Analysis in Euclidean Spaces

Mark Sullivan

April 4, 2018

Contents

1 Real Analysis, part I **1**
- 1.1 Fundamentals of $\mathbb{R}$ . . . 1
  - 1.1.1 Construction of $\mathbb{R}$ . . . 1
  - 1.1.2 Algebraic properties of $\mathbb{R}$ . . . 4
  - 1.1.3 Order properties of $\mathbb{R}$ . . . 6
- 1.2 Metric spaces . . . 13
  - 1.2.1 Definition and examples . . . 13
  - 1.2.2 Topology of metric spaces . . . 16
  - 1.2.3 Sequences in metric spaces . . . 21
  - 1.2.4 Functions in metric spaces . . . 25
  - 1.2.5 Exercises . . . 29
- 1.3 Sequences in $\mathbb{R}$ . . . 30
  - 1.3.1 Exercises . . . 33
- 1.4 Topology of $\mathbb{R}$ . . . 34
  - 1.4.1 Exercises . . . 38
- 1.5 Real-valued functions . . . 39
  - 1.5.1 Continuous real-valued functions . . . 42
  - 1.5.2 Differentiable functions . . . 44
  - 1.5.3 Exercises . . . 49
- 1.6 Riemann integrals in $\mathbb{R}$ . . . 50

2 Real Analysis, part II **53**
- 2.1 From linear algebra to topology . . . 53
  - 2.1.1 Spaces . . . 53
  - 2.1.2 Maps between spaces . . . 58
  - 2.1.3 Exercises . . . 60
- 2.2 Differentiation . . . 61
  - 2.2.1 Defining differentiability . . . 61
  - 2.2.2 C^k functions . . . 71
  - 2.2.3 Exercises . . . 78
- 2.3 Integration . . . 79
  - 2.3.1 Definition of Integrability . . . 79
  - 2.3.2 Criteria for integrability . . . 83
  - 2.3.3 Exercises . . . 88

1 Real Analysis, part I

1.1 Fundamentals of $\mathbb{R}$

1.1.1 Construction of $\mathbb{R}$

Note to Math 431 students: this topic is not required for the class, but if you've ever wondered how the real numbers are defined, you can find the answer here.

In order to properly study the properties of the real numbers in terms of things like limits, functions and infinite sequences, it would be nice to have a concrete definition of the real numbers. Pretty much all of us have some idea of what's meant by the real numbers, but we'd get choked up if we were asked to give their definition. In order to properly define the real numbers, we'll need to delve into some rather abstract set theory.

Definition 1.1 *Let $\alpha \subsetneq \mathbb{Q}$. We say that α is a <u>Dedekind cut</u> provided that the following statements are true.*
(i) $\alpha \neq \varnothing$.
(ii) Given $a, b \in \mathbb{Q}$, if $b \in \alpha$ and $a < b$, then $a \in \alpha$.
(iii) $\forall\, a \in \alpha$, $\exists\, b \in \alpha$ such that $a < b$.

Definition 1.2 *Let $\alpha \subsetneq \mathbb{Q}$. We say that α is a <u>rational cut</u> provided that $\exists\, q \in \mathbb{Q}$ such that $\alpha = \left\{a \in \mathbb{Q} \middle| a < q\right\}$.*

This might seem redundant at first. Aren't *all* Dedekind cuts rational cuts? Absolutely not. What if we looked at all of the rational numbers q such that $q^2 < 2$? That would certainly not be a rational cut, but it certainly *would* be some kind of Dedekind cut. Therefore, there are potentially many more Dedekind cuts than just the rational ones.

From now on, we'll denote the set of Dedekind cuts by $\mathcal{R}$.

Definition 1.3 *Let $\alpha, \beta \in \mathcal{R}$. We say that <u>$\alpha$ is less than or equal to β</u>, denoted $\alpha \leq \beta$, provided that $\alpha \subseteq \beta$.*

To unravel this, this says that for two Dedekind cuts α and β, $\alpha \leq \beta$ if and only if $\forall\, x \in \alpha$, $\exists\, y \in \beta$ such that $x \leq y$. This allows us to give an ordering to $\mathcal{R}$. In fact, it is a *total* ordering; for two Dedekind cuts α and β, $\alpha \subsetneq \beta$, $\beta \subsetneq \alpha$, or $\alpha = \beta$.

Now, let's try to construct the algebraic properties of $\mathbb{R}$.

Definition 1.4 *The addition of Dedekind cuts is a binary operation + on $\mathcal{R}$ defined via the following: given $\alpha, \beta \in \mathcal{R}$,*

$$\alpha + \beta = \left\{ a + b \in \mathbb{Q} \middle| a \in \alpha \textit{ and } b \in \beta \right\}. \tag{1}$$

That wasn't too hard. Unfortunately, the same cannot be done with multiplication; if we were to say

$$\alpha\beta = \left\{ ab \in \mathbb{Q} \middle| a \in \alpha \text{ and } b \in \beta \right\}, \tag{2}$$

then we would have the following problem: if $0 \notin \alpha$ and $0 \in \beta$ (in other words, α is negative, while β is positive), then we would have to allow that $0b \in \alpha\beta$ for any $b \in \beta$. Of course, $0b = 0$, so this would mean that $0 \in \alpha\beta$ (which would correspond to $\alpha\beta$ being positive).

In order to fix this, we need to tinker with the idea of positivity for a moment. First, let's define negation in $\mathcal{R}$.

Definition 1.5 *Let $\alpha \in \mathcal{R}$. The additive inverse or negation of α is the Dedekind cut $-\alpha = \mathbb{Q} \setminus \left\{ -q \in \mathbb{Q} \middle| q \in \alpha \right\}$.*

Perhaps this definition requires a bit of explanation. To obtain an additive inverse of a Dedekind cut α, we will first take the set of additive inverses of the elements of α. This is not enough. Simply doing this gives us a set that is not a Dedekind cut. Instead, it gives us a set S of rationals so that if $p \in S$ and $q > p$, then $q \in S$. In order to obtain a true Dedekind cut, we'll "flip" this set by taking its complement in $\mathbb{Q}$. The result is the Dedekind cut we desired.

Next, let's define absolute value in $\mathcal{R}$.

Definition 1.6 *Let $\alpha \in \mathcal{R}$. The absolute value of α is the Dedekind cut*

$$|\alpha| = \begin{cases} \alpha & \textit{if } 0 \in \alpha \\ -\alpha & \textit{if } 0 \notin \alpha \end{cases}.$$

This does exactly what we'd expect. If $0 \in \alpha$, then this means that α contains rational numbers greater than 0; it is a "positive" Dedekind cut, and so the absolute value does nothing to it. On the other hand, if $0 \notin \alpha$, then α contains only rational numbers less than 0; it is a "negative" Dedekind cut, and so we must negate it to reach a positive one.

Now, as for multiplication, if the two cuts being multiplied are both "positive," then we can define a product in the normal way.

Definition 1.7 *Let $P = \{\alpha \in \mathcal{R} \big| 0 \in \alpha\}$. Given $\alpha, \beta \in P$, we define the product of α and β as the Dedekind cut $\alpha \cdot \beta = \{a \cdot b \in \mathbb{Q} \big| a \in \alpha \text{ and } b \in \beta\}$.*

Based on this, let's take the absolute value of every Dedekind cut in order to obtain positive ones. From this, we can construct the multiplication that we're used to.

Definition 1.8 *The multiplication of Dedekind cuts is a binary operation $+$ on $\mathcal{R}$ defined via the following: given $\alpha, \beta \in \mathcal{R}$,*

$$\alpha \cdot \beta = \begin{cases} |\alpha| \cdot |\beta| & \text{if } a \geq 0 \text{ and } b \geq 0 \\ |\alpha| \cdot |\beta| & \text{if } a < 0 \text{ and } b < 0 \\ -|\alpha| \cdot |\beta| & \text{if } a \geq 0 \text{ and } b < 0 \\ -|\alpha| \cdot |\beta| & \text{if } a < 0 \text{ and } b \geq 0 \end{cases} \tag{3}$$

With all of this, we can finally define the real numbers. The set $\mathcal{R}$ of Dedekind cuts forms what is called a *field* with respect to these definitions of addition and multiplication. The set of Dedekind cuts, together with their total ordering and the operations of addition and multiplication, is what we mean by the real numbers.

Definition 1.9 *The field of real numbers is the totally ordered field $\mathbb{R}$ of Dedekind cuts under addition and multiplication.*

From now on, we'll refer to a rational Dedekind cut $\{q \in \mathbb{Q} \big| q < a\}$ as simply a. In this way, we can think of the field of rational numbers as being a subset of the field of real numbers, just as the set of rational cuts is a subset of the set of Dedekind cuts.

1.1.2 Algebraic properties of $\mathbb{R}$

For those unfamiliar with abstract algebra, we'll state the algebraic properties of $\mathbb{R}$ as a field.

Proposition 1.10 *The following statements are true.*
(i) Addition of real numbers is associative. This means that $\forall\, x, y, z \in \mathbb{R}$,

$$x + (y + z) = (x + y) + z.$$

(ii) The number 0 *is an additive identity for* $\mathbb{R}$*. This means that* $\forall\, x \in \mathbb{R}$,

$$0 + x = x + 0 = x.$$

(iii) Each real number has an additive inverse. This means that $\forall\, x \in \mathbb{R}$, $\exists\, y \in \mathbb{R}$ *such that*

$$x + y = y + x = 0.$$

(iv) Addition of real numbers is commutative. This means that $\forall\, x, y \in \mathbb{R}$,

$$x + y = y + x$$

(v) Multiplication of real numbers is associative: $\forall\, x, y, z \in \mathbb{R}$,

$$x(yz) = (xy)z.$$

(vi) Multiplication of real numbers distributes over addition from the left. This means that $\forall\, x, y, z \in \mathbb{R}$,

$$x\,(y + z) = xy + xz$$

(vii) Multiplication of real numbers distributes over addition from the right. This means that $\forall\, x, y, z \in \mathbb{R}$,

$$(x + y)z = xz + yz$$

(iix) Multiplication of real numbers is commutative: $\forall\, x, y \in \mathbb{R}$*,*

$$xy = yx.$$

(ix) The number 1 *is a multiplicative identity for* $\mathbb{R}$*:* $\forall\, x \in \mathbb{R}$*,*

$$1x = x1 = x.$$

(x) Each real number except 0 *has a multiplicative inverse:* $\forall\, x \in \mathbb{R}$*, if* $x \neq 0$*, then* $\exists\, y \in \mathbb{R}$ *such that*

$$xy = yx = 1.$$

Proof One can prove these properties from the definitions of addition and multiplication of real numbers. □

From now on, we'll refer to the multiplicative inverse of a real number x as x^{-1}. We define "subtraction" via $x - y = x + (-y)$ and "division" via $\frac{x}{y} = xy^{-1}$ when $y \neq 0$. These two operations are completely superfluous, as all of mathematics can be done using neither subtraction nor division, but to know them is necessary to reading most mathematical literature.

There are many other algebraic properties of $\mathbb{R}$, but all of them follow either directly from the definitions or from these properties listed above. In any case, it is likely that they are all intensely familiar to the reader.

1.1.3 Order properties of $\mathbb{R}$

We will now turn our attention to the order properties of $\mathbb{R}$.

Definition 1.11 *Let $S \subseteq \mathbb{R}$ such that $S \neq \varnothing$. We say that S is bounded above provided that $\exists\ b \in \mathbb{R}$ such that $\forall\ s \in S$, $s \leq b$. In that case, we say that b is an upper bound for S.*

Definition 1.12 *Let $S \subseteq \mathbb{R}$ such that $S \neq \varnothing$. We say that S is bounded below provided that $\exists\ a \in \mathbb{R}$ such that $\forall\ s \in S$, $r \leq a$. In that case, we say that a is a lower bound for S.*

Definition 1.13 *Let $S \subseteq \mathbb{R}$ such that $S \neq \varnothing$. We say that S is bounded provided that S is bounded above and bounded below.*

These three definitions are necessary to giving a new idea of "size" to subsets of $\mathbb{R}$. After all, the sets $[0, 1]$ and $[0, \infty) = \{x \in \mathbb{R} | x \geq 0\}$ have the same amount of elements (in terms of cardinality), but there is some sense in which $[0, \infty)$ is "larger" than $[0, 1]$. Boundedness is a first approach to this idea.

Given a set that is bounded above, like, say, $(0, 4)$, there are many possible upper bounds, such as: 10, 100, 1000, 5, $\frac{21}{5}$, and so on. However, there is one particular upper bound which is special: 4. This idea of "best" upper and lower bound will be formalized in the following two definitions.

Definition 1.14 *Let $S \subseteq \mathbb{R}$ be bounded above. A supremum or least upper bound of S is a $\beta \in \mathbb{R}$ satisfying the following conditions.*
(i) β is an upper bound of S.
(ii) If $\gamma \in \mathbb{R}$ is an upper bound of S, then $\beta \leq \gamma$.

Definition 1.15 *Let $S \subseteq \mathbb{R}$ be bounded below. An infimum or greatest lower bound of S is a $\alpha \in \mathbb{R}$ satisfying the following conditions.*
(i) α is a lower bound of S.
(ii) If $\gamma \in \mathbb{R}$ is a lower bound of S, then $\gamma \leq \alpha$.

Now, given a bounded set, are its infimum and supremum unique? The answer is yes.

Proposition 1.16 *Let $S \subseteq \mathbb{R}$ such that $S \neq \varnothing$. The following statements are true.*
(i) If S is bounded above, and β_1 and β_2 are suprema of S, then $\beta_1 = \beta_2$.
(ii) If S is bounded below, and α_1 and α_2 are infima of S, then $\alpha_1 = \alpha_2$.

Proof One should be able to reproduce these proofs as an **exercise**. □

From now on, given a set $S \subseteq \mathbb{R}$, we will denote *the* supremum of S by $\sup S$ and *the* infimum of S by $\inf S$.

The following is a much deeper question: suppose a set has an upper bound. Must it have a least upper bound? The answer is not at all obvious, and its proof is rather abstract. However, it is yes. The following is known as the [order] completeness property of $\mathbb{R}$.

Theorem 1.17 *Let $S \subseteq \mathbb{R}$ such that $S \neq \varnothing$. The following statements are true.*
(i) If S is bounded above, then $\sup S \in \mathbb{R}$.
(ii) If S is bounded below, then $\inf S \in \mathbb{R}$.

Proof (i) Note that this proof involves the definition of $\mathbb{R}$ using Dedekind cuts. For more information, see Section 1.1.1.

Let $S \subseteq \mathbb{R}$ be bounded above. Define the following set.

$$\beta = \left\{ q \in \mathbb{Q} \middle| \exists\, \sigma \in S \text{ such that } q < \sigma \right\}. \tag{4}$$

We claim that $\beta \in \mathbb{R}$. First of all, $S \neq \varnothing$, so $\exists\, \sigma \in S$. Since σ is a Dedekind cut, $\sigma \neq \varnothing$, so $\exists\, q \in \mathbb{Q}$ such that $q \in \sigma$. Moreover, by taking $p \in \mathbb{Q}$ such that $p < q$, we see that $p \in \sigma$. Thus, the rational cut $\left\{ a \in \mathbb{Q} \middle| a < p \right\} \subsetneq \sigma$, so $p < \sigma$. Therefore, $p \in \beta$, so $\beta \neq \varnothing$. Now, let $a, b \in \mathbb{Q}$ such that $a < b$. If $b \in \beta$, then $\exists\, \sigma \in S$ such that $a < b < \sigma$, so $a \in \beta$. Finally, given $a \in \beta$, $\exists\, \sigma \in S$ such that $a \in \sigma$. Since σ is a Dedekind cut, $\exists\, b \in \sigma$ such that $a < b$. This implies that the rational cut $\left\{ q \in \mathbb{Q} \middle| q < b \right\} \subsetneq \sigma$, which indicates that $b < s$. Ergo, $b \in \beta$. This shows that $\beta \in \mathbb{R}$.

We claim that $\beta = \sup S$. First, let $\sigma \in S$. Given $q \in \sigma$, we know that $\left\{ a \in \mathbb{Q} \middle| a < q \right\} \subseteq \sigma$. In that case, $q < \sigma$, and so $q \in \beta$. We deduce that $\sigma \subseteq \beta$, and so $\sigma \leq \beta$. Thus, β is an upper bound for S.

Suppose that $\forall\, \sigma \in S$, $\sigma \leq \gamma$. Let $q \in \beta$. We know that $\exists\, \sigma \in S$ such that $q < \sigma < \gamma$. Therefore, $\left\{a \in \mathbb{Q} \middle| a < q\right\} \subseteq \gamma$, which indicates that $q \in \gamma$. Ergo, $\beta \subseteq \gamma$, and so $\beta \leq \gamma$.

(ii) One should be able to assume (i) and prove (ii) as an **exercise**. □

This theorem has an enormous variety of both theoretical and practical applications. In fact, the real numbers were defined for the sole purpose of making the order completeness property true, but we'll discuss that later.

For now, we will prove one of the most basic and intuitive facts about $\mathbb{R}$. The following is known as the Archimedean property of the real numbers.

Theorem 1.18 *The set $\mathbb{Z}^+$ is not bounded above in $\mathbb{R}$.*

Proof Assume, with the expectation of a contradiction, that $\mathbb{Z}^+$ is bounded in $\mathbb{R}$. In that case, by the completeness of $\mathbb{R}$ (Theorem 1.17), we know that $\exists\, \beta \in \mathbb{R}$ such that $\beta = \sup\left(\mathbb{Z}^+\right)$. In that case, $\beta - 1 < \beta$ cannot be the supremum of $\mathbb{Z}^+$, so $\exists\, n \in \mathbb{Z}^+$ such that $\beta - 1 < n$. We deduce that $\beta < n + 1$, despite that $\beta \geq n + 1$ by assumption. This contradiction leads us to conclude that our assumption that $\mathbb{Z}^+$ is bounded above is false; $\mathbb{Z}^+$ is not bounded above in $\mathbb{R}$. □

As pedestrian as it may seem, the Archimedean property leads to a wide variety of important properties that we would expect from $\mathbb{R}$. First of all, the property can be used to show that positive real numbers always have real square roots.

Proposition 1.19 *Let $n \in \mathbb{Z}^+$. There exists $r \in \mathbb{R}$ such that $r^2 = n$.*

Proof Define $S = \left\{r \in \mathbb{R} \middle| r^2 < n\right\}$. We know that $1 \in S$, so $S \neq \varnothing$. At the same time, $n^2 \geq n$, so n is an upper bound for S. By the order completeness of $\mathbb{R}$ (Theorem 1.17), $\exists\, \beta \in \mathbb{R}$ such that $\beta = \sup S$. We claim that $\beta^2 = n$. Assume, with the expectation of a contradiction, that $\beta^2 \neq n$. We consider two cases: either $\beta^2 < n$ or $\beta^2 > n$.

Consider the case that $\beta^2 < n$. By the Archimedean property (Theorem 1.18), we know that $\exists\, m \in \mathbb{Z}^+$ such that $m > \frac{2\beta+1}{n-\beta^2}$. In that case, $\frac{1}{m}\left(2\beta + 1\right) < n - \beta^2$.

Now, consider $\beta + \frac{1}{m}$. We note that

$$\left(\beta + \frac{1}{m}\right)^2 = \beta^2 + \frac{2}{m}\beta + \frac{1}{m^2} = \beta^2 + \frac{1}{m}\left(2\beta + \frac{1}{m}\right)$$
$$\leq \beta^2 + \frac{1}{m}(2\beta + 1) < \beta^2 + \left(n - \beta^2\right) = n. \quad (5)$$

We deduce that $\beta + \frac{1}{m} \in S$, despite that $\beta < \beta + \frac{1}{m}$ and $\beta = \sup S$. This is a contradiction.

Consider the case that $\beta^2 > n$. By the Archimedean property (Theorem 1.18), we know that $\exists\, m \in \mathbb{Z}^+$ such that $m > \frac{2\beta}{\beta^2 - n}$. In that case, $\frac{2}{m}\beta < \beta^2 - n$. Now, consider $\beta - \frac{1}{m}$. We note that

$$\left(\beta - \frac{1}{m}\right)^2 = \beta^2 - \frac{2}{m}\beta + \frac{1}{m^2} > \beta^2 - \frac{2}{m}\beta > \beta^2 - \left(\beta^2 - n\right) = n. \quad (6)$$

We deduce that $\beta - \frac{1}{m}$ is an upper bound for S, despite that $\beta = \sup S$. This is a contradiction.

Whatever the case, we reach a contradiction that leads us to conclude that our assumption that $\beta^2 \neq n$ is false; $\beta^2 = n$. $\square$

We would like to address the following question that may or may not have already occurred to the reader: why couldn't we just study analysis in $\mathbb{Q}$? What was the point of constructing $\mathbb{R}$? Besides that, why is $\mathbb{R}$ sufficient for our needs? We will now begin to discuss the answers to these questions.

The rationals are "defective" in the following sense: there exist bounded subsets of $\mathbb{Q}$ that do not have suprema or infima in $\mathbb{Q}$. For example, consider the following set:

$$S = \left\{q \in \mathbb{Q} \middle| q^2 < 2\right\}. \quad (7)$$

Let's try to think of the upper bounds of S. Certainly 100, 50, 25, and $\frac{25}{2}$ are all rational upper bounds of S. Even 2 is a rational upper bound of S. So is 1.5. So is 1.42. So is 1.415. In fact, any rational number larger than $\sqrt{2}$ would be an upper bound of S. Now then, of all the upper bounds (of which there are infinitely many), which is the least? What is the supremum of S?

The order completeness property of $\mathbb{R}$ guarantees that a supremum will exist for S in $\mathbb{R}$. In fact, $\sqrt{2}$ is a real number, by Proposition 1.19. Moreover, Proposition 1.16 indicates that no other real number *will* be a least upper bound for S. Ergo, there is no rational number that serves as a least upper bound of S. Thus, the set of rationals has many "holes," points where a least upper bound should exist, but cannot be rational. It is for this reason that we must do analysis in $\mathbb{R}$ rather than $\mathbb{Q}$; the order completeness property guarantees us that we will never have an upper bounded set that has no least upper bound.

We can think of the rational numbers as being like a skeleton, and the real numbers as being like the flesh surrounding it. Of course, for any piece of the flesh, there is some bone nearby. Therefore, given a real number, how close is the closest rational number? We will introduce a few definitions before answering this question.

Definition 1.20 *Let $x, y \in \mathbb{R}$ such that $x \leq y$. The open interval from x to y is the set $(x, y) = \{r \in \mathbb{R} | x < r < y\}$.*

Definition 1.21 *Let $x, y \in \mathbb{R}$ such that $x \leq y$. The closed interval from x to y is the set $[x, y] = \{r \in \mathbb{R} | x \leq r \leq y\}$.*

Definition 1.22 *Let $x \in \mathbb{R}$.*
(i) The open interval from x to ∞ is the set $(x, \infty) = \{r \in \mathbb{R} | x < r\}$.
(ii) The closed interval from x to ∞ is the set $[x, \infty) = \{r \in \mathbb{R} | x \leq r\}$.
(iii) The open interval from ∞ to x is the set $(\infty, x) = \{r \in \mathbb{R} | r < x\}$.
(iv) The closed interval from x to ∞ is the set $(\infty, x] = \{r \in \mathbb{R} | r \leq x\}$.

In addition to these, we will sometimes refer to the somewhat absurd notion of a "half-open interval" like

$$[x, y) = \{r \in \mathbb{R} | x \leq r < y\} \tag{8}$$

or

$$(x, y] = \{r \in \mathbb{R} | x < r \leq y\}. \tag{9}$$

Now then, what is the distance between rational numbers? The answer is contained in the following proposition.

Proposition 1.23 *Let $x, y \in \mathbb{R}$ such that $x < y$. The intersection $(x, y) \cap \mathbb{Q} \neq \varnothing$.*

Proof We consider two cases: either $x \geq 0$ or $x < 0$. Consider the case that $x \geq 0$. By the Archimedean property (Theorem 1.18), $\exists\, n \in \mathbb{Z}^+$ such that $n > \frac{1}{y-x}$. In that case, $nx + 1 < ny$.

We claim that $\exists\, m \in \mathbb{Z}$ such that $m - 1 \leq nx < m$. By the Archimedean property (Theorem 1.18), $S = \left\{n \in \mathbb{Z}^+ \middle| nx < m\right\} \neq \varnothing$. On the other hand, by the well-ordering of $\mathbb{Z}^+$, there exists a least element m of S. Now $m - 1 \leq nx$; otherwise $m - 1 \in S$, despite that m is the least element. Ergo, $m - 1 \leq nx < m$.

By the claim, $m \leq nx + 1$, and so $m < ny$. We deduce that $\frac{m}{n} < y$. Since by the claim, $nx < m$, we also deduce that $x < \frac{m}{n}$. Thus, $\frac{m}{n} \in (x, y)$.

Consider the case that $x < 0$. Using the Archimedean property (Theorem 1.18), select $n \in \mathbb{Z}^+$ such that $-x < n$. Now $0 < n + x$, and so, by the other case, we can say that $\exists\, q \in (n + x, n + y) \cap \mathbb{Q}$. In that case, $n + x < q < n + y$, and so $x < q - n < y$. Yet $q - n \in \mathbb{Q}$, so $q - n \in (x, y) \cap \mathbb{Q}$. $\square$

Note that this can be extended: for any $x, y \in \mathbb{R}$, if $x < y$, then it is not difficult to show that $(x, y) \cap \mathbb{Q}$ is actually an infinite set. Additionally, one can show that $(x, y) \cap (\mathbb{R} \setminus \mathbb{Q})$ is also an infinite set.

There is one more property of the ordering of $\mathbb{R}$ that we should discuss: one which allows us to discuss metric properties of the real line.

Theorem 1.24 *Given $x, y \in \mathbb{R}$, $|x + y| \leq |x| + |y|$.*

Proof We know that $-|x| \leq x \leq |x|$ and $-|y| \leq y \leq |y|$. Adding these, we obtain

$$-|x| - |y| \leq x + y \leq |x| + |y|. \tag{10}$$

Therefore, $-\left(|x| + |y|\right) \leq x+y \leq |x|+|y|$. We consider two cases: either $x+y \geq 0$ or $x + y < 0$. If $x + y \geq 0$, then

$$|x + y| = x + y \leq |x| + |y|. \tag{11}$$

On the other hand, if $x + y < 0$, then

$$|x + y| = -(|x| + |y|) \leq x + y \leq |x| + |y|. \tag{12}$$

Whatever the case, we must have $|x + y| \leq |x| + |y|$. $\square$

1.2 Metric spaces

1.2.1 Definition and examples

Note to Math 431 students: this topic is not required for the class, but I would highly recommend you at least take a look at it.

The concept of metric spaces is one of the most important ideas in the foundation of real analysis.

Definition 1.25 *Let X be a set, and let $d : X \times X \to \mathbb{R}$ be a function. We say that (X, d) is a metric space provided that the following statements are true.*
(i) $\forall\, x, y \in X$, $d(x, y) \geq 0$.
(ii) $\forall\, x, y \in X$, $d(x, y) = d(y, x)$.
(iii) $\forall\, x, y \in X$, $d(x, y) = 0$ if and only if $x = y$.
(iv) $\forall\, x, y, z \in X$, $d(x, z) \leq d(x, y) + d(y, z)$.
In that case, we say that d is a metric on X.

To summarize: a metric space is a set with a notion of distance. This distance needs to make sense in four ways: first, no negative distances are possible, second, the distance between two points depends only on the points themselves, third, a distance of zero corresponds to equality, and fourth, any single side of a triangle must be, at most, as large as the sum of the lengths of the other two sides. The axiom (iv) is called the triangle inequality. Its importance is less clear than the other axioms, but it is vital to making sure that the properties of distance to which we're accustomed will hold true.

Why are metric spaces important for analysis? The answer is that any concept of limits must rely on a notion of distance. Moreover, many of the important results of real analysis are due to the fact that the real line is a metric space. It is possible to do a first course in real analysis without mentioning metric spaces, but it is the opinion of the author that such an approach is horrifically awkward.

Here are some examples of metric spaces.

Example 1.26 $\mathbb{R}^2$ *with distance defined via*

$$d\left(\left(x_1, y_1\right), \left(x_2, y_2\right)\right) = \sqrt{\left(x_2 - x_1\right)^2 + \left(y_2 - y_1\right)^2}. \tag{13}$$

(A proof that this is, in fact, a metric is given in Theorem 2.6.)

Example 1.27 $\mathbb{R}^2$ *with distance defined via*

$$d\left(\left(x_1, y_1\right), \left(x_2, y_2\right)\right) = |x_1 - x_2| + |y_1 - y_2|. \tag{14}$$

Example 1.28 $\mathbb{R}^2$ *with distance defined via*

$$d\left(\left(x_1, y_1\right), \left(x_2, y_2\right)\right) = \max\left\{|x_1 - x_2|, |y_1 - y_2|\right\}. \tag{15}$$

Example 1.29 $\mathbb{R}^2$ *with distance defined via*

$$d\left(\left(x_1, y_1\right), \left(x_2, y_2\right)\right) = \begin{cases} \sqrt{{x_1}^2 + {y_1}^2} + \sqrt{{x_2}^2 + {y_2}^2} & \textit{if } (x_1, y_1) \neq (x_2, y_2) \\ 0 & \textit{if } (x_1, y_1) = (x_2, y_2) \end{cases}. \tag{16}$$

Example 1.30 *The set* $F_r = \left\{f : \mathbb{R} \to \mathbb{R} \middle| \forall\, x \in \mathbb{R},\ |f(x)| < r\right\}$ *for some* $r \in \mathbb{R}^+$*, with distance defined via*

$$d\left(f, g\right) = \sup\left\{f(x) - g(x) \middle| x \in \mathbb{R}\right\}. \tag{17}$$

Example 1.31 *The set* $S = \left\{f : [0,1] \to \mathbb{R} \middle| f \textit{ is continuous}\right\}$ *with distance defined via*

$$d\left(f, g\right) = \int_0^1 \left|f(x) - g(x)\right| \,\mathrm{d}x. \tag{18}$$

In $\mathbb{R}$, the distance between two points, $x, y \in \mathbb{R}$, is usually thought of as $d(x, y) = |x - y|$. Therefore, our next task is to show that the real line is, in fact, a metric space under this notion of distance. (We promise that we would not have mentioned metric spaces if it weren't.)

Theorem 1.32 *For each* $x, y \in \mathbb{R}$*, define* $d(x, y) = |x - y|$*. The pair* $(\mathbb{R}, d)$ *is a metric space.*

Proof We must show that the metric $d : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ satisfies the axioms (i)-(iv) in Definition 1.25. We know that $\forall\ a \in \mathbb{R}$, $|a| \geq 0$, so in particular, $\forall\ x, y \in \mathbb{R}$, $d(x, y) = |x - y| \geq 0$. This proves axiom (i). Given $x, y \in \mathbb{R}$, we know that

$$d(x, y) = |x - y| = \big| -(y - x) \big| = |y - x| = d(y, x). \tag{19}$$

This proves axiom (ii). Given $x, y \in \mathbb{R}$, if $d(x, y) = 0$, then $|x - y| = 0$, which implies that $x - y = 0$, thus $x = y$. On the other hand, if $x = y$, then

$$d(x, y) = |x - y| = |x - x| = |0| = 0 \tag{20}$$

This proves axiom (iii). Finally, given $x, y, z \in \mathbb{R}$,

$$d(x, z) = |x - z| = \big|x + (y - y) - z\big| = \big|\,(x - y) + (y - z)\,\big|. \tag{21}$$

However, by Theorem 1.24,

$$d\,(x, z) = \big|\,(x - y) + (y - z)\,\big| \leq |x - y| + |y - z| = d(x, y) + d(y, z), \tag{22}$$

which proves axiom (iv). $\square$

1.2.2 Topology of metric spaces

In the course of studying real analysis, we'll need to discuss some properties of subsets of metric spaces. The study of spaces via looking at small patches of those spaces is a mathematical field known as topology. Of course, topology is a topic that could constitute an entire document (or two, or three) of its own, but for real analysis, we'll only be concerned with the topological properties of metric spaces.

First, in a space that has a notion of distance, there is a notion of "finitely large."

Definition 1.33 *Let (X, d) be a metric space. Given $S \subseteq X$ such that $S \neq \varnothing$, we say that S is* *[metrically] bounded in X* *provided that $\exists\ a \in X$ and $r \in \mathbb{R}^+$ such that $\forall\ s \in S$, $d(a, s) < r$.*

As **Exercise 1.63** indicates, this is a generalization of boundedness to metric spaces other than $\mathbb{R}$.

Next, let's look at a very specific type of set in a metric space.

Definition 1.34 *Let (X, d) be a metric space. Given $a \in X$ and $\varepsilon \in \mathbb{R}^+$, we define the* *open ball centered at a with radius ε* *as the set $B(a, \varepsilon) = \{x \in X \mid d(x, a) < \varepsilon\}$.*

An open ball is just the set of points in a metric space that are less than a fixed distance from a certain point, known as the center of the ball. In $\mathbb{R}^2$, this looks like a disk, the interior of a circle. In $\mathbb{R}^3$, an open ball looks like the interior of a sphere. In $\mathbb{R}$, open balls are just open intervals.

Now, let's generalize the notion of openness just a bit.

Definition 1.35 *Let (X, d) be a metric space. Given $U \subseteq X$, we say that* *U is open in X* *provided that $\forall\ x \in U$, $\exists\ \varepsilon \in \mathbb{R}^+$ such that $B(x, \varepsilon) \subseteq U$.*

Definition 1.36 *Let (X, d) be a metric space, and let $Y \subseteq X$. Given $V \subseteq Y$, we say that* *V is open in Y as a subspace of X* *provided that $V = U \cap Y$ for some $U \subseteq X$ that is open in X.*

In other words, given a metric space (X, d), a set $U \subseteq X$ is open in X provided that $\forall\ x \in U$, $\exists\ \varepsilon \in \mathbb{R}^+$ such that if $|x - y| < \varepsilon$, then $y \in U$ as well. Intuitively, this means that a set is open if every point in the set is surrounded by points that are also in the set.

Here are some examples of open sets.

Example 1.37 *The following sets are open sets in their respective metric spaces.*
(i) Any open interval (a, b) is open in $\mathbb{R}$, since for any $x \in (a, b)$, we can select $\varepsilon = \min\{|x-a|, |x-b|\}$ so that if $|x-y| < \varepsilon$, then $y \in (a, b)$.
(ii) The set $U = \{(x, y) \in \mathbb{R}^2 \big| x < 0\}$ is open in $\mathbb{R}^2$; given a point (x, y) such that $x < 0$, one can define $\varepsilon = \frac{-x}{2}$ so that if $\sqrt{(x-u)^2 + (y-v)^2} < \varepsilon$, then $(u, v) \in U$.
(iii) The set $U = \{(x, y) \in \mathbb{R}^2 \big| x \neq 0\}$ is open in $\mathbb{R}^2$; given a point (x, y), such that $x \neq 0$, one can find $\varepsilon = \frac{|x|}{2}$ such that if $\sqrt{(x-u)^2 + (y-v)^2} < \varepsilon$, then $(u, v) \in U$.
(iv) Any infinite open interval (a, ∞) is open in $\mathbb{R}$, since for any $x \in (a, \infty)$, we can take $\varepsilon = |x-a|$ so that if $|x-y| < \varepsilon$, then $y \in (a, \infty)$.
(v) The set $[0, \frac{1}{2})$ is open in $[0, 1]$ as a subspace of $\mathbb{R}$, since $[0, \frac{1}{2}) = (-\frac{1}{2}, \frac{1}{2}) \cap [0, 1]$.

Now, let's state an important theorem about open sets in a metric space.

Theorem 1.38 *Let (X, d) be a metric space. The following statements are true.*
(i) X and $\varnothing$ are open in X.
(ii) If $\{U_\alpha\}_{\alpha \in J}$ is a collection of sets that are open in X, then their union, $\bigcup_{\alpha \in J} U_\alpha$, is also open in X.
(iii) If $\{U_1, U_2, ..., U_n\}$ is a finite collection of sets that are open in X, then their intersection, $\bigcap_{k=1}^{n} U_k$, is also open in X.

Proof This is **Exercise 1.64**. □

Sets can have different notions of what defines "openness." When Theorem 1.38 is true for a set X, then we say that X is a "topological space." The topic of redefining openness is one of the main goals of point-set topology, but we will not digress into that.

Next, we should define the dual of the open sets: the closed sets.

Definition 1.39 *Let (X, d) be a metric space, and let $F \subseteq X$. We say that F is closed in X provided that $X \setminus F$ is open in F.*

There is a problem here. It is possible for a set to be neither open nor closed, as **Exercise 1.66** indicates. At the same time, there sets that are both open and closed, such as $\varnothing$. A set that is both open and closed is called "clopen." (At this point, the author would like to clarify that this is, in fact, actual mathematical terminology used in the literature, and that he is not, in fact, just screwing with you.)

There is another way to define closedness.

Definition 1.40 *Let (X, d) be a metric space, and let $S \subseteq X$. Given $a \in X$, we say that a is a limit point, or cluster point, or accumulation point of S provided that $\forall\, \varepsilon \in \mathbb{R}^+$, $\exists\, x \in S$ such that $x \neq a$ and $x \in B\,(a, \varepsilon)$.*

A limit point is a point that is arbitrarily close to the points in the set. The following proposition indicates that the statement "a set is closed" is exactly equivalent to the statement "a set contains all of its limit points."

Proposition 1.41 *Let (X, d) be a metric space. Given $A \subseteq X$, A is closed in X if and only if $\forall\, a \in X$, if a is a limit point of A, then $a \in A$.*

Proof ($\Rightarrow$) Let $a \in X$ be a limit point of A. Assume, with the expectation of a contradiction, that $a \notin A$. We note that $X \setminus A$ is open, since A is closed. Ergo, $\exists\, \varepsilon \in \mathbb{R}^+$ such that $B\,(a, \varepsilon) \subseteq X \setminus A$. However, a is a limit point of A, so we must have that $B\,(a, \varepsilon) \cap A \neq \varnothing$. Thus, $(X \setminus A) \cap A \neq \varnothing$, which is an obvious contradiction. This leads us to conclude that our assumption that $a \notin A$ is false; $a \in A$.

($\Leftarrow$) Assume that A contains all of its limit points. Let $x \in X \setminus A$. We know that x is not a limit point of A, so $\exists\, \varepsilon \in \mathbb{R}^+$ such that $\forall\, a \in A$, either $a = x$ or $a \notin B\,(x, \varepsilon)$. Clearly $a \neq x$, since $x \notin A$. Therefore, $\forall\, a \in A$, $a \notin B\,(x, \varepsilon)$, so $B\,(x, \varepsilon) \subseteq X \setminus A$. This demonstrates that $X \setminus A$ is open in X. $\square$

Having dealt with the technical background, let's discuss some of the problems that point-set topology was designed to address. There are three main questions about a topological space. The first of these is: is the space "in one piece," or is it "broken up into multiple pieces?"

Definition 1.42 *Let (X, d) be a metric space. Given $S \subseteq X$, we say that S is a disconnected set provided that there exist open sets $U, V \subseteq X$ satisfying the following conditions.*
(i) $S \cap U \neq \varnothing$ and $S \cap V \neq \varnothing$.
(ii) $S \subseteq U \cup V$.
(iii) $U \cap V = \varnothing$.
We say that S is a connected set provided that S is not a disconnected set.

In other words, a set is disconnected if it can be partitioned by nonempty open sets. Here are some examples of disconnected sets.

Example 1.43 *The following sets are disconnected.*
(i) $(0,1) \cup (2,3)$ is disconnected, since $(0,1)$ and $(2,3)$ are nonempty open sets that partition the set.
(ii) $\{0,4\}$ is disconnected, since $(-1,1)$ and $(3,5)$ are nonempty open sets that partition the set.
(iii) $[0,1] \cup \{3\}$ is disconnected, since $(-1,2)$ and $(2,4)$ are nonempty open sets that partition the set.

The second question from point-set topology that will concern us is: does the space "go up to its boundary," or does it "get arbitrarily close to its boundary without touching it?"

Definition 1.44 *Let (X, d) be a metric space. Given $C \subseteq X$, we say that C is a compact set provided that for all collections $\{U_\alpha\}_{\alpha \in J}$ of open sets of X, if the set $C \subseteq \bigcup_{\alpha \in J} U_\alpha$, then there exists a finite subcollection $\{U_1, U_2, ..., U_n\}$ of $\{U_\alpha\}_{\alpha \in J}$ such that $C \subseteq \bigcup_{k=1}^{n} U_k$.*

We say that a set is compact if "every open cover of the set has a finite subcover." Compactness is a bit more difficult to intuitively see than connectedness, but it will be of importance to us nonetheless. Here are some examples of non-compact sets.

Example 1.45 *The following sets are not compact.*
(i) The set $\mathbb{R}$ is not compact, because the collection $\{(-n, n)\}_{n=1}^{\infty}$ is an open cover

of $\mathbb{R}$ with no finite subcover.
(ii) The set $(-1,1)$ is not compact, because the collection $\left\{\left(\frac{1}{n}-1, 1-\frac{1}{n}\right)\right\}_{n=1}^{\infty}$ is an open cover of the set that has no finite subcover.
(iii) The set $\mathbb{R}^2$ is not compact, because the collection $\{B\left((0,0),n\right)\}_{n=1}^{\infty}$ is an open cover with no finite subcover.

There is a property of compact sets that will come to be relevant to our study of real analysis: closed subsets of compact spaces are compact.

Proposition 1.46 *Let (X,d) be a metric space, and let $S \subseteq X$ be compact. If $A \subseteq S$ is closed in S, then A is compact.*

Proof Let $\mathcal{U}$ be a collection of sets that are open in X such that $A \subseteq \bigcup \mathcal{U}$. Since A is closed in S, we have that $S \setminus A$ is open in S. Therefore,

$$S \subseteq \bigcup \left(\mathcal{U} \cup \{S \setminus A\}\right). \tag{23}$$

As S is compact, there exists a finite subset $\mathcal{V} \subseteq \mathcal{U} \cup \{S \setminus A\}$ such that $S \subseteq \bigcup \mathcal{V}$. However, $A \subseteq S$, so $A \subseteq \bigcup \mathcal{V}$. Thus, A is compact. □

1.2.3 Sequences in metric spaces

In real analysis, the term "sequence" will always refer to an "infinite sequence," formalized as follows.

Definition 1.47 *Let X be a set. A sequence in X is a function $s : \mathbb{Z}^+ \to X$.*

Although a sequence is only a typical function, our notation for a sequence will be somewhat different. We establish the following convention: given a sequence $K : \mathbb{Z}^+ \to X$, we will commonly write $K(n)$ as k_n. In general, if we write $K = (k_n)$, then we mean that K is a sequence and $\forall\, n \in \mathbb{Z}^+$, $K(n)$ is denoted k_n. This point, k_n, is often called the "nth term" or "nth element" of the sequence K.

In a metric space, we have a notion of distance, so we can talk about closeness. Therefore, given a sequence in a metric space, we can talk about a particular point to which the terms of the sequence "get close."

Definition 1.48 *Let (X, d) be a metric space, and let $S = (s_n)$ be a sequence in X. Given a point $x \in X$, we say that S converges to x in X, denoted $S \to x$ or $(s_n) \to x$, provided that $\forall\, \varepsilon \in \mathbb{R}^+$, $\exists\, n_\varepsilon \in \mathbb{Z}^+$ such that $\forall\, m \in \mathbb{Z}$, if $m \geq n_\varepsilon$, then $d(s_m, x) < \varepsilon$. In that case, we say that S is a convergent sequence. If $\forall\, x \in X$, $S \not\to x$, then we say that S is divergent in X.*

A convergent sequence is one whose terms "get closer and closer" to a particular point. Note that a sequence could be convergent in one metric space, but divergent in another. For example, consider the sequence

$$S = (3, 3.1, 3.14, 3.141, 3.1415, 3.14159, 3.141592, 3.1415926, ...) . \tag{24}$$

In $\mathbb{Q}$, S is divergent, since $\pi \notin \mathbb{Q}$. On the other hand, in $\mathbb{R}$, S is convergent, since $\pi \in \mathbb{R}$. (This is another consequence of the order completeness property of $\mathbb{R}$, which $\mathbb{Q}$ lacks.) Therefore, whether or not a sequence is convergent in X just depends on whether X contains a point of convergence of the sequence in question.

Now, given a sequence, how many points of convergence can it have?

Proposition 1.49 *Let (X, d) be a metric space, and let S be a sequence in X. Given $x, y \in X$, if $S \to x$ and $S \to y$, then $x = y$.*

Proof Let $\varepsilon \in \mathbb{R}^+$. We know that $\exists\, n_x \in \mathbb{R}^+$ such that $\forall\, m \geq n_x$, $d\left(s_m, x\right) < \frac{\varepsilon}{2}$. At the same time, $\exists\, n_y \in \mathbb{R}^+$ such that $\forall\, m \geq n_y$, $d\left(s_m, y\right) < \frac{\varepsilon}{2}$. Select a number $n = \max\left\{n_x, n_y\right\}$. We note that $\forall\, m \geq n$, the triangle inequality implies that

$$d\left(x, y\right) \leq d\left(x, s_m\right) + d\left(s_m, y\right) < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon. \tag{25}$$

We deduce that $d\left(x, y\right) < \varepsilon$, $\forall\, \varepsilon \in \mathbb{R}^+$. In that case, $d\left(x, y\right) = 0$, so $x = y$. $\square$

This is our first big application of the triangle inequality.

Now, how can we know when a sequence converges? The definition is often difficult to verify, since it only answers the question of whether a proposed limit of a sequence is truly a limit; it does not provide any hint as to what the limit should be, or that there even is one. We will begin to answer this question by first considering some properties that sequences *must* have in order to be convergent.

Definition 1.50 *Let (X, d) be a metric space, and let $S = (s_n)$ be a sequence in X. We say that X is a Cauchy sequence provided that $\forall\, \varepsilon \in \mathbb{R}^+$, $\exists\, n_\varepsilon \in \mathbb{Z}^+$ such that $\forall\, m_1, m_2 \in \mathbb{Z}^+$, if $m_1, m_2 \geq n$, then $d\left(x_{m_1}, x_{m_2}\right) < \varepsilon$.*

To summarize, a Cauchy sequence is a sequence whose terms "get closer and closer together."

Aren't convergent sequences Cauchy? Is it not reasonable to expect that a sequence whose terms get closer and closer to a particular point would also be a sequence whose terms get closer and closer to each other? The following theorem shows that the answer is yes.

Theorem 1.51 *Let (X, d) be a metric space. Given $x \in X$ and a sequence S in X, if $S \to x$, then S is Cauchy.*

Proof Suppose that $(x_n) \to x$. Let $\varepsilon \in \mathbb{R}^+$. There exists $n \in \mathbb{Z}^+$ such that if $m \in \mathbb{Z}^+$ such that $m \geq n$, then $d\left(x_m, x\right) < \frac{\varepsilon}{2}$. Let $m_1, m_2 \in \mathbb{Z}^+$ such that $m_1, m_2 \geq n$. By the triangle inequality,

$$d\left(x_{m_1}, x_{m_2}\right) \leq d\left(x_{m_1}, x\right) + d\left(x, x_{m_2}\right) < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon. \tag{26}$$

Ergo, (x_n) is Cauchy. □

Now, let's define a notion of boundedness for sequences in metric spaces. See Definition 1.33.

Definition 1.52 *Let (X, d) be a metric space. Given a sequence (x_n) of X, we say that S is a* <u>bounded sequence</u> *provided that the set $\left\{x_n \in X \middle| n \in \mathbb{Z}^+\right\}$ is metrically bounded in X.*

A bounded set is one that is a finite "distance across" in every direction. A bounded sequence is just a sequence whose range is bounded. The following lemma indicates that the "distance across" finitely many points is also finite.

Lemma 1.53 *Let (X, d) be a metric space, and let $A \subseteq X$ such that $A \neq \varnothing$. If $|A| = n$ for some $n \in \mathbb{Z}^+$, then A is metrically bounded in X.*

Proof This is **Exercise 1.67**. □

Lemma 1.54 *Let (X, d) be a metric space, and let $A, B \subseteq X$ such that $A, B \neq \varnothing$. The following statements are true.*
(i) If A and B are bounded in X, then $A \cap B$ is bounded in X.
(ii) If A and B are bounded in X, then $A \cup B$ is bounded in X.

Proof This is **Exercise 1.68**. □

Are Cauchy sequences bounded? After all, Cauchy sequences have a way of taking up a small amount of "space." The following proposition indicates that the answer is yes.

Proposition 1.55 *Let (X, d) be a metric space, and let S be a sequence in X. If S is Cauchy, then S is bounded.*

Proof Suppose that (x_n) is Cauchy. Given $\varepsilon \in \mathbb{R}^+$, we know that $\exists\, n \in \mathbb{Z}^+$ such that $\forall\, m_1, m_2 \in \mathbb{Z}^+$, if $m_1, m_2 \geq n$, then $d\left(x_{m_1}, x_{m_2}\right) < \varepsilon$. Consider

$$A = \left\{ x_m \in X \middle| d\left(x_m, x_n\right) < \varepsilon \right\} \tag{27}$$

and

$$B = \left\{ x_k \in X \middle| d\left(x_k, x_n\right) \geq \varepsilon \right\}. \tag{28}$$

We note that $B \subseteq \{x_1, x_2, ..., x_{n-1}\}$; after all, if $m \geq n$, then $x_m \in A$. Therefore, B is finite, and so B is bounded by Lemma 1.53. Moreover, A is bounded, since $\forall\, x_m \in A$, $d\left(x_m, x_n\right) < \varepsilon$. Therefore, since $\left\{ x_n \in X \middle| n \in \mathbb{Z}^+ \right\} = A \cup B$, Lemma 1.54 indicates that (x_n) is bounded. $\square$

To summarize: convergent implies Cauchy, which implies bounded. (Ergo, if one can show that a sequence is not bounded, then we know that it is not convergent.)

Aren't all Cauchy sequences convergent? After all, how can a sequence's terms get closer and closer together without them converging on some particular point? The answer is not so simple in general. When this is true, we have a special name for such metric spaces.

Definition 1.56 *A metric space (X, d) is [metrically] complete provided that for any Cauchy sequence (x_n) of X, $\exists\, x \in X$ such that $(x_n) \to x$.*

1.2.4 Functions in metric spaces

One of the defining characteristics of analysis as a field of mathematics is that analysis is concerned with limits. In metric spaces, this is possible because metric spaces have a notion of distance. However, before defining what's meant by a limit, we must first define what points in a metric space are valid for limits. Refer to Definition 1.40. In general, whenever we discuss $\lim_{x \to a} f(x)$ for a function $f : E \to Y$ and a point a, a must *always* be a limit point of E. Otherwise, this will not make sense.

What is the point of this? The point is to avoid pathologies like the following: define $f : (0, 1) \cup \{2\} \to \mathbb{R}$ via $f(t) = t^2$. Now, what is $\lim_{t \to 2} f(t)$? By defining what we mean by a limit point, we can formally say that this question is meaningless, because 2 is not a limit point of $(0, 1) \cup \{2\}$. On the other hand, $\lim_{t \to 1} f(t)$ *is* meaningful, because 1 is a limit point of $(0, 1) \cup \{2\}$. Thus, we can take the limit approaching points that are not necessarily in the domain of f, but they must be limit points. (At the same time, there could be some points in the domain that we can't approach in a limit.)

So, having dealt with the pathologies, let's define what we mean by a limit.

Definition 1.57 *Let (X, d_X) and (Y, d_Y) be metric spaces, and let $E \subseteq X$. Given $f : E \to Y$, a point $y \in Y$, and a limit point $a \in X$ of E, we say that y is the limit of f as the independent variable approaches a provided that $\forall\, \varepsilon \in \mathbb{R}^+$, $\exists\, \delta_\varepsilon \in \mathbb{R}^+$ such that if $x \in E$ and $d_X(x, a) < \delta_\varepsilon$, then $d_Y(f(x), y) < \varepsilon$.*

In introductory calculus, it is commonly taught that a limit must be unique; that is, from every approach, the limit is the same value. Otherwise, the limit is said to fail to exist. Our first result, therefore, is to show that this is the case.

Proposition 1.58 *Let (X, d_X) and (Y, d_Y) be metric spaces, and let $E \subseteq X$. Given $f : E \to Y$ and a limit point $a \in X$ of E, if y_1 is the limit of f as x approaches a and y_2 is the limit of f as x approaches a, then $y_1 = y_2$.*

Proof This is **Exercise 1.71**. □

Given that uniqueness is guaranteed, we will write $\lim_{x \to a} f(x) = y$ from now on.

We'll now define the most important type of function in analysis.

Definition 1.59 *Let (X, d_X) and (Y, d_Y) be metric spaces, and let $E \subseteq X$. Given $f : E \to Y$ and $a \in E$ that is a limit point of E, we say that f is continuous at a provided that $\forall\, \varepsilon \in \mathbb{R}^+$, $\exists\, \delta_\varepsilon \in \mathbb{R}^+$ such that $\forall\, x \in X$, if $d_X(x, a) < \delta_\varepsilon$, then $d_Y(f(x), f(a)) < \varepsilon$. If $\forall\, a \in E$, f is continuous at E, then we say that f is a continuous function.*

As **Exercise 1.73** indicates, this definition can be rephrased. However, notice that in the definition of $\lim_{x \to a} f(x)$, the point a had to be a limit point of the domain of f. For continuity, this condition is strengthened; a must be an accumulation point *and* a has to be contained in the domain.

The following theorem compiles all of the possible definitions of continuity of a function.

Theorem 1.60 *Let (X, d_X) and (Y, d_Y) be metric spaces, and let $E \subseteq X$. Given $f : E \to Y$, the following statements are equivalent.*
(i) f is continuous.
(ii) $\forall\, a \in E$, $\lim_{x \to a} f(x) = f(a)$.
(iii) $\forall\, a \in E$, $\forall\, V \subseteq Y$, if V is open in Y and $f(a) \in V$, then $\exists\, U \subseteq E$ that is open in E such that $a \in U$ and $f(U) \subseteq V$.
(iv) $\forall\, V \subseteq Y$, if V is open in Y, then $f^{-1}(V)$ is open in X.

Proof $(i \Rightarrow ii)$ This is part of **Exercise 1.73**.

$(ii \Rightarrow iii)$ Let $a \in E$. Suppose that $V \subseteq Y$ is open in Y and $f(a) \in V$. Since V is open, $\exists\, \varepsilon \in \mathbb{R}^+$ such that if $d_Y(y, f(a)) < \varepsilon$, then $y \in V$. As $\lim_{x \to a} f(x) = f(a)$, $\exists\, \delta \in \mathbb{R}^+$ such that if $d_X(x, a) < \delta$, then $d_Y(f(x), f(a)) < \varepsilon$. Ergo, if $x \in B(a, \delta)$, then $f(x) \in B(f(a), \varepsilon)$. This means that $f(B(a, \delta)) \subseteq B(f(a), \varepsilon) \subseteq V$. Since $B(a, \delta)$ is open in X, we can take $U = B(a, \delta)$.

$(iii \Rightarrow iv)$ Let $V \subseteq Y$ be open in Y, and let $a \in f^{-1}(V)$. This means that $f(a) \in V$. In that case, $\exists\, U \subseteq E$ that is open in E such that $a \in U$ and $f(U) \subseteq V$. We deduce that $U \subseteq f^{-1}(V)$. Now, since U is open, we know that $\exists\, \varepsilon \in \mathbb{R}^+$ such that if $d_X(x, a) < \delta$, then $x \in U \subseteq f^{-1}(V)$. Thus, $f^{-1}(V)$ is open in E.

$(iv \Rightarrow i)$ Let $a \in E$, and let $\varepsilon \in \mathbb{R}^+$. We know that $B(f(a), \varepsilon)$ is open in Y. Therefore, $f^{-1}(B(f(a), \varepsilon))$ is open in X. Additionally, $a \in f^{-1}(B(f(a), \varepsilon))$. Thus, $\exists\, \delta \in \mathbb{R}^+$ such that if $d_X(x, a) < \delta$, then $x \in f^{-1}(B(f(a), \varepsilon))$. This indicates that $f(x) \in B(f(a), \varepsilon)$, which means that $d_Y(f(x), f(a)) < \varepsilon$. Therefore, f is continuous. □

We now prove two crucial properties of continuous functions: preservation of connected and compact sets.

Theorem 1.61 *Let (X, d_X) and (Y, d_Y) be metric spaces, and let $E \subseteq X$. Given $f : E \to Y$, if $S \subseteq E$ is connected, then $f(S)$ is connected.*

Proof Let $S \subseteq E$ be connected. Let $U, V \subseteq Y$ be open in Y such that $f(S) \subseteq U \cup V$ and $U \cap V = \varnothing$. We will show that $f(S) \cap U = \varnothing$ or $f(S) \cap V = \varnothing$. By Theorem 1.60, we know that $f^{-1}(U)$ and $f^{-1}(V)$ are open in E. Additionally,

$$S \subseteq f^{-1}(U \cup V) = f^{-1}(U) \cup f^{-1}(V). \tag{29}$$

However,

$$f^{-1}(U) \cap f^{-1}(V) = f^{-1}(U \cap V) = f^{-1}(\varnothing) = \varnothing. \tag{30}$$

Since S is connected, we must have that $S \cap f^{-1}(U) = \varnothing$ or $S \cap f^{-1}(V) = \varnothing$. This implies that $f(S) \cap U = \varnothing$ or $f(S) \cap V = \varnothing$. Therefore, $f(S)$ is connected. □

Theorem 1.62 *Let (X, d_X) and (Y, d_Y) be metric spaces, and let $E \subseteq X$. Given $f : E \to Y$, if $C \subseteq E$ is compact, then $f(C)$ is compact.*

Proof Let $C \subseteq E$ be compact, and let $\{V_\alpha\}_{\alpha \in J}$ be a collection of open sets of Y such that $f(C) \subseteq \bigcup_{\alpha \in J} V_\alpha$. By Theorem 1.60, we know that $\forall\, \alpha \in J$, $f^{-1}(V_\alpha)$ is open in E. Additionally,

$$C \subseteq f^{-1}\left(\bigcup_{\alpha \in J} V_\alpha\right) = \bigcup_{\alpha \in J} f^{-1}(V_\alpha). \tag{31}$$

Now, since C is compact, there must exist a finite subcollection $\{f^{-1}(V_k)\}_{k=1}^{n}$ such that $C \subseteq \bigcup_{k=1}^{n} f^{-1}(V_k)$. However,

$$f(C) \subseteq f\left(\bigcup_{k=1}^{n} f^{-1}(V_k)\right) = f\left(f^{-1}\left(\bigcup_{k=1}^{n} V_k\right)\right) = \bigcup_{k=1}^{n} V_k \tag{32}$$

Therefore, $f(C)$ is compact. $\square$

1.2.5 Exercises

Example 1.63 *Let $A \subseteq \mathbb{R}$ such that $A \neq \varnothing$. Show that A is metrically bounded in $\mathbb{R}$ if and only if A is bounded above and bounded below in $\mathbb{R}$.*

Example 1.64 *Prove Theorem 1.38.*

Example 1.65 *Prove the following statements.*
(i) $U = \left\{(x, y) \in \mathbb{R}^2 \middle| y > 0\right\}$ is open in $\mathbb{R}^2$.
(ii) $U = \left\{(x, y) \in \mathbb{R}^2 \middle| xy = 0\right\}$ is closed in $\mathbb{R}^2$.
(iii) $U = \left\{(x, y) \in \mathbb{R}^2 \middle| x + y = 1\right\}$ is closed in $\mathbb{R}^2$.
(iv) $U = \bigcup_{n=1}^{\infty}(2n, 2n+1)$ is open in $\mathbb{R}$.
(v) $U = \bigcup_{n=1}^{\infty}[2n, 2n+1]$ is closed in $\mathbb{R}$.

Example 1.66 *Let $a, b \in \mathbb{R}$ such that $a < b$. Prove that $[a, b)$ is neither open nor closed in $\mathbb{R}$.*

Example 1.67 *Prove Lemma 1.53.*

Example 1.68 *Prove Lemma 1.54.*

Example 1.69 *Prove or disprove: if S is a sequence in a metric space (X, d) and S is bounded, then S is Cauchy.*

Example 1.70 *Give an example of a metric space that is not metrically complete.*

Example 1.71 *Prove Proposition 1.58.*

Example 1.72 *Let (X, d_X) and (Y, d_Y) be metric spaces, and let $E \subseteq X$. Given $f : E \to Y$ and a limit point $a \in E$, prove that $\lim\limits_{x \to a} f(x) = y$ if and only if for every sequence (x_n) of A such that $(x_n) \to a$, the sequence $(f(x_n)) \to y$.*

Example 1.73 *Let (X, d_X) and (Y, d_Y) be metric spaces, and let $E \subseteq X$. Given $f : E \to Y$ and $a \in E$, show that f is continuous at a if and only if the limit $\lim\limits_{x \to a} f(x) = f(a)$.*

1.3 Sequences in $\mathbb{R}$

One of the features of $\mathbb{R}$ that makes it so interesting among metric spaces is the fact that it has an ordering. This ordering allows us to talk about sequences that "go in a single direction," so to speak.

Definition 1.74 *Let (x_n) be a sequence in $\mathbb{R}$. We say that the sequence (x_n) is an increasing sequence provided that $\forall\, n \in \mathbb{Z}^+$, $x_n \leq x_{n+1}$. If $\forall\, n \in \mathbb{Z}^+$, $x_n < x_{n+1}$, then we say that (x_n) is strictly increasing.*

Definition 1.75 *Let (x_n) be a sequence in $\mathbb{R}$. We say that the sequence (x_n) is an decreasing sequence provided that $\forall\, n \in \mathbb{Z}^+$, $x_n \geq x_{n+1}$. If $\forall\, n \in \mathbb{Z}^+$, $x_n > x_{n+1}$, then we say that (x_n) is strictly decreasing.*

Definition 1.76 *Let (x_n) be a sequence in $\mathbb{R}$. We say that the sequence (x_n) is a monotone sequence provided that (x_n) is increasing or (x_n) is decreasing.*

Monotone sequences are interesting to us because for these types of sequences, bounded will imply convergent.

Theorem 1.77 *Let (x_n) be a monotone sequence in $\mathbb{R}$. Define $S = \{x_n \in \mathbb{R} | n \in \mathbb{Z}^+\}$. The following statements are true.*
(i) If S is increasing and bounded above, then $(x_n) \to \sup S$.
(ii) If S is decreasing and bounded below, then $(x_n) \to \inf S$.

Proof (i) Let (x_n) be an increasing sequence in $\mathbb{R}$ with $b = \sup \left\{x_n \in \mathbb{R} \middle| n \in \mathbb{Z}^+\right\}$. Let $\varepsilon \in \mathbb{R}^+$. Since b is the least upper bound, $\exists\, n_\varepsilon \in \mathbb{Z}^+$ such that $x_{n_\varepsilon} > b - \varepsilon$. In fact, as x_n is increasing, this means that $\forall\, m \in \mathbb{Z}^+$ such that $m \geq n_\varepsilon$, $x_m > b - \varepsilon$. We deduce that $\varepsilon > b - x_m$. Since $x_m < b$, $b - x_m > 0$, so $b - x_m = |x_m - b|$. This indicates that

$$d\,(x_m, b) = |x_m - b| = b - x_m < \varepsilon, \tag{33}$$

and so $(x_n) \to b$.

(ii) This is **Exercise 1.82**. $\square$

Notice that this theorem is entirely a consequence of the order completeness of the real line. Without order completeness, we cannot guarantee the existence of an infimum or supremum. The order completeness property is the reason that you are now studying "real analysis," as opposed to "rational analysis" or something of the sort.

Monotone sequences are useful, but at first glance, we might think them rather rare. After all, these are sequences which *only* decrease or *only* increase, for all $n \in \mathbb{Z}^+$. Given a random sequence, wouldn't it be extremely unlikely for it to be monotone? Indeed, that is true. However, monotone sequences appear more often than one might think.

Definition 1.78 *Let X be a set, and let S be a sequence in X. A* <u>*subsequence of S*</u> *is a sequence T in X such that $T = S \circ f$, where $f : \mathbb{N} \to \mathbb{N}$ is an injective function.*

In other words, a subsequence is just a selection of some elements of a sequence, in the order they appear. As it turns out, when we look at subsequences, monotone sequences become a bit more common.

Theorem 1.79 *Let S be a sequence in $\mathbb{R}$. There exists a subsequence T of S that is monotone.*

Proof Let (x_n) be a sequence in $\mathbb{R}$. Consider the set

$$P = \left\{ p \in \mathbb{Z}^+ \middle| \forall\, m \in \mathbb{Z}^+, \text{ if } p \leq m, \text{ then } x_p \geq x_m \right\}. \tag{34}$$

We consider two cases: either P is infinite, or P is finite.

Consider the case that P is infinite. We define a sequence (p_n) in P such that $\forall\, m, n \in \mathbb{Z}^+$, $p_m \leq p_n$ if and only if $m \leq n$. Now we consider the subsequence (x_{p_n}) of (x_n). For each $n \in \mathbb{Z}^+$, $n \leq n+1$, so $p_n \leq p_{n+1}$, which implies that $x_{p_n} \geq x_{p_{n+1}}$. Ergo, (x_{p_n}) is a decreasing sequence, which is monotone.

Consider the case that P is finite. Suppose that $P = \{p_1, p_2, ..., p_n\}$, where $p_1 \leq p_2 \leq ... \leq p_n$. Let $r_1 = p_n + 1$. Given $n \in \mathbb{Z}^+$, since $r_n \notin P$, $\exists\, r_{n+1} \in \mathbb{Z}^+$ such that $x_{r_{n+1}} > x_{r_n}$. Therefore, (x_{r_n}) is an increasing subsequence of (x_n), which is monotone. □

The following is known as the Bolzano-Weierstrass theorem.

Theorem 1.80 *Let S be a sequence in $\mathbb{R}$. If S is bounded, then there exists a subsequence T of S such that $T \to x$ for some $x \in \mathbb{R}$.*

Proof By Theorem 1.79, there exists a monotone susbsequence T of S. By Theorem 1.77, T must converge in $\mathbb{R}$. □

In **Exercise 1.70**, we asked you to provide an example of a metric space that is not metrically complete. In the special case of sequences in $\mathbb{R}$, Cauchy sequences *do* converge. Therefore, the following theorem says that $\mathbb{R}$ is metrically complete.

Theorem 1.81 *Given a sequence S in $\mathbb{R}$, S is convergent in $\mathbb{R}$ if and only if S is Cauchy.*

Proof ($\Rightarrow$) If S is convergent, then Theorem 1.51 indicates that S must be Cauchy.

($\Leftarrow$) Assume that (x_n) is a Cauchy sequence in $\mathbb{R}$. By Proposition 1.55, we know that the set $X = \left\{x_n \in \mathbb{R} \middle| n \in \mathbb{Z}^+\right\}$ is metrically bounded in $\mathbb{R}$. Therefore, the Bolzano-Weierstrass theorem (Theorem 1.80) indicates that there exists a convergent subsequence (y_n) of (x_n). Suppose $(y_n) \to y$.

We claim that $(x_n) \to y$. Let $\varepsilon \in \mathbb{R}^+$. Select $n_1 \in \mathbb{Z}^+$ such that $\forall\, m \in \mathbb{Z}^+$ with $m \geq n_1$, $d(x_m, y) < \frac{\varepsilon}{2}$. Also select $n_2 \in \mathbb{Z}^+$ such that $\forall\, m_1, m_2 \in \mathbb{Z}^+$, if $m_1, m_2 \geq n_2$, then $d(x_{m_1}, x_{m_2}) < \frac{\varepsilon}{2}$. Define $n_\varepsilon = \max\{n_1, n_2\}$. Now, select $m, k \in \mathbb{Z}^+$ such that $m, k \geq n_\varepsilon$ and $x_k = y_l$ for some $l \in \mathbb{Z}^+$. In that case,

$$d(x_m, y) \leq d(x_m, x_k) + d(x_k, y) = d(x_m, x_k) + d(y_l, y) < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon. \quad (35)$$

Ergo, $(x_n) \to y$. □.

We have already emphasized this, but just to beat the dead horse some more, as **Exercise 1.70** indicates, this result is *not* true in general. It is a special property of $\mathbb{R}$.

1.3.1 Exercises

Example 1.82 *Prove that if* (x_n) *is a bounded and decreasing sequence in* $\mathbb{R}$*, then* $(x_n) \to a$*, where* $a = \inf \left\{ x_n \in \mathbb{R} \middle| n \in \mathbb{Z}^+ \right\}$.

1.4 Topology of $\mathbb{R}$

Proposition 1.83 *Let $a, b \in \mathbb{R}$ such that $a < b$. The closed interval $[a, b]$ is connected.*

Proof Assume, with the expectation of a contradiction, that $[a, b]$ is not connected. In that case, $\exists\, U_0, V_0 \subseteq \mathbb{R}$ that are nonempty and open in $\mathbb{R}$ such that $[a, b] \subseteq U_0 \cup V_0$ and $U_0 \cap V_0 = \varnothing$. Define $U = U_0 \cap [a, b]$ and $V = V_0 \cap [a, b]$. Since $U \cup V = [a, b]$, we must have that $b \in U$ or $b \in V$. Assume, with the understanding that the other choice is similar, that $b \in V$. Now, we define $u = \sup U$. We know that $u \in [a, b]$, so $u \in U$ or $u \in V$.

Consider the case that $u \in U$. We know that $u < b$, since $b \in V$. Since U_0 is open and $u \in U_0$, $\exists\, \varepsilon \in \mathbb{R}^+$ such that $(u - \varepsilon, u + \varepsilon) = B\,(u, \varepsilon) \subseteq U_0$. Thus, if we define $\delta = \min\{\varepsilon, b - u\}$, then $[u, u + \delta) \subseteq [a, b]$ and at the same time, $[u, u + \delta) \subseteq (u - \varepsilon, u + \varepsilon) \subseteq U_0$. This implies that $[u, u + \delta) \subseteq U_0 \cap [a, b] = U$. However, this means that $u + \frac{\delta}{2} \in U$. By assumption, $u = \sup U$, so this means that $u + \frac{\delta}{2} \leq u$, hence $\delta \leq 0$, despite that $\delta > 0$ by definition. This is a contradiction.

Consider the case that $u \in V$. We know that $u \in V_0$, so $\exists\, \varepsilon \in \mathbb{R}^+$ such that $B(u, \varepsilon) = (u - \varepsilon, u + \varepsilon) \subseteq V_0$. We also know that $u > a$, since if $u = a$, then $U = \{a\}$, which is not open in $[a, b]$. Choose $\delta = \min\{\varepsilon, u - a\}$. We now have that $(u - \delta, u] \subseteq (u - \varepsilon, u + \varepsilon) \subseteq V_0$, while at the same time, $(u - \delta, u] \subseteq [a, b]$. We deduce that $(u - \delta, u] \subseteq V_0 \cap [a, b] = V$. As $U \cap V = \varnothing$, $U \cap (u - \delta, u] = \varnothing$. However, this implies that $\forall\, x \in U$, $x \leq u - \frac{\delta}{2}$, despite that $u - \frac{\delta}{2} < u$ and $u = \sup U$, This is a contradiction.

Whatever the case, we reach a contradiction that leads us to conclude that our assumption that $[a, b]$ is disconnected is false; $[a, b]$ is connected. □

Proposition 1.84 *Let $a, b \in \mathbb{R}$ such that $a < b$. The closed interval $[a, b]$ is compact.*

Proof Let $\mathcal{U}$ be a collection of sets that are open in $[a, b]$ such that $[a, b] \subseteq \bigcup \mathcal{U}$. We define

$$S = \left\{ x \in [a,b] \,\middle|\, \exists\, U_1, U_2, ..., U_n \in \mathcal{U} \text{ such that } [a,x] \subseteq \bigcup_{k=1}^{n} U_k \right\}. \tag{36}$$

We know that $S \neq \varnothing$, since $a \in S$. Let $s_0 = \sup S$.

We claim that $s_0 > a$. Since $a \in [a,b]$, $\exists\, U \in \mathcal{U}$ such that $a \in U$. As U is open in $[a,b]$, $\exists\, \varepsilon \in \mathbb{R}^+$ such that $[a, a+\varepsilon) \subseteq U$. This implies that $\left[a, a+\frac{\varepsilon}{2}\right] \subseteq \bigcup_{k=1}^{1} U$, and so $a + \frac{\varepsilon}{2} \in S$. Therefore, $s_0 \geq a + \frac{\varepsilon}{2} > a$.

We claim that $s_0 \in S$. Since $s_0 \in [a,b]$, $\exists\, U_0 \in \mathcal{U}$ such that $s_0 \in U_0$. As U_0 is open in $[a,b]$, $\exists\, \varepsilon \in \mathbb{R}^+$ such that $(s_0 - \varepsilon, s_0] \subseteq U_0$. Since $s_0 - \frac{\varepsilon}{2} \neq \sup S$, we know that $\exists\, s_1 \in \left(s_0 - \frac{\varepsilon}{2}, s_0\right] \cap S$. This means that $[a, s_1] \subseteq \bigcup_{k=1}^{n} U_k$ for some $U_1, U_2, ..., U_n \in \mathcal{U}$. However,

$$[a, s_0] = [a, s_1] \cup (s_0 - \varepsilon, s_0] \subseteq \left(\bigcup_{k=1}^{n} U_k \right) \cup U_0 = \bigcup_{k=0}^{n} U_k. \tag{37}$$

Therefore, $s_0 \in S$.

We claim that $s_0 = b$. Assume, with the expectation of a contradiction, that $s_0 < b$. As U_0 is open in $[a,b]$, $U_0 = V \cap [a,b]$ for some $V \subseteq \mathbb{R}$ that is open in $\mathbb{R}$. Therefore, $\exists\, \varepsilon \in \mathbb{R}^+$ such that $(s_0 - \varepsilon, s_0 + \varepsilon) = B(s_0, \varepsilon) \subseteq V$. We now define $\delta = \min\{\varepsilon, b - s_0\}$. As $s_0 < b$, we know that $\delta > 0$. Since $\delta \leq \varepsilon$, we know that $[s_0, s_0 + \delta) \subseteq (s_0 - \varepsilon, s_0 + \varepsilon) \subseteq V$. At the same time, since $\delta \leq b - s_0$, $[s_0, s_0 + \delta) \subseteq [a,b]$. Thus, $[s_0, s_0 + \delta) \subseteq V \cap [a,b] = U_0$. This tells us that $s_0 + \frac{\delta}{2} \in U_0$. However,

$$\left[a, s_0 + \frac{\delta}{2}\right] \subseteq [a, s_0] \cup [s_0, s_0 + \delta) \subseteq \left(\bigcup_{k=0}^{n} U_k \right) \cup U_0 = \bigcup_{k=0}^{n} U_k. \tag{38}$$

This indicates that $s_0 + \frac{\delta}{2} \in S$, despite that $s_0 = \sup S$. This contradiction leads us to conclude that our assumption that $s_0 < b$ is false; $s_0 = b$.

Finally, since $b = \sup S \in S$, we have that $[a,b] \subseteq \bigcup_{k=0}^{n+1} U_k$ for some finitely many $U_0, U_1, ..., U_n \in \mathcal{U}$. Ergo, $[a,b]$ is compact. $\square$

The following is known as the Heine-Borel theorem.

Theorem 1.85 *Given $S \subseteq \mathbb{R}^n$, S is compact if and only if S is bounded and closed in $\mathbb{R}^n$.*

Proof ($\Rightarrow$) Let $S \subseteq \mathbb{R}^n$ be compact.

We will show that S is bounded. Given any $s \in S$, consider the collection $\{B(s,k)\}_{k\in\mathbb{Z}^+}$. We know that $\bigcup_{k=1}^{\infty} B(s,k) = \mathbb{R}^n$, so $S \subseteq \bigcup_{k=1}^{\infty} B(s,k)$. Since S is compact, $\exists\ B(s,k_1), B(s,k_2), ..., B(s,k_r)$ such that $S \subseteq \bigcup_{i=1}^{r} B(s,k_i)$. By taking $k = \max\{k_1, k_2, ..., k_r\}$, we deduce that $S \subseteq B(s,k)$, and so S is bounded.

We will show that S is closed in $\mathbb{R}^n$. Let $x \in \mathbb{R}^n \setminus S$. For each $s \in S$, we define $\varepsilon_s \in \mathbb{R}^+$ such that $x \notin B(s,\varepsilon_s)$. We note that $S \subseteq \bigcup_{s\in S} B(s,\varepsilon_s)$. Therefore, as S is compact, $\exists\ s_1, s_2, ..., s_r \in S$ such that $S \subseteq \bigcup_{k=1}^{r} B(s_k, \varepsilon_{s_k})$. For $k \in \{1,2,...,r\}$, define $d_k = d(x,s_k)$. By assumption, $x \notin B(s_k,\varepsilon_{s_k})$, which indicates that $d_k = d(x,s_k) > \varepsilon_{s_k}$. Take $\delta = \min\left\{d_k - \varepsilon_{s_k} \middle| k \in \{1,2,...,r\}\right\}$. By construction, $B(x,\delta) \cap B(s_k,\varepsilon_{s_k}) = \varnothing$, so

$$B(x,\delta) \subseteq \mathbb{R}^n \setminus \left(\bigcup_{k=1}^{r} B(s_k,\varepsilon_{s_k})\right) \subseteq \mathbb{R}^n \setminus S. \tag{39}$$

Ergo, $\mathbb{R}^n \setminus S$ is open, so S is closed.

($\Leftarrow$) Let $S \subseteq \mathbb{R}^n$ be closed in $\mathbb{R}^n$ and bounded. In that case, we know that $\exists\ x = (x_1, x_2, ..., x_n) \in S$ and $\alpha \in \mathbb{R}^+$ such that $S \subseteq B(x,\alpha)$. We note that

$$S \subseteq [x_1 - \alpha, x_1 + \alpha] \times [x_2 - \alpha, x_2 + \alpha] \times ... \times [x_n - \alpha, x_n + \alpha]. \tag{40}$$

By Proposition 1.84, we know that for each $i \in \{1,2,...,n\}$, $[x_i - \alpha, x_i + \alpha]$ is compact. One can show that a product of finitely many compact sets is also compact. (This is a result from topology that we will not prove here.) Therefore, since S is closed, Proposition 1.46 indicates that S must also be compact. $\square$

Proposition 1.86 *Let $S \subseteq \mathbb{R}$ be nonempty. If S is bounded and closed in $\mathbb{R}$, then* $\sup S \in S$.

Proof Since S is bounded, $\sup S$ and $\inf S$ are both finite. Define $b = \sup S$. Let $\varepsilon \in \mathbb{R}^+$. Take $x = b - \varepsilon$. If $\forall\, s \in S$, $s \leq x$, then this would mean that $x = \sup S$. Since $x \neq \sup S$, $\exists\, s \in S$ such that $x < s$. Ergo, $b - \varepsilon < s$, which indicates that $d(s, b) = |b - s| < \varepsilon$. Thus, b is a limit point of S. Therefore, by Proposition 1.41, we have that $b \in S$. $\square$

1.4.1 Exercises

Example 1.87 *Let* $S \subseteq \mathbb{R}$ *be nonempty. Prove that if* S *is bounded and closed in* $\mathbb{R}$*, then* $\inf S \in S$.

1.5 Real-valued functions

Let's try to rigorously verify the limit laws that we learned first in calculus.

Definition 1.88 *Let X be a set, and let $f, g : X \to \mathbb{R}$.*
(i) The sum of f and g is the function $f+g : X \to \mathbb{R}$ via $(f+g)(x) = f(x)+g(x)$.
(ii) The product of f and g is the function $fg : X \to \mathbb{R}$ via $(fg)(x) = f(x)g(x)$.
(iii) The quotient of f and g is the function $\frac{f}{g} : E \to \mathbb{R}$ via $\left(\frac{f}{g}\right)(x) = \frac{f(x)}{g(x)}$, where we define $E = \{x \in X \,|\, g(x) \neq 0\}$.

Proposition 1.89 *Let (X, d) be a metric space, and let $E \subseteq X$. Given functions $f, g : X \to \mathbb{R}$ and a limit point $a \in X$ of E, suppose that $\lim_{x \to a} f(x) = y$ and $\lim_{x \to a} g(x) = z$. The following statements are true.*
(i) $\lim_{x \to a} (f+g)(x) = y + z$.
(ii) $\lim_{x \to a} (fg)(x) = yz$.
(iii) If $z \neq 0$ and $\forall\, x \in E$, $g(x) \neq 0$, then $\lim_{x \to a} \left(\frac{f}{g}\right)(x) = \frac{y}{z}$.

Proof (i) Let $\varepsilon \in \mathbb{R}^+$. Let $\delta_1 \in \mathbb{R}^+$ such that if $d(x,a) < \delta_1$, then $d(f(x), y) < \frac{\varepsilon}{2}$. Also let $\delta_2 \in \mathbb{R}^+$ such that if $d(x,a) < \delta_2$, then $d(g(x), z) < \frac{\varepsilon}{2}$. Now, define $\delta_\varepsilon = \min\{\delta_1, \delta_2\}$. If $d(x,a) < \delta_\varepsilon$, then both $|f(x) - y| < \frac{\varepsilon}{2}$ and $|g(x) - z| < \frac{\varepsilon}{2}$. In that case,

$$|f(x) - y| + |g(x) - z| < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon. \tag{41}$$

However, by Theorem 1.24, this implies that

$$\begin{aligned} \big|\, (f(x) + g(x)) - (y+z) \,\big| &= \big|\, (f(x) - y) + (g(x) - z) \,\big| \\ &\leq |f(x) - y| + |g(x) - z| < \varepsilon, \end{aligned} \tag{42}$$

which proves that $\lim_{x \to a} (f+g)(x) = y + z$.

(ii) Let $\varepsilon \in \mathbb{R}^+$. We consider two cases: either $y = 0$ or $y \neq 0$.

Consider the case that $y \neq 0$. Select $\delta_1 \in \mathbb{R}^+$ such that if $d(x,a) < \delta_1$, then

$$|f(x) - y| < \frac{\varepsilon}{2\,(|z| + 1)}. \tag{43}$$

At the same time, select $\delta_2 \in \mathbb{R}^+$ such that if $d(x,a) < \delta_2$, then

$$|g(x) - z| < \min\left\{1, \frac{\varepsilon}{2|y|}\right\}. \tag{44}$$

Now, we define $\delta_\varepsilon = \min\{\delta_1, \delta_2\}$. Given $x \in X$, if $d(x,a) < \delta_\varepsilon$, then the following statements are true. First,

$$|g(x)| - |z| \leq |g(x) - z| < 1, \tag{45}$$

which implies that $|g(x)| < |z| + 1$. Therefore,

$$|f(x) - y||g(x)| < \frac{\varepsilon}{2\left(|z| + 1|\right)}\left(|z| + 1\right) = \frac{\varepsilon}{2}. \tag{46}$$

Second,

$$|g(x) - z||y| < \frac{\varepsilon}{2|y|}|y| = \frac{\varepsilon}{2}. \tag{47}$$

By adding these,

$$|f(x) - y||g(x)| + |g(x) - z||y| < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon. \tag{48}$$

Ergo, $|(f(x) - y)g(x)| + |(g(x) - z)y| < \varepsilon$. By Theorem 1.24,

$$|(f(x) - y)g(x) + (g(x) - z)y| \leq |(f(x) - y)g(x)| + |(g(x) - z)y| < \varepsilon. \tag{49}$$

However, this means that

$$\begin{aligned}\varepsilon > |(f(x) - y)g(x) + (g(x) - z)y| &= |f(x)g(x) - yg(x) + g(x)y - yz| \\ &= |f(x)g(x) - g(x)y + g(x)y - yz| = |f(x)g(x) - yz|,\end{aligned} \tag{50}$$

and so $\lim\limits_{x \to a}(fg)(x) = yz$.

Consider the case that $y = 0$. Select $\delta_1 \in \mathbb{R}^+$ such that if $d(x,a) < \delta_1$, then $|f(x) - y| < \frac{\varepsilon}{|z|+1}$. Also select $\delta_2 \in \mathbb{R}^+$ such that if $d(x,a) < \delta_2$, then

$|g(x) - z| < 1$. Now let $\delta_\varepsilon = \min\{\delta_1, \delta_2\}$. We know that

$$|g(x)| - |z| \leq |g(x) - z| < 1. \tag{51}$$

Therefore, $|g(x)| < |z| + 1$. In that case,

$$|f(x)g(x) - yg(x)| = |f(x) - y||g(x)| < \frac{\varepsilon}{|z|+1}(|z|+1) = \varepsilon. \tag{52}$$

However, $y = 0$, so

$$|f(x)g(x) - yz| = |f(x)g(x) - 0z| = |f(x)g(x)| < \varepsilon, \tag{53}$$

which demonstrates that $\lim_{x\to a} f(x)g(x) = yz$.

(iii) By part (ii), if $\lim_{x\to a} \frac{1}{g(x)} = \frac{1}{z}$, then we can say that $\lim_{x\to a} \frac{f(x)}{g(x)} = \frac{y}{z}$. This is **Exercise 1.111**. □

1.5.1 Continuous real-valued functions

See Definition 1.59. In the context of the metric space $\mathbb{R}$, this definition becomes the following.

Definition 1.90 *Let $f : [a, b] \to \mathbb{R}$. Given $x_0 \in [a, b]$, we say that f is continuous at x_0 provided that $\forall\, \varepsilon \in \mathbb{R}^+$, $\exists\, \delta \in \mathbb{R}^+$ such that $\forall\, x \in [a, b]$, if $|x - x_0| < \delta$, then $|f(x) - f(x_0)| < \varepsilon$. If $\forall\, x_0 \in [a, b]$, f is continuous at x_0, then we say that f is a continuous function.*

We remind the reader that, as Theorem 1.60 indicates, this is equivalent to saying that $\lim\limits_{x \to x_0} f(x) = f(x_0)$.

Let's rigorously prove a few of the familiar theorems about continuous functions that we first saw in introductory calculus.

Definition 1.91 *Let X be a set, and let $f : X \to \mathbb{R}$ be a function. Given $a \in X$, we say that f has an absolute maximum or global maximum at a provided that $\forall\, x \in X$, $f(x) \leq f(a)$. In that case, we say that $f(a)$ is an absolute maximum value of f, and that $(a, f(a)$ is an absolute maximum point of f.*

Definition 1.92 *Let X be a set, and let $f : X \to \mathbb{R}$ be a function. Given $a \in X$, we say that f has an absolute minimum or global minimum at a provided that $\forall\, x \in X$, $f(x) \geq f(a)$. In that case, we say that $f(a)$ is an absolute minimum value of f, and that $(a, f(a)$ is an absolute minimum point of f.*

The following is the extreme value theorem.

Theorem 1.93 *Let $f : [a, b] \to \mathbb{R}$. If f is continuous, then $\exists\, c_1, c_2 \in [a, b]$ such that f has an absolute minimum at c_1 and f has an absolute maximum at c_2.*

Proof 1 By Proposition 1.84, we know that $[a, b]$ is compact. By Theorem 1.62, this means that $f([a, b])$ must be compact. The Heine-Borel theorem (Theorem 1.85) now indicates that $f([a, b])$ must be closed and bounded. Therefore, Proposition 1.86 and Exercise 1.87 indicate that $\exists\, y_1, y_2 \in f([a, b])$ such that $\forall\, y \in f([a, b])$, $y_1 \leq y \leq y_2$. By definition of $f([a, b])$, this indicates that $\exists\, c_1, c_2 \in [a, b]$ such that

$\forall\, x \in [a, b]$, $f(c_1) \leq f(x) \leq f(c_2)$. $\square$

Proof 2 See Bartle and Sherbert, page 136. $\square$

The following theorem is the primary motivation for the definition of continuity in the real numbers, which proves that the graph of a continuous function is connected. The following is the intermediate value theorem.

Theorem 1.94 *Let $f : [a, b] \to \mathbb{R}$ be continuous, and let $y \in \mathbb{R}$. If $f(a) \leq y \leq f(b)$ or $f(b) \leq y \leq f(a)$, then $\exists\, c \in [a, b]$ such that $f(c) = y$.*

Proof 1 Suppose that $f(a) \leq f(b)$, and let $f(a) \leq y \leq f(b)$. Assume, with the expectation of a contradiction, that $y \notin f([a, b])$. Proposition 1.83 indicates that $[a, b]$ is connected. Therefore, Theorem 1.61 requires that $f([a, b])$ is connected. However, $f([a, b]) \subseteq (-\infty, y) \cup (y, \infty)$, $f(a) \in (-\infty, y)$, and $f(b) \in (y, \infty)$. This shows that $f([a, b])$ is disconnected. This contradiction leads us to conclude that our assumption that $y \notin f([a, b])$ is false; $y \in f([a, b])$. $\square$

Proof 2 See Bartle and Sherbert, pages 137-138. $\square$

1.5.2 Differentiable functions

One of the primary triumphs of the theory of limits is that it allows us to formally define the rate of change of one variable with respect to another: the derivative of a function.

Definition 1.95 *Let $f : (a, b) \to \mathbb{R}$. Given $c \in (a, b)$, we say that f is <u>differentiable at c</u> provided that $\exists\, d \in \mathbb{R}$ such that*

$$\lim_{x \to c} \frac{f(x) - f(c)}{x - c} = d. \tag{54}$$

In that case, we say that d is the <u>derivative of f at c</u>, denoted $f'(c)$. If $\forall\, x \in (a, b)$, f is differentiable at x, then we say that <u>f is a differentiable function</u>.

Definition 1.96 *Let $f : (a, b) \to \mathbb{R}$ be a differentiable function. The <u>derivative function of f</u> is the function $Df : (a, b) \to \mathbb{R}$ such that $\forall\, c \in (a, b)$, $Df(c)$ is the derivative of f at c.*

If we've done everything right so far, then calculus tells us that the derivative of a constant ought to be zero, and so it is.

Proposition 1.97 *Given $f : (a, b) \to \mathbb{R}$, if $\exists\, c \in \mathbb{R}$ such that $\forall\, x \in (a, b)$, $f(x) = c$, then $\forall\, x \in (a, b)$, $f'(x) = 0$.*

Proof This is **Exercise 1.114**. □

Additionally, differentiable functions should be continuous.

Proposition 1.98 *Let $f : (a, b) \to \mathbb{R}$. Given $c \in (a, b)$, if f is differentiable at c, then f is continuous at c.*

Proof Suppose that $f'(c)$ exists. We note that $f(x) - f(c) = \frac{f(x)-f(c)}{x-c}(x - c)$, so using Proposition 1.89,

$$\begin{aligned}\lim_{x \to c} (f(x) - f(c)) &= \lim_{x \to c} \left(\frac{f(x) - f(c)}{x - c} (x - c) \right) \\ &= \lim_{x \to c} \frac{f(x) - f(c)}{x - c} \lim_{x \to c} (x - c) = f'(c)(0) = 0. \end{aligned} \tag{55}$$

We deduce that $\lim_{x \to c} f(x) = f(c)$. □

However, as Exercise **1.115** demands, the converse of this is not true.

Let's discuss critical points. First, we'll define local extrema.

Definition 1.99 *Let (X, d) be a metric space, and let $f : X \to \mathbb{R}$. Given $c \in X$, we say that f has a relative maximum or local maximum at c provided that $\exists\, \delta \in \mathbb{R}^+$ such that $\forall\, x \in X$, if $d\,(x, c) < \delta$, then $f(x) \leq f(c)$.*

Definition 1.100 *Let (X, d) be a metric space, and let $f : X \to \mathbb{R}$. Given $c \in X$, we say that f has a relative minimum or local minimum at c provided that $\exists\, \delta \in \mathbb{R}^+$ such that $\forall\, x \in X$, if $d\,(x, c) < \delta$, then $f(x) \geq f(c)$.*

The entire procedure of optimization is based around the idea that the derivative at relative extrema must either be zero or fail to exist. We must prove that this is true. The following is known as Fermat's theorem on relative extrema.

Theorem 1.101 *Let $f : (a, b) \to \mathbb{R}$ be a differentiable function. Given $c \in (a, b)$, if f has a relative minimum or relative maximum at c, then $f'(c) = 0$.*

Proof Coming soon.

The following is known as Rolle's theorem.

Theorem 1.102 *Let $f : [a, b] \to \mathbb{R}$ be continuous. If f is differentiable on (a, b) and $f(a) = f(b)$, then $\exists\, c \in (a, b)$ such that $f'(c) = 0$.*

Proof By the extreme value theorem (Theorem 1.93), we know that $\exists\, c \in [a, b]$ such that $\forall\, x \in [a, b]$, $f(x) \leq f(c)$. If $f(c) = f(a)$ or $f(c) = f(b)$, then f is a constant function, in which case Proposition 1.97 indicates that $f'(c) = 0$. Consider the case that $f(a) < f(c)$. In that case, $c \in (a, b)$ and f has a relative maximum at c, so $f'(c) = 0$ by Fermat's theorem on relative extrema (Theorem 1.101). □

The following is the mean value theorem.

Theorem 1.103 *Let* $f : [a, b] \to \mathbb{R}$ *be continuous. If* f *is differentiable on* (a, b)*, then* $\exists\, c \in (a, b)$ *such that*

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

Proof We define $g : [a, b] \to \mathbb{R}$ via

$$g(x) = f(x) - \left(\frac{f(b) - f(a)}{b - a} \right) (x - a). \tag{56}$$

We note that $g(b) = f(a) = g(a)$. Therefore, by Rolle's theorem (Theorem 1.102), $\exists\, c \in (a, b)$ such that $g'(c) = 0$. However,

$$g'(x) = f'(x) - \frac{f(b) - f(a)}{b - a}. \tag{57}$$

Thus, $0 = f'(c) - \frac{f(b)-f(a)}{b-a}$, so $f'(c) = \frac{f(b)-f(a)}{b-a}$. □

Proposition 1.104 *Given* $f : (a, b) \to \mathbb{R}$*,* $\exists\, c \in \mathbb{R}$ *such that* $\forall\, x \in (a, b)$*,* $f(x) = c$ *if and only if* $\forall\, x \in (a, b)$*,* $f'(x) = 0$*.*

Proof $(\Rightarrow)$ This is exactly Proposition 1.97.

$(\Leftarrow)$ Suppose that $\forall\, x \in (a, b)$, $f'(x) = 0$. Let $x_1, x_2 \in (a, b)$. We note that f is continuous on $[x_1, x_2]$ and differentiable on (x_1, x_2). By the mean value theorem (Theorem 1.103), $\exists\, c \in (x_1, x_2)$ such that

$$\frac{f(x_2) - f(x_1)}{x_2 - x_1} = f'(c). \tag{58}$$

However, by assumption, $f'(c) = 0$. Therefore, $\frac{f(x_2)-f(x_1)}{x_2-x_1} = 0$, which indicates that $f(x_2) - f(x_1) = 0$. Thus, $f(x_1) = f(x_2)$. □

Proposition 1.105 *Given* $f, g : (a, b) \to \mathbb{R}$*,* $\forall\, x \in (a, b)$*,* $f'(x) = g'(x)$ *if and only if* $\exists\, c \in \mathbb{R}$ *such that* $\forall\, x \in (a, b)$*,* $f(x) - g(x) = c$*.*

Proof This is **Exercise 1.113**. □

Proposition 1.106 *Let* $f, g : (a, b) \to \mathbb{R}$*. If* f *and* g *are differentiable, then*

$$D(f + g) = Df + Dg$$

Proof The is **Exercise 1.116**. □

Proposition 1.107 *Let* $f : (a, b) \to \mathbb{R}$*. If* f *is differentiable, then* $\forall\, r \in \mathbb{R}$*,*

$$D(rf) = rDf.$$

Proof This is **Exercise 1.117**. □

The following is the product rule.

Theorem 1.108 *Let* $f, g : (a, b) \to \mathbb{R}$*. If* f *and* g *are differentiable, then*

$$D(fg) = fDg + gDf.$$

Proof This is **Exercise 1.118**. □

The following is the quotient rule.

Theorem 1.109 *Let* $f, g : (a, b) \to \mathbb{R}$*. If* f *and* g *are differentiable and for every* $x \in (a, b)$*,* $g(x) \neq 0$*, then*

$$D\left(\frac{f}{g}\right) = \frac{gDf - fDg}{g^2}.$$

Proof This is **Exercise 1.119**.

The following is the chain rule.

Theorem 1.110 *Let* $f : (a, b) \to \mathbb{R}$ *and* $g : (c, d) \to \mathbb{R}$ *such that* $f(a, b)) \subseteq (c, d)$. *If* f *and* g *are differentiable, then,*

$$D(g \circ f) = (Dg \circ f) Df.$$

Proof

1.5.3 Exercises

Example 1.111 *Let (X, d) be a metric space, and let $E \subseteq X$. Given a function $g : E \to \mathbb{R}$ and a limit point $a \in X$ of E, suppose that $\lim_{x \to a} g(x) = z$, where $z \neq 0$ and $\forall\, x \in E$, $g(x) \neq 0$. Show that $\lim_{x \to a} \frac{1}{g(x)} = \frac{1}{z}$.*

Example 1.112 *Prove that the following functions are continuous.*
(i) $f : \mathbb{R} \to \mathbb{R}$ via $f(x) = mx + b$ for some $m, b \in \mathbb{R}$.
(ii) $f : \mathbb{R} \to \mathbb{R}$ via $f(x) = x^2$.
(iii) $f : \mathbb{R} \to \mathbb{R}$ via $f(x) = x^3$.
(iv) $f : \mathbb{R} \to \mathbb{R}$ via $f(x) = x^n$.
(v) $f : \mathbb{R} \to \mathbb{R}$ via

$$f(x) = a_n x^n + a_{n-1} x^{n-1} + ... + a_1 x + a_0$$

for some $a_n, a_{n-1}, ..., a_1, a_0 \in \mathbb{R}$.

Example 1.113 *Prove Proposition 1.105.*

Example 1.114 *Prove Proposition 1.97.*

Example 1.115 *Prove that there exist functions $f : (a, b) \to \mathbb{R}$ which are continuous, but not differentiable.*

Example 1.116 *Prove Proposition 1.106.*

Example 1.117 *Prove Proposition 1.107.*

Example 1.118 *Prove the product rule (Theorem 1.108).*

Example 1.119 *Prove the quotient rule (Theorem 1.109).*

1.6 Riemann integrals in $\mathbb{R}$

Definition 1.120 *Let $[a,b]$ be a closed interval. A partition of $[a,b]$ is a finite sequence $(t_0, t_1, ..., t_n)$ such that the following statements are true.*
(i) $t_0 = a$.
(ii) For each $i \in \{1, 2, ..., n\}$, $t_{i-1} < t_i$.
(iii) $t_n = b$.

Definition 1.121 *Let $[a,b]$ be a closed interval, and let $P = (t_0, t_1, ..., t_n)$ be a partition of $[a,b]$. Given a function $f : [a,b] \to \mathbb{R}$, the upper Riemann sum of f with respect to P is the value*

$$U(f;P) = \sum_{k=1}^{n} (t_k - t_{k-1}) \sup \left\{ f(t) \in \mathbb{R} \middle| t_{k-1} \le t \le t_k \right\}.$$

Definition 1.122 *Let $[a,b]$ be a closed interval, and let $P = (t_0, t_1, ..., t_n)$ be a partition of $[a,b]$. Given a function $f : [a,b] \to \mathbb{R}$, the lower Riemann sum of f with respect to P is the value*

$$L(f;P) = \sum_{k=1}^{n} (t_k - t_{k-1}) \inf \left\{ f(t) \in \mathbb{R} \middle| t_{k-1} \le t \le t_k \right\}.$$

Definition 1.123 *Let $[a,b]$ be a closed interval. Given a function $f : [a,b] \to \mathbb{R}$, the upper Riemann integral of f over $[a,b]$ is the value*

$$\overline{\int_a^b} f(t)\, \mathrm{d}t = \inf \left\{ U(f;P) \middle| P \text{ is a partition of } [a,b] \right\}.$$

Definition 1.124 *Let $[a,b]$ be a closed interval. Given a function $f : [a,b] \to \mathbb{R}$, the lower Riemann integral of f over $[a,b]$ is the value*

$$\underline{\int_a^b} f(t)\, \mathrm{d}t = \sup \left\{ L(f;P) \middle| P \text{ is a partition of } [a,b] \right\}.$$

Definition 1.125 *Let $f : [a,b] \to \mathbb{R}$. We say that f is Riemann integrable over*

$\underline{[a,b]}$ *provided that*

$$\overline{\int_a^b} f(t)\,\mathrm{d}t = \underline{\int_a^b} f(t)\,\mathrm{d}t.$$

In that case, we define

$$\int_a^b f(t)\,\mathrm{d}t = \overline{\int_a^b} f(t)\,\mathrm{d}t = \underline{\int_a^b} f(t)\,\mathrm{d}t,$$

called the Riemann integral of f over $[a,b]$.

Given a closed interval $[a,b]$, we often refer to the set of functions that are Riemann integrable over $[a,b]$ as $\mathcal{R}\,[a,b]$.

A large portion of the field of mathematical analysis was invented for the sole purpose of determining which functions are integrable. The following theorem shows that this class of functions at least includes continuous functions.

Theorem 1.126 *Let $f:[a,b]\to\mathbb{R}$. If f is continuous, then f is Riemann integrable.*

Proof Coming soon.

The following is the fundamental theorem of calculus.

Theorem 1.127 *Let $f:[a,b]\to\mathbb{R}$. If f is continuous, then*

$$\frac{\mathrm{d}}{\mathrm{d}t}\int_a^x f(t)\,\mathrm{d}t = f(x).$$

Proof Coming soon.

Corollary 1.128 *Let $f:[a,b]\to\mathbb{R}$ be continuous. There exists a differentiable function $F:[a,b]\to\mathbb{R}$ such that $\forall\, x\in(a,b)$, $F'(x)=f(x)$ and*

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

Proof Define a function $g : [a, b] \to \mathbb{R}$ via $g(x) = \int_a^x f(t)\, \mathrm{d}t$. By Theorem 1.127, $g'(x) = f(x)$. Moreover,

$$\int_a^b f(x)\, \mathrm{d}x = g(b) = g(b) - g(a). \tag{59}$$

□

2 Real Analysis, part II

2.1 From linear algebra to topology

2.1.1 Spaces

First, let's address the elephant in the room: what is a Euclidean space?

Definition 2.1 *Given $n \in \mathbb{Z}^+$, the $\underline{n\text{-dimensional Euclidean space}}$ is the real vector space $\mathbb{R}^n$.*

Why is this type of space so important? The answer is simple: they're easier to work with than most other types of infinite spaces. In this section, we'll introduce some of the properties that make this true.

First of all, $\mathbb{R}^n$ is a finite-dimensional real vector space, so any vector $v \in \mathbb{R}^n$ can be written using finitely many coordinates. Suppose that $\{u_1, u_2, ..., u_n\}$ is a basis for $\mathbb{R}^n$. This means that there exist some unique $r_1, r_2, ..., r_n \in \mathbb{R}$ such that the vector $v = \sum_{i=1}^{n} r_i u_i$. As a result, we can identify the vector v with the n-tuple $(r_1, r_2, ..., r_n)$. For the rest of this document, unless otherwise stated, we will always work with the "standard basis" for $\mathbb{R}^n$:

$$\begin{aligned} e_1 &= (1, 0, 0, 0..., 0, 0) \\ e_2 &= (0, 1, 0, 0..., 0, 0) \\ e_3 &= (0, 0, 1, 0..., 0, 0) \\ &\vdots \\ e_{n-1} &= (0, 0, 0, 0, ..., 1, 0) \\ e_n &= (0, 0, 0, 0, ..., 0, 1) \end{aligned} . \tag{60}$$

Now, any finite-dimensional real vector space is also an inner product space, and $\mathbb{R}^n$ is no exception. Let's review the definition of an inner product space.

Definition 2.2 *Let V be a vector space. Given a function $\langle , \rangle : V \times V \to \mathbb{R}$, we say that $\langle , \rangle$ is an $\underline{\text{inner product on } V}$ provided that the following statements are true.*
(i) $\forall\, v \in V$, $\langle v, v \rangle \geq 0$.
(ii) $\forall\, u, v \in V$, $\langle u, v \rangle = \langle v, u \rangle$.

(iii) $\forall\, v \in V$, $\langle v, v\rangle = 0$ *if and only if* $v = 0$.
(iv) $\forall\, u, v \in V$ *and* $\forall\, r \in \mathbb{R}$, $\langle ru, v\rangle = r\,\langle u, v\rangle$.
(v) $\forall\, u, v, w \in V$, $\langle u + v, w\rangle = \langle u, w\rangle + \langle v, w\rangle$.
In that case, $(V, \langle , \rangle)$ *is called an inner product space.*

We will now take advantage of our ability to identify any vector with its coordinates in the standard basis in order to define an inner product on $\mathbb{R}^n$.

Definition 2.3 *Let* $u, v \in \mathbb{R}^n$. *Suppose that* $u = \sum_{i=1}^n a_i e_i$ *and* $v = \sum_{i=1}^n b_i e_i$, *where* $(e_1, e_2, ..., e_n)$ *is the standard basis for* $\mathbb{R}^n$. *The Euclidean inner product or dot product of* u *and* v *is the value* $u \cdot v = \sum_{i=1}^n a_i b_i$.

Definition 2.4 *Let* $v \in \mathbb{R}^n$. *The length or magnitude of* v *is the value* $||v|| = \sqrt{v \cdot v}$.

As **Exercise 2.18** indicates, the dot product is an inner product on $\mathbb{R}^n$. We will now concern ourselves with something that will give a geometric structure to $\mathbb{R}^n$. For each pair of points $u, v \in \mathbb{R}^n$, let us define the "distance" from u to v as $||u - v||$, the magnitude of the difference between them.

We refer to the definition of metric space, Definition 1.25. We claim that the notion of distance that we have just defined makes $\mathbb{R}^n$ into a metric space. However, in order to prove it, we'll need a result from linear algebra. The following is the Cauchy-Schwarz inequality.

Theorem 2.5 *Let* V *be an inner product space. Given any vectors* $u, v \in V$, $|\langle u, v\rangle| \leq ||u||||v||$.

Proof Let $u, v \in V$. Given $\lambda \in \mathbb{R}$,

$$\langle u, u\rangle - 2\lambda\,\langle u, v\rangle + \lambda^2\,\langle v, v\rangle = \langle u - \lambda v, u - \lambda v\rangle = ||u - \lambda v||^2 \geq 0. \qquad (61)$$

This is a quadratic equation in λ. Thus, in order for $||u - \lambda v||^2 \geq 0$ to be true for all $\lambda \in \mathbb{R}$, we must have that the discriminant $(-2\,\langle u, v\rangle)^2 - 4\,\langle v, v\rangle\,\langle u, u\rangle \leq 0$. This is equivalent to saying that

$$\begin{aligned} 4\langle u, v\rangle^2 - 4\,\langle v, v\rangle\,\langle u, \rangle &\leq 0 \\ \langle u, v\rangle^2 &\leq \langle u, u\rangle\,\langle v, v\rangle \\ \big|\,\langle u, v\rangle\,\big| &\leq \sqrt{\langle u, u\rangle\,\langle v, v\rangle} = ||u||||v||. \end{aligned} \qquad (62)$$

□

In particular, for Euclidean space, $|u \cdot v| \leq ||u||||v||$. This will be important to the following theorem.

Theorem 2.6 *Define* $d : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ *via* $d(u, v) = ||u - v||$. *The pair* $(\mathbb{R}^n, d)$ *is a metric space.*

Proof We must show that each of the four requirements of metric spaces are satisfied. The proof of the first three axioms is **Exercise 2.19**.

Let $u, v, w \in \mathbb{R}^n$. By the Cauchy-Schwarz inequality (Theorem 2.5), we know that

$$\big|\,(u - v) \cdot (v - w)\,\big| \leq ||u - v||||v - w||. \tag{63}$$

We expand this into

$$(u \cdot v) - (u \cdot w) + (v \cdot w) - (v \cdot v) \leq ||u - v||||v - w||. \tag{64}$$

Equivalently,

$$-2\,(u \cdot w) \leq -2\,(u \cdot v) + 2\,(v \cdot v) - 2\,(v \cdot w) + 2||u - v||||v - w||. \tag{65}$$

To both sides of this equality, we add $u \cdot u$ and $w \cdot w$, obtaining

$$\begin{aligned}(u \cdot u) - (u \cdot w) - (w \cdot u) + (w \cdot w) &\leq (u \cdot u) - (u \cdot v) - (v \cdot u) + (v \cdot v) \\ &\quad + (v \cdot v) - (v \cdot w) - (w \cdot v) + (w \cdot w) + 2||u - v||||v - w||. \end{aligned} \tag{66}$$

We can rewrite this as

$$\begin{aligned}(u \cdot (u - w)) + (w \cdot (w - u)) &\leq (u \cdot (u - v)) + (v \cdot (v - u)) \\ &\quad + (v \cdot (v - w)) + (w \cdot (w - v)) + 2||u - v||||v - w||. \end{aligned} \tag{67}$$

In other words,

$$\begin{array}{c}(u-w)\cdot(u-w)\leq((u-v)\cdot(u-v))+((v-w)\cdot(v-w))+2||u-v||||v-w|| \\ ||u-w||^2\leq||u-v||^2+||v-w||^2+2||u-v||||v-w||\end{array}. \tag{68}$$

By taking the square root of both sides,

$$||u-w||\leq||u-v||+||v-w||, \tag{69}$$

which proves $d(u,w)\leq d(u,v)+d(v,w)$, the triangle inequality. □

Every metric space has a notion of open sets. In $\mathbb{R}^n$, these sets take on a particularly convenient form, due to the intuitive nature of the Euclidean metric. These open sets define a "topology" on Euclidean space. A topology is simply a rule for determining which sets are "open." Openness is a property that must satisfy three conditions, stated below.

Definition 2.7 *Let X be a set, and let $\mathcal{T}\subseteq\mathcal{P}(X)$, the power set of X. We say that $\mathcal{T}$ is a* <u>*topology on X*</u> *provided that the following statements are true.*
(i) $X,\varnothing\in\mathcal{T}$.
(ii) $\forall\, S\subseteq\mathcal{T}$, $\bigcup S\in\mathcal{T}$.
(iii) $\forall\, U,V\in\mathcal{T}$, $U\cap V\in\mathcal{T}$.
In that case, we say that $(X,\mathcal{T})$ is a <u>*topological space*</u>*. A subset $U\subseteq X$ such that $U\in\mathcal{T}$ is then said to be* <u>*open in X*</u>*.*

Definition 2.8 *Let X be a topological space. Given $F\subseteq X$, we say that* <u>*F is closed in X*</u> *provided that $X\setminus F$ is open in X.*

Topological spaces have a broad range of applications in mathematics, due to their versatility. Part of this versatility is due to the ideas of interior and closure.

Definition 2.9 *Let X be a topological space. Given $S\subseteq X$, the* <u>*interior of S in X*</u> *is the set $U\subseteq X$ that is open in X satisfying the following conditions.*
(i) $U\subseteq S$.
(ii) If $V\subseteq X$ is open in X and $V\subseteq S$, then $V\subseteq U$.

Definition 2.10 *Let X be a topological space. Given $S \subseteq X$, the* <u>*closure of S in X*</u> *is the set $F \subseteq X$ that is closed in X satisfying the following conditions.*
(i) $S \subseteq F$.
(ii) If $E \subseteq X$ is closed in X and $S \subseteq E$, then $F \subseteq E$.

We will denote the interior of S by $\operatorname{int} S$, and the closure of S by $\overline{S}$.

2.1.2 Maps between spaces

In discussing properties of types of sets, it is a gross oversight to neglect properties of functions between sets. We'll begin with the most basic, and one which is likely familiar to the reader: linear transformations.

Definition 2.11 *Let V_1 and V_2 be real vector spaces. Given $f : V_1 \to V_2$, we say that f is a linear transformation, or linear map, or linear function provided that $\forall\, u, v \in V_1$ and $\forall\, r, s \in \mathbb{R}$,*

$$f\left(ru + sv\right) = rf\left(u\right) + sf\left(v\right).$$

In other words, a linear transformation is a map that "preserves" addition and scalar multiplication.

We can say a bit more about linear transformations when the vector spaces in question are inner product spaces: linear transformations between inner product spaces are "bounded."

Proposition 2.12 *Let $f : \mathbb{R}^m \to \mathbb{R}^n$ be a linear transformation. There exists $k \in \mathbb{R}$ such that $\forall\, v \in \mathbb{R}^m$, $||f(v)|| \leq k||v||$.*

Proof Since f is a linear transformation, there exists a matrix representation M_f of f such that $\forall\, v \in \mathbb{R}^m$, $f(v) = M_f v$. Let $k = \max\left\{M_f(i,j)\right\}_{(i,j)=(1,1)}^{(m,n)}$. Now $\forall\, v \in \mathbb{R}^m$, $kv \geq M_f v = f(v)$. $\square$

There are many sorts of interesting functions between vector spaces: one for each way of morphing the distance between two points.

Definition 2.13 *Let (X, d_X) and (Y, d_Y) be metric spaces. Given $f : X \to Y$, we say that the map f is an isometry from X to Y provided that $\forall\, x_1, x_2 \in X$, $d_X\left(x_1, x_2\right) = d_Y\left(f\left(x_1\right), f\left(x_2\right)\right)$.*

Definition 2.14 *Let (X, d_X) and (Y, d_Y) be metric spaces. Given $f : X \to Y$, we say that the function f is a Lipschitz map or a contractive map provided that $\exists\, K \in [0, 1)$ such that $\forall\, x_1, x_2 \in X$, $d_Y\left(f\left(x_1\right), f\left(x_2\right)\right) \leq K d_X\left(x_1, x_2\right)$. In that case, we say that K is a Lipschitz constant of f.*

Definition 2.15 *Let* (X, d_X) *and* (Y, d_Y) *be metric spaces. Given* $f : X \to Y$*, we say that* f *is uniformly continuous provided that* $\forall\, \varepsilon \in \mathbb{R}^+$*,* $\exists\, \delta \in \mathbb{R}^+$ *such that* $\forall\, x_1, x_2 \in X$*, if* $d\,(x_1, x_2) < \delta$*, then* $d\,(f\,(x_1,)\,, f\,(x_2)) < \varepsilon$*.*

Finally, for topological spaces, the most general geometric construction possible, there is the continuous map.

Definition 2.16 *Let* X *and* Y *be topological spaces, and let* $f : X \to Y$*. Given* $a \in X$*, we say that* f *is continuous at* a *provided that* $\forall\, V \subseteq Y$*, if* V *is open in* Y *and* $f(a) \in V$*, then* $\exists\, U \subseteq X$ *that is open in* X *such that* $a \in U$ *and* $f(U) \subseteq V$*. If* $\forall\, a \in X$*,* f *is continuous at* a*, then we say that* f *is a continuous map.*

However, we will most commonly deal with continuous functions in metric spaces. For this reason, we would like to restate the definition in a more palatable form, as seen in Definition 1.59 and the following proposition.

Proposition 2.17 *Let* (X, d_X) *and* (Y, d_Y) *be metric spaces, and let* $f : X \to Y$*. Given* $a \in X$*, the following statements are equivalent.*
(i) f *is continuous at* a*.*
(ii) $\forall\, \varepsilon \in \mathbb{R}^+$*,* $\exists\, \delta \in \mathbb{R}^+$ *such that* $f\,(B\,(a, \delta)) \subseteq B\,(f(a), \varepsilon)$*.*
(iii) $\forall\, \varepsilon \in \mathbb{R}^+$*,* $\exists\, \delta \in \mathbb{R}^+$ *such that if* $d_X\,(x, a) < \delta$*, then* $d_Y\,(f\,(x)\,, f\,(a)) < \varepsilon$*.*
(iv) $\lim\limits_{x \to a} f(x) = f(a)$*.*

Proof This is **Exercise 2.23**.

To summarize: $\mathbb{R}^n$ is a finite-dimensional real vector space, which makes it an inner product space, which makes it a metric space, which makes it a topological space. This means that we can talk about linear transformations, isometries, Lipschitz maps, uniformly continuous maps, and continuous maps between Euclidean spaces. Even so, it's none of these properties that explain what is so interesting about $\mathbb{R}^n$. The fact is that there are plenty of vector spaces, metric spaces, topological spaces and so on that are not nearly as interesting as $\mathbb{R}^n$. What makes $\mathbb{R}^n$ so special can be summarized in a single word: calculus.

2.1.3 Exercises

Example 2.18 *Show that the dot product on $\mathbb{R}^n$ is an inner product.*

Example 2.19 *Show that $\mathbb{R}^n$, with distance defined via $d(u, v) = ||u - v||$, satisfies axioms (i), (ii) and (iii) of the definition of metric space.*

Example 2.20 *Let X and Y be metric spaces, and let $f : X \to Y$.*
(i) Let f be an isometry. Show that f is uniformly continuous.
(ii) Let f be a Lipschitz map. Show that f is uniformly continuous.

Example 2.21 *Let $f : \mathbb{R}^m \to \mathbb{R}^n$ be a linear transformation. Prove that f is uniformly continuous.*

Example 2.22 *Let X and Y be metric spaces. Show that if $f : X \to Y$ is uniformly continuous, then f is continuous.*

Example 2.23 *Prove Proposition 2.17.*

2.2 Differentiation

2.2.1 Defining differentiability

Euclidean spaces are interesting because they grant us the unique ability to deal with both algebraic properties (like addition and multiplication) and topological properties (like distance, or openness) at the same time. This means that it is possible to define things like limits of algebraic expressions. With this, we can define differentiability. However, the definition from introductory calculus will not be sufficient. Perhaps you thought that the derivative would be defined simply as the limit of the quotient of $f(x) - f(a)$ and $x - a$. However, there is an issue with that definition: it is only sensible in $\mathbb{R}$. If we want to generalize to $\mathbb{R}^n$, then we need something different.

However, there is still some interest in the limit of a difference quotient. We will give that a separate name: the directional derivative.

Definition 2.24 *Let $U \subseteq \mathbb{R}^m$ be open in $\mathbb{R}^m$, and let $f : U \to \mathbb{R}^n$. Given $a \in U$ and $v \in \mathbb{R}^m$, the directional derivative of f at a in the direction of v is the limit*

$$D_v f(a) = \lim_{t \to 0} \frac{f(a + tv) - f(a)}{t}.$$

If v is a standard basis vector of $\mathbb{R}^n$, we call $D_v f(a)$ a partial derivative of f at a, denoted $D_i f(a)$ or $\frac{\partial f}{\partial x_i}(a)$.

Why not just say that a function is differentiable if all of its directional derivatives exist? The answer is that general Euclidean spaces are more difficult to deal with than $\mathbb{R}$. Let's study an example of why Definition 2.26 is so necessary.

Example 2.25 *We define $f : \mathbb{R}^2 \to \mathbb{R}$ via the piecewise relationship*

$$f(x, y) = \begin{cases} \frac{x^2 y}{x^4 + y^2} & \text{if } (x, y) \neq (0, 0) \\ 0 & \text{if } (x, y) = (0, 0) \end{cases}. \tag{70}$$

Let $v \in \mathbb{R}^2$. We will show that $D_v f(0)$ exists. Suppose that $v = (r, s)$ for some

$r, s \in \mathbb{R}$. we note that

$$D_{(r,s)}f(0,0) = \lim_{t\to 0} \frac{f\left((0,0)+t(r,s)\right) - f\left(0,0\right)}{t} = \lim_{t\to 0} \frac{1}{t} \frac{\left(tr\right)^2 \left(ts\right)}{\left(tr\right)^4 + \left(ts\right)^2}$$
$$= \lim_{t\to 0} \frac{t^2 r^2 s}{t^4 r^4 + t^2 s^2} = \lim_{t\to 0} \frac{t^2 r^2 s}{t^2\left(t^2 r^4 + s^2\right)} = \lim_{t\to 0} \frac{r^2 s}{t^2 r^4 + s^2}. \quad (71)$$

This limit depends on whether $s = 0$ or $s \neq 0$:

$$D_{(r,s)}f(0,0) = \lim_{t\to 0} \frac{r^2 s}{t^2 r^4 + s^2} = \begin{cases} 0 & \text{if } s = 0 \\ \frac{r^2}{s} & \text{if } s \neq 0 \end{cases}. \quad (72)$$

Ergo, $D_v f(0)$ exists; every directional derivative of f exists at $(0,0)$.

We will show that f is not continuous at 0. Assume, with the expectation of a contradiction, that f is continuous at $(0,0)$. In that case, $\lim_{(x,y)\to(0,0)} f(x,y) = 0$. This indicates that $\forall\ \varepsilon \in \mathbb{R}^+$, $\exists\ \delta \in \mathbb{R}^+$ such that if $\left|\left|(x,y)-(0,0)\right|\right| < \delta$, then $|f(x,y) - f(0,0)| < \varepsilon$. In other words, if $\sqrt{x^2+y^2} < \delta$, then $|f(x,y)| < \varepsilon$.

Select $\varepsilon = \frac{1}{2}$. Consider $(x,y) \in \mathbb{R}^2$ via

$$(x,y) = \left(\frac{-1+\sqrt{1+\delta}}{2}, \frac{2+\delta-2\sqrt{1+\delta}}{4}\right). \quad (73)$$

We note two things: first, $y = x^2$, and second, $x^2 + x - \frac{1}{4}\delta = 0$. This means that

$$x^2 + y^2 = x^2 + x^4 \leq x^4 + x^2 + 2x^3 = \left(x^2 + x\right)^2 = \left(\frac{1}{4}\delta\right)^2 < \delta^2. \quad (74)$$

We deduce that $\sqrt{x^2+y^2} < \delta$. However,

$$f(x,y) = \frac{x^2 y}{x^4 + y^2} = \frac{x^2 x^2}{x^4 + x^4} = \frac{1}{2}. \quad (75)$$

Since $f(x,y) < \varepsilon$, this means that $\frac{1}{2} < \frac{1}{2}$, which is an obvious contradiction. This implies that our assumption that f is continuous is false; f is not continuous.

This example displays the nightmarish possibility that all directional derivatives could exist, even in the absence of continuity. This is undesirable; we would want our notion of differentiability to be a kind of smoothness that can only exist with continuity.

The true definition of differentiability in Euclidean space is as follows.

Definition 2.26 *Let $U \subseteq \mathbb{R}^m$ be open in $\mathbb{R}^m$, and let $f : U \to \mathbb{R}^n$. Given $a \in U$, we say that $\underline{f \text{ is differentiable at } a}$ provided that there exists a linear transformation $D : \mathbb{R}^m \to \mathbb{R}^n$ such that*

$$\lim_{x \to a} \frac{f(x) - f(a) - D(x-a)}{||x-a||} = 0.$$

In that case, we say that D is the $\underline{\text{derivative of } f \text{ at } a}$, denoted $Df(a)$. If $\forall\, a \in U$, $Df(a)$ exists, then we say that f is a $\underline{\text{differentiable function}}$.

With this in mind, let's begin the process of proving that this new definition of continuity has all the features we desire. First, *if* a function is differentiable, then all of its partial derivatives should exist.

Proposition 2.27 *Let $U \subseteq \mathbb{R}^m$ be open in $\mathbb{R}^m$, and let $f : U \to \mathbb{R}^n$. Given $a \in U$, if f is differentiable at a, then $\forall\, v \in \mathbb{R}^m$ such that $v \neq 0$, $D_v f(a) = Df(a)(v)$.*

Proof Let $\varepsilon \in \mathbb{R}^+$. We know that $\exists\, \delta \in \mathbb{R}^+$ such that if $||x-a|| < \delta$, then

$$\left|\left|\frac{f(x) - f(a) - Df(a)(x-a)}{x-a} - 0\right|\right| < \frac{\varepsilon}{||v||}. \tag{76}$$

Therefore, select $t \in \mathbb{R}^+$ such that $t < \frac{\delta}{||v||}$, and define $x = a + tv$. We note that

$$||x-a|| = ||tv|| = |t|||v|| = t||v|| < \frac{\delta}{||v||}||v|| = \delta, \tag{77}$$

which implies that

$$\left|\left|\frac{f(a+tv) - f(a) - Df(a)(tv)}{tv}\right|\right| < \frac{\varepsilon}{||v||}. \tag{78}$$

We multiply both sides of this inequality by $||v||$ to obtain

$$\left|\left|\frac{f(a+tv)-f(a)-Df(a)(tv)}{t}\right|\right| < \varepsilon. \tag{79}$$

Now, since $Df(a)$ is a linear transformation, we know that $Df(a)(tv) = tDf(a)(v)$. Ergo,

$$\left|\left|\frac{f(a+tv)-f(a)}{t} - Df(a)(v)\right|\right| < \varepsilon. \tag{80}$$

Yet this says exactly that $\lim\limits_{t\to 0} \frac{f(a+tv)-f(a)}{t} = Df(a)(v)$. $\square$

In particular, if $v = e_i$, where e_i is one of the standard basis vectors, then the directional derivative $D_{e_i}f = \frac{\partial f}{\partial x_i}$. Therefore, the matrix representation of the derivative of a differentiable function $f : U \to \mathbb{R}^m$, with components given by $f(x) = (f_1(x), f_2(x), ..., f_n(x))$, is the Jacobi matrix:

$$J(f) = \begin{pmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \cdots & \frac{\partial f_1}{\partial x_m} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} & \cdots & \frac{\partial f_2}{\partial x_m} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_n}{\partial x_1} & \frac{\partial f_n}{\partial x_2} & \cdots & \frac{\partial f_n}{\partial x_m} \end{pmatrix} \tag{81}$$

(with respect to the standard bases on $\mathbb{R}^m$ and $\mathbb{R}^n$). Our gains from the previous proposition are multiple, then: we know that differentiability implies existence of directional derivatives *and* we now have a way of computing the matrix representation of the derivative.

Now, is the function from Example 2.25 differentiable under this definition?

Example 2.28 *We define $f : \mathbb{R}^2 \to \mathbb{R}$ via the piecewise relationship*

$$f(x,y) = \begin{cases} \frac{x^2y}{x^4+y^2} & \textit{if } (x,y) \neq (0,0) \\ 0 & \textit{if } (x,y) = (0,0) \end{cases}. \tag{82}$$

We will show that f is not differentiable at $(0,0)$. Assume, with the expectation of a contradiction, that f is differentiable at $(0,0)$. From Example 2.25, we know *that*

$$D_v f(0) = \begin{cases} 0 & \text{if } s = 0 \\ \frac{r^2}{s} & \text{if } s \neq 0 \end{cases}, \tag{83}$$

where $v = (r, s)$*. By Proposition 2.27, we must have that* $\forall\ v \in \mathbb{R}^2$ *such that* $v \neq (0, 0)$*,* $D_v f(0, 0) = Df(0, 0)(v)$*. Take* $v = (2, 3) = (1, 1) + (1, 2)$*. We know that* $Df(0, 0)$ *is a linear transformation, so*

$$\begin{gathered} Df(0,0)(2,3) = Df(0,0)(1,1) + Df(0,0)(1,2) \\ D_{(2,3)}f(0,0) = D_{(1,1)}f(0,0) + D_{(1,2)}f(0,0) \\ \frac{2^2}{3} = \frac{1^2}{1} + \frac{1^2}{2} \\ \frac{4}{3} = \frac{3}{2} \end{gathered}, \tag{84}$$

which is an obvious contradiction. This leads us to conclude that our original assumption that f *is differentiable is false;* f *is not differentiable.*

Good, then; a function that is not continuous also happens to be not differentiable. Is this true in general? The next proposition shows that the answer is yes.

Proposition 2.29 *Let* $U \subseteq \mathbb{R}^m$ *be open in* $\mathbb{R}^m$*, and let* $f : U \to \mathbb{R}^n$*. Given* $a \in U$*, if* f *is differentiable at* a*, then* f *is continuous at* a*.*

Proof Let $\varepsilon \in \mathbb{R}^+$. By the definition of differentiability, $\exists\ \delta \in \mathbb{R}^+$ such that if $||x - a|| < \delta$, then

$$\left|\left|\frac{f(x) - f(a) - Df(a)(x - a)}{||x - a||} - 0\right|\right| < \varepsilon. \tag{85}$$

Additionally,

$$\begin{aligned} ||f(x) - f(a)|| &= \left|\left|\frac{f(x) - f(a) - Df(a)(x - a)}{||x - a||}||x - a|| + Df(a)(x - a)\right|\right| \\ &\leq \left|\left|\frac{f(x) - f(a) - Df(a)(x - a)}{||x - a||} - 0\right|\right| ||x - a|| + \big|\big|Df(a)(x - a)\big|\big|, \end{aligned} \tag{86}$$

due to the triangle inequality. Further, by Proposition 2.12, we know that $\exists\, k \in \mathbb{R}$ such that

$$||Df(a)(x-a)|| \leq k||x-a||. \tag{87}$$

Therefore, if $||x-a|| < \min\left\{\delta, \frac{\varepsilon}{\varepsilon+k}\right\}$, then

$$||f(x)-f(a)|| < \varepsilon||x-a|| + k||x-a|| \\ = (\varepsilon+k)\,||x-a|| < (\varepsilon+k)\,\frac{\varepsilon}{\varepsilon+k} = \varepsilon. \tag{88}$$

This shows that $\lim\limits_{x\to a} f(x) = f(a)$, and so f is continuous. □

This shows that our definition of differentiability is more selective than the simple criterion of existence of directional derivatives; Example 2.25 allowed the existence of partial derivatives without continuity of the function. At the same time, if a function *is* differentiable, then the directional derivatives *must* exist, by Proposition 2.27.

We ask the following question concerning the adequacy of our definition: is the composition of two differentiable functions also a differentiable function? We'll now establish a lemma leading to a theorem that shows that the answer is yes.

Lemma 2.30 *Let $U \subseteq \mathbb{R}^m$ be open in $\mathbb{R}^m$. Given $a \in U$, if $f : U \to \mathbb{R}^n$ is differentiable at a, then $\exists\, \delta \in \mathbb{R}^+$ and $m \in \mathbb{R}$ such that $\forall\, h \in B(0,\delta) \setminus \{0\}$,*

$$\frac{||f(a+h)-f(a)||}{||h||} \leq m.$$

Proof Consider $\varepsilon_1 = 1$. By definition of differentiability, $\exists\, \delta \in \mathbb{R}^+$ such that if $||h|| < \delta$, then

$$||f(a+h)-f(a)-Df(a)(h)|| < \varepsilon||h|| = ||h||. \tag{89}$$

This implies that

$$||f(a+h)-f(a)|| - ||Df(a)(h)|| < ||h||. \tag{90}$$

By Proposition 2.12, $\exists\, m \in \mathbb{R}$ such that $||Df(a)(h)|| \leq m||h||$. Therefore,

$$||f(a+h) - f(a)|| < m||h|| + ||h|| = (m+1)\,||h||. \tag{91}$$

Ergo, $\frac{||f(a+h)-f(a)||}{||h||} < m+1$, for any $h \in \mathbb{R}^m$ satisfying $0 < ||h|| < \delta$. $\square$

The following is known as the chain rule.

Theorem 2.31 *Let $U \subseteq \mathbb{R}^m$ be open in $\mathbb{R}^m$ and $V \subseteq \mathbb{R}^n$ open in $\mathbb{R}^n$. Define $f : U \to \mathbb{R}^n$ and $g : V \to \mathbb{R}^p$ such that $f(U) \subseteq V$. Given $a \in U$, if f is differentiable at a and g is differentiable at $f(a)$, then $g \circ f$ is differentiable at a and*

$$D\,(g \circ f)\,(a) = Dg\,(f(a)) \circ Df(a).$$

Proof Define $f(a) = b$. Let $\varepsilon \in \mathbb{R}^+$ such that $B\,(f(a), \varepsilon) \subseteq V$. By Proposition 2.29, we know that f is continuous at a, so $\exists\, \delta \in \mathbb{R}^+$ such that if $||h|| < \delta$, then $||f(a+h) - f(a)|| < \varepsilon$. We define $F : B\,(0, \delta) \to \mathbb{R}^n$ via

$$F\,(h) = \begin{cases} \frac{f(a+h)-f(a)-Df(a)(h)}{||h||} & \text{if } h \neq 0 \\ 0 & \text{if } h = 0 \end{cases}. \tag{92}$$

As f is differentiable at a, we know that $\lim\limits_{h \to 0} F(h) = 0$, so F is continuous at 0. Additionally, we notice that

$$f(a+h) - f(a) = Df(a)(h) + F(h)||h||. \tag{93}$$

We also define $G : B\,(0, \varepsilon) \to \mathbb{R}^p$ via

$$G(k) = \begin{cases} \frac{g(b+k)-g(b)-Dg(b)(k)}{||k||} & \text{if } k \neq 0 \\ 0 & \text{if } k = 0 \end{cases}. \tag{94}$$

Again, since g is differentiable at b, G is continuous at 0. We notice that, given $k \in B\,(0, \varepsilon) \setminus \{0\}$,

$$g(b+k) - g(b) = Dg(b)(k) + G(k)||k||. \tag{95}$$

At the same time, $||f(a+h)-f(a)|| < \varepsilon$, so this implies that if $0 < ||h|| < \delta$, then

$$\begin{aligned} &g\left(b+f(a+h)-f(a)\right)-g\left(b\right) \\ &\quad = Dg(b)\left(f(a+h)-f(a)\right)+G\left(f(a+h)-f(a)\right)||f(a+h)-f(a)|| \\ &\quad = Dg(b)\left(Df(a)(h)+F(h)||h||\right)+G\left(f(a+h)-f(a)\right)||f(a+h)-f(a)|| \end{aligned} \tag{96}$$

As $Dg(b)$ is a linear transformation, this can be written as

$$\begin{aligned} &g\left(f(a+h)\right)-g\left(f(a)\right) \\ &= Dg(b)\circ Df(a)(h)+||h||Dg(b)\left(F(h)\right)+G\left(f(a+h)-f(a)\right)||f(a+h)-f(a)|| \end{aligned} \tag{97}$$

Subtracting $Dg(b) \circ Df(a)(h)$ and dividing by $||h||$, this becomes

$$\begin{aligned} &\frac{g\left(f(a+h)\right)-g\left(f\left(a\right)\right)-Dg(b)\circ Df(a)(h)}{||h||} \\ &\qquad = Dg(b)\left(F\left(h\right)\right)+G\left(f(a+h)-f(a)\right)\frac{||f(a+h)-f(a)||}{||h||} \end{aligned} \tag{98}$$

However, by Lemma 2.30, $\exists\, m \in \mathbb{R}$ such that

$$\begin{aligned} &\lim_{h\to 0} Dg(b)\left(F\left(h\right)\right)+G\left(f(a+h)-f(a)\right)\frac{||f(a+h)-f(a)||}{||h||} \\ &\qquad \le Dg(b)\left(\lim_{h\to 0} F(h)\right)+mG\left(\lim_{h\to 0} f\left(a+h\right)-f\left(a\right)\right) \\ &\qquad = Dg(b)(0)+mG(0)=0+0=0. \end{aligned} \tag{99}$$

(We know that $Dg(b)(0) = 0$ because $Dg(b)$ is a linear transformation.) This shows that

$$\lim_{h\to 0}\frac{g\left(f(a+h)\right)-g\left(f\left(a\right)\right)-Dg(b)\circ Df(a)(h)}{||h||}=0, \tag{100}$$

and so $Dg(b) \circ Df(a)$ is the derivative of $g \circ f$. $\Box$

As a final check on the validity of our definition of the derivative, we ask the following: is this definition of derivative truly the rate of change of the function? The answer is yes for functions $f : \mathbb{R} \to \mathbb{R}$. For higher dimensions, the answer is not so simple. However, based on our intuition, we would expect that the derivative of a function that is locally constant should be the trivial linear transformation. In **Exercise 2.47**, we ask the reader to show that this is true when the function in question is globally constant. Let's now consider a more nuanced question.

Definition 2.32 *Let X be a metric space, and let $f : X \to \mathbb{R}$. Given $a \in X$, we say that $(a, f(a))$ is a* <u>*local minimum of f*</u> *provided that $\exists\, \varepsilon \in \mathbb{R}^+$ such that $\forall\, x \in B(a, \varepsilon)$, $f(x) \geq f(a)$.*

Definition 2.33 *Let X be a metric space, and let $f : X \to \mathbb{R}$. Given $a \in X$, we say that $(a, f(a))$ is a* <u>*local minimum of f*</u> *provided that $\exists\, \varepsilon \in \mathbb{R}^+$ such that $\forall\, x \in B(a, \varepsilon)$, $f(x) \leq f(a)$.*

In calculus, we learned that the derivative of a differentiable function at a local minimum or maximum should be 0. Now our derivative is a linear transformation; in this context, our "zero" is the trivial linear transformation. Should we expect that the derivative of a differentiable function at a local minimum or maximum is the trivial linear transformation? As the following proposition indicates, the answer is yes.

Proposition 2.34 *Let $U \subseteq \mathbb{R}^m$ be open in $\mathbb{R}^m$, and let $f : U \to \mathbb{R}$ be differentiable. Given $a \in U$, if $(a, f(a))$ is a local minimum of f, then $Df(a)$ is the trivial linear transformation.*

Proof Let $u \in \mathbb{R}^m$. Suppose $D_u f(a) = L$. Suppose, with the understanding that the other choice is similar, that $L \geq 0$. Let $\varepsilon \in \mathbb{R}^+$ such that $\forall\, x \in B(a, \varepsilon)$, $f(x) \geq f(a)$. By definition, $\exists\, \delta \in \mathbb{R}^+$ such that if $t \in (-\delta, \delta)$, then

$$\left| \frac{f(a + tu) - f(a)}{t} - L \right| < \varepsilon. \tag{101}$$

As $(a, f(a))$ is a local minimum, we can guarantee that $f(a + tu) - f(a) \geq 0$. Therefore, if $t < 0$, then $\frac{f(a+tu)-f(a)}{t} < 0$. In that case, since $|L| \geq 0$, by taking

$t = -\frac{\delta}{2}$,

$$\varepsilon > \left| \frac{f\left(a - \frac{\delta}{2}u\right) - f(a)}{-\frac{\delta}{2}} - L \right| \geq |L|. \tag{102}$$

Ergo, $|L| < \varepsilon$ for every $\varepsilon \in \mathbb{R}^+$. We deduce that $|L| = 0$, so $L = 0$.

By Proposition 2.27, we now have that $\forall\ u \in \mathbb{R}^m$, $Df(a)(u) = 0$. $\square$

2.2.2 C^k functions

We know that existence of partial derivatives is a *necessary* condition for differentiability, but can we find a *sufficient* condition? Perhaps one that would be easier to verify than going back to the definition? The following shows that while the mere *existence* of directional derivatives is not sufficient, their *continuity* is.

Theorem 2.35 *Let $U \subseteq \mathbb{R}^m$ be open in $\mathbb{R}^m$, and let $f : U \to \mathbb{R}^n$. If $\forall\, a \in U$, $\forall\, j \in \{1, 2, ..., m\}$, $D_j f(a)$ exists and is continuous, then f is differentiable.*

Proof We will use the fact that f is differentiable if and only if every component of f is differentiable. (The proof of this fact is **Exercise 2.49**.) With this fact in mind, we need only consider a real-valued function $f : U \to \mathbb{R}$. Let $a \in U$. We know that $\exists\, \varepsilon \in \mathbb{R}^+$ such that $B(a, \varepsilon) \subseteq U$. Let $h = (h_1, h_2, ..., h_m) \in \mathbb{R}^m$ such that $0 < ||h|| < \varepsilon$. We define

$$\begin{aligned} p_0 &= a \\ p_1 &= a + h_1 e_1 \\ p_2 &= a + h_1 e_1 + h_2 e_2 \\ &\vdots \\ p_m &= a + h_1 e_1 + h_2 e_2 + ... + h_m e_m \end{aligned} , \tag{103}$$

where $(e_1, e_2, ..., e_j)$ is the standard basis on $\mathbb{R}^m$. We note that $p_m = a + h$, so the set of points $\{p_0, p_1, ..., p_m\}$ spans a closed rectangular prism C of diagonal length $||h||$. As $||h|| < \varepsilon$, we see that $C \subseteq B(a, \varepsilon) \subseteq U$. Additionally,

$$f(a + h) - f(a) = \sum_{i=1}^{m} \left(f(p_i) - f(p_{i-1}) \right). \tag{104}$$

Now, given $i \in \{1, 2, ..., m\}$, define $\phi : [0, h_i] \to \mathbb{R}$ via $\phi(t) = f(p_{i-1} + te_i)$. We know that $p_{i-1} + te_i \in C$ for any $t \in [0, h_i]$; this line segment defines the edge of C from p_{i-1} to p_i. Ergo, $D_i f$ exists at each point in this line segment (since $C \subseteq U$), and so ϕ is differentiable on the interval $(0, h_i)$. Therefore, by the mean value theorem, $\exists\, c_i \in [0, h_i]$ such that

$$\phi(h_i) - \phi(0) = \phi'(c_i)\, h_i. \tag{105}$$

We define $q_i = p_{i-1} + c_i e_i$ in order to rephrase this as

$$f(p_i) - f(p_{i-1}) = D_i f(q_i)\, h_i. \tag{106}$$

With this, we can say that for some $q_1, q_2, ..., q_m \in C$,

$$f(a+h) - f(a) = \sum_{i=1}^{m} \left(f(p_i) - f(p_{i-1}) \right) = \sum_{i=1}^{m} D_i f(q_i)\, h_i. \tag{107}$$

Now, define B as the matrix

$$B = \begin{pmatrix} D_1 f(a) & D_2 f(a) & ... & D_m f(a) \end{pmatrix}. \tag{108}$$

In that case,

$$\begin{aligned} \lim_{h\to 0} \frac{f(a+h) - f(a) - Bh}{||h||} &= \lim_{h\to 0} \frac{\sum_{i=1}^{m} D_i f(q_i)\, h_i - \sum_{i=1}^{m} D_i f(a) h_i}{||h||} \\ = \lim_{h\to 0} \sum_{i=1}^{m} \frac{\left(D_i f(q_i) - D_i f(a)\right)(h_i)}{||h||} &\leq \lim_{h\to 0} \sum_{i=1}^{m} D_i f(q_i) - D_i f(a) \\ &= \sum_{i=1}^{m} \lim_{h\to 0} \left(D_i f(q_i) - D_i f(a) \right). \end{aligned} \tag{109}$$

As $D_i f$ is continuous at a, $\exists\, \delta \in \mathbb{R}^+$ such that if $||h|| < \delta$, then the distance $\big|\big|D_i f(a+h) - D_i f(a)\big|\big| < \frac{\varepsilon}{2}$. By the triangle inequality,

$$\begin{aligned} \big|\big|D_i f(q_i) - D_i f(a)\big|\big| &\leq \big|\big|D_i f(q_i) - D_i f(a+h)\big|\big| + \big|\big|D_i f(a+h) - D_i f(a)\big|\big| \\ &\leq \big|\big|D_i f(a) - D_i f(a+h)\big|\big| + \big|\big|D_i f(a+h) - D_i f(a)\big|\big| \\ &= 2\big|\big|D_i f(a+h) - D_i f(a)\big|\big| < 2\left(\frac{\varepsilon}{2}\right) = \varepsilon. \end{aligned} \tag{110}$$

We deduce that

$$\lim_{h \to 0} \frac{f(a+h) - f(a) - Bh}{||h||} = 0, \tag{111}$$

and so $B = Df(a)$. Thus, f is differentiable. $\square$

We should come up with a name for these functions. In fact, they fit into a broader context of functions that have many desirable properties. First, let's define what we mean by an higher-order partial derivative.

Definition 2.36 *Let $U \subseteq \mathbb{R}^m$ be open in $\mathbb{R}^m$, and let $f : U \to \mathbb{R}^n$. Given a number $i \in \{1, 2, ..., m\}$, a first order ith partial derivative of f is the function $D_i f : U \to \mathbb{R}^n$ such that $\forall\ a \in U$, $D_i f(a)$ is the ith partial derivative of f at a. Given $n \in \mathbb{Z}^+$, a $n+1$th order ith partial derivative of f is a function $D_i F : U \to \mathbb{R}^n$ such that $\forall\ a \in U$, $D_i F(a) = D_i D_{j_{n-1}} D_{j_{n-2}} ... D_{j_1} f(a)$, where $j_1, j_2, ..., j_{n-1} \in \{1, 2, ..., m\}$.*

This notation is positively ghastly, and unfortunately, there is no notation for a high-order partial derivative that is exactly wonderful to look at. Thankfully, we will not commonly be concerned with writing these partial derivatives explicitly.

On the other hand, defining nth order partial derivatives leads us to the following definition.

Definition 2.37 *Let $U \subseteq \mathbb{R}^m$ be open in $\mathbb{R}^m$, and let $f : U \to \mathbb{R}^n$. Given $k \in \mathbb{Z}^+$, we say that f is an element of the space of class C^k functions from U to $\mathbb{R}^n$, denoted $f \in C^k(U, \mathbb{R}^n)$, $\forall\ a \in U$, $\forall\ i \in \{1, 2, ..., k\}$, the ith order partial derivatives at a exist and are continuous.*

Definition 2.38 *Let $U \subseteq \mathbb{R}^m$ be open in $\mathbb{R}^m$, and let $f : U \to \mathbb{R}^n$. The space of class C^∞ functions or smooth functions from U to $\mathbb{R}^n$ is the intersection*

$$C^\infty(U, \mathbb{R}^n) = \bigcap_{k=1}^{\infty} C^k(U, \mathbb{R}^n). \tag{112}$$

Therefore, Theorem 2.35 applies to all C^1-class functions. (Some authors refer to these as "continuously differentiable functions," but the author of the present document considers this terminology rather unwieldy.) Note that for any $k \in \mathbb{Z}^+$, $C^{k+1}(U, \mathbb{R}^n) \subseteq C^k(U, \mathbb{R}^n)$, so this also applies to any functions of class C^2 or higher.

Are there other interesting properties of C^k-class functions? Yes, indeed. We'll now prove a crucial result about C^2 functions. We'll begin with a lemma that is a kind of multivariable analogue of the mean value theorem.

Lemma 2.39 *Let $U \subseteq \mathbb{R}^m$ be open in $\mathbb{R}^m$, and let $f \in C^2(U, \mathbb{R})$. Given indices $i, j \in \{1, 2, ..., m\}$, define a closed square*

$$Q_{(h,k)} = \left\{ a + se_i + te_j \in U \middle| 0 \leq s \leq h,\ 0 \leq t \leq k \right\},$$

where e_i and e_j are the ith and jth standard basis vectors of $\mathbb{R}^m$, respectively, and $h, k \in \mathbb{R}$ such that $a + he_i + ke_j \in U$. In that case, $\exists\ (s_0, t_0) \in [0, h] \times [0, k]$ such that

$$f(a+he_i+ke_j) - f(a + he_i) - f(a + ke_j) + f(a) = D_jD_if(a + s_0e_i + t_0e_j)\, kh.$$

Proof For each $h, k \in \mathbb{R}$ such that $a + he_i + ke_j \in U$, let

$$\lambda(h, k) = f(a) - f(a + he_i) - f(a + ke_j) + f(a + he_i + ke_j). \tag{113}$$

We define $\phi : [0, h] \to \mathbb{R}$ via $\phi(s) = f(a + se_i + ke_j) - f(a + se_i)$. We note that

$$\phi(h) - \phi(0) = f(a + he_i + ke_j) - f(a + he_i) - f(a + ke_j) + f(a) = \lambda(h, k). \tag{114}$$

We know that ϕ is differentiable, since D_if exists and is continuous on U. Therefore, the mean value theorem indicates that $\exists\ s_0 \in [0, h]$ such that

$$\lambda(h, k) = \phi(h) - \phi(0) = \phi'(s_0)\, h = (D_if(a + s_0e_i + ke_j) - D_if(a + s_0e_i))\, h. \tag{115}$$

Now, define $\varphi : [0, k] \to \mathbb{R}$ via $\varphi(t) = D_if(a + s_0e_i + te_j)$. We know that φ is differentiable, since D_jD_if exists an is continuous on U. Therefore, the mean value theorem indicates that $\exists\, t_0 \in [0, k]$ such that

$$D_i f\,(a + s_0 e_i + k e_j) - D_i f\,(a + s_0 e_i) = \varphi(k) - \varphi(0)$$
$$= \varphi'\,(t_0)\,k = D_j D_i f\,(a + s_0 e_i + t_0 e_j)\,k \quad (116)$$

We deduce that

$$\lambda(h, k) = D_j D_i f\,(a + s_0 e_i + t_0 e_j)\,kh. \quad (117)$$

□

The following is known as Clairaut's theorem.

Theorem 2.40 *Let $U \subseteq \mathbb{R}^m$ be open in $\mathbb{R}^m$, and let $f \in C^2\,(U, \mathbb{R})$. Given $a \in U$ and $i, j \in \{1, 2, ..., m\}$, $D_i D_j f(a) = D_j D_i f(a)$.*

Proof Let $t \in \mathbb{R}^+$ such that the closed square

$$Q_{(t,t)} = \left\{a + b e_i + c e_j \in \mathbb{R}^m \middle| 0 \leq b \leq t,\ 0 \leq c \leq t\right\} \subseteq U. \quad (118)$$

By Lemma 2.39, $\exists\ (b_0, c_0) \in [0, t]$ such that

$$D_j D_i f\,(a + b_0 e_i + c_0 e_j)\,t^2 = f(a + t e_i + t e_j) - f\,(a + t e_i) - f\,(a + t e_j) + f(a). \quad (119)$$

However, we also know that $\lim\limits_{t \to 0} a + b_0 e_i + c_0 e_j = a$. As $D_j D_i f$ is continuous, this implies that

$$\lim_{t \to 0} \frac{f\,(a + t e_i + t e_j) - f\,(a + t e_i) - f\,(a + t e_j) + f\,(a)}{t^2} = D_j D_i f\,(a)\,. \quad (120)$$

At the same time, Lemma 2.39 indicates that $\exists\ (u_0, v_0) \in [0, t]$ such that

$$D_i D_j f\,(a + u_0 e_i + v_0 e_j)\,t^2 = f(a + t e_i + t e_j) - f\,(a + t e_i) - f\,(a + t e_j) + f(a). \quad (121)$$

Again, $\lim_{t\to 0} a + u_0 e_i + v_0 e_j = a$, so since $D_i D_j f$ is continuous,

$$\lim_{t\to 0} \frac{f(a+te_i+te_j) - f(a+te_i) - f(a+te_j) + f(a)}{t^2} = D_i D_j f(a). \quad (122)$$

We deduce that $D_j D_i f(a) = D_i D_j f(a)$. □

Another interesting property of C^k functions deals with inverse functions. First, we define a particularly nice condition for C^k functions with inverse functions.

Definition 2.41 *Let $U \subseteq \mathbb{R}^m$ be open in $\mathbb{R}^m$, and let $V \subseteq \mathbb{R}^n$ be open in $\mathbb{R}^n$. Given a function $f : U \to V$, we say that f is a diffeomorphism [of class C^k] provided that $f \in C^k(U, V)$, f^{-1} exists, and $f^{-1} \in C^k(V, U)$.*

Naturally, we want to know when an invertible C^k function will be a diffeomorphism. After all, the inverse function of a bijective continuous function is not always continuous, so why should we expect that C^k functions have C^k inverses?

The skepticism is appropriate. **Exercise 2.50** indicates that, in order for a differentiable function to have an inverse function, the derivative must be an invertible linear transformation. Therefore, a C^k diffeomorphism must have a non-singular derivative. However, as the following proposition shows, that is the only other requirement.

Proposition 2.42 *Let $U \subseteq \mathbb{R}^m$ be open in $\mathbb{R}^m$, and let $f \in C^k(U, \mathbb{R}^n)$ be injective. Define $V = f(U)$ and $F : U \to V$ via $F(x) = f(x)$. If $\forall\, a \in U$, $Df(a)$ is non-singular, then F is a diffeomorphism of class C^k.*

Proof [Coming soon.] □

Now, given a function $f : X \to Y$, is it possible to find an appropriate nonempty $S \subseteq X$ so that the restriction function $f\big|_S$ is injective? In general set theory, the answer is trivially yes; given any function, just take a subset of the domain that contains only one point that maps to each value in the range. It would be more impressive if X were a topological space and we could find an open neighborhood

U around every point so that $f\big|_U$ were injective. The following theorem shows that this is, indeed, possible. The following is the inverse function theorem.

Theorem 2.43 *Let $U \subseteq \mathbb{R}^m$ be open in $\mathbb{R}^m$, and let $f \in C^k(U, \mathbb{R}^n)$. Define the set $W = f(U)$. Given $a \in U$, suppose that $Df(a)$ is a non-singular linear transformation. In that case, $\exists\, V \subseteq U$ that is open in U such that $a \in V$ and the function $F : V \to W$ via $F(x) = f(x)$ is a diffeomorphism of class C^r.*

Proof [Coming soon.] □

Definition 2.44 *Let $U \subseteq \mathbb{R}^m \times \mathbb{R}^n$ be open in $\mathbb{R}^m \times \mathbb{R}^n$, and let $f : U \to \mathbb{R}^k$. Given $(a, b) \in U$, define $U_b = \left\{x \in \mathbb{R}^m \middle| (x, b) \in U\right\}$ and $U_a = \left\{y \in \mathbb{R}^n \middle| (a, y) \in U\right\}$. Define the functions $h_1 : U_b \to \mathbb{R}^k$ via $h_1(x) = f(x, b)$ and $h_2 : U_a \to \mathbb{R}^k$ via $h_2(y) = f(a, y)$. The partial derivative of f with respect to $\mathbb{R}^m$ is the derivative $\frac{\partial f}{\partial(x_1, x_2, \ldots, x_m)}(a) = Dh_1(a)$. Similarly, the partial derivative of f with respect to $\mathbb{R}^n$ is the derivative $\frac{\partial f}{\partial(y_1, y_2, \ldots, y_n)}(b) = Dh_2(b)$.*

The following is the implicit function theorem.

Theorem 2.45 *Let $U \subseteq \mathbb{R}^m \times \mathbb{R}^n$ be open in $\mathbb{R}^m \times \mathbb{R}^n$, and let $F \in C^r(U, \mathbb{R}^n)$. Suppose that $F(a, b) = 0$ for some $(a, b) \in U$. If $\frac{\partial F}{\partial(y_1, y_2, \ldots, y_n)}$ is a non-singular linear transformation, then $\exists\, V \subseteq \mathbb{R}^m$ that is open in $\mathbb{R}^m$ such that $a \in V$ and $\exists\, g \in C^r(V, \mathbb{R}^n)$ such that $g(a) = b$ and $\forall\, x \in V$, $F(x, g(x)) = 0$.*

Proof [Coming soon.]

2.2.3 Exercises

Example 2.46 *Let $U \subseteq \mathbb{R}^m$ be open in $\mathbb{R}^m$, and let $f : U \to \mathbb{R}^n$. Given $a \in U$, suppose that $A, B : \mathbb{R}^m \to \mathbb{R}^n$ are derivatives of f at a. Prove that $A = B$.*

Example 2.47 *Let $U \subseteq \mathbb{R}^m$ be open in $\mathbb{R}^m$, and let $f : U \to \mathbb{R}^n$. Given $a \in U$, prove the following.*
(i) If $\exists\ v \in \mathbb{R}^n$ such that $\forall\ u \in U$, $f(u) = v$, then $Df(a)$ is the trivial linear transformation.
(ii) If f is a linear transformation, then $Df(a) = f$.

Example 2.48 *Let $U \subseteq \mathbb{R}^m$ be open in $\mathbb{R}^m$, and let $f, g : U \to \mathbb{R}^n$. Given $a \in U$, prove the following.*
(i) $D(f+g)(a) = Df(a) + Dg(a)$.
(ii) $D(fg)(a) = f(a)Dg(a) + g(a)Df(a)$.

Example 2.49 *Let $U \subseteq \mathbb{R}^m$ be open in $\mathbb{R}^m$, and let $f : U \to \mathbb{R}^n$. For each $i \in \{1, 2, ..., n\}$, define $\pi_i : \mathbb{R}^n \to \mathbb{R}$ via $\pi_i(x_1, x_2, ..., x_n) = x_i$. Show that f is differentiable if and only if $\forall\ i \in \{1, 2, ..., n\}$, $\pi_i \circ f$ is differentiable.*

Example 2.50 *Let $U \subseteq \mathbb{R}^m$ be open in $\mathbb{R}^m$ and $V \subseteq \mathbb{R}^n$ be open in $\mathbb{R}^n$. Show that if $f : U \to V$ is differentiable and invertible, then $\forall\ a \in U$, the derivative $Df(a)^{-1} = D(f^{-1})(f(a))$.*

2.3 Integration

2.3.1 Definition of Integrability

A fundamental question of mathematical analysis is: when does the integral exist? A large amount of the field of analysis was created for the sole purpose of answering this question and/or enlarging the class of sets and functions that have defined integrals. We will start with the most basic notions of integrability: existence of the Riemann integral. As in Definition 1.125, we will define the Riemann integral in higher-dimensional spaces as being the common value of an upper and lower Riemann integral.

Definition 2.51 *Given $S \subseteq \mathbb{R}^n$, we say that S is a rectangle in $\mathbb{R}^n$ provided that $\exists\, a_1, a_2, ..., a_n, b_1, b_2, ..., b_n \in \mathbb{R}$ such that*

$$S = [a_1, b_1] \times [a_2, b_2] \times ... \times [a_n, b_n] .$$

In that case, the volume of S in $\mathbb{R}^n$ is the value

$$v(S) = (b_1 - a_1)(b_2 - a_2) ... (b_n - a_n) .$$

Definition 2.52 *Let $Q = [a_1, b_1] \times [a_2, b_2] \times ... \times [a_n, b_n]$. A partition of Q is an ordered n-tuple $(P_1, P_2, ..., P_n)$ such that for each $i \in \{1, 2, ..., n\}$, P_i is a partition of $[a_i, b_i]$. (See Definition 1.120 for the definition of a partition of a closed interval.)*

Definition 2.53 *Let $Q = [a_1, b_1] \times [a_2, b_2] \times ... \times [a_n, b_n]$, and let $P = (P_1, P_2, ..., P_n)$ be a partition of Q. Suppose that for each $i \in \{1, 2, ..., n\}$, $P_i = (t_{i0}, t_{i1}, ..., t_{im_i})$. A subrectangle of P is any rectangle*

$$\left[t_{1(j_1-1)} - t_{1(j_1)}\right] \times \left[t_{2(j_2-1)} - t_{2(j_2)}\right] \times ... \times \left[t_{n(j_n-1)} - t_{n(j_n)}\right] ,$$

where for each $i \in \{1, 2, ..., n\}$, $j_i \in \{1, 2, ..., m_i\}$.

Definition 2.54 *Let $Q = [a_1, b_1] \times [a_2, b_2] \times ... \times [a_n, b_n]$, and let $f : Q \to \mathbb{R}$. Let P be a partition of Q, and suppose that $\{R_1, R_2, ..., R_m\}$ is the set of subrectangles* *of P. The upper Riemann sum of f with respect to P is the value*

$$U\left(f;P\right)=\sum_{i=1}^{m}v\left(R_i\right)\sup\left\{f(x)\middle|x\in R_i\right\}.$$

Definition 2.55 *Let $Q=[a_1,b_1]\times[a_2,b_2]\times\ldots\times[a_n,b_n]$, and let $f:Q\to\mathbb{R}$. Let P be a partition of Q, and suppose that $\{R_1,R_2,\ldots,R_m\}$ is the set of subrectangles of P. The lower Riemann sum of f with respect to P is the value*

$$L\left(f;P\right)=\sum_{i=1}^{m}v\left(R_i\right)\inf\left\{f(x)\middle|x\in R_i\right\}.$$

It's worth showing that these definitions make sense. Given $f:[a,b]\to\mathbb{R}$ and a partition P of $[a,b]$, we should show that $L\left(f;P\right)\leq U\left(f;P\right)$. First, we make a technical definition, for no other reason than to simplify the argument.

Definition 2.56 *Let $Q=[a_1,b_1]\times[a_2,b_2]\times\ldots\times[a_n,b_n]$ be a rectangle, and let $P_1=(P_{11},P_{12},\ldots,P_{1n})$ and $P_2=(P_{21},P_{22},\ldots,P_{2n})$ be partitions of Q. The common refinement of P_1 and P_2 is the partition $P=(P_1,P_2,\ldots,P_n)$ defined as follows: for each $i\in\{1,2,\ldots,n\}$, if $P_{1i}=(s_1,s_2,\ldots,s_k)$ and $P_{2i}=(t_1,t_2,\ldots,t_l)$, then let P_i be the partition of $[a_i,b_i]$ whose elements are $\{s_1,s_2,\ldots,s_k,t_1,t_2,\ldots,t_l\}$.*

Proposition 2.57 *Let $Q=[a_1,b_1]\times[a_2,b_2]\times\ldots\times[a_n,b_n]$, and let $f:Q\to\mathbb{R}$. Given partitions P_1 and P_2 of Q, $L\left(f;P_1\right)\leq U\left(f;P_2\right)$.*

Proof Suppose that $P_1=(P_{11},P_{12},\ldots,P_{1n})$ and $P_2=(P_{21},P_{22},\ldots,P_{2n})$. Let P be the common refinement of P_1 and P_2. One can show that $L\left(f;P_1\right)\leq L\left(f;P\right)$ and $U\left(f;P\right)\leq U\left(f;P_2\right)$. We know that for any subrectangle R_i of P,

$$\inf\left\{f(x)\middle|x\in R\right\}\leq\sup\left\{f(x)\middle|x\in R\right\}, \tag{123}$$

so we deduce that

$$L(f;P_1) \le L(f;P) = \sum_{i=1}^{m} v(R_i) \inf\left\{f(x) \middle| x \in R_i\right\}$$
$$\le \sum_{i=1}^{m} v(R_i) \sup\left\{f(x) \middle| x \in R_i\right\} = U(f;P) \le U(f;P_2). \quad (124)$$

□

Next, let's define the upper and lower Riemann integrals, in analogy with Definitions 1.123 and 1.124.

Definition 2.58 *Let Q be a rectangle in $\mathbb{R}^n$, and let $f : Q \to \mathbb{R}$ be bounded on Q. The upper Riemann integral of f over Q is the value*

$$\overline{\int_Q} f = \inf\left\{U(f;P) \,\middle|\, P \text{ is a partition of } Q\right\}.$$

Definition 2.59 *Let Q be a rectangle in $\mathbb{R}^n$, and let $f : Q \to \mathbb{R}$ be bounded on Q. The lower Riemann integral of f over Q is the value*

$$\underline{\int_Q} f = \sup\left\{L(f;P) \,\middle|\, P \text{ is a partition of } Q\right\}.$$

By Proposition 2.57, it is a simple matter to show that the lower Riemann integral is at most as large as the upper Riemann integral. When they are equal, they define the general Riemann integral.

Definition 2.60 *Let Q be a rectangle in $\mathbb{R}^n$, and let $f : Q \to \mathbb{R}$ be bounded on Q. We say that f is Riemann integrable over Q provided that*

$$\overline{\int_Q} f = \underline{\int_Q} f.$$

In that case, we say that $\underline{\int_Q f \text{ exists}}$, where we define

$$\int_Q f = \overline{\int_Q} f = \underline{\int_Q} f,$$

called the Riemann integral of f over $[a,b]$.

2.3.2 Criteria for integrability

When is the Riemann integral defined? As it is, the definition of integrability is rather abstract and difficult to verify, since it requires data about infinitely many partitions of a rectangle. The following theorem will allow us to distill this into a question of whether only one particular partition exists. The following is known as the Riemann condition.

Theorem 2.61 *Let Q be a rectangle in $\mathbb{R}^n$, and let $f : Q \to \mathbb{R}$ be bounded on Q. The function f is Riemann integrable over Q if and only if $\forall\, \varepsilon \in \mathbb{R}^+$, there exists a partition P of Q such that $U(f; P) - L(f; P) < \varepsilon$.*

Proof ($\Rightarrow$) Assume that $\int_Q f$ exists. Let $\varepsilon \in \mathbb{R}^+$. Since by definition, the lower Riemann integral is $\underline{\int_Q} f = \sup \left\{ L(f; P) \,\middle|\, P \text{ is a partition} \right\}$, there must exist a partition P_1 such that $\underline{\int_Q} f - L(f; P_1) < \frac{\varepsilon}{2}$. (If no such partition were to exist, then $\underline{\int_Q} f$ could not actually be the supremum of the lower Riemann sums.) Similarly, let P_2 be a partition of Q such that $U(f; P_2) - \overline{\int_Q} f < \frac{\varepsilon}{2}$. Now, let P be the common refinement of P_1 and P_2. We note that

$$U(f; P) - \frac{\varepsilon}{2} \leq U(f; P_2) - \frac{\varepsilon}{2} < \int_Q f < L(f; P_1) + \frac{\varepsilon}{2} \leq L(f; P) + \frac{\varepsilon}{2}, \quad (125)$$

which implies that $U(f; P) - L(f; P) < \varepsilon$.

($\Leftarrow$) Assume that for every $\varepsilon \in \mathbb{R}^+$, there exists a partition P of Q such that $U(f; P) - L(f; P) < \varepsilon$. We know that $\overline{\int_Q} f \leq U(f; P)$ and $L(f; P) \leq \underline{\int_Q} f$. Therefore, $\forall\, \varepsilon \in \mathbb{R}^+$,

$$\overline{\int_Q} f - \underline{\int_Q} f \leq U(f; P) - \underline{\int_Q} f \leq U(f; P) - L(f; P) < \varepsilon. \quad (126)$$

We deduce that $\overline{\int_Q} f - \underline{\int_Q} f \notin \mathbb{R}^+$, so $\overline{\int_Q} f - \underline{\int_Q} f \leq 0$. As $\overline{\int_Q} f \geq \underline{\int_Q} f$, we must have that $\overline{\int_Q} f - \underline{\int_Q} f = 0$, and so f is Riemann integrable. $\square$

There is yet another way to phrase integrability of a function over a rectangle, but first, we must introduce the notion of measure zero sets.

Definition 2.62 *Let $A \subseteq \mathbb{R}^n$. We say that* <u>*A has Lebesgue measure zero in $\mathbb{R}^n$*</u> *provided that $\forall\, \varepsilon \in \mathbb{R}^+$, there exists a countable collection $\{Q_k\}_{k\in\mathbb{Z}^+}$ of rectangles in $\mathbb{R}^n$ such that $A \subseteq \bigcup_{k=1}^{\infty} Q_k$ and $\sum_{k=1}^{\infty} v(Q_k) < \varepsilon$.*

First, we prove that there is a slightly different way to define measure zero.

Lemma 2.63 *Let $A \subseteq \mathbb{R}^n$. The set A has Lebesgue measure zero in $\mathbb{R}^n$ if and only if there exists a countable collection $\{Q_k\}_{k\in\mathbb{Z}^+}$ of rectangles in $\mathbb{R}^n$ such that $A \subseteq \bigcup_{k=1}^{\infty} \operatorname{int} Q_k$ and $\sum_{k=1}^{\infty} v(Q_k) < \varepsilon$.*

Proof $(\Rightarrow)$ Assume that A has Lebesgue measure zero in $\mathbb{R}^n$. In that case, there exists a collection $\{Q_k\}_{k\in\mathbb{Z}^+}$ of rectangles in $\mathbb{R}^n$ such that $A \subseteq \bigcup_{k=1}^{\infty} Q_k$ and $\sum_{k=1}^{\infty} v(Q_k) < \frac{\varepsilon}{2^n}$. Now, for each $k \in \mathbb{Z}^+$, suppose that

$$Q_k = [a_{k1}, b_{k1}] \times [a_{k2}, b_{k2}] \times ... \times [a_{kn}, b_{kn}]. \tag{127}$$

We define

$$R_k = \left[a_{k1} - \frac{b_{k1}-a_{k1}}{2}, b_{k1} + \frac{b_{k1}-a_{k1}}{2}\right] \times \left[a_{k2} - \frac{b_{k2}-a_{k2}}{2}, b_{k2} + \frac{b_{k2}-a_{k2}}{2}\right] \times ... \times \left[a_{kn} - \frac{b_{kn}-a_{kn}}{2}, b_{kn} + \frac{b_{kn}-a_{kn}}{2}\right] \tag{128}$$

We note that each component interval of R_k contains the corresponding interval in Q_k and has twice the length. Also, $Q_k \subseteq \operatorname{int} R_k$, so $A \subseteq \bigcup_{k=1}^{\infty} \operatorname{int} R_k$. Moreover,

$$v(R_k) = \prod_{k=1}^{n} 2(b_{ki} - a_{ki}) = 2^n \prod_{k=1}^{n} (b_{ki} - a_{ki}) = 2^n v(Q_k). \tag{129}$$

As a result,

$$\sum_{k=1}^{\infty} v(R_k) = \sum_{k=1}^{\infty} 2^n v(Q_k) = 2^n \sum_{k=1}^{\infty} v(Q_k) < 2^n \frac{\varepsilon}{2^n} = \varepsilon. \tag{130}$$

($\Leftarrow$) Since for each $k \in \mathbb{Z}^+$, $\operatorname{int} Q_k \subseteq Q_k$, we must have that

$$\bigcup_{k=1}^{\infty} \operatorname{int} Q_k \subseteq \bigcup_{k=1}^{\infty} Q_k. \tag{131}$$

Therefore, if $A \subseteq \bigcup_{k=1}^{\infty} \operatorname{int} Q_k$ and $\sum_{k=1}^{\infty} v(Q_k) < \varepsilon$ for any $\varepsilon \in \mathbb{R}^+$, then A must have measure zero in $\mathbb{R}^n$. $\square$

With this, we can define a new condition that is equivalent to Riemann integrability.

Theorem 2.64 *Let $Q \subseteq \mathbb{R}^n$ be a rectangle in $\mathbb{R}^n$, and let $f : Q \to \mathbb{R}$ be a bounded function. Define $S = \{x \in Q \,|\, f \text{ is continuous at } x\}$. The function f is Riemann integrable over Q if and only if $Q \setminus S$ has measure zero in $\mathbb{R}^n$.*

Proof Define $D = Q \setminus S$. Additionally, since f is bounded, define $M \in \mathbb{R}$ such that $\forall\, x \in Q$, $|f(x)| \leq M$.

($\Rightarrow$) Coming soon.

($\Leftarrow$) Assume that D has measure zero in $\mathbb{R}^n$. Let $\varepsilon \in \mathbb{R}^+$. According to Lemma 2.63, there exists a collection $\{Q_m\}_{m \in \mathbb{Z}^+}$ such that $D \subseteq \bigcup_{m=1}^{\infty} \operatorname{int} Q_m$ and $\sum_{m=1}^{\infty} v(Q_m) < \frac{\varepsilon}{2M+2v(Q)}$.

For each $a \in Q \setminus D$, we know that f is continuous at a, by assumption. Therefore, select $\delta_a \in \mathbb{R}^+$ such that if $d(x,a) < \delta_a$, then $d(f(x), f(a)) < \frac{\varepsilon}{2M+2v(Q)}$. Now, let $Q_a \subseteq B(a, \delta_a)$ be a rectangle in $\mathbb{R}^n$ such that $a \in \operatorname{int} Q_a$. (See **Exercise 2.67**.) This assures that, given any $a \in Q \setminus D$, $\forall\, x \in Q_a \cap Q$, $d(x,a) < \delta_a$, and so $d(f(x), f(a)) < \frac{\varepsilon}{2M+2v(Q)}$.

We note that

$$Q = D \cup (Q \setminus D) \subseteq \left(\bigcup_{m=1}^{\infty} \operatorname{int} Q_m \right) \cup \left(\bigcup_{a \in Q \setminus D} \operatorname{int} Q_a \right). \tag{132}$$

However, the Heine-Borel theorem (Theorem 1.85) indicates that Q must be compact, so $\exists\, m_1, m_2, ..., m_r \in \mathbb{Z}^+$ and $a_1, a_2, ..., a_s \in Q \setminus D$ such that

$$Q \subseteq \left(\bigcup_{i=1}^{r} \operatorname{int} Q_{m_i} \right) \cup \left(\bigcup_{j=1}^{s} \operatorname{int} Q_{a_j} \right). \tag{133}$$

Further, we notice that

$$\sum_{i=1}^{r} v\left(Q_{m_i}\right) \leq \sum_{m=1}^{\infty} v\left(Q_m\right) < \frac{\varepsilon}{2M + 2v(Q)}. \tag{134}$$

Additionally, for each $j \in \{1, 2, ..., s\}$, the triangle inequality implies that for any $x, y \in Q_{a_j} \cap Q$,

$$\begin{aligned} d\left(f\left(x\right), f\left(y\right)\right) &\leq d\left(f(x), f\left(a_j\right)\right) + d\left(f\left(a_j\right), f\left(y\right)\right) \\ &< \frac{\varepsilon}{2M + 2v(Q)} + \frac{\varepsilon}{2M + 2v(Q)} = \frac{\varepsilon}{M + v(Q)}. \end{aligned} \tag{135}$$

For each $i \in \{1, 2, ..., r\}$, define $Q_i = Q_{m_i} \cap Q$ and for each $j \in \{1, 2, ..., s\}$, define $Q_j' = Q_{a_j} \cap Q$. We seek to produce a certain partition P of Q such that each Q_i and Q_j' is a union of subrectangles of P. For each $i \in \{1, 2, ..., r\}$ and $j \in \{1, 2, ..., s\}$, suppose that

$$\begin{aligned} Q_i &= [a_{i1}, b_{i1}] \times [a_{i2}, b_{i2}] \times ... \times [a_{in}, b_{in}] \\ Q_j' &= \left[a_{j1}', b_{j1}'\right] \times \left[a_{j2}', b_{j2}'\right] \times ... \times \left[a_{jn}', b_{jn}'\right] \end{aligned}. \tag{136}$$

For each $t \in \{1, 2, ..., n\}$, define the partition P_t whose elements are the elements of the following set:

$$S_t = \left(\bigcup_{i=1}^{r} \{a_{it}, b_{it}\}\right) \cup \left(\bigcup_{j=1}^{s} \left\{a_{jt}', b_{jt}'\right\}\right). \tag{137}$$

We define the partition P of Q by $P = (P_1, P_2, ..., P_t)$, and we define $\mathcal{R}$ as the set of subrectangles of P. We are guaranteed, by the construction of P, that each Q_i and Q_j' will be a union of elements of $\mathcal{R}$. On the other hand, each element of $\mathcal{R}$ is contained in one of the Q_i or Q_j', since the collection $\{Q_1, Q_2, ..., Q_r, Q_1', Q_2', ..., Q_s'\}$ covers Q. Therefore, define the subsets

$$\begin{aligned} R &= \left\{S \in \mathcal{R} \middle| S \subseteq Q_i \text{ for some } i \in \{1, 2, ..., r\}\right\} \\ R' &= \left\{S \in \mathcal{R} \middle| S \subseteq Q_j' \text{ for some } j \in \{1, 2, ..., s\}\right\} \end{aligned}. \tag{138}$$

Now, we note that for each $S \in R$, $\left(\sup_{x\in S} f(x) - \inf_{x\in S} f(x)\right) \leq 2M$. Thus,

$$\sum_{S\in R} \left(\sup_{x\in S} f(x) - \inf_{x\in S} f(x)\right) v(S) \leq \sum_{S\in R} 2Mv(S) = 2M \sum_{S\in R} v(S). \tag{139}$$

At the same time, for each $S \in R'$, $\exists\, j \in \{1, 2, ..., s\}$ such that $S \subseteq Q_j = Q_{a_j} \cap Q$, so we know that $\left(\sup_{x\in S} f(x) - \inf_{x\in S} f(x)\right) < \frac{\varepsilon}{M+v(Q)}$. This implies that

$$\sum_{S\in R'} \left(\sup_{x\in S} f(x) - \inf_{x\in S} f(x)\right) v(S) \leq \frac{\varepsilon}{M + v(Q)} \sum_{S\in R'} v(S). \tag{140}$$

Meanwhile,

$$\sum_{S\in R} v(S) \leq \sum_{i=1}^{r} \sum_{S\subseteq Q_i} v(S) = \sum_{i=1}^{r} v\left(Q_i\right) \leq \sum_{m=1}^{\infty} v\left(Q_m\right) < \frac{\varepsilon}{2M + 2v(Q)}, \tag{141}$$

and

$$\sum_{S\in R'} v(S) \leq \sum_{S\subseteq Q} v(S) = v(Q). \tag{142}$$

Thus,

$$\begin{aligned} U\left(f; P\right) - L\left(f; P\right) &= \sum_{S\in \mathcal{R}} \left(\sup_{x\in S} f(x) - \inf_{x\in S} f(x)\right) \\ &= \sum_{S\in R} \left(\sup_{x\in S} f(x) - \inf_{x\in S} f(x)\right) v(S) + \sum_{S\in R'} \left(\sup_{x\in S} f(x) - \inf_{x\in S} f(x)\right) v(S) \\ &\leq 2M \sum_{S\in R} v(S) + \frac{\varepsilon}{M + v(Q)} \sum_{S\in R'} v(S) \\ &< 2M \frac{\varepsilon}{2M + 2v(Q)} + \frac{\varepsilon}{M + v(Q)} v(Q) = \varepsilon. \end{aligned} \tag{143}$$

By Riemann's condition (Theorem 2.61), this indicates that f is integrable over Q.
□

2.3.3 Exercises

Example 2.65 *Let Q be a rectangle in $\mathbb{R}^n$, and let $f : Q \to \mathbb{R}$ be a constant function. Show that f is Riemann integrable over Q.*

Example 2.66 *Let $\{A_n\}_{n\in\mathbb{Z}^+}$ be a collection of subsets of $\mathbb{R}^n$. Show that if for each $n \in \mathbb{Z}^+$, A_n has measure zero in $\mathbb{R}^n$, then $\bigcup_{n=1}^{\infty} A_n$ has measure zero in $\mathbb{R}^n$.*

Example 2.67 *Let $\varepsilon \in \mathbb{R}^+$, and let $a \in \mathbb{R}^n$. Show there exists a rectangle $Q \subseteq \mathbb{R}^n$ such that $Q \subseteq B(a, \varepsilon)$ and $a \in \operatorname{int} Q$.*