mathbf{h}\to 0} \dfrac{f(\mathbf{a}+\mathbf{h}) - f(\mathbf{a}) - A(\mathbf{h})}{\|\mathbf{h}\|_\infty} = 0$ where $\|\mathbf{x}\|_\infty = \max\limits_{i=1,\ldots,n} |x_i|,$

(c) $\lim\limits_{\mathbf{h}\to 0} \dfrac{\|f(\mathbf{a}+\mathbf{h}) - f(\mathbf{a}) - A(\mathbf{h})\|}{\|\mathbf{h}\|_\infty} = 0 \ .$

Show these definition are all equivalent to (3.2)

3-3. Suppose $\|\cdot\|_1$ and $\|\cdot\|_2$ are equivalent norms on a real vector space V, and let $f : V \to \mathbb{R}^n$ be a map. Show that f is differentiable in $\|\cdot\|_1$ if and only if it is differentiable in $\|\cdot\|_2$.

3-4. Suppose $\mathbf{f} : \mathbb{R}^n \to \mathbb{R}^m$ is a linear map. Show that $D\mathbf{f}(\mathbf{a}) = \mathbf{f}(\mathbf{a})$.

3-5. Prove Proposition 3.12; that is, show that $f : \mathbb{R}^n \to \mathbb{R}^m$, $\mathbf{f}(\mathbf{x}) = (f_1(\mathbf{x}), f_2(\mathbf{x}), \ldots, f_m(\mathbf{x}))$ is differentiable if and only each of its component functions $f_i : \mathbb{R}^n \to \mathbb{R}$ is differentiable, and moreover $[D\mathbf{f}(\mathbf{a})]_{i,j} = \partial_j f_i(\mathbf{a})$.

3-6. Consider the map $\mathbf{f} : \mathbb{R}^2 \to M_2(\mathbb{R})$ given by

$$(x,y) \mapsto \begin{bmatrix} x & y \\ -y & x \end{bmatrix}.$$

Show that $\mathbf{f}$ is differentiable and compute $D\mathbf{f}$.

3-7. Let $(V, \|\cdot\|_V)$ and $(W, \|\cdot\|_W)$ be normed vector spaces.

(a) Determine what it should mean for $f : V \to W$ to be a differentiable function.

(b) Fix a basis for V. Use this to define partial derivatives and C^1 functions. Generalize the proof of Theorem 3.10 to this case.

(c) Define a suitable notion of the directional derivative. Show that the directional derivative is a linear map.

3-8. Compute the partial derivatives $\partial_i f$, and the total derivative Df for each of the following functions $f : \mathbb{R}^n \to \mathbb{R}^k$:

(a) $f(x, y, z) = x^y$

(b) $f(x, y) = x^y$

(c) $f(x, y, z) = (x^y, z)$

(d) $f(x, y) = \sin(x \sin(y))$

(e) $f(x, y, z) = (x + y)^z$

(f) $f(x, y, z) = \left(\log\left(x^2 + y^2 + z^2\right), xyz\right)$

(g) $f(x, y) = \sin(xy)$

3-9. Show that the converse of Theorem 3.17 is false; that is, there is a function in which every directional derivative exists at a point, but the function is not differentiable.

3-10. Suppose $\langle \cdot, \cdot \rangle$ is an inner product on $\mathbb{R}^n$, and $\mathbf{f}, \mathbf{g} : \mathbb{R} \to \mathbb{R}^n$. Show that

$$\frac{\mathrm{d}}{\mathrm{d}t} \langle \mathbf{f}(t), \mathbf{g}(t) \rangle = \left\langle \mathbf{f}'(t), \mathbf{g}(t) \right\rangle + \left\langle \mathbf{f}(t), \mathbf{g}'(t) \right\rangle .$$

3-11. In Example 3.27 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 3.27.

3-12. Prove Corollary 3.32: Suppose $U \subseteq \mathbb{R}^n$ is convex and $\mathbf{f} : U \to \mathbb{R}^m$ is differentiable. If there exists an $M > 0$ such that $\|D\mathbf{f}(\mathbf{x})\| \leq M$ for all $\mathbf{x} \in U$, show that $\|\mathbf{f}(\mathbf{b}) - \mathbf{f}(\mathbf{a})\| \leq M\|\mathbf{b} - \mathbf{a}\|$ for all $\mathbf{a}, \mathbf{b} \in U$.

3-13. Generalize Corollary 3.33 by showing that if U is open and connected with $\mathbf{f} : U \to \mathbb{R}^m$ differentiable satisfying $D\mathbf{f}(\mathbf{x}) = \mathbf{0}$ for all $\mathbf{x} \in U$, then $\mathbf{f}$ is constant.

3-14. Suppose $U \subseteq \mathbb{R}^n$ is open and connected. Let $\mathbf{f}, \mathbf{g} : U \to \mathbb{R}^m$ be differentiable and satisfy $D\mathbf{f}(\mathbf{x}) = D\mathbf{g}(\mathbf{x})$ for all $\mathbf{x} \in U$. Show that $\mathbf{f} = \mathbf{g}$ on U.

3-15. Prove Corollary 3.33: Suppose $U \subseteq \mathbb{R}^n$ is convex and $\mathbf{f} : U \to \mathbb{R}^m$ is differentiable. If $D\mathbf{f}(\mathbf{x}) = \mathbf{0}$ for all $\mathbf{x} \in U$, then $\mathbf{f}$ is a constant function.

3-16. Suppose $\mathbf{f} : \mathbb{R}^n \to \mathbb{R}^m$ is twice differentiable. Show that

$$D^2\mathbf{f}(\mathbf{v}, \mathbf{w}) = \lim_{t \to 0} \frac{f(\mathbf{a} + t\mathbf{v} + t\mathbf{w}) - f(\mathbf{a} + t\mathbf{v}) - f(\mathbf{a} + t\mathbf{w}) + f(\mathbf{a})}{t^2}.$$

3-17. Let $\mathbf{f} : \mathbb{R}^2 \to \mathbb{R}^2$ be the map $f(x, y) = (x^2 - y^2, 2xy)$.

(a) Calculate $D\mathbf{f}$ and $\det D\mathbf{f}$

(b) Let $S = \left\{(x, y) \in \mathbb{R}^2 : x^2 + y^2 \leq 1, x \geq 0, y \geq 0\right\}$. Make a sketch of $f(S)$ by showing where some of the coordinate curves get mapped.

3-18. Suppose that $\mathbf{f} : \mathbb{R}^3 \to \mathbb{R}^2$ is a function such that $\mathbf{f}(0, 0, 0) = (1, 2)$ and:

$$D\mathbf{f}_{(0,0,0)} = \begin{pmatrix} 1 & 2 & 3 \\ 0 & 0 & 1 \end{pmatrix}$$

Let $\mathbf{g} : \mathbb{R}^2 \to \mathbb{R}^2$ be the map $\mathbf{g}(x, y) = (x + 2y + 1, 3xy)$. Find $D(\mathbf{f} \circ \mathbf{g})_{(0,0,0)}$

3-19. Find an equation for the tangent plane T_pS to the following surfaces at the indicated point

(a) $S = \{(x, y, z) : x^2 + 2y^2 + 3z^2 = 6\}$ at $(1, 1, -1)$.

(b) $S = \{(x, y, z) : xyz^2 - \log(z - 1) = 8\}$ at $(-2, -1, 2)$.

(c) $S = \{(x, y, z) : x^2 + y^2 = 1\}$ at $(1/\sqrt{2}, 1/\sqrt{2}, 1)$

3-20. Suppose that $f(x, y, z, t)$, $x(t)$, $y(x, t, s)$, and $z(y, x)$. Use the chain rule to find an expression for $\frac{\partial f}{\partial t}$ and $\frac{\partial f}{\partial s}$.

3-21. Finish the proof of Proposition 3.22.

3-22. Show that $f : \mathbb{R}^2 \to \mathbb{R}$, $f(x, y) = \sqrt{|xy|}$ is not differentiable at $(x, y) = (0, 0)$.

3-23. Let $F(\varphi, \theta) = (x(\varphi, \theta), y(\varphi, \theta), z(\varphi, \theta))$ be the spherical polar coordinate system on the unit sphere $S^2 = \{(x, y, z)\}\, x^2 + y^2 + z^2 = 1$.

(a) Sketch the coordinate curves $\theta = \pi/4$ and $\varphi = \pi/2$

(b) Compute the derivative to the coordinate curves from part (1) at the point $(\varphi, \theta) = (\pi/2, \pi/4)$. Add these arrows to your plot.

(c) Prove that $\partial_\varphi F \times \partial_\theta F$ is parallel to $\nabla(x^2 + y^2 + z^2)$.

3-24. Let $f : \mathbb{R}^3 \to \mathbb{R}$, and suppose that $\nabla f \neq 0$. Prove that if $F : U \subseteq \mathbb{R}^2 \to \mathbb{R}^3$ is a differentiable parameterization of a level set $S = f^{-1}(c)$, then $\exists\, \lambda \in \mathbb{R}$ such that $\partial_u F \times \partial_v F = \lambda \nabla f$.

3-25. Let $\mathbf{f}, \mathbf{g} : \mathbb{R}^n \to \mathbb{R}^m$. If $\mathbf{f}$ and $\mathbf{g}$ are differentiable at $\mathbf{x} \in \mathbb{R}^n$, then $D(\mathbf{f} + \mathbf{g})_\mathbf{x} = D\mathbf{f}_\mathbf{x} + D\mathbf{g}_\mathbf{x}$ and $D(\mathbf{fg})_\mathbf{x} = \mathbf{f}(\mathbf{x})D\mathbf{g}_\mathbf{x} + \mathbf{g}(\mathbf{x})D\mathbf{f}_\mathbf{x}$. Notice these are generalizations of the sum and product rules for differentiation from last year.

3-26. (a) For each function in Exercise 3-1, find $D^2 f(\mathbf{a})$.

(b) For each function in Exercise 3-1, find $D^3 f(\mathbf{a})$.

3-27. Let $k \in \mathbb{N}$ and U an open subset of $\mathbb{R}^n$.

(a) Show that $C^k(U, \mathbb{R})$ is a real vector space.

(b) If $m > k$, show that $C^m(U, \mathbb{R})$ is a subspace of $C^k(U, \mathbb{R})$.

3-28. Show that every symmetric matrix $A \in M_n(\mathbb{R})$ arises as $D^2 f$ for some function $f : \mathbb{R}^n \to \mathbb{R}$.

3-29. A function $f : \mathbb{R}^n \to \mathbb{R}$ is *homogeneous of order* k if for every $\lambda \in \mathbb{R}$ and $\mathbf{x} \in \mathbb{R}^n$, $f(\lambda \mathbf{x}) = \lambda^k f(\mathbf{x})$. Show that if f is homogeneous of order k, then $\mathbf{x} \cdot \nabla f(\mathbf{x}) = k f(\mathbf{x})$.

3-30. Let $f : U \subseteq \mathbb{R}^n \to \mathbb{R}$ be a differentiable function, where U is star convex about a point $\mathbf{p} \in U$, and $f(\mathbf{p}) = 0$. Show that $\exists\, g_1, \ldots, g_n : U \to \mathbb{R}$ such that $f(\mathbf{x}) = \sum_{i=1}^n x_i g_i(\mathbf{x})$, where $\mathbf{x} = (x_1, \ldots, x_n)$. *Hint:* You may as well assume that U contains the origin and $p = 0$. Consider the function $f(t\mathbf{x})$.

3-31. Let $C([0, 1])$ be the collection of all continuous functions on the closed unit interval, given the sup norm as above. Consider the map:

$$F = \int_0^t : C([0, 1]) \to C([0, 1]), f \mapsto \int_0^t f(x)\, \mathrm{d}x$$

Compute DF_f. *Hint:* What quantity do you need to estimate when computing a derivative? Think about the properties of integration.

3-32. Fix a rectangle $R \subseteq \mathbb{R}^n$ and an interval $I \subseteq \mathbb{R}$. Write $(\mathbf{x}, t) \in R \times I \subseteq \mathbb{R}^n \times \mathbb{R}$ for our coordinates, and let $f : R \times I \to \mathbb{R}$ be such that $\partial_t f$ is continuous. Define

$$\phi : R \to \mathbb{R}, \qquad \phi(\mathbf{x}) = \int_I f(\mathbf{x}, t)\, \mathrm{d}t.$$

Show that $\partial_k \phi$ exists for all $\mathbf{x} \in R^{\text{int}}$, and moreover

$$\frac{\partial \phi}{\partial k}(\mathbf{x}) = \int_I \frac{\partial f}{\partial k}(\mathbf{x}, t)\, \mathrm{d}t.$$

> "That book also showed how to differentiate parameters under the integral sign – it's a certain operation. It turns out that's not taught very much in the universities; they don't emphasize it. But I caught on how to use that method, and I used that one damn tool again and again. So because I was self-taught using that book, I had peculiar methods of doing integrals.
>
> The result was, when guys at MIT or Princeton had trouble doing a certain integral, it was because they couldn't do it with the standard methods they had learned in school. If it was contour integration, they would have found it; if it was a simple series expansion, they would have found it. Then I come along and try differentiating under the integral sign, and often it worked. So I got a great reputation for doing integrals, only because my box of tools was different from everybody else's, and they had tried all their tools on it before giving the problem to me."
>
> – Richard Feynman, Surely You're Joking, Mr. Feynman.

3-33. Let $S \subseteq \mathbb{R}^n$ be an open set and $\mathbf{f} : S \to \mathbb{R}^m$ be a C^1 function. We say that $\mathbf{f}$ is C^1 on $\overline{S}$ if for every $\mathbf{p} \in \overline{S}$, there is an open neighbourhood $U_\mathbf{p} \subseteq \mathbb{R}^n$ of $\mathbf{p}$ and a C^1 function $\tilde{\mathbf{f}} : U_\mathbf{p} \to \mathbb{R}^m$ such that $\mathbf{f}|_{S \cap U_\mathbf{p}} = \tilde{\mathbf{f}}|_{S \cap U_\mathbf{p}}$. Show that $D\tilde{\mathbf{f}}(\mathbf{p})$ is independent of the choice $\tilde{\mathbf{f}}$ and $U_\mathbf{p}$, and therefore there is no ambiguity in writing $D\mathbf{f}$ to mean the derivative of $\mathbf{f}$ on $\overline{S}$.

3-34. Let $f, g : \mathbb{R}^2 \to \mathbb{R}$ and $h : \mathbb{R}^3 \to \mathbb{R}$ be C^1 functions and define $F(x, y) = \displaystyle\int_{f(x,y)}^{g(x,y)} h(x, y, t) \mathrm{d}t$. Compute $\dfrac{\partial F}{\partial x}, \dfrac{\partial F}{\partial y}$.

3-35. Let $\mathbf{G} : \mathbb{R}^2 \to \mathbb{R}^2$ be a C^1 function satisfying $\mathbf{G}(\mathbf{0}) = \mathbf{0}$. Define a function $\mathbf{H}(\mathbf{x}) = \mathbf{G}^{\circ n}(\mathbf{x})$ where $\mathbf{G}^{\circ n}$ denotes the n-fold composition of $\mathbf{G}$ with itself. If

$$D\mathbf{G}(\mathbf{0}) = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$$

determine $D\mathbf{H}(\mathbf{0})$.

3-36. Let $f(x, y) = e^x \cos(y)$. Verify by hand that $\partial_x \partial_y f = \partial_y \partial_x f$.

3-37. Prove the following version of the binomial theorem: For any pair of points $x, y \in \mathbb{R}^n$,

$$(x + y)^\alpha = \sum_{\beta + \gamma = \alpha} \frac{\alpha!}{\beta! \gamma!} x^\beta y^\gamma.$$

3-38. Derive the following version of the "product rule" for partial derivatives; if α is any multi-index, then:

$$\partial^\alpha(fg) = \sum_{\beta+\gamma=\alpha} \frac{\alpha!}{\beta!\gamma!}\partial^\beta f \partial^\gamma g$$

3-39. Define $f : \mathbb{R}^2 \to \mathbb{R}$ by

$$f(x,y) = \begin{cases} \frac{x^3y-xy^3}{x^2+y^2} & (x,y) \neq (0,0) \\ 0 & (x,y) = (0,0) \end{cases}.$$

Compute the mixed partial derivatives $\partial_{xy}f, \partial_{yx}f$. Why does this not contradict Clairut's theorem?

3-40. Find the 3rd order Taylor polynomial of the following functions:

- $f(x,y) = \sin(x)\cos(y)$ based at the point $(0,0)$.
- $f(x,y) = \frac{1}{1+x-y}$ based at the point $(0,0)$.
- $f(x,y) = \log(1+x-y)$ based at the point $(0,0)$.
- $f(x,y) = x + \cos(\pi y) + x\log y$ based at the point $(3,1)$
- $f(x,y,z) = x^2y + z$ based at $(1,2,1)$. Why should the remainder be zero?

3-41. Find all the critical points of the following functions. Say whether the critical points are local maxima, local minima, saddle points, or otherwise.

(a) $f(x,y) = x^4 - 2x^2 + y^3 - 6y$

(b) $f(x,y) = (x-1)(x^2-y^2)$

(c) $f(x,y) = x^2y^2(1-x-y)$

(d) $f(x,y) = (2x^2+y^2)e^{-x^2-y^2}$

(e) $f(x,y,z) = xyz(4-x-y-z)$

3-42. Find the extreme values of $f(x,y,z) = x^2 + 2y^2 + 3z^2$ on the unit sphere, $x^2+y^2+z^2 = 1$.

3-43. What conditions on a, b, and c guarantee that $f(x,y) = ax^2 + bxy + cy^2$ has local max, a local min, or a saddle point at $(0,0)$?

3-44. What is the volume of the largest box which can be fit inside of the sphere $x^2+y^2+z^2 = 1$?

3-45. Use the method of Lagrange multipliers to find the smallest distance between the parabola $y = x^2$ and the line $y = x - 1$.

3-46. Let $f(x,y) = (y-x^2)(y-2x^2)$. Show that the origin is a degenerate critical point of f. Prove that f restricted to any line through the origin has a local minimum, but f does not have a local minimum at the origin. *Hint:* Consider the regions where $f > 0$ and $f < 0$.

3-47. Find the minimum of the function $f(x,y,z) = x^2+y^2+z^2$ on the surface $x^2+y^2-z^2 = c$, where $c \in \mathbb{R}$. *Hint:* You'll need to break this into cases, depending on the value of c.

3-48. Let $A : \mathbb{R}^n \to \mathbb{R}^n$ be a linear map. Show that the on the set $S = \{v \in \mathbb{R}^n : \|v\| = 1\}$, the maximum and minimum of A are the largest and smallest eigenvalues of A, respectively.

3-49. In this exercise we give the complete proof to Theorem 3.72. Be sure to read over the given proof for $k = 2$, as we will be imitating it. We will proceed by induction on the number of variables $(y_1, \ldots, y_k)$. The base case is the Implicit Function Theorem, so assume the result holds in $k - 1$ variables.

(a) Suppose B is an invertible matrix, and let B_{ij} denote B with the ith row and jth column deleted. Argue that there must be a $(k-1) \times (k-1)$ submatrix that is invertible. Argue that you can assume B_{kk} is invertible.

(b) Apply the induction hypothesis to write $(y_1, \ldots, y_{k-1}) = \mathbf{f}(\mathbf{x}, y_k)$ for some function $\mathbf{f}$.

(c) Define $G(\mathbf{x}, y_k) = F_k(\mathbf{x}, \mathbf{f}(\mathbf{x}, y_k), y_k)$. Argue that it's sufficient to show that $\partial_{y_k} G(\mathbf{a}, b_k) \neq 0$.

(d) The hard part is actually showing that $\partial_{y_k} G(\mathbf{a}, b_k) \neq 0$.

i. Apply the chain rule to find a formula for $\partial_{y_k} G(\mathbf{x}, y_k)$.

ii. Differentiate the equation $F_i(\mathbf{x}, \mathbf{f}(\mathbf{x}, y_k), y_k) = 0$ with respect to y_k. Solve this equation for $\partial_{y_k} f_i$ by using Cramer's Rule.

iii. Conclude that $\partial_{y_k} G(\mathbf{a}, b_k) \neq 0$.

3-50. Determine whether the equation $\sin(xyz) = z$ may be solved for z as a function of x and y in a neighbourhood of the point $(x, y, z) = (\pi/2, 1, 1)$.

3-51. Find conditions on x and y which guarantee that one can locally solve the following for $u(x, y)$ and $v(x, y)$:

$$\begin{aligned} xu^2 + yv^2 &= 9 \\ xv^2 - yu^2 &= 7. \end{aligned}$$

3-52. Consider the function $\mathbf{f} : \mathbb{R}^3 \to \mathbb{R}^3$ given by

$$\mathbf{f}(x, y, z) = \left(ye^x + \sin(\pi y)\cos(z),\ \cos(yz),\ z^2 \right).$$

Determine whether $\mathbf{f}$ is invertible in a neighbourhood of $(0, 1, \pi/2)$.

3-53. If $U, V \subseteq \mathbb{R}^n$, a map $\mathbf{f} : U \to V$ is said to be a *diffeomorphism* if $\mathbf{f}$ is bijective and both $\mathbf{f}$ and $\mathbf{f}^{-1}$ are C^1.

(a) Show that if $\mathbf{f} : U \to V$ is a C^1 homeomorphism such that $\det D\mathbf{f}(\mathbf{x}) \neq 0$ for all $\mathbf{x} \in U$, then $\mathbf{f}$ is in fact a diffeomorphism.

(b) Show that the requirement that $\det D\mathbf{f}(\mathbf{x}) \neq 0$ is necessary by giving an example of a C^1 homeomorphism which is not a diffeomorphism.

3-54. Show that the following system always has a solution for sufficiently small a,

$$\begin{aligned} x + y + \sin(xy) &= a \\ \sin(x^2 + y) &= 2a \end{aligned}$$

3-55. **Diffeomorphic Invariance of Domain:** Suppose $U \subseteq \mathbb{R}^n$ is open, and $\mathbf{f} : U \to \mathbb{R}^n$ is an injective C^1 map such that $\det D\mathbf{f}(\mathbf{x}) \neq 0$ for all $\mathbf{x} \in U$. We will show that $\mathbf{f}(U)$ is an open subset of $\mathbb{R}^n$.

We need to show that every point $\mathbf{b} \in \mathbf{f}(A)$ contains an open ball which remains in $\mathbf{f}(A)$.

(a) Fix a closed ball $K = \overline{B_r(\mathbf{b})} \subseteq A$ such that $\mathbf{a} = \mathbf{f}^{-1}(\mathbf{b}) \in B_r(\mathbf{b})$. Argue that you can find an open ball $B_\delta(\mathbf{b})$ such that $B_\delta(\mathbf{b}) \cap \mathbf{f}(\partial K) = \emptyset$.

(b) Set $\epsilon = \delta/2$. We want to show that $B_\epsilon(\mathbf{b}) \subseteq \mathbf{f}(A)$. Fix a $\mathbf{c} \in B_\epsilon(\mathbf{b})$ and define the function $\phi_{\mathbf{c}} : A \to \mathbb{R}$ by $\phi(\mathbf{x}) = \|\mathbf{f}(\mathbf{x}) - \mathbf{c}\|^2$. Argue that ϕ achieves a minimum $\mathbf{x}_0$ on K, and that $\mathbf{f}^{-1}(\mathbf{c}) = \mathbf{x}_0$. This will complete the proof.

Note: This result is actually true if $\mathbf{f} : A \to \mathbb{R}^n$ is simply a continuous injective function, though the proof is much harder.

3-56. **Diffeomorphic Invariance of Boundary** Suppose $U, V \subseteq \mathbb{H}^k$ such that $U \cap \partial\mathbb{H}^k \neq \emptyset$ and $V \cap \partial\mathbb{H}^k \neq \emptyset$. Let $\mathbf{G} : U \to V$ be a diffeomorphism in the sense of Exercise 3-33, so that $\mathbf{G}$ extends to a diffeomorphism in a neighbourhood of every point in $\partial\mathbb{H}^k$. Show that $\mathbf{G}(U \cap \partial\mathbb{H}^k) = V \cap \partial\mathbb{H}^k$. *Hint:* There are a couple of ways to do this. One is to employ Execise 3-55, the other is to employ Execise 2-54.

3-57. Let $f : \mathbb{R} \to \mathbb{R}$ be a non-constant C^1 function such that $f'(0) \neq 0$ and $f(x+y) = f(x)f(y)$. Define $F : \mathbb{R}^2 \to \mathbb{R}$ by $F(x,y) = f(x)f(y)$. Determine what conditions (if any) must be imposed upon y to ensure that y can be solved as a function of x on the set $\{(x,y) : F(x,y) = 1\}$. *Bonus:* Write down an explicit formula for y as a function of x.

3-58. Define the set

$$M_2(\mathbb{R}) = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} : a, b, c, d \in \mathbb{R} \right\}$$

to be the set of 2×2 matrices. Define a map $g : M_2(\mathbb{R}) \to M_2(\mathbb{R})$ by $g(A) = A^2$. Determine whether g is invertible in a neighbourhood of $I = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$.

3-59. In this problem, we will show that for a special class of polynomials, slightly perturbing the coefficients will preserve the number of roots of the equation.

(a) Let $f : \mathbb{R} \to \mathbb{R}$ be a degree n *polynomial.* Show that if all the roots of f are distinct, then for any root r we necessarily have $f'(r) \neq 0$. *Hint:* We say that a root r has multiplicity k if $f(x) = (x-r)^k q(x)$ and $q(r) \neq 0$. All the roots of a polynomial are distinct if every root has multiplicity 1.

(b) For fixed $(c_0, \ldots, c_{n-1})$ let $f(x) = x^n + c_{n-1}x^{n-1} + \cdots + c_1 x + c_0$ be a function with distinct roots. Identify the $(c_0, \ldots, c_{n-1})$ with a point in $\mathbb{R}^n$. Show that for each root r, there is a neighbourhood U_r of $(c_0, c_1, \ldots, c_{n-1}) \in \mathbb{R}^n$ and a neighborhood V_r of r such that if $(d_0, \ldots, d_{n-1}) \in U_r$ then $x^n + d_{n-1}x^{n-1} + \cdots + d_1 x + d_0$ has a root in V_r.

(c) Use part *b* to conclude that a degree n polynomial with exactly $m < n$ roots, all of which are distinct, has the same number of roots under small perturbation of its coefficients.

(d) Does this result still hold if the roots are no longer distinct? Prove the result or give a counter-example.

3-60. A map $\mathbf{f} : \mathbb{R}^n \to \mathbb{R}^n$ is said to be open if whenever U is open then $\mathbf{f}(U)$ is open. Show that if $\mathbf{f}$ is C^1 and $D\mathbf{f}(\mathbf{x}_0)$ is invertible for all $\mathbf{x}_0 \in \mathbb{R}^n$ then $\mathbf{f}$ is an open map.

3-61. We say that a function $f : \mathbb{R}^n \to \mathbb{R}$ is strictly convex if for every $t \in [0,1]$ and $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$

$$f(t\mathbf{x} + (1-t)\mathbf{y}) < tf(\mathbf{x}) + (1-t)f(\mathbf{y}).$$

Let $F : \mathbb{R}^2 \to \mathbb{R}$ be a C^2 function which is strictly convex, non-negative, and satisfies $F(0,0) = 0$. Show that there is an *everywhere concave-down* function $f : \mathbb{R} \to \mathbb{R}$, such that in a neighbourhood of $(0,1)$, $F(x, f(x)) = F(0,1)$.

3-62. Consider the function

$$f : \mathbb{R} \to \mathbb{R}, \quad f(x) = \begin{cases} e^{-1/x} & x > 0 \\ 0 & \text{otherwise} \end{cases}.$$

Show that f is a C^∞ function as follows: Define

$$f_n(x) = \begin{cases} e^{-1/x}/x^n & x > 0 \\ 0 & \text{otherwise.} \end{cases}$$

(a) Show that f_n is continuous at 0.

(b) Show that f_n is differentiable at 0

(c) Show that $f_n'(x) = f_{n+2}(x) - nf_{n+1}(x)$

(d) Conclude that f_n is C^∞.

3-63. Let f be the function from Exercise 3-62, and define $\tilde{f}(x) = f(1-x)f(x)$.

(a) Show that $\tilde{f}$ is positive on $(0,1)$ and zero everywhere else.

(b) Let $\epsilon > 0$ be given. Use $\tilde{f}$ to construct a function $g_\epsilon : \mathbb{R} \to \mathbb{R}$ such that $g_\epsilon(x) = 0$ if $x < 0$ and $g_\epsilon(x) = 1$ if $x \geq \epsilon$.

(c) Let $R = [a_1, b_1] \times \cdots [a_n, b_n] \subseteq \mathbb{R}^n$. Show that

$$\phi(x) = \tilde{f}\left(\frac{x_1 - a_1}{b_1 - a_1}\right) \cdots \tilde{f}\left(\frac{x_n - a_n}{b_n - a_n}\right)$$

is a non-negative C^∞ function which is positive on R^{int} and vanishes for all $\mathbf{x} \notin R$.

(d) Suppose $U \subseteq \mathbb{R}^n$ and $K \subseteq U$ is compact. Show there is a non-negative C^∞ function $f : U \to \mathbb{R}$ such that $f(x) > 0$ on K and $f(x) = 0$ outside of of V. *Hint:* Exercise 2-55.

(e) Show that in part (d), we could take $f(x) = 1$ on K. *Hint:* Compose your function from part (d) with your function from part (b), for some appropriate choice of ϵ.

Given an open set $U \subseteq \mathbb{R}^n$ and a compact subsets $K \subseteq U$, $f : U \to \mathbb{R}$ is a *bump function* if $f(K) = 1$ and $f(\mathbf{x}) = 0$ for all $\mathbf{x}$ outside of U. You've just proven the existence of bump functions.

3-64. Suppose $A \subseteq \mathbb{R}^n$ is an open set with open cover $\mathcal{O}$. Let $(\phi_i)_{i=1}^n$ be a C^1, compactly supported partition of unity subordinate to $\mathcal{O}$. Fix a compact subset $K \subseteq A$. Show there exists an $M \in \mathbb{N}$ such that $\phi_i(\mathbf{x}) = 0$ for all $i > M$ and $\mathbf{x} \in K$.

3-65. Suppose $K \subseteq \mathbb{R}^n$ is a set, and $\{U_i : i \in I\}$ is a (relatively) open covering of K. Show that there is a compactly supported C^1 partition of unity subordinate to the U_i.

4 Integration of Scalar Valued Functions

We will begin with a review of integration on the line before moving onto the general theory for integrating variables in several dimensions.

4.1 Jordan Measure

Integration is concerned with measuring area under a graph, which we've done by using rectangular approximations and taking a limit. However, we're getting ahead of ourselves: We don't even know how to measure lengths let alone areas. In case it's not obvious that our grasp on length is tenuous, think about the set $\{1/n : n \in \mathbb{N}\}$. What is the length of this set, and does it even make sense to say that it has a length?

There is an entire field of mathematics known as *measure theory* which is concerned with assigning numbers to sets in a way that describes their size. The idea of measuring a set is essential in many fields of mathematics, physics, and statistics, so is a well-studied and deep topic. Here I'll introduce the Jordan measure. This is hardly the "correct" measure with which to do integration in $\mathbb{R}$; however, it is the simplest measure which does not require its own book to introduce.

Let me make a few comments here. The first is that if you don't like epsilon arguments and inequalities, this section is going to be your worst nightmare. Almost every proof in measure theory is technical. Secondly, we don't need a lot of this theory, so I'm going to introduce only what I feel is absolutely necessary to our study, and ignore the rest.

We start with a fundamental shape whose length we should already know, in this case a rectangular prism.

Definition 4.1

Let $R = (a_1, b_1) \times (a_2, b_2) \times \cdots (a_n, b_n)$ be a rectangle in $\mathbb{R}^n$. The *volume* of R is $|R| = \prod_{i=1}^n (b_i - a_i)$. As a convention, we will say that $|\emptyset| = 0$. Define a *poly-rectangle* as any set which is a finite union of open rectangles. If $P = R_1 \cup R_2 \cup \cdots \cup R_k$ is a poly-rectangle, its volume is $|P| = |R_1| + |R_2| + \cdots + |R_k|$.

The definition of a poly-rectangle does not require the constituent rectangles to be disjoint. For example, if $I = (5, 15)$ then $|I| = 15 - 5 = 10$. On the other hand, if $I = (-1, 1) \cup (-2, 2) \cup (-3, 3)$ then $|I| = 2 + 4 + 6 = 12$. We use our poly-rectangles as building blocks to both over-approximate and under-approximate the size of our sets, which leads to the notion of Jordan measure:

Definition 4.2

Let $S \subseteq \mathbb{R}^n$ be any set. We define the *Jordan outer measure* of S as

$$\mu^*(S) = \inf\left\{|P| : P \text{ is a poly-rectangle such that } S \subseteq P\right\}$$

and the *Jordan inner measure* of S as

$$\mu_*(S) = \sup\left\{|P| : P \text{ is a poly-rectangle such that } P \subseteq S\right\}.$$

The set S is said to be *Jordan measurable* if $\mu^*(S) = \mu_*(S)$, in which case we will just write $\mu(S)$. By convention, $\mu(\emptyset) = 0$.

The length of a poly-rectangle is non-negative. Consequently, $0 \leq \mu_*(S) \leq \mu^*(S)$, a fact which we'll take for granted (though it really should be proved). Moreover, one can quickly see that if P is a poly rectangle, it is measurable and $\mu(P) = |P|$.

We could just as easily use closed rectangles, by showing that we can bound a closed rectangle by open rectangles to within a volume of ϵ. More concretely, if $C = [a_1, b_1] \times \cdots \times [a_n, b_n]$ is an open rectangle, let $L_i = b_i - a_i$ and $L = \max\{L_1, \ldots, L_n\}$. Set $\delta = \min\{\epsilon/N, 1\}$ where

$$N = \sum_{k=1}^{n} \binom{n}{k} L^k.$$

Take $R = (a_1 - \delta/2, b_1 + \delta/2) \times \cdots \times (a_n - \delta/2, b_n + \delta/2)$, so that

$$|R| = \prod_{i=1}^{n}\left[\left(b_i + \frac{\delta}{2}\right) - \left(a_i - \frac{\delta}{2}\right)\right] = \prod_{k=1}^{n}(L_i + \delta) \overset{(*)}{\leq} \left[\prod_{i=1}^{n} L_i\right] + \delta \sum_{k=1}^{n}\binom{n}{k} L^k = \left[\prod_{i=1}^{n} L_i\right] + \epsilon.$$

The inequality $(*)$ takes some thinking about. In the product $\prod(L_i + \delta)$ we just keep the highest order term $\prod L_i$, replace all other instances of L_i with L and δ^k with δ. The argument shows we can bound C by an open rectangle contained inside of C by replacing $(a_i - \delta/2, b_i + \delta/2)$ with the $(a_i + \delta/2, b_i - \delta/2)$, with the additional caveat that we must ensure that the latter interval is non-empty. In fact, this shows that if R is any rectangle, $\overline{R}$ is measurable and $\mu(R) = \mu(\overline{R})$.

Proposition 4.3: Translation Invariance of the Jordan Measure

If $S \subseteq \mathbb{R}^n$ is Jordan measurable and $\mathbf{a} \in \mathbb{R}^n$, then $S + \mathbf{a} = \{\mathbf{x} + \mathbf{a} : \mathbf{x} \in S\}$ is Jordan measurable and $\mu(S + \mathbf{a}) = \mu(S)$.

This is straightforward and left to Exercise 4-6. Effectively, if R is a rectangle, $R + \mathbf{a}$ is rectangle and they have the same measure. Once you prove this, the rest of the proof comes together quickly.

Proposition 4.4

If $S = \{\mathbf{p}\} \subseteq \mathbb{R}^n$ is the singleton set consisting of a single element, then S is measurable and $\mu(S) = 0$.

Proof. Without loss of generality and by translation invariance of the Jordan measure, we may assume $\mathbf{p} = \mathbf{0}$. There are no poly-rectangles which live inside of S so $\mu_*(S) = \sup\{|\emptyset|\} = 0$. Fix some $\epsilon > 0$. For the outer measure, the poly-rectangle $R_\epsilon = (-\sqrt[n]{\epsilon}/2, \sqrt[n]{\epsilon}/2)^n$ covers S, and $|R| = \epsilon$. Hence $0 < |R| < \epsilon$, showing that $\mu^*(S) = 0$. We've thus shown $\mu^*(S) = \mu_*(S) = 0$, so S is Jordan measurable and $\mu(S) = 0$. □

Jordan measures are difficult to determine in general, but for our purposes we're most interested in sets with zero Jordan measure. This situation is so special that we assign it a special name.

Definition 4.5

If S is a measurable set such that $\mu(S) = 0$, we say that S has *Jordan measure zero*, or *zero content*.

Conveniently, one need only check the outer measure to determine whether a set has zero content.

Proposition 4.6

If $S \subseteq \mathbb{R}$ satisfies $\mu^*(S) = 0$, then S is Jordan measurable and $\mu(S) = 0$.

Proof. The inner and outer measures always satisfy the relation

$$0 \leq \mu_*(S) \leq \mu^*(S),$$

so if $\mu^*(S) = 0$ then $\mu_*(S) = \mu^*(S) = 0$. □

Since $\mu^*(S)$ is the infimum over poly-rectangles which cover S, to show that $\mu^*(S) = 0$ it suffices to show that for every $\epsilon > 0$, there exists a poly-rectangle P such that $S \subseteq P$ and $\mu(P) < \epsilon$. The following also aids us, but I've left the proof to Exercise 4-12.

Proposition 4.7

If $S \subseteq \mathbb{R}^n$ is measurable and $\mu(S) = 0$, then for any $T \subseteq S$, T is measurable and $\mu(T) = 0$ as well.

Proposition 4.4 interestingly shows that non-empty sets can have zero-content. The Jordan measure of a set can be thought of as a way of measuring a set's "width," and as a single point has no width, it has Jordan measure zero. By adapting our proof, we can show more generally that any finite set has zero-content.

Proposition 4.8

Any finite set $S \subseteq \mathbb{R}$ has zero content.

Proof. We can prove this directly or by induction.

Induction: We will induct over the number of elements in S. The base case occurs when S has a single point, which we covered in Proposition 4.4.

Now assume that any set with $n-1$ elements has zero content and let $S = \{\mathbf{p}_1, \ldots, \mathbf{p}_n\}$. Pick an arbitrary $\epsilon > 0$ and define $S' = \{\mathbf{p}_1, \ldots, \mathbf{p}_{n-1}\} \subseteq S$, which contains $n-1$ points. By hypothesis $\mu(S') = 0$, so there is a poly-rectangle P'_ϵ such that $S' \subseteq P'_\epsilon$ and $\mu(P'_\epsilon) < \epsilon/2$. Let $P_\epsilon = \mathbf{p} + R_{\epsilon/2}$, where $R_{\epsilon/2}$ is the rectangle of volume $\epsilon/2$ from Proposition 4.4, so that $P'_\epsilon \cup P_\epsilon$ is a poly-rectangle, $S \subseteq P'_\epsilon \cup P_\epsilon$, and

$$0 \leq \mu(P'_\epsilon \cup P_\epsilon) \leq \mu(P'_\epsilon) + \mu(P_\epsilon) < \epsilon.$$

Hence $\mu(S) = 0$, completing the induction. □

We'd like to avoid using poly-rectangles and the definition of Jordan measure every time we want to say something about the measure of a set. If we will allow ourselves the intuition of thinking of Jordan measure as width, then the following properties should not seem unreasonable.

Proposition 4.9

Let $S_1, S_2 \subseteq \mathbb{R}^n$ be Jordan measurable.

1. If $S_1 \subseteq S_2$ then $\mu(S_1) \leq \mu(S_2)$.
2. $\mu(S_1 \cup S_2) \leq \mu(S_1) + \mu(S_2)$.
3. If $T : \mathbb{R}^n \to \mathbb{R}^n$ is a linear transformation, then $\mu(T(S)) = |\det T|\mu(S)$.
4. Jordan measure is closed under unions and intersections; that is, $S_1 \cup S_2$ and $S_1 \cap S_2$ are Jordan measurable.

Proof. I'll prove (2) and leave the remainder to the exercises. Since each S_i is measurable, there exists a poly-rectangle I_i such that $S_i \subseteq I_i$ and $\mu(S_i) \leq |I_i| \leq \mu(S_i) + \epsilon/2$. The union $I = I_1 \cup I_2$ is also a poly-rectangle, and covers $S_1 \cup S_2$, so $\mu(S_1 \cup S_2) \leq |I_1 \cup I_2| = |I_1| + |I_2|$. Combining this with our previous discussion, we get

$$\mu(S_1 \cup S_2) \leq |I_1| + |I_2| \leq \mu(S_1) + \mu(S_2) + \epsilon.$$

Since ϵ is arbitrary, we conclude that $\mu(S_1 \cup S_2) \leq \mu(S_1) + \mu(S_2)$ as required. □

Like integrable functions, most sets you think of are Jordan measurable. Are there any non-measurable sets?

Example 4.10

Show that the set $S = \mathbb{Q} \cap [0, 1]$ is not Jordan measurable in $\mathbb{R}$.

Proof. Because the rationals are dense in $[0, 1]$, any interval which contains every rational in $[0, 1]$ must also contain $[0, 1]$ itself. As such, $\mu^*(S) \geq \mu^*([0, 1]) = 1$. On the other hand, S contains no intervals so $\mu_*(S) = \sup |\emptyset| = 0$. Since it's impossible for $\mu^*(S)$ to equal $\mu_*(S)$, we conclude that S is not measurable. □

A nice alternative characterization of Jordan measurability is attained simply by looking at the measure of the boundary.

Theorem 4.11

A set $S \subseteq \mathbb{R}^n$ is Jordan measurable if and only if $\mu(\partial S) = 0$.

We'll use this theorem but not prove it. I've outline the proof in Exercise 4-10.

Theorem 4.12

Suppose $n < m$. If $K \subseteq \mathbb{R}^n$ is compact and $\mathbf{f} : K \to \mathbb{R}^m$ is C^1, then $\mathbf{f}(K)$ has zero content.

Proof. The idea is that covering K with cubes with side length $1/k$ will require on the order of k^n such cubes. However, since K is compact and f is C^1, we can limit how much these cubes can grow in size, thus covering the image $f(K)$ with cubes whose side length is proportional to $1/k$ as well. As the image will require k^m such cubes, the difference between them will be $1/k^{m-n}$, which will go to zero as $k \to \infty$.

We begin by reducing the complexity of the problem statement. First of all, since K is compact it is bounded, and hence we can find a cube C containing K, of side length say L. It suffices to show the result on C, for if $\mu(\mathbf{f}(C)) = 0$ then $\mathbf{f}(K) \subseteq \mathbf{f}(C)$ and Proposition 4.7 implies $\mu(\mathbf{f}(K)) = 0$.

Since C is compact and $\mathbf{f}$ is C^1, $\|D\mathbf{f}(\mathbf{x})\|$ achieves its maximum on C, which we'll call M. Divide C into $L^n k^n$ subcubes of side length $1/k$, and fix an arbitrary subcube $\tilde{C}$. Since $\tilde{C}$ is convex, by the Mean Value Inequality we know that for any $\mathbf{x}, \mathbf{y} \in \tilde{C}$ and $i \in \{1, \ldots, n\}$

$$|f_i(\mathbf{x}) - f_i(\mathbf{y})| \leq \|\mathbf{f}(\mathbf{x}) - \mathbf{f}(\mathbf{y})\| \leq M\|\mathbf{x} - \mathbf{y}\| \leq \frac{M\sqrt{n}}{k}.$$

In the last inequality I've used Exercise 4-14, which shows that the diameter of the cube of side length L in $\mathbb{R}^n$ is $L\sqrt{n}$. Hence $\mathbf{f}(\tilde{C})$ is contained in a cube whose side length is $M\sqrt{n}/k$, which has a volume $M^m n^{m/2}/k^m$. Doing this for every subcube of C shows that we can cover $\mathbf{f}(c)$ with $L^n k^n$ cubes of volume $M^m n^{m/2} k^m$, giving us a poly-rectangle R of size

$$|R| = (\text{number of cubes}) \times (\text{measure of cubes}) = (L^n k^n) \times \left(\frac{M^m n^{m/2}}{k^m}\right) = \frac{L^n M^m n^{m/2}}{k^{m-n}}.$$

Everything in the numerator is a fixed constant, so by taking k sufficiently large, we can make this as small as we like. Thus $\mu(\mathbf{f}(C)) = 0$ as required. □

If you're having trouble seeing why the proof of Theorem 4.12 works, think about the case when $f : [a, b] \to \mathbb{R}^2$, visualized in Figure 4.1. The quantity M represents the maximum speed we can travel in any direction, meaning that on a subinterval $[t_i, t_{i+1}]$ we have travelled at most $M|t_{i+1} - t_i|$ in the x-direction, and similar for the y-direction. Thus $f([t_i, t_{i+1}])$ fits into a small square. The areas of these squares tend to zero quadratically, while the number of required squares tends to zero linearly, and the combination of the two means we can make the squares have a total area as small as desired.

Proposition 4.13

If $K \subseteq \mathbb{R}^n$ is bounded and $f : K \to \mathbb{R}$ is a continuous function, then the graph $\Gamma(f) = \{(\mathbf{x}, f(\mathbf{x})) : \mathbf{x} \in K\} \subseteq \mathbb{R}^{n+1}$ has zero measure in $\mathbb{R}^{n+1}$.

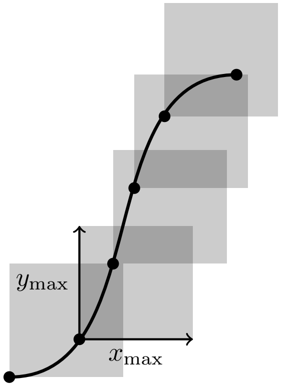


Figure 4.1: By looking at the maximum speed that the function attains, one can find the worst case box the each subinterval (marked by black dots) fits into. As we increase the number of subintervals, the number of necessary boxes increases linearly, while the area of each box decreases proportional to the nth power.

Proposition 4.13 is a similarly useful, but somewhat different result, which you will prove in Exercise 4-13. The regularity of f has been weakened to only continuity, but this has been tempered by requiring the f be real-valued. Of note is that Theorem 4.12 does *not* hold if the C^1 condition is relaxed. The Peano Space filling curve is an example of a continuous *surjective* map $p : [0,1] \to [0,1] \times [0,1]$, so its image has non-zero measure.

4.2 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 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.

It turns out that Riemann integration – or integrating by partitions of the domain – is an inferior way of doing things: There are many functions which are not integrable. A more prudent choice is to break up the range of the function and integrate that way, in a manner known as *Lebesgue integration*. Any functions which is Riemann integrable is Lebesgue integrable, and in that case the values of the integral agree. However, the collection of Lebesgue integrable functions is strictly larger than the Riemann integrable functions. Unfortunately, Lebesgue integration is beyond the scope of the course.

4.2.1 Riemann Sums

Definition 4.14

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 $P = \{0 < 1/2 < 2/3 < 1\}$ 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. The idea is that when integrating, we are going to want to take partitions whose length between endpoints gets smaller, corresponding to letting the width of our approximating triangles get smaller. The number $\ell(P)$ then describes the widest width, corresponding to the "worst" rectangle.

Definition 4.15

If P and Q are two partitions of $[a, b]$, then Q is a *refinement* of P if $P \subseteq Q$.

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.

Definition 4.16

A partially ordered set $(S, \leq)$ is said to be a directed set if for every $a, b \in S$ there exists a $c \in S$ such that $a \leq c$ and $b \leq c$.

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 $Q \subseteq R$. Hence refinements define a partial order on $\mathcal{P}_{[a,b]}$ and turn it into a directed set.

Definition 4.17

Given a function $f : [a, b] \to \mathbb{R}$, a *(tagged) 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 you have likely seen amount to particular choices of the t_i. For example, the left- and right- endpoint Riemann sums are

$$\ell(f,P) = \sum_{i=1}^n f(x_{i-1})(x_i - x_{i-1}) \quad \text{and} \quad r(f,P) = \sum_{i=1}^n f(x_i)(x_i - x_{i-1}).$$

Definition 4.18

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 $I = \int f$ and is called the *Riemann integral of f*.

Note that the definition of a Riemann integrable function must hold for any choice of tagging. Roughly speaking, a function is Riemann integrable with integral I if we can approximate I arbitrarily well by taking a sufficiently fine partition P.

Theorem 4.19

If $f,g : [a,b] \to \mathbb{R}$ are Riemann integrable and $c \in \mathbb{R}$, then

1. If f,g are integral on $[a,b]$ then $f+g$ is integrable on $[a,b]$ and
$$\int [f+g] = \int f + \int g$$
2. If f is integrable on $[a,b]$ and $c \in \mathbb{R}$, then cf is integrable on $[a,b]$ and
$$\int cf = c \int f.$$

Proof. The proof is similar to the Limit Laws, but adapted for integrals. Since f,g are integrable, set

$$I = \int f \quad \text{and} \quad J = \int g.$$

1. Let $\epsilon > 0$ be given. Choose $\delta_f > 0$ and $\delta_g > 0$ such that
$$\begin{aligned} |S(f,P) - I| &< \frac{\epsilon}{2} \text{ whenever } P \in \mathcal{P}_{[a,b]} \text{ and } \ell(P) < \delta_f \\ |S(g,P) - I| &< \frac{\epsilon}{2} \text{ whenever } P \in \mathcal{P}_{[a,b]} \text{ and } \ell(P) < \delta_g. \end{aligned}$$
Set $\delta = \min\{\delta_f, \delta_g\}$ and take R to be any partition satisfying $\ell(R) < \delta$. We know that $S(f+g,R) = S(f,R) + S(g,R)$, and consequently
$$\begin{aligned} |S(f+g,R) - (I+J)| &= |S(f,R) + S(g,R) - (I+J)| \leq |S(f,R) - I| + |S(g,R) - J| \\ &< \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon. \end{aligned}$$

2. Let $c \in \mathbb{R}$ and $\epsilon > 0$ be arbitrary. Choose $\delta > 0$ such that $|S(f,P) - I| < \epsilon/|c|$ whenever $\ell(P) < \delta$. We know that $S(cf,P) = cS(f,P)$, so
$$|S(cf,P) - cI| = |c||S(f,P) - I| < |c|\frac{\epsilon}{|c|} = \epsilon$$
as required. □

Note that writing $\int f$ is totally unambiguous – the domain of the function is built into the definition, and the name of the variable is of no consequence. This being said, there are times when it's convenient to use the notation
$$\int_{[a,b]} f$$
so that the interval of integration is explicitly mentioned. If $f : [a,b] \to \mathbb{R}$ is integrable and $[c,d] \subseteq [a,b]$, we may want to discuss the integral of $f|_{[c,d]} : [c,d] \to \mathbb{R}, x \mapsto f(x)$. This notation is then convenient for writing
$$\int_{[c,d]} f = \int f|_{[c,d]}.$$
An important note for later is that the above integral is *not* oriented; namely, the integral is independent of whether we're moving from a to b or from b to a. You would've learned previously that there is also a notion of the oriented integral
$$\int_a^b f = -\int_b^a f.$$
Finally, when we want to be explicit about the variable with which we're integrating, we'll write
$$\int_{[a,b]} f(x)\,\mathrm{d}x \quad \text{or} \quad \int_a^b f(x)\,\mathrm{d}x.$$

A major obstacle of this definition is it requires us to know the value of the integral $\int f$ ahead of time. This can be avoided by appealing to a Cauchy-sequence notion for Riemann Sums (see Theorem 4.21). If we're willing to move away from Riemann sums, there's an alternative definition which is often convenient.

Definition 4.20

Given a function $f : [a,b] \to \mathbb{R}$ and a partition $P = \{x_0, x_1, \ldots, x_n\}$, the *lower- and upper-Darboux sums of f with respect to P* are
$$L(f,P) = \sum_{i=1}^N \left[\inf_{x \in [x_{i-1},x_i]} f(x)\right](x_i - x_{i-1}) \quad \text{and} \quad U(f,P) = \sum_{i=1}^N \left[\sup_{x \in [x_{i-1},x_i]} f(x)\right](x_i - x_{i-1}).$$

If f is a continuous function, then the upper and lower Darboux sums are actually Riemann sums. However, because the infimum and supremum need not be achieved, this is not guaranteed to be true if f is not continuous.

Theorem 4.21: Equivalent Definitions of Integrability

If $f : [a,b] \to \mathbb{R}$ is a function, then the following are equivalent:

1. f is Riemann integrable,
2. $\displaystyle\sup_{P \in \mathcal{P}_{[a,b]}} L(f,P) = \inf_{P \in \mathcal{P}_{[a,b]}} U(f,P)$,
3. For every $\epsilon > 0$ there exists a partition $P \in \mathcal{P}_{[a,b]}$ such that $U(f,P) - L(f,P) < \epsilon$,
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 $|S(f,P) - S(f,Q)| < \epsilon$.

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. The equivalence of these definitions is left to Exercise 4-17.

The geometric interpretation of condition (3) – that the difference between the upper and lower Darboux sums can be made arbitrarily small, is visualized in Figure 4.2. You'll show later that the upper Darboux sum is larger than any Riemann sum on the same partition, and similarly the lower Darboux sum is smaller than any Riemann sum. Their difference is thus the worst-case difference in approximating the area under the curve. If this error tends to zero, the area can be found perfectly.

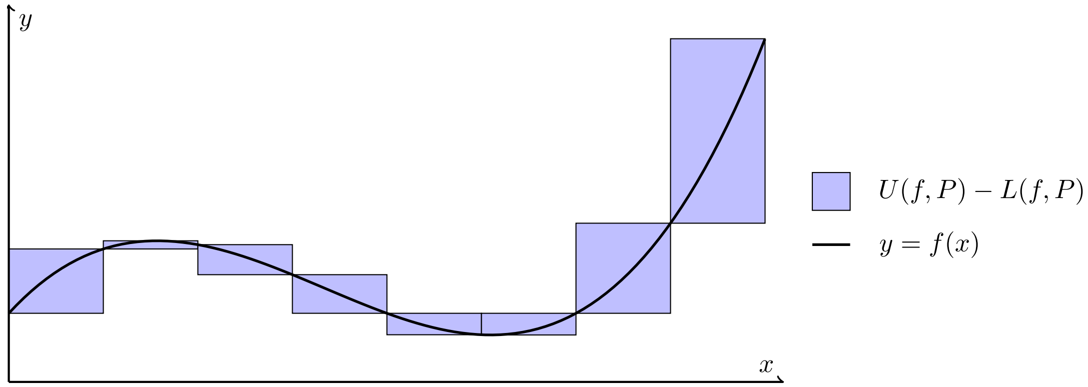


Figure 4.2: The difference between the upper and lower Darboux sums is the area shaded in blue. It represents the worst case difference between over and under approximating the area under the curve.

Example 4.22

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 $c(b-a)^2/n < \epsilon$ (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

$$\begin{aligned} L(f,P) &= \sum_{k=0}^{n-1} f(x_k)(x_{k+1}-x_k) = \frac{c(b-a)}{n}\sum_{k=0}^{n-1} x_k \\ U(f,P) &= \sum_{k=0}^{n-1} f(x_{k+1})(x_{k+1}-x_k) = \frac{c(b-a)}{n}\sum_{k=0}^{n-1} x_{k+1}. \end{aligned}$$

Hence their difference yields

$$\begin{aligned} U(f,P) - L(f,P) &= \frac{c(b-a)}{n}\sum_{k=0}^{n-1}(x_{k+1}-x_k) \\ &= \frac{c(b-a)}{n}(x_n - x_0) = \frac{c(b-a)}{n}(b-a) \\ &= \frac{c}{n} < \epsilon. \end{aligned}$$

which is what we wanted to show. ■

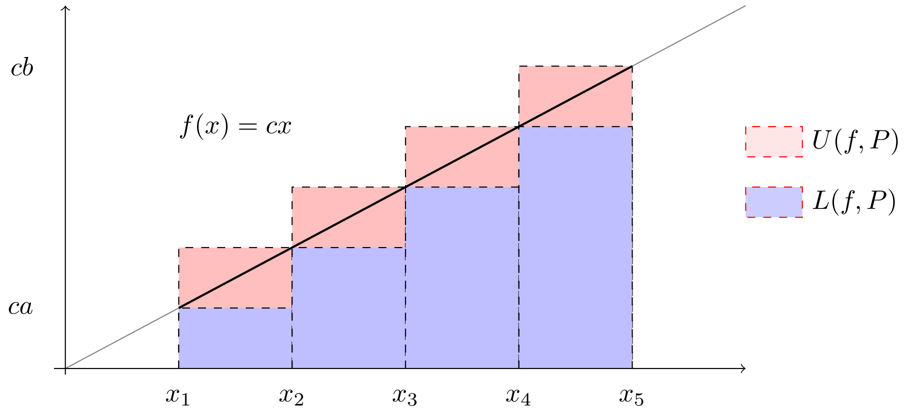


Figure 4.3: One can visually see why the difference between $U(f,P)$ and $L(f,P)$ results in a telescoping sum. For example, the red rectangle on $[x_1,x_2]$ is the same area as the blue rectangle on $[x_2,x_3]$, so they cancel in the difference.

Example 4.23

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

$$\begin{aligned} U(f, P) &= \sum_{i=1}^{n} M_i(x_i - x_{i-1}) = \sum_{i=1}^{n} (x_i - x_{i-1}) = x_1 - x_0 = 1 \\ L(f, P) &= \sum_{i=1}^{n} m_i(x_i - x_{i-1}) = 0 \end{aligned}$$

so that $U(f, P) - L(f, P) = 1$. Since this holds for arbitrary partitions, any $\epsilon < 1$ will fail the definition of integrability, so $\chi_{\mathbb{Q}}$ is not integrable. ■

4.2.2 Properties of the Integral

The following properties will be left as exercises.

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. **Inherited Integrability:** If f is integrable on $[a, b]$ then f is integrable on any subinterval $[c, d] \subseteq [a, b]$.
3. **Monotonicity of Integral:** If f, g are integrable on $[a, b]$ and $f(x) \le g(x)$ for all $x \in [a, b]$ then
$$\int f \le \int g.$$
4. **Subnormality:** If f is integrable on $[a, b]$ then $|f|$ is integrable on $[a, b]$ and satisfies
$$\left|\int f\right| \le \int |f|.$$

These proofs have been left to Exercises 4-18, 4-19, 4-20, 4-21.

Theorem 4.24: The Fundamental Theorem of Calculus

1. If f is integrable on $[a, b]$ and $x \in [a, b]$ define $F(x) = \int_a^x f(t)\,\mathrm{d}t$. 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.

4.2.3 Sufficient Conditions for Integrability

Theorem 4.21 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 4.23 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 4.25

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 4.22 (see Figure 4.3).

More formally, assume without loss of generality that f is increasing on $[a, b]$; otherwise, replace f with $-f$ use linearity of the integral. 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:

$$L(f, P) = \sum_{i=1}^{n} f(x_{i-1})(x_i - x_{i-1}), \qquad U(f, P) = \sum_{i=1}^{n} f(x_i)(x_i - x_{i-1}).$$

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

$$\begin{aligned} U(f, P) - L(f, P) &= \sum_{i=1}^{n} \left[f(x_i) - f(x_{i-1})\right](x_i - x_{i-1}) \leq \delta \sum_{i=1}^{n} \left[f(x_i) - f(x_{i-1})\right] \\ &\leq \delta(f(b) - f(a)) \leq \frac{\epsilon}{f(b) - f(a)}(f(b) - f(a)) < \epsilon. \end{aligned}$$

Since ϵ was arbitrary, Theorem 4.21 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 4.26

Every continuous function on $[a, b]$ is integrable.

It is tempting to use Theorem 4.25, since f is certainly bounded and we should be able to restrict f to intervals on which it is monotone. Applying Additivity of Domain 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

$$\begin{aligned} U(f, P) - L(f, P) &= \sum_{i=1}^{n} (M_i - m_i)(x_i - x_{i-1}) \leq \sum_{i=1}^{n} \left[\frac{\epsilon}{b-a}\right](x_i - x_{i-1}) \\ &\leq \frac{\epsilon}{b-a} \sum_{i=1}^{n} (x_i - x_{i-1}) = \frac{\epsilon}{b-a}(b-a) = \epsilon. \end{aligned}$$

Applying Theorem 4.21 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?

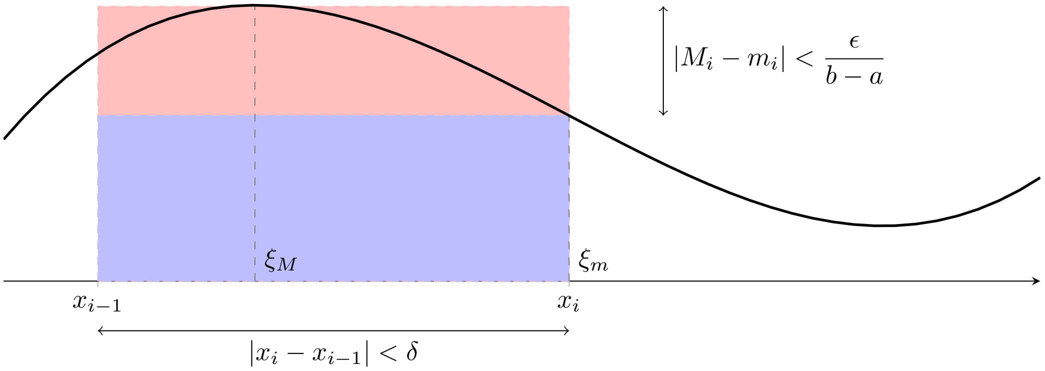

$|M_i - m_i| < \frac{\epsilon}{b-a}$
ξ_M
ξ_m
x_{i-1}
x_i
$|x_i - x_{i-1}| < \delta$

Figure 4.4: 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.

Example 4.27

Let $f(x) = x$ on $[0,2]$ and define

$$g(x) = \begin{cases} f(x) & x \neq 1 \\ 10^6 & x = 1 \end{cases}.$$

Show that g is integrable and $\int f = \int g$.

Solution. 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. By Additivity of Domain we have

$$\int_{[0,2]} g = \int_{V_\epsilon} g + \int_{U_\epsilon} g = \int_{V_\epsilon} f + \int_{U_\epsilon} g.$$

We want to show that in the limit $\epsilon \to 0$ we get $\int_{U_\epsilon} g \to 0$, so that $\int f = \int g$. 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 $0 \leq \int_{U_\epsilon} g \leq 2\epsilon 10^6$. By the Squeeze Theorem, it then follows that

$$\int_{U_\epsilon} g \xrightarrow{\epsilon \to 0} 0. \qquad \blacksquare$$

Theorem 4.28

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 $\sum_j \ell(I_j) < \epsilon/(2(M-m))$. 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 $U(f|_V, P) - L(f|_V, P) < \epsilon/2$. 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

$$U(f,P) = U(f|_W, P) + U(f|_V, P), \qquad L(f,P) = L(f|_W, P) + L(f|_V, P).$$

Since we already know how to bound the V contribution, we need now only look at the W contribution. Notice on W we have

$$U(f|_W, P) - L(f|_W, P) < \sum_{j=1}^k (M-m)\ell(I_j) \leq (M-m)\frac{\epsilon}{2(M-m)} = \frac{\epsilon}{2},$$

thus

$$U(f,P) - L(f,P) = [U(f|_W,P) - L(f|_W,P)] + [U(f|_V,P) - L(f|_V,P)] \leq \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon. \qquad \square$$

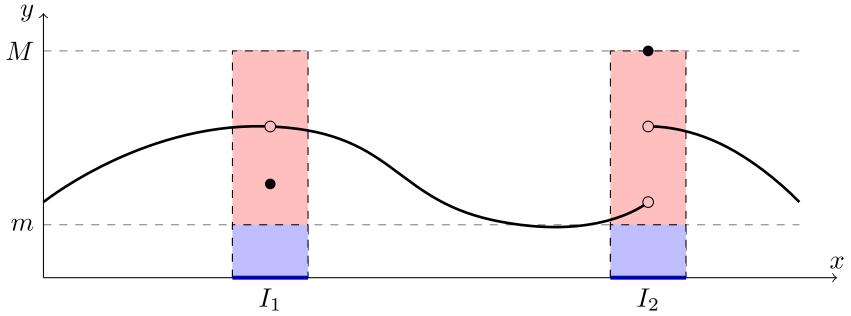


Figure 4.5: The set $W = I_1 \cup I_2$ contains the discontinuities of our function. Since our function is continuous away from W, we can make the difference between the upper and lower sums as small as we want, hence we need only bound the function on W. The difference in height will always be at worst $M - m$, 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 4.29

If f, g are integrable on $[a,b]$ and $f = g$ up to a set of Jordan measure zero, then $\int f = \int g$.

This is an easy corollary, whose proof effectively emulates that of Example 4.27, so I'll leave it as an exercise.

4.3 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.

4.3.1 Integration in the Plane

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

$$R_{ij} = [x_{i-1}, x_i] \times [y_{j-1}, y_j], \qquad {\scriptstyle i=1,\ldots,n \atop j=1,\ldots,m}.$$

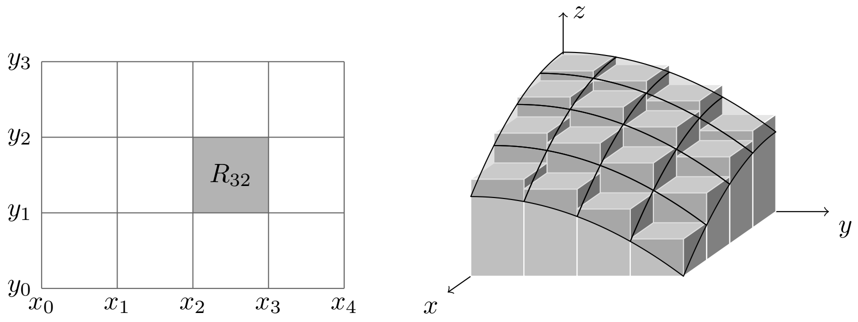


Figure 4.6: Left: The index of a rectangle is specified by its right endpoint in the corresponding sub-partitions in the x and y directions. Right: A Riemann sum for a function $z = f(x, y)$, where the tags are taken to be the midpoints of each rectangle.

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

$$S(f, P) = \sum_{\substack{i=1,\ldots,n \\ j=1,\ldots,m}} f(t_{ij}) A(R_{ij}), \qquad t_{ij} \in R_{ij}.$$

The notion of left- and right-Riemann sums no longer make sense, but certainly the upper and lower Riemann sums are still well-defined:

$$U(f, P) = \sum_{\substack{i=1,\ldots,n \\ j=1,\ldots,m}} \left[\sup_{\mathbf{x} \in R_{ij}} f(\mathbf{x}) \right] A(R_{ij}), \quad L(f, P) = \sum_{\substack{i=1,\ldots,n \\ j=1,\ldots,m}} \left[\inf_{\mathbf{x} \in R_{ij}} f(\mathbf{x}) \right] A(R_{ij}).$$

The usual definitions of Riemann integrability then carry over directly from Definition 4.18. 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 $U(f, P) - L(f, P) < \epsilon$. It is still unambiguous to denote the integral as $\int f$, but alternatives include

$$\iint_R f \, \mathrm{d}A \quad \text{or} \quad \iint f(x, y) \, \mathrm{d}A.$$

Here the $\mathrm{d}A$ term represents the "area element."

4.3.2 Integration in $\mathbb{R}^n$

While it will be hard to visualize, it's easy to see how to generalize this to $\mathbb{R}^n$ in general. 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 the any sum of the form

$$S(f, P) = \sum_{(i_1,\ldots,i_n)} f(t_i) V(R_{(i_1,\ldots,i_n)}), \qquad t \in R_{(i_1,\ldots,i_n)}.$$

As usual, one can define the upper $U(f, P)$ and lower $L(f, P)$ 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

$$U(f, P) - L(f, P) < \epsilon.$$

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:

$$\int f = \int f \, \mathrm{d}V = \int \cdots \int_S f \, \mathrm{d}V = \int \cdots \int f(\mathbf{x}) \, \mathrm{d}^n x = \int \cdots \int f(x_1, \ldots, x_n) \, \mathrm{d}V.$$

4.3.3 Properties of the Integral in $\mathbb{R}^n$

The usual results of integration apply, with minimal change to the proofs themselves.

1. **Linearity of the Integral:** If f_1, f_2 are integrable on $R \subseteq \mathbb{R}^n$ and $c_1, c_2 \in \mathbb{R}$ then $c_1 f_1 + c_2 f_2$ is integrable on R and

$$\int_R [c_1 f_1 + c_2 f_2] \, \mathrm{d}V = c_1 \int_R f_1 \, \mathrm{d}V + c_2 \int_R f_2 \, \mathrm{d}V.$$

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
$$\int_{R_1\cup R_2} f\,\mathrm{d}V = \int_{R_1} f\,\mathrm{d}V + \int_{R_2} f\,\mathrm{d}V.$$
3. **Monotonicity:** If $f_1 \leq f_2$ are integrable functions on R then
$$\int_R f_1\,\mathrm{d}V \leq \int_R f_2\,\mathrm{d}V.$$
4. **Subnormality:** If f is integrable on R and $|f|$ is integrable on R and
$$\left|\int f\,\mathrm{d}V\right| \leq \int |f|\,\mathrm{d}V.$$

Theorem 4.30

If $R \subseteq \mathbb{R}^n$ 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 4.28. □

4.3.4 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 $S \subseteq \mathbb{R}^n$ is a bounded set, we can always find a sufficiently large rectangle R containing S. We thus need only extend $f : S \to \mathbb{R}$ in a way that should not affect which rectangle we take. The way to do this is to define the *characteristic function of S*:

$$\chi : R \to \mathbb{R}, \quad \chi_S(\mathbf{x}) = \begin{cases} 1 & \mathbf{x} \in S \\ 0 & \text{otherwise} \end{cases}.$$

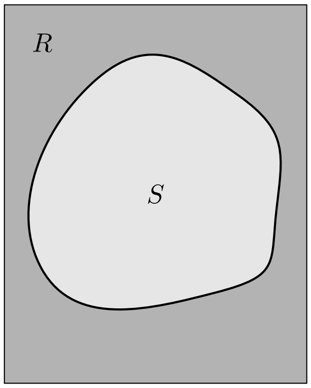


Figure 4.7: Every bounded set can be placed inside a rectangle.

Thus the function $f\chi_S : R \to \mathbb{R}, \mathbf{x} \mapsto f(\mathbf{x})\chi_S(\mathbf{x})$ is just $f(\mathbf{x})$ on S and 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 4.31

If $S \subseteq \mathbb{R}^n$ is Jordan measurable and the set of discontinuities of $f : S \to \mathbb{R}$ 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 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 4.30. □

In particular, we have the following corollary:

Corollary 4.32

If $S \subseteq \mathbb{R}^n$ is Jordan measurable then $\mu(S) = \displaystyle\int_S \chi_S$.

Theorem 4.33: Mean Value Theorem for Integrals

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

$$\int \cdots \int_S f(\mathbf{x})g(\mathbf{x}) \, \mathrm{d}^n\mathbf{x} = f(\mathbf{a}) \int \cdots \int_S g(\mathbf{x}) \, \mathrm{d}^n\mathbf{x}.$$

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

$$m \int \cdots \int_S g(\mathbf{x}) \, \mathrm{d}^n\mathbf{x} \leq \int \cdots \int_S f(\mathbf{x})g(\mathbf{x}) \, \mathrm{d}^n\mathbf{x} \leq M \int \cdots \int_S g(\mathbf{x}) \, \mathrm{d}^n\mathbf{x}.$$

or equivalently

$$m \leq \frac{\int \cdots \int_S f(\mathbf{x})g(\mathbf{x}) \, \mathrm{d}^n\mathbf{x}}{\int \cdots \int_S g(\mathbf{x}) \, \mathrm{d}^n\mathbf{x}} \leq M.$$

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. □

You have 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-dimensions simultaneously. This is not easy to do, so our next section will introduce a method by which we integrate our function in 'slices.'

4.4 Iterated Integrals

In developing the theory of integration in the plane and higher, it was necessary to consider partitions of rectangles and hence to consider the area of a function with respect to an infinitesimal volume $\mathrm{d}V$. 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 analyzing the change of the function *in a single, specific direction.* 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.

From our single variable calculus days, we know that integration is often more difficult than 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 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 $\Delta x_i = (x_i - x_{i-1}), \Delta y_j = (y_j - y_{j-1})$. If we look at this sum, we can decompose it as

$$\sum_{\substack{i \in \{1,\ldots,n\} \\ j \in \{1,\ldots,m\}}} f(\tilde{x}_i, \tilde{y}_j) \Delta y_j \Delta x_i = \sum_{i=1}^{n} \underbrace{\left[\sum_{j=1}^{m} f(\tilde{x}_i, \tilde{y}_j) \Delta y_j \right]}_{\approx \int_a^b f(\tilde{x}_i, y)\, \mathrm{d}y} \Delta x_i. \tag{4.1}$$

The heuristic idea is as follows: If we define the function

$$g_k(x) = \lim_{\ell(P_y) \to 0} S(f, P_x \times P_x) = \int_a^b f(x, y)\, \mathrm{d}y$$

then (4.1) gives

$$\int_R f(x, y)\, \mathrm{d}x = \lim_{\ell(P) \to 0} \sum_{i=1}^{n} \left[\sum_{j=1}^{m} f(\tilde{x}_i, \tilde{y}_j) \Delta y_j \right] \Delta x_i = \lim_{\ell(P_x) \to 0} \sum_{i=1}^{n} g_k(\tilde{x}_k) \Delta x_k = \int_c^d \left[\int_a^b f(x, y)\, \mathrm{d}x \right] \mathrm{d}y$$

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, we get

Theorem 4.34: Fubini's Theorem

Let $R = R_{\mathbf{x}} \times R_{\mathbf{y}}$ where $R_{\mathbf{x}} \subseteq \mathbb{R}^n, R_{\mathbf{y}} \subseteq \mathbb{R}^m$ are rectangles. Suppose $f : R \to \mathbb{R}$ is integrable, and assign it coordinates $f(\mathbf{x}, \mathbf{y})$ for $(\mathbf{x}, \mathbf{y}) \in R_{\mathbf{x}} \times R_{\mathbf{y}}$. For each $\mathbf{x}_0 \in R_{\mathbf{x}}$ define the function $f_{\mathbf{x}_0} : R_{\mathbf{y}} \to \mathbb{R}$ by $\mathbf{y} \mapsto f(\mathbf{x}_0, \mathbf{y})$. If $f_{\mathbf{x}_0}$ is integrable for each $\mathbf{x}_0 \in R_{\mathbf{x}}$, then

$$\int_R f \, \mathrm{d}V = \int_{R_{\mathbf{x}}} \left[\int_{R_{\mathbf{y}}} f(\mathbf{x}, \mathbf{y}) \, \mathrm{d}\mathbf{y} \right] \mathrm{d}\mathbf{x}.$$

Proof. Define the function

$$I : R_{\mathbf{x}} \to \mathbb{R}, \qquad I(\mathbf{x}) = \int_{R_{\mathbf{y}}} f_{\mathbf{x}}(\mathbf{y}) \, \mathrm{d}\mathbf{y}.$$

It suffices to show that I is integrable on $R_{\mathbf{x}}$, and that $\int I \, \mathrm{d}\mathbf{x} = \int f \, \mathrm{d}V$. The crux of the entire argument comes down to the fact that infima become larger over subsets, and the suprema become smaller. More precisely, if $A \subseteq B$ and $g : B \to \mathbb{R}$, then

$$\inf_{x \in A} g(x) \geq \inf_{x \in B} g(x) \quad \text{and} \quad \sup_{x \in A} g(x) \leq \sup_{x \in B} g(x).$$

Be sure to convince yourself of this fact, using a formal proof if necessary,

Let $\epsilon > 0$ be given. Since f is integrable, fix a partition $P = P_{\mathbf{x}} \times P_{\mathbf{y}}$ of R such that $U(f, P) - L(f, P) < \epsilon$. This same partition will do the trick for I. Indeed, fix a subrectangle $R_{ij} = R_i^{\mathbf{x}} \times R_j^{\mathbf{y}}$ of P and an $\mathbf{x}_0 \in R_{ij}$. We know that

$$\inf_{(\mathbf{x},\mathbf{y}) \in R_{ij}} f(\mathbf{x}, \mathbf{y}) \leq \inf_{\mathbf{y} \in R_j^{\mathbf{y}}} f(\mathbf{x}_0, \mathbf{y}) \quad \text{and} \quad \sup_{(\mathbf{x},\mathbf{y}) \in R_{ij}} f(\mathbf{x}, \mathbf{y}) \geq \sup_{\mathbf{y} \in R_j^{\mathbf{y}}} f(\mathbf{x}_0, \mathbf{y}).$$

This is easily justified by realizing the infimum/supremum on the right hand side is being taken over the set $\{\mathbf{x}_0\} \times R_j^{\mathbf{y}} \subseteq R_{ij}$. Multiplying by the volume of $V(R_j^{\mathbf{y}})$ and summing over all such rectangles gives

$$\sum_j \left[\inf_{(\mathbf{x},\mathbf{y}) \in R_{ij}} f(\mathbf{x}, \mathbf{y}) \right] V(R_j^{\mathbf{y}}) \leq \sum_j \left[\inf_{\mathbf{y} \in R_j^{\mathbf{y}}} f(\mathbf{x}_0, \mathbf{y}) \right] V(R_j^{\mathbf{y}}) = L(f_{\mathbf{x}_0}, P_{\mathbf{y}}) \leq \int_{R_{\mathbf{y}}} f_{\mathbf{x}_0}(\mathbf{y}) \, \mathrm{d}\mathbf{y} = I(\mathbf{x}_0).$$

$$\sum_j \left[\sup_{(\mathbf{x},\mathbf{y}) \in R_{ij}} f(\mathbf{x}, \mathbf{y}) \right] V(R_j^{\mathbf{y}}) \geq \sum_j \left[\sup_{\mathbf{y} \in R_j^{\mathbf{y}}} f(\mathbf{x}_0, \mathbf{y}) \right] V(R_j^{\mathbf{y}}) = U(f_{\mathbf{x}_0}, P_{\mathbf{y}}) \geq \int_{R_{\mathbf{y}}} f_{\mathbf{x}_0}(\mathbf{y}) \, \mathrm{d}\mathbf{y} = I(\mathbf{x}_0).$$

This must hold for all $\mathbf{x}_0$ – of which the left hand side is independent – so in turn we find that

$$\sum_j \left[\inf_{(\mathbf{x},\mathbf{y})\in R_{ij}} f(\mathbf{x},\mathbf{y}) \right] V(R_j^{\mathbf{y}}) \leq \inf_{\mathbf{x}\in R_i^{\mathbf{x}}} I(\mathbf{x}), \tag{4.2}$$

$$\sum_j \left[\sup_{(\mathbf{x},\mathbf{y})\in R_{ij}} f(\mathbf{x},\mathbf{y}) \right] V(R_j^{\mathbf{y}}) \geq \sup_{\mathbf{x}\in R_i^{\mathbf{x}}} I(\mathbf{x}). \tag{4.3}$$

This pretty much completes the proof, for if we multiply (4.2) and (4.3) by $V(R_i^{\mathbf{x}})$ and sum over all rectangles in $P_{\mathbf{x}}$, we get

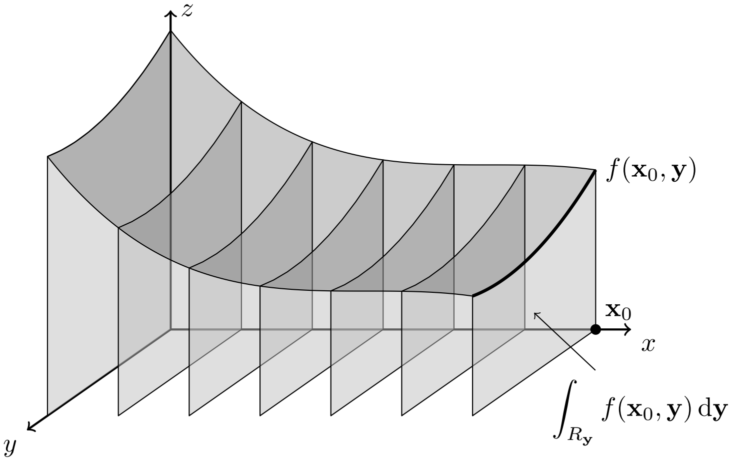
$$L(f,P) = \sum_i \sum_j \left[\inf_{(\mathbf{x},\mathbf{y})\in R_{ij}} f(\mathbf{x},\mathbf{y}) \right] V(R_j^{\mathbf{y}})V(R_i^{\mathbf{x}}) \leq \sum_i \inf_{\mathbf{x}\in R_i^{\mathbf{x}}} I(\mathbf{x})V(R_i^{\mathbf{x}}) = L(I,P_{\mathbf{x}}),$$

$$U(f,P) = \sum_i \sum_j \left[\sup_{(\mathbf{x},\mathbf{y})\in R_{ij}} f(\mathbf{x},\mathbf{y}) \right] V(R_j^{\mathbf{y}})V(R_i^{\mathbf{x}}) \geq \sum_i \sup_{\mathbf{x}\in R_i^{\mathbf{x}}} I(\mathbf{x})V(R_i^{\mathbf{x}}) = U(I,P_{\mathbf{x}}).$$

That is, $L(f,P) \leq L(I,P_{\mathbf{x}}) \leq U(I,P_{\mathbf{x}}) \leq U(f,P)$, from which

$$U(I,P_{\mathbf{x}}) - L(I,P_{\mathbf{x}}) \leq U(f,P) - L(f,P) < \epsilon,$$

shows that I is integrable. Moreover, $\int I(\mathbf{x})\,\mathrm{d}\mathbf{x} = \int f\,\mathrm{d}V$ (Exercise 4-22), as required. □



z
$f(\mathbf{x}_0,\mathbf{y})$
$\mathbf{x}_0$
x
y
$\int_{R_{\mathbf{y}}} f(\mathbf{x}_0,\mathbf{y})\,\mathrm{d}\mathbf{y}$

Figure 4.8: Fixing an $\mathbf{x}_0$, we look at the function $f(\mathbf{x}_0,\mathbf{y})$. The vertical planes represent $I(\mathbf{x}_0) = \int_{R_{\mathbf{y}}} f_{\mathbf{x}_0}(\mathbf{y})\,\mathrm{d}\mathbf{y}$. If this function is integrable for each $\mathbf{x}_0$, then the value of $\int f\,\mathrm{d}V$ is the sum of all the vertical planes.

Of course, the theorem also holds with the roles of x and y reversed.

Example 4.35

Determine the volume under the function $f(x,y) = xe^{x^2-y}$ on the rectangle $R = [0,1]\times[0,1]$.

Solution. Since $f(x,y)$ 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 yields the same results. If we integrate first with respect to x then y, we have

$$\begin{aligned}\int_0^1 \left[\int_0^1 xe^{x^2-y}\,\mathrm{d}x\right]\mathrm{d}y &= \int_0^1 \left[\frac{1}{2}e^{x^2-y}\right]_{x=0}^1 \mathrm{d}y = \frac{1}{2}(e-1)\int_0^1 e^{-y}\,\mathrm{d}y \\ &= \frac{1}{2}(e-1)\left[-e^{-y}\right]_0^1 = -\frac{1}{2}(e-1)(e^{-1}-1) \\ &= \cosh(1)-1.\end{aligned}$$

Conversely, let us instead integrate with respect to y first. We have

$$\int_0^1\left[\int_0^1 xe^{x^2-y}\,\mathrm{d}y\right]\mathrm{d}x = -(e^{-1}-1)\int_0^1 xe^{x^2}\,\mathrm{d}x = -(e^{-1}-1)(e-1) = \cosh(1)-1.$$

As expected, the result was the same either way. ■

Now rectangles are rather boring objects about which to integrate, so we again look at Jordan measurable sets $S \subseteq \mathbb{R}^{n+1}$. In particular, we will suppose that $S \subseteq \mathbb{R}^{n+1}$ can be written as

$$S = \{(\mathbf{x},y) \subseteq \mathbb{R}^n \times \mathbb{R} : \mathbf{x} \in K, \alpha(\mathbf{x}) \leq y \leq \beta(\mathbf{x})\},$$

where $K \subseteq \mathbb{R}^n$ is compact and Jordan measurable, with $\alpha, \beta : K \to \mathbb{R}$ piecewise C^1 functions. I will refer to this as a *y-simple* region. Note that y-simple regions are necessarily bounded, so that $S \subseteq R \times [-M,M]$ for some rectangle $R \subseteq \mathbb{R}^n$ and $M > 0$. If $f : S \to \mathbb{R}$ is a continuous function, then for any $\mathbf{x}_0 \in K$ its slices $f_{\mathbf{x}_0}$ are continuous at all but at most two points (the y-values $\alpha(\mathbf{x}_0)$ and $\beta(\mathbf{x}_0)$), and hence integrable on R. Moreover, beyond these two y-values, $f\chi_S$ is identically zero, so that

$$\int_S f\,\mathrm{d}A = \int_K \left[\int_{\alpha(\mathbf{x})}^{\beta(\mathbf{x})} f(\mathbf{x},y)\,\mathrm{d}y\right]\mathrm{d}\mathbf{x}.$$

In this way, we can peel the layers of a function to express it as a collected of iterated one-dimensional integrals.

The hardest part of setting up iterated integrals is decomposing a region into x_k-simple pieces. Note in particular that the relation $\alpha(\mathbf{x}) \leq y \leq \beta(\mathbf{x})$ says that the range of y depends explicitly upon $\mathbf{x}$. As we peel away further layers, this dependence decreases until only a single variable remains. The simplest way to determine the bounding functions α and β is to choose an arbitrary representative $\mathbf{x}_0$, and draw a line in the direction of increasing y. Note the function through which you pass when you first enter S, and the function you pass when you leave S. These are your α and β respectively.

Example 4.36

Determine the integral of $f(x,y) = \dfrac{y}{x^5+1}$ on the region $S \subseteq \mathbb{R}^2$ bounded by $y = x^2$ and $x = 1$ in the first quadrant.

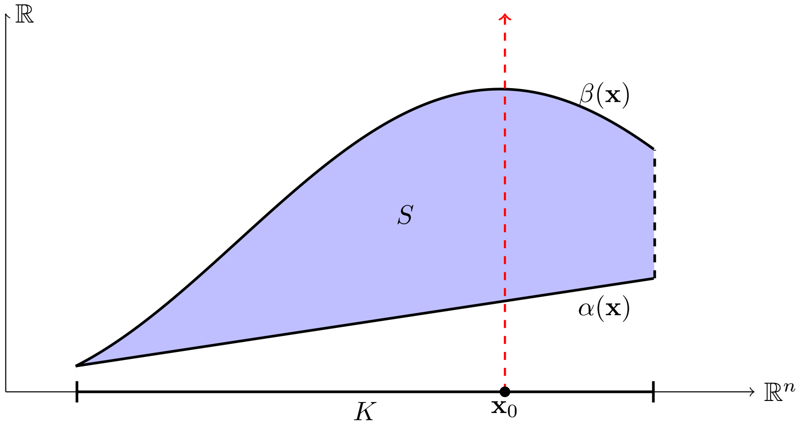


Figure 4.9: To determine α and β, choose a representative point $\mathbf{x}_0 \in K$, and draw an arrow in the direction of positive y. The function α is the curve you pass upon entering S, and β is the curve you pass upon leaving.

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 the y-simple set

$$S = \left\{(x,y) : 0 \leq x \leq 1, 0 \leq y \leq x^2\right\}.$$

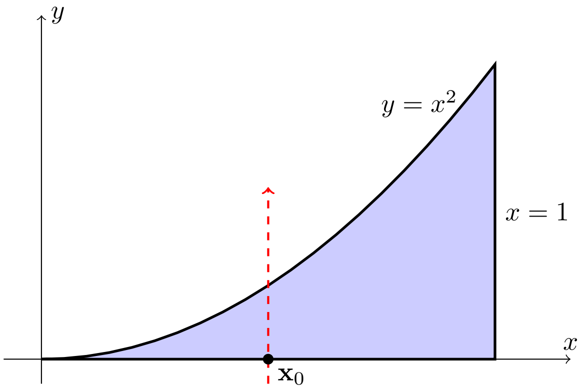


Figure 4.10: The figure corresponding to Example 4.36.

.

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:

$$\begin{aligned}\iint_S f \, \mathrm{d}A &= \int_0^1 \left[\int_0^{x^2} \frac{y}{x^5+1} \, \mathrm{d}y\right] \mathrm{d}x \\ &= \frac{1}{2}\int_0^1 \left[\frac{y^2}{x^5+1}\right]_0^{x^2} \mathrm{d}x = \frac{1}{2}\int_0^1 \frac{x^4}{x^5+1} \, \mathrm{d}x \\ &= \frac{1}{10} \ln|x^5+1|_0^1 = \frac{\ln(2)}{10}.\end{aligned}$$

■

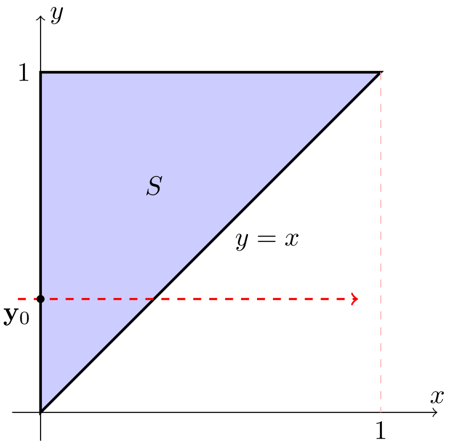


Figure 4.11: The figure for Example 4.37.

Note that the region in Example 4.36 also could have been described by an x-simple set

$$S = \left\{(x,y) : 0 \leq y \leq 1,\ \sqrt{y} \leq x \leq 1\right\},$$

so we also could have (attempted to) compute the integral as

$$\iint_S f \, \mathrm{d}A = \int_0^1 \left[\int_{\sqrt{y}}^1 \frac{y}{x^5+1} \, \mathrm{d}x \right] \mathrm{d}y.$$

This would not have worked as nicely, since $1/(x^5+1)$ is not easy to integrate. Being able to rewrite our domain is a useful skill, as sometimes we are given the boundary, but the problem is not amenable to the given description.

Example 4.37

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 function $f(x,y) = e^{y^2}$ is everywhere continuous, and the region is a simple triangle given in Figure 4.11. We can write our domain as either an x- or y-simple set as follows:

$$\begin{aligned} S &= \left\{(x,y) : 0 \leq x \leq 1,\ x \leq y \leq 1\right\} \\ &= \left\{(x,y) : 0 \leq y \leq 1,\ 0 \leq x \leq y\right\}. \end{aligned}$$

If we use the first description,

$$\int_S f \, \mathrm{d}A = \int_0^1 \left[\int_x^1 e^{y^2} \, \mathrm{d}y \right] \mathrm{d}x,$$

but the function e^{y^2} has no elementary anti-derivative, and we are stuck. Using the second description instead gives

$$\int_S f\,\mathrm{d}A = \int_0^1 \left[\int_0^y e^{y^2}\,\mathrm{d}x\right]\mathrm{d}y = \int_0^1 \left[xe^{y^2}\right]_{x=0}^{x=y}\mathrm{d}y = \int_0^1 ye^{y^2}\,\mathrm{d}y$$
$$= \left[\frac{1}{2}e^{y^2}\right]_{y=0}^1 = \frac{1}{2}(e-1). \qquad \blacksquare$$

Example 4.38

Determine $\iint_S xy\,\mathrm{d}A$ where S is the region bounded by $y = x-1$ and $y^2 = 2x+6$.

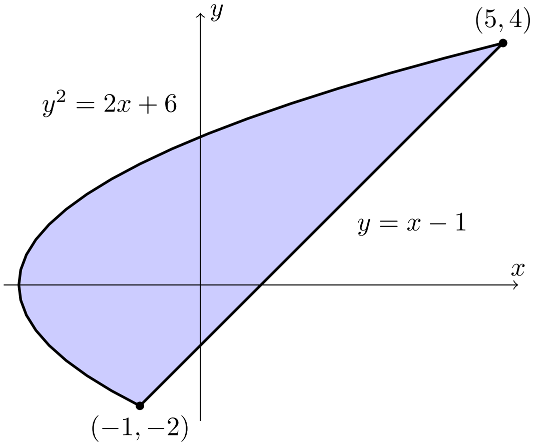


Figure 4.12: The figure for Example 4.38.

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 $(5,4)$. We can write S as the y-simple set

$$S = \left\{(x,y) : -2 \le y \le 4,\ \frac{1}{2}y^2 - 3 \le x \le y+1\right\}.$$

Integrating we get

$$\iint_S xy\,\mathrm{d}A = \int_{-2}^4 \left[\int_{y^2/2-3}^{y+1} xy\,\mathrm{d}x\right]\mathrm{d}y = \frac{1}{2}\int_{-2}^4 \left[x^2y\right]_{y^2/2-3}^{y+1}\mathrm{d}y$$
$$= \frac{1}{2}\int_{-2}^4 y\left[(y+1)^2 - \left(\frac{1}{2}y^2-3\right)^2\right]\mathrm{d}y$$
$$= \frac{1}{2}\int_{-2}^4 \left[-\frac{y^5}{4} + 4y^3 + 2y^2 - 8y\right]\mathrm{d}y$$
$$= \frac{1}{2}\left[-\frac{y^6}{24} + y^4 + \frac{2y^3}{3} - 4y^2\right]_{-2}^4 = 36. \qquad \blacksquare$$

The region S in Example 4.38 was x simple but not y-simple. That being said, we can write S as the union of two y-simple sets by making a cut at the line $x = -1$. In doing this, the integral becomes

$$\int_{-3}^{-1}\left[\int_{-\sqrt{2x+6}}^{\sqrt{2x+6}} xy\,\mathrm{d}y\right]\mathrm{d}x + \int_{-1}^{5}\left[\int_{x-1}^{\sqrt{2x+6}} xy\,\mathrm{d}y\right]\mathrm{d}x.$$

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 4.39

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: Our discussion thus far was limited to functions of two variables. 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.

Let's see what happens when we peel more layers using Fubini's Theorem. Suppose $f : S \subseteq \mathbb{R}^3 \to \mathbb{R}$ is continuous and S is z-simple, written as

$$S = \left\{ (x,y,z); (x,y) \in \tilde{K}, \varphi(x,y) \leq z \leq \psi(x,y) \right\}.$$

for some compact Jordan measurable set K and piecewise C^1 α, β. In this case,

$$\int_S f(x,y,z)\,\mathrm{d}V = \int_{\tilde{K}} \left[\int_{\varphi(x,y)}^{\psi(x,y)} f(x,y,z)\,\mathrm{d}z \right] \mathrm{d}A = \int_{\tilde{K}} I(x,y)\,\mathrm{d}A,$$

where $I(x,y) = \int_{\varphi(x,y)}^{\psi(x,y)} f(x,y,z)\,\mathrm{d}z$. If $\tilde{K}$ itself is y-simple, say $\tilde{K} = \{(x,y) : x \in K, \alpha(x) \leq y \leq \beta(x)\}$, then

$$\int_{\tilde{K}} I(x,y)\,\mathrm{d}A = \int_K \left[\int_{\alpha(x)}^{\beta(x)} I(x,y)\,\mathrm{d}y \right] \mathrm{d}x.$$

Putting everything together, the corresponding integral becomes

$$\iiint_S f(x,y,z)\,\mathrm{d}V = \int_K \left[\int_{\alpha(x)}^{\beta(x)} \left[\int_{\varphi(x,y)}^{\psi(x,y)} f(x,y,z)\,\mathrm{d}z \right] \mathrm{d}y \right] \mathrm{d}x.$$

Example 4.40

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 $(0,0,0)$. We begin by writing S as a z-simple set

$$S = \{(x,y,z) : (x,y) \in K, 0 \leq z \leq 1 - x - y\},$$

where K is the set bounded by $x = 0, y = 0$ and $x + y = 1$. We then write K as a y-simple set:

$$K = \{(x,y) : x \in [0,1], 0 \leq y \leq 1 - x\}.$$

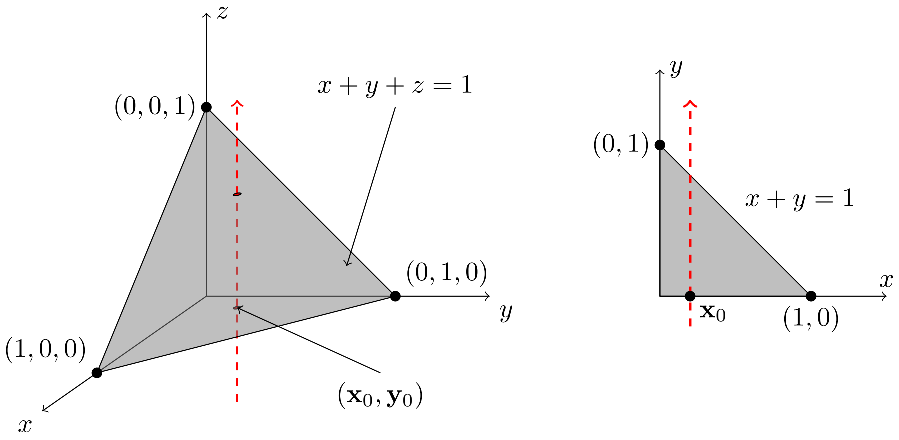


Figure 4.13: The diagram for Example 4.40. We begin by writing S as a z-simple, as illustrated in the left image. Having peeled away the z layer, we then write compact space in which (x,y) lives as a y-simple set. This decomposes the integral into an iterated integral.

The function z is clearly integrable, so applying Fubini's Theorem yields the iterated integral

$$\begin{aligned}\iiint_S z\ \mathrm{d}V &= \int_0^1\left[\int_0^{1-x}\left[\int_0^{1-x-y} z\,\mathrm{d}z\right]\mathrm{d}y\right]\mathrm{d}x\\ &= \int_0^1\left[\int_0^{1-x}\left[\frac{z^2}{2}\right]_0^{1-x-y}\mathrm{d}y\right]\mathrm{d}x\\ &= \frac{1}{2}\int_0^1\left[\int_0^{1-x}(1-x-y)^2\,\mathrm{d}y\right]\mathrm{d}x = \frac{1}{2}\int_0^1\left[-\frac{(1-x-y)^3}{3}\right]_0^{1-x}\mathrm{d}x\\ &= \frac{1}{6}\int_0^1(1-x)^3\,\mathrm{d}x = \frac{1}{6}\left[-\frac{(1-x)^4}{4}\right]_0^1 = \frac{1}{24}.\end{aligned}$$

■

Example 4.41

Determine $\iiint_S(2x+4z)\,\mathrm{d}V$ where S is the region bounded by the planes $y=x$, $z=x$, $z=0$, and $y=x^2$.

Solution. Stare at these equations to visualize the space. We have several nice decompositions of the, but one can be given as

$$S=\left\{(x,y,z): 0\leq x\leq 1,\ x^2\leq y\leq x,\ 0\leq z\leq x\right\}.$$

Our function is clearly C^1 on this set, so we can apply Fubini to get

$$\begin{aligned}\iiint_S f\,\mathrm{d}V &= \int_0^1\left[\int_{x^2}^x\left[\int_0^x(2x+4z)\,\mathrm{d}z\right]\mathrm{d}y\right]\mathrm{d}x = \int_0^1\left[\int_{x^2}^x 2x^2+2x^2\,\mathrm{d}y\right]\mathrm{d}x\\ &= 2\int_0^1\left(4x^3-4x^4\right)\mathrm{d}x = 4\left[\frac{1}{4}x^4-\frac{1}{5}x^5\right]_0^1 = \frac{1}{5}.\end{aligned}$$

■

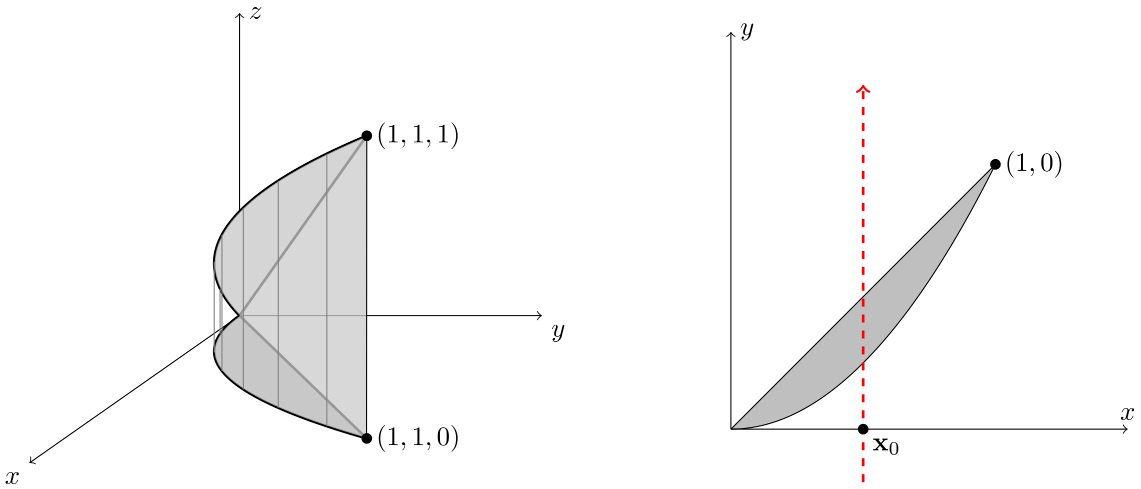


Figure 4.14: The diagram for Example 4.41.

4.5 Improper Integrals

Here we'll examine how to integrate unbounded functions and how to integrate on unbounded sets. Just as with single variable calculus, this requires that we extend the notion of an integral, but we'll see that the extended and traditional agree when they should.

Definition 4.42

Let $U \subseteq \mathbb{R}^n$ be an open (not necessarily bounded) set, and $f : U \to \mathbb{R}$ a non-negative continuous function. Let $\mathcal{K} = \{K \subseteq U : K \text{ is compact, Jordan measurable}\}$. We define the *extended integral of f over U* to be

$$⨍_A f := \sup_{K \in \mathcal{K}} \int_K f \quad \text{if the supremum exists.}$$

When f is not non-negative, let $f_+ = \max\{f, 0\}$ and $f_- = \max\{-f, 0\}$, so that $f_\pm$ are non-negative and $f = f_+ - f_-$. We define

$$⨍_A f := ⨍_A f_+ - ⨍_A f_-, \tag{4.4}$$

provided the extended integrals on the right both exist.

We restrict our attention to non-negative functions in order to ensure the supremum is doesn't do anything perverse. Once we show that the extended integral is linear, it will suffice to show all of our proofs exclusively for non-negative functions, since Equation (4.4) will then follow by linearity.

Taking the supremum over the collection of all measurable compact subsets of A is going to be a hassle. Recall from Exercise 2-70 that every open set admits a compact exhaustion; that is, a countable collection of compact sets $K_i, i \in \mathbb{N}$ such that $\int K_i \subseteq K_{i+1}$. The next result shows that we can compute extended integrals using any compact exhaustion.

Proposition 4.43

Let $A \subseteq \mathbb{R}^{n+1}$ be an open set, and $f : A \to \mathbb{R}$ a continuous function. If $(K_i)_{i=1}^{\infty}$ is any compact exhaustion of A, then

$$\fint_A f \quad \text{exists if and only if the sequence} \quad \int_{K_i} |f| \quad \text{is bounded.}$$

Proof. Note immediately that $\fint_{K_i} |f|$ is an increasing sequence, so boundedness will imply convergence of the sequence. Suppose for now that f is a non-negative function, so $|f| = f$. Let $\mathcal{K}$ denote the set of compact, Jordan measurable subsets of U.

[⇒] Suppose that $\fint_U f$ exists. Note that $K_i \in \mathcal{K}$ for each $i \in \mathbb{N}$, so

$$\int_{K_n} f \leq \sup_{K \in \mathcal{K}} \int_K f = \fint_U f. \tag{4.5}$$

[⇐] Assume the sequence $\int_{K_i} |f|$ is bounded, so that it converges. It suffices to show that $\left\{\int_K f : K \in \mathcal{K}\right\} \subseteq \mathbb{R}$ is bounded from above, for then the Completeness Axiom will imply the existence of a supremum. Fix an arbitrary compact Jordan measurable subset $K \subseteq U$. Note that K is covered by the K_i^{int}, and hence by finitely many of the K_i^{int}. However, as $K_i^{\text{int}} \subseteq K_{i+1}^{\text{int}}$, there exists an $N \in \mathbb{N}$ such that $K \subseteq K_N^{\text{int}} \subseteq K_N$. This in turn implies that

$$\int_K f \leq \int_{K_N} f \leq \lim_{n\to\infty} \int_{K_n} f. \tag{4.6}$$

The bound on the right hand side is independent of K, and hence gives the desired upper bound. The Completeness Axiom now guarantees that $\fint_U f$ exists.

Now suppose f is not non-negative. Since $0 \leq f_\pm \leq |f|$, if $\int_{K_i} |f|$ is bounded then so too are $\int_{K_i} f_\pm$, by Monotonicity of the Integral. Conversely, if both $\int_{K_i} f_\pm$ are bounded, then

$$\int_{K_i} |f| = \int_{K_i} [f_+ + f_-] = \int_{K_i} f_+ + \int_{K_i} f_i$$

is bounded by the sum of the bounds of the two sequences. Thus $\int_{K_i} |f|$ is bounded if and only if both $\int_{K_i} f_\pm$ are bounded, which immediately generalizes the [⇒] direction. On the other hand, if $\int_{K_i} f_\pm$ are bounded, then (4.6) generalizes to

$$\int_K f = \int_K [f_+ - f_-] \leq \lim_{n\to\infty} \int_{K_n} [f_+ - f_-] \leq \left[\lim_{n\to\infty} \int_{K_n} f_+\right] - \left[\lim_{n\to\infty} \int_{K_n} f_-\right],$$

completing the proof. □

The usual results regarding integration now hold; namely, if $U \subseteq \mathbb{R}^n$ is an open set and $f : U \to \mathbb{R}$ is continuous and extended integrable, then

1. **Linearity:** If $g : U \to \mathbb{R}$ is continuous and extended integrable, then for any $c \in \mathbb{R}$, $f + g$ and cf are extended integrable and
$$\fint_U [f+g] = \fint_U f + \fint_U g \quad \text{and} \quad \fint_U [cf] = c \fint_U f.$$

2. **Additivity of Domain:** If $V \subseteq \mathbb{R}^n$ is another open set and f can be extended to a continuous function on B, then
$$\fint_{U \cup V} f = \fint_U f + \fint_V f - \fint_{U\cap V} f.$$

3. **Monotonicity of Integral:** If $g : U \to \mathbb{R}$ is continuous and $f(x) \leq g(x)$ for all $x \in [a, b]$ then
$$\fint_U f \leq \fint_U g.$$
Notably, $\left|\fint_U f\right| \leq \fint_U |f|$.

4. **Monotonicity of Domain:** If $V \subseteq U$ is open and f is non-negative, then
$$\fint_V f \leq \fint_U f.$$

The proofs are left to you as an exercise.

Example 4.44

Let $S = \{(x, y) : x, y > 0\}$. Determine $\displaystyle\iint_S e^{-(x+y)}\,\mathrm{d}A$.

Solution. It suffices to determine the result for any compact exhaustion of S. Let $K_n = [1/n, n] \times [1/n, n]$. Certainly $K_n \subseteq K_{n+1}^{\text{int}}$ and $\bigcup_n K_n = S$, so
$$\begin{aligned}\iint_S e^{-(x+y)}\,\mathrm{d}A &= \lim_{n\to\infty} \iint_{K_n} e^{-(x+y)}\,\mathrm{d}A = \lim_{n\to\infty} \int_{1/n}^n \int_{1/n}^n e^{-(x+y)}\,\mathrm{d}x\,\mathrm{d}y \\ &= \lim_{n\to\infty} \left[\int_{1/n}^n e^{-x}\,\mathrm{d}x\right]^2 = \lim_{n\to\infty} \left[e^{-1/n} - e^{-n}\right]^2 = 1.\end{aligned}$$
■

Example 4.45

Let $S = \left\{(x, y) : x > 0, x^2 < y < 3x^2\right\}$ and $f(x, y) = x/(y + 1)^3$. Determine $\displaystyle\iint_S f(x, y)\,\mathrm{d}A$.

Solution. Choose as a compact exhaustion the sets
$$K_n = \left\{(x, y) : 1/n \leq x \leq n, x^2 + 1/n \leq y \leq 2x^2 - 1/n\right\},$$

which are conveniently written in y-simple form. Applying Fubini's Theorem, we get

$$\begin{aligned}\iint_{K_n} \frac{x}{(y+1)^3}\,\mathrm{d}A &= \int_{1/n}^{n}\int_{x^2+1/n}^{2x^2-1/n} \frac{x}{(y+1)^3}\,\mathrm{d}y\,\mathrm{d}x \\ &= \int_{1/n}^{n}\left[\frac{x}{(x^2+1/n+1)^2} - \frac{x}{(2x^2-1/n+1)^2}\right]\mathrm{d}x\end{aligned}$$

You can quickly check that

$$\lim_{n\to\infty}\int_{1/n}^{n}\frac{x}{(ax^2+b/n+1)^2}\,\mathrm{d}x = \lim_{n\to\infty}\frac{1}{2a}\left[\frac{n^2}{1+bn+n^2} - \frac{n}{an^3+b+n}\right] = \frac{1}{2a},$$

so that

$$\iint_S \frac{x}{(y+1)^3}\,\mathrm{d}A = \frac{1}{4}. \qquad \blacksquare$$

The extended integral is defined for any open set, but for bounded open sets we already have a definition of the integral. At the very least, one would hope that the two definitions agree on bounded sets.

Proposition 4.46

If $U \subseteq \mathbb{R}^n$ is an open bounded set and $f : U \to \mathbb{R}$ is a bounded continuous function, then $\fint_U f$ exists. Moreover, if $\int_U f$ exists, then $\fint_U f = \int_U f$.

Proof. We begin by showing that $\fint_U f$ exists. Since f is bounded, choose an $M > 0$ such that $|f(\mathbf{x})| \leq M$ for all $x \in U$. Moreover, pick a rectangle R containing U. It suffices to show that $\int_{K_i} |f|$ is a bounded sequence, and to this effect we have

$$\int_{K_n} |f| \leq \int_{K_n} M \leq \int_R M = \mu(R)M.$$

The bound on the right is independent of K_n, showing that the sequence is bounded.

Now suppose that $\int_U f$ exists. We'll show a double inequality, establishing that $\int_U f = \fint_U f$. It suffices to show the result for non-negative f, since linearity of the corresponding integrals will complete the result. Fix a $K \in \mathcal{K}$, noting that $f(\mathbf{x}) = f(\mathbf{x})\chi_U(\mathbf{x})$ on K, so that

$$\int_K f = \int_K f\chi_U \leq \int_R f\chi_U = \int_U f.$$

The right hand side is independent of K, so in taking the supremum over $\mathcal{K}$ we get $\fint_U f \leq \int_U f$.

For the opposing inequality, choose a partition P of the enveloping rectangle R. Let $\mathcal{R}_P$ be those subrectangles defined by the partition which live strictly inside of U, and set $K_P = \bigcup \mathcal{R}_P$ which is compact by assumption. The infimum of $f\chi_U$ on any subrectangle not contained strictly in U is necessarily zero, so

$$\begin{aligned}L(f\chi_U, P) &= \sum_{R\in\mathcal{R}}\left[\inf_{\mathbf{x}\in R} f(\mathbf{x})\chi_U(\mathbf{x})\right]\mu(R) \leq \sum_{R\in\mathcal{R}}\int_R f\chi_U \\ &= \int_{K_P} f\chi_U = \int_{K_P} f \leq \fint_U f & \text{since } K_p \in \mathcal{K}.\end{aligned}$$

Taking the supremum over P thus shows that $\int_U f \leq \fint_U f$. Both inequalities now give equality. □

An alternative approach to using a compact exhaustion of a set is to use a partition of unity (with compact supports).

Proposition 4.47

Suppose $U \subseteq \mathbb{R}^n$ is an open set, and $f : U \to \mathbb{R}^n$ is a continuous function. If $(\phi_i)_{i=1}^n$ is any compactly supported C^1 partition of unity on U, then

$$\fint_U f \quad \text{exists if and only if} \quad \sum_{i=1}^\infty \int_U \phi_i f \quad \text{converges.}$$

In this case, the value of the extended integral is equal to the value of the series.

Because of Exercise 4-36, we can interpret elements of the above series as either $\int_A \phi_i f = \int_{\operatorname{supp}(\phi_i)} f$ or $\fint_A \phi_i f$.

Proof. Once again, it suffices to show the result assuming that f is non-negative. Let $\mathcal{K}$ denote the compact, Jordan measurable subsets of A, and $S_i = \operatorname{supp}(\phi_i)$.

[⇐] Suppose that $\sum_i \int_A \phi_i f$ converges. It suffices to show that $\left\{\int_K f : K \in \mathcal{K}\right\}$ is bounded, for then the result will follow by the Completeness Axiom. Fix a $K \in \mathcal{K}$. By Exercise 3-64, there exists an $M \in \mathbb{N}$ such that $\phi_i(\mathbf{x}) = 0$ for all $\mathbf{x} \in K$ and $i > M$, in which case $f(\mathbf{x}) = \sum_{i=1}^M \phi_i(\mathbf{x}) f(\mathbf{x})$ on K. Integrating we get

$$\int_K f = \int_K \sum_{i=1}^M \phi_i f = \sum_{i=1}^M \int_K \phi_i f \leq \sum_{i=1}^\infty \int_K \phi_i f,$$

establishing the bound as required.

[⇒] Conversely, suppose $\fint_A f$ exists. Let $K_n = S_1 \cup S_2 \cup \cdots \cup S_n$, a finite union of compact sets and hence compact. Note that

$$\int_A \phi_i f = \int_{K_n} \phi_i f.$$

If σ_n is the nth partial sum of the series, then σ_n is increasing since

$$\begin{aligned} \sigma_n &= \sum_{i=1}^n \int_A \phi_i f = \sum_{i=1}^n \int_{K_n} \phi_i f \\ &\leq \sum_{i=1}^n \int_{K_{n+1}} \phi_i f + \int_{K_{n+1}} \phi_{n+1} f && \text{Monotonicity of Domain and adding a non-negative term} \\ &= \sum_{i=1}^{n+1} \int_{K_{n+1}} \phi_i f = \sigma_{n+1}. \end{aligned}$$

The partial sums are also bounded, since

$$\sigma_n = \sum_{i=1}^n \int_A \phi_i f = \sum_{i=1}^n \int_{K_n} \phi_i f = \int_{K_n} \sum_{i=1}^n \phi_i f \leq \int_{K_n} f \leq \fint_A f.$$

Hence the series converges by the Monotone Convergence Theorem. □

4.6 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 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.

4.6.1 Coordinate Systems

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 in $\mathbb{R}^2$ is a familiar example, wherein a point $\mathbf{p} \in \mathbb{R}^2$ is described by (x, y), describing the horizontal and vertical displacement of $\mathbf{p}$ from the origin.

But there are many other types of coordinate systems. For example, *polar coordinates* indicate a point $\mathbf{p} \in \mathbb{R}^2$ using the numbers (r, θ), where r is the distance of $\mathbf{P}$ from the origin, and θ is the angle that $\mathbf{p}$ makes with respect to the positive x-axis.

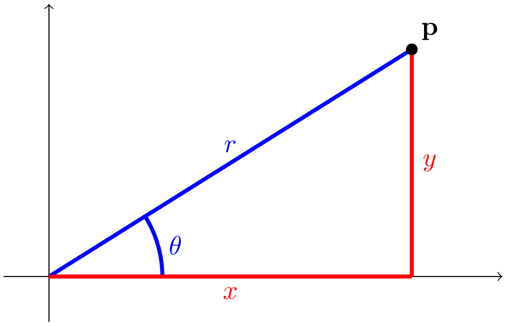


Figure 4.15: The point $\mathbf{p}$ can be described in terms of its Cartesian coordinates (x, y) in red, or its polar coordinates (r, θ) in blue.

With this point of view, the object $\mathbf{p}$ has not changed; rather, only how we've chosen to represent it has changed. This is similar to having two different bases for a vector space. The relationship between Cartesian and polar coordinates is given by $(x, y) = \mathbf{G}(r, \theta) = (r\cos(\theta), r\sin(\theta))$, as can quickly be discerned from Figure 4.15. One can think of $\mathbf{G}$ as an object which translates (r, θ) coordinates into (x, y) coordinates, or as a function from two different coordinate spaces (Figure 4.16).

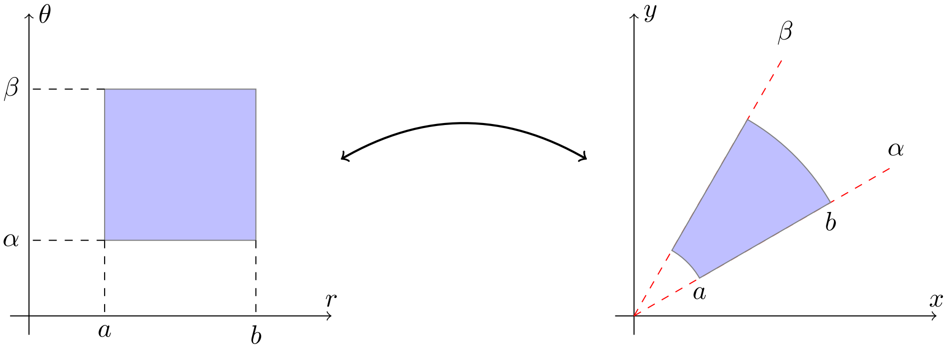


Figure 4.16: How a simple square in polar coordinates (left) changes under the map $(x, y) = \mathbf{G}(r, \theta) = (r\cos(\theta), r\sin(\theta))$.

However, I mentioned earlier that coordinate systems must uniquely identify points, and polar coordinates will not do this without restricting the values for r and θ. For example, $(-1, 0)_{\text{polar}}$ and $(1, \pi)_{\text{polar}}$ both correspond to $(-1, 0)_{\text{Cart}}$, while $(1, \pi)_{\text{polar}}$ and $(1, 3\pi)_{\text{polar}}$ both correspond to $(1, 0)_{\text{Cart}}$. For this reason, we demand $r > 0$ and $\theta \in [0, 2\pi)$.

In $\mathbb{R}^3$ there are three popular coordinate systems:

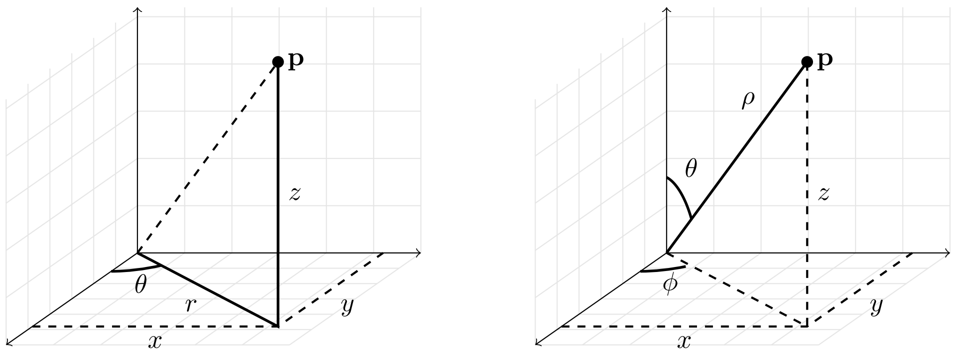
1. Our usual Cartesian coordinate system (x, y, z), with $x, y, z \in \mathbb{R}$,
2. *Cylindrical Coordinates* of the form (r, θ, z) with $r > 0, \theta \in [0, 2\pi), z \in \mathbb{R}$. Here we use polar coordinates (r, θ) to define the (x, y) coordinates of the point, and z agrees with its usual Cartesian interpretation.
3. *Spherical Coordinates* prescribed by (ρ, ϕ, θ), where $\rho > 0, \phi \in [0, 2\pi), \theta \in [0, \pi)$. This is the three dimensional version of polar coordinates, where ρ represents the distance of the point $\mathbf{p}$ from the origin, ϕ the azimuthal angle, and θ the polar angle.

These are all pictured in Figure 4.17.

Definition 4.48

Let U be an open set.

1. If $\mathbf{G} : U \to \mathbb{R}^n$ is a C^1 injective map, and $|\det D\mathbf{G}(\mathbf{x})| \neq 0$ for all $\mathbf{x} \in U$, we say $\mathbf{G}$ is a *change of variables*.
2. If $\mathbf{G} : U \to V \subseteq \mathbb{R}^n$ is C^1 and invertible with C^1 inverse $\mathbf{G}^{-1}$, we say $\mathbf{G}$ is a *diffeomorphism*.



Cylindrical

Spherical

Figure 4.17: Left: Cylindrical coordinates are a combination of polar coordinates in the xy-plane and Cartesian coordinates. Right: Spherical coordinates in $\mathbb{R}^3$ generalize polar coordinates.

Theorem 4.49

If $U \subseteq \mathbb{R}^n$, there is a bijective correspondence between changes of variable, and diffeomorphisms on U.

I'm going to leave the proof of Theorem 4.49 to you. It's fairly straightforward, but is good practice. The transformations for polar, cylindrical, and spherical coordinates into Cartesian are below

$$\begin{aligned}(x, y) &= \mathbf{G}_{\text{pol}}(r, \theta) = (r\cos(\theta), r\sin(\theta))\\ (x, y, z) &= \mathbf{G}_{\text{cyl}}(r, \theta, z) = (r\cos(\theta), r\sin(\theta), z)\\ (x, y, z) &= \mathbf{G}_{\text{sph}}(\rho, \phi, \theta) = (\rho\cos(\phi)\sin(\theta), \rho\sin(\phi)\sin(\theta), \rho\cos(\theta)).\end{aligned}$$

In Exercise 4-38 you'll determine the corresponding inverse maps.

The quantity $|\det D\mathbf{G}(\mathbf{x})|$ will be appearing frequently in the next few pages, and is called the *Jacobian*. Note the difference between the Jacobian matrix $D\mathbf{G}(\mathbf{x})$ written in the standard basis of $\mathbb{R}^n$, and the Jacobian we just defined. A few important Jacobians for the diffeomorphisms translating into Cartesian coordinate systems are computed below. Note that while the (co)domain are important in determining whether $\mathbf{G}$ is a diffeormophism, the computation of the Jacobian is independent of them.

1. **Polar Coordinates:** The transformation is $(x, y) = \mathbf{G}(r, \theta) = (r\cos(\theta), r\sin(\theta))$, from which
$$|\deg D\mathbf{G}(r, \theta)| = \left|\det\begin{bmatrix}\cos(\theta) & \sin(\theta)\\ -r\sin(\theta) & r\cos(\theta)\end{bmatrix}\right| = |r\cos^2(\theta) + r\sin^2(\theta)| = r.$$

2. **Cylindrical Coordinates:** Recall that cylindrical coordinates are related to Cartesian coordinates by $(x, y, z) = D\mathbf{G}(r, \theta, z) = (r\cos(\theta), r\sin(\theta), z)$. Hence
$$|\det D\mathbf{G}(r, \theta, z)| = \left|\det\begin{bmatrix}\cos(\theta) & -r\sin(\theta) & 0\\ \sin(\theta) & r\cos(\theta) & 0\\ 0 & 0 & 1\end{bmatrix}\right| = r.$$

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.

3. **Spherical Coordinates:** Cartesian and Spherical coordinates are related by $(x,y,z) = \mathbf{G}(\rho,\phi,\theta) = (\rho\sin\phi\cos\theta, \rho\sin\phi\sin\theta, \rho\cos\phi)$, and

$$\begin{aligned} |\det D\mathbf{G}(r,\phi,\theta)| &= \left|\det\begin{bmatrix} \sin\phi\cos\theta & \rho\cos\phi\cos\theta & -\rho\sin\phi\sin\theta \\ \sin\phi\sin\theta & \rho\cos\phi\sin\theta & \rho\sin\phi\cos\theta \\ \cos\phi & -\rho\sin\phi & 0 \end{bmatrix}\right| \\ &= \cos\phi\left[\rho^2\cos\phi\sin\phi\cos^2\theta - \rho^2\cos\phi\sin\phi\sin^2\theta\right] \\ &\quad + \rho\sin\theta\left[\rho\sin^2\phi\cos^2\theta + \rho\sin^2\phi\sin^2\theta\right] \\ &= \rho^2\cos^2\phi\sin\theta + \rho^2\sin\theta\sin^2\phi \\ &= \rho^2\sin\theta. \end{aligned}$$

Once we have a diffeomorphism $\mathbf{G}: U \to V$ we know that the 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.

Recall from Proposition 4.9(4) that if $T: \mathbb{R}^n \to \mathbb{R}^n$ is a linear map and $S \subseteq \mathbb{R}^n$ is measurable, then $\mu(T(S)) = |\det T|\mu(S)$. The same idea will hold for non-linear maps; namely, that $|\det D\mathbf{G}(\mathbf{x})|$ represents the infinitesimal change in scale under the diffeomorphism $\mathbf{G}$.

This can be seen at the level of Riemann sums. Fix a rectangle $R = [a,b]\times[c,d]$ in (x,y) space and an integrable function $f: R \to \mathbb{R}$. Let $P = P_x \times P_y = \{(x_i, y_j) : i \in I, j \in J\}$ denote a partition of this rectangle, and write a typical Riemann sum

$$S(f,P) = \sum_{i,j} f(x_{ij}, j_{ij})\Delta x_i \Delta y_j, \tag{4.7}$$

where $\Delta x_i = x_i - x_{i-1}$ and $\Delta y_j = y_j - y_{j-1}$ are the dimensions of the (i,j)th subrectangle.

Realizing $(x,y) = \mathbf{G}_{\text{pol}}(r,\theta) = (r\cos(\theta), r\sin(\theta))$, let (r_{ij},θ_{ij}) be such that $(x_{ij}, y_{ij}) = \mathbf{G}_{\text{pol}}(r_{ij},\theta_{ij})$. Ignoring for the moment how the rectangle transforms, let's see how $\Delta x_i \Delta y_i$ relates to $\Delta r_i \Delta\theta_j$. Look again at Figure 4.16. The area of a circular sector of radius r and angle θ is $A = r^2\theta/2$, meaning

$$\Delta x_i \Delta y_i = \frac{r_i^2\Delta\theta_j}{2} - \frac{r_{i-1}^2\Delta\theta_j}{2} = \frac{1}{2}(r_i + r_{i-1})\Delta r_i\Delta\theta_j$$

Substituting this into (4.7) yields

$$S(f,P) = \sum_{i,j}\frac{r_i + r_{i-1}}{2} f(r_{ij},\theta_{ij})\Delta r_i\Delta\theta_j.$$

This is *not* a Riemann sum. However, by Bliss' theorem, we know that it nonetheless converges to a Riemann sum of the form

$$\int_R f(x,y)\,\mathrm{d}x\,\mathrm{d}y = \int_{G^{-1}(R)} \frac{r+r}{2} f(r,\theta)\,\mathrm{d}r\,\mathrm{d}\theta = \int_{G^{-1}(R)} f(r,\theta) r\,\mathrm{d}r\,\mathrm{d}\theta. \tag{4.8}$$

Effectively, $\mathrm{d}x\,\mathrm{d}y = r\,\mathrm{d}r\,\mathrm{d}\theta$, and the relationship between them is precisely given by $|\det D\mathbf{G}_{\text{pol}}(r,\theta)|$.

4.6.2 The Change of Variables Theorem

Equation (4.9) should remind you of integration by substitution, and fleshing out the multivariable version of substitution is the object of this section. I will warn you ahead of time that the proof of the theorem is significantly more complicated and technical than its one dimensional analog.

Recall that in one-dimension, the Method of Substitution tells us that if $g : [a, b] \to \mathbb{R}$ is differentiable, $g'(x) > 0$, and $f : g([a, b]) \to \mathbb{R}$ is integrable, then

$$\int_a^b f(g(x))g'(x)\,\mathrm{d}x = \int_{g(a)}^{g(b)} f(u)\,\mathrm{d}u \tag{4.9}$$

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.

Integration by substitution hinged upon an orientable integral. Namely, if $g'(x) < 0$ then the bounds of integration would reverse, though we could fix this by introducing an additional minus sign:

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

An equivalent statement using non-oriented integrals is to write

$$\int_{[a,b]} f(g(x))|g'(x)|\,\mathrm{d}x = \int_{g([a,b])} f(u)\,\mathrm{d}u \quad \text{or} \quad \int_{g^{-1}(J)} f'(g(x))|g'(x)|\,\mathrm{d}x = \int_J f(u)\,\mathrm{d}u,$$

where $J = g([a, b])$. The latter notation is convenient, as it keeps the right hand side independent of any g-terms.

Theorem 4.50: 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

$$\int_{\mathbf{G}(S)} f(\mathbf{u})\,\mathrm{d}\mathbf{u} = \int_S f(\mathbf{G}(\mathbf{x}))|\det D\mathbf{G}(\mathbf{x})|\,\mathrm{d}\mathbf{x}.$$

Remark 4.51

1. The Change of Variables Theorem requires that $\mathbf{G}$ be a diffeomorphism, which does not seem to be the case for Substitution. Indeed, this is true, but we can get around this by remembering that the integral does not care about what happens on a measure zero set. Hence there will be times when either $\mathbf{G}$ will not be a diffeomorphism, but the failure will only be on a zero measure set, or the dual paradigm wherein we'll restrict ourselves to subsets $T' \subseteq T$ and $S' \subseteq S$ which differ from their parents by only zero measure sets.

2. Just as setting $u = g(x)$ gave rise to $\mathrm{d}u = g'(x)\,\mathrm{d}x$, in Change of Variables we have $\mathrm{d}u = |\det D\mathbf{G}(\mathbf{x})|\,\mathrm{d}x$. We thus see that $g'(x)$ is changing the scale of the variables u

and x.

3. While Substitution and Inverse Substitution must be treated differently in one dimension, the restriction that $\mathbf{G}$ be a diffeomorphism means Change of Variables gives both results for free. In this case, the theorem becomes

$$\int_S f(\mathbf{x})\,\mathrm{d}\mathbf{x} = \int_{\mathbf{G}^{-1}(S)} f(\mathbf{G}(\mathbf{u}))|\det D\mathbf{G}(\mathbf{u})|\,\mathrm{d}\mathbf{u}.$$

Proof. I will sketch the proof, as the complete proof is long, technical, and not terribly intuitive. We will make several arguments to reduce the complexity of the statement to showing something much simpler.

Claim 1: Suppose S admits an open cover $\mathcal{O}$ such that every element of $\mathcal{O}$ lies strictly within S. If the theorem holds on each element $U \in \mathcal{O}$, then it holds on all of S as well. In effect, we can build up to S by looking at the theorem locally.

The set $\mathbf{G}(\mathcal{O}) = \{\mathbf{G}(U) : U \in \mathcal{O}\}$ is a cover of $\mathbf{G}(S) = T$. Let $\Phi = \{\phi_i : T \to \mathbb{R}\}$ be a partition of unity subordinate to $\mathbf{G}(\mathcal{O})$. Fix a $\phi \in \Phi$, and note that if $\phi(\mathbf{y}) = 0$ for all $\mathbf{y} \notin \mathbf{G}(U)$, then since $\mathbf{G}$ is a diffeomorphism, $[(\phi f)\circ \mathbf{G}](\mathbf{x}) = \phi(\mathbf{G}(\mathbf{x}))f(\mathbf{G}(\mathbf{x})) = 0$ for all $\mathbf{x} \notin U$. This in turn implies that

$$\int_{\mathbf{G}(S)} \phi f = \int_{\mathbf{G}(U)} \phi f = \int_U [(\phi f)\circ \mathbf{G}]\,|\det D\mathbf{G}| = \int_S [(\phi f)\circ \mathbf{G}]\,|\det D\mathbf{G}|.$$

Thus by Proposition 4.47 we know

$$\int_{\mathbf{G}(S)} f = \sum_{\phi\in\Phi} \int_{\mathbf{G}(S)} \phi f = \sum_{\phi\in\Phi}\int_A [(\phi f)\circ \mathbf{G}]\,|\det D\mathbf{G}| = \int_A (f\circ \mathbf{G})|\det D\mathbf{G}|.$$

Claim 2: If the theorem is true when $f(\mathbf{x}) \equiv 1$, then the proof is true in general.

If the theorem is true when $f \equiv 1$, then it's true for all constant functions since the integral is a linear operator. Let R' be a rectangle in $\mathbf{G}(S)$ with partition P, and R denote an arbitrary subrectangle, so that

$$\begin{aligned}
L(f,P) &= \sum_R \left[\inf_{\mathbf{x}\in R} f(\mathbf{x})\right] V(R) = \sum_R \int_{R^{\text{int}}} \left[\inf_{\mathbf{x}\in R} f(\mathbf{x})\right] \\
&= \sum_R \int_{\mathbf{G}^{-1}(R^{\text{int}})} \left(\left[\inf_{\mathbf{x}\in R} f(\mathbf{x})\right]\circ \mathbf{G}\right) |\det D\mathbf{G}| && \text{by assumption} \\
&\leq \sum_R \int_{\mathbf{G}^{-1}(R^{\text{int}})} (f\circ \mathbf{g})\,|\det D\mathbf{G}| \leq \int_{\mathbf{G}^{-1}(R')} (f\circ \mathbf{G})\,|\det D\mathbf{G}|.
\end{aligned}$$

Here we've applied our assumption to the constant function $\inf_R f(\mathbf{x})$. Precisely the same argument but with the supremum instead gives

$$U(f,P) \geq \int_{\mathbf{G}^{-1}(R')} (f\circ \mathbf{G})\,|\det D\mathbf{G}|,$$

which we combine with the first result to conclude that

$$L(f,P) \leq \int_{\mathbf{G}^{-1}(R')} (f\circ \mathbf{G})\,|\det D\mathbf{G}| \leq U(f,P),$$

showing that

$$\int_{R'} f = \int_{\mathbf{G}^{-1}(R')} (f\circ\mathbf{G})|\det D\mathbf{G}|.$$

Combined with Claim 1, we've now proven Claim 2.

Claim 3: If the theorem is true for diffeomorphisms $\mathbf{G}: S\to W$ and $\mathbf{H}: W\to T$, then the theorem holds true for the combined diffeomorphism $\mathbf{H}\circ\mathbf{G}: S\to T$.

This is pretty straightforward, effectively amounting to the fact that the determinant is multiplicative on matrix multiplication. Indeed,

$$\begin{aligned}\int_{\mathbf{H}(\mathbf{G}(S))} f = \int_{\mathbf{G}(S)} (f\circ\mathbf{H})|\det\mathbf{H}| &= \int_S [(f\circ\mathbf{H})\circ\mathbf{G}]\,[|\det D\mathbf{H}|\circ\mathbf{G}]\,|\det D\mathbf{G}| \\ &= \int_S [f\circ(\mathbf{H}\circ\mathbf{G})]\,|\det D(\mathbf{H}\circ\mathbf{G})|.\end{aligned}$$

We've used associativity of function composition several times. The first is to write $(f\circ\mathbf{G})\circ\mathbf{H} = f\circ(\mathbf{G}\circ\mathbf{H})$, but we've also used it to identify $|\det D\mathbf{H}|\circ\mathbf{G} = |\det(D\mathbf{H}\circ\mathbf{G})|$. In the last equality we've used the Chain Rule and the fact that determinants respect multiplication to write

$$\det[D(\mathbf{H}\circ\mathbf{G})] = \det[(D\mathbf{H}\circ\mathbf{G})\,D\mathbf{G}] = \det[D\mathbf{H}\circ\mathbf{G}]\det[D\mathbf{G}].$$

Claim 4: The theorem is true if $\mathbf{G}$ is an invertible linear transformation.

By Claim 2 it suffices to show the result for $f\equiv 1$, and by Claim 1 we can show the result on any open set U; that is,

$$\int_{\mathbf{G}(U)} 1 = \int_U |\det D\mathbf{G}|.$$

However, $D\mathbf{G} = \mathbf{G}$ since $\mathbf{G}$ is a linear transformation, hence this amounts to showing $\mu(\mathbf{G}(U)) = |\det\mathbf{G}|\mu(U)$, which is precisely Exercise 4-7.

Claim 5: The theorem holds for diffeomorphisms $\mathbf{H},\mathbf{K}: S\to T$ of the following form:

$$\mathbf{H}(\mathbf{x}) = (H_1(\mathbf{x}), H_2(\mathbf{x}),\ldots,H_{n-1}(\mathbf{x}), x_n) \quad\text{and}\quad \mathbf{K}(\mathbf{x}) = (x_1,x_2,\ldots,x_{n-1},K_n(\mathbf{x})).$$

Fix a rectangle $\tilde{R} = R\times[a,b]$ in S, with coordinates $(\mathbf{s},t)$. For a fixed value of $t\in[a,b]$ define the map $\tilde{\mathbf{H}}_t : R\subseteq\mathbb{R}^{n-1}\to\mathbb{R}^{n-1}$ by $\tilde{\mathbf{H}}(\mathbf{s}) = (H_1(\mathbf{s}),\ldots,H_{n-1}(\mathbf{s}))$. Note that

$$\mathbf{H}(\tilde{R}) = \left\{(\mathbf{s},t) : t\in[a,b], \mathbf{s}\in\tilde{\mathbf{H}}_t(R)\right\}.$$

Hence if $\mathbf{x} = (\mathbf{s},t)$ then

$$\det D\mathbf{H}(\mathbf{x}) = \begin{bmatrix} \partial_1 H_1(\mathbf{x}) & \partial_2 H_1(\mathbf{x}) & \cdots & \partial_{n-1}H_1(\mathbf{x}) & \partial_n H_1(\mathbf{x}) \\ \partial_1 H_2(\mathbf{x}) & \partial_2 H_2(\mathbf{x}) & \cdots & \partial_{n-1}H_2(\mathbf{x}) & \partial_n H_2(\mathbf{x}) \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \partial_1 H_{n-1}(\mathbf{x}) & \partial_2 H_{n-1}(\mathbf{x}) & \cdots & \partial_{n-1}H_{n-1}(\mathbf{x}) & \partial_n H_{n-1}(\mathbf{x}) \\ 0 & 0 & \cdots & 0 & 1\end{bmatrix} = \det\tilde{\mathbf{H}}_t(\mathbf{s}).$$

By Claim 4, it suffices to show the result when $f \equiv 1$, and by Fubini's Theorem we have

$$\begin{aligned}\int_{\tilde{R}} |\det D\mathbf{H}| &= \int_{[a,b]} \left[\int_R |\det D\mathbf{H}(\mathbf{s},t)|\,\mathrm{d}\mathbf{s}\right]\,\mathrm{d}t = \int_{[a,b]} \left[\int_R |\det D\tilde{\mathbf{H}}_t(\mathbf{s})|\,\mathrm{d}\mathbf{s}\right]\,\mathrm{d}t \\ &= \int_{[a,b]} \left[\int_{\tilde{\mathbf{H}}_t(R)} \mathrm{d}\mathbf{s}\right]\,\mathrm{d}t = \int_{\mathbf{H}(\tilde{R})} 1.\end{aligned}$$

The proof for $\mathbf{K}$ is nearly identical, so is omitted.

The Proof: We proceed by induction on the number of variables. When $n = 1$, the Method of Substitution establishes the base case. Therefore assume that the result holds for $n-1$ variables. By Claim 1, it suffices to show that for any $\mathbf{x}_0 \in S$, there is an open neighbourhood U of $\mathbf{x}_0$ on which the theorem holds. By Claims 3, 4, and 5, it suffices to show that we can write $\mathbf{G} = T \circ \mathbf{H} \circ \mathbf{K}$ on U, where T is linear and $\mathbf{H}, \mathbf{K}$ are as in Claim 5.

Let $\mathbf{x}_0$ be given. As $\mathbf{G}$ is a diffeomorphism, $\det D\mathbf{G}(\mathbf{x}_0) \neq 0$, so we can define a map $\tilde{\mathbf{G}}(\mathbf{x}) = D\mathbf{G}(\mathbf{x}_0)^{-1} \circ \mathbf{G}$. Note that

$$D\tilde{\mathbf{G}}(\mathbf{x}_0) = D\mathbf{G}(\mathbf{x}_0)^{-1}(\mathbf{G}(\mathbf{x}_0))D\mathbf{G}(\mathbf{x}_0) = D\mathbf{G}(\mathbf{x}_0)^{-1}D\mathbf{G}(\mathbf{x}_0) = I.$$

Define $\mathbf{H}(\mathbf{x}) = (\tilde{G}_1(\mathbf{x}), \ldots, \tilde{G}_{n-1}(\mathbf{x}), x^n)$ where $\tilde{G}_i$ is the ith component of $\tilde{\mathbf{G}}$. Since $\det \mathbf{G}(\mathbf{x}_0) = I$ we quickly see that $\det \mathbf{H}(\mathbf{x}) = I$ as well. By the Inverse Function Theorem, there are neighbourhoods U of $\mathbf{x}_0$ and V of $\mathbf{H}(\mathbf{x}_0)$ such that $\mathbf{H}$ is a diffeomorphism from U to V. In turn, define $\mathbf{K}(\mathbf{x}) = (x_1, \ldots, x_{n-1}, \tilde{G}_n(\mathbf{H}^{-1}(\mathbf{x})))$, which is similarly a diffeomorphism. Thus $\tilde{\mathbf{G}} = \mathbf{K} \circ \mathbf{H}$, and so

$$\mathbf{G} = D\mathbf{G}(\mathbf{x}_0) \circ D\mathbf{G}(\mathbf{x}_0)^{-1} \circ \mathbf{G} = D\mathbf{G}(\mathbf{x}_0) \circ \tilde{\mathbf{G}} = D\mathbf{G}(\mathbf{x}_0) \circ \mathbf{K} \circ \mathbf{H}$$

on the neighbourhood U, which is what we wanted to show. □

Example 4.52

Let $(u, v) = \mathbf{G}(r, \theta) = (e^r \cos(\theta), e^r \sin(\theta))$. 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

$$\det D\mathbf{G}(r,\theta) = \det \begin{bmatrix} e^r\cos(\theta) & e^r\sin(\theta) \\ -e^r\sin(\theta) & e^r\cos(\theta)\end{bmatrix} = e^{2r},$$

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

$$\mathrm{d}r\,\mathrm{d}\theta = \frac{1}{e^{2r}}\,\mathrm{d}u\,\mathrm{d}v = \frac{\mathrm{d}u\,\mathrm{d}v}{u^2+v^2}.$$ ■

Example 4.52 illustrates an important point. At times we will employ an inverse substitution, though solving for the inverse coordinate transformation might be tricky. If $\mathbf{G}$ is a diffeomorphism and $\mathbf{y} = \mathbf{G}(\mathbf{x})$, then $\mathbf{x} = \mathbf{G}^{-1}(\mathbf{y})$. By the Inverse Function Theorem, $D\mathbf{G}^{-1}(\mathbf{y}) = [D\mathbf{G}(\mathbf{x})]^{-1}$, so the Jacobian in turn satisfies

$$\left|\det D\mathbf{G}^{-1}(\mathbf{y})\right| = \left|\det\left[D\mathbf{G}(\mathbf{x})\right]^{-1}\right| = \frac{1}{|\det D\mathbf{G}(\mathbf{x})|} = \frac{1}{|\det D\mathbf{G}(\mathbf{G}^{-1}(\mathbf{y}))|}.$$

This is often easier to compute that the inverse transform.

While the Method of Substitution was often used to simplify the integrand, Change of Variables can also be used to simplify the domain of integration.

Example 4.53

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 $\iint_S (x^2 + y^2)\,\mathrm{d}A$.

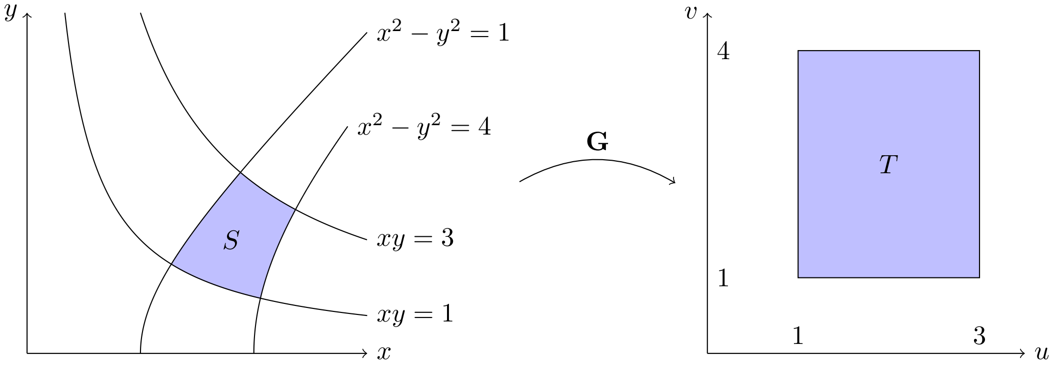


Figure 4.18: The Change of Variables for Example 4.53. The region S looks like a skewed rectangle, and the map $\mathbf{G}$ defined in the solution converts this into a proper rectangle.

Solution. Whether we take S to be the open or closed region bounded by these curves does not matter, as the difference is a set of measure zero. For convenience, let's assume the region is open. It is plotted in Figure 4.18. The region looks like a skewed rectangle, so our goal is to straighten these lines out. By making $u = xy$ and $v = x^2 - y^2$, S will transform into the rectangle $T = (1, 3) \times (1, 4)$. Thus set $\mathbf{G} : S \to T$ and $(u, v) = \mathbf{G}(x, y) = (xy, x^2 - y^2)$, so that

$$|\det D\mathbf{G}(x, y)| = \left|\det\begin{bmatrix} y & x \\ 2x & -2y \end{bmatrix}\right| = 2(x^2 + y^2).$$

Thus $\mathrm{d}u\,\mathrm{d}v = \frac{1}{2}(x^2 + y^2)\,\mathrm{d}x\,\mathrm{d}y$ and our integral becomes

$$\iint_T (x^2 + y^2)\,\mathrm{d}x\,\mathrm{d}y = \frac{1}{2}\int_S \mathrm{d}u\,\mathrm{d}v = 3. \qquad \blacksquare$$

Example 4.54

Let $R = \left\{(x, y) \in \mathbb{R}^2 : 1 \leq x^2 + y^2 \leq 3\right\}$. Evaluate $\displaystyle\iint_R e^{x^2+y^2}\,\mathrm{d}A$.

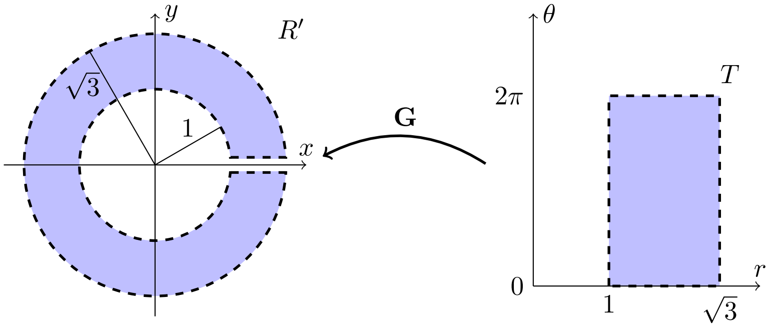


Figure 4.19: The diffeomorphism $\mathbf{G} : T \to R'$.

Solution. Our region R is the annulus of inner radius 1 and outer radius $\sqrt{3}$, suggesting polar coordinates would make a good substitution. Note that as R is not open, we cannot write down a diffeomorphism directly. Instead, set $(x, y) = \mathbf{G}(r, \theta) = (r\cos(\theta), r\sin(\theta))$ where $\mathbf{G} : T \to R'$ is a diffeomorphism, R' is the open annulus with radii 1 and $\sqrt{3}$ with the positive x-axis excised, and T is the rectangle $(1, \sqrt{3}) \times (0, 2\pi)$. This is illustrated in Figure 4.19. Note that this is an inverse substitution. We already know that $\mathrm{d}x\,\mathrm{d}y = r\,\mathrm{d}r\,\mathrm{d}\theta$, so applying the Change of Variables Theorem we get

$$\begin{aligned}\iint_{R'} e^{x^2+y^2}\,\mathrm{d}A_{\text{Cart}} &= \iint_{\mathbf{G}^{-1}(R')} e^{r^2\cos^2(\theta)+r^2\sin^2(\theta)} r\,\mathrm{d}A_{\text{pol}} = \iint_T re^{r^2}\,\mathrm{d}A_{\text{pol}} \\ &= \int_1^{\sqrt{3}}\int_0^{2\pi} e^{r^2} r\,\mathrm{d}r\,\mathrm{d}\theta = 2\pi\int_1^{\sqrt{3}} re^{r^2}\,\mathrm{d}r\ \mathrm{d}\theta \\ &= \pi\left[e^{r^2}\right]_{r=1}^{\sqrt{3}} = \pi\left[e^3 - e\right].\end{aligned}$$

■

Example 4.55

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, which looks like a long cylinder with slightly rounded caps. With some thinking, or by staring at a picture, we see that the volume is symmetric about reflection in the xy-plane, so we need only find the volume bounded by the upper-half hemisphere and the cylinder. Let S be the interior of this region with the plane $z \geq 0$ excised (Figure 4.20). Using cylindrical coordinates $(x, y, z) = \mathbf{G}(r, \theta, z) = (r\cos(\theta), r\sin(\theta), z)$, we know $|\det D\mathbf{G}(r, \theta, z)| = r$, and that S is the image of a z-simple set T, which we can write as

$$T = \left\{(r, \theta, z) : r \in (0, 1), \theta \in (0, 2\pi), 0 \leq z \leq \sqrt{4 - r^2}\right\}.$$

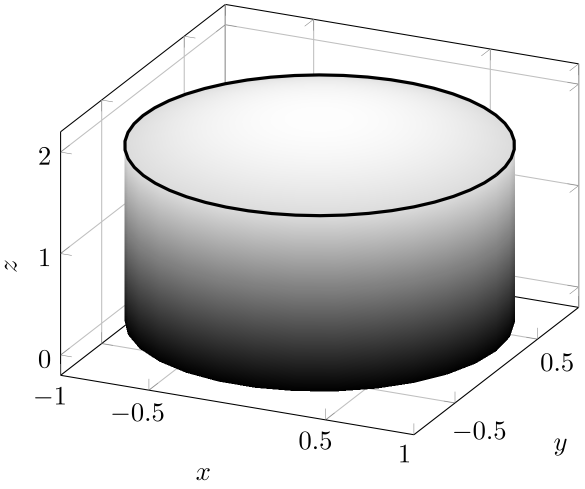


Figure 4.20: The intersection of the upper half hemisphere $x^2+y^2+z^2=4$ with $z\geq 0$, and the cylinder $x^2+y^2=1$. It looks like a cylinder with a small rounded cap.

Applying the Change of Variables Theorem, we thus get

$$\begin{aligned}\iiint_S \mathrm{d}V_{\text{Cart}} &= \iiint_{\mathbf{G}^{-1}(S)} r\,\mathrm{d}V_{\text{cyl}} = \int_0^{2\pi}\int_0^1\int_0^{\sqrt{4-r^2}} r\,\mathrm{d}z\,\mathrm{d}r\,\mathrm{d}\theta \\ &= \int_0^{2\pi}\int_0^1 r\sqrt{4-r^2}\,\mathrm{d}r\,\mathrm{d}\theta = \int_0^{2\pi}\left[-\frac{1}{3}(4-r^2)^{3/2}\right]_{r=0}^1 \mathrm{d}\theta \\ &= \frac{2\pi}{3}\left[8-3\sqrt{3}\right].\end{aligned}$$

Hence the fully bounded area is $2V(S)=\frac{4\pi}{3}\left[8-3\sqrt{3}\right]$. ■

4.7 Exercises

4-1. Show that the set $\{1/n : n\in\mathbb{N}\}\subseteq\mathbb{R}$ is Jordan measurable in $\mathbb{R}$, and determine its measure.

4-2. Here we generalize Exercise-4-1. Let $(a_n)_{n=1}^\infty$ be a convergent sequence, and let $S=\{a_n : n\in\mathbb{N}\}$. Show that this set is Jordan measurable, and determine its measure. *Hint*: Try to emulate Proposition 4.8, but note that you cannot use infinitely many intervals. Instead, put all but finitely many points into a single interval.

4-3. Let $S\subseteq\mathbb{R}$. Show that the following are equivalent:

(a) S is Jordan measurable,

(b) For every $\epsilon>0$, there exist poly-rectangles A and B such that $A\subseteq S\subseteq B$ and $\mu(B\setminus A)<\epsilon$,

4-4. True or False: Every bounded, open set $S\subseteq\mathbb{R}^n$ is Jordan measurable? Prove the result or provide a counter example.

4-5. **Monotonicity of measure:** Show that if S_1, S_2 are both Jordan measurable with $S_1 \subseteq S_2$, then $\mu(S_1) \leq \mu(S_2)$.

4-6. **Translation invariance of measure:** Show that if S is Jordan measurable and $\mathbf{a} \in \mathbb{R}^n$ then $\mu(S + \mathbf{a}) = \mu(S)$. *Hint:* Show that if P is a poly-rectangle which covers S, then $P + \mathbf{a}$ covers $S + \mathbf{a}$.

4-7. **Scaling of measure:** Show that if $S \subseteq \mathbb{R}^n$ is Jordan measurable and $T : \mathbb{R}^n \to \mathbb{R}^n$ is a linear transformation, then $\mu(T(S)) = |\det T|\mu(S)$. *Hint:* Show the result is true for rectangles first. You may use the fact that the volume of a parallelepiped with vertices $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ in $\mathbb{R}^n$ is $\det[\mathbf{v}_1 | \cdots | \mathbf{v}_n]$.

4-8. For two sets $A, B \subseteq \mathbb{R}$, we define the *distance* between A, B to be

$$d(A, B) = \inf \{|x - y| : x \in A, y \in B\}.$$

Let S_1, S_2 be measurable with $d(S_1, S_2) > 0$. Show that $\mu(S_1 \cup S_2) = \mu(S_1) + \mu(S_2)$. *Hint:* One direction is given by Proposition 4.9. To show the opposite direction, argue that you can find a poly-rectangle which covers $S_1 \cup S_2$ and which consists of intervals of length less than $d(S_1, S_2)$.

4-9. Change the definition of a poly-rectangle to allow a countably infinite number of open intervals. The *Lebesgue outer measure* of a set $S \subseteq S$ is

$$m^*(S) = \inf \{|I| : I \text{ is a poly-rectangle such that } S \subseteq I\}$$

while its *Lebesgue inner measure* is

$$m_*(S) = \sup \{|I| : I \text{ is a poly-rectangle such that } I \subseteq S\}.$$

A set S is *Lebesgue measurable* if $m^*(S) = m_*(S)$, in which case we write its measure as $m(S) = m^*(S) = m_*(S)$. Let $J(\mathbb{R})$ be the Jordan measurable subsets of $\mathbb{R}$, and $L(\mathbb{R})$ be the Lebesgue measurable subsets.

(a) Show that $J(\mathbb{R}) \subseteq L(\mathbb{R})$.

(b) Show that the above inclusion is strict; that is, there is a Lebesgue measurable set which is not Jordan measurable. *Hint:* Example 4.10.

Note: The Lebesgue measure is more often used in advanced analysis, but does not play as nicely with the Riemann integral, as seen in Example 4.23. The reason is that Riemann integrals require finite partitions, and the Jordan measure also deals with finite poly-rectangles. On the other hand, some theorems are strengthened by using the Lebesgue measure. For example, Theorem 4.28 will become an 'if and only if' theorem if the Lebesgue measure is used instead of the Jordan measure.

4-10. For some fixed $k \in \mathbb{N}$, define

$$\mathcal{C}_k = \left\{ \left[\frac{p_1 - 1}{k}, \frac{p_1}{k}\right] \times \cdots \times \left[\frac{p_n - 1}{k}, \frac{p_n}{k}\right] : p_i \in \mathbb{Z}, i = 1, \ldots, n \right\},$$

which are the collection of all possible cubes with rational vertices and side-length $1/k$ in $\mathbb{R}^n$. For any set $S \subseteq \mathbb{R}^n$, we can classify each element $C \in \mathcal{C}_n$ as an interior or exterior cube if C

lies entirely in the interior of S or in the interior of S^c, and a boundary cube otherwise. Let $\underline{\mu}(S,k)$ denote the sum of the volumes of the interior cubes of side length $1/k$, and $\overline{\mu}(S,k)$ denote the sums of the interior and boundary cubes of side length $1/k$.

(a) Show that if $R \subseteq \mathbb{R}^n$ is a rectangle, then

$$\lim_{k\to\infty} \underline{\mu}(R,k) = \mu(R) = \lim_{k\to\infty} \overline{\mu}(R,k).$$

(b) Show that if $S \subseteq \mathbb{R}^n$ is a bounded set, then

$$\lim_{k\to\infty} \underline{\mu}(S,k) = \mu_*(S) \quad \text{and} \quad \lim_{k\to\infty} \overline{\mu}(S,k) = \mu^*(S).$$

(c) Show that if $S \subseteq \mathbb{R}^n$ is a bounded set, then

$$\mu^*(S) = \mu^*(\overline{S}) = \mu_*(S) + \mu^*(\partial S).$$

(d) Conclude that a bounded set $S \subseteq \mathbb{R}^n$ is Jordan measurable if and only if $\mu(\partial S) = 0$.

4-11. Let $\mathbf{G} : S \to T$ be a diffeomorphism.

(a) Show that if $K \subseteq S$ is compact and measure zero, then $\mathbf{G}(K)$ is compact and measure zero.

(b) Show that if $U \subseteq S$ is measurable, then $\mathbf{G}(U)$ is measurable.

(c) Suppose $f : T \to \mathbb{R}$ is continuous everywhere except possibly on a compact set of measure zero. Show that $f \circ \mathbf{G}$ is also continuous everywhere except possibly a set of measure zero.

4-12. Prove Proposition 4.7.

4-13. Prove Proposition 4.13.

4-14. Define the *diameter* of a set $S \subseteq \mathbb{R}^n$ to be

$$\operatorname{diam}(S) = \sup_{\mathbf{x},\mathbf{y}\in S} \|\mathbf{x} - \mathbf{y}\|.$$

(a) If R is a cube with side length ℓ in $\mathbb{R}^n$, show that $\operatorname{diam}(R) = s\sqrt{n}$.

(b) Show that for any $\delta > 0$, any rectangle $R \subseteq \mathbb{R}^n$ can be decomposed into non-overlapping closed rectangles of

4-15. Let $\mathcal{R}_{[a,b]}$ denote the collection of Riemann integrable functions on the interval $[a,b]$.

(a) Show that $\mathcal{R}_{[a,b]}$ is a vector space.

(b) Show that $\mathcal{R}_{[a,b]}$ is infinite dimensional.

(c) Show that $\int : \mathcal{R}_{[a,b]} \to \mathcal{R}_{[a,b]}, f \mapsto \int f$ is a linear operator.

4-16. Let $f : [a,b] \to \mathbb{R}$ be integrable and $P = \{x_0, \ldots, x_n\}$ be a fixed partition of $[a,b]$. Show that for any $\epsilon > 0$ there is a Riemann sum $S(f,P)$ such that $U(f,P) - S(f,P) < \epsilon$. Argue that similarly, there is a Riemann sum $S(f,P)$ such that $S(f,P) - L(f,P) < \epsilon$.

4-17. Prove Theorem 4.21 as follows:

(a) Show that (1) $\Leftrightarrow$ (4). (1) $\Rightarrow$ (4) is straightforward. For (4) $\Rightarrow$ (1), construct a Cauchy sequence of real numbers.

(b) Prove that (2) $\Leftrightarrow$ (3). Both directions amount to arguments about supremum and infimum.

(c) Prove that (1) $\Rightarrow$ (3). You'll need Exercise 4-16.

(d) Prove that (3) $\Rightarrow$ (1). This is the hardest proof. Fix an $\epsilon > 0$ and choose P_ϵ satisfying (3). If P is any other partition, break P into those subintervals Σ_1 which contain endpoints of P_ϵ and those which do not, Σ_2. Bound the contribution from Σ_1 by choosing $\ell(P)$ to be sufficiently small.

4-18. Show that integrability is inherited. Namely, if $f : [a,b] \to \mathbb{R}$ is integrable and $[c,d] \subseteq [a,b]$, then $f|_{[c,d]} : [c,d] \to \mathbb{R}$ defined by $x \mapsto f(x)$ is also integrable.

4-19. Prove that integration is additive on domains. Namely, if $f : [a,c] \to \mathbb{R}$ is integrable on $[a,b]$ and on $[b,c]$, then f is integrable on $[a,c]$ and

$$\int_{[a,c]} f = \int_{[a,b]} f + \int_{[b,c]} f.$$

4-20. Show that the integral is monotone. Namely, if $f, g : [a,b] \to \mathbb{R}$ are integrable and satisfy $f(x) \leq g(x)$ for all $x \in [a,b]$, then

$$\int f \leq \int g.$$

4-21. Show that the integral is subnormal. Namely, if $f : [a,b] \to \mathbb{R}$ is integrable, then $|f|$ is also integrable and

$$\left|\int f\right| \leq \int |f|.$$

4-22. Let $R \subseteq \mathbb{R}^n$ be a rectangle, and let $f, g : R \to \mathbb{R}$ be functions. Suppose that for every partition P of R, there exists a partition Q such that

$$L(f,P) \leq L(g,Q) \leq U(g,Q) \leq U(f,P).$$

Show that if f is integrable, then g is integrable, and $\int f = \int g$.

4-23. Determine the integral of each function on the specified rectangle:

(a) $f(x,y) = e^x \cos(y)$ on $R = \{(x,y) : 0 \leq x \leq, \pi/2 \leq y \leq \pi\}$,

(b) $f(x,y) = e^{x-y}$ on $R = \{(x,y) : 0 \leq x \leq 1, -2 \leq y \leq -1\}$

(c) $f(x,y) = x^2 6 - 3xy^2$ on $R = \{(x,y) : 1 \leq x \leq 2, -1 \leq y \leq 1\}$,

(d) $f(x,y) = 1/(x+y)$ on $R = \{(x,y) : 0 \leq x \leq 1, 1 \leq y \leq 2\}$,

(e) $f(x,y) = y\cos(xy)$ on $R = \{(x,y) : 0 \leq x \leq 1, 0 \leq\leq \pi\}$,

(f) $f(x,y) = e^y \sin(xy^{-1})$ on $R = \{(x,y) : -\pi/2 \leq x \leq \pi/2, 1 \leq y \leq 2\}$,

(g) $f(x,y) = \sin(x+y)$ on $R = \{(x,y) : 0 \leq x, y \leq \pi/2\}$,

(h) $f(x,y) = 4xy\sqrt{x^2+y^2}$ on $R = \{(x,y) : 0 \leq x \leq 3, 0 \leq y \leq 1\}$.

4-24. In this question we will generalize the notions of even and odd, and show multivariable analogs of single variable results. Let $r > 0$ and set $R = \left\{(x,y) \in \mathbb{R}^2 : -r \leq x, y \leq r\right\}$.

(a) Let f be an integrable function such that $f(-x,-y) = -f(x,y)$. Show that

$$\iint_R f(x,y)\,\mathrm{d}A = 0.$$

(b) Let f be an integrable function such that $f(x,-y) = -f(x,y)$. Show that

$$\iint_R f(x,y)\,\mathrm{d}A = 0.$$

4-25. Let R be the region in the xy-plane bounded by the curves $y = 2x$ and $y = x^2$. Determine the area bounded by R and the paraboloid $z = x^2 + y^2$.

4-26. Determine the area given by the intersection of the two cylinders $x^2+y^2 = r^2$ and $y^2+z^2 = r^2$ for any $r > 0$.

4-27. In each case, determine $\iint_S f(x,y)\,\mathrm{d}A$ where f and S are specified:

(a) $f(x,y) = 1 + x + y$, S is the region in the first quadrant bounded by $x = 1$ and $y = e^x$,

(b) $f(x,y) = (x-y)^2$, S is the region bounded between x^2 and x^3,

(c) $f(x,y) = y$, $S = \left\{(x,y) : x^2+y^2 \leq 1\right\} \cap \left\{(x,y) : x^2 + (y-1)^2 \leq 1\right\}$,

(d) $f(x,y) = x^2y^2$, $S = \left\{(x,y) : -y^2 \leq x \leq y^2, 0 \leq y \leq 1\right\}$,

(e) $f(x,y) = xy$, S the area bounded by the lines $y = x - 1$ and $y^2 = 2x + 6$,

(f) $f(x,y) = 1 + x$, S is the area bounded between $x + y = 0$ and $y + x^2 = 1$,

(g) $f(x,y) = \frac{\sin(y)}{y}$, S is the area bounded between $y = x$ and $y = \sqrt{x}$.

4-28. Find the volume of the solid bounded by the surfaces $z = 3x^2 + 3y^2$ and $x^2 + y^2 + z = 4$.

4-29. Let $f : \mathbb{R}^2 \to \mathbb{R}$ be an integrable function, and define

$$G(x) = \int_a^x \int_a^s f(s,t)\,\mathrm{d}t\,\mathrm{d}s.$$

Show that one can equivalently write

$$G(x) = \int_a^x \int_t^x f(s,t)\,\mathrm{d}s\,\mathrm{d}t.$$

4-30. Let $R = [0,1] \times [0,1]$ and consider the function $f : R \to \mathbb{R}$ given by $f(x,y) = \dfrac{x^2-y^2}{(x^2+y^2)^2}$.

(a) Show that

$$\iint_R f(x,y)\,\mathrm{d}x\,\mathrm{d}y \neq \iint_R f(x,y)\,\mathrm{d}y\,\mathrm{d}x.$$

(b) Is this a contradiction to Fubini's Theorem? Why or why not?

4-31. (a) Let $\alpha \in \mathbb{R}$ be an arbitrary non-zero constant. Compute
$$\int \frac{x-\alpha}{(x+\alpha)^3}\,\mathrm{d}x.$$
Hint: To integrate $x/(x+\alpha)^3$ make the substitution $u = x+\alpha$.

(b) Let R be the rectangle $R = [0,1]\times[0,1]$ and compute the iterated integrals
$$\iint_R \frac{x-y}{(x+y)^3}\,\mathrm{d}x\,\mathrm{d}y, \qquad \iint_R \frac{x-y}{(x+y)^3}\,\mathrm{d}y\,\mathrm{d}x.$$
Notice that the order of integration is changed!

(c) You should have found in part (c) that the integrals did not agree. Explain why this is not a contradiction to Fubini's theorem.

4-32. Evaluate the integral of the following functions on the specified domain:

(a) $f(x,y,z) = y$ over the region bounded by the planes $x = 0$, $y = 0$, $z = 0$, and $2x+2y+z = 4$,

(b) $f(x,y,z) = z$ over the region bounded by $y^2+z^2 = 9$, $x = 0$, $y = 3x$ and $z = 0$ in the first octant.

(c) $f(x,y,z) = 1$ over the region bounded by $y = x^2$, $z = 0$ and $y+z = 1$.

4-33. Compute the given interval on the given domain:

(a) $\displaystyle\iint_R \left[2 + x^2y^3 - y^2\sin(x)\right]\,\mathrm{d}A$ where $R = \{(x,y) : |x|+|y| \leq 1\}$,

(b) $\displaystyle\iint_R \left[ax^2 + by^3 + \sqrt{a^2-x^2}\right]\,\mathrm{d}A$ where $R = \{(x,y) : |x| \leq a, |y| \leq b\}$.

4-34. Evaluate the following triple integrals on the given regions:

(a) $f(x,y,z) = z$ where S is the region bounded by $y^2+z^2 = 9$ and the planes $x = 0, y = 3x$ and $z = 0$, in the first octant,

(b) $f(x,y,z) = 1$ where S is the region bounded by $y = x^2$ and the planes $z - 0, z = 4$, and $y = 9$,

(c) $f(x,y,z) = z$ where S is portion of $x^2+y^2+z^2 \leq 4$ in the first octant.

4-35. Show the following results hold for extended integrals. If $U \subseteq \mathbb{R}^n$ is an open set and $f : U \to \mathbb{R}$ is continuous and extended integrable, then

(a) **Linearity:** If $g : U \to \mathbb{R}$ is continuous and extended integrable, then for any $c \in \mathbb{R}$, $f+g$ and cf are extended integrable and
$$⨍_U [f+g] = ⨍_U f + ⨍_U g \quad\text{and}\quad ⨍_U [cf] = c ⨍_U f.$$

(b) **Additivity of Domain:** If $V \subseteq \mathbb{R}^n$ is another open set and f can be extended to a continuous function on B, then
$$⨍_{U\cup V} f = ⨍_U f + ⨍_V f - ⨍_{U\cap V} f.$$

(c) **Monotonicity of Integral:** If $g : U \to \mathbb{R}$ is continuous and $f(x) \leq g(x)$ for all $x \in [a,b]$ then

$$\fint_U f \leq \fint_U g.$$

Notably, $\left|\fint_U f\right| \leq \fint_U |f|$.

(d) **Monotonicity of Domain:** If $V \subseteq U$ is open and f is non-negative, then

$$\fint_V f \leq \fint_U f.$$

The proofs are left to you as an exercise.

4-36. Let $U \subseteq \mathbb{R}^n$ and $K \subseteq U$ be an compact subset. Suppose $f : A \to \mathbb{R}$ is continuous such that $f(\mathbf{x}) = 0$ if $\mathbf{x} \notin K$. Show that

$$\int_k f = \fint_A f.$$

4-37. Determine the Jacobian of the following transformations. Whenever possible, write the infinitesimal area/volume element in terms of one another:

(a) $(x,y) = (e^\xi, \eta^3)$,

(b) $(x,y) = (5u - 2v, u + v)$,

(c) $(x,y) = (\sin(u^2 v), \cos(v^2 u))$,

(d) $(x,y,z) = (v + w^2, w + u^2, u + v^2)$,

(e) $(x,y,z) = (u^3 - v^2, u^3 + v^2, u^3 + v^2 + w)$.

4-38. The transformation from polar, cylindrical, and sphereical coordinates into Cartesian coordinates is given by

$$\begin{aligned}
(x,y) &= \mathbf{G}_{\text{pol}}(r,\theta) = (r\cos(\theta), r\sin(\theta)) \\
(x,y,z) &= \mathbf{G}_{\text{cyl}}(r,\theta,z) = (r\cos(\theta), r\sin(\theta), z) \\
(x,y,z) &= \mathbf{G}_{\text{sph}}(\rho,\phi,\theta) = (\rho\cos(\phi)\sin(\theta), \rho\sin(\phi)\sin(\theta), \rho\cos(\theta)).
\end{aligned}$$

Determine the cooresponding inverse transformations.

4-39. Let R be the region bounded by the curves $y = x^2$, $4y = x^2$, $xy = 1$ and $xy = 2$. Compute the integral

$$\iint_R x^2 y^2 \, \mathrm{d}A.$$

4-40. Determine $\displaystyle\iint_S \frac{(x+y)^4}{(x-y)^5} \, \mathrm{d}A$ where $S = \{-1 \leq x + y \leq 1,\ 1 \leq x - y \leq 3\}$.

4-41. Compute $\displaystyle\iint_R (4x+8y) \, \mathrm{d}A$ where R is the quadrilateral with endpoints $(-1,3), (1,-3), (3,-1), (-3,1)$.

4-42. Compute $\displaystyle\iint_R \sin(9x^2 + 4y^2) \, \mathrm{d}A$ where R is the circle $9x^2 + 4y^2 = 36$.

4-43. Compute $\iint_R x^2 \, \mathrm{d}A$ where R is the region $a^2x^2 + b^2y^2 = c^2$, $a, b, c > 0$.

4-44. Let $R \subseteq \mathbb{R}^2$ be the region in the 1^{st} quadrant bounded by the curves $y = x^2$, $y = x^2/5$, $xy = 2$, and $xy = 4$.

(a) Define the variables $u = x^2/y$ and $v = xy$. Compute $\mathrm{d}x \, \mathrm{d}y$ in terms of $\mathrm{d}u \, \mathrm{d}v$.

(b) Compute the area of R by changing variables from (x, y) to (u, v).

4-45. (a) Let $R = [a, b] \times [c, d]$ be a rectangle and $f : R \subseteq \mathbb{R}^2 \to \mathbb{R}$ be a continuous function. Assume there exist functions $f_1, f_2 : \mathbb{R} \to \mathbb{R}$ such that for all $x, y \in R$, $f(x, y) = f_1(x)f_2(y)$. Show that

$$\iint_R f(x, y) \, \mathrm{d}A = \left[\int_{[a,b]} f_1(x) \, \mathrm{d}x \right] \left[\int_{[c,d]} f_2(y) \, \mathrm{d}y \right].$$

(b) It is known that the function $f(x) = e^{-x^2}$ does not have an elementary anti-derivative; however, this function is Riemann integrable on all of $\mathbb{R}$ (by using improper integrals). Compute

$$\int_{-\infty}^{\infty} e^{-x^2} dx.$$

Hint: Consider the function $e^{-x^2-y^2}$ on $\mathbb{R}^2$ and use part (b).

4-46. Look at Claim 5 in the proof of Theorem 4.50. Prove the result for the function $\mathbf{K}$ defined therein.

5 Vector Calculus

We now approach the classical study of integrating vector fields. This topic has a beautiful generalization using *differential forms*, but the study of differential forms is subtle and sophisticated. I will mention differential forms towards the end of these notes for those interested, but the classical treatment is still rich and interesting enough for our purposes.

5.1 Curves, Surfaces, and Manifolds

We begin with a small introduction to the idea of smooth curves, surfaces, and manifolds in general. Intuitively, an object is smooth if it contains no corners, such as a sphere. We want to reject something like a cube, whose vertices and edges form sharp boundaries. However, avoiding corners and edges is still not sufficient. For example, the *lemniscate* (Figure 5.1) forms a figure-8 pattern. When we draw the lemniscate, we can do it in 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.

The general notion of smoothness comes down to looking at tangent spaces. In one dimension we saw that the derivative was the slope of the tangent line, and we've discussed that differentiable functions $f : \mathbb{R}^n \to \mathbb{R}$ admit tangent hyperplanes. These spaces are all isomorphic to vector spaces, and a space is smooth if its tangent spaces all have the same dimension.

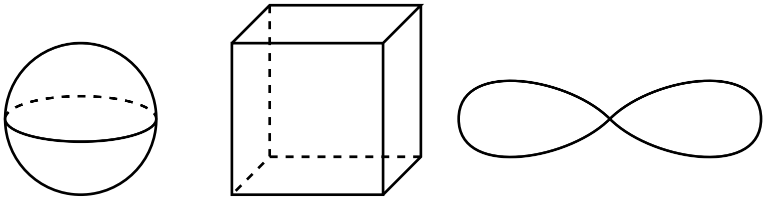


Figure 5.1: The sphere should be smooth, the cube should not be, but who knows about the lemniscate?

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.

However, it's often impractical to look at the dimension of every tangent space, so we define smooth manifolds in alternative ways, 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.

Definition 5.1

A set $M \subseteq \mathbb{R}^n$ is said to be a *smooth k-manifold* if around every point $\mathbf{p} \in M$ there is an open neighbourhood $V \subseteq M$ on which V is the graph of a C^1 function $\mathbf{f} : U \to \mathbb{R}^{n-k}$ for some $U \subseteq \mathbb{R}^k$.

Note that $V \subseteq M$ is open in the relative topology M inherits from the ambient space $\mathbb{R}^n$, hence $V = M \cap \tilde{V}$ for some open set $\tilde{V} \subseteq \mathbb{R}^n$. The word *curve* is often used to refer to a 1-manifold, and *surface* for a 2-manifold. This definition appeals to our intuition that smooth k-manifolds look locally like the graphs of smooth functions. The graph of any smooth function $f : U \subseteq \mathbb{R}^k \to \mathbb{R}^n$ is a trivially a smooth k-manifold in $\mathbb{R}^{n+k}$, though most manifolds cannot be defined as such. In practice, manifolds are constructed using parameterizations and level sets.

Theorem 5.2

Suppose $\mathbf{F} : \mathbb{R}^{k+n} \to \mathbb{R}^n$ is a C^1 function, and $M = \mathbf{F}^{-1}(\mathbf{0})$. If $\operatorname{rank} D\mathbf{F}(\mathbf{p}) = n$ for every $\mathbf{p} \in M$, then M is a smooth k-manifold.

Proof. We need to show that for any $\mathbf{p} \in M$, there is a neighbourhood $V \subseteq M$ containing $\mathbf{p}$ such that V is the graph of a C^1 function $\mathbf{f} : U \subseteq \mathbb{R}^k \to \mathbb{R}^n$. Let $(\mathbf{x}, \mathbf{y}) \in \mathbb{R}^k \times \mathbb{R}^n$ denote the coordinates of $\mathbb{R}^{k+n}$, and fix a point $\mathbf{p} \in M$. Since $\operatorname{rank} D\mathbf{F}(\mathbf{p}) = n$, the Implicit Function Theorem ensures the existence of a neighbourhood $U \subseteq \mathbb{R}^k$ and function $\mathbf{f} : U \to \mathbb{R}^n$ such that $\mathbf{F}(\mathbf{x}, \mathbf{f}(\mathbf{x})) = \mathbf{0}$. This implies that $(\mathbf{x}, \mathbf{f}(\mathbf{x})) \in M$ for every $\mathbf{x} \in U$; namely, that these points are the graph of a C^1 function. □

The converse of this theorem is locally true as well. If $M \subseteq \mathbb{R}^n$ is a k-manifold, fix a point $\mathbf{p} \in M$ and a neighbourhood $V \subseteq M$ of $\mathbf{p}$ such that V is the graph of a C^1 function $\mathbf{f} : U \subseteq \mathbb{R}^k \to \mathbb{R}^n$. Define the function $\mathbf{F} : U \times \mathbb{R}^n \to \mathbb{R}^n$ by $\mathbf{F}(\mathbf{x}, \mathbf{y}) = \mathbf{y} - \mathbf{f}(\mathbf{x})$, so that $V = \mathbf{F}^{-1}(\mathbf{0})$ (though there is something to be said to ensure that the preimage is not larger than V).

There is nothing special about $\mathbf{0}$ in the statement of Theorem 5.2. Indeed, if $\mathbf{c} \in \mathbb{R}^n$ is any other point, define a function $\mathbf{G}(\mathbf{x}) = \mathbf{F}(\mathbf{x}) - \mathbf{c}$. We have $\{\mathbf{x} : G(\mathbf{x}) = \mathbf{0}\} = \{\mathbf{x} : F(\mathbf{x}) = \mathbf{c}\} = M$, and $D\mathbf{G} = D\mathbf{F}$, so the theorem holds true here as well.

Example 5.3

Show that the sphere of radius $r > 0$ is a smooth 2-manifold in $\mathbb{R}^3$.

Solution. Define the function $F_r : \mathbb{R}^3 \to \mathbb{R}$ by $F(x, y, z) = x^2 + y^2 + z^2 - r^2$, so that $F_r^{-1}(0) = \left\{(x, y, z) : x^2 + y^2 + z^2 = r^2\right\} = S_r^2$. To show that this is a smooth manifold, note that

$$DF_r(x, y, z) = (2x, 2y, 2z).$$

The maximal rank of DF_r is 1, and the only way it can be zero is if $(x, y, z) = (0, 0, 0)$, which is not an element of S_r^2. Hence by Theorem 5.2, S_r^2 is a smooth 2-manifold in $\mathbb{R}^3$. ■

The rank condition in Theorem 5.2 is important. To see what kind of things can go wrong, consider $F : \mathbb{R}^2 \to \mathbb{R}$ given by $F(x, y) = y^3 - x^2$ (Figure 5.2). This is certainly a C^1 function, but the zero locus it defines is the curve $y^3 = x^2$. Indeed, the corresponding curve has a cusp at $(0, 0)$.

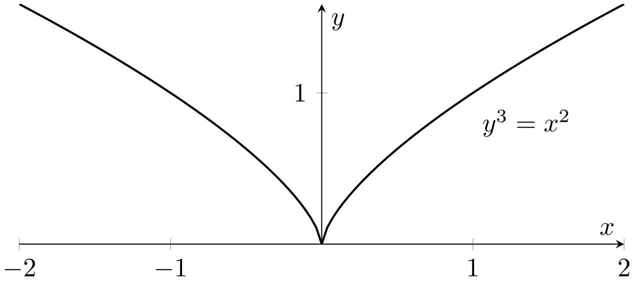


Figure 5.2: The curve defined by the graph of $f(x) = \sqrt[3]{x^2}$, the zero locus of $F(x, y) = y^3 - x^2$ and the parametric function $p(t) = (t^3, t^2)$.

Using level sets and zero loci to define smooth manifolds represents an easy criterion for determining smoothness. However, this framework is not useful for computation. The second method for constructing smooth manifolds is by means of a parameterization. The trade off is that several conditions are required to guarantee smoothness, and they are trickier to check.

Definition 5.4

Suppose $U \subseteq \mathbb{R}^k$ is a non-empty open set. We say that a C^1 map $\phi : U \to \mathbb{R}^n$ is a *regular embedding* if $\operatorname{rank} D\phi(\mathbf{x}) = k$ for all $\mathbf{x} \in U$, and ϕ is a homeomorphism onto its image (in the subspace topology).

Theorem 5.5

Let $M \subseteq \mathbb{R}^n$. If for every $\mathbf{p} \in M$ there exists open sets $U \subseteq \mathbb{R}^k$ and $V \subseteq M$, and a regular embedding $\phi : U \to V$ such that $\mathbf{p} \in V$, then M is a smooth k-manifold in $\mathbb{R}^n$.

Proof. Once again, it suffices to show that for every $\mathbf{p} \in M$, there is a neighbourhood $V \subseteq M$ such that V is the graph of smooth function $\mathbf{f} : U \to \mathbb{R}^{n-k}$ for some $U \subseteq \mathbb{R}^k$. Since $\operatorname{rank} D\phi(\mathbf{p}) = k$, $D\phi(\mathbf{p})$ has a $k \times k$ minor which is invertible. By rearranging the variables if necessary, we can assume the principal $k \times k$ minor is invertible.

Let $(\mathbf{u}, \mathbf{v})$ be the coordinates for $\mathbb{R}^n$, and write $(\mathbf{u}, \mathbf{v}) = \phi(\mathbf{x}) = (\mathbf{F}(\mathbf{x}), \mathbf{G}(\mathbf{x}))$ where $\mathbf{F} : \mathbb{R}^k \to \mathbb{R}^k$ and $\mathbf{G} : \mathbb{R}^k \to \mathbb{R}^{n-k}$. We know $D\mathbf{F}(\mathbf{p})$ is invertible by the above argument. By the Inverse Function Theorem, there are neighbourhoods $\tilde{U}, U \subseteq \mathbb{R}^k$ containing $\mathbf{p}$ and $\mathbf{F}(\mathbf{p})$ respectively, such that $\mathbf{F} : \tilde{U} \to U$ is a diffeomorphism. Since $\mathbf{u} = \mathbf{F}(\mathbf{x})$ and $\mathbf{F}$ is a diffeomorphism, we can in turn write $\mathbf{x} = \mathbf{F}^{-1}(\mathbf{u})$ for $\mathbf{u} \in U$. Hence $(\mathbf{u}, \mathbf{v}) \in M$ can be written as

$$(\mathbf{u}, \mathbf{v}) = \phi(\mathbf{x}) = (\mathbf{F}(\mathbf{x}), \mathbf{G}(\mathbf{x})) = (\mathbf{u}, \mathbf{G}(\mathbf{F}^{-1}(\mathbf{u})))$$

for all $\mathbf{u} \in U$. Setting $\mathbf{f} : U \to \mathbb{R}^{n-k}$ defined by $\mathbf{f} = \mathbf{G} \circ \mathbf{F}^{-1}$ shows that M can locally be written as the graph of a C^1 function. Moreover, since ϕ is a homeomorphism, $\phi(U)$ is open in the subspace topology, and hence can be written as $\phi(U) = V$ for some open set $V \subseteq M$, as required. □

The fact that ϕ is a homeomorphism was used only at the very end of the proof. In fact, if ϕ is just a C^1 map with rank $D\phi(\mathbf{x}) = k$ (omitting the requirement that ϕ^{-1} is continuous) then the proof of Theorem 5.5 shows that $\phi(V)$ is the graph of a smooth function, but that $\phi(U)$ might not be open in the subspace topology. To see why this is an issue, see Example 5.8.

The converse of Theorem 5.5 is true. Suppose $M \subseteq \mathbb{R}^n$ is a smooth manifold, and pick a point $\mathbf{p} \in M$ and an open set $V \subseteq M$ so that V looks like the graph of a C^1 function $\mathbf{f} : U \to \mathbb{R}^{n-k}$. Define $\phi : U \to V$ by $\phi(\mathbf{x}) = (\mathbf{x}, \mathbf{f}(\mathbf{x}))$, which is quickly seen to be a regular embedding.

Because of Theorem 5.5, the maps $\phi : U \to V$ are sometimes referred to as *coordinate charts.* A collection of coordinate charts which show that a set M is a smooth manifold is called an *atlas.* These terms will be referred to in the future.

Example 5.6

Show that the unit circle $S^1 = \left\{(x_1, x_2) : x_1^2 + x_2^2 = 1\right\} \subseteq \mathbb{R}^2$ is a smooth 1-manifold.

Solution. This is straightforward using level sets, but let's see how to do it with parameterizations. Let

$$S_i^+ = \left\{(x_1, x_2) : x_1^2 + x_2^2 = 1, x_i > 0\right\} \quad \text{and} \quad S_i^- = \left\{(x_1, x_2) : x_1^2 + x_2^2 = 1, x_i < 0\right\}, \quad i = 1, 2$$

be the four hemispheres of S^1. Define

$$I_1^+ = (-\pi/2, \pi/2), \quad I_1^- = (\pi, 3\pi/2), \quad I_2^+ = (0, \pi), \quad I_2^- = (\pi, 2\pi),$$

and $\phi_i^\pm : I_i^\pm \to S_i^\pm$ via $\phi_i^\pm(t) = (\cos(t), \sin(t))$. I claim these four maps are C^1 functions with rank 1 and are homeomorphisms onto their image. I'll do the work for ϕ_2^+ as the rest have similar arguments. This is certainly a C^1 map, with $D\phi_2^+(t) = (-\sin(t), \cos(t))$. It is impossible for both components to be zero simultaneously, so $D\phi_2^+$ has rank 1 everywhere. Its inverse is $(\phi_2^+)^{-1}(x, y) = \operatorname{arccot}(x/y)$, which is well defined since $y \neq 0$. If $(x, y) \in S^1$, then

$$[\phi_2^+ \circ (\phi_2^+)^{-1}](x, y) = (\cos(\operatorname{arccot}(x/y)), \sin(\operatorname{arccot}(x/y))) = \left(\frac{x}{x^2 + y^2}, \frac{y}{x^2 + y^2}\right) = (x, y)$$

while if $t \in (0,\pi)$ then

$$[(\phi_2^+)^{-1} \circ \phi_2^+](t) = \operatorname{arccot}\left(\frac{\cos(t)}{\sin(t)}\right) = t,$$

so indeed these maps are inverses of one another. Moreover, $(\phi_2^+)^{-1}$ is C^1 because $y \neq 0$, and so is actually a diffeomorphism, and not just a homeomorphism.

As any point $\mathbf{p} \in S^1$ lies in at least one of these four sets, the unit circle satisfies Theorem 5.5, and hence is a smooth manifold. ■

Remark 5.7

1. We might specify a manifold using a parameterization which is not a regular embedding. For example, the unit circle is the image of the map $\phi : [0, 2\pi) \to \mathbb{R}^2, t \mapsto (\cos(t), \sin(t))$. While this is similar in form to the maps $\phi_i^\pm$ in Example 5.6, its domain is not an open set.

2. It is possible to cover S^1 with only two homeomorphisms, as you'll do in Exercise 5-8.

Example 5.8

Consider the image of the parametric equation

$$\phi : \left(-\frac{\pi}{2}, \frac{3\pi}{2}\right) \to \mathbb{R}^2, \qquad t \mapsto \frac{1}{1+\sin^2 t}\left(\cos(t), \sin(t)\cos(t)\right). \tag{5.1}$$

Determine whether this defines a smooth manifold.

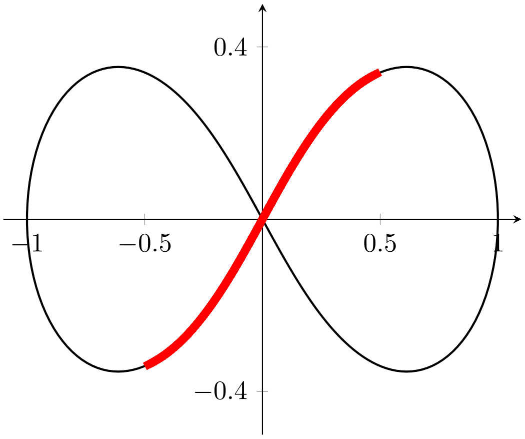


Figure 5.3: The lemniscate drawn with the parametric equation given by (5.1). The thick red line is the set $\phi((\pi/4, 3\pi/4))$, which is the graph of a C^1-function. However, the whole curve fails to be smooth.

Solution. Let M be the image of ϕ, shown in Figure 5.3. The map ϕ is clearly C^1, and moreover it begins and ends at the origin $(0,0)$, but only hits the origin at $t = \pi/2$. It's difficult to show algebraically, but ϕ is bijective, and its derivative

$$D\phi(t) = \frac{1}{(1+\sin^2(t))^2}\left(-\sin(t)[2+\cos^2(t)], \cos(2t)[1+\sin^2(t)] - \sin(t)\cos(t)\sin(2t)\right),$$

everywhere has rank 1.

But something is wrong here. There is no neighbourhood of $(0,0)$ where the curve looks like the graph of a C^1 function, despite the fact that ϕ is a C^1 bijection with constant rank. So what happened? The answer is that ϕ is *not* a homeomorphism onto its image, since ϕ^{-1} is not continuous. If this were the case, $\phi(U)$ would be relatively open for any open $U \subseteq (-\pi/2, 3\pi/2)$, but $\phi((\pi/4, 3\pi/4))$ (the red area in Figure 5.3) is not relatively open in C. ■

Example 5.9

Consider the surface S defined as the image of the the function $\mathbf{p} : \mathbb{R} \to \mathbb{R}^2$,

$$\mathbf{p}(s,t) = (s\cos(t), s\sin(t), s^2).$$

Find a zero-locus description of the image and determine if this is a smooth surface.

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. The derivative of F is given by

$$DF(x,y,z) = (-2x, -2y, 1)$$

and this is certainly never zero, hence S is a smooth 2-manifold. ■

Example 5.10

Find a parametric and level set description of the torus with major radius R and inner radius r.

Solution. Fix $r, R > 0$ and consider Figure 5.4. In the xy-plane, the coordinate of a point on the sphere $(x-R)^2 + y^2 = r^2$ is of the form

$$(x,y,z) = (R + r\cos(\phi), r\sin(\phi), 0).$$

We now rotate this about the y-axis, for which we use a rotation matrix

$$R_\theta = \begin{bmatrix} \cos(\theta) & 0 & -\sin(\theta) \\ 0 & 1 & 0 \\ \sin(\theta) & 0 & \cos(\theta) \end{bmatrix}$$

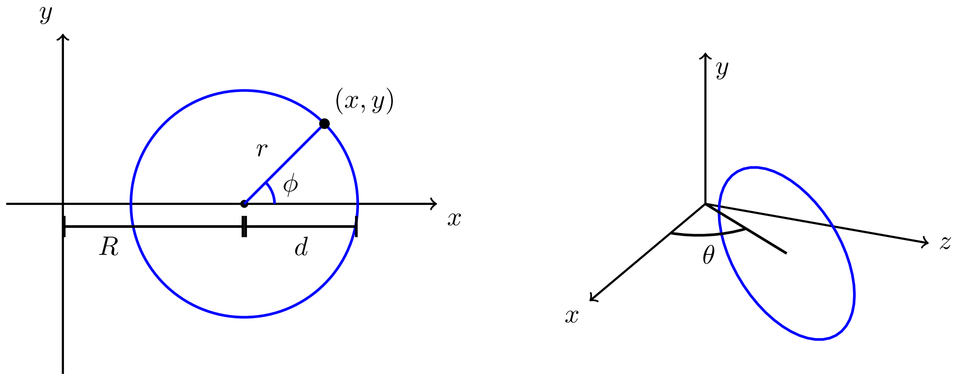


Figure 5.4: A slice of the torus.

to get

$$\begin{bmatrix}\cos(\theta) & 0 & -\sin(\theta)\\ 0 & 1 & 0\\ \sin(\theta) & 0 & \cos(\theta)\end{bmatrix}\begin{bmatrix}R+r\cos(\phi)\\ r\sin(\phi)\\ 0\end{bmatrix}=\begin{bmatrix}(R+r\cos(\phi))\cos(\theta)\\ r\sin(\phi)\\ (R+r\cos(\phi))\sin(\theta)\end{bmatrix}.$$

yielding a parameterization $\phi : [0,2\pi)\times[0,2\pi)\to\mathbb{R}^3$ given by

$$g(\phi,\theta)=[(R+r\cos(\phi))\cos(\theta), r\sin(\phi), (R+r\cos(\phi))\sin(\theta)].$$

The zero locus is somewhat harder to come by. Choose a point $\mathbf{p}=(x,0,z)$ in the xz-plane, whose distance from the origin is $d_{xz}=\sqrt{x^2+z^2}$. The difference in d_{xz} from R is $d=|R-d_{xz}|$. In the xy-pane, d represents the x distance, so (x,y,z) lives on the torus if $d^2+y^2=r^2$. Putting this all together gives

$$F(x,y,z)=\left[R-\sqrt{x^2+z^2}\right]^2+y^2-r^2=0. \qquad \blacksquare$$

5.1.1 Tangent Spaces

It was mentioned previously that smoothness is related to the dimension of the tangent space, which requires a definition of the tangent space in the first place. There are several ways of doing this, but the simplest from a geometric standpoint is to look at velocity vectors.

Definition 5.11

Let $M\subseteq\mathbb{R}^n$ be a smooth k-manifold, and fix a point $\mathbf{p}\in M$. Let $\mathcal{C}_{\mathbf{p}}$ denote the collection C^1 maps of the form $\gamma:(-1,1)\to M$ such that $\gamma(0)=\mathbf{p}$. Fix a regular embedding $\phi:U\to V$ such $\mathbf{p}\in V$, and let $\phi^{-1}:V\to U$ denote its inverse. Endow $\mathcal{C}_{\mathbf{p}}$ with an equivalence relation such that $\gamma_1\sim\gamma_2$ if $(\phi^{-1}\circ\gamma_1)'(0)=(\phi^{-1}\circ\gamma_2)'(0)$. The *tangent space to* M *at* $\mathbf{p}$, denoted $T_{\mathbf{p}}M$, as the collection of equivalence classes $[\gamma]\in\mathcal{C}_{\mathbf{p}}/\sim$, which we call *tangent vectors to* M *at* $\mathbf{p}$.

This won't seem intuitive at first, but let's shed some light on the definition. What we would like to say is that if $\gamma:(-1,1)\to M$ is a C^1 curve such that $\gamma(0)=\mathbf{p}$, then its velocity vector at $t=0$ is $\gamma'(0)=\mathbf{v}$, and this should be tangent to M at $\mathbf{p}$. Great, why didn't I define the tangent space this way? In reality I could have, but the mathematics becomes more difficult. Two different curves γ_1 and γ_2 could both define the same tangent vector $\mathbf{v}$, and because $\mathbf{v}$ is defined in terms of paths, we would have trouble seeing that $\gamma_1'(0) = \gamma_2'(0)$. The equivalence relation fixes this for us.

However, there are a few more things to check. We need to ensure that this definition does not depend upon the choice of the regular embedding ϕ, and that the relation defined above is really an equivalence relation. I have left these to Exercise 5-11.

Proposition 5.12

If $M \subseteq \mathbb{R}^n$ is a smooth k-manifold and $\mathbf{p} \in M$, we can realize $T_{\mathbf{p}}M$ as a k-dimensional vector space.

Proof. Let $\phi : U \to M \cap V$ be a regular embedding such that $\mathbf{p} \in V$. Denote by $\phi^{-1} : V \to U$ its inverse, and define a map

$$\mathbf{F} : T_{\mathbf{p}}M \to \mathbb{R}^k, \quad [\gamma] \mapsto \left.\frac{\mathrm{d}}{\mathrm{d}t}\right|_{t=0} \left(\phi^{-1} \circ \gamma\right),$$

for some representative $\gamma \in [\gamma]$. Note that $\phi^{-1} \circ \gamma : (-1, 1) \to \mathbb{R}^k$, so it makes sense to differentiate in one variable, and we've bypassed the manifold in between. This map is well defined, for if $\gamma_1, \gamma_2 \in [\gamma]$ then

$$\left.\frac{\mathrm{d}}{\mathrm{d}t}\right|_{t=0} \left(\phi^{-1} \circ \gamma_1\right) = \left.\frac{\mathrm{d}}{\mathrm{d}t}\right|_{t=0} \left(\phi^{-1} \circ \gamma_2\right),$$

which is precisely the definition of the equivalence relation. This same idea shows $\mathbf{F}$ is injective, since if $[\gamma], [\lambda] \in T_{\mathbf{p}}M$, then $\mathbf{F}([\gamma])) = \mathbf{F}([\lambda])$ means that for $\gamma \in [\gamma]$ and $\lambda \in [\lambda]$ we have

$$\left.\frac{\mathrm{d}}{\mathrm{d}t}\right|_{t=0} \left(\phi^{-1} \circ \gamma\right) = \left.\frac{\mathrm{d}}{\mathrm{d}t}\right|_{t=0} \left(\phi^{-1} \circ \lambda\right),$$

hence $\lambda \in [\gamma]$, showing that $[\lambda] = [\gamma]$. Finally, we show that $\mathbf{F}$ is surjective. Fix some $\mathbf{v} \in \mathbb{R}^k$ and define the map $\gamma : (-1, 1) \to M, t \mapsto \phi(\phi^{-1}(\mathbf{p}) + t\mathbf{v})$. This curve is a composition of C^1 functions and so C^1 itself, $\gamma(0) = \phi(\phi^{-1}(\mathbf{p})) = \mathbf{p}$, and

$$\mathbf{F}([\gamma]) = \left.\frac{\mathrm{d}}{\mathrm{d}t}\right|_{t=0} \phi^{-1}\left(\phi\left(\phi^{-1}(\mathbf{p}) + t\mathbf{v}\right)\right) = \left.\frac{\mathrm{d}}{\mathrm{d}t}\right|_{t=0} \phi^{-1}(\mathbf{p}) + t\mathbf{v} = \mathbf{v}.$$

If we use a different chart $\psi : \tilde{U} \to \tilde{V}$ and define $\mathbf{G}([\gamma]) = (\psi^{-1} \circ \gamma)'(0)$, then

$$\mathbf{G}([\gamma]) = \left.\frac{\mathrm{d}}{\mathrm{d}t}\right|_{t=0} \left(\psi^{-1} \circ \gamma\right) = \left.\frac{\mathrm{d}}{\mathrm{d}t}\right|_{t=0} \left(\left(\psi^{-1} \circ \phi\right) \circ \phi^{-1} \circ \gamma\right) = D(\psi^{-1} \circ \phi)(\phi^{-1}(\mathbf{p})) \left[\left.\frac{\mathrm{d}}{\mathrm{d}t}\right|_{t=0} \phi^{-1} \circ \gamma\right].$$

By Exercise 5-10, we know $\psi^{-1} \circ \phi$ is a diffeomorphism, so that $D(\psi^{-1} \circ \phi)(\phi^{-1}(\mathbf{p}))$ is a vector space isomorphism. For brevity, write this as T, so that $\mathbf{G}([\gamma]) = T\mathbf{F}([\gamma])$. Everything here is bijective, so this relationship will tell us precisely when two tangent vectors are the same in different charts. That is, if $\mathbf{v}, \mathbf{w} \in \mathbb{R}^k$ and $\mathbf{w} = T\mathbf{v}$, then $\mathbf{G}^{-1}(\mathbf{w}) = \mathbf{F}^{-1}(\mathbf{v})$ since

$$\mathbf{G}(\mathbf{F}^{-1}(\mathbf{v})) = T\mathbf{F}(\mathbf{F}^{-1}(\mathbf{v})) = T\mathbf{v} = \mathbf{w} = \mathbf{G}(\mathbf{G}^{-1}(\mathbf{w})),$$

and $\mathbf{G}$ is injective.

Now we use the bijection $\mathbf{F}$ to define the vector space structure on $T_{\mathbf{p}}M$. For $[\gamma], [\lambda] \in T_{\mathbf{p}}M$, let $\mathbf{v} = \mathbf{F}([\gamma])$ and $\mathbf{w} = \mathbf{F}([\lambda])$, defining

$$[\gamma] + [\lambda] = \mathbf{F}^{-1}(\mathbf{v} + \mathbf{w}) \quad \text{and} \quad c[\gamma] = \mathbf{F}^{-1}(c\mathbf{v}).$$

This structure makes $\mathbf{F}$ a linear map. All that remains to be seen is that the choice of chart does not matter. To see this, let $[\sigma_1] = \mathbf{F}^{-1}(\mathbf{v} + \mathbf{w})$, and let $\mathbf{G}$ be defined for the chart ψ above. If $\mathbf{v}' = \mathbf{G}([\gamma])$ and $\mathbf{w}' = \mathbf{G}([\lambda])$, set $[\sigma_2] = \mathbf{G}^{-1}(\mathbf{v}' + \mathbf{w}')$ so that

$$\mathbf{G}([\sigma_2]) = \mathbf{v}' + \mathbf{w}' = T\mathbf{v} + T\mathbf{w} = T\left(\mathbf{v} + \mathbf{w}\right) = T\mathbf{F}([\sigma_1])$$

showing that $[\sigma_1] = [\sigma_2]$.

Thus $T_{\mathbf{p}}M$ has a vector space structure, and in particular $\mathbf{F}$ is a vector space isomorphism, so $\dim T_{\mathbf{p}}M = k$. □

I mentioned we could think of $[\gamma]$ as the velocity vector of one of its representatives. Let's see that this is the case and flesh out the map $\mathbf{F}$ defined in the proof of Proposition 5.12. Let M be a smooth k-manifold, $\mathbf{p} \in M$, and fix a regular embedding $\phi : U \subseteq \mathbb{R}^k \to V \subseteq M$. Let $\gamma_1, \gamma_2 \in [\gamma]$, and restrict the map $D\phi^{-1}(\mathbf{p})$ to $T_{\mathbf{p}}M$. Then

$$\left.\frac{\mathrm{d}}{\mathrm{d}t}\right|_{t=0} (\phi^{-1} \circ \gamma_1) = D\phi^{-1}(\mathbf{p})\gamma_1'(0) = D\phi^{-1}(\mathbf{p})\gamma_2'(0) = \left.\frac{\mathrm{d}}{\mathrm{d}t}\right|_{t=0} (\phi^{-1} \circ \gamma_2).$$

But we know $D\phi^{-1}(\mathbf{p})$ restricted to $T_{\mathbf{p}}M$ is a bijection, so in particular $\gamma_1'(0) = \gamma_2'(0)$. Thus we can identify elements $[\gamma] \in T_{\mathbf{p}}M$ with vectors $\mathbf{v} \in \mathbb{R}^n$ which arise as velocity vectors of curves in M.

Note the distinction between the vector space $T_{\mathbf{p}}M$ and the geometric tangent space $T_{\mathbf{p}}M$. Namely, vector spaces must always pass through the origin, but the tangent spaces generally will not. For example, if $S^1 = \left\{(x, y) : x^2 + y^2 = 1\right\} \subseteq \mathbb{R}^2$, then the geometric tangent space to $\mathbf{p} = (1, 0)$ is the vertical line passing through $(1, 0)$, but the vector space to which it is isomorphic is $\mathbb{R}$, which we identify with the subspace $\{0\} \times \mathbb{R}$ translated so that the origin sits at $\mathbf{p}$.

Tangent spaces now give an alternative interpretation of the derivative. Suppose M and N are two manifolds of any dimension, and $f : M \to N$ is a C^1 map. For each $\mathbf{p} \in M$, there is an induced map $Df_{\mathbf{p}} : T_{\mathbf{p}}M \to T_{f(\mathbf{p})}N$ given by $Df_{\mathbf{p}}(\mathbf{v}) = Df(\mathbf{p})(\mathbf{v})$. We need only check that $Df(\mathbf{p})(\mathbf{v}) \in T_{f(\mathbf{p})}N$ for this map to be well defined.

Proposition 5.13

Let M, N be two smooth manifolds and suppose $f : M \to N$ is a C^1 map. If $\mathbf{p} \in M$ and $\mathbf{v} \in T_{\mathbf{p}}M$, then $Df(\mathbf{p})(\mathbf{v}) \in T_{f(\mathbf{p})}N$. Hence $Df_{\mathbf{p}} : T_{\mathbf{p}}M \to T_{f(\mathbf{p})}N, \mathbf{v} \to Df(\mathbf{p})(\mathbf{v})$ is a well-defined map.

Proof. Fix a $\mathbf{v} \in T_{\mathbf{p}}M$ and let γ be a representative, so that $\gamma : (-1, 1) \to M$ is a C^1 map, $\gamma(0) = \mathbf{p}$ and $\gamma'(0) = \mathbf{v}$. The composition $f \circ \gamma : (-1, 1) \to N$ is a C^1 curve in N with $(f \circ \gamma)(0) = f(\mathbf{p})$, so $(f \circ \gamma)'(0) \in T_{f(\mathbf{p})}N$, but

$$(f \circ \gamma)'(0) = Df(\mathbf{p})(\gamma'(0)) = Df(\mathbf{p})(\mathbf{v})$$

showing that $Df_{\mathbf{p}}(\mathbf{v}) = Df(\mathbf{p})(\mathbf{v}) \in T_{f(\mathbf{p})}N$ as required. □

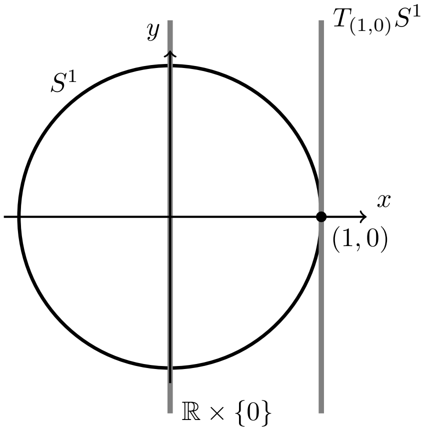


Figure 5.5: The tangent space $T_{(1,0)}S^1$ of the unit circle. This is not a subspace of the ambient vector space $\mathbb{R}^2$, but is nonetheless a vector space in its own right. It is isomorphic to the subspace $\mathbb{R} \times \{0\}$.

How do we compute tangent spaces in practice? One way is to actually differentiate curves $\gamma : (-1,1) \to M$, but there are alternative methods depending on whether we defined M using level sets or parameterizations.

Proposition 5.14

Suppose $M \subseteq \mathbb{R}^n$ is a smooth k-manifold, and $\mathbf{p} \in M$. Let $U \subseteq \mathbb{R}^k$ and $V \subseteq M$ be open sets such that $\mathbf{p} \in V$, and take $\alpha : U \to V$ to be a regular embedding. If $\mathbf{q} = \phi^{-1}(\mathbf{p})$, then $T_{\mathbf{p}}M = \operatorname{span}\{\partial_i \phi(\mathbf{q})\}_{i=1}^k$.

Proof. For each $i \in \{1, \ldots, k\}$, define the straight line path $\tilde{\gamma}_i(t) = \mathbf{q} + t\mathbf{e}_i$, where $\mathbf{e}_i$ is the standard basis vector in $\mathbb{R}^k$. Naturally, we can restrict ourselves to sufficiently small t to ensure that $\tilde{\gamma}_i(t) \subseteq U$ for all t. Define $\gamma_i = \phi \circ \tilde{\gamma}_i$, which is a C^1 curve in M satisfying $\gamma_i(0) = \phi(\tilde{\gamma}_i(0)) = \phi(\mathbf{q}) = \mathbf{p}$, and

$$\gamma_i'(0) = D\phi(\mathbf{q})\tilde{\gamma}_i'(0) = D\phi(\mathbf{q})\mathbf{e}_i = \frac{\partial \phi}{\partial x_i}(\mathbf{q}) \in T_{\mathbf{p}}M.$$

Since we know $\dim T_{\mathbf{p}}M = k$, it suffices to show that the $\partial_i \phi(\mathbf{q})$ are all linearly independent. But since ϕ is a regular embedding, we know $\operatorname{rank} D\phi(\mathbf{q}) = k$, meaning $D\phi(\mathbf{q})$ has precisely k linearly independent columns. Thus $T_{\mathbf{p}}M = \operatorname{span}\{\partial_i \phi(\mathbf{q}) : i = 1, \ldots, k\}$ as required. □

Example 5.15

Let M be the cylinder bounded by $x^2 + y^2 = 4$, $z = 0$, and $z = 1$. Find a basis for the tangent space at the point $\mathbf{p} = (2, 0, 1/2)$.

Solution. Let's fix a regular embedding $\phi : (-\pi, \pi) \times (0,1) \to M$ by $\phi(\theta, z) = (2\cos(\theta), 2\sin(\theta), z)$, so that $\phi(0,1/2) = (2,0,1/2) = \mathbf{p}$. Now

$$D\phi(\theta,z) = \begin{bmatrix} -\sin(\theta) & 0 \\ \cos(\theta) & 0 \\ 0 & 1 \end{bmatrix} \qquad \text{so} \qquad D\phi(0,1/2) = \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

Hence $T_{(2,0,1/2)}M = \operatorname{span}\left\{(0,1,0)^T, (0,0,1)^T\right\}$, as shown in Figure 5.6. ■

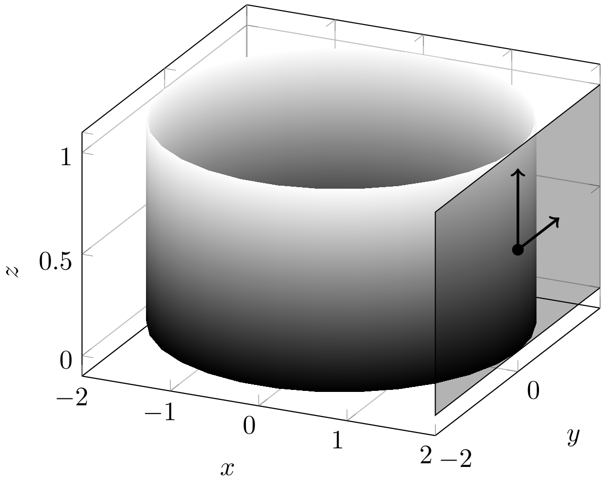


Figure 5.6: The tangent space to the cylinder from Example 5.15.

Proposition 5.16

If $\mathbf{F} : \mathbb{R}^{k+n} \to \mathbb{R}^n$ is a C^1 function, $M = \mathbf{F}^{-1}(\mathbf{0})$, and $\operatorname{rank} D\mathbf{F}(\mathbf{p}) = n$ for all $\mathbf{p} \in M$, then $T_\mathbf{p}M = \ker D\mathbf{F}(\mathbf{p})$.

Solution. By Theorem 5.2 we know that M is a smooth manifold, so $\dim T_\mathbf{p}M = k$. By rank-nullity, $\dim(\ker D\mathbf{F}(\mathbf{p})) = k$ as well, so it remains to show that $T_\mathbf{p}M \subseteq \ker D\mathbf{F}(\mathbf{p})$, from which the result will follow. Fix a $\gamma : (-1,1) \to M$ with $\gamma(0) = \mathbf{p}$ and $\gamma'(0) = \mathbf{v} \in T_\mathbf{p}M$. Since $\gamma(t) \in M$ for all $t \in (-1,1)$, this implies $\mathbf{F}(\gamma(t)) = \mathbf{0}$. Differentiating gives

$$\mathbf{0} = D\mathbf{F}(\gamma(0))\gamma'(0) = D\mathbf{F}(\mathbf{p})\mathbf{v}$$

so that $\mathbf{v} \in \ker D\mathbf{F}(\mathbf{p})$ as required. ■

In fact, if you think about your linear algebra, you'll realize that the rows of $D\mathbf{F}(\mathbf{p})$ span the orthogonal complement of $T_\mathbf{p}M$, which is why the dimension of the tangent space is the difference between the dimension of the domain and the codomain.

Example 5.17

Find the tangent space $T_\mathbf{p}M$ where $M = S^2 \subseteq \mathbb{R}^2$, and $\mathbf{p} = (1,0,0)$.

Solution. Let $F(x,y,z) = x^2+y^2+z^2-1$, so that $DF(1,0,0) = (2,0,0)$ has kernel

$$\ker DF(1,0,0) = \operatorname{span}\left\{ \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \right\}.$$

This is the yz-plane, though again we must translate it so that the origin sits at $\mathbf{p} = (1,0,0)$. This agrees with our intuition. ■

5.1.2 Manifolds with Boundary

Consider the set $B^2 = \left\{(x,y) : x^2+y^2 \leq 1\right\}$, the closed unit disk in $\mathbb{R}^2$. We know that $S^1 = \left\{(x,y) : x^2+y^2 = 1\right\}$ is a smooth curve, while Exercise 5-13 shows that $B_1(\mathbf{0}) = \left\{(x,y) : x^2+y^2 < 1\right\}$ is a smooth surface. It seems as though these should be related. Sure, their dimensions are different, but S^1 should represent the "boundary" of $B_1(\mathbf{0})$. We can extend the notion of a manifold to allow B^2 to be a smooth manifold, but in doing so we must abandon the idea that a manifold looks locally like the graph of a smooth function. Instead, we'll rely entirely upon the parametric and level set notions of a manifold. To do this, we determine a basic example and use C^1 functions to shape it to our liking.

If $k \in \mathbb{N}$, define $\mathbb{H}^k = \left\{(\mathbf{x}, y) \in \mathbb{R}^{k-1} \times \mathbb{R} : y \geq 0\right\} \subseteq \mathbb{R}^k$. You should think of this set as being analogous to the upper-half plane in $\mathbb{R}^2$, which has a geometric boundary corresponding to the x-axis. Note that

$$(\mathbb{H}^k)^{\text{int}} = \{(\mathbf{x}, y) : y > 0\} \quad \text{and} \quad \partial\mathbb{H}^k = \{(\mathbf{x}, y) : y = 0\}.$$

We endow $\mathbb{H}^k$ with the relative topology it inherits from Euclidean $\mathbb{R}^k$.

Finally, recall that that if $S \subseteq \mathbb{R}^n$ is a not necessarily open set, we say that $\mathbf{f} : S \to \mathbb{R}^m$ is of type C^k if for every $\mathbf{p} \in S$ there is an open neighbourhood $U_\mathbf{p} \subseteq \mathbb{R}^n$ on which $\mathbf{f}$ can be extended to a C^k function. Moreover, the derivative of $\mathbf{f}$ is independent of the extension and is determined entirely by the value of $D\mathbf{f}$ on S^{int} (Exercise 3-33). This allows us to define regular embeddings on non-open sets: If $U \subseteq \mathbb{R}^k$ is non-empty, a C^1 map $\phi : U \to \mathbb{R}^n$ is a regular embedding if $\operatorname{rank} D\phi(\mathbf{x}) = k$ for all $\mathbf{x} \in U$, and ϕ is a homeomorphism onto its image.

Definition 5.18

A set $M \subseteq \mathbb{R}^n$ is a *smooth k-manifold with boundary* if for every $\mathbf{p} \in M$ there exists open sets $V \subseteq M$, $U \subseteq \mathbb{H}^k$, and a regular embedding $\phi : U \to V$. We say that $\mathbf{p}$ is an *interior point* of M if there is a regular embedding $\phi : U \to V$ – with $\mathbf{p} \in \phi(U)$ – where $U \subseteq (\mathbb{H}^k)^{\text{int}}$, and a *boundary point* otherwise. We will denote the collection of boundary points of M by ∂M.

If $U \subseteq (\mathbb{H}^k)^{\text{int}}$ we will call $\phi : U \to V \subseteq M$ an *interior chart*, and a *boundary chart* otherwise. The idea is that points along the boundary $\partial\mathbb{H}^k$ will map to those points that should constitute the *geometric boundary ∂M*. Note that the geometric boundary will be different than the topological boundary we discussed in Section 2, a point I'll make clear once we've developed our toolbox further. You will show in Exercise 5-16 that instead of using just $\mathbb{H}^k$ to parameterize your manifold, you can use a combination of $\mathbb{R}^k$ and $\mathbb{H}^k$. In this framework, $\mathbb{R}^k$ can be used to parameterize interior points while $\mathbb{H}^k$ will be used to parameterize boundary points. I'll use this result freely hereafter.

Example 5.19

Show that $B^2 = \{(x,y) : x^2 + y^2 \leq 1\}$ is a smooth 2-manifold with boundary.

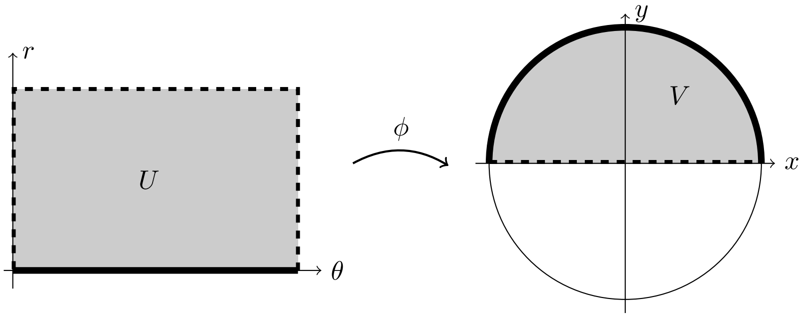


Figure 5.7: A boundary chart for the unit ball B^2. The boundaries, indicated by the thickened lines, are mapped to one another.

Solution. Let's break B^2 into $S^1 = \{(x,y) : x^2 + y^2 = 1\}$ and $B_1(\mathbf{0}) = \{(x,y) : x^2 + y^2 < 1\}$. If $\mathbf{p} \in B_1(\mathbf{0})$, take as the regular embedding the identity map $\operatorname{id} : B_1(\mathbf{0}) \to B_1(\mathbf{0})$. Thus we only need to show that points in S^1 admit local regular embeddings. For the moment, let $V = \{(x,y) : x^2 + y^2 \leq 1, y > 0\}$ be the upper hemisphere of the disk. Define $U = (0,\pi) \times [0,1)$ and define $\phi : U \to V$ by $\phi(\theta, r) = ((1-r)\cos(\theta), (1-r)\sin(\theta))$. You can quickly check that U and V are open in their respective ambient spaces, and that ϕ is a regular embedding. By rotating V to account for the left-, right-, and lower-hemispheres of the disk, we've ensured every point in S^1 admits a local regular embedding. ■

Proposition 5.20

Let $M \subseteq \mathbb{R}^n$ be a smooth k-manifold with boundary, and fix a point $\mathbf{p} \in M$. Let $\phi : U \to V$ be a regular embedding for a neighbourhood $V \subseteq M$ of $\mathbf{p}$ and an open set $U \subseteq \mathbb{H}^k$. If $\phi^{-1}(\mathbf{p}) \in (\mathbb{H}^k)^{\text{int}}$ then $\mathbf{p}$ is an interior point of M, and if $\phi^{-1}(\mathbf{p}) \in \partial\mathbb{H}^k$, then $\mathbf{p}$ is a boundary point of M.

Proof. The first part is straightforward: Since $\mathbf{p} \in (\mathbb{H}^k)^{\text{int}}$, there exists an $r > 0$ such that $B_r(\mathbf{p}) \subseteq (\mathbb{H}^k)^{\text{int}}$, in which case restricting ϕ to $B_r(\mathbf{p}) \cap U$ gives a regular embedding whose domain is contained in $(\mathbb{H}^k)^{\text{int}}$.

Now suppose $\phi^{-1}(\mathbf{p}) \in \partial\mathbb{H}^k$. Assume $\psi : \tilde{U} \to \tilde{V}$ is another regular embedding, with $\tilde{U} \subseteq (\mathbb{H}^k)^{\text{int}}$ and $\tilde{V}$ an open neighbourhood of $\mathbf{p}$ in M. Let $Y = V \cap \tilde{V}$, an open neighbourhood of $\mathbf{p}$ in M, and define $X_1 = \phi^{-1}(Y), X_2 = \psi^{-1}(Y)$. The map $\alpha = \phi^{-1} \circ \psi : X_2 \to X_1$ is a diffeomorphism by Exercise 5-18. Thinking of α as a map $X_2 \to \mathbb{R}^n$, Exercise 3-55 tells us that $X_1 = \alpha(X_2)$ is open in $\mathbb{R}^n$, but this is a contradiction, since $X_1 \subseteq \mathbb{H}^k$ contains $\mathbf{p} \in \partial\mathbb{H}^k$, showing that X_1 cannot be open in $\mathbb{R}^n$. □

The previous proposition is quite technical, owing to the use of Diffeomorphic Invariance of

Domain (Exercise 3-55). The use of level sets to define manifolds with boundary can be made possible by the following proposition.

Proposition 5.21

Let $F : \mathbb{R}^n \to \mathbb{R}$ be a C^1 function, and define $M = F^{-1}([0, \infty))$. If $\nabla F(\mathbf{x}) \neq \mathbf{0}$ for all $\mathbf{x} \in M$, then M is a smooth $(n-1)$-manifold with boundary, and $\partial M = F^{-1}(0)$.

Proof. The set $F^{-1}((0, \infty))$ is open, and so admits a regular embedding by Exercise 5-13. Thus assume $\mathbf{p} \in F^{-1}(0)$. Since $DF(\mathbf{p}) \neq \mathbf{0}$, at least one of its partials is non-zero. By re-ordering the variables if necessary, suppose $\partial_n F(\mathbf{p}) \neq 0$, and define the function $\phi : \mathbb{R}^n \to \mathbb{R}^n$ by $\phi(\mathbf{x}) = (x_1, x_2, \ldots, x_{n-1}, F(\mathbf{x}))$. Quick computation reveals that $\det D\phi(\mathbf{p}) = \partial_n F(\mathbf{p}) \neq 0$, so by the Inverse Function Theorem there are open neighbourhoods $U, V \subseteq \mathbb{R}^n$ with $\mathbf{p} \in U$ and $\phi(\mathbf{p}) \in V$ such that $\phi : U \to V$ admits a C^1-inverse ϕ^{-1}. Moreover, $\phi(U \cap M) = V \cap \mathbb{H}^n$ since $\mathbf{x} \in M$ means that $F(\mathbf{x}) \geq 0$, so $\phi(\mathbf{x}) = (x_1, \ldots, x_{n-1}, F(\mathbf{x})) \in \mathbb{H}^n$. Thus $\phi^{-1} : V \cap \mathbb{H}^n \to U \cap M$ is the desired regular embedding. Clearly $\phi(\mathbf{p}) = (p_1, \ldots, p_{n-1}, 0) \in \partial\mathbb{H}^n$, showing that $\mathbf{p} \in \partial M$. Hence $\partial M = F^{-1}(0)$ as required. □

As with level set results in general, we can be flexible in how we define the preimages. For example, instead of using $[0, \infty)$, we could use the bounded set $[0, r)$. To see this, note that the proof only required that $(0, r)$ be open and positive. From here, one could use $[a, b)$ for any $a < b$. Indeed, define $G(\mathbf{x}) = F(\mathbf{x} + \mathbf{a})$ so that $G^{-1}([0, b-a)) = F^{-1}([a, b))$, and $DG = DF$. Finally, if one wanted to use a set of the form $(a, b]$, set $G(\mathbf{x}) = -F(\mathbf{x})$ so that $G([-b, -a)) = F((a, b])$, and $DG = -DF$ does not affect the hypotheses of the theorem.

Example 5.22

Show that the 2-ball $B^2 = \left\{(x, y, z) : x^2 + y^2 + z^2 \leq 1\right\}$ is a smooth 2-manifold with boundary.

Proof. Let $F(x, y, z) = 1 - x^2 - y^2 - z^2$, which is certainly a C^1 function. It's straightforward to show that $B^2 = F^{-1}([0, \infty))$, showing that B^2 is a 2-manifold with boundary. Moreover, the boundary is $S^2 = F^{-1}(0) = \left\{(x, y, z) : x^2 + y^2 + z^2 = 1\right\}$. □

Here now we can make the distinction between topological and geometric boundaries. Every smooth k-manifold is a k-manifold with boundary (Exercise 5-17), but it may be the case that the boundary is empty. For example, the set S^2 is a smooth 2-manifold with boundary, but it's boundary is $\partial_{\text{geom}} S^2 = \emptyset$. This disagrees with its topological boundary $\partial_{\text{top}} S^2 = S^2$. The majority of the remainder of these notes deals with the geometric boundary, and that should be contextually clear. However, in cases where it is not, I will mention which interpretation of the boundary we are using.

Manifolds with Corners: Despite saying earlier that the cube in $\mathbb{R}^3$ is not smooth, we may still want to work with it; after all, it only fails to be smooth at a small collection of points. I'll run through this portion quite quickly, since the proofs mimic those above. We take as our model space the set

$$\overline{\mathbb{R}^k_+} = \{(x_1,\ldots,x_k) : x_1 \geq 0, x_2 \geq 0, \ldots, x_k \geq 0\}.$$

Definition 5.23

A set $M \subseteq \mathbb{R}^n$ is said to be a *smooth k-manifold with corners* if for every $p \in M$ there exist open sets $V \subseteq M$, $U \subseteq \overline{\mathbb{R}^k_+}$, and a regular embedding $\phi : U \to V$ such that $p \in V$. We say that $\mathbf{p}$ is a *corner point* of M if $\mathbf{q} = \phi^{-1}(\mathbf{p})$ has at least two coordinates identically zero. We'll denote the corner set as $\angle M$.

Certainly $\overline{\mathbb{R}^k_+}$ is a smooth k-manifold with corners, and its corner set consists of those points with at least two coordinates being zero. Let

$$H_i^k = \left\{(x_1,\ldots,x_k) \in \overline{\mathbb{R}^k_+} : x_i = 0\right\}. \tag{5.2}$$

Any chart $\phi : U \to V$ such that $U \cap \angle\overline{\mathbb{R}^k_+} \neq \emptyset$ is said to be a *corner chart.* If ϕ is not a corner chart, but $U \cap H_i^k \neq \emptyset$ for some $i \in \{1,\ldots,k\}$, then ϕ is a boundary chart, and if $U \subseteq (\overline{\mathbb{R}^k_+})^{\text{int}}$, it is an interior chart. Invariance of Domain can be used to prove Invariance of Corners; that is, a point $\mathbf{p} \in M$ is a corner point precisely if there is a chart $\phi : U \to V$ such that $\phi^{-1}(\mathbf{p})$ is a corner point of $\overline{\mathbb{R}^k_+}$. Note that every smooth manifold with boundary is a smooth manifold with corners, though the corner set might be empty (Exercise 5-20).

Unfortunately, the boundary of a smooth manifold with corners is not itself a smooth manifold with corners, and this will complicate some of the future discussion. At the end of the day, don't worry too much about the technical details here, but focus on absorbing the intuition.

5.1.3 Orientation

Orientation is the notion of relative position. For example, when a train pulls into a station, an automated voice might tell you that "the doors will open on the left." But what does "left" mean? The assumption is that left is relative to a commuter standing upright and facing the direction in which the train is travelling. However, if I were standing on my head, left would mean the opposite direction. Alternatively, stand in front of a mirror, holding your left hand in front of you, and your right hand to the side. Mirror-you is doing the opposite: his/her right hand is facing frontward, and his/her left hand is to the side. We aim to translate this idea to manifolds. For example, an orientation of a curve is a choice as to which direction along the curve is 'forward.' An orientation along a surface is a choice of 'forward-left,' and an orientation of a 3-manifold is a choice of 'forward-left-up.'

Exercise 1-40 discussed how vector spaces are endowed with orientations, and the idea is to use the tangent space to define the orientation locally. From Proposition 5.14, if $\phi : U \to V$ is a regular embedding for a manifold $M \subseteq \mathbb{R}^n$, then the columns of $D\phi(\mathbf{p})$ form a basis for $T_{\mathbf{p}}M$ whenever $\mathbf{p} \in V$, and hence define an orientation on $T_{\mathbf{p}}M$. To define an orientation globally, we just need to ensure we choose regular embeddings in a consistent manner.

Definition 5.24

If $U, V \subseteq \mathbb{R}^k$ are two open sets, a diffeomorphism $\mathbf{G} : U \to V$ is said to be *orientation-preserving* if $\det D\mathbf{G}(\mathbf{x}) > 0$ for every $\mathbf{x} \in U$.

As discussed in Proposition 5.13, the map $\mathbf{G} : U \to V$ induces a map on tangent spaces, $DG_{\mathbf{p}} : T_{\mathbf{p}}U \to T_{\mathbf{G}(\mathbf{p})}V, \mathbf{v} \mapsto D\mathbf{G}(\mathbf{p})(\mathbf{v})$. Under the above definition, if $\mathbf{G}$ is orientation preserving, then the linear transformation $D\mathbf{G}_{\mathbf{p}}$ is orientation preserving in the classical sense.

Definition 5.25

Let $M \subseteq \mathbb{R}^n$ be a smooth k-manifold (with boundary). Let $\phi_i : U_i \to V_i, i = 1, 2$ be regular embeddings of M such that $V_1 \cap V_2 \neq \emptyset$. We say that ϕ_1 and ϕ_2 are *consistently oriented* if their transition map $\phi_2^{-1} \circ \phi_1 : \phi_1^{-1}(V_1 \cap V_2) \subseteq \mathbb{R}^k \to \phi_2^{-1}(V_1 \cap V_2) \subseteq \mathbb{R}^k$ is orientation-preserving.

Regular embeddings therefore induce orientations on tangent spaces, and consistently oriented embeddings maintain the same orientation through the orientation preserving diffeomorphism that is their transition map.

Definition 5.26

Let $M \subseteq \mathbb{R}^n$ be a smooth k-manifold (with boundary). If M can be covered by consistently oriented regular embeddings, we say that M is *orientable*. An atlas consisting of all consistently oriented regular embeddings on M is said to be an *orientation* for M, and M together with an orientation is said to be an *oriented manifold*.

Example 5.27

Show that S^1 is an orientable manifold.

Solution. Let's cover S^1 with the four charts from Example 5.6, which I claim are all consistently oriented. I'll do one example, but the arguments for the others are nearly identical. Consider $\phi_1^+ : (-\pi/2, \pi/2) \to \mathbb{R}^2$ and $\phi_2^+ : (0, \pi) \to \mathbb{R}^2$, both of which are of the form $\phi_i^+(t) = (\cos(t), \sin(t))$. Their inverses are different,

$$(\phi_1^+)^{-1}(x, y) = \arctan(y/x) \quad \text{and} \quad (\phi_2^+)^{-1}(x, y) = \operatorname{arccot}(x/y),$$

but this won't matter. Their transition function $\alpha = (\phi_2^+)^{-1} \circ \phi_1^+ : (0, \pi/2) \to \mathbb{R}^2$ is the identity map $\alpha(t) = t$, for which $\det D\alpha(t) = 1 > 0$. Thus ϕ_2^+ and ϕ_1^+ are consistently oriented. The other transition functions are also the identity, thus we can cover S^1 with a set of consistently oriented charts, showing that S^1 is an orientable manifold. ■

If ϕ_2^+ is the map given in Example 5.27, two embeddings that are inconsistently oriented with respect to ϕ_2^+ are

$$\begin{aligned} \psi_1 : (0, \pi) &\to \mathbb{R}^2 \\ t &\mapsto (\sin(t), \cos(t)) \end{aligned} \quad \text{and} \quad \begin{aligned} \psi_2 : (-\pi/2, \pi/2) &\to \mathbb{R}^2 \\ t &\mapsto (-\cos(t), \sin(t)) \end{aligned},$$

though they are consistently oriented with respect to each other.

In the special case where $M \subseteq \mathbb{R}^n$ is a smooth n-manifold, any regular embedding $\phi : U \to V$ has an $n \times n$ derivative $D\phi$. If the columns of $D\phi(\mathbf{p})$ define a positive orientation for each $\mathbf{p} \in V$ (that is, $\det D\phi(\mathbf{p}) > 0$), then we say that M has the *natural orientation.*

In the special case of curves (1-manifolds) and hypersurfaces ($(n-1)$-manifolds), orientations can be thought of as a single vector tracing its way through the manifold, as follows:

Oriented Curves: An orientation of a one dimensional vector space can be thought of as an orientation of $\mathbb{R}$, which boils down to the choice of facing in the positive or negative direction. To remove the length of the vector from consideration, an orientation on $\mathbb{R}$ is a choice of a number from $\{\pm 1\}$.

Suppose $C \subseteq \mathbb{R}^n$ is a smooth orientable curve, so that the tangent spaces of C are consistently oriented. This amounts to the choice of a unit tangent vector at each point along C, which we can think of as specifying the direction of travel of a particle along that curve. Therefore, an oriented curve is a curve with a specified direction of travel.

Oriented Hypersurfaces: Let S be an oriented smooth $(n-1)$-manifold in $\mathbb{R}^n$, so that for each $\mathbf{p} \in S$, $T_{\mathbf{p}}M$ is an oriented $(n-1)$-dimensional vector space, and the orientations are consistent. As $T_{\mathbf{p}}M$ is $(n-1)$-dimensional, it is entirely specified by a unit normal vector $\hat{\mathbf{n}}_{\mathbf{p}}$ from Exercise 1-38. Hence an oriented hypersurface is equivalent to choosing a continuous choice of normal vector for every point along the hypersurface.

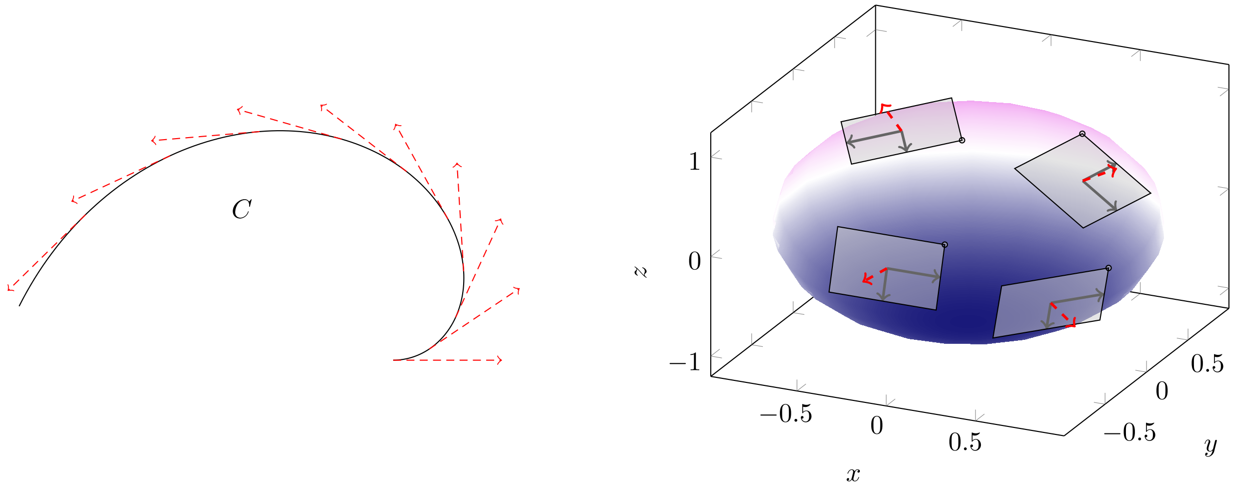


Figure 5.8: Left: The orientation of a curve C is equivalent to specifying a unit tangent vector at every point on the curve in a continuous fashion. Right: An orientation on a hypersurface is equivalent to specifying a unit normal vector at every point on the hypersurface in a continuous fashion.

With this disucssion in mind, we can give an example of a non-orientable manifold. Consider the Möbius band, constructed by identifying the specified edges of the rectangle shown in Figure 5.9. Walking around the Mbius band will flip the orientation of any normal vector, meaning it is impossible to define a consistent set of covering charts.

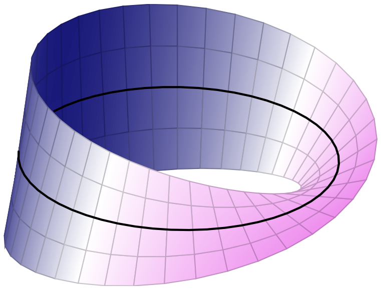


Figure 5.9: The Mobius band. Consider the central circle and specify a normal vector in any direction you like. Translating this normal vector around the circle will cause the normal to flip direction after a 2π rotation.

Finally and most importantly, we discuss the Stokes' orientation, which is the orientation induced on ∂M given an orientation on M. You will show in Exercise 5-21 that if M is orientable then the charts defining the orientation restrict to consistently oriented charts on ∂M, making ∂M orientable.

Definition 5.28

Suppose $M \subseteq \mathbb{R}^n$ is an oriented smooth k-manifold with non-empty boundary ∂M. The *Stokes' Orientation* on ∂M is that induced by the restricted charts if k is even, and the opposite of the restricted charts if k is odd.

The Stokes' orientation is chosen so that Theorems 5.41, 5.55, and 5.58, omit extraneous minus signs, but the proofs of those theorems involve a $(-1)^k$ term. Hence when k is odd, we choose the opposite orientation so that two minus signs will annihilate one another. The two cases we'll be using most extensively are when $M \subseteq \mathbb{R}^2$ is a smooth surface, and when $M \subseteq \mathbb{R}^3$ is a smooth 3-manifold. In both cases, M will be given the natural orientation, and ∂M will be a smooth hypersurface.

Let's develop a heuristic to see the boundary orientations. If $M \subseteq \mathbb{R}^2$ is an oriented smooth surface with the natural orientation, let $\phi : U \subseteq \mathbb{H}^2 \to V \subseteq M$ be a boundary chart with $\det D\phi(\mathbf{x}) > 0$ for all $\mathbf{x} \in U$. We can therefore either think of the columns of $D\phi(\mathbf{x})$ as specifying a positive orientation on $T_{\phi(\mathbf{x})}M$, or think of $D\phi(\mathbf{x})$ as an orientation preserving map. In the latter case, $D\phi(\mathbf{x})$ maps the standard basis $\{\mathbf{e}_1, \mathbf{e}_2\}$ of $T_{\mathbf{x}}U$ to the basis of column vectors of $D\phi(\mathbf{x})$ in $T_{\phi(\mathbf{x})}M$, which are thus in the same orientation. The restriction of $\{\mathbf{e}_1, \mathbf{e}_2\}$ to $\partial\mathbb{H}^2$ is just $\{\mathbf{e}_1\}$, and since $k = 2$ is even, this induces the Stokes' orientation on the boundary. In effect, when we walk along the boundary, the interior of the set should be to our left (Figure 5.10).

If $M \subseteq \mathbb{R}^3$ is an oriented 3-manifold and $\phi : U \subseteq \mathbb{R}^3 \to V \subseteq M$ is a naturally oriented boundary chart, the same argument above says that the induced orientation in $\partial\mathbb{H}^3$ is the standard basis $\{\mathbf{e}_1, \mathbf{e}_2\} \subseteq \mathbb{R}^2 \times \{0\}$. The normal vector which defines this orientation points towards the interior of the set, but since $k = 3$ is odd, we use the reverse orientation. Hence the Stokes' orientation in this case is defined by a normal vector which points outwards from the interior of M.

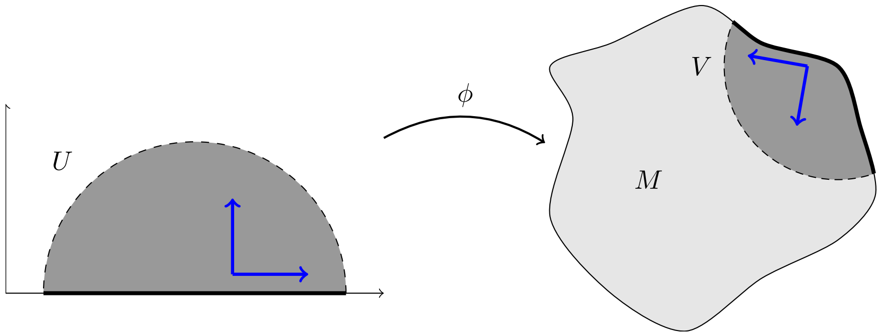


Figure 5.10: For a 2-manifold M with the natural orientation, the Stokes' orientation of the boundary curve is to travel in the direction such that the interior of M is to the left.

5.2 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 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 $\mathbf{F}(\mathbf{x}) \in \mathbb{R}^n$ means there is no measure of what is "largest" or "smallest." Instead, we'll realize $\mathbf{F}$ as a *vector field*. A vector field $\mathbf{F}$ is function which prescribes to every point $\mathbf{x} \in \mathbb{R}^n$ the 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 5.12.

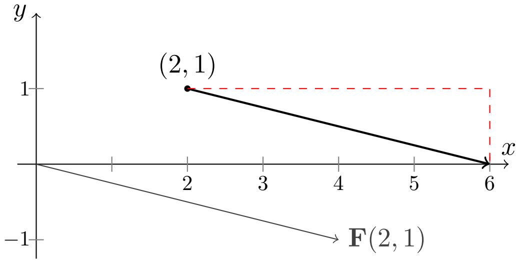
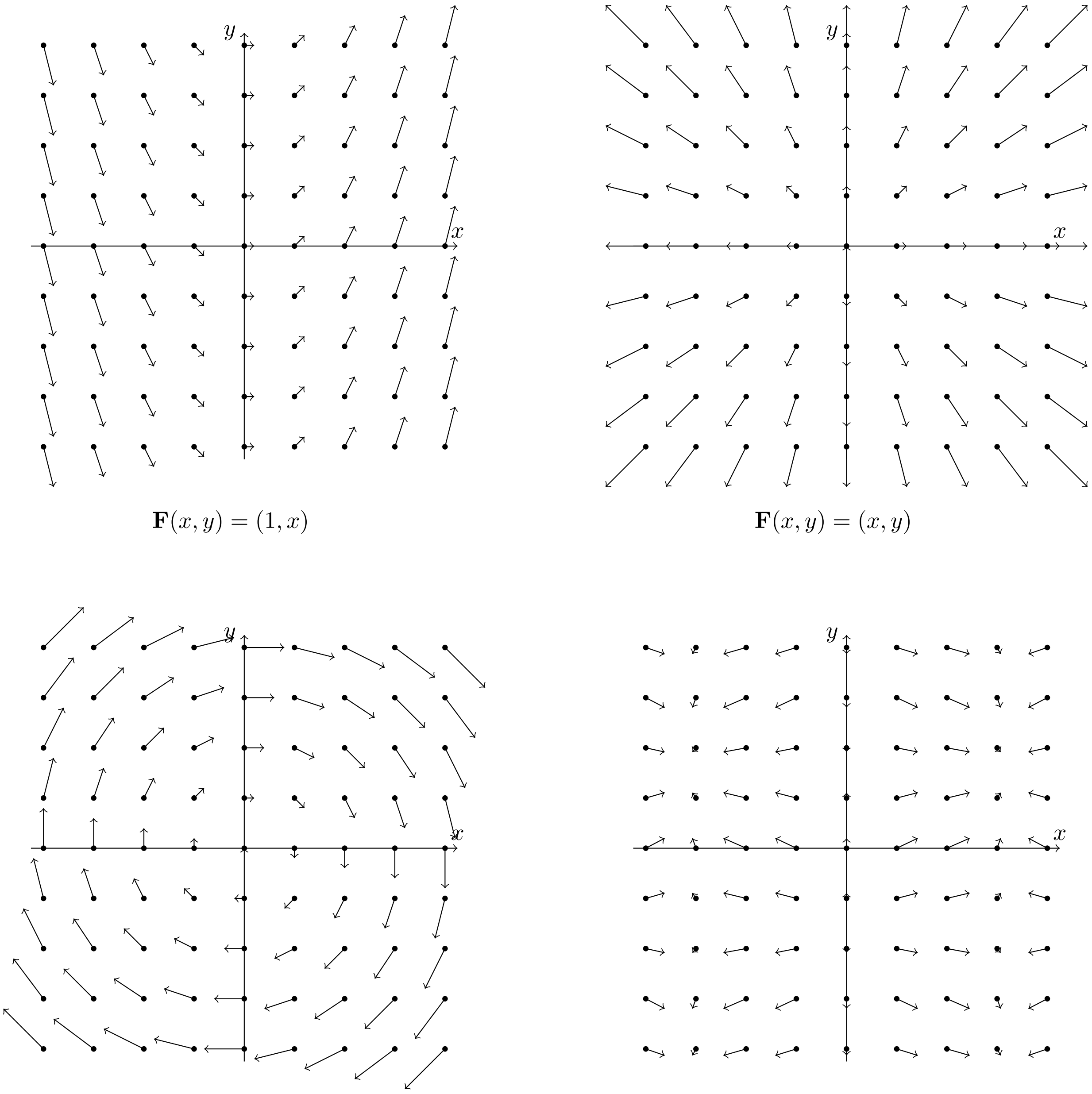


Figure 5.11: A single vector for the vector field $\mathbf{F}(x, y) = (x^2, -y)$. We assign the point $\mathbf{p} = (2, 1)$ the vector $\mathbf{v} = F(\mathbf{p}) = (4, -1)$, and visualize this by drawing $\mathbf{v}$ with its base at $\mathbf{p}$.

Vector fields can be used to describe physical fields and forces: The force exhibited by an electromagnetic field or gravity may be conveyed as a vector field, or 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.



$\mathbf{F}(x, y) = (1, x)$

$\mathbf{F}(x, y) = (x, y)$

$\mathbf{F}(x, y) = (y, -x)$

$\mathbf{F}(x, y) = (\sin(x), \cos(y))$

Figure 5.12: A visualization of several vector fields. Using many points gives us an intuitive idea for how these vectors ”flow.”

5.2.1 Vector Derivatives

The notion of a derivative becomes harder to pin down with vector fields, as there are several interpretation of what you might want a derivative to do. In this section, we'll to look at four such operators: the gradient (which you have already seen), divergence, curl, and the Laplacian. The first three of these are all actually the same operator in disguise, but that requires some knowledge of differential forms. In each of these cases, we'll abuse notation and think of the nabla operator ∇ as a vector whose components are the partial derivative operators. 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

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

   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

$$\operatorname{div} \mathbf{F} = \nabla \cdot \mathbf{F} = \frac{\partial F_1}{\partial x_1} + \cdots + \frac{\partial F_n}{\partial x_n}.$$

   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 $\mathbb{R}^3$, the *curl* of $\mathbf{F}$ is

$$\operatorname{curl} \mathbf{F} = \nabla \times \mathbf{F} = \left(\frac{\partial F_3}{\partial x_2} - \frac{\partial F_2}{\partial x_3}, \frac{\partial F_1}{\partial x_3} - \frac{\partial F_3}{\partial x_1}, \frac{\partial F_2}{\partial x_1} - \frac{\partial F_1}{\partial x_2}\right).$$

   The curl measures the *infinitesimal circulation* of the vector field. The cross-product means that this definition of curl is restricted to vector fields in $\mathbb{R}^3$, though this operator does generalize to higher dimensions. 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, the *Laplacian* of f is
$$\nabla^2 f = \Delta f = \frac{\partial^2 f}{\partial x_1^2} + \cdots + \frac{\partial^2 f}{\partial x_n^2}.$$
We can 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
$$\nabla^2 f(x,y,z) = \left[\frac{\partial^2}{\partial x^2}\left(x^2y + z^3\right)\right] + \left[\frac{\partial^2}{\partial y^2}\left(x^2y + z^3\right)\right] + \left[\frac{\partial^2}{\partial z^2}\left(x^2y + z^3\right)\right] = 2y + 6z.$$

All of these vector derivatives are important in physics and mathematics. While it won't be the focus of our ongoing conversation, the Laplacian 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). You will show this in Exercises 5-54 and 5-61. 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 $\operatorname{div}\mathbf{F} = 0$ we say that $\mathbf{F}$ is *incompressible*.

Proposition 5.29

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). The gradient, divergence, and curl 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
$$\begin{aligned}
\operatorname{curl}(f\mathbf{G})_1 &= \frac{\partial}{\partial y}(fG_3) - \frac{\partial}{\partial z}(fG_2) = \frac{\partial f}{\partial y}G_3 + f\frac{\partial G_3}{\partial y} - \frac{\partial f}{\partial z}G_2 - f\frac{\partial G_2}{\partial z} \\
&= f\left(\frac{\partial G_3}{\partial y} - \frac{\partial G_2}{\partial z}\right) + \left(\frac{\partial f}{\partial y}G_3 - \frac{\partial f}{\partial z}G_2\right) = f(\operatorname{curl}\mathbf{G})_1 + [\operatorname{grad} f\times\mathbf{G})]_1 \\
&= [f(\operatorname{curl}\mathbf{G})_1 + \operatorname{grad} f\times\mathbf{G})]_1.
\end{aligned}$$

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. □

Two identities worth pointing out are that for any function $f : \mathbb{R}^3 \to \mathbb{R}$ and $\mathbf{F} : \mathbb{R}^3 \to \mathbb{R}^3$,

$$\operatorname{curl}(\operatorname{grad} f) = 0 \quad \text{and} \quad \operatorname{div}(\operatorname{curl} \mathbf{F}) = 0,$$

which you'll show in Exercise 5-44. In higher level mathematics this is effectively contained within the definition of divergence and curl. A very nice diagram is the following:

$$\begin{tikzcd} \begin{array}{c}\text{scalar}\\ \text{function}\end{array} \arrow[r, "\text{grad}"] & \begin{array}{c}\text{vector}\\ \text{fields}\end{array} \arrow[r, "\text{curl}"] & \begin{array}{c}\text{vector}\\ \text{fields}\end{array} \arrow[r, "\text{div}"] & \begin{array}{c}\text{scalar}\\ \text{functions}\end{array} \end{tikzcd},$$

which is (up to renaming some things) called the *de Rham complex*.

5.3 Integrating on Manifolds

We're going to integrate scalar valued functions on manifolds, and tie these into to vector valued functions later. To see how this is done, we to recall the following fact from linear algebra:

Theorem 5.30

Suppose $\mathbf{x}_1, \ldots, \mathbf{x}_k \in \mathbb{R}^n$ form the vertices of a k-parallelepiped P, and let $X = [\mathbf{x}_1 | \cdots | \mathbf{x}_k]$. The volume of P is

$$\operatorname{Vol}(P) = \sqrt{\det(X^T X)}.$$

Notably, if P is an n-parallelepiped in $\mathbb{R}^n$, then $\operatorname{Vol}(P) = |\det(X)|$.

To define the integral over M, we need a notion of infinitesimal volume on M. Manifolds generally aren't rectangles, so we deal with this the way we always do – We look at the image of rectangles in a chart of M, and use that area. Suppose then that $M \subseteq \mathbb{R}^n$ is a smooth k-manifold and $\phi : U \to V$ is a chart. For any point $\mathbf{p} \in U$, the k columns of $D\phi(\mathbf{p})$ are linearly independent, and so span a non-singular parallelepiped in $\mathbb{R}^n$. By Theorem 5.30, the volume of this parallelepiped is $\operatorname{Vol}(D\phi(\mathbf{p}))$.

Definition 5.31

Suppose $M \subseteq \mathbb{R}^n$ is a compact smooth k-manifold, and $f : M \to \mathbb{R}$ is a continuous function. Since $S = \operatorname{supp}(f)$ is compact, suppose $\phi : U \to V$ is a chart such that $S \subseteq V$ and U is bounded. We define the *integral of f over M* to be

$$\int_M f \, \mathrm{d}V = \int_{U^{\text{int}}} (f \circ \phi) \operatorname{Vol}(D\phi).$$

We call $\mathrm{d}V$ a *volume element*, and will refer to both length and area as generic volumes as well. Note that $(f \circ \phi)\operatorname{Vol}(D\phi) : \mathbb{R}^k \to \mathbb{R}$, and hence the integral on the right of Definition 5.31 is our usual integral. That this is invariant under the choice of chart follows quickly from Change of Variables, and is left to Exercise 5-22.

Unfortunately, Definition 5.31 only holds if the support of f fits within a single chart. To extend the notion to general functions, we use a partition of unity. Note if that $M \subseteq \mathbb{R}^n$ is a compact k-manifold covered by a collection of charts $\phi_i : U_i \to V_i$, we can always find a finite partition of unity subordinate to the V_i. Indeed, as each $V_i \subseteq M$ is relatively open in M, we know $V_i = M \cap \tilde{V}_i$ for some open set $\tilde{V}_i \subseteq \mathbb{R}^n$. Fix a C^1 compactly supported partition of unity $\psi_i : \tilde{V}_i \to \mathbb{R}$ subordinate to $\{\tilde{V}_i\}$. It should be clear that restricting the $\psi_i|_{V_i}$ results in a partition of unity (Exercise 3-65), and since M is compact, only finitely many are required.

Definition 5.32

Suppose $M \subseteq \mathbb{R}^n$ is a compact smooth k-manifold and $f : M \to \mathbb{R}$ is a continuous function. Fix a maximal atlas $\mathcal{A}$ on M and a finite C^1 compactly supported partition of unity $\{\psi_i : U_i \to V_i, i = 1, \ldots, m\}$ subordinate to the V_i. The *integral of f over M* is

$$\int_M f \, \mathrm{d}V = \sum_{i=1}^m \left[\int_M (\psi_i f) \, \mathrm{d}V \right].$$

A quick calculation (Exercise 5-23) demonstrates that the partition of unity does not affect the value of the integral. On the other hand, this is not a practical definition for computing. I'll state but not prove the following result, which is somewhat technical but quite reasonable to believe.

Proposition 5.33

Suppose $M \subseteq \mathbb{R}^n$ is a compact smooth k-manifold and $f : M \to \mathbb{R}$ is a continuous function. If $\phi_i : U_i \to V_i, i = 1, \ldots, m$ is a finite collection of coordinate charts such that the V_i are disjoint and cover M up to a set of measure zero[a], then

$$\int_M f \, \mathrm{d}V = \sum_{i=1}^m \int_{U_i} \left[(f \circ \phi_i) \mathrm{Vol}(D\phi_i) \right].$$

[a] By this we mean that $S \subseteq M$ is a set of measure zero if $\phi^{-1}(V_i \cap S)$ is zero measure in $\mathbb{R}^k$, for if $k < n$ then $\mu(M) = 0$ in $\mathbb{R}^n$

If M is a smooth manifold with boundary, then as boundary points are preserved under the coordinate charts, ∂M is measure zero in M and so doesn't affect the value of the integral. When $f \equiv 1$, the integral $\int_M \mathrm{d}V = \mathrm{Vol}(M)$ is the *volume* of M. To see that this makes sense, let's compute this in the special case where M is a curve or surface, and can be defined in terms of a single coordinate patch.

Example 5.34

Suppose $C \subseteq \mathbb{R}^n$ is a parameterized C^1 curve; that is, it is the image of a C^1 function $\gamma : (a, b) \to \mathbb{R}^n$. Find the volume of C.

Solution. The parameterization $\gamma(t) = (\gamma_1(t), \gamma_2(t))$ is the desired coordinate chart, for which

$$\mathrm{Vol}(D\gamma) = \sqrt{\det(D\gamma^T D\gamma)} = \left[\sum_{i=1}^n \gamma_i'(t)^2\right]^{1/2}.$$

Hence the volume of C is

$$\mathrm{Vol}(C) = \int_a^b \left[\sum_{i=1}^n \gamma_i'(t)^2\right]^{1/2} \mathrm{d}t. \tag{5.3}$$

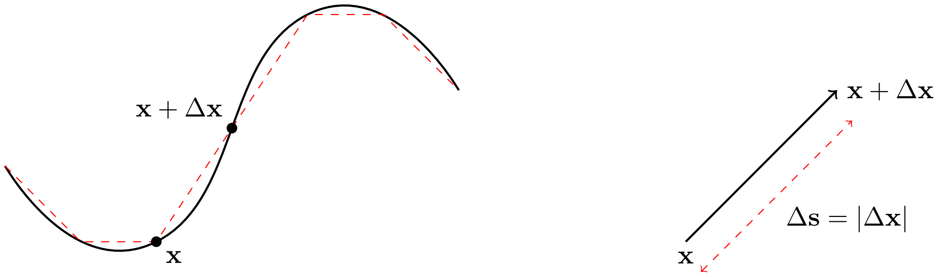

$\mathbf{x} + \Delta\mathbf{x}$
$\mathbf{x}$
$\mathbf{x} + \Delta\mathbf{x}$
$\Delta\mathbf{s} = |\Delta\mathbf{x}|$
$\mathbf{x}$

Figure 5.13: 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 $\Delta s = \|\Delta\mathbf{x}\|$ with corresponding infinitesimals $\mathrm{d}s = \|\,\mathrm{d}\mathbf{x}\|$.

The term inside of the brackets looks a lot like the Euclidean norm. Indeed, let's approximate the length of this curve by using straight line components. Fix a point $\mathbf{x} \in C$, and move along a straight line in the direction $\Delta\mathbf{x} = (\Delta x_1, \ldots, \Delta x_n)$ to the point $\mathbf{x} + \Delta\mathbf{x}$. The distance between these two points is then

$$\Delta\mathbf{s} = d(\mathbf{x}, \mathbf{x} + \Delta\mathbf{x}) = \|\Delta\mathbf{x}\|_= \sqrt{\Delta x_1^2 + \cdots + \Delta x_n^2}. \tag{5.4}$$

The total length of the curve is given by taking a limit as the $\Delta\mathbf{x} \to 0$, in which case Equation 5.4 converges to what is sometimes written as $\mathrm{d}s$ and is called an *element of arc*. When $\mathbf{x} = \gamma(t) = (\gamma_1(t), \ldots, \gamma_n(t))$,

$$\mathrm{d}s = |\,\mathrm{d}\mathbf{x}| = |g'(t)|\,\mathrm{d}t = \sqrt{\gamma_1'(t)^2 + \cdots + \gamma_n'(t)^2}\,\mathrm{d}t,$$

which is precisely the term in Equation 5.3. Hence the volume of a smooth curve C is precisely its arclength. ■

Example 5.35

Show that the circumference of a circle with radius r is precisely $2\pi r$.

Solution. Our curve in question is the circle of radius r, which we know admits a simple parametric descriptions as

$$(x, y) = g(t) = (r\cos(t), r\sin(t)), \qquad 0 \leq t \leq 2\pi.$$

The velocity is $g'(t) = (-r\sin(t), r\cos(t))$ and the speed is

$$|g'(t)| = \sqrt{r^2\sin^2(t) + r^2\cos^2(t)} = r.$$

Our arc length formula then gives

$$\text{Arclength}(C) = \int_0^{2\pi} |g'(t)|\,\mathrm{d}t = \int_0^{2\pi} r\,\mathrm{d}t = 2\pi r,$$

as required. ■

The case of surface area is somewhat more complicated, but has a very nice picture when $M \subseteq \mathbb{R}^3$.

Example 5.36

Suppose that $M \subseteq \mathbb{R}^n$ is a parameterized smooth surface; that is, there is a single regular embedding $\mathbf{G} : U \subseteq \mathbb{R}^2 \to M$. Find the volume of M, and consider the special case when $n = 3$.

Solution. Let $\mathbf{G}(u,v) = (G_1(u,v), \ldots, G_n(u,v))$ so that

$$D\mathbf{G} = \begin{bmatrix} \partial_u \mathbf{G}_1 & \partial_v \mathbf{G}_2 \\ \vdots & \vdots \\ \partial_u \mathbf{G}_n & \partial_v \mathbf{G}_n \end{bmatrix} \quad \Rightarrow \quad \det(D\mathbf{G}^T D\mathbf{G}) = \|\partial_u \mathbf{G}\|^2 \|\partial_v \mathbf{G}\|^2 - (\partial_u \mathbf{G} \cdot \partial_v \mathbf{G})^2,$$

and the volume of M is then

$$\text{Vol}(M) = \int_U \sqrt{\|\partial_u \mathbf{G}\|^2 \|\partial_v \mathbf{G}\|^2 - (\partial_u \mathbf{G} \cdot \partial_v \mathbf{G})^2}\,\mathrm{d}A. \tag{5.5}$$

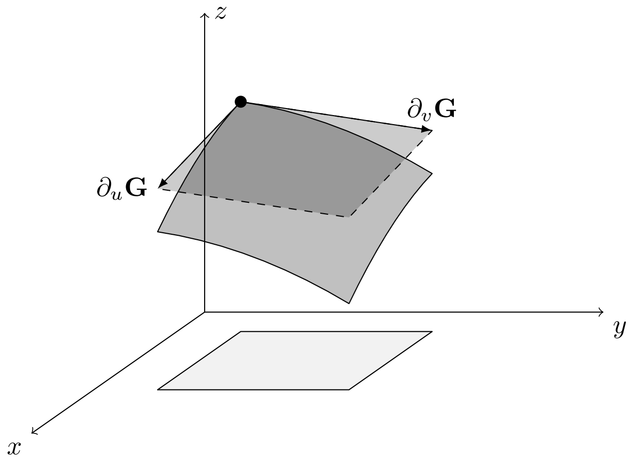


Figure 5.14: For a surface S embedded in $\mathbb{R}^2$, the parallelogram spanned by $\partial_u \mathbf{G}$ and $\partial_v \mathbf{G}$ are a first order approximation to the surface area.

Just as with arclength, let's determine the geometric interpretation. Fix some $\mathbf{u} = (u_0, v_0) \in M$ and translate by $\Delta\mathbf{u} = (\Delta u, \Delta v)$. Taking $\mathbf{G}(u,v)$ as a base point, the following displacement vectors form a parallelogram in $\mathbb{R}^3$, which we can approximate via

$$\mathbf{G}(u, v+\Delta v) - \mathbf{G}(u,v) \approx \frac{\partial \mathbf{G}}{\partial v}\Delta v \quad \text{and} \quad \mathbf{G}(u+\Delta u, v) - \mathbf{G}(u,v) \approx \frac{\partial \mathbf{G}}{\partial u}\Delta u.$$

It is known that the area of a parallelogram in $\mathbb{R}^3$ can be computed as the norm of the cross-product, giving

$$\Delta S = \left\| \frac{\partial \mathbf{G}}{\partial u} \times \frac{\partial \mathbf{G}}{\partial v} \right\| \Delta u \Delta v.$$

Limiting as $\Delta u, \Delta v \to 0$ turns ΔS into a *surface element* $\mathrm{d}S$, and we get the *surface area* of M

$$\text{Surface Area}(M) = \int_U \mathrm{d}S = \int_U \left\| \frac{\partial \mathbf{G}}{\partial u} \times \frac{\partial \mathbf{G}}{\partial v} \right\| \mathrm{d}A$$

This looks pretty different from (5.5), but it's actually the same. If $\mathbf{v}, \mathbf{w} \in \mathbb{R}^3$ we have the identity $\|\mathbf{v} \times \mathbf{w}\|^2 = \|\mathbf{v}\|^2 \|\mathbf{w}\|^2 - \langle \mathbf{v}, \mathbf{w} \rangle^2$, so that

$$\text{Surface Area}(M) = \int_U \|\partial_u \mathbf{G} \times \partial_v \mathbf{G}\| \,\mathrm{d}A = \int_U \sqrt{\|\partial_u \mathbf{G}\|^2 \|\partial_v \mathbf{G}\|^2 - (\partial_u \mathbf{G} \cdot \partial_v \mathbf{G})^2}\, \mathrm{d}A = \mathrm{Vol}(M),$$

showing that the volume of a surface agrees with its surface area. ■

The cross product $\partial_u \mathbf{G} \times \partial_v \mathbf{G}$ is not appealing to write out in coordinates, so we introduce a short hand notation. If $(x,y,z) = \mathbf{G}(u,v)$, we set

$$\frac{\partial(y,z)}{\partial(u,v)} = \frac{\partial y}{\partial u}\frac{\partial z}{\partial v} - \frac{\partial y}{\partial v}\frac{\partial z}{\partial u} \quad \text{so that} \quad \frac{\partial \mathbf{G}}{\partial u} \times \frac{\partial \mathbf{G}}{\partial v} = \left[\frac{\partial(y,z)}{\partial(u,v)}, \frac{\partial(z,x)}{\partial(u,v)}, \frac{\partial(x,y)}{\partial(u,v)} \right].$$

Example 5.37

Find the surface area of surface defined by $x^2 + y^2 + z = 25$, lying above the xy-plane.

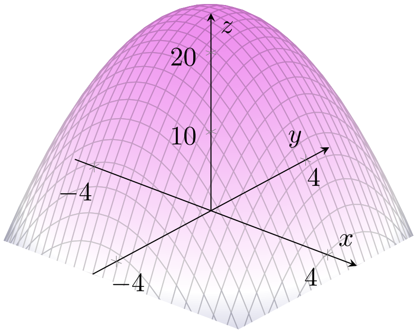


Figure 5.15: The surface described in Example 5.37.

Solution. The surface S can be written as the graph $z = 25 - x^2 - y^2$, admitting a quick parameterization $\mathbf{G}(u, v) = (u, v, 25 - u^2 - v^2)$ with $u^2 + v^2 \leq 25$. From $\mathbf{G}$ we find

$$\frac{\partial \mathbf{G}}{\partial u} = \begin{bmatrix} 1 \\ 0 \\ -2u \end{bmatrix} \quad \text{and} \quad \frac{\partial \mathbf{G}}{\partial v} = \begin{bmatrix} 0 \\ 1 \\ 2v \end{bmatrix} \quad \text{so} \quad \frac{\partial \mathbf{G}}{\partial u} \times \frac{\partial \mathbf{G}}{\partial v} = \begin{bmatrix} -2u \\ 2v \\ 1 \end{bmatrix}$$

giving a surface element is $\mathrm{d}S = \|\partial_u \mathbf{G} \times \partial_v \mathbf{G}\| = \sqrt{1 + 4u^2 + 4v^2}$. The easiest way to integrate this is going to be through polar coordinates. Set $(u, v) = \mathbf{H}(r, \theta) = (r\cos(\theta), r\sin(\theta))$. The preimage of U under $\mathbf{H}$ is quickly seen to be the rectangle $[0, 5] \times [0, 2\pi]$, and so our integral becomes

$$\begin{aligned} A(S) &= \iint_S \sqrt{1 + 4u^2 + 4v^2}\,\mathrm{d}u\,\mathrm{d}v = \int_0^{2\pi} \int_0^5 r\sqrt{1 + 4r^2}\,\mathrm{d}r\,\mathrm{d}\theta \\ &= \frac{\pi}{4} \int_1^{101} \sqrt{w}\,\mathrm{d}w = \frac{\pi}{6} u^{3/2}\Big|_1^{101} = \frac{\pi}{6}(101^{3/2} - 1), \end{aligned}$$

where in the second last equality we used the substitution $w = 1 + 4r^2$. ■

Remark 5.38 When $M \subseteq \mathbb{R}^n$ is a compact smooth n-manifold, the identity map $\phi : M \to M$ is a chart. Hence integration on M is equivalent to the usual integral.

Manifolds with Corners: If M is a compact smooth k-manifold with corners, I mentioned above that ∂M need not be a smooth manifold with corners. Since we want to integrate over such manifolds, we need to deal with this case separately. Without going into excruciating detail, the idea is that $\angle M$ should be "measure zero" within ∂M. Suppose ∂M fits within a single coordinate chart $\phi : U \subseteq \overline{\mathbb{R}^k_+} \to V \subseteq M$, and let $\hat{\phi}$ denote the restriction of ϕ to $\partial\overline{\mathbb{R}^k_+}$. Let H^k_i be as defined in (5.2). If $f : M \to \mathbb{R}$ is a continuous function, we define

$$\int_{\partial M} f\,\mathrm{d}V = \sum_{i=1}^{k} \int_{U \cap H_i} (f \circ \hat{\phi})\mathrm{Vol}(D\hat{\phi}).$$

From here, one can use partitions of unity to build to the general case when ∂M is not covered by a single chart, but in practice we're not going to worry about this.

5.4 Line Integrals

The set up is as follows: Let $\mathbf{F} : \mathbb{R}^n \to \mathbb{R}^n$ be a continuous vector field, and $C \subseteq \mathbb{R}^n$ be some oriented smooth curve. We want to integrate the vector field along this curve. To do this, let $\mathbf{T}_\mathbf{p}$ denote the unit tangent vector at the point $\mathbf{p} \in C$ defined by the orientation of C. The contribution of the field at $\mathbf{p}$ is given by $\langle \mathbf{F}(\mathbf{p}), \mathbf{T}_\mathbf{p} \rangle$ – the projection of $\mathbf{F}(\mathbf{p})$ onto the $\mathbf{T}_\mathbf{p}$ vector. The function $\langle \mathbf{F}, \mathbf{T} \rangle : C \to \mathbb{R}$ is a composition of continuous functions, and can be integrated to yield the *line integral of* $\mathbf{F}$ *along the curve* C:

$$\int_C \langle \mathbf{F}, \mathbf{T} \rangle\,\mathrm{d}V.$$

In more classical settings, this is often written

$$\int_C \mathbf{F} \cdot \mathrm{d}\mathbf{x} = \int_C (F_1 \, \mathrm{d}x_1 + \cdots + F_n \, \mathrm{d}x_n).$$

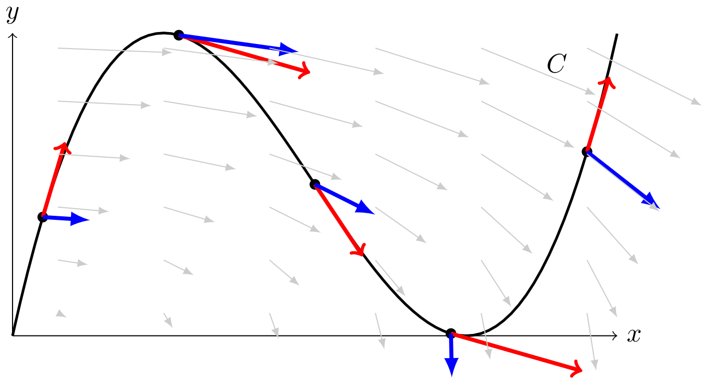


Figure 5.16: An oriented curve C passing through a vector field $\mathbf{F}$. The red arrows indicate the unit tangent vectors defining the orientation on C, while the blue arrows are the values of the vector field at each point. The line integral computes the projection of the blue arrow onto the red arrow, then integrates this function.

One way to think about this is to visualize the vector field as describing the force of a current in a stream, and the curve C as the path a fish takes swimming that stream. The integral is the amount of work the fish must do swimming that path. This should be orientation dependent: Swimming with the stream is much easier than swimming against it. The relationship between the two orientations of the curve is easy to see. Let C_1 denote the curve C with orientation given by the unit tangent $\mathbf{T}$, and C_2 the same curve with the reverse orientation $-\mathbf{T}$ at every point. Their line integrals are related via

$$\int_{C_1} \langle \mathbf{F}, -\mathbf{T} \rangle = - \int_{C_2} \langle \mathbf{F}, \mathbf{T} \rangle ,$$

where I've abused the line integral notation to make it clear which orientation is taken into consideration.

For computational purposes, suppose C is covered by a single oriented chart $\gamma : (a,b) \to \mathbb{R}^n$ up to a zero measure set. Since $\gamma'(t) \neq 0$, the orientation coincides with the unit tangent defined by

$$\mathbf{T}(t) = \frac{\gamma'(t)}{\|\gamma'(t)\|},$$

and the integral is thus

$$\int_C \langle \mathbf{F}, \mathbf{T} \rangle = \int_a^b (\langle \mathbf{F}, \mathbf{T} \rangle \circ \gamma) \operatorname{Vol}(D\gamma) = \int_a^b \left\langle \mathbf{F}(\gamma(t)), \frac{\gamma'(t)}{\|\gamma'(t)\|} \right\rangle \|\gamma'(t)\| \, \mathrm{d}t = \int_a^b \left\langle \mathbf{F}(\gamma(t)), \gamma'(t) \right\rangle \mathrm{d}t.$$

Example 5.39

Find the line integral over $\mathbf{F}(x,y,z) = (xyz, y^2, z)$ if C is the curve parameterized by $\gamma(t) = (t, t^2, t^2)$ for $0 \leq t \leq 1$.

Solution. Note that it does not matter whether we use $t \in [0,1]$ or $t \in (0,1)$, since their difference is a zero measure set. Clearly $\gamma'(t) = (1, 2t, 2t)$ and $\mathbf{F}(\gamma(t)) = \mathbf{F}(t, t^2, t^2) = (t^5, t^4, t^2)$ so their inner product yields

$$\left\langle \mathbf{F}(g(t)), g'(t) \right\rangle = (t^5, t^4, t^2) \cdot (1, 2t, 2t) = t^5 + 2t^5 + 2t^3 = 3t^5 + 2t^3.$$

Integrating gives

$$\int_0^1 \left\langle \mathbf{F}(\gamma(t)), \gamma'(t) \right\rangle \, \mathrm{d}t = \int_0^1 3t^5 + 2t^3 \, \mathrm{d}t = \frac{1}{2} \left[t^6 + t^4 \right]_0^1 = 1.$$

■

Example 5.40

Let $\mathbf{F}$ be the same vector field in Example 5.39. Evaluate the line integral of $\mathbf{F}$ over C if 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 $\gamma(t, z) = (\cos(t), \sin(t), 1)$ where $0 \leq t \leq 2\pi$. Again, it does not matter whether we use $[0, 2\pi]$ or $(0, 2\pi)$ because their difference is a measure zero set. We can compute

$$\mathbf{F}(\gamma(t)) = \big(\cos(t) \sin(t), \sin^2(t), 1 \big) \quad \text{and} \quad \gamma'(t) = (-\sin(t), \cos(t), 0),$$

$$\left\langle \mathbf{F}(\gamma(t)), \gamma'(t) \right\rangle = -\cos(t) \sin^2(t) + \cos(t) \sin^2(t) + 0 = 0$$

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 5.39 was simple because everything was written as polynomials, while Example 5.40 magically became zero before having to integrate. In general, line integrals may yield nasty integrands, necessitating that we expand our line integral toolbox.

5.4.1 Green's Theorem

Line integrals can be tricky to compute, so we'd like to develop tools to facilitate their computation. Effectively, what we'll prove is the Fundamental Theorem of Calculus, but for line integrals.

Theorem 5.41: Green's Theorem

Let $M \subseteq \mathbb{R}^2$ be a compact smooth 2-manifold (with corners) and boundary ∂M. Endow M with the natural orientation, and ∂M with the Stokes orientation. If $\mathbf{F} : M \to \mathbb{R}^2$ is a C^1-vector field, then

$$\int_{\partial M} \langle \mathbf{F}, \mathbf{T} \rangle = \iint_M \left(\frac{\partial F_2}{\partial x_1} - \frac{\partial F_1}{\partial x_2} \right).$$

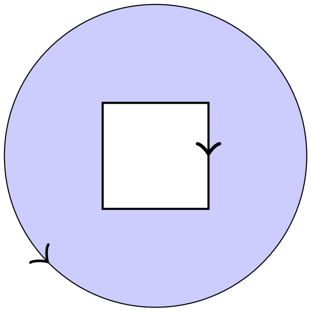


Figure 5.17: A compact smooth 2-manifold with corners M, endowed with the natural orientation. The arrows indicate the Stokes orientation or ∂M. Notice that the orientation on the internal boundary is the opposite of the external boundary.

Proof. Suppose for the moment that M can be written as both an x-simple and y-simple set:

$$S = \{(x,y) : a \leq x \leq b, f(x) \leq y \leq g(x)\} \quad \text{and} \quad S = \{(x,y) : c \leq y \leq d, r(y) \leq x \leq s(y)\}.$$

Label the edges $\partial M = C_1 + C_2 + C_3 + C_4$ as illustrated in Figure 5.18, and let's compute $\int_{\partial M} \langle \mathbf{F}, \mathbf{T} \rangle$. On the components C_1 and C_3 the orientation vectors are of the form $\mathbf{T}(\mathbf{x}) = (0, T_2(\mathbf{x}))$, while on the components C_2 and C_4 the tangent vector is of the form $\mathbf{T}(\mathbf{x}) = (T_1(\mathbf{x}), 0)$, so

$$\begin{aligned} \int_{\partial M} \langle \mathbf{F}, \mathbf{T} \rangle &= \int_{C_1} \langle \mathbf{F}, \mathbf{T} \rangle + \int_{C_2} \langle \mathbf{F}, \mathbf{T} \rangle + \int_{C_3} \langle \mathbf{F}, \mathbf{T} \rangle + \int_{C_4} \langle \mathbf{F}, \mathbf{T} \rangle A \\ &= \int_{C_1} F_2 T_2 + \int_{C_2} F_1 T_1 + \int_{C_3} F_2 T_2 + \int_{C_4} F_1 T_1. \end{aligned}$$

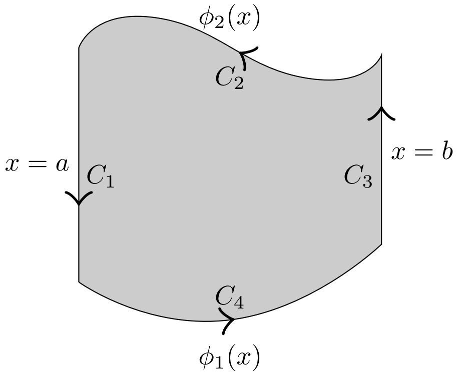


Figure 5.18: An x-simple description of our set S.

Let's focus on the integrals over C_2 and C_4. We can parameterize these as $\phi_1(t) = (t, f(t))$ and $\phi_2(t) = (t, g(t))$, with the Stokes orientation on C_2 inducing a minus sign to give

$$\int_{C_2} F_1 T_1 + \int_{C_4} F_1 T_1 = -\int_a^b F_1(t, f(t))\,\mathrm{d}t + \int_a^b F_1(t, g(t))\,\mathrm{d}t = \int_a^b [-F_1(t, f(t)) + F_2(t, g(t))]\,\mathrm{d}t. \tag{5.6}$$

On the other hand, applying the Fundamental Theorem of Calculus to the following iterated integral gives

$$\iint_S \frac{\partial F_1}{\partial y}(x,y)\,\mathrm{d}A = \int_a^b \int_{f(t)}^{g(t)} \frac{\partial F_1}{\partial y}(x,y)\,\mathrm{d}y\,\mathrm{d}x = \int_a^b \left[F_1(x,f(x)) - F_1(x,g(x))\right]\,\mathrm{d}x. \tag{5.7}$$

Comparing (5.6) and (5.7) yields

$$\int_{C_2\cup C_4} \langle \mathbf{F}, \mathbf{T}\rangle = -\iint_S \frac{\partial F_1}{\partial y}.$$

Proceeding in precisely the same manner but using x-simple description of S results in

$$\int_{C_1\cup C_3} \langle \mathbf{F}, \mathbf{T}\rangle = \iint_S \frac{\partial F_2}{\partial x}.$$

Thus combining these two results tells us that

$$\int_{\partial M} \langle \mathbf{F}, \mathbf{T}\rangle = \iint_M \left[\frac{\partial F_2}{\partial x} - \frac{\partial F_1}{\partial y}\right].$$

More generally, the remainder of the proof hinges upon the ability to decompose M into subsets which are both x- and y-simple; namely $M = M_1 \cup \cdots \cup M_n$ where the M_i have disjoint interior and are 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 ∂M_i but not of ∂M) have orientations which "cancel" each other out. By the additivity of line integrals and iterated integrals, the result then follows. □

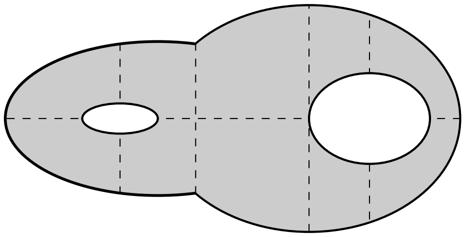


Figure 5.19: To prove Green's Theorem on more general regions, we decompose the region into subregions which are both x- and y-simple.

Example 5.42

Compute the line integral of $\mathbf{F}(x,y) = \left(2y + \sqrt{1+x^5}, 5x - e^{y^2}\right)$ over the curve C given by $x^2 + y^2 = R^2$ for some $R > 0$.

Solution. This would be a difficult integral to calculate in the absence of 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

$$\int_C \left[\underbrace{\left(2y + \sqrt{1+x^5}\right)}_{F_1} \mathrm{d}x + \underbrace{\left(5x - e^{y^2}\right)}_{F_2} \mathrm{d}y \right] = \iint_D \left(\frac{\partial F_2}{\partial x} - \frac{\partial F_1}{\partial y} \right) \mathrm{d}A$$
$$= \iint_D (5-2)\, \mathrm{d}A = 3\pi R^2.$$

We didn't even have to compute the iterated integral since we knew the area of D! ■

Example 5.43

Determine the line integral $\int_C \mathbf{F} \cdot \mathrm{d}\mathbf{x}$ where $\mathbf{F}(x,y) = (1, xy)$ and C is the triangle whose vertices are $(0,0), (1,0)$ and $(1,1)$, oriented counter clockwise.

Solution. We can write the interior of the triangle S as an x-simple set

$$T = \left\{ (x,y) \in \mathbb{R}^2 : 0 \leq x \leq 1,\, 0 \leq y \leq x \right\}.$$

The Stokes orientation is already consistent with the counter clockwise description, so Green's theorem implies

$$\int_{\partial T} \mathbf{F} \cdot \mathrm{d}\mathbf{x} = \iint_S \left[\frac{\partial F_2}{\partial x} - \frac{\partial F_1}{\partial y} \right] \mathrm{d}A = \iint_S y \,\mathrm{d}A$$
$$= \int_0^1 \int_0^x y \,\mathrm{d}y \,\mathrm{d}x = \frac{1}{6}.$$

Let's compute the line integral explicitly and verify that we get the same result. Let C_1 be the portion of T lying on the x-axis, parameterized as $g_1(t) = (t, 0)$ for $0 \leq t \leq 1$. The line integral over C_1 is

$$\int_{C_1} \mathbf{F} \cdot \mathrm{d}\mathbf{x} = \int_0^1 \mathbf{F}(g_1(t)) \cdot g_1'(t) \,\mathrm{d}t = \int_0^1 (1,0) \cdot (1,0) \,\mathrm{d}t = 1$$

Let C_2 we the portion of T on the line $x = 1$, for which we use $g_2(t) = (1, t)$ for $0 \leq t \leq 1$ to get

$$\int_{C_2} \mathbf{F} \cdot \mathrm{d}\mathbf{x} = \int_0^1 (1,t) \cdot (0,1) \,\mathrm{d}t = \int_0^1 t \,\mathrm{d}t = \frac{1}{2}.$$

Finally, set C_3 to be the portion of T lying on the line $y = x$, parameterized as $g_3(t) = (1-t, 1-t)$ for $0 \leq t \leq 1$ (we choose this over $g_3(t) = (t,t)$ to keep the correct orientation), so that

$$\int_{C_3} \mathbf{F} \cdot \mathrm{d}\mathbf{x} = \int_0^1 (1, (1-t)^2) \cdot (-1,-1) \,\mathrm{d}t$$
$$= \int_0^1 -1 - (1-t)^2 \,\mathrm{d}t = -\frac{4}{3}.$$

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 have some surprising properties; for example, line integrals can be used to tell you something about the geometry of a surface. We'll set up the ground work for that study here.

5.5.1 Exact Vector Fields

Our first result is a version of the Fundamental Theorem of Calculus:

Theorem 5.44: Fundamental Theorem of Calculus for Line Integrals

If $C \subseteq \mathbb{R}^n$ is a smooth oriented curve with boundary points $\mathbf{a}$ and $\mathbf{b}$ (in that order), and $\mathbf{F} : \mathbb{R}^n \to \mathbb{R}^n$ is a vector field such that $\mathbf{F} = \nabla f$ for some C^1 function $f : \mathbb{R}^n \to \mathbb{R}$, then

$$\int_C \mathbf{F} \cdot \mathrm{d}\mathbf{x} = f(\mathbf{b}) - f(\mathbf{a}).$$

In particular, the integral only depends on the endpoints $\mathbf{a}$ and $\mathbf{b}$ of the curve C.

Proof. Assume that $\mathbf{F} = \nabla f$ and let C be some oriented curve with parameterization $\gamma : [0,1] \to \mathbb{R}^n$, so that $\gamma(0) = \mathbf{a}$ and $\gamma(1) = \mathbf{b}$. Straightforward computation then reveals that

$$\begin{aligned}
\int_C \mathbf{F} \cdot \mathrm{d}\mathbf{x} &= \int_0^1 \mathbf{F}(\gamma(t)) \cdot \gamma'(t)\,\mathrm{d}t = \int_0^1 \nabla f(\gamma(t)) \cdot \gamma'(t)\,\mathrm{d}t && \text{by assumption} \\
&= \int_0^1 \frac{\mathrm{d}}{\mathrm{d}t} f(\gamma(t))\,\mathrm{d}t && \text{the Chain Rule} \\
&= f(\gamma(1)) - f(\gamma(0)) = f(\mathbf{b}) - f(\mathbf{a}) && \text{the Fundamental Theorem of Calculus.}
\end{aligned}$$

This is precisely what we wanted to show. □

We know that the choice of curve generally makes a difference to the value of the line integral, so Theorem 5.44 tells us 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 5.45

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 referred to as a *scalar potential*.

Example 5.46

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 $F_1 = \partial_1 f$. We thus integrate the first 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. You can quickly check that $f(x,y,z) = e^{xyz}$ gives $\nabla f = \mathbf{F}$.
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 $\partial_2 f = G_2$ 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 $\partial_3 f = -y\sin(z)$ 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 $f(x,y,z) = x^2/2 + yx + g(y,z)$. Differentiating with respect to y gives $\partial_2 f(x,y,z) = x + \partial_2 g(x,y,z)$. Equating to H_2 tells us that $\partial g(x,y,z) = z$, so that $f(x,y,z) = x^2/2 + yx + yz + h(z)$. Finally, $\partial_3 f(x,y,z) = y + \partial h(x,y,z) = H_3(x,y,z) = y + z$, so it must be the case that $\partial_3 h(x,y,z) = z$, and we conclude that
$$f(x,y,z) = \frac{1}{2}x^2 + yx + yz + \frac{1}{2}z^2. \qquad \blacksquare$$

Example 5.47

Determine the line integral $\int_C \mathbf{F} \cdot d\mathbf{x}$ 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 resulting integral is intractable. Instead, all one needs to realize is that the endpoints of this curve are $(\pm 1, 0, 0)$. Furthermore, in Example 5.46 we showed that $\mathbf{F} = \nabla f$ where $f(x, y, z) = x^2 y + y\cos(z)$. Consequently, the line integral is just

$$\int_C \mathbf{F} \cdot \,\mathrm{d}\mathbf{x} = f(1, 0, 0) - f(-1, 0, 0) = 0. \qquad ■$$

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. We say that a smooth curve $C \subseteq \mathbb{R}^n$ is a *closed curve* if it is closed in the subspace topology of $\mathbb{R}^n$, but has empty boundary; namely, it is a loop. To this end, we have the following lemma:

Lemma 5.48

If $U \subseteq \mathbb{R}^n$ is an open set, and $\mathbf{F} : U \to \mathbb{R}^n$ is a continuous vector field, 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

$$0 = \int_C \mathbf{F} \cdot \,\mathrm{d}\mathbf{x} = \int_{C_1} \mathbf{F} \cdot \,\mathrm{d}\mathbf{x} - \int_{C_2} \mathbf{F} \cdot \,\mathrm{d}\mathbf{x},$$

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 the energy difference and the system ignores anything else which happens 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 5.49

If $U \subseteq \mathbb{R}^n$ is an open set, then a continuous vector field $\mathbf{F} : U \to \mathbb{R}^n$ is conservative if and only if there is a C^1 function $f : U \to \mathbb{R}$ such that $\mathbf{F} = \nabla f$. More concisely, conservative and exact vector fields are the same thing.

Proof. ($\Leftarrow$) This follows from the Fundamental Theorem of Calculus for Line Integrals.

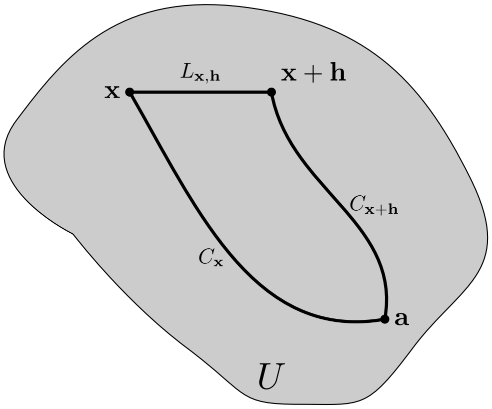


Figure 5.20: The scalar potential $f(\mathbf{x})$ is given by finding the line integral over any curve from a fixed point $\mathbf{a}$ to $\mathbf{x}$.

($\Rightarrow$) Conversely, assume that $\mathbf{F} : U \to \mathbb{R}^n$ is a conservative vector field. Without loss of generality, we may assume that U is connected and hence path connected; otherwise perform the following operations on each connected component. Fix some point $\mathbf{a} \in U$ and for each $\mathbf{x} \in U$ let $C_{\mathbf{x}}$ be a curve from $\mathbf{a}$ to $\mathbf{x}$. Define the function

$$f(\mathbf{x}) = \int_{C_{\mathbf{x}}} \mathbf{F} \cdot d\mathbf{x}.$$

This is well defined since, by assumption, the definition is invariant of our choice of curve $C_{\mathbf{x}}$. I claim f is a C^1 function and is the scalar potential for $\mathbf{F}$; namely, $\nabla f = \mathbf{F}$. Both claims will follow if we show that $\partial_i f = F_i$ for each $i = 1, \ldots, n$.

To this end, let $i \in \{1, \ldots, n\}$. Fix $\mathbf{x} \in U$ and choose $h > 0$ sufficiently small so that if $\mathbf{h} = h\mathbf{e}_i$, then the line $L_{\mathbf{x},\mathbf{h}}$ between $\mathbf{x}$ and $\mathbf{x} + \mathbf{h}$ remains in U (Figure 5.20). Let $C_{\mathbf{x}+\mathbf{h}}$ be $C_{\mathbf{x}}$ followed by $L_{\mathbf{x},\mathbf{h}}$ so that

$$\begin{aligned} f(\mathbf{x}+\mathbf{h}) &= \int_{C_{\mathbf{x}+\mathbf{h}}} \mathbf{F} \cdot d\mathbf{x} = \int_{C_{\mathbf{x}}} \mathbf{F} \cdot d\mathbf{x} + \int_{L_{\mathbf{x},\mathbf{h}}} \mathbf{F} \cdot d\mathbf{x} \\ &= f(\mathbf{x}) + \int_{L_{\mathbf{x},\mathbf{h}}} \mathbf{F} \cdot d\mathbf{x}. \end{aligned}$$

Parameterize the line $L_{\mathbf{x},\mathbf{h}}$ by $g(t) = \mathbf{x} + t\mathbf{e}_i$ for $0 \leq t \leq h$ so that $g'(t) = \mathbf{e}_i$ and

$$\begin{aligned}\int_{L_{\mathbf{x},\mathbf{h}}} \mathbf{F} \cdot \mathrm{d}\mathbf{x} &= \int_0^h \mathbf{F}(x_1, \ldots, x_{i-1}, x_i + t, x_{i+1}, \ldots, x_n) \cdot (0, \ldots, 0, 1, 0, \ldots, 0)\, dt \\ &= \int_0^h F_i(x_1, \ldots, x_i + t, \ldots, x_n)\, \mathrm{d}t.\end{aligned}$$

Computing $\partial_i f$ we have

$$\begin{aligned}\frac{\partial f}{\partial x_i}(\mathbf{x}) = \lim_{h\to 0} \frac{f(\mathbf{x} + h\mathbf{e}_i) - f(\mathbf{x})}{h} &= \lim_{h\to 0} \frac{1}{h} \int_{L_{\mathbf{x},\mathbf{h}}} \mathbf{F} \cdot \mathrm{d}\mathbf{x} \\ &= \lim_{h\to 0} \frac{1}{h} \int_0^h F_i(x_1, \ldots, x_i + t, \ldots, x_n)\, \mathrm{d}t \\ &= F_i(\mathbf{x}).\end{aligned}$$

Hence f is differentiable and is the scalar potential for $\mathbf{F}$. Since $\mathbf{F}$ is continuous, it also follows that f is C^1. □

5.5.3 Closed Vector Fields

Theorem 5.49 is a nice condition, but it is intractable to compute all possible line integrals, and it can be difficult to ascertain whether a vector field is the gradient of a function. We aim to create a litmus test for testing the feasibility of an exact vector field. If $\mathbf{F}$ is a C^1 vector field with $\mathbf{F} = \nabla f$, then $F_i = \partial_i f$. Since mixed partials commute by Clairut's Theorem, we have

$$\partial_i F_j = \partial_i \partial_j f = \partial_j \partial_i f = \partial_j F_i,$$

or alternatively

$$\frac{\partial F_i}{\partial x_j} - \frac{\partial F_j}{\partial x_i} = 0, \qquad i \neq j. \tag{5.8}$$

Vector fields which satisfy (5.8) are called *closed vector fields*. If we are working in $\mathbb{R}^3$, the components of (5.8) correspond to those of the curl. Hence closed vector fields of $\mathbb{R}^3$ are irrotational.

By construction, all exact vector fields are closed, but the question remains whether all closed vector fields are exact. The answer is no, and that in effect we can detect the number of "holes" in an n-manifold $M \subseteq \mathbb{R}^n$ by looking at how many closed vector fields fail to be exact.

Example 5.50

Show that the vector field $\mathbf{F} : \mathbb{R}^2 \setminus \{(0,0)\} \to \mathbb{R}^2$ given by

$$F(x,y) = \frac{1}{x^2 + y^2}(-y, x)$$

Is closed but not exact.

Solution. Showing that $\mathbf{F}$ is closed amounts to the computation

$$\begin{aligned} \frac{\partial F_2}{\partial x} &= \frac{\partial}{\partial x}\frac{x}{x^2+y^2} = \frac{y^2-x^2}{(x^2+y^2)^2}, \\ \frac{\partial F_1}{\partial y} &= \frac{\partial}{\partial y}\frac{-y}{x^2+y^2} = \frac{y^2-x^2}{(x^2+y^2)^2}, \end{aligned}$$

so that $\partial_1 F_2 = \partial_2 F_1$, showing that $\mathbf{F}$ is a closed vector field.

To see that $\mathbf{F}$ is not exact, we'll show that the line integral over $\mathbf{F}$ of a closed curve is non-zero. Let C be any circle containing the origin, say parameterized by $\gamma(\theta) = (r\cos(\theta), r\sin(\theta))$ for some $r > 0$. We know $\gamma'(t) = (-r\sin(\theta), r\cos(\theta))$ 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? The vector field $\mathbf{F}$ is C^1 because of the hole at the origin $(0,0)$, and our 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 5.51

A set $U \subseteq \mathbb{R}^n$ is said to be *star-convex* if there exists a point $\mathbf{a} \in U$ such that for every point $\mathbf{x} \in U$ the straight line connected $\mathbf{x}$ to $\mathbf{a}$ is contained in U.

Every convex set is star shaped, though the converse need not be true. For example, Figure 5.21 gives an example of a star shaped set in $\mathbb{R}^2$ that is not convex.

Theorem 5.52: 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, assume that U is star shaped about the origin. For any $\mathbf{x} \in U$ let $\gamma_{\mathbf{x}}(t) = t\mathbf{x}$ be the straight line connecting the origin to $\mathbf{x}$, and define the function

$$f(\mathbf{x}) = \int_{\gamma_{\mathbf{x}}} \mathbf{F}\cdot \mathrm{d}\mathbf{x} = \int_0^1 F_1(t\mathbf{x})x_1 + \cdots + F_n(t\mathbf{x})x_n\,\mathrm{d}t.$$

This is well defined since $\gamma_{\mathbf{x}}(t) \in U$ for all t, and there is a unique straight line connecting 0 to $\mathbf{x}$.

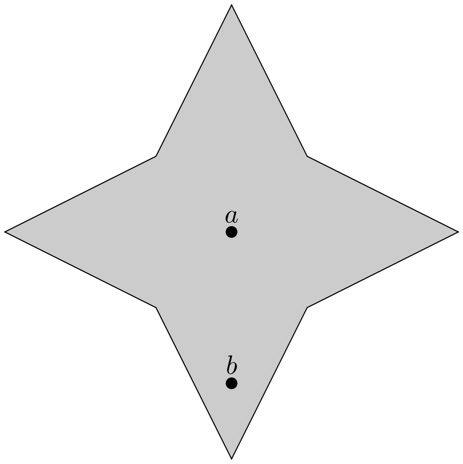


Figure 5.21: An example of a star shaped set which is not convex. The point a satisfies the required definition, while the point b does not.

I claim that $\mathbf{F} = \nabla f$ on U. Inspecting one component at a time, we have

$$\begin{aligned}\frac{\partial f}{\partial x_k}(\mathbf{x}) &= \int_0^1 \left[\sum_{i=1}^n \frac{\partial F_i}{\partial x_k}(t\mathbf{x})tx_i + F_k(t\mathbf{x})\right] \mathrm{d}t \\ &= \int_0^1 \left[\sum_{i=1}^n \frac{\partial F_k}{\partial x_i}(t\mathbf{x})tx_i + F_k(t\mathbf{x})\right] \mathrm{d}t \qquad \text{since } F \text{ is closed} \\ &= \int_0^1 \frac{\mathrm{d}}{\mathrm{d}t}\left[tF_k(t\mathbf{x})\right] \mathrm{d}t = F_k(\mathbf{x}).\end{aligned}$$

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 *flux* of a vector field passing through a surface. In this section, we'll assume our surfaces are always embedded in $\mathbb{R}^3$.

5.6.1 Surface Integrals over Vector Fields

As with line integrals, surface integrals are going to depend on a choice of orientation. Recall that if $S \subseteq \mathbb{R}^3$ is a compact smooth surface, an *orientation* of S is a consistent choice of normal vector to the surface. Of particular use is that if $\mathbf{G} : U \to V$ is a chart of S, then $\{\partial_u \mathbf{G}, \partial_v \mathbf{G}\}$ forms a basis for the tangent space, so the normal vectors at a point $\mathbf{p}$ are $\pm\mathbf{n}_\mathbf{p} = \pm[\partial_u \mathbf{G}(\mathbf{p}) \times \partial_v \mathbf{G}(\mathbf{p})]$. The *unit normal vector* $\hat{\mathbf{n}}_\mathbf{p}$ is

$$\hat{\mathbf{n}}_\mathbf{p} = \frac{\partial_u \mathbf{G}(\mathbf{p}) \times \partial_v \mathbf{G}(\mathbf{p})}{\|\partial_u \mathbf{G}(\mathbf{p}) \times \partial_v \mathbf{G}(\mathbf{p})\|}, \tag{5.9}$$

and an orientation is specified by one of $\pm\hat{\mathbf{n}}_\mathbf{p}$.

The idea of a surface integral is thus the following: Given a vector field $\mathbf{F} : S \to \mathbb{R}^3$, we want to compute the *flux* of the vector field through the surface. If we 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 $\hat{\mathbf{n}}_{\mathbf{p}}$ is given by $\langle \mathbf{F}(\mathbf{p}), \hat{\mathbf{n}}_{\mathbf{p}} \rangle$. This is a continuous function in $\mathbf{p}$, so the flux integral is given by integrating over all of S:

$$\int_S \langle \mathbf{F}, \hat{\mathbf{n}} \rangle \, \mathrm{d}V.$$

Of course, this is not easily computed without a parameterization. If $\mathbf{G} : U \subseteq \mathbb{R}^2 \to V \subseteq S$ is local chart, we can use (5.9) to write the flux integral as

$$\begin{aligned} \int_S \langle \mathbf{F}, \hat{\mathbf{n}} \rangle \, \mathrm{d}V &= \int_U (\langle \mathbf{F}, \hat{\mathbf{n}} \rangle \circ \mathbf{G}) \operatorname{Vol}(D\mathbf{G}) \\ &= \int_U \left\langle \mathbf{F}(\mathbf{G}(u,v)), \frac{\partial_u \mathbf{G}(u,v) \times \partial_v \mathbf{G}(u,v)}{\|\partial_u \mathbf{G}(u,v) \times \partial_v \mathbf{G}(u,v)\|} \right\rangle \|\partial_u \mathbf{G}(u,v) \times \partial_v \mathbf{G}(u,v)\| \, \mathrm{d}A \\ &= \int_U \langle \mathbf{F}(\mathbf{G}(u,v)), \partial_u \mathbf{G}(u,v) \times \partial_v \mathbf{G}(u,v) \rangle \, \mathrm{d}A. \end{aligned}$$

Notationally, the surface integral is sometimes written as

$$\int_S \mathbf{F} \cdot \hat{\mathbf{n}} \, \mathrm{d}A.$$

Example 5.53

Evaluate the flux of $\mathbf{F}(x,y,z) = (x^2+y, y^2z, x^2y)$ through the surface $S = [0,1]\times[0,1]\times\{0\} \subseteq \mathbb{R}^3$, oriented pointing in the positive z-direction.

Solution. The surface is quickly parameterized as

$$\mathbf{G}(s,t) = (s,t,0) \quad \text{for} \quad 0 \leq s \leq 1, 0 \leq t \leq 1,$$

from which we find that

$$\frac{\partial \mathbf{G}}{\partial s} \times \frac{\partial \mathbf{G}}{\partial t} = \begin{vmatrix} i & j & k \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{vmatrix} = (0,0,1).$$

This is oriented in the correct direction, so we proceed with the surface integral to get

$$\begin{aligned} \iint_S \mathbf{F} \cdot \hat{\mathbf{n}} \, \mathrm{d}A &= \int_0^1 \int_0^1 (s^2+t, t^2, s^2t^2) \cdot (0,0,1) \, \mathrm{d}s \, \mathrm{d}t \\ &= \int_0^1 \int_0^1 s^2t^2 \, \mathrm{d}s \, \mathrm{d}t = \frac{1}{9}. \end{aligned}$$ ■

Sometimes it is necessary to break our surfaces into several pieces in order to determine the integral, as the next example demonstrates.

Example 5.54

Evaluate the flux 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, and $\mathbf{F}(x, y, z) = (0, y, -z)$

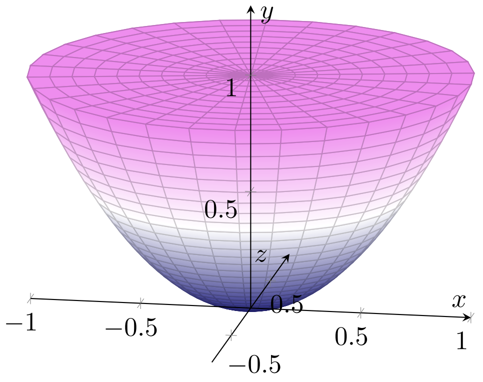


Figure 5.22: The surface S for Example 5.54.

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 $0 \leq \theta \leq 2\pi$, so that

$$\frac{\partial \mathbf{G}}{\partial r} = \begin{bmatrix} \cos(\theta) \\ 2r \\ \sin(\theta) \end{bmatrix} \quad \text{and} \quad \frac{\partial \mathbf{G}}{\partial \theta} = \begin{bmatrix} -r\sin(\theta) \\ 0 \\ r\cos(\theta) \end{bmatrix} \quad \text{so} \quad \frac{\partial \mathbf{G}}{\partial r} \times \frac{\partial \mathbf{G}}{\partial \theta} = \begin{bmatrix} 2r^2\cos(\theta) \\ -r \\ 2r^2\sin(\theta) \end{bmatrix}.$$

The y-component is negative, which it should be for the normal vector to be pointing outwards. Hence we have the correct orientation, and

$$\begin{aligned} \mathbf{F}(\mathbf{G}(r,\theta)) \cdot \left[\frac{\partial \mathbf{G}}{\partial r} \times \frac{\partial \mathbf{G}}{\partial \theta} \right] &= (0, r^2, -r\sin(\theta)) \cdot (2r^2\cos(\theta), -r, 2r^2\sin(\theta)) \\ &= -r^3 \left(1 + 2\sin^2(\theta) \right), \end{aligned}$$

which we integrate to give

$$\iint_S \mathbf{F} \cdot \hat{\mathbf{n}} \, \mathrm{d}A = - \left[\int_0^1 r^3 \, \mathrm{d}r \right] \left[\int_0^{2\pi} 1 + 2\sin^2(\theta) \, \mathrm{d}\theta \right] = -\pi.$$

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, the normal vector should always points in the positive y-direction. You can verify that

$$\mathbf{G}(r, \theta) = (r\cos(\theta), 1, r\sin(\theta)) \quad \text{for} \quad 0 \leq r \leq 1, 0 \leq \theta \leq 2\pi$$

satisfies

$$\frac{\partial G}{\partial r} \times \frac{\partial G}{\partial \theta} = (0, -r, 0),$$

so that this is 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. ■

5.6.2 The Divergence Theorem

The Divergence Theorem – also known as Gauss' Theorem – is the analog of Green's theorem for surface integrals.

Theorem 5.55: Divergence Theorem

Let $M \subseteq \mathbb{R}^3$ be a compact 3-manifold (with corners) and boundary ∂M. If $\mathbf{F} : M \to \mathbb{R}^3$ is a C^1 vector field and ∂M has the Stokes orientation with respect to the natural orientation on M, then

$$\iint_{\partial M} \mathbf{F} \cdot \hat{\mathbf{n}} \, \mathrm{d}A = \iiint_M \operatorname{div} \mathbf{F} \, \mathrm{d}V.$$

Proof. As with Green's Theorem, I will only provide a simplified proof which captures the idea of the Divergence Theorem. It suffices to show that

$$\iint_{\partial M} F_i \langle \mathbf{e}_i, \hat{\mathbf{n}} \rangle \, \mathrm{d}A = \iiint_M \frac{\partial F_i}{\partial x_i} \, \mathrm{d}V \quad \text{for} \quad i = 1, 2, 3.$$

Let's focus on $i = 3$. Assume that M is an z-simple set, written as

$$M = \left\{ (x, y, z) \in \mathbb{R}^3 : (x, y) \in K, \psi_1(x, y) \leq z \leq \psi_2(x, y) \right\}$$

for some compact region $K \in \mathbb{R}^2$ and C^1 functions ψ_1 and ψ_2. We can write $\partial M = S_1 \cup S_2 \cup S_3$, where $S_i = \{(x, y, z) : (x, y) \in K, z = \psi_i(x, y)\}$ for $i = 1, 2$ are the top and bottom of ∂M, and S_3 is the vertical component. Note that $\mathbf{e}_3 \cdot \hat{\mathbf{n}} = 0$ along S_3, while $\hat{\mathbf{e_3}}$ is consistent with the orientation of the top surface S_2 and is the opposite orientation of the bottom surface S_1. Both S_1 and S_2 admit straightforward parameterizations $\mathbf{G}_i(s, t) = (s, t, \psi_i(s, t))$, so that

$$\partial_s \mathbf{G}_i(s, t) \times \partial_t \mathbf{G}_i(s, t) = (-\partial_s \psi_i(s, t), -\partial_t \psi_i(s, t), 1)^T,$$

and the surface integrals then become

$$\begin{aligned}
\iint_{\partial M} F_3 \left\langle \hat{\mathbf{e}_3}, \hat{\mathbf{n}} \right\rangle \, \mathrm{d}A &= \iint_{S_2} F_3 \left\langle \hat{\mathbf{e}_3}, \hat{\mathbf{n}} \right\rangle \, \mathrm{d}A + \iint_{S_1} F_3 \left\langle \hat{\mathbf{e}_3}, \hat{\mathbf{n}} \right\rangle \, \mathrm{d}A \\
&= \iint_K \left[F_3(x, y, \psi_2(x, y)) - F_3(x, y, \psi_1(x, y)) \right] \, \mathrm{d}x \, \mathrm{d}y \\
&= \iint_K \int_{\psi_1(x,y)}^{\psi_2(x,y)} \frac{\partial F_3}{\partial x_3}(x, y, z) \, \mathrm{d}z \\
&= \iiint_M \frac{\partial F_3}{\partial x_3} \, \mathrm{d}V.
\end{aligned}$$

If M admits an x-simple and a y-simple decomposition, then the same result holds for $i = 1$ and $i = 2$, and the sum of the three components gives the theorem. Finally, one concludes that any smooth 3-manifold with corners can be decomposed into a union of such pieces. □

Example 5.56

Evaluate the flux of $\mathbf{F}(x, y, z) = (y^2 z, y^3, xz)$ through the surface S defined to be the boundary of the cube

$$C = \left\{ (x, y, z) \in \mathbb{R}^3 : -1 \leq x \leq 1, -1 \leq y \leq 1, 0 \leq z \leq 2 \right\},$$

oriented so that the normal points outwards.

Solution. This would normally be a tedious exercise: 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. However, with the Divergence Theorem it becomes straightforward. Noting that $\operatorname{div} \mathbf{F}(x, y, z) = 3y^2 + x$, we get

$$\begin{aligned}
\iint_S \mathbf{F} \cdot \hat{\mathbf{n}} \, \mathrm{d}A = \iiint_C \operatorname{div} \mathbf{F} \, \mathrm{d}V &= \int_{-1}^{1} \int_{-1}^{1} \int_0^2 \left[3y^2 + x \right] \, \mathrm{d}z \, \mathrm{d}y \, \mathrm{d}x \\
&= 2 \int_{-1}^{1} \left[y^3 + xy \right]_{y=-1}^{y=1} \, \mathrm{d}x = 4 \int_{-1}^{1} [1 + x] \, \mathrm{d}x = 8.
\end{aligned}$$ ■

Example 5.57

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

$$S = \left\{ (x, y, z) \in \mathbb{R}^3 : x^2 + y^2 + z = 4 \right\} \cup \left\{ (x, y, z) \in \mathbb{R}^3 : x^2 + y^2 \leq 4, z = 0 \right\}.$$

Solution. This is an almost impossible exercise to approach from the definition, so instead we use the Divergence Theorem. One can easily compute that

$$\operatorname{div}\mathbf{F}(x,y,z) = \big(z\sin(yz) + 3x^3\big) + (-z\sin(yz)) + (3y^2) = 3x^3 + 3y^2.$$

Hence if V is the filled paraboloid so that $\partial V = S$ then our surface integral becomes

$$\iint_S \mathbf{F}\cdot\hat{\mathbf{n}}\,\mathrm{d}A = \iiint_V (3x^2+3y^2)\,\mathrm{d}V.$$

To determine this integral, notice we can write

$$\iiint_V (3x^2+3y^2)\,\mathrm{d}V = \iint_D \int_0^{4-x^2-y^2} (3x^2+3y^2)\,\mathrm{d}z\,\mathrm{d}A$$

where D is the unit disk. Changing to polar coordinates (or cylindrical if we skip the previous step) gives

$$\int_0^2\int_0^{2\pi}\int_0^{4-r^2} 3r^3\,\mathrm{d}z\,\mathrm{d}\theta\,\mathrm{d}r = 6\pi\int_0^2 r^3(4-r^2)\,\mathrm{d}r\,\mathrm{d}\theta = 6\pi\left(16-\frac{64}{6}\right) = 32\pi.$$ ■

5.6.3 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 discuss this version of the theorem here. Hence I 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.

Theorem 5.58: Stokes' Theorem

Let $M \subseteq \mathbb{R}^3$ be a compact, oriented smooth surface with corners, and give its boundary ∂M the induced orientation. If $\mathbf{F} : M \to \mathbb{R}^3$ is a C^1 vector field, then

$$\int_{\partial M} \mathbf{F}\cdot\,\mathrm{d}\mathbf{x} = \iint_M (\operatorname{curl}\mathbf{F})\cdot\hat{\mathbf{n}}\,\mathrm{d}A.$$

Proof. If M is just a region in the xy-plane, then $\hat{\mathbf{n}} = (0,0,1)$ and

$$(\operatorname{curl}\mathbf{F})\cdot\hat{\mathbf{n}} = \frac{\partial F_2}{\partial x_1} - \frac{\partial F_1}{\partial x_2}.$$

Hence Green's Theorem gives

$$\int_{\partial M} \mathbf{F}\cdot\,\mathrm{d}\mathbf{x} = \iint_M \left[\frac{\partial F_2}{\partial x_1} - \frac{\partial F_1}{\partial x_2}\right]\,\mathrm{d}A,$$

showing that Stokes' theorem in the xy-plane is just Green's theorem.

Now assume that M is a surface which does not live in one of the coordinate planes, and let $\mathbf{G} : U \subseteq \mathbb{R}^2 \to M$ be a parameterization of M. Assume that $\mathbf{G}$ gives an orientation which coincides with the orientation of M (if $\mathbf{G}(u,v)$ gives the opposite orientation, just switch the roles of u and

v). Since the boundaries are preserved under $\mathbf{G}$, and Stokes' theorem is just Green's theorem, we will "pullback" the calculation to the uv-plane and apply Green's Theorem. As always, we'll do this component by component. Take $\mathbf{F} = (F_1, 0, 0)$, so that the proof amounts to showing

$$\int_{\partial M} F_1 \, \mathrm{d}x = \iint_M \left(0, \frac{\partial F_1}{\partial x_3}, -\frac{\partial F_1}{\partial x_2}\right) \cdot \hat{\mathbf{n}} \; \mathrm{d}A. \tag{5.10}$$

Applying our parameterization, the right hand side becomes

$$\begin{aligned}\iint_M \left(0, \frac{\partial F_1}{\partial x_3}, -\frac{\partial F_1}{\partial x_2}\right) \cdot \hat{\mathbf{n}} \; \mathrm{d}A &= \iint_U \left(0, \frac{\partial F_1}{\partial x_3}, -\frac{\partial F_1}{\partial x_2}\right) \cdot \left(\frac{\partial \mathbf{G}}{\partial u} \times \frac{\partial \mathbf{G}}{\partial v}\right) \mathrm{d}A \\ &= \iint_U \left(\frac{\partial F_1}{\partial x_3}\frac{\partial(z,x)}{\partial(u,v)} - \frac{\partial F_1}{\partial x_2}\frac{\partial(x,y)}{\partial(u,v)}\right) \mathrm{d}A.\end{aligned}$$

The other side is trickier. Let $\gamma : \tilde{U} \subseteq \mathbb{R} \to \mathbb{R}^2$ parameterize ∂U with an orientation so that ∂M is parameterized as $\mathbf{G}(\gamma(t))$, and this has the induced orientation. The line integral becomes

$$\begin{aligned}\int_{\partial M} \mathbf{F} \cdot \mathrm{d}\mathbf{x} &= \int_{\partial U} \mathbf{F}(\mathbf{G}(\gamma(t))) \cdot \left[\frac{\mathrm{d}}{\mathrm{d}t}\mathbf{G}(\gamma(t))\right] \mathrm{d}t \\ &= \int_{\partial U} \mathbf{F}(\mathbf{G}(\gamma(t))) \cdot D\mathbf{G}(\gamma(t))\gamma'(t)\, \mathrm{d}t \\ &= \int_{\partial U} F_1(\mathbf{G}(\gamma(t))) \left[\frac{\partial G_1}{\partial u}(\gamma(t))\gamma_1'(t) + \frac{\partial G_1}{\partial v}(\gamma(t))\gamma_2'(t)\right] \mathrm{d}t.\end{aligned}$$

Now here is where we pullback the computation to $\mathbb{R}^2$. Let $\mathbf{H} : U \to \mathbb{R}^2$ be given by

$$\mathbf{H}(\mathbf{x}) = F_1(\mathbf{G}(\mathbf{x})) \left(\frac{\partial G_1}{\partial u}(\mathbf{x}), \frac{\partial G_1}{\partial v}(\mathbf{x})\right),$$

so that the line integral over ∂U, still parameterized by γ_1, is

$$\int_{\partial U} \mathbf{H} \cdot \mathrm{d}\mathbf{x} = \int_{\partial U} F_1(\mathbf{G}(\gamma(t))) \left[\frac{\partial G_1}{\partial u}(\gamma(t))\gamma_1'(t) + \frac{\partial G_1}{\partial v}(\gamma(t))\gamma_2'(t)\right] \mathrm{d}t = \int_{\partial M} \mathbf{F} \cdot \mathrm{d}\mathbf{x}.$$

Applying Green's theorem to the left hand side gives

$$\begin{aligned}\int_{\partial U} \mathbf{H} \cdot \mathrm{d}\mathbf{x} = \iint_U \left[\frac{\partial H_2}{\partial u} - \frac{\partial H_1}{\partial v}\right] \mathrm{d}A &= \iint_U \left(\frac{\partial}{\partial u}\left[(F_1 \circ \mathbf{G})\frac{\partial G_1}{\partial v}\right] - \frac{\partial}{\partial v}\left[(F_1 \circ \mathbf{G})\frac{\partial G_1}{\partial u}\right]\right) \mathrm{d}A \\ = \iint_U \Bigg[\left(\frac{\partial F_1}{\partial x}\frac{\partial G_1}{\partial u} + \frac{\partial F_1}{\partial y}\frac{\partial G_2}{\partial u} + \frac{\partial F_1}{\partial z}\frac{\partial G_3}{\partial u}\right)\frac{\partial G_1}{\partial v} &+ (F_1 \circ \mathbf{G})\frac{\partial G_1}{\partial u \partial v}\Bigg] \\ - \Bigg[\left(\frac{\partial F_1}{\partial x}\frac{\partial G_1}{\partial v} + \frac{\partial F_1}{\partial y}\frac{\partial G_2}{\partial v} + \frac{\partial F_1}{\partial z}\frac{\partial G_3}{\partial v}\right)\frac{\partial G_1}{\partial u} &+ (F_1 \circ \mathbf{G})\frac{\partial G_1}{\partial v \partial u}\Bigg] \mathrm{d}A \\ = \iint_U \left(\frac{\partial F_1}{\partial z}\frac{\partial(z,x)}{\partial(u,v)} - \frac{\partial F_1}{\partial y}\frac{\partial(x,y)}{\partial(u,v)}\right) \mathrm{d}A,\end{aligned}$$

which is precisely the right hand side we had computed before. Tracing through the equalities, we've shown that (5.10) holds. Replicating this proof with the other two components of $\mathbf{F}$ gives the same result, and additivity of the integral thus yields

$$\int_{\partial M} \mathbf{F} \cdot \mathrm{d}x = \iint_M (\operatorname{curl} \mathbf{F}) \cdot \hat{\mathbf{n}} \, \mathrm{d}A$$

as required. □

Example 5.59

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

$$\int_C \mathbf{F} \cdot \mathrm{d}\mathbf{x} \quad \text{and} \quad \iint_S (\operatorname{curl} \mathbf{F}) \cdot \hat{\mathbf{n}}\, \mathrm{d}A.$$

Solution. We can parameterize C as

$$\gamma(\theta) = (\cos(\theta), \sin(\theta), \cos(\theta)), \qquad 0 \leq \theta \leq 2\pi$$

so that

$$\begin{aligned}\int_C \mathbf{F} \cdot d\mathbf{x} &= \int_0^{2\pi} (\cos(\theta), \cos(\theta), 2\sin(\theta)) \cdot (-\sin(\theta), \cos(\theta), -\sin(\theta)) d\theta \\ &= \int_0^{2\pi} -\cos(\theta)\sin(\theta) + \cos^2(\theta) - 2\sin^2(\theta) d\theta \\ &= 0 + \pi - 2\pi = -2\pi.\end{aligned}$$

On the other hand, the curl of $\mathbf{F}$ is easily computed to be

$$\operatorname{curl} \mathbf{F} = \det \begin{bmatrix} \mathbf{e}_1 & \mathbf{e}_2 & \mathbf{e}_3 \\ \partial_x & \partial_y & \partial_z \\ x & z & 2y \end{bmatrix} = (1, 0, 0).$$

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

$$\int_0^{2\pi} \int_0^{\sqrt{2}} (1, 0, 0) \cdot (-r, 0, r)\, \mathrm{d}r\, \mathrm{d}\theta = \int_0^{2\pi} \int_0^{\sqrt{2}} -r\, \mathrm{d}r\, \mathrm{d}\theta = 2\pi \left[-\frac{1}{2} r^2 \right]_{r=0}^{\sqrt{2}} = -\pi. \qquad \blacksquare$$

Example 5.60

Let $S = \left\{(x,y,z) : x^2+y^2+z^2 = 1, z \geq 0\right\}$. If ∂S is oriented counter clockwise when viewed from $(0,0,1)$, and

$$\mathbf{F}(x,y,z) = \left(xy + xe^z, \frac{1}{6}\left(2x^3+3x^2+y^2z\right), \sqrt{1+x^2+zy}\right),$$

compute $\int_{\partial S} \mathbf{F} \cdot \mathrm{d}\mathbf{x}$.

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 $S' = \left\{(x,y,0) : x^2+y^2 = 1\right\}$, 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

$$\int_C \mathbf{F} \cdot \mathrm{d}\mathbf{x} = \iint_{S'} (\operatorname{curl} \mathbf{F}) \cdot \hat{\mathbf{n}} \, \mathrm{d}A = \iint_{S'} x^2 \, \mathrm{d}A$$

This integral is much easier done. Converting to polar coordinates, we get

$$\int_{S'} x^2 \, \mathrm{d}A = \left[\int_0^1 r^3 \, \mathrm{d}r\right]\left[\int_0^{2\pi} \sin^2(\theta) \, \mathrm{d}\theta\right] = \frac{\pi}{4}.$$

■

5.7 Exercises

5-1. For each of the following manifolds, determine which can be written as the graph of a function, the zero-locus of a function, or as a collection of coordinate charts. Give the appropriate functions in each case.

(a) The ellipse $ax^2 + by^2 = c^2$, for $a, b, c > 0$.

(b) The set $S = \left\{(t^2+t, 2t-1) : t \in \mathbb{R}\right\}$,

(c) The plane $3x - 4y + 3z = 10$,

(d) The sphere of radius r, $S_r = \left\{\mathbf{x} \in \mathbb{R}^3 : \|\mathbf{x}\| = r\right\}$,

(e) The cylinder $\{(x,y,z) : x^2+y^2=4\}$,

(f) The intersection of the plane $x+z=1$ with the sphere $x^2+y^2+z^2=1$,

(g) If $f : [a,b] \to \mathbb{R}$ let S be the space defined by revolving the graph of f about the x-axis.

5-2. Determine whether the following spaces are smooth:

(a) The set $C = \{(\cos(t), \sin(2t) : t \in [0, 2\pi)\}$.

(b) For $a > 0$ define $S = \{a(t - \sin t, 1 - \cos t) : t \in \mathbb{R}\}$,

(c) Let S be the surface defined by the image of $g : \mathbb{R}^2 \to \mathbb{R}^3$, $g(s,t) = (3s, s^2 - 2t, s^3 + t^2)$.

(d) Let $S = F^{-1}(0)$ where $F(x,y,z) = 3xy + x^2 + z$.

(e) Let $S = F^{-1}(0)$ where $F(x,y,z) = \cos(xy) + e^z$

5-3. Let $U = \{(x,y,z) \in \mathbb{R}^3 : x > 0, y > 0, z > 0\}$ be the first octant, and let $\mathbf{G} : U \to \mathbb{R}^4$ be given by

$$G(t,u,v) = (tu, tv, uv, tuv).$$

Determine whether the image of g defines a smooth surface.

5-4. For what values of c do the following equations define a C^1 surface in $\mathbb{R}^3$?

(a) $x^2+y^2+z^2 = c_1$, $x^2+y^2-z^2 = c_2$,

(b) $xyz = c$

5-5. Let $F_1, F_2 : \mathbb{R}^n \to \mathbb{R}$ be C^1 functions and let $F_3(\mathbf{x}) = F_1(\mathbf{x})F_2(\mathbf{x})$. If $S_i = \{\mathbf{x} \in \mathbb{R}^n : F_i(\mathbf{x}) = 0\}$ show that $S_3 = S_1 \cup S_2$. In particular, this shows that it when analyzing the zero-locus of a function which is the product of two functions, it is sufficient to look at each constituent function separately.

5-6. Find the arclength of the following curves

(a) The straight line between $(1,2,3)$ and $(3,1,2)$,

(b) The curve given by $y^2 = x^3$ between $(1,1)$ and $(4,8)$,

(c) The curve given by $(x,y) = (t - \sin(t), 1 - \cos(t))$ for $0 \leq t \leq 2\pi$,

(d) The curve given by $(x,y) = (\cos^3(t), \sin^3(t))$ for $0 \leq t \leq \frac{\pi}{2}$,

(e) The graph of a C^1 function $y = f(x)$ for $a \leq x \leq b$.

5-7. Consider the curve $\gamma : \mathbb{R} \to \mathbb{R}^2$ give by $\gamma(t) = (2e^{-t/2}\cos(t), 2e^{-t/2}\sin(t))$.

(a) Show that $\gamma(t)$ defines a C^1 curve.

(b) Calculate the speed of this curve as a function of t. *Hint:* The velocity is $\gamma'(t)$ so the speed is $\|\gamma'(t)\|$

(c) We define the unit tangent vector to the curve to be $T(t) = \gamma'(t)/\|\gamma'(t)\|$. Compute the unit tangent vector.

(d) We will see later in the course that the arc-length of a curve on the interval $[0,t]$ is given by

$$s(t) = \int_0^t \|\gamma'(u)\| du.$$

Compute the arc-length function $s(t)$ for the curve γ.

(e) Inverting the arc-length formula gives a function $t(s)$ (time as a function of arc-length). The reparameterization of the curve $\gamma(t)$ using $t = t(s)$ is known as the *arclength parameterization*. Compute the arc-length parameterization of $\gamma(t)$; that is, compute $\gamma(t(s))$.

5-8. Let S^1 be the unit circle inside of $\mathbb{R}^2$. Let $\mathbf{N} = (0,1)$ and $\mathbf{S} = (0,-1)$ be the North and South Poles respectively.

(a) Define a map $\phi_{\mathbf{N}} : S^1 \setminus \{\mathbf{N}\} \to \mathbb{R}$ as follows: For each point $\mathbf{p} \in S^1$, consider the straight line connecting $\mathbf{N}$ and $\mathbf{p}$. Define $\phi_{\mathbf{N}}(\mathbf{p})$ to the point where this line passes the x-axis. Show that $\phi_{\mathbf{N}}$ is a diffeomorphism.

(b) The map $\phi_{\mathbf{S}} : S^1 \setminus \{\mathbf{S}\} \to \mathbb{R}$ is defined in the same way, but lines pass through the South Pole. Conclude that these two homeomorphisms show that S^1 is a smooth 1-manifold.

(c) Compute the transition maps $\phi_{\mathbf{N}}^{-1} \circ \phi_{\mathbf{S}}$ and $\phi_{\mathbf{S}}^{-1} \circ \phi_{\mathbf{N}}$ and show these are diffeomorphisms.

5-9. Read the definitions of $GL_n(\mathbb{R}), SL_n(\mathbb{R}), O(n)$, and $SO(n)$ from Exercise 2-38. Show that each of these is a smooth manifold of $\mathbb{R}^{n^2}$, and determine the dimension of that manifold.

5-10. Suppose $M \subseteq \mathbb{R}^n$ is a smooth k-manifold, and fix a point $\mathbf{p} \in M$. For $i = 0, 1$, let $U_i \subseteq \mathbb{R}^k$ and $V_i \subseteq \mathbb{R}^m$ be such that $\phi_i : U_i \subseteq \mathbb{R}^k \to M \cap V_i$ are regular embeddings and $\mathbf{p} \in \phi_i(U)$. If $M \cap V_0 \cap V_1 \neq \emptyset$, we define the *transition map between* ϕ_0 *and* ϕ_1 as

$$\phi_1^{-1} \circ \phi_0 : \phi_0^{-1}(M \cap V_0 \cap V_1) \to \phi_1^{-1}(M \cap V_0 \cap V_1).$$

Show that the transition map is a diffeomorphism of constant rank, and hence is really just a change of variables.

5-11. Suppose $M \subseteq \mathbb{R}^n$ is a smooth k-manifold.

(a) Show that the relation $\gamma_1 \sim \gamma_2$ defined on $\mathcal{C}_{\mathbf{p}}$ in Definition 5.11 is an equivalence relation.

(b) Show that Definition 5.11 does not depend on the choice of regular embedding. *Hint:* Exercise 5-10.

(c) Show that the vector space structure on $T_{\mathbf{p}}M$ is independent of the choice of regular embedding.

5-12. Find an equation for the tangent plane $T_{\mathbf{p}}S$ to the following surfaces at the indicated point

(a) $S = \left\{(x,y,z) \in \mathbb{R}^3 : x^2 + 2y^2 + 3z^2 = 6\right\}$ at $\mathbf{p} = (1,1,-1)$.

(b) $S = \left\{(x,y,z) \in \mathbb{R}^3 : xyz^2 - \log(z-1) = 8\right\}$ at $\mathbf{p} = (-2,-1,2)$.

(c) $S = \left\{(x,y,z) \in \mathbb{R}^3 : x^2 + y^2 = 1\right\}$ at $\mathbf{p} = (1/\sqrt{2}, 1/\sqrt{2}, 1)$

(d) $S = GL_n(\mathbb{R})$ at $\mathbf{p} = I_n$, where $GL(n)$ is defined in Exercise 2-38.

(e) $S = SL_n(\mathbb{R})$ at $\mathbf{p} = I_n$, where $GL(n)$ is defined in Exercise 2-38.

(f) $S = O(n)$ at $\mathbf{p} = I_n$, where $GL(n)$ is defined in Exercise 2-38.

(g) $S = SO(n)$ at $\mathbf{p} = I_n$, where $GL(n)$ is defined in Exercise 2-38.

5-13. Let $U \subseteq \mathbb{R}^n$ be an open set. Show that U is a smooth n-manifold. *Note:* Don't overthink this. It's very straightforward.

5-14. Suppose $M \subseteq \mathbb{R}^n$ is a smooth k-manifold. An *atlas* on M is a collection of regular embeddings $\phi_i : U_i \to V_i$ such that $M \subseteq \bigcup_{i\in I} V_i$ for some indexing set I. Argue that one can show M is a smooth k-manifold by providing an atlas for M.

5-15. In Example 5.6 and Exercise 5-8, we showed that S^1 is a smooth curve in $\mathbb{R}^2$ by providing regular embeddings. In each case, it was necessary to specify at least two such embeddings. Argue why it's impossible to cover S^1 with a single regular embedding.

5-16. Suppose we make the following definition:

> "If $M \subseteq \mathbb{R}^n$, we say that M is a *pseudo-smooth k-manifold with boundary* if for every point $\mathbf{p} \in M$ there exists an open neighbourhood $V \subseteq M$ of $\mathbf{p}$, and an open set U of *either* $\mathbb{R}^k$ or $\mathbb{H}^k$ such that $\phi : U \to V$ is a regular embedding."

Show that this definition is equivalent to Definition 5.18.

5-17. Show that every smooth k-manifold in $\mathbb{R}^n$ is a smooth k-manifold with boundary.

5-18. Repeat Exercise 5-10, but now for manifolds with boundary; that is, show that the transition map of any two regular embeddings of a manifold with boundary result in diffeomorphisms. *Hint:* Show that ϕ_1^{-1} restricted to $V_0 \cap V_1 \subseteq M$ is a C^1 function. To do this, you need to show that ϕ_1^{-1} extends to a C^1 function locally at every point in $V_0 \cap V_1$.

5-19. Suppose $M \subseteq \mathbb{R}^n$ is a smooth k-manifold with boundary, such that $\partial M \neq \emptyset$. Show that ∂M is a smooth $(k-1)$-manifold, and that $\partial(\partial M) = \emptyset$. *Hint:* Show that every chart $\phi : U \subseteq \mathbb{H}^k \to V \subseteq M$ restricts to a chart $\phi : \mathbb{R}^{k-1} \to V \cap \partial M$.

5-20. Show that every smooth k-manifold with boundary is a smooth manifold with corners.

5-21. Suppose $M \subseteq \mathbb{R}^k$ is an orientable smooth k-manifold with non-empty boundary ∂M. For $i = 1, 2$ let $\phi_i : U_i \subseteq \mathbb{H}^k \to V_i \subseteq M$ be two boundary charts such that $V_1 \cap V_2 \neq \emptyset$. Let $\alpha = \phi_2^{-1} \circ \phi_1 : \phi_1^{-1}(V_1 \cap V_2) \to \phi_2^{-1}(V_1 \cap V_2)$ be their transition function.

(a) Using Invariance of Boundary (Exercise 3-56), argue that $\partial_i \alpha(\mathbf{p}) = 0$ for all $i = 1, \ldots, k-1$ and $\mathbf{p} \in \partial\mathbb{H}^k \cap \phi_1^{-1}(V_1 \cap V_2)$.

(b) In Exercise 5-19 you showed that charts on M restrict to charts on ∂M. Assuming ϕ_1 and ϕ_2 are consistently oriented, show that the restricted charts are also consistently oriented. *Hint:* Use part (a),

5-22. Show that Definition 5.31 is independent on the choice of chart containing the support of f. *Hint:* The transition map between two charts is a diffeomorphism (change of variable), so apply the Change of Variables Theorem.

5-23. Show that Definition 5.32 does not depend on the choice of partition of unity.

5-24. Prove Proposition 5.33.

5-25. Plot the following vector fields:

(a) $F(x,y) = (\sqrt{x}, y)$,
(b) $F(x,y) = (y, x)$,
(c) $F(x,y) = (1/x, y)$,
(d) $F(x,y) = (y/x, y)$,
(e) $F(x,y) = (x+y, x-y)$.

5-26. For each of the following functions, compute the gradient, Laplacian, divergence, and curl, if it makes sense to do so.

(a) $F(x,y) = \sin(xe^y)$,
(b) $F(x,y) = x^2 + y^2 - 2z^2$,
(c) $F(x,y,z) = (\sin^2(x), \cos^2(z), xyz)$,
(d) $F(x,y,z) = x/(x^2+y^2)$,
(e) $F(x,y) = \log(x^2+y^2)$,
(f) $F(x,y,z) = (2xy^2z^2, 2yx^2z^2, 2zx^2y^2)$,
(g) $F(x,y,z,w) = (e^{zw}, e^{xz}, e^{xw}, e^{yz})$.

5-27. Show that the following identities hold:

(a) $\nabla(fg) = f\nabla g + g\nabla f$
(b) $\operatorname{curl}(f\mathbf{G}) = f\operatorname{curl}\mathbf{G} + (\nabla f) \times \mathbf{G}$
(c) $\operatorname{div}(f\mathbf{G}) = f\operatorname{div}\mathbf{G} + (\nabla f) \cdot \mathbf{G}$
(d) $\operatorname{div}(\mathbf{F} \times \mathbf{G}) = \mathbf{G} \cdot (\operatorname{curl}\mathbf{F}) - \mathbf{F} \cdot (\operatorname{curl}\mathbf{G})$

5-28. Show that the following identities hold:

(a) $\operatorname{curl}(\nabla f) = 0$
(b) $\operatorname{div}(\operatorname{curl}\mathbf{F}) = 0$

5-29. Let (x,y) be the standard Cartesian coordinates in $\mathbb{R}^2$. Changing to polar coordinates we set

$$x = r\cos(\theta), \quad y = r\sin(\theta).$$

(a) If $e_x = (1,0)$ and $e_y = (0,1)$ are the standard Cartesian unit vectors of $\mathbb{R}^2$, determine e_r and e_θ, the standard polar unit vectors.
(b) Using the multivariable chain rule, determine the ∇ operator in polar coordinates.
(c) Compute the Laplacian ∇^2 in polar coordinates.

5-30. Compute $\displaystyle\int_C \frac{1}{1+z/2}\,\mathrm{d}z$ where $C \subseteq \mathbb{R}^3$ is the smooth curve defined by the image of

$$\gamma(t) = (-2\sin(t), 2\cos(t), 2t^2) \quad \text{for} \quad t \in [0,1].$$

5-31. Let $M \subseteq \mathbb{R}^n$ be a connected, open, smooth n-manifold. We say that two C^1 curves $\gamma_0, \gamma_1 : [0,1] \to M$ are *homotopic* if there exists a continuous map $H : [0,1] \times [0,1] \to M$ such that

- $H(0,t) = \gamma_0(t)$ for all $t \in [0,1]$,
- $H(1,t) = \gamma_1(t)$ for all $t \in [0,1]$,
- For any $s \in [0,1]$, $\gamma_s(t) = H(s,t)$ is a C^1-curve in M.

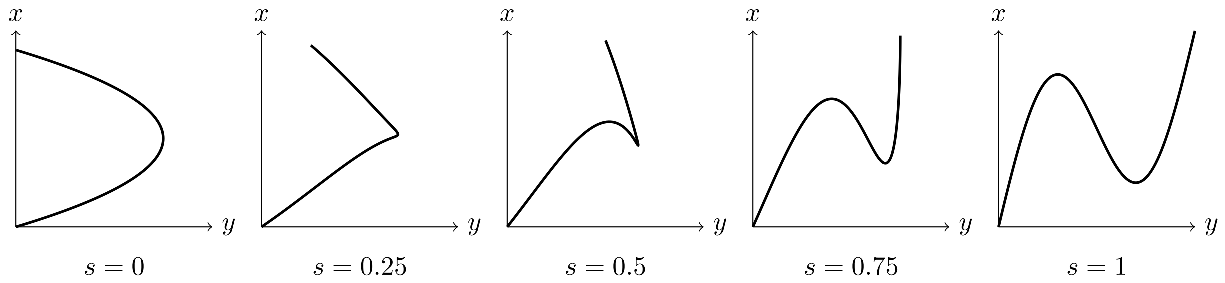


Figure 5.23: A homotopy between two curves.

The map H is called a *homotopy*.

In this exercise, we'll demonstrate the homotopy invariance of the line integral. Fix a C^1 vector field $\mathbf{F} : M \to \mathbb{R}^n$. Suppose $\gamma_0, \gamma_1 : [0,1] \to M$ are homotopic with image curves C_0 and C_1 respectively.

(a) Argue that $H([0,1] \times [0,1])$ can be covered in finitely many open star-convex sets $U_1, \ldots, U_m$.

(b) Argue that $[0,1] \times [0,1]$ can be partitioned into rectangles $R_{ij} = [s_i, s_{i+1}] \times [t_j, t_{j+1}]$ sufficiently small so that $H(R_{ij}) \subseteq U_k$ for one of star-convex sets U_k. Hint: Exercise 2-48.

(c) Define the closed curve C_{ij} to the be the image of the boundary of R_{ij} under H, traversed from $(s_i, t_j) \to (s_{i+1}, t_j) \to (s_{i+1}, t_{j+1}) \to (s_i, t_{j+1}) \to (s_i, t_j)$. Argue that $\int_{C_{ij}} \mathbf{F} \cdot \mathrm{d}x = 0$.

(d) Show that $\int_{C_0} \mathbf{F} \cdot \mathrm{d}x = \int_{C_1} \mathbf{F} \cdot \mathrm{d}x$.

5-32. Let $\mathbf{F}(x,y) = (-y^2, xy)$ and

$$C = \left\{ \frac{x^2}{a^2} + \frac{y^2}{b^2} = 1 : y \geq 0 \right\}.$$

Determine $\int_C \mathbf{F} \cdot \mathrm{d}\mathbf{x}$ if C is oriented counter clockwise when viewed from above.

5-33. Determine $\int_C \mathbf{F} \cdot \mathrm{d}\mathbf{x}$ where $\mathbf{F}(x,y) = (x^2, -y)$ and C is the graph of $y = e^x$ from $(2, e^2)$ to $(0,1)$.

5-34. Determine $\int_C \mathbf{F} \cdot \mathrm{d}\mathbf{x}$ where $\mathbf{F}(x,y,z) = (z, -y, x)$ and C is the line segment between the points $(5,0,2)$ and $(5,3,4)$.

5-35. Let $\mathbf{F}(x,y,z) = (x, y, z^2)$ and C be the curve given by the intersection of the cylinder $x^2 + y^2 = 1$ and $z = x$, with any orientation. Determine $\int_C \mathbf{F} \cdot \mathrm{d}\mathbf{x}$.

5-36. By explicitly parametrizing, show that if C is the constant curve (that is, the curve which consists of a single point), then for any vector field $\mathbf{F}$ we have

$$\int_C \mathbf{F} \cdot \mathrm{d}\mathbf{x} = 0.$$

5-37. Consider a vertical line segment $S = \{(x,y) : a \leq y \leq b, x = c\}$ in $\mathbb{R}^2$, for constants a, b, c. Show that for any vector field $\mathbf{F}(x,y) = (F_1(x,y), F_2(x,y))$ that the line integral does not depend on F_1. Similarly conclude that for a horizontal line segment, the line integral does not depend on F_2.

5-38. Let $\mathbf{F}(x,y,z) = (yz, xz, xy)$ and define

$$C_{r,h} = \left\{(x,y,z) : x^2 + y^2 = r^2, z = h\right\}.$$

Show that for and $r > 0$ and $h \in \mathbb{R}$,

$$\int_{C_{r,h}} \mathbf{F} \cdot \mathrm{d}\mathbf{x} = 0.$$

5-39. (a) Determine the values of α and β such that

$$\mathbf{F}(x,y,z) = (y + z, \alpha x + z, x + \beta y)$$

is a conservative vector field. Write down the potential function $f : \mathbb{R}^3 \to \mathbb{R}$ so that $\mathbf{F} = \nabla f$.

(b) Let C be the curve given parametrically by

$$\gamma(t) = (t\cos(t), t\sin(t), t^2), \qquad 0 \leq t \leq \frac{\pi}{2}$$

Compute the line integral $\int_C \mathbf{F} \cdot \mathrm{d}\mathbf{x}$ where $\mathbf{F}$ is the your solution from part (a).

5-40. Compute the line integral of curve $\left\{x^2 + y^2 = 1, y \geq 0\right\}$ (oriented counter clockwise) through the vector field $\mathbf{F}(x,y) = (2xy^2 + 1, 2x^2y + 2y)$.

5-41. (a) Show that the vector field $\mathbf{F}(x,y,z) = (y\cos(xy), x\cos(xy) + e^z, ye^z)$ is conservative and find its scalar potential.

(b) Consider the curve C given by the intersection of the sets

$$C = \left\{x^2 + y^2 + z^2 = 4\right\} \cap \{x = 0\} \cap \{z \geq 0\}$$

oriented so that its tangent point is the positive y-direction. Determine $\int_C \mathbf{F} \cdot \mathrm{d}\mathbf{x}$ if $\mathbf{F}$ is the vector field given in part (a).

5-42. Evaluate the following line integrals both directly and by using Green's Theorem.

(a) $\displaystyle\int_C (x + 2y)\,\mathrm{d}x + (x - 2y)\,\mathrm{d}y$ where C is given by the union of the images of the following two functions on $[0,1]$: $f(x) = x^2$ and $g(x) = x$, positively oriented with respect to the area the curves bound.

(b) $\displaystyle\int_C (3x - 5y)\,\mathrm{d}x + (x + 6y)\,\mathrm{d}y$ where C is the ellipse $x^2/4 + y^2 = 1$ oriented counterclockwise.

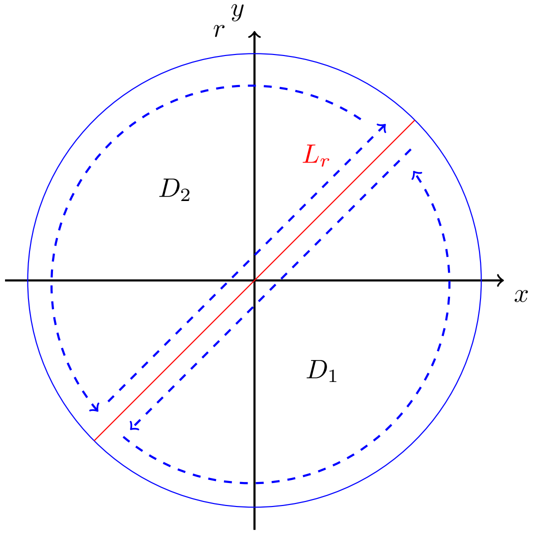
5-43. (a) Let $D \subseteq \mathbb{R}^2$ be a regular region and $\partial D = C$ be a piece-wise smooth simple closed curve, oriented positively. If $A(D)$ is the area of D, show that

$$A(D) = \int_C x\,\mathrm{d}y = \int_C y\,\mathrm{d}x = \int_C \frac{1}{2}(y\,\mathrm{d}x - x\,\mathrm{d}y).$$

(b) Consider the disk of radius r, $D_r = \left\{x^2+y^2 \leq r\right\} \subseteq \mathbb{R}^2$. Use any of the formulae from part (a) to compute the area of this disk. Of course, you already know what the result should be!

(c) In this question we will show that artificially adding boundaries does not affect the line integral. Let L_r be any diameter of D. Show that if we break D into two regions $D = D_1 \cup D_2$ and give the boundary of D_1 and D_2 positive orientations, then for any C^1 vector field $\mathbf{F}(x,y)$ we have

$$\int_{\partial D} \mathbf{F}\cdot d\mathbf{x} = \int_{\partial D_1} \mathbf{F}\cdot d\mathbf{x} + \int_{\partial D_2} \mathbf{F}\cdot d\mathbf{x}.$$



(d) Use part (a) to compute the area of the lemniscate $x^4 = x^2 - y^2$. *Note*: Be careful, as the lemniscate's boundary is not a simple closed curve.

5-44. For this problem, we will always be working in $\mathbb{R}^3$. Let $\Omega^0(\mathbb{R}^3)$ be the set of C^1 functions $\mathbb{R}^3 \to \mathbb{R}$ and $\Omega^1(\mathbb{R}^3)$ be the set of C^1 vector fields $\mathbb{R}^3 \to \mathbb{R}^3$.

(a) Show that $\Omega^0(\mathbb{R}^3)$ and $\Omega^1(\mathbb{R}^3)$ are both $\mathbb{R}$-vector spaces. This should be short proof.

(b) Show that $\operatorname{grad} : \Omega^0 \to \Omega^1$ and $\operatorname{curl} : \Omega^1 \to \Omega^1$ are linear operators.

(c) Recall that if Z is the set of closed vector fields, then $Z = \ker(\operatorname{curl})$ and if B is the set of exact vector fields, then $B = \operatorname{im}(\operatorname{grad})$. Show that $B \subseteq Z$.

(d) If $S \subseteq \mathbb{R}^3$, we define the *de Rham cohomology of S of degree one* to be

$$H^1(S) = Z/B.$$

Show that if S is the complement of the z-axis in $\mathbb{R}^3$ then $H^1(S)$ is not the trivial vector space. Hint: It suffices to show that $Z \neq B$.

5-45. Let S be the capless cylinder $S = \left\{(x,y) : x^2+y^2 = 9, 0 \leq z \leq 5\right\}$, and $\mathbf{F}(x,y,z) = (2x, 2y, 2z)$. Determine the flux of $\mathbf{F}$ through S.

5-46. Let S be the disk of radius 3, sitting in the plane $z = 3$. Determine the flux of $\mathbf{F}(x,y,z) = (0, 0, x^2+y^2)$ through S.

5-47. Evaluate the flux of $\mathbf{F}(x,y,z) = (3x^2, 2y, 8)$ over the plane $-2x+y+z=0$ for $(x,y) \in [0,2]\times[0,2]$, oriented pointing in the $-z$-direction.

5-48. Let S be the triangle with vertices $(1,0,0),(0,2,0),(0,1,1)$, and $\mathbf{F}(x,y,z) = (xyz, xyz, 0)$. Find the flux of $\mathbf{F}$ through S.

5-49. Let $S \subseteq \mathbb{R}^3$ be the capped upper half of the unit sphere; that is,

$$S = \left\{x^2+y^2+z^2=1, z\geq 0\right\} \cup \left\{x^2+y^2 \leq 1, z=0\right\}.$$

Let $\mathbf{F}(x,y,z) = (2x,2y,2z)$ be a given vector field.

(a) Compute $\iint_S \mathbf{F}\cdot\hat{\mathbf{n}}\,\mathrm{d}A$ using the Divergence Theorem.

(b) Compute $\iint_S \mathbf{F}\cdot\hat{\mathbf{n}}\,\mathrm{d}A$ by parameterizing the surface explicitly. Check that this agrees with your answer from part (a).

5-50. Assume that $\mathbf{F}:\mathbb{R}^3\to\mathbb{R}^3$ is a C^1-vector field and can be written as $\mathbf{F} = \operatorname{curl}\mathbf{G}$ for some C^2-vector field G. Show that if S is any piecwise smooth surface which bounds a regular region in $\mathbb{R}^3$, then the flux of $\mathbf{F}$ through S is zero.

5-51. Let $\mathbf{F}(x,y,z) = (xy, yz, xz)$ and $S = \left\{x^2+y^2\leq 1, 0\leq z\leq 1\right\}$ be the solid cylinder.

(a) Directly (without using any theorems) compute the flux of $\mathbf{F}$ through ∂S if ∂S has the Stokes' orientation relative to S.

(b) Compute the flux of $\mathbf{F}$ through S by using the Divergence theorem.

5-52. (a) Let $E:\mathbb{R}^3\to\mathbb{R}^3$ be given by

$$E(x,y,z) = \frac{k}{(x^2+y^2+z^2)^{3/2}}(x,y,z)$$

for some $k>0$. Show that $\operatorname{div} E = 0$.

(b) Let E be the vector-field given in part (b), and let $S_r = \{x^2+y^2+z^2=r^2\}\subseteq\mathbb{R}^3$ be the sphere of radius r. Compute the flux of E through S_r.

(c) You got different answers in part (b) and (c). Explain why this is not a contradiction to the Divergence Theorem.

5-53. Consider the cube

$$C = \{0\leq x\leq a,\ 0\leq y\leq b,\ 0\leq z\leq c\}, \qquad a,b,c>0.$$

and let $\mathbf{F}(x,y,z) = [(x-a)yz, x(y-b)z, xy(z-c)]$.

(a) Directly compute the flux of $\mathbf{F}$ through ∂C; that is, compute $\iint_{\partial C}\mathbf{F}\cdot\hat{\mathbf{n}}\,\mathrm{d}A$. *Hint:* Use symmetry to reduce this whole computation to a single integral.

(b) Directly compute the integral $\iiint_C \operatorname{div}\mathbf{F}\,\mathrm{d}V$.

(c) These two quantities are equal by a theorem. State that theorem.

5-54. Let $\mathbf{F}(\mathbf{x}) = (F_1(\mathbf{x}), F_2(\mathbf{x}), F_3(\mathbf{x}))$ be a vector field in $\mathbb{R}^3$.

(a) For arbitrary $h > 0$, let $S_h = \left\{(x, y, z) : x^2 + y^2 + z^2 = h^2\right\}$ be the sphere of radius h. Parameterize S_h by a function $\mathbf{g} : [a, b] \times [c, d] \to \mathbb{R}^3$. Compute $\frac{\partial \mathbf{g}}{\partial s} \times \frac{\partial \mathbf{g}}{\partial t}$.

(b) Under the assumption that h is very small, we can use a first order approximation on the functions F_i. Write out the linear approximations for $F_i(\mathbf{x})$ at $(0, 0, 0)$ and evaluate these on the parameterization.

(c) Use parts (a) and (b) to determine $\mathbf{F}(\mathbf{g}(t)) \cdot \left(\frac{\partial \mathbf{g}}{\partial s} \times \frac{\partial \mathbf{g}}{\partial t}\right)$. Ignore terms in order h^4, or keep track of them by writing $O(h^4)$

(d) Compute $\lim_{h \to 0} \dfrac{1}{\frac{4}{3}\pi h^3} \displaystyle\int_{S_h} \mathbf{F} \cdot \hat{\mathbf{n}} \, \mathrm{d}S$. Compare this to the divergence. Conclude that divergence is the infinitesimal flux. Notice that $\frac{4}{3}\pi h^3$ is the volume of the sphere, so we are 'normalizing' by the volume in our limit.

5-55. Let C be the circle $\left\{(x, y, 0) : x^2 + y^2 = r^2\right\}$ for some $r > 0$, and set $\mathbf{F}(x, y, z) = (x, x, y)$. Determine $\int_C \mathbf{F} \cdot \mathrm{d}\mathbf{x}$.

5-56. Let C be the curve which is the intersection of the cylinder $\left\{(x, y, z) : x^2 + y^2 = 1\right\}$ with the plane $z = 0$, in the first octant; that is, C is a quarter circle in the xy-plane. Let $\mathbf{F}(x, y, z) = (y, z, x)$.

(a) By explicitly parametrizing the curve, determine $\int_C \mathbf{F} \cdot \mathrm{d}\mathbf{x}$,

(b) Evaluate the line integral using Stokes theorem and confirm that you get the same answer.

5-57. Let $\mathbf{F}(x, y, z) = (z - y, -x - z, -x - y)$ and C be the curve given by the intersection of $x^2 + y^2 + z^2 = 4$ and the plane $z = y$; oriented counter clockwise when viewed from $(0, 0, 1)$.

(a) By explicit computation, determine $\int_C \mathbf{F} \cdot \mathrm{d}\mathbf{x}$,

(b) Using Stokes theorem, verify your result from part (a).

5-58. Let S be a smooth oriented surface in $\mathbb{R}^3$ with piecewise smooth, compatibly oriented boundary ∂S. Show that if f in C^1 and g is C^2 on S then

$$\int_{\partial S} f \nabla g \cdot \mathrm{d}\mathbf{x} = \iint_S (\nabla f \times \nabla g) \cdot \hat{\mathbf{n}} \, \mathrm{d}A.$$

5-59. (a) Let S_1 and S_2 be smooth surfaces in $\mathbb{R}^3$ with piecewise smooth boundary satisfying $\partial S_1 = \partial S_2$. If $\mathbf{F} : \mathbb{R}^3 \to \mathbb{R}^3$ is a C^1 vector field, and S_1 and S_2 are oriented so that the Stokes orientation on ∂S_1 is the same as that of ∂S_2, show that

$$\iint_{S_1} (\operatorname{curl} \mathbf{F}) \cdot \hat{\mathbf{n}} \, \mathrm{d}A = \iint_{S_2} (\operatorname{curl} \mathbf{F}) \cdot \hat{\mathbf{n}} \, \mathrm{d}A.$$

(b) Assume that $S \subseteq \mathbb{R}^4$ is an oriented smooth surface with boundary

$$\partial S = \{(x, y, 0) : x^2 + y^2 = 1\}.$$

If $\mathbf{F} : \mathbb{R}^3 \to \mathbb{R}^3$ is a C^1 vector field such that $(\operatorname{curl} \mathbf{F}) \cdot (0, 0, 1) = 0$, show that

$$\iint_S (\operatorname{curl} \mathbf{F}) \cdot \hat{\mathbf{n}} \, \mathrm{d}A = 0.$$

5-60. Show that if $S \subseteq \mathbb{R}^3$ is a smooth, closed, boundaryless surface oriented so that the unit normal points outwards, and $\mathbf{F} : \mathbb{R}^3 \to \mathbb{R}^3$ is a C^1-vector field, then

$$\iint_S (\operatorname{curl} \mathbf{F}) \cdot \hat{\mathbf{n}} \, \mathrm{d}A = 0.$$

Hint: Draw any closed curve C in S and use this to break S into two disjoint parts. Now think about what the Stokes orientation is on C

Note: If you cannot do the general case, try the following special case: Let $R \subseteq \mathbb{R}^3$ be such that $\partial R = S$ and assume that $\mathbf{F}$ is C^1 on R.

5-61. Let $\mathbf{F}(\mathbf{x}) = (F_1(\mathbf{x}), F_2(\mathbf{x}), F_3(\mathbf{x}))$ be a vector field in $\mathbb{R}^3$.

(a) For arbitrary $h > 0$, let $S_h = \left\{(x, y, 0) : x^2 + y^2 = h^2\right\}$ be the circle of radius h in the xy-plane. Parameterize S_h by a function $\mathbf{g} : [a, b] \to \mathbb{R}^3$.

(b) Under the assumption that h is very small, we can use a first order approximation on the functions F_i. Write out the linear approximations for $F_i(\mathbf{x})$ at $(0, 0, 0)$ and evaluate these on the parameterization.

(c) Use parts (a) and (b) to determine $\mathbf{F}(\mathbf{g}(t)) \cdot \mathbf{g}'(t)$. Ignore terms in order h^3, or keep track of them by writing $O(h^3)$

(d) Compute $\displaystyle\lim_{h \to 0} \frac{1}{\pi h^2} \int_{S_h} \mathbf{F} \cdot \mathrm{d}\mathbf{x}$. Compare this to the curl. Conclude that curl is the infinitesimal circulation. Notice that πh^2 is the area of the circle, so we are ‘normalizing’ by the area in our limit.