LECTURE NOTES MATH 4423 - INTRODUCTION TO REAL ANALYSIS I & II

EMANUELE ZAPPALA

1. Introduction

This is an introductory course in Mathematical Analysis. Roughly speaking, Analysis is the study of functions and spaces of functions along with their properties. The main objects of study are continuous and smooth functions, limits, sequences, convergence and metric spaces (i.e. spaces where there is a notion of distance between objects). Our main reference will be [4], which is colloquially known in the mathematical community as "Baby Rudin". Other excellent references include [1–3]. Any book of Calculus – e.g. [5] – can be considered as a prerequisite to this course. We will focus on analysis where functions and spaces are over the real numbers, as opposed to complex numbers, hence the name "Real Analysis".

2. Real Numbers

As the main topic of the course relies on the real numbers, it is important that they are properly introduced. The approach we will follow is to first define the natural numbers $\mathbb{N}$ using an axiomatic procedure based on the *Peano axioms.* From the natural numbers we will construct the integers $\mathbb{Z}$, and next the rational numbers $\mathbb{Q}$. We will see that the rationals have "holes" and we will construct the real numbers $\mathbb{R}$ to "fill these holes". We will not prove several properties during such constructions for the sake of simplicity.

The natural numbers are the numbers $0, 1, 2, \dots$ and Peano's axioms are used to define their arithmetic.

Axioms 2.1 (Peano Axioms)**.** *We consider the symbol* 0*, and a function* S *whose properties we require with the following axioms.*

1. 0 *is a natural number.*
2. *For any natural number* n*, we have* $n = n$*.*
3. *For any pair of natural numbers* n *and* m*,* $n = m$ *implies* $m = n$*.*
4. *For all natural numbers* n, m, l*, we have that* $n = m$ *and* $m = l$ *implies* $n = l$*.*
5. *For any natural number* n*, if* $n = y$ *for some* y*, then* y *is a natural number.*

The function S *is called the "successor function", and it gives the "next" natural number of its input. So, for instance,* $S(0) = 1$*,* $S(1) = 2$ *and so on.*

6. *For any natural number* n*,* $S(n)$ *is a natural number as well.*
7. *If* $S(n) = S(m)$*, then* $n = m$*.*
8. *There is no natural number* n *such that* $S(n) = 0$*.*
9. *If* K *is a set such that* 0 *is in* K*, and* n *being in* K *implies that* $S(n)$ *is in* K *as well, then* K *is the set of natural numbers* $\mathbb{N}$*.*

The last axiom, number 9, is particularly used in mathematics and it is called Peano's Induction, or simply the Induction Axiom.

Using these axioms it is possible to define the standard operations on the natural numbers $\mathbb{N}$, such as sum and multiplication. They are, of course, what we already know since our young age. Addition is defined recursively by means of the two defining properties

- $n + 0 = n$,
- $n + S(m) = S(n + m)$.

Then, we can show, for instance, that $n + 1 = S(n)$ as follows:

$$\begin{aligned} n + 1 &= n + S(0) \\ &= S(n + 0) \\ &= S(n). \end{aligned}$$

Similarly, one can show that $n + 2 = S(S(2))$, $n + 3 = S(S(S(n)))$ and so on. This allows us to define addition for all natural numbers (using induction!).

Multiplication is defined similarly from the two properties

- $n \cdot 0 = 0$,
- $n \cdot S(m) = n \cdot m + n$.

All the properties of multiplication are derived from the two equations just listed.

Exercise 2.2. Show that $S(0)$ acts as the unity. In other words, show that $n \cdot S(0) = n$ for all n. In fact, observe that $S(0)$ is precisely what we called 1, after all...

We have therefore defined the set of natural numbers $\mathbb{N}$ and endowed it with a notion of sum and multiplication. It should be clear, at this point, that we do not have anything such as subtraction, or division. This is in fact obvious from the fact that we have only "positive integers" so far, so if we take a difference such as $5 - 7$ we obtain a negative number, which does not make sense since we have not invented them yet. Similarly, we incur in troubles with divisions since we do not have anything like fractions as we did not invent them yet either.

It is clear that we need to invent something that allows us to subtract and divide. Let us do one step at the time, and invent the numbers that allow us to subtract first. These are called the integers, and indicated as $\mathbb{Z}$. In everyday language, you represent them by putting a positive or a negative sign in front of natural numbers. However, we like to do simple things in complicated ways, and we therefore follow another route.

Let us define the set of pairs (n, m) where n and m are natural numbers. The intuitive meaning of a pair (n, m) is the value $n - m$, as we shall see, but we did not invent subtraction so we cannot do it just yet.

We declare two pairs (n, m) and (n', m') to be the same if it happens that $n + m' = n' + m$. Before ruling that this is nonsense, we invite the reader to think of the previous analogy of (n, m) and subtraction. This defines an equivalence relation between pairs, which decomposes the set of pairs in equivalence classes. The class of pairs that are the same, according to the equality $n + m' = n' + m$ is indicated with the extra brackets $[(n, m)]$.

We now introduce the operation of addition and multiplication between pairs. The addition is defined simply as

$$[(n, m)] + [(n', m')] = [n + n', m + m'].$$

Multiplication is defined as

$$[(n, m)] \cdot [(n', m')] = [(nn' + mm', nm' + n'm)].$$

Observe that we are only using addition and multiplication in the natural numbers (in each pair!) to define these operations. In this setting, we can easily define the negative of an element by switching the order of the elements in a pair:

$$-[(n, m)] = [(m, n)].$$

Now we can simply define subtraction as addition of the negative element:

$$[(n, m)] - [(n', m')] = [(n, m)] + [(m', n')],$$

where the sum is the one defined above.

The set of classes $[(a, b)]$ is denoted by $\mathbb{Z}$ and it is called the set of integer numbers, or just the integers. Here we identify the natural number n with $[(n, 0)]$, and the negative of n with $[(0, n)]$. Now that we have defined integers and subtraction formally, we will use the standard symbols $-n$ and $n - m$.

We now turn to our other problem: division. We will play the same game and define the rational numbers $\mathbb{Q}$ using pairs of integers. We now consider the set of pairs of integers (a, b) where $b \neq 0$. We define an equivalence relation between these pairs by saying that (a, b) and (c, d) are equivalent if $ad = cd$. Once again we obtain classes of elements which we denote as $[(a, b)]$. The intuitive meaning of the pairs (a, b) is the familiar notion of fraction $\frac{a}{b}$. The equivalence relation simply states the common fact that $\frac{qa}{qb} = \frac{a}{b}$, because we can factor out q.

We define the addition of two classes as

$$[(a, b)] + [(c, d)] = [(ad + bc, bd)],$$

while we define multiplication as

$$[(a, b)] \cdot [(c, d)] = [(ac, bd)].$$

Obsereve that thinking of $[(a, b)]$ as the usual fraction $\frac{a}{b}$ one readily sees that the two operations above simply coincide with the usual sum of fractions (using the common denominator), and usual product of fractions! We now have a natural notion of division, simply saying that the inverse of an element is given by swapping the order of the elements in a pair, and saying that dividing amounts to multiplying by the inverse:

$$\frac{[(a, b)]}{[(c, d)]} = [(a, b)] \cdot [(d, c)].$$

The set of numbers just constructed is the set of rational numbers, indicated by the symbol $\mathbb{Q}$. This is just the set of fractions we are all familiar with.

We now will see why the set of rational numbers is not completely satisfactory, in a sense.

Example 2.3. Consider the equation

$$x^2 = 2.$$

We want to show that this has no solutions in $\mathbb{Q}$. Suppose that we have a rational solution $x = \frac{n}{m}$, which we can assume in such a way that n and m are coprime, since if they have a common factor we could simplify it. Being a solution of the equation, we get that

$$\frac{n^2}{m^2} = 2,$$

which implies

$$n^2 = 2m^2.$$

This means that n^2 is even. But if n^2 is even, i.e. if 2 divides n^2, then it must also divide n. So, n is even. Therefore, n^2 is actually divisible by 4, not only by 2. This implies that $2m^2$ is divisible by 4 as well, which means that m^2 is divisible by 2. Therefore, n and m are not coprime, being both divisible by 2. This is in contradiction with our initial choice of n and m, and this contradiction shows that a solution to $x^2 = 2$ cannot be in $\mathbb{Q}$.

Of course, we know that a solution to $x^2 = 2$ would be the square root of 2, $\sqrt{2}$, and all the roots of the equation are $\pm\sqrt{2}$. The previous example shows that we need more than just rational numbers. We need real numbers.

Example 2.4. Let A be the set of positive rational numbers p such that $p^2 < 2$. Let B be the set of positive rational numbers p such that $p^2 > 2$. We can show that A does not contain any largest number, and B does not contain any smallest number.

In other words, it is possible to show that for any $p \in A$ there exists a $q \in A$ with $q > p$, and similarly for any $p \in B$ we can find a $q \in B$ with $q < p$. Given a p, we consider the fraction

$$q = \frac{2p+2}{p+2}, \tag{1}$$

from which we get

$$q^2 - 2 = p - \frac{p^2-2}{p+2} = \frac{2(p^2-2)}{(p+2)^2}. \tag{2}$$

If p is in A, then $p^2 - 2 < 0$ and Equation (1) shows that $q > p$, while Equation (2) shows that $q^2 < 2$, and therefore $q \in A$ as well. Similar discussion can be done for B.

Example 2.4 shows that there are "holes" in the rational numbers. The set of real numbers substantially fills these holes.

Before starting with the construction of real numbers, we will take a small digression on *ordered sets* and *ordered fields*.

Definition 2.5. Let S be a set. An order on S, denoted by the symbol $<$, is a relation satisfying the following properties:

- Given $x, y \in S$ only one of the following three cases happen, $x < y$, $x = y$ or $x > y$.
- The transitive property holds: $if x < y$ and $y < z$, then $x < z$.

As usual, we can say that $x < y$ or $x = y$ in a compact way as $x \leq y$.

Definition 2.6. An ordered set is a set S along with an order defined on it.

Definition 2.7. Let S be an ordered set, and let $E \subset S$ be a subset. We say that E is bounded above if there exists $\beta \in S$ such that $x \leq \beta$ for all $x \in E$. We say that β is an upper bound for E. Similar definitions hold for bounded below, and lower bound. The reader is invited to adapt the definition to that case.

Definition 2.8. Let S be an ordered set, and let E be a subset which is bounded above. An element α is said to be a *supremum*, or *least upper bound*, if it satisfies the following properties:

- It is an upper bound.
- If $\gamma < \alpha$, then γ is not an upper bound for E.

The supremum of E is denoted by the symbol $\sup E$. Similar definitions hold for *infimum*, or *greatest lower bound*. Once again, the reader is invited to adapt the definition to that case.

Definition 2.9. We say that an ordered set S has the least-upper-bound property if the following holds true. For any subset $E \subset S$ which is not empty and bounded above, then $\sup E$ exists.

Theorem 2.10. *Let S be an ordered set with the least-upper-bound property, and let B be a nonempty set bounded below. Let L be the set of all lower bounds of B. Then $\alpha = \sup L$ exists, and $\alpha = \inf B$.*

Proof. First, we want to show that $\alpha = \sup L$ exists. Since S has the least-upper-bound property, this would follow by showing that L is not empty, and that L is bounded above. To show that L is not empty, we need to show that B has at least a lower bound. However, since B is bounded below by assumption, this follows immediately. To show that L is bounded above, we need to show that there exists at least an element x in S such that $x \geq y$ for all $y \in L$. Since L is a set of lower bounds of B, it follows that if $x \in B$ and $y \in L$, we have $y \leq x$. Also, since B is not the empty set by assumption, such an x exists. This completes the first part of the proof.

Next, we show that $\alpha = \inf B$. This means, by definition of infimum, that α is a lower bound of B with the property that whenever we pick a $\beta > \alpha$ then β cannot be a lower bound of B. So, to complete the proof, we need to show the two facts that α is a lower bound of B, and that no β larger than α is a lower bound. For the first fact we proceed as follows. First, let $\gamma < \alpha$. Since α is the supremum of L, it follows (from the definition of supremum) that γ is not an upper bound for L. This latter fact in turn implies that γ cannot be in B. So, in particular, no element of B can be smaller than α. This means that $\alpha \leq \gamma$ for all γ in B, and therefore α is a lower bound of B, i.e. that $\alpha \in L$. We are just left to show that α is such that no $\beta > \alpha$ can be a lower bound for B, i.e. there cannot be an element in L which is larger than α. But this follows from the fact that α is the sup of L, by definition. This completes the proof. $\square$

The following is a fundamental definition in all mathematics (and applications such as theoretical physics).

Definition 2.11. A field is a set $\mathbb{k}$ with two operations, usually denoted by $+$ and $\cdot$, satisfying the following properties.

- Addition $+$ is commutative.
- Addition $+$ is associative.
- There is an element in $\mathbb{k}$, denoted by 0, with the property that $x + 0 = x$ for all x.
- For every $x \in \mathbb{k}$ there is an element in $-x \in \mathbb{k}$ such that $x + (-x) = 0$.
- Multiplication $\cdot$ is commutative.
- Multiplication $\cdot$ is associative.
- There is an element $1 \neq 0$ such that $x \cdot 1 = x$ for all x.
- For every $x \in \mathbb{k}^\times$, the set of nonzero elements, there exists an element $\frac{1}{x} \in \mathbb{k}$ such that $x \cdot \frac{1}{x} = 1$.
- Multiplication distributes over addition, i.e. $x(y + z) = xy + xz$.

Remark 2.12. The usual conventions used for $\mathbb{R}$ will hold, for instance we will write $x - y$ for $x + (-y)$ and so on.

Exercise 2.13. Check the following (precisely stating what conditions fail):

- $\mathbb{N}, \mathbb{Z}$ are not fields.
- $\mathbb{Q}, \mathbb{R}, \mathbb{C}$ are fields.
- $\mathbb{H}$ (the quaternions) is not a field.

- $\mathbb{Z}_n$ is a field if and only if n is prime.

Note that we did not really invent yet $\mathbb{R}$, and therefore $\mathbb{C}$ and $\mathbb{H}$. However, since the reader has already had enough practice with them, we will forgive this anachronistic exercise.

Definition 2.14. An ordered field is a field $\Bbbk$ endowed with an ordering (i.e. it is also an ordered set) such that the following two conditions hold:

- If $y < z$, then $x + y < x + z$ for all x.
- If $x, y > 0$, then $xy > 0$ as well.

The main result of this first part of the course is the following.

Theorem 2.15. *There exists an ordered field $\mathbb{R}$ which contains $\mathbb{Q}$ as a subfield, and has the least-upper-bound property.*

We will defer the proof of this fundamental theorem to a later part of the course, but we will derive some important consequences nonetheless.

Theorem 2.16. *The following two facts hold.*

(a) *For every $x > 0$ and $y \in \mathbb{R}$, there exists a natural number $n \in \mathbb{N}$ such that $nx > y$. As a special case, for any real number y, we can find $n \in \mathbb{N}$ such that $n > y$.*
(b) *For every $x, y \in \mathbb{R}$ with $x \neq y$, there exists a $p \in \mathbb{Q}$ such that $x < p < y$.*

Proof. We prove (a) first. Let A be the set of all numbers of the form nx, where n runs over all natural numbers $\mathbb{N}$. Let us assume by way of contradiction that the assertion is false. Then, y is an upper bound for A, since $nx \leq y$ would hold for all n. Since $\mathbb{R}$ has the least-upper-bound property, there exists the supremum $\alpha = \sup A$. Since $x > 0$ by hypothesis, we have that $\alpha - x < \alpha$. This means that $\alpha - x$ is not an upper bound (by definition of supremum), and therefore there is an element of A, say mx for some $m \in \mathbb{N}$, such that $\alpha - x < mx$. However, this means that $\alpha < x + mx = (m+1)x$. Since $m + 1 \in \mathbb{N}$, it follows that α is not the supremum of A, which is a contradiction. We leave the proof of the special case to the reader. Therefore, (a) is proved.

Let us now prove (b). Since $x < y$, we have that $y - x > 0$. Using (a) we find a natural number n such that $n(y - x) > 1$. Applying (a) again twice (or better to say, its special case) we can find m_1 and m_2 such that $m_1 > nx$ and $m_2 > -nx$. This means that we have the inequalities

$$-m_2 < nx < m_1. \tag{3}$$

We can therefore find an integer m with the property that $-m_2 \leq m \leq m_1$, and such that

$$m - 1 \leq nx < m. \tag{4}$$

Using (4) and the fact that $n(y - x) > 1$, we find that

$$nx < m \leq 1 + nx < ny. \tag{5}$$

It then follows that $x < \frac{m}{n} < y$, which is what we wanted to prove by setting $p = \frac{m}{n}$. □

Exercise 2.17. In the previous proof, we used the fact that from $-m_2 < nx < m_1$ it follows that we can find m such that $m - 1 \leq nx < m$. Prove this fact. [*Hint: Use induction wisely.*]

Now we show a fundamental result of the real numbers, which was our motivation to embark in the construction of $\mathbb{R}$.

Theorem 2.18. *For every real number $x > 0$, and every integer $n > 0$ there is one and only one positive real y such that $y^n = x$.*

Proof. First, we mention that if such a number y exists, this has to be unique. In fact, assuming otherwise, let y_1 and y_2 be two such numbers. Since $\mathbb{R}$ is ordered and $y_1 \neq y_2$, it must hold that $y_1 < y_2$ or $y_2 < y_1$. Without loss of generality, assume that $y_1 < y_2$. Since $\mathbb{R}$ is an ordered field, it follows that $y_1^n < y_2^n$, and therefore $x < x$, which obviously nonsense. This means that y has to be unique, if it exits.

We now show existence. Let E denote the set of all positive real numbers t such that $t^n < x$. Consider the number $t = \frac{x}{1+x}$. Then, it follows that $0 \leq t < 1$ and therefore $t^n \leq t < x$. Therefore, E is not the empty set. Consider now any number $t > 1 + x$. Then, $t^n \geq t > x$, which means that such t would not be elements of E. This means that $1 + x$ is an upper bound of E. Using the least-upper-bound-property of $\mathbb{R}$, there is an upper bound $y = \sup E$. We now claim that y is the number we seek.

To show that, we proceed by contradiction. We first assume that $y^n \neq x$. Therefore, it must hold either $y^n < x$, or $y^n > x$. We show that either of the inequalities will lead to a contradiction, which would conclude the proof.

First observe that if $b - a > 0$, with both $a, b > 0$ one has that $b^n - a^n = (b-a)(b^{n-1} + ab^{n-2} + \cdots + a^{n-1})$ implies $b^n - a^n < (b-a)nb^{n-1}$. This fact will be useful in the rest of the proof.

Case: $y^n < x$. We choose $0 < h < 1$ such that $h < \frac{x - y^n}{n(y+1)^{n-1}}$. Put $a = y$ and $b = y + h$. Then, we have that

$$(y+h)^n - y^n < hn(y+h)^{n-1} < hn(y+1)^{n-1} < x - y^n. \tag{6}$$

It therefore follows that $(y+h)^n < x$. But since $y + h > y$, this means that y cannot be the supremum of E, which is absurd.

Case: $y^n > x$. We put $k = \frac{y^n - x}{ny^{n-1}}$, which implies that $0 < k < y$. If $t \geq y - k$, we get

$$y^n - t^n \leq y^n - (y-k)^n < kny^{n-1} = y^n - x. \tag{7}$$

So, we find that $t^n > x$, which means that t is not in E. Therefore, $y - k$ is an upper bound for E. But $y - k < y$, against the fact that y is the supremum of E. This means that $y^n > x$ is false. The proof is now complete. □

3. Topology of Metric Spaces

In this section we will introduce some basic notions of topology. While our focus will be on metric spaces (which we will introduce in this section), some elementary definitions will be given in general terms for more abstract topological spaces.

Topologies over a set give a notion of "being close" even when the space does not have a numerical quantification for such a term. For instance, for real numbers one can consider the absolute value of the difference of two numbers as the quantification of how far apart they are. On abstract sets, however, it is unclear how such a quantification would be defined. In particular, topologies allow us to define "neighbor" elements, which is of significant importance throughout the sciences.

Definition 3.1. Let X be a set, and $\mathcal{T}$ a collection of subsets of X. Then, we say that $\mathcal{T}$ is a *topology* on X if the following condition are satisfied:

- The empty set $\emptyset$ and the set X are in $\mathcal{T}$.
- Arbitrary union of elements in $\mathcal{T}$ is in $\mathcal{T}$.
- Finite intersection of elements in $\mathcal{T}$ is in $\mathcal{T}$.

Elements in $\mathcal{T}$ are called *open sets*, while complements of open sets are called *closed sets*.

The fundamental idea in topology is that elements that can be separated "easily" by open sets are far apart, while if two points are not easily put in two different open sets are close to each others. Let us consider a very familiar example. We will denote by $\mathcal{P}(X)$ the power set of a given set X, so a topology $\mathcal{T}$ is a subset of the power set $\mathcal{P}(X)$ of X.

Example 3.2. Let X be a set. The simplest topologies on X are the two most obvious ones:

- $\mathcal{T} = \mathcal{P}(X)$.
- $\mathcal{T} = \{\emptyset, X\}$.

It is immediate to show that they are both topologies.

Example 3.3. A less boring topology on a set with 4 elements $X = \{1,2,3,4\}$ is the following, $\mathcal{T} = \{\emptyset, \{2\}, \{1,2\}, \{2,3\}, \{1,2,3\}, X\}$. One can verify directly that the union of elements in $\mathcal{T}$ is still in $\mathcal{T}$, and that intersections of elements in $\mathcal{T}$ are still in $\mathcal{T}$.

Example 3.4. Consider $\mathbb{R}$ with the following set $\mathcal{T} \subset \mathcal{P}(\mathbb{R})$. We say that A is in $\mathcal{T}$ if for every point $x \in A$, we can find an interval (p,q) containing x and contained in A, $x \in (p,q) \subset A$.

Let us verify that this is indeed a topology. First, observe that $\emptyset$ vacuously satisfies the definition of being in $\mathcal{T}$, and it is therefore open. It is also clear that $\mathbb{R}$ is in $\mathcal{T}$ as well, since any interval is a subset of $\mathbb{R}$, and we can always find an interval containing any chosen element of $\mathbb{R}$. Let $A_{t t\in\tau}$ be a family of elements in $\mathcal{T}$, where the index set has arbitrary cardinality (we need arbitrary unions!). We want to show that the union $A = \bigcup_{t\in\tau} A_t$ is in $\mathcal{T}$ as well. Let $x \in A$ be an element of the union. Then, $x \in A_t$ for some $t \in \tau$. Since A_t is open, there is an interval (p,q) containing x such that $(p,q) \subset A_t$. But then (p,q) is an interval contained also in the union A. Therefore A is open, since x was arbitrary. Similar reasoning can be used to show that finite intersections of open sets are still open. We show this for two open sets. The reader then can either directly generalize the argument to arbitrary (but finitely many) open sets, or can use an inductive argument to show that it holds for any (finite) number of open sets. Let A and B be open sets. Then, let $A \cap B$ denote their intersection. If $A \cap B = \emptyset$ there is nothing to show, since we already know that $\emptyset \in \mathcal{T}$. Suppose $x \in A \cap B \neq \emptyset$. Since A is open, we can find $x \in (p_1, q_1) \subset A$. Similarly, since B is open, we can also find $x \in (p_2, q_2) \subset B$. Since $\mathbb{Q}$ is dense in $\mathbb{R}$ by Theorem 2.16 (b), it follows that we can find $x < q < \min\{q_1, q_2\}$ and $\max\{p_1, p_2\} < p < x$. It is clear that $(p,q) \subset A, B$ and $x \in (p,q)$, showing that $A \cap B$ is open as well.

Therefore, $\mathcal{T}$ is a topology on the real numbers. This topology is called the *Euclidean topology* of $\mathbb{R}$.

Exercise 3.5. Show that if $\{C_j\}_{j\in J}$ is an arbitrary family of closed sets, then $\bigcap_{j\in J} C_j$ is closed. So, while only finite intersections of open sets are open, arbitrary intersections of closed sets are closed.

What can you say about unions of closed sets?

There is a general way of definining a topology by listing a set of subsets with some additional properties.

Theorem 3.6. *Let X be a set and let $\mathcal{B}$ be a set. Then, the set of all unions of elements of $\mathcal{B}$ defines a topology on X if and only if $\mathcal{B}$ satisfies the following two properties:*

(1) $\bigcup_{B\in\mathcal{B}} B = X$.
(2) *If B_1 and B_2 are elements in $\mathcal{B}$, and $x \in B_1 \cap B_2$, then there is an element $B \in \mathcal{B}$ such that $x \in B \subset B_1 \cap B_2$.*

Proof. Suppose first that $\mathcal{T}$, defined as the set of all unions of elements of $\mathcal{B}$ is a topology. We want to verify that (1) and (2) hold. Since X is always an element of a topology, it means that $X \in \mathcal{T}$, which means that X can be written as the union of elements in $\mathcal{B}$. This shows (1). To show (2), since finite intersection of elements in $\mathcal{T}$ is in $\mathcal{T}$, and since $\mathcal{B} \subset \mathcal{T}$, we find that $B_1 \cap B_2 \in \mathcal{T}$. This means that $B_1 \cap B_2 = \cup_{t\in\tau} B_t$ for some index set τ and elements B_t of $\mathcal{B}$. Given $x \in B_1 \cap B_2$, there must exist some B_q such that $x \in B_t$. Clearly $B_q \subset \cup_{t\in\tau} B_t = B_1 \cap B_2$, which shows (2).

Conversely, let us now consider a family $\mathcal{B}$ satisfying both (1) and (2). Defining $\mathcal{T}$ to be the family of all unions of elements of $\mathcal{B}$, we want to verify that such $\mathcal{T}$ is indeed a topology. Observe that $X \in \mathcal{T}$ by (1), and also $\emptyset \in \mathcal{T}$ since it is the empty union of elements from $\mathcal{B}$. It is clear that the union of A_t, with $t \in \tau$ some index set, and A_t in $\mathcal{T}$ results in an element of $\mathcal{T}$. We need to verify that the intersection of two open sets is still open. Observe that from (2), we find that for any $B_1, B_2 \in \mathcal{B}$, we have the equality

$$B_1 \cap B_2 = \bigcup \{C \in \mathcal{B} \mid C \subset B_1 \cap B_2\}. \tag{8}$$

If A_1, A_2 are elements of $\mathcal{T}$, then by definition we have $A_1 = \cup_i U_i$ and $A_2 = \cup_i V_j$ for open sets U_i and V_j. Then, we have that

$$A_1 \cap A_2 = \bigcup_{i,j} U_i \cap V_j = \bigcup \{C \in \mathcal{B} \mid C \subset U_i \cap V_j\}. \tag{9}$$

This completes the proof. □

Definition 3.7. A family $\mathcal{B}$ satisfying the properties (1) and (2) in Theorem 3.6 is called a *basis* for the topology $\mathcal{T}$.

Example 3.8. Let us define now the *Sorgenfrey line*. This is the set of real numbers $\mathbb{R}$ with the topology $\mathcal{T}$ where the basis $\mathcal{B}$ is defined as the set of all semi-closed intervals $[a, b)$. One can immediately see that such $\mathcal{B}$ is indeed a basis, and therefore $\mathcal{T}$ is a topology.

The following definition plays a fundamental role in all analysis.

Definition 3.9. Let X be a topological space, and $B \subset X$. We then define the following concepts.

(1) The set obtained by taking the union of all open sets contained in B, denoted by B° and called the *interior* of B.
(2) The set obtained as the intersection of all closed sets containing B, denoted by $\bar{B}$, and called the *closure* of B.
(3) The set $\partial B := \bar{B} - B^\circ$, called the *boundary* of B.

The elements of B° are also called *interior points* of B.

Exercise 3.10. Show that a set B is open if and only if $B = B^\circ$, and it is closed if and only if $B = \bar{B}$.

Definition 3.11. A subset A of a topological space X is said to be *dense* in X if $\bar{A} = X$.

Example 3.12. In Theorem 2.16 we have shown that given two real numbers x, y, we can find a rational p such that $x < p < y$. This means that $\mathbb{Q}$ is dense in $\mathbb{R}$ with the Euclidean topology. In fact, A intersects any open set of $\mathbb{R}$ (because it intersects all intervals!). So, the only closed set of $\mathbb{R}$ that contains $\mathbb{Q}$ is $\mathbb{R}$ itself. Therefore, $\bar{\mathbb{Q}} = \mathbb{R}$.

Definition 3.13. Let X be a topological space, and let x be an element of X. We say that the subset $U \subset X$ is a *neighborhood* of x if $x \in U^\circ$. In other words, if x is an interior point of U, we say that U is a neighborhood of x.

The fundamental idea of using topological spaces, is that it allows one to define continuous functions in a very general way. Here we will also call functions simply as *maps*, to indicate that the domain and target are arbitrary (not necessarily numerical sets).

Definition 3.14. Let X, Y be topological spaces, and let $f : X \longrightarrow Y$ be a map between them. Then, we say that f is *continuous* if $f^{-1}(V)$ is an open set of X whenever V is an open set of Y. In other words, the preimage of open sets is open.

Remark 3.15. Observe that continuity is equivalently defined as $f^{-1}(C)$ is closed in X whenever C is a closed subset of Y. This is shown by considering the complements.

Theorem 3.16. *Composition of continuous maps is continuous.*

Proof. Left to the reader as an (important) exercise. □

Definition 3.17. Let X, Y be topological spaces. A continuous bijective map $f : X \longrightarrow Y$ is said to be a *homeomorphism* if the inverse f^{-1} is continuous as well. The spaces X and Y are said to be *homeomorphic*, if an homeomorphism between them exists.

Definition 3.18. Let X be a set. Then, a *metric*, or *distance* on X, is a map $d : X \times X \longrightarrow \mathbb{R}$ satisfying the following conditions.

(1) $d(x, y) \geq 0$ for all x, y, and $d(x, y) = 0$ if and only if $x = y$.
(2) $d(x, y) = d(y, x)$ for all x, y.
(3) $d(x, z) \leq d(x, z) + d(y, z)$ for all x, y, z.

Condition (3) is called the *traingular inequality.* A set X endowed with a metric d is called a *metric space*, and often denoted as a pair (X, d).

Example 3.19. Let X be a set. Define the following map

$$d(x, y) = \begin{cases} 0 & x = y \\ 1 & x \neq y. \end{cases}$$

Then, d defines a metric on X.

Example 3.20. Consider the space $\mathbb{R}^n$ and define

$$d(x, y) = \sqrt{(x_1 - y_1)^2 + \cdots + (x_n - y_n)^2}.$$

Then $\mathbb{R}^n$ is a metric space with such d. This is called the Euclidean space, and the metric is called the Euclidean metric of $\mathbb{R}^n$. It is clear that when $n = 1$.

The only property of metrics that is not trivially satisfied in this example is the triangular inequality. To this purpose, define $\|x\| = d(x, 0)$, which is called the norm of x. Observe that $d(x, y) = \|x - y\|$ follows from the definitions. Then, showing the triangular inequality is equivalent to showing that $\|u + v\| \leq \|u\| + \|v\|$, where we are setting $x - z = u$ and $z - y = v$. Equivalently, this is the same as showing that $(\|u\| + \|v\|)^2 - \|u + v\|^2 \geq 0$. Defining $x \cdot y = \sum x_i y_i$, we have $2(\|u\|^2\|v\|^2 - (u \cdot v)^2) = (\|u\| + \|v\|)^2 - \|u + v\|^2$. So, we will just prove that $\|u\|^2\|v\|^2 - (u \cdot v)^2 \geq 0$, which is known as *Cauchy-Schwarz inequality*, and complete.

Observe that when $v = 0$, there is nothing to prove. So, we can assume that $v \neq 0$, and define $w := u\|v\| - (u \cdot v)\frac{v}{\|v\|}$. One directly sees that $\|w\|^2 = \|u\|^2\|v\|^2 - (u \cdot v)^2$. Since $\|w\|^2 \geq 0$, the Cauchy-Schwarz inequality follows, and therefore the triangular inequality follows as well.

Metric spaces are of fundamental importance in analysis. They are automatically endowed with a natural notion of topology induced by the metric, as we now see.

Definition 3.21. Let (X, d) be a metric space. Then, given $x \in X$ and $\rho > 0$, the set $B(x, \rho) = \{y \in Y \mid d(x, y) < \rho\}$ is called an *open ball around x with radius ρ*, or also just an open ball, or a ball.

Definition 3.22. Let (X, d) a metric space. We say that a subset $A \subset X$ is open if whenever $x \in A$, there exists a $\rho > 0$ such that $B(x, \rho) \subset A$. The topology induced by d is the family $\mathcal{T}$ consisting of all open sets so defined.

Exercise 3.23. Let $X = \mathbb{R}$, $Y = \mathbb{R}$, and let $f : X \longrightarrow Y$ be a continuous map (according to Definition 3.14), where X and Y have the topology induced by the metric d defined in Example 3.20 with $n = 1$. Show that continuity in this sense gives the (in)famous $\epsilon - \delta$ definition of continuity that is (usually not) seen in Calculus I.

Exercise 3.24. Show that the topology induced by d as in Example 3.20 with $n = 1$ gives the topology of Example 3.4.

Exercise 3.25. Show that open balls in a metric space satisfy the conditions of Theorem 3.6 to induce a topology. Then, show that the topology induced by a metric d is the same as the topology having open balls as its basis.

Definition 3.26. Let (X, d) be a metric space and let B and U be subsets of X.

- Then, we say that B is *bounded* if there is a ball $B(x, \rho)$ such that $B \subset B(x, \rho)$. Note that x does not need to be taken inside B.
- An element x is said to be a *limit point*, or *accumulation point*, for the subset U in X, if every neighborhood of x contains elements of U that are distinct from x itself. If x is not accumulation point for U, but it belongs to U, then it is said to be *isolated*.

Theorem 3.27. *Let X be a topological space. Let $E \subset X$ be a subset, and denote by E' the set of accumulation points of E. Then,*

- $\bar{E} = E \cup E'$.
- $E = E \cup E'$ *if and only if E is closed.*

Proof. Observe that according to the definitions, $\bar{E}$ is the intersection of all closed sets containing E. So, the first equality is proved if we show that E' is contained in any closed set that contains E, and also that $E \cup E'$ is closed (i.e. it is itself one of the closed sets containing E).

Let F be a closed set containing E, and let x be an element of E' (i.e. an accumulation point) which is not in F. Since F is closed, $X - F$ is open, and since x is not in F, $x \in X - F$. This means that $X - F$ is a neighborhood of x. But F contains E, so that $X - F$ does not intersect E, meaning that x has a neighborhood that does not intersect E, against the fact that it is an accumulation point. This is absurd, and therefore there cannot be any accumulation point of E not contained in a closed set containing E. This means that E' is a subset of the intersection of all closed sets containing E. We want to show now that $E \cup E'$ is closed, which would complete the proof of the first part. To show that it is closed, we need to show that its complement is open. Then, let $x \in X - (E \cup E')$. We need to find a neighborhood of x contained in $X - (E \cup E')$. To this scope, take a closed set F containing E, and such that x is not in F. This can be done because x is not an accumulation point. As already shown, $E' \subset F$. This means that $X - F$ is open and contained in $X - (E \cup E')$. But then it follows that $X - F$ is a neighborhood of x contained in $X - (E \cup E')$. Therefore, $X - (E \cup E')$ is open since we can perform the previous procedure for any $x \in X - (E \cup E')$. This shows that $E \cup E'$ is closed, and therefore $\bar{E} = E \cup E'$.

We already know that a subset of a topological space is closed if and only if it is equal to its closure, i.e. E is closed if and only if $E = \bar{E}$. Since $\bar{E} = E \cup E'$ from the previous part of the proof, this part also follows. □

Corollary 3.28. *Let E be a nonempty set of real numbers which is bounded above, and let $y = \sup E$. Then, $y \in \bar{E}$. In particular, if E is closed, $y \in E$.*

Proof. First, observe that if $y \in E$ (i.e. y is the maximum of E) then $y \in \bar{E}$ by definition. So, we can simply consider the case $y \notin E$. For any choice of $h > 0$, by definition of supremum, we have that $y - h$ is not an uppear bound for E. Therefore, this means that there is at least an element $x \in E$ with $y > x > y - h$. In other words, any ball around y contains elements of E. This means that y is an accumulation point for E, so $y \in E'$, where E' denotes the set of accumulation points of E. By Theorem 3.27 it follows that $y \in \bar{E}$. When E is closed we know, as already seen before, that $E = \bar{E}$ and therefore the second statement follows immediately too. □

Remark 3.29. Analogous result holds for a subset bounded below, where sup needs to be changed to inf.

Remark 3.30. If $Y \subset X$ is a subset of a topological space, then there is a natural topology automatically defined on Y. This is the topology given by the family $\mathcal{T}_Y$ obtained by taking the intersection of the open sets of X with Y. So, $U \in \mathcal{T}_Y$ if and only if $U = V \cap Y$ for some open V of X. Therefore, we will talk of properties relative to subsets of a topological space, meaning that the definition applies to them with respect to the induced topology. For instance, we will talk about closed subsets of a subset Y of X and so on.

We now consider the notion of compact spaces.

Definition 3.31. Let $E \subset X$ be a asubset of the topological space X. Then, an *open cover* of E is a family of open sets $\{G_\alpha\}$ such that $E \subset \bigcup_\alpha G_\alpha$.

Definition 3.32. A subset K of the topological space K is said to be *compact* if for any open cover $\{G_\alpha\}$ of E, there is a finite subcover, i.e. we can find finitely many indices $\alpha_1, \dots, \alpha_n$ such that $E \subset \bigcup_{i=1}^n G_{\alpha_i}$.

Remark 3.33. Compact sets play a fundamental role in Analysis. They give us a pathway to pass from infintely many objects to only finitely many.

Example 3.34. We will show later that there is a very important class of compact sets in $\mathbb{R}^n$. These are the closed balls, i.e. $\bar{B}(x, \rho) = \{y \in \mathbb{R}^n \mid d(x, y) \leq \rho\}$, where d is the Euclidean distance.

Theorem 3.35. *Let $f : X \longrightarrow Y$ be a continuous function, and let X be compact. Then, $f(X)$ is compact in Y.*

Proof. Let $\{V_\alpha\}$ be a collection of open sets of Y covering $f(X)$. Then, since f is continuous, every preimage $U_\alpha := f^{-1}(V_alpha)$ is open in X. Moreover, the U_α's cover X. Since X is compact, we can find a finite subcover. The image of this subcover consists of finitely many V_α's that cover $f(X)$, showing that $f(X)$ is also compact. □

Theorem 3.36. *Compact subsets of metric spaces are closed.*

Proof. Let K be compact in X, metric space. We want to show that $X - K$ is open (which would mean that K is closed by definition). Let $x \in X - K$, and for every $y \in K$ consider the ball around

y of radius $\rho_y := \frac{d(x,y)}{2}$, $B(y, \rho_y)$. For any chosen y, $\rho_y < d(x, y)$, therefore x is not in $B(y, \rho_y)$. Since each y belongs to one of the balls – it is certainly in the ball $B(y, \rho_y)$ of which it is center – the family of balls covers K: $K \subset \bigcup_y B(y, \rho_y)$. Being compact, K can be covered with only finitely many of them, so we can find $y_1, \ldots, y_n$ such that $K \subset \bigcup_{i=1}^n B(y_i, \rho_{y_i})$. Of course, we find that $A = \bigcup_{i=1}^n B(y_i, \rho_{y_i})$ is an open set, being union of open sets. Also, by construction x is not in A (because it is not in any of the balls!). We can also say more, if $0 < r < \min\{\rho_{y_i}\}$, $B(x, r)$ does not intersect any of the balls $B(y_i, \rho_{y_i})$, and therefore it does not intersect A. This means that x is an interior point to $X - K$. Since we can do this for any $x \in X - K$, it means that $X - K$ consists only of interior points, which means it is open. This completes the proof. □

Theorem 3.37. *Closed subsets of compact sets are compact.*

Proof. Let $F \subset K$ be closed in X, and let $K \subset X$ be compact. Let $\{V_\alpha\}$ be an open cover of F. Since F is closed, $A = X - F$ is open, and $\{V_\alpha\} \cup \{A\}$ is an open cover of K. So, we can find an open subcover, i.e. we can find finitely many α_i such that $K \subset \bigcup_{i=1}^n V_{\alpha_i} \bigcup A$. Since $F \subset K$, we also have $F \subset \bigcup_{i=1}^n V_{\alpha_i} \bigcup A$. But $A = X - F$, so A does not contribute to the previous inclusion and we can remove it without changing it. We have $F \subset \bigcup_{i=1}^n V_{\alpha_i}$, but this means that we have found an open subcover of F from $\{V_\alpha\}$. So, F is compact. □

Corollary 3.38. *If F is closed, and K is compact, then $F \cap K$ is compact.*

Proof. Since K is compact, it is also closed by Theorem 3.36. We have that $F \cap K$ is closed as well, being the intersection of two closed. Since $F \cap K \subset K$ is a closed subset of a compact, by Theorem 3.37 it is also compact. □

Theorem 3.39. *If $\{K_\alpha\}$ is a collection of compact subsets of X with the property any intersection of a finite subcollection of K_α's is nonempty, then $\bigcap_\alpha K_\alpha \neq \emptyset$ as well.*

Proof. We fix a member K_* of $\{K_\alpha\}$. Set $G_\alpha = X - K_\alpha$ for all $\alpha \neq *$. Suppose that K_* does not intersect every K_α. Then, it means that every element of K_* belongs to some G_α. In other words, $K_* \subset \bigcup_{\alpha \neq *} G_\alpha$. Since K_* is compact, we can find finitely many α's, say $\alpha_1, \ldots, \alpha_n$ such that $K_* \subset G_{\alpha_1} \cup \cdots \cup G_{\alpha_n}$. However, this means that $K_* \cap K_{\alpha_1} \cap \cdots \cap K_{\alpha_n} = \emptyset$, against the hypothesis that every choice of finitely many of K_α have nonempty intersection. This contradiction proves the result. □

The following corollary will be very useful in this course.

Corollary 3.40. *If $\{K_n\}$ is a sequence of nonempty compact sets such that $K_{n+1} \subset K_n$ for all n, then $\bigcap_{n=1}^\infty K_n \neq$.*

Theorem 3.41. *If E is an infinite subset of a compact set K, then E has an accumulation point in K.*

Proof. Since no point of K is accumulation point of E, for any choice of $x \in E$ we can find a neighborhood V_x of x which intersects E at most in a single point (which is x if $x \in E$). Since E is infinite, no finite subcollection of $\{V_x\}$ can cover E, and therefore no finite subcollection can cover K, against the fact that K is compact. So, K must have some accumulation point of E. □

Theorem 3.42. *Let $\{I_n\}$ be a sequence of closed intervals in $\mathbb{R}$ with the property that $I_{n+1} \subset I_n$ for all n. Then, $\bigcap_{n=1}^\infty I_n \neq \emptyset$.*

Proof. Let $I_n = [a_n, b_n]$ for all n. Define $E = \{a_i \mid i \in \mathbb{N}\}$. The set E is clearly nonempty, and since $I_{n+1} \subset I_n$ it is bounded above by b_1 (why?). Let $x := \sup E$. For all natural numbers $n, m \in \mathbb{N}$ we have the inequalities

$$a_n \leq a_{n+m} \leq b_{n+m} \leq b_m,$$

which implies that $x \leq b_m$ for all m. Obviously, $x \geq a_m$ for all m by definition. Therefore, $x \in I_m$ for all m, and the intersection is not empty. □

Theorem 3.43. *Let $k \in \mathbb{N}$ and let $\{I_n\}$ be a sequence of k-dimensional closed hypercubes of $\mathbb{R}^k$ with the property that $I_{n+1} \subset I_n$. Then, $\bigcap_{n=1}^\infty I_n \neq \emptyset$.*

Proof. For every n, let the hypercube I_n consist of the product of intervals $I_n = [a_1^n, b_1^n] \times \cdots \times [a_k^n, b_k^n] := I_n^1 \times \cdots \times I_n^k$. Every interval satisfies the hypotheses of Theorem 3.42, so we can find a point x_j^* in I_n^j for $j = 1, \ldots, k$ and all $n \in \mathbb{N}$. It follows that $x^* = (x_1^*, \ldots, x_k^*)$ is in the intersection of all I_n. □

Theorem 3.44. *Every closed k-dimensional hypercube is compact.*

Proof. Let $I = [a_1, b_1] \times \cdots \times [a_k, b_k]$. Define $\delta = \sqrt{(b_1 - a_1)^2 + \cdots + (b_k - a_k)^2}$. Then, if $x, y \in I$ we have that $d(x, y) \leq \delta$. We proceed by way of contradiction. Suppose that there is an open cover G_α of I which does not contain any finite subcover. We set $c_j = \frac{a_j + b_j}{2}$ for $j = 1, \ldots, k$. Define the intervals $[a_j, c_j]$ and $[c_j, b_j]$ for all $j = 1, \ldots, k$. Then, this gives us 2^k smaller hypercubes Q_i whose union is equal to I. Since I cannot be covered by finitely many G_α's, it follows that at least one of the hypercubes Q_i cannot be covered by finitely many of the G_α's. Let us call this hypercube I_2, and denote the initial I as I_1. It is clear that $I_2 \subset I_1$. Iterate this process and subdivide I_1 in 2^k sub-hypercubes and continue. At each step, we obtain $I_{n+1} \subset I_n$. We have the following additional properties of our sequence of hypercubes:

- No I_n can be covered by finitely many G_α's.
- For each $x, y \in I_n$ we have $d(x, y) \leq 2^{-n+1}\delta$.

Applying Theorem 3.43 it follows that there is a point $x^* \in \bigcap_{n=1}^\infty I_n$. Since the G_α's cover I, and the intersection of all I_n is contained in I, it follows that $x^* \in G_\alpha$ for some α. Since G_α is open, it follows that there is a ball around x^* of some radius, say $r > 0$, such that $B(x^*, r) \subset G_\alpha$. Taking n large enough to satisfy the inequality $2^{-n+1}\delta < r$, then it follows that $I_n \subset B(x^*, r) \subset G_\alpha$ against the fact that no I_n can be covered by finitely many G_α's. This contradiction proves the result. □

Theorem 3.45 (Heine-Borel)**.** *Let $E \subset \mathbb{R}^k$ be a subset. Then, the following properties are equivalent.*

(i) *E is closed and bounded.*
(ii) *E is compact.*
(iii) *Every infinite subset of E has an accumulation point in E.*

Proof. $[(i) \implies (ii)]$ If E is bounded, then we can find a ball $B(x, \rho)$ containing E. This means that we can find a closed hypercube I of side ρ containing E as well. Since I is compact and E is closed, $E = E \cap I$ is compact by Corollary 3.38.

$[(ii) \implies (iii)]$ This is the content of Theorem 3.41.

$[(iii) \implies (i)]$ We proceed by contradiction. If E is not bounded, then it has a sequence of points x_n where $d(x_n, 0) > n$ for all n. This is because whenever we choose a ball around 0 of radius n, the ball cannot contain E, so we can find an x_n outside of the ball. The set of such elements x_n

cannot have an accumulation point in E. This contradiction means that E is bounded. Suppose now that E is not closed. Then, this means that $\bar{E} \neq E$. Since $E \subset \bar{E}$ is always true, it follows that E not being closed implies that there is an acumulation point of E which does not belong to E. Let us call this element x^*. By definition of accumulation point for every n we can find x_n in the ball $B(x^*, \frac{1}{n})$. In other words, $d(x_n, x^*) < \frac{1}{n}$. Denote by S the set of the points x_n. It is clear that S is an infinite set having only x^* as limit point. This contradicts property (iii), which completes the proof. □

Theorem 3.46 (Weierstrass)**.** *Every bounded infinite subset E of $\mathbb{R}^k$ has an accumulation point in $\mathbb{R}^k$.*

Proof. Since E is bounded, we can find a hypercube I which contains E. This hypercube is compact. Then, E is an infinite subset of a compact and bounded. Therefore, by Theorem 3.45 there is a limit point of E in I. □

We can now prove a very important result of what we have done so far.

Theorem 3.47. *Let $f : X \longrightarrow \mathbb{R}$ be a continuous function where X is compact. Then, f attains its maximum and minimum on X.*

Proof. Since $f(X)$ is compact in $\mathbb{R}$ (Theorem 3.35), it is closed and bounded. Since it is bounded, it has a supremum and a infimum. Being closed, supremum and infimum are in $f(X)$, so they are maximum and minimum (Corollary 3.28 and Remark 3.29). □

Corollary 3.48 (Extreme Value Theorem)**.** *Lef $f : [a, b] \longrightarrow \mathbb{R}$ be continuous. Then, f attains its maximum and minimum values in $[a, b]$.*

Proof. Intervals are compact by Theorem 3.44, and we can therefore apply Theore 3.47. □

Definition 3.49. We say that a set is *perfect* if it is closed, and it has no isolated points.

The main property of perfect sets is that every point is an accumulation point, and therefore we can approximte any point in the set by means of other points in the set.

Exercise 3.50. Show that if P is perfect, then we have $P = P'$, where P' indicates the set of accumulation points as before.

Perfect sets in $\mathbb{R}^k$ are large. We show this in the following.

Theorem 3.51. *Let P be a nonempty perfect set in $\mathbb{R}^k$. Then P is uncountable.*

Proof. Since P has at least one limit point, P must be infinite. By way of contradiction, we suppose P countable. Pick one element x_1 of P, and let V_1 denote any open neighborhood of x_1. We want to construct a sequence of open sets V_n with bounded closure containing an element x_n of P by induction, where V_1 serves as the induction basis. We can take V_1 to be bounded, upon possibly choosing a ball around x_1. Suppose that V_n has been constructed. Since every point of P is a limit point of P, we can find an open neighborhood V_{n+1} with bounded $\bar{V}_{n+1}$ such that

- $\bar{V}_{n+1} \subset V_n$.
- $x_n \notin \bar{V}_{n+1}$.
- $V_{n+1} \cap P \neq \{x_{n+1}\}$.

This can be shown as follows. Since x_n is accumulation point for P, V_n has to contain also other elements of P rather than x_n itself. Call this element y. Since V_n is open, y has a ball of radius ρ contained in V_n. Take $0 < r < \min\{d(x_n, y), \rho\}$, and set $V_{n+1} = B(y, r)$. It is clear by construction that V_{n+1} has the first two properties listed above. The third one follows from the fact that $P = P'$, and therefore any neighborhood of y must contain some element of P which is not y itself. We can then call $y = x_{n+1}$ and continue our inductive construction. Put $K_n := \bar{V}_n \cap P$. We have that $\bar{V}_n$ is closed and bounded in $\mathbb{R}^k$, then $\bar{V}_n$ is compact for all n. Since $x_n \notin K_{n+1}$, no point of P lies in all K_n, and therefore $\cap K_n = \emptyset$. But this is in contrast with Corollary 3.38, since the family $\{K_n\}$ satisfies also all the hypotheses of the corollary. □

Corollary 3.52. *Every interval $[a, b]$ is uncountable. Therefore, $\mathbb{R}$ is uncountable as well.*

We now introduce another fundamental notion, akin to compactness. This will allow us to obtain another important result that is usually not proved in Calculus I courses. This is the intermediate value theorem.

Definition 3.53. A topological space X is said to be *connected* if the only subsets of X that are both open and closed are X and $\emptyset$.

The following lemma gives a characterization of connectedness.

Lemma 3.54. *The following are equivalent:*

- *X is not connected.*
- *X is disjoint union of two proper open subsets.*
- *X is disjoint union of two proper closed subsets.*

Proof. Left to the reader as a very useful exercise. □

Theorem 3.55. *Let $f : X \longrightarrow Y$ be a continuous function. Then, if X is connected, $f(X)$ is connected as well.*

Proof. Suppose that $Z \subset f(X)$ is a nonempty subset which is both open and closed. This means that $Z = f(X) \cap A$ for some open A, and $Z = f(X) \cap C$ for some closed C, since $f(X)$ is a subspace of Y. By continuity, $f^{-1}(Z) = f^{-1}(A)$ is open in X, and $f^{-1}(Z) = f^{-1}(C)$ is closed in X. Therefore, $f^{-1}(Z)$ is a set that is both open and closed in the connected X, which means that $f^{-1}(Z) = X$, and therefore $Z = f(X)$. Since we have showed that any nonempty subset of $f(X)$ that is both open and closed is $f(X)$ itself, it follows that $f(X)$ is connected. □

Exercise 3.56. Prove Theorem 3.55 using the characterization of Lemma 3.54.

Theorem 3.57. *Any interval $[a, b]$ of $\mathbb{R}$ is connected (in the Euclidean topology).*

Proof. We will prove it for $[0, 1]$, since the same proof works for any choice of a and b. Suppose that C, D are two nonempty closed subspaces of $[0, 1]$ with $C \cup D = [0, 1]$. We need to show that $C \cap D \neq \emptyset$, which would complete the proof by Lemma 3.54. Suppose without loss of generality that $0 \in C$, and let $d := \inf(D)$. Since D is closed, we have that $d \in D$. If $d = 0$, we have already found that $C \cap D \neq \emptyset$, and we would be done. So, we can assume that $d > 0$. In this case, set $E = C \cap [0, d]$. Since C is closed, and $[0, d) \subset E$, it follows that $d \in E$ as well, which shows that $d \in C$ and that the intersection $C \cap D \neq \emptyset$. This completes the proof. □

Corollary 3.58 (Intermediate Value Theorem)**.** *Let $f : [a, b] \longrightarrow \mathbb{R}$ be continuous. Then, for any value $y \in [\min f, \max f]$, there exists $c \in [a, b]$ such that $f(c) = y$.*

Proof. By Theorem 3.55 and Theorem 3.57 we find that the image of f, $f([a,b])$ is connected in $\mathbb{R}$. Suppose that there is a point y that violates the statement of the theorem. This means that y is not in $f([a,b])$. Let $z = \sup f([a,b]) \cap (-\infty, y)$. We have that $[\inf f, z]$ is both open and closed in $f([a,b])$ which means that the latter would be disconnected. This is not possible, and the corollary is proved. □

4. Sequences and Series

Definition 4.1. A sequence in a set X is a function $f : \mathbb{N} \longrightarrow X$. Instead of writing $f(1), f(2), \ldots$ for the elements of a sequence, it is customary to write $x_1, x_2, \ldots$ where $x_n := f(n)$. When we refer to a sequence, we also write the whole image of f as $\{x_n\}_{n\in\mathbb{N}}$, or simply $\{x_n\}_n$ or even just $\{x_n\}$.

Definition 4.2. Let $\{x_n\}$ be a sequence in a metric space (X, d). Then, we say that $\{x_n\}$ converges to $x \in X$ if for any $\epsilon > 0$, we can find $\nu \in \mathbb{N}$ such that $d(x_n, x) < \epsilon$ whenever $n > \nu$. In this case we write $\lim_{n\to\infty} x_n = x$, or $\lim_n x_n = x$, $\lim x_n = x$ or simply $x_n \longrightarrow x$.

Intuitively, the definition of convergence means that eventually (for large n) the elements x_n lie in any chosen ball around x, no matter how small the ball has been chosen.

From now on, we will discuss sequences only in metric spaces.

Definition 4.3. A sequence in X is said to be bounded if $\{x_n\}$ is bounded as a set.

Theorem 4.4. *Let $\{x_n\}$ be a sequence in the metric space X.*

(a) *$\{x_n\}$ converges to x if and only if for any ball around x, all but finitely many x_n lie in the ball.*
(b) *If $x_n \longrightarrow x$ and $x_n \longrightarrow x'$, then $x = x'$.*
(c) *If $\{x_n\}$ is convergent, then it is bounded.*
(d) *If $E \subset X$ and x is an accumulation point of E, then there is a sequence in E which is convergent to x.*

Proof. (a) Suppose that $x_n \longrightarrow x$. Then, let $B(x, \rho)$ be a ball of radius ρ around x. By definition of convergence, we find $\nu \in \mathbb{N}$ such that $x_n \in B(x, \rho)$ for all $n > \nu$. Then, only finitely many x_n,wthose ith $n \le \nu$, can be outside of the ball $B(x, \rho)$. Viceversa, consider $\epsilon > 0$ and take the ϵ-ball around x $B(x, \epsilon)$. Only finitely many x_n can lie outside of the ball. Say ν is the largest index of the x_n outside of $B(x, \epsilon)$. Clearly, for $n > \nu$ one has $x_n \in B(x, \epsilon)$. So, $x_n \longrightarrow x$.

(b) Suppose otherwise. So, x_n converges both to x and x' with $x \ne x'$. Let $\rho = d(x, x')$, and $\epsilon = \frac{\rho}{2}$. Clearly, no element y can be in $B(x, \epsilon) \cap B(x', \epsilon)$. By definition of convergence, we can find ν_1 such that $x_n \in B(x, \epsilon)$ for all $n > \nu_1$, and similarly ν_2 such that $x_n \in B(x', \epsilon)$ for all $n > \nu_2$. For $n > \max\{\nu_1, \nu_2\}$ we find that x_n is both in $B(x, \epsilon)$ and $B(x', \epsilon)$ which is a contradiction, since they have empty intersection. We must have $x = x'$.

(c) Let x be such that $x_n \longrightarrow x$. Then, we can find ν such that $x_n \in B(x, 1)$ for all $n > \nu$. This means that $x_1, \ldots, x_\nu$ are the only elements that can possibly be outside $B(x, 1)$. Let $\rho' = \max\{d(x, x_1), \ldots, d(x, x_\nu)\}$, and set $\rho = \max\{\rho', 1\}$. Then it is clear that $x_n \in B(x, \rho)$ for all n, and therefore $\{x_n\}$ is bounded.

(d) For each n, let us set $B_n := B(x, \frac{1}{n})$. Since x is an accumulation point for E, it follows that we can find $x_n \in B_n$ for all n, where $x_n \in E$. Then, given $\epsilon > 0$, we can choose ν large enough such that $\frac{1}{\nu} < \epsilon$. Therefore, $B_\nu \subset B(x, \epsilon)$. Since $B_{n+1} \subset B_n$ by construction, it follows that $x_n \in B(x, \epsilon)$ for all $n > \nu$. This shows that $x_n \longrightarrow x$. □

We next consider the special case of sequences of complex numbers and study the interaction of convergence with the algebraic operations over $\mathbb{C}$, where $\mathbb{C}$ has the Euclidean metric defined over $\mathbb{R}^2$ (this is just the modulus of a complex number).

Theorem 4.5. *Let $\{s_n\}$ and $\{t_n\}$ are sequences in $\mathbb{C}$ or $\mathbb{R}$ with $s_n \longrightarrow s$ and $t_n \longrightarrow t$. Then:*

(a) $\lim s_n + t_n = s + t$.
(b) $\lim cs_n = cs$.
(c) $\lim s_n t_n = st$.
(d) $\lim \frac{1}{s_n} = \frac{1}{s}$ *in the assumption that* $s_n, s \neq 0$ *(for all n).*

Proof. We prove only (c) in the complex case, and leave the simple proof of the other statements to the reader as an exercise.

For a given choice of $\epsilon > 0$, we can find ν_1 and ν_2 such that $|s_n - s| < \sqrt{\epsilon}$ for $n > \nu_1$, and $|t_n - t| < \sqrt{\epsilon}$ for $n > \nu_2$. For $\nu > \max\{\nu_1, \nu_2\}$, we find that $n > \nu$ implies $|(s_n - s)(t_n - s)| = |s_n - s| \cdot |t_n - t| < \epsilon$. This completes the proof of (c). The other cases are proved similarly. □

We now consider sequences of vectors in $\mathbb{R}^k$.

Theorem 4.6. *Let $\{x_n\}$ be a sequence in $\mathbb{R}^k$, where each term x_n is written as a k-tuple $x_n = (x_{1,n}, \ldots, x_{k,n})$. Then:*

(a) $x_n \longrightarrow x := (x_1, \ldots, x_n)$ *if and only if* $x_{j,n} \longrightarrow x_j$ *for all* $j = 1, \ldots, k$.
(b) *Suppose that y_n is another sequence in $\mathbb{R}^k$ and that β_n is a sequence in $\mathbb{R}$. Assume that $x_n \longrightarrow x$, $y \longrightarrow y$ and $\beta_n \longrightarrow \beta$. Then we have*

$$\begin{aligned} \lim x_n + y_n &= x + y \\ \lim x_n \cdot y_n &= x \cdot y \\ \lim \beta_n x_n &= \beta x. \end{aligned}$$

Proof. (a) Suppose $x_n \longrightarrow x$. Since we have $|x_{j,n} - x_j| \leq d(x_n, x)$ by definition of Euclidean metric, it follows that each $x_{j,n}$ converges to x_j.

Vice versa, suppose that each limit $x_{j,n} \longrightarrow x_j$ holds for all $j = 1, \ldots, k$. Then, for each $j = 1, \ldots, k$, we can find ν_j such that $|x_{j,n} - x_j| < \frac{\epsilon}{\sqrt{k}}$ whenever $n > \nu_j$. Choosing $\nu > \max\{\nu_1, \ldots, \nu_k\}$, it follows that for each $n > \nu$ we have

$$d(x_n, x) = \sqrt{\sum_{j=1}^{k} |x_{j,n} - x_j|^2} < \sqrt{\sum_{j=1}^{k} (\frac{\epsilon}{\sqrt{k}})^2} = \epsilon.$$

This shows that $x_n \longrightarrow x$.

(b) is proved either analogously to Theorem 4.5, or applying Theorem 4.5 and part (a) of the present theorem. It is left as an exercise to the reader. □

We now introduce an important notion.

Definition 4.7. Let $\{x_n\}$ be a sequence. ELet n_k be a sequence of natural numbers with the condition that $n_k < n_{k+1}$ for all k. Then, restricting n to the n_k we obtain a sequence x_{n_k} which is a called a *subsequence* of x_n. Observe that x_{n_k} is a sequence obtained by considering only some elements of x_n, those for which $n = n_k$ for some k. If $\lim_k x_{n_k} \longrightarrow x$, we say that x is a *subsequential limit* of x_n.

Remark 4.8. $x_n \longrightarrow x$ if and only if $x_{n_k} \longrightarrow x$ for every subsequence x_{n_k}. The proof of this fact is left to the reader as a very good exercise.

Theorem 4.9. *The following facts hold.*

(a) *If $\{x_n\}$ is a sequence in a compact metric space X, then some subsequence of $\{x_n\}$ converges in X.*
(b) *Every bounded subsequence in $\mathbb{R}^k$ contains a convergent subsequence.*

Proof. (a) Let E the set of elements of type x_n for some n (i.e. the range of the sequence). If E is finite, there is at least one element x which is repeated infinitely many times. So, we can construct a constant (and therefore convergent) subsequence. Suppose now that E is not finite. By Theorem 3.41 there is an accumulation point. By Theorem 4.4 part (d), we can construct a sequence of elements of $\{x_n\}$, i.e. a subsequence, converging to the accumulation point.

(b) This can be proved as a consequence of (a). In fact, being bounded, we can inscribe the sequence in a closed ball, which is compact in $\mathbb{R}^k$, as we have previously seen. Then we can use (a) to find a convergent subsequence. □

Theorem 4.10. *The subsequential limits of a sequence $\{x_n\}$ in a metric space X form a closed subset of X.*

Proof. We call E^* the set of subsequential limits of $\{x_n\}$, and we let y be a limit point of E^*. We just need to show that y is also in E^*.

We choose an arbitrary element x_{n_1} of $\{x_n\}$ which is different from y (if this is not possible there is nothing to prove: why?). We let $\delta := d(y, x_{n_1})$. We proceed inductively, and assume that we have found $n_1, \dots, n_{i-1}$. Since y is a limit point of E^*, we can find $x \in E^*$ with $d(x, q) < 2^{-i}\delta$. But $x \in E^*$, and therefore there is an $n_i > n_{i-1}$ such that $d(x, x_{x_{n_i}}) < 2^{-i}\delta$. Then, $d(y, x_{n_i}) < d(y, x) + d(x, x_{n_i}) < 2^{-i}\delta + 2^{-i}\delta = 2^{-i+1}\delta$. We obtain a subsequence $\{x_{n_i}\}_{i\in\mathbb{N}}$ which is convergent to y, showing that $y \in E^*$ and completing the proof. □

The following definition is of paramount importance in all analysis and more generally mathematics.

Definition 4.11. Let (X, d) be a metric space and let $\{x_n\}$ be a sequence in X. We say that $\{x_n\}$ is a *Cauchy sequence* if for every $\epsilon > 0$ there is an integer ν such that $d(x_n, x_m) < \epsilon$ for all $n, n > \nu$.

Definition 4.12. Let E be a nonempty subset of the metric space (X, d). Then, we define

$$S := \{d(x, y) \mid x, y \in E\} \subset \mathbb{R}_0^+.$$

We define the *diameter* of E to be $\mathrm{diam}E = \sup S$, which is infinite if the set E is unbounded, and it is a number if E is bounded.

Exercise 4.13. Show the following two facts.

- $\mathrm{diam}E$ is finite if and only if E is bounded.
- Let $\{x_n\}$ be a sequence and let E_ν consist of all x_n with $n \geq \nu$. Then, $\{x_n\}$ is Cauchy if and only if $\lim_\nu \mathrm{diam}E_\nu = 0$.

Theorem 4.14. *The following facts hold.*

(1) $\mathrm{diam}E = \mathrm{diam}\bar{E}$*, where $\bar{E}$ is the closure of E.*
(2) *Let K_n be a sequence of compact sets in X such that $K_{n+1} \subset K_n$ for all n. Then, if $\lim_n \mathrm{diam}K_n = 0$, $\bigcap_n K_n$ consists of a single point.*

Proof. (1) Since $E \subset \bar{E}$, it follows immediately that $\mathrm{diam}E \leq \mathrm{diam}\bar{E}$. For fixed $\epsilon > 0$ and given points $\bar{E}$, say x and y, we can find elements of E, x' and y', satisfying $d(x, x') = d(y, y') < \epsilon$.

It follows that $d(x,y) \leq d(x,x') + d(x',y') + d(y',x) < 2\epsilon + d(x',y') \leq 2\epsilon + \text{diam}E$. Therefore, $\text{diam}\bar{E} \leq 2\epsilon + \text{diam}E$ for all $\epsilon > 0$. This means that $\text{diam}\bar{E} \leq \text{diam}E$, which completes the proof of (1).

(2) Let $K = \bigcap_n K_n$, which we know to be nonempty, applying Theorem 3.39. If K is not just a singleton, then obviously $\text{diam}K > 0$. Since $K \subset K_n$ for all n, it follows that $\text{diam}K \leq \text{diam}K_n$ for all n. It therefore cannot happen that $\lim_n \text{diam}K_n = 0$. This contradiction proves the result. □

Theorem 4.15. *The following facts hold.*

1. *In a metric space X, any convergent sequence is a Cauchy sequence.*
2. *If X is compact, and $\{x_n\}$ is a Cauchy sequence in X, then $\{x_n\}$ converges in X.*
3. *In $\mathbb{R}^k$, every Cauchy sequence converges.*

Proof. (1) Suppose that $x_n \longrightarrow x$. For fixed $\epsilon > 0$, by definition of sequence we can find ν such that $d(x,x_n) < \frac{\epsilon}{2}$ whenever $n > \nu$. Therefore, we have that for all $n, m > \nu$

$$\begin{aligned} d(x_n, x_m) &\leq d(x_n, x) + d(x, x_m) \\ &< \frac{\epsilon}{2} + \frac{\epsilon}{2} \\ &= \epsilon. \end{aligned}$$

This shows that $\{x_n\}$ is Cauchy.

(2) Let $\{x_n\}$ be Cauchy in the compact X. Using Exercise 4.13, we know that being Cauchy means that $\text{diam}E_n \longrightarrow 0$. Using Theorem 4.14 we also find also that $\text{diam}\bar{E}_n \longrightarrow 0$. Since closed subsets of compact spaces are compact, it follows that $\bar{E}_n$ is compact for all n, and we clearly have that $\bar{E}_n \subset \bar{E}_{n+1}$ as well. Therefore, using Theorem 4.14 we find that $\bigcap_n \bar{E}_n = \{x\}$. We want to show now that x is the limit of x_n, i.e. $x_n \longrightarrow x$. Let $\epsilon > 0$ be given arbitrarily. Since $\text{diam}\bar{E}_n \longrightarrow 0$, we can find ν such that $\text{diam}\bar{E}_n < \epsilon$ for all $n > \nu$. Since $x \in E_n$ for all n, whenever $n > \nu$ we have that $d(x,x_n) < \text{diam}E_\nu < \epsilon$, which shows convergence of x_n to x.

(3) Let $\{x_n\}$ be a Cauchy sequence in $\mathbb{R}^k$. We can find n such that $\text{diam}E_n < 1$ since the diameters of the sets E_n converge to zero. Then, consider the elements $x_1, \ldots, x_n$, and consider a ball centered at zero whose radius ρ is larger than $d(0,x_1), \ldots, d(0,x_n)$ and $d(0,x_{\nu+1}) + 1$. The set $\bar{B}(0,\rho)$ is closed an bounded, hence compact, and contains the sequence $\{x_n\}$. Using (2) the sequence converges in $\bar{B}(0,\rho)$, completing the proof. □

Corollary 4.16. *In $\mathbb{R}^k$ we have that a sequence converges if and only if it is a Cauchy sequence.*

Exercise 4.17. Prove the corollary, using the theorem that precedes it.

Definition 4.18. A metric space in which every Cauchy sequence converges is said to be *complete*.

Remark 4.19. As an application of Theorem 4.15, we have that all compact metric spaces and the Euclidean spaces are complete.

We now focus on sequences of real numbers.

Definition 4.20. A sequence $\{x_n\}$ of real numbers is said to be

- *Monotonically increasing* if $x_n \leq x_{n+1}$ for all n.
- *Monotonically decreaasing* is $x_n \geq x_{n+1}$ for all n.

If the inequalities hold starting from some value $\nu \in \mathbb{N}$, e.g. $x_n \leq x_{n+1}$ for all $n \geq \nu$, then we say that the sequence is *eventually monotonic* (increasing or decreasing depending on the inequality). Sometimes we omit saying eventuallty altogether.

Theorem 4.21. *Suppose $\{x_n\}$ is monotonic. Then, the sequence converges if and only if it is bounded.*

Proof. For the sake of simplicity, we consider the case of monotonically increasing sequences. The other case is proved following the same procedure and reversing the inequalities.

We first show that if the sequence is bounded, then it converges. Since x_n takes bounded values in $\mathbb{R}$, there is a finite upper bound $y := \sup_n\{x_n\}$. We want to show that y is the limit. For a choice of $\epsilon > 0$, from the properties of suprema we have that $y - \epsilon$ is not an upper bound, and therefore we can find $\nu \in \mathbb{N}$ such that $x_\nu > y - \epsilon$. Since $\{x_n\}$ is increasing, for all $n > \nu$ we also have $x_n > y - \epsilon$. Since y is upper bound, it must also hold that $x_n \leq y < y + \epsilon$ for all n. So, for $n > \nu$ we have that $x_n \in (y - \epsilon, y + \epsilon)$. Therefore, $x_n \longrightarrow y$.

The converse was already proved in Theorem 4.4 (c), since we have shown that any convergent sequence (whether is monotinic or not) is bounded. □

Definition 4.22. Let $\{x_n\}$ be a sequence of real numbers, and suppose that for any choice of $M \in \mathbb{R}$ there is an index ν such that $x_n \geq M$ for all $n > \nu$. Then we say that $\{x_n\}$ diverges at $+\infty$ and we write $x_n \longrightarrow +\infty$.

Similarly, if for any $m \in \mathbb{R}$ we can find ν such that $x_n < m$ whenever $n > \nu$ we say that $\{x_n\}$ diverges at $-\infty$ and write $x_n \longrightarrow -\infty$.

Definition 4.23. Let $\{x_n\}$ we a sequence of real numbers. Let E denote the subset of elements x in $\mathbb{R} \bigcup \{-\infty, +\infty\}$ such that $x_{n_k} \longrightarrow x$ for some subsequence of x_n. This is the set of subsequential limits with the possible addition of $-\infty$ and $+\infty$. Then, we set

- $\limsup x_n := \sup E$, and call it the *upper limit* (or supremum limit).
- $\liminf x_n := \inf E$, and call it the *lower limit* (or infimum limit).

Theorem 4.24. *Let $\{x_n\}$ be a sequence of real numbers. Let E be defined as above. Then,* $\limsup x_n$ *has the following two properties:*

(a) $\limsup x_n \in E$.

(b) *If $x > \limsup x_n$ there is an integer ν such that $x_n < x$ for all $n > \nu$.*

Moreover, $\limsup$ *is the only number which satisfies both properties (a) and (b). Similar facts also hold for* $\liminf$.

Proof. (a) The case when $\limsup x_n$ is finite was already proved, since we have shown that the set of subsequential limits is closed, and that closed sets contain their supremum.

If $\limsup x_n = +\infty$, then E is not bounded above, which also implies (why?) that x_n is not bounded above. Therefore, we can construct a subsequence x_{n_k} such that $x_{n_k} \longrightarrow +\infty$, so that $+\infty \in E$.

If $\limsup x_n = -\infty$ then E does not contain any finite element. For any choice of $m > 0$, only finitely many x_n can satisfy the property that $x_n > m$. So, $x_n \longrightarrow -\infty$ and $-\infty \in E$.

(b) Suppose by way of contradiction that the exists a number x such that $x_n > x$ for infinitely many n. Then, we can construct a subsequence of x_n convergent to some $y \geq x > \limsup x_n$. Such y would be an element of E which is larger than $\limsup x_n$ (the supremum of E by definition). This contradiction shows that such x cannot exist.

We now prove uniqueness. We proceed again by contradiction. Let p and q be two numbers satisfying properties (a) and (b) with $p < q$. We pick x such that $p < x < q$. Then, by (b) applied to p we find that there are no infinitely many x_n's above x. This means that (a) cannot hold for q. This contradiction completes the proof. □

Theorem 4.25. *The following limits hold.*

(a) *If* $p > 0$, $\lim \frac{1}{n^p} = 0$.
(b) *If* $p > 0$, *then* $\lim \sqrt[n]{p} = 1$.
(c) $\lim \sqrt[n]{n} = 1$.
(d) *If* $p > 0$ *and* α *is real, then* $\lim \frac{n^\alpha}{(1+p)^n} = 0$.
(e) *If* $|x| < 1$, *then* $\lim x^n = 0$.
(f)

$$\lim \frac{a_\alpha n^\alpha + a_{\alpha-1}n^{\alpha-1} + \cdots + a_0}{b_\beta n^\beta + b_{\beta-1}n^{\beta-1} + \cdots + b_0} = \begin{cases} \frac{a_\alpha}{b_\beta} & \text{if} \quad \alpha = \beta \\ 0 \ \text{if} & \alpha < \beta \\ +\infty & \text{if} \quad \alpha > \beta \end{cases}.$$

Proof. This is done by direct investigation, using the definition. We will not cover this proof in detail. □

We will now consider the importnat notion of series. We will focus on series of complex numbers, unless otherwise stated.

Definition 4.26. Given a sequence $\{x_n\}$, we will associate to it another sequence, obtained by summing all elements of x_n up to some fixed value:

$$s_k = \sum_{n=0}^{k} x_n. \tag{10}$$

For $\{x_n\}$ we also use the symbolic expression $\sum_{n=0}^\infty x_n$ to indicate the sum of all the elements x_n. We call such an object *series*, and the elements s_n *partial sums*.

If s_n is a convergent sequence, i.e. it converges to s, then we set $\sum_{n=0}^\infty x_n = s$. In this case we say that the series is convergent. If s_n is divergent, we say that the series is divergent.

Remark 4.27. Depending on the definition of x_n and context, we can consider series starting from a different number rather than 0 or 1. So, something like $\sum_{n=p}^\infty x_n$ are also allowed, where p is some natural number larger than 1. In some cases, we will just write $\sum x_n$ when we are not interested in the first element of the series. After all, the behavior of a series is not affected by few (finitely many) elements at the beginning of it, and our main question (does it converge or diverege?) does not depend on the starting point. Observe, however, that removing some elements from a series affects the precise value the series converges to.

We have the following criterion of convergence of a series, which is just a reformulation of the criterion of convergence of sequences we have found before, adapted to the case of series.

Theorem 4.28 (Cauchy Criterion)**.** *The series* $\sum x_n$ *converges if and only if for every* $\epsilon > 0$ *there is an integer* ν *such that*

$$|\sum_{k=n}^{m} x_k| < \epsilon,$$

whenever $m \geq n > \nu$. *In particular, if* $\sum x_n$ *converges, then* $\lim x_n = 0$.

Proof. This is a consequence of the Cauchy Criterion of convergence of sequences applied to s_n. □

Remark 4.29 (Warning)**.** Observe that the first part of the thoerem is an if and only if, while the second statement is an single implication. So, it is not generally true that if $\lim x_n = 0$ the series converges.

Theorem 4.30. *Let x_n be a series of real nonnegative numbers. Then $\sum x_n$ converges if and only if the partial sums form a bounded series.*

Proof. This follows from the fact that since the elements x_n are positive, the partial sums s_n constitute a monotonically increasing sequence. Then Theorem 4.21 completes the result. □

The following test for convergence is very useful.

Theorem 4.31 (Comparison Test). (a) *Suppose $|x_n| \leq y_n$ (real y_n) for all $n \geq \nu_0$ for some fixed natural number ν_0, and $\sum y_n$ is convergent. Then $\sum x_n$ is convergent as well.*
(b) *Let $x_n \geq y_n \geq 0$ for all $n \geq \nu_0$ for some ν_0. Then, if $\sum y_n$ diverges, the series $\sum x_n$ diverges as well.*

Proof. (a) Given $\epsilon > 0$ we can find $\nu \in \mathbb{N}$ such that $m \geq n > \nu$ implies $\sum_{k=n}^{m} y_k < \epsilon$ by the Cauchy Criterion. Then, we also have

$$\begin{aligned} |\sum_{k=n}^{m} x_n| &\leq \sum_{k=n}^{m} |x_k| \\ &\leq \sum_{k=n}^{m} y_k \\ &< \epsilon, \end{aligned}$$

and by the Cauchy Criterion $\sum x_n$ is also a convergent series.

(b) This follows from (a) because if $\sum x_n$ is convergent, then by (a) $\sum y_n$ is convergent as well, which is a contradiction. Alternatively, one can argue that $\lim x_n \neq 0$, which is not possible by the Cauchy Criterion. □

We will now focus on a fundamental class of series. Namely, the series of nonnegative terms.

Theorem 4.32 (Geometric Series). *Let $0 \leq x < 1$. Then*

$$\sum_{n=0}^{\infty} x^n = \frac{1}{1-x}. \tag{11}$$

If $x \geq 1$ the series diverges.

Proof. The case $x \geq 1$ is obvious since $\lim x^n \neq 0$, so the Cauchy Criterion gives divergence.

When $0 \leq x < 1$, from $1 + x + \cdots + x^n = (1-x)^{-1}(1-x^{n+1})$ we have that

$$s_n = \sum_{k=0}^{n} x^k = \frac{1-x^{n+1}}{1-x}.$$

Taking the limit as $n \longrightarrow \infty$, and considering that $x^n \longrightarrow 0$ since $0 \leq x < 1$, we obtain the result. □

Theorem 4.33. *Let x_n be nonnegative and monotonically descreasing. Then the series $\sum_{n=1}^{\infty} x_n$ converges if and only if the series $\sum_{k=0}^{\infty} 2^k x_{2^k}$ converges.*

Proof. Since this is a sequence of nonnegative terms, it is enough to consider boundedness of partial sums. We let

$$s_n = x_1 + \cdots x_n$$
$$t_k = x_1 + 2x_2 + \cdots + 2^k x_{2^k}.$$

For $n < 2^k$ we have that

$$\begin{aligned} s_n &\leq x_1 + (x_2 + x_3) + \cdots + (x_{2^k} + \cdots + x_{2^{k+1}-1}) \\ &\leq x_1 + 2x_2 + \cdots + 2^k x_{2^k} \\ &= t_k. \end{aligned}$$

Therefore, $s_n \leq t_k$. Also, if $n > 2^k$, we have

$$\begin{aligned} s_n &\geq x_1 + x_2 + (x_3 + x_4) \cdots + (x_{2^{k-1}+1} + \cdots + x_{2^k}) \\ &\geq \frac{1}{2}x_1 + x_2 + 2x_4 + \cdots 2^{k-1} x_{2^k} \\ &= \frac{1}{2}t_k. \end{aligned}$$

Therefore, $2s_n \geq t_k$.

It follows that the series are either both bounded (and therefore convergent), or both unbounded, and therefore divergent. The proof is complete. □

Theorem 4.34. *The series $\sum \frac{1}{n^p}$ converges if $p > 1$, and diverges if $p \leq 1$.*

Proof. For $p \leq 0$ the series obviously diverges, since the sequence $\frac{1}{n^p}$ does not converge to zero. For $p > 0$ we can apply Theorem 4.33. So, the convergence of $\sum \frac{1}{n^p}$ is equivalent to the convergence of the series

$$\sum_{k=0}^{\infty} 2^k \frac{1}{2^{kp}} = \sum_{k=0}^{\infty} 2^{(1-p)k}.$$

If $2^{1-p} < 1$, i.e. if $1 - p < 0$, then $\sum_{k=0}^{\infty} 2^{(1-p)k}$ is a geometric series with $x = 2^{1-p} < 1$, and it is therefore convergent. If $2^{1-p} \geq 1$, i.e. $1 - p \geq 0$ using again the geometric series with $x = 2^{1-p} \geq 1$ we find that the series is divergent. This completes the proof. □

The following result follows from Theorem 4.33 as well.

Theorem 4.35. *The series*

$$\sum_{n=2}^{\infty} \frac{1}{n(\ln n)^p}$$

converges when $p > 1$, and diverges when $p \leq 1$.

Proof. The function $f(x) = \frac{1}{x \ln x}$ decreases, and therefore we can use Theorem 4.33. We therefore consider the convergence/divergence of the series

$$\begin{aligned} \sum_{k=1}^{\infty} \frac{2^k}{2^k (\ln 2^k)^p} &= \sum_{k=1}^{\infty} \frac{1}{(k \ln 2)^p} \\ &= \frac{1}{(\ln 2)^p} \sum_{k=1}^{\infty} \frac{1}{k^p}. \end{aligned}$$

This completes the proof using Theorem 4.34. □

The previous result, technically speaking, uses the logarithm with base e which we have not yet defined. Let us now solve this issue.

Definition 4.36. We define the *Napier constant* (or also *Euler number*), written e, as

$$e := \sum_{n=0}^{\infty} \frac{1}{n!},$$

where the factorial, $n!$, is defined inductively as $0! = 1$, $n! = n \cdot (n-1)!$, for all $n \in \mathbb{N}$.

Exercise 4.37. Show that the series $\sum_{n=0}^{\infty} \frac{1}{n!}$ converges, and that the definition therefore makes sense.

Theorem 4.38. *We have*

$$\lim(1 + \frac{1}{n})^n = e. \tag{12}$$

Proof. We set $s_n = \sum_{k=0}^{n} \frac{1}{k!}$, and $t_n = (1 + \frac{1}{n})^n$.

The binomial theorem gives us

$$\begin{aligned}(1 + \frac{1}{n})^n &= \sum_{q=0}^{n} \binom{n}{q} 1^{n-q} (\frac{1}{n})^q \\ &= 1 + 1 + \frac{n!}{2!(n-2)!}(\frac{1}{n})^2 + \cdots + \frac{n!}{(n-1)!(n-(n-1))!}(\frac{1}{n})^{n-1} + (\frac{1}{n})^n \\ &= 1 + 1 + \frac{1}{2!}(1 - \frac{1}{n}) + \cdots + \frac{1}{n!}(1 - \frac{1}{n})(1 - \frac{2}{n}) \cdots (1 - \frac{n-1}{n}).\end{aligned}$$

It therefore follows that $t_n \le s_n$, which in turn implies that $\limsup t_n \le e$ (why?).

Suppose now that $n \ge m$. Then we have

$$t_n \ge 1 + 1 + \frac{1}{2!}(1 - \frac{1}{n}) + \cdots + \frac{1}{m!}(1 - \frac{1}{n})(1 - \frac{2}{n}) \cdots (1 - \frac{m-1}{n}),$$

and letting $n \longrightarrow \infty$, but keeping m fixed, we obtain $\liminf t_n \ge 1 + 1 + \frac{1}{2!} + \cdots + \frac{1}{m!}$. Therefore, we have obtained that $s_m \le \liminf t_n$. Taking the limit $m \longrightarrow \infty$ we find that $e \le \liminf t_n$. So, $e = \lim t_n$, which completes the proof. □

We now show that e is an irrational number.

Theorem 4.39. *The Napier's constant is irrational.*

Proof. We have

$$\begin{aligned}e - s_n &= \frac{1}{(n+1)!} + \frac{1}{(n+2)!} + \cdots \\ &< \frac{1}{(n+1)!}[1 + \frac{1}{n+1} + \frac{1}{(n+1)^2} + \cdots] \\ &= \frac{1}{n!n},\end{aligned}$$

where in the last equality we have used the geometric series sum with $x = \frac{1}{n+1}$ for the term in square brackets.

Suppose that e is rational. Then $e = \frac{p}{q}$ for some $p, q \in \mathbb{N}$. Using the above inequality we therefore have

$$0 < q!(e - s_q) < \frac{1}{q}.$$

Clearly, $q!(e - s_q)$ is an integer, since we assumed that $e = \frac{p}{q}$. Since $q \ge 1$, we have therefore found an integer between 0 and 1, which is nonsense. This contradiction completes the proof. □

We now discuss two very important tests for convergence of series.

Theorem 4.40 (Root Test)**.** *Given* $\sum x_n$*, define* $\alpha = \limsup \sqrt[n]{|x_n|}$*. Then*

(a) *If* $\alpha < 1$*,* $\sum x_n$ *converges;*
(b) *If* $\alpha > 1$*,* $\sum x_n$ *diverges;*
(c) *If* $\alpha = 1$ *the series can be either divergent or convergent (no information).*

Proof. We prove (a) first. If $\alpha < 1$, we can choose $\beta < 1$ such that $\alpha < \beta < 1$. In correspondence to such β we can find an integer ν such that whenever $n > \nu$, $\sqrt[n]{|x_n|} < \beta$, by Theorem 4.24. Therefore, for $n > \nu$ we have $|x_n| < \beta^n$. Since $\beta < 1$, the series $\sum \beta^n$ is convergent (it is a geometric series!), and $\sum x_n$ converges as well by the comparison thest, i.e. Theorem 4.31.

Next, we prove (b). If $\alpha > 1$, again by Theorem 4.24 we hav a subsequence x_{n_k} such that $\sqrt[n_k]{|x_{n_k}|} \longrightarrow \alpha$. Consequently, $|x_n| > 1$ for infinitely many values of n, which means that $x_n \longrightarrow 0$ cannot hold. The second statement of the Cauchy criterion (Theorem 4.28) gives the divergence of the series $\sum x_n$.

Lastly, to prove (c), consider the series $\sum \frac{1}{n}$ and $\sum \frac{1}{n^2}$. In both cases one can see that $\alpha = 1$. But from Theorem 4.34 we know that in one case the series diverges, and in the other case it converges. □

Theorem 4.41 (Ratio Test)**.** *The series* $\sum x_n$ *is*

(a) *Convergent, if* $\limsup |\frac{x_{n+1}}{x_n}| < 1$*;*
(b) *Divergent, if* $|\frac{x_{n+1}}{x_n}| \geq 1$ *for all* $n \geq \nu$ *for some* $\nu \in \mathbb{N}$*.*

Proof. We prove (a) first. We can find $\beta < 1$ and $\nu \in \mathbb{N}$ such that

$$|\frac{x_{n+1}}{x_n}| < \beta,$$

for all $n \geq \nu$. Therefore, for any $p > 0$, we have

$$\begin{aligned} |x_{\nu+1}| &< \beta|x_\nu| \\ |x_{\nu+2}| &< \beta|x_{\nu+1}| < \beta^2|x_\nu| \\ &\vdots \\ |x_{\nu+p}| &< \beta^p|x_\nu|. \end{aligned}$$

Therefore, $|x_n| < \beta^n|x_\nu|$ for all $n > \nu$. The comparison test (with the geometric series) shows convergence of $\sum x_n$.

To prove (b), one simply notes that x_n cannot converge to zero because from the inequality it follows that $|x_n| \geq |x_\nu| > 0$ for infinitely many terms. note that $x_\nu \neq 0$ since otherwise the ratio $x_{\nu+1}/x_\nu$ would not be defined, but by assumption it is defined. □

The following result clarifies the relation between the ratio and root tests.

Theorem 4.42. *For any sequence* $\{x_n\}$ *of positive numbers, it holds*

(a) $\liminf \frac{x_{n+1}}{x_n} \leq \liminf \sqrt[n]{c_n}$.
(b) $\limsup \sqrt[n]{x_n} \leq \limsup \frac{x_{n+1}}{x_n}$.

Proof. We prove only (b), and leave (a) to the reader.

We put $\alpha = \limsup \frac{x_{n+1}}{x_n}$. If $\alpha = \infty$, then there is nothing to prove. So, we can assume that α is a finite number. We choose $\beta > \alpha$, and take an integer ν such that $n \geq \nu$ implies

$$\frac{x_{n+1}}{x_n} \leq \beta.$$

Therefore, for any $p > 0$, we have $x_{\nu+p+1} < \beta x_{\nu+p}$. So, proceeding in a similar way as in the proof of the ratio test we find that $x_{\nu+p} < \beta^p x_\nu$ for all $p > 0$. This inequality can also be rewritten as $x_n \le x_\nu \beta^{-\nu} \beta^n$ for all $n \ge \nu$. Taking n roots of both sides we find

$$\sqrt[n]{x_n} \le \sqrt[n]{x_\nu \beta^{-\nu}} \cdot \beta,$$

which implies that $\limsup \sqrt[n]{x_n} \le \beta$ (since $\lim \sqrt[n]{x} = 1$ for all positive numbers x). Since we have proved that $\limsup \sqrt[n]{x_n} \le \beta$ for all $\beta > \alpha$, we find that $\limsup \sqrt[n]{x_n} \le \alpha$, which completes the proof. □

Definition 4.43 (Power series)**.** Given a sequence $\{c_n\}$ of complex numbers, the series

$$\sum_{n=0}^{\infty} c_n z^n,$$

where z is a complex variable, is called a power series. Depending on z, the series might converge or diverge. On those values of z such that the series is convergent, the power series defines a complex function.

We will now see that associated to a given series, there is a radius of convergence R, i.e. a number such that inside the ball $B(0, R)$ the series converges, and it diverges outside of it. In this case, we also allow $R = \infty$, in which case $B(0, \infty)$ is set to be the whole complex plane. We also allow $R = 0$, and $B(0, 0)$ would just be $\{0\}$. On the boundary of the ball, i.e. on the circle, the convergence is not simple to determine, and should be analyzed on a case by case basis.

Theorem 4.44. *Given a power series $\sum c_n z^n$, set $\alpha := \limsup \sqrt[n]{|c_n|}$ and $R = \frac{1}{\alpha}$ (with the convention that $0 = \frac{1}{\infty}$ and $\infty = \frac{1}{0^+}$). Then, $\sum c_n z^n$ converges if $|z| < R$ and diverges if $|z| > R$.*

Proof. Setting $a_n = c_n z^n$, we find

$$\begin{aligned}\limsup \sqrt[n]{|a_n|} &= \limsup \sqrt[n]{|c_n| \cdot |z^n|} \\ &= |z| \limsup \sqrt[n]{|c_n|} \\ &= \frac{|z|}{R}\end{aligned}$$

. By the root test applied to a_n, we obtain that the original series is convergent of $|z| < R$ and it is divergent if $|z| > R$. □

Definition 4.45. The value R of Theorem 4.44 is called *radius of convergence* of the power series $\sum c_n z^n$.

We now investigate a class of series of fundamental importance. Namely, the absolute convergent series.

Definition 4.46. A series $\sum a_n$ is said to be *absolutely convergent* if the series $\sum |a_n|$ converges. In this situation we also say that $\sum a_n$ *converges absolutely.*

Theorem 4.47. *If $\sum a_n$ converges absolutely, then it converges.*

Proof. We apply the Cauchy criterion. In fact, if $\sum |a_n|$ converges, i.e. $\sum a_n$ converges absolutely, then we can find ν such that for all $n, m > \nu$ we have $|\sum_{k=n}^{m} |a_k|| < \epsilon$. In correspondence of such

ν it also holds

$$\begin{aligned} |\sum_{k=n}^{m} a_k| &\leq \sum_{k=n}^{m} |a_k| \\ &= |\sum_{k=n}^{m} |a_k|| \\ &< \epsilon. \end{aligned}$$

Therefore, $\sum a_n$ is Cauchy. □

Remark 4.48. While absolute convergence and convergence coincide for a sequence of positive terms, this is not the case for sequences whose terms are not all positive. For instance, the series $\sum \frac{(-1)^n}{n}$ converges but does not converge absolutely. In this case we say also that the series *converges nonabsolutely.*

Exercise 4.49. Let the sum of two series be defined as $\sum a_n + \sum b_n := \sum(a_n + b_n)$. Show that if the series are convergent, say to A and B, then their sum converges to $A + B$.

Definition 4.50. Given two series $\sum a_n$ and $\sum b_n$, we define their product as $\sum c_n$ where $c_n = \sum_{k=0}^{n} a_k b_{n-k}$.

Theorem 4.51. *Suppose that $\sum a_n$ converges absolutely, $\sum a_n = A$ and $\sum b_n = B$. Then their product series converges as well. Moreover, $\sum c_n = AB$, where $\sum c_n$ is the product series.*

5. Continuous Functions

We begin with the definition of limit in metric spaces.

Definition 5.1. Let $f : E \longrightarrow Y$, where $E \subset X$, and $X$and Y are metric spaces with matrics d_X and d_Y, respectively. Assume that p is an accumulation point of E. We define the limit of $f(x)$, as x goes to p, written

$$\lim_{x \to p} f(x) = q,$$

if for all $\epsilon > 0$, there is a $\delta > 0$ such that

$$d_Y(f(x), q) < \epsilon,$$

whenever $d_X(x, p) < \delta$, with $x \in E$ and $x \neq p$.

Observe that p needs to be an accumulation point, but does not need to be itself a point of E.

Theorem 5.2. *Let X, Y, E, f and p be as above. Then,*

$$\lim_{x \to p} f(x) = q \tag{13}$$

if and only if for every sequence $\{x_n\} \subset E$ with $x_n \longrightarrow p$ and $x_n \neq p$ for all n we have

$$f(x_n) \longrightarrow q. \tag{14}$$

Proof. Suppose first that the limit (13) holds. For a choice of $\epsilon > 0$, we can find $\delta > 0$ such that for all $x \in E$ with $d_X(x, p) < \delta$ we have $d_Y(f(x), q) < \epsilon$. In correspondence to such δ, we can find $\nu \in \mathbb{N}$ such that $d_X(x_n, p) < \delta$ whenever $n > \nu$, since $x_n \longrightarrow p$. Then, for all $n > \nu$, $d_Y(f(x_n), q) < \epsilon$. Since ϵ is arbitrary, we have shown that $f(x_n) \longrightarrow q$.

Viceversa, suppose that $f(x_n) \longrightarrow q$ whenever $x_n \longrightarrow p$ with $x_n \neq p$. By way of contradiction, we assume that (13) does not hold. Then, we can find an $\epsilon > 0$ such that for each $\delta > 0$ we can find an $x_\delta \in E$, with $x_\delta \neq p$ and $d_X(f(x_\delta), q) \geq \epsilon$. Define $\delta_n = \frac{1}{n}$. The corresponding x_n, associated to δ_n as above, will satisfy the properties that $x_n \longrightarrow p$, $x_n \neq p$ for all n, $x_n \in E$, and $d_Y(f(x_n), q) \geq \epsilon$. The last fact, in particular, shows that $f(x_n) \longrightarrow q$ does not hold, contradicting (14). This completes the proof. □

We have previously defined continuity as the property of have open preimage of open sets. We now reformulate this fact in terms of metrics for metric spaces.

Theorem 5.3. *Let X and Y be metric spaces. Then, $f : X \longrightarrow Y$ is continuous if and only if for all open balls $B(f(x), \epsilon)$ with $f(x)$ in the image of f, there exists a ball $B(x, \delta)$ such that $f(B(x, \delta)) \subset B(f(x), \epsilon)$.*

Proof. This follows from the fact that open sets of metric spaces are those sets that are union of open balls.

Suppose that f is continuous. Then, since $B(f(x), \epsilon)$ is open, the preimage of it $f^{-1}(B(f(x), \epsilon))$ is open as well. Therefore, every point in the preimage is an interior point. In particular, x is an interior point. This means that we can find a ball $B(x, \delta) \subset f^{-1}(B(f(x), \epsilon))$. Taking images, we find $f(B(x, \delta)) \subset B(f(x), \epsilon)$.

Viceversa, if V is an open set in Y, let $f^{-1}(V)$ be the preimage of it. We want to show that any point $x \in f^{-1}(V)$ is interior to it. This means that we want to show that we can find a ball $B(x, \delta) \subset f^{-1}(V)$ or, in other words, $f(B(x, \delta)) \subset V$. Since V is open, we can find a ball $B(f(x), \epsilon) \subset V$. Choosing δ such that $f(B(x, \delta)) \subset B(f(x), \epsilon)$ therefore completes the proof. □

Definition 5.4. We say that a function is continuous at a point p if for every $\epsilon > 0$ we can find $\delta > 0$ such that $d_X(x, p) < \delta$ implies $d_Y(f(x), f(p)) < \epsilon$.

Corollary 5.5. *$f : X \longrightarrow Y$ is continuos if and only if it is continuous at every point of its domain.*

Proof. Using the reformulation of Theorem 5.3 continuity is the same as being continuous at every point. □

Remark 5.6. Observe that while taking limits does not require that the point be part of the domain (only accumulation point), being continuous does indeed require being part of the domain. This, in particular, implies that the function $f(x) = \frac{1}{x}$ is continuous. However, the function defined as

$$f(x) = \begin{cases} \frac{1}{x} & x \neq 0 \\ 0 & x = 0 \end{cases}$$

is not continuous.

Question 5.7. What happens if the domain of f has an isolated point?

Theorem 5.8. *Let $f : X \longrightarrow Y$ be a continuous function which is injective and surjective, where X is compact. Then, the inverse function $f^{-1} : Y \longrightarrow X$ is continuous.*

Proof. We just need to show that for any open set $U \subset X$, $(f^{-1})^{-1}(U) = f(U)$ is open in Y. This is equivalent to showing that $f(C)$ is closed in Y for any closed set $C \subset X$, since this is an equivalent formulation of continuity. If C is closed in X, then C is compact. Its image $f(C)$ is the image of a compact through a continuous map, therefore compact. But compacts in a metric space are closed. This completes the proof. □

Definition 5.9. We say that a map between two metric spaces X and Y is *uniformly continuous* if for any $\epsilon > 0$ we can find a $\delta > 0$ such that $d_Y(f(x), f(y)) < \epsilon$ whenever $d_X(x, y) < \delta$.

Remark 5.10. Clearly, the a uniformly continuous function is also continuous. However, a continuous function is not necessarily continuous. The reason for this is that the value δ that allows one to have $d_(x, y) < \epsilon$ generally depends on the point x considered. If we can choose a δ that works for all x in the domain, a function would be uniformly continuous, but this is not necessarily possible. Consider for instance the function $f(x) = \frac{1}{x}$. This is continuous, but not uniformly continuous.

When a function is defined over a compact, uniform continuity and continuity coincide, as the following result shows.

Theorem 5.11 (Heine-Cantor)**.** *Let $f : X \longrightarrow Y$ be a continuous map defined over a compact X. Then, X is uniformly continuous.*

Proof. For an arbitrary choice of $\epsilon > 0$, for any $x \in X$ we can find $\delta_x > 0$ such that $d_X(x, z) < \delta_x$ implies $d_Y(f(x), f(z)) < \frac{\epsilon}{2}$, since f is continuous at x for all $x \in X$. Define the balls $B(x, \frac{\delta_x}{2})$. It is clear that $X = \bigcup_{x\in X} B(x, \frac{\delta_x}{2})$, and since X is compact, we can find finitely many points $x_1, \dots, x_n$ such that $X = \bigcup_{i=1}^{n} B(x_i, \frac{\delta_i}{2})$, where we have set $\delta_i = \delta_{x_i}$. We choose

$$0 < \delta < \frac{1}{2}\min\{\delta_1, \dots, \delta_n\}.$$

We now show that δ defined above satisfies the requirement for uniform continuity. Let $x, y \in X$ satisfy $d_X(x, y) < \delta$. Since the balls $B(x_i, \frac{\delta_i}{2})$ cover X, we can find a k, $1 \leq k \leq n$, such that $x \in B(x_k, \frac{\delta_k}{2})$. Therefore, $d_X(x, x_k) < \frac{\delta_k}{2}$. Moreover, we have

$$\begin{aligned} d_X(x_k, y) &\leq d_X(x_k, x) + d_X(x, y) \\ &\leq \frac{\delta_k}{2} + \frac{\delta}{2} \\ &< \frac{\delta_k}{2} + \frac{\delta_k}{2} \\ &= \delta_k. \end{aligned}$$

Therefore, by the choice of δ_k, we have

$$\begin{aligned} d_Y(f(x), f(y)) &\leq d_Y(f(x), f(x_k)) + d_Y(f(x_k), f(y)) \\ &< \frac{\epsilon}{2} + \frac{\epsilon}{2} \\ &= \epsilon. \end{aligned}$$

The proof is complete. □

Definition 5.12. Let f be a function befined on $(a, b) \subset \mathbb{R}$. Let $t \in [a, b)$. Then, we define $f(t^+) = q$ if $f(x_n) \longrightarrow q$ for all sequences $\{x_n\}$ in (t, b) with $x_n \longrightarrow t$. Similarly one defines $f(t^-)$.

Remark 5.13. We have that $\lim_{x\to t} f(x)$ exists if an only if $f(t^-) = f(t+)$. In this case, $f(t^-) = f(t+) = \lim_{x\to t} f(x)$.

Definition 5.14. If f is discontinuous at a point t, but $f(t^-)$ and $f(t^+)$ both exist, then f is said to have a *discontinuity of the first kind*. Otherwise, we say that f has a *discontinuity of the second kind* if f is discontinuous at t, and at least one between $f(t^-)$ and $f(t^+)$ does not exist.

Definition 5.15. Let f be a real function defined on the interval (a,b). Then, we say that f is *monotonically increasing* if $f(x) \leq f(y)$ whenever $x < y$. If $f(x) \geq f(y)$ whenever $x < y$ we say that f is *monotonically decreasing.*

Theorem 5.16. *Let f be a monotonically increasing function on (a,b). Then, $f(x^+)$ and $f(x^-)$ exist at every point x of (a,b). We have*

$$\sup_{t\in(a,x)} f(t) = f(x^-) \leq f(x) \leq f(x^+) = \inf_{t\in(x,b)} f(t). \tag{15}$$

Moreover, if $x < y$, we have

$$f(x^+) \leq f(y^-). \tag{16}$$

Similar results hold for a monotonically decreasing function.

Proof. Since f is monotonic, the set of numbers $A = \{f(t) \mid t \in (a,x)\}$ is bounded above by $f(x)$. Therefore, it has a supremum. Let us call such number $a := \sup A$. Obviously we have $a \leq f(x)$. We need to show that $a = f(x^-)$. For $\epsilon > 0$, since a is an upper bound, we can find points of A between $a - \epsilon$ and a. This means that there is a point $a < t < x$ such that $a - \epsilon < f(t) < a$. Setting $\delta = x - t$, we have found $\delta > 0$ such that $a - \epsilon < f(x - \delta) < a$. Since f is monotonic, for all $x > t > x - \delta$ we have that $f(x - \delta) \leq f(t) \leq a$. This implies that $f(x^-) = a$, since if $t_n \longrightarrow x$, with $t_n \in (a,x)$, we find that eventually $x - \delta < t_n < x$ and $a - \epsilon \leq f(t_n) \leq a$. So, $f(t_n) \longrightarrow a$. The same reasoning shows that the second half of the chain (15) holds too.

To show the last part of the statement, suppose that $x < y$. From the first part of the proof, we have that $f(x^+) = \inf_{t\in(x,b)} f(t) = \inf_{t\in(x,y)} f(t)$, where the last equality is obtained by the monotonicity assumption. In a similar way, we have $f(y^-) = \sup_{t\in(a,y)} f(t) = \sup_{t\in(x,y)} f(t)$. Then, (16) follows. □

We have therefore obtained the following interesting result showing that monotonic functions are relatively well behaved.

Corollary 5.17. *Monotonic functions have no discontinuities of the second kind.*

The following strengthens the results on the good behavior of monotonic functions.

Theorem 5.18. *Let f be monotonic on (a,b). Then, the set of points at which f is discontinuous is at most countable.*

Proof. Suppose without loss of generality that f is increasing. Let E be the set of points at which f is discontinuous. Since each point of E is a discontinuity of the first kind, we can find for each point $x \in E$ a rational number $q(x) \in (f(x^-), f(x^=))$. Since $x < y$ implies $f(x^+) \leq f(y^-)$, we obtain that $q(x) \neq q(y)$. So, E is in 1-1 correpsondence with a subset of $\mathbb{Q}$, and has to be countable. □

Definition 5.19. Let f be a real function defined on $[a,b]$. For any $x \in [a,b]$ define

$$\frac{\Delta f}{\Delta x}(t) = \frac{f(t) - f(x)}{t - x}, \tag{17}$$

where $t \in (a,b)$ and $t \neq x$. We define, if it exists, the limit

$$f'(x) = \lim_{t\to x} \frac{\Delta f}{\Delta x}(t).$$

For all those points x such that f' exists, we say that f is *differentiable* at x, and we call f' the derivative of f. Sometimes we also write, for $f'(x)$, $\frac{d}{dx}(f)(x)$, or $\frac{df}{dx}(x)$, $D_x(f)(x)$.

Remark 5.20. Considering only left or right limits for f', we obtain the notion of left and right derivative.

This is a well known result.

Theorem 5.21. *Let f be differentiable at $x \in [a, b]$, which is the domain of f. Then f is continuous at x.*

Proof. Taking $t \longrightarrow x$ we have

$$f(t) - f(x) = \frac{f(t) - f(x)}{t - x} \cdot (t - x) \to f'(x) \cdot 0 = 0.$$

□

Remark 5.22. The converse is not true. In fact, Weierstrass constructed a function which is continuous everywhere, but not differentiable anywhere.

Theorem 5.23. *Let f and g be defined on $[a, b]$ and suppose that they are differentiable at $x \in [a, b]$. Then, $f+g$, fg and $\frac{f}{g}$ (assuming $g(x) \neq 0$) are differentiable as well and the derivatives are obtained as follows.*

(a) $(f + g)'(x) = f'(x) + g'(x)$.
(b) $(fg)'(x) = f'(x)g(x) + g(x)f'(x)$ *(Leibniz rule).*
(c) $(\frac{f}{g})'(x) = \frac{g(x)f'(x) - g'(x)f(x)}{g(x)^2}$.

Proof. The proof of (a) is more or less trivial, and left as an exercise to the reader. To prove (b), observe that

$$f(t)g(t) - f(x)g(x) = f(t)[g(t) - g(x)] + g(x)[f(t) - f(x)].$$

Therefore, dividing by $(t - x)$ and taking the limit $t \longrightarrow x$ we obtain

$$\begin{aligned} \lim_{t\to x} \frac{(fg)(t) - (fg)(x)}{t - x} &= \lim_{t\to x} \frac{f(t)g(t) - f(x)g(x)}{t - x} \\ &= \lim_{t\to x} \frac{f(t)[g(t) - g(x)] + g(x)[f(t) - f(x)].}{t - x} \\ &= f(x)g'(x) + g(x)f'(x), \end{aligned}$$

having used the fact that $f(t) \to f(x)$ since f is continuous at x by Theorem 5.21.

To prove (c), we proceed similarly. First, observe that since g is continuous at x and $g(x) \neq 0$, there is a neighborhood of x where $g \neq 0$ (why?). So, quotienting by $g(t)$ also makes sense because taking the limit $t \longrightarrow x$ we can restrict ourselves in the neighborhood where $g \neq 0$. Then, note that

$$\frac{\frac{f(t)}{g(t)} - \frac{f(x)}{g(x)}}{t - x} = \frac{1}{g(t)g(x)}[g(x)\frac{f(t) - f(x)}{t - x} - f(x)\frac{g(t) - g(x)}{t - x}].$$

The result therefore follows by taking $t \longrightarrow x$ as in (b). □

Example 5.24. Observe that the derivative of any constant function is obviously zero by definition. Applying the previous result we find that if $f(x) = x$, then $f'(x) = 1$. Using the previous result several times, one finds that for $f(x) = x^n$ with $n \in \mathbb{Z}$ (assuming $x \neq 0$ when $n < 0$), one has $f'(x) = nx^{n-1}$. So, polynomial and rational functions are all differentiable (wherever the denominator does not vanish).

We now prove the chain rule.

Theorem 5.25. *Suppose that f is continuous on $[a,b]$ and that $f'(x)$ exists at some point $x \in [a,b]$. Let g be defined on some interval I which contains the image of f, and assume that g is differentiable at $f(x)$. Setting $h(t) = g \circ f(t) = g(f(t))$, we have that h is differentiable at x and the derivative is given by the so called chain rule:*

$$h'(x) = g'(f(x))f'(x).$$

Proof. We set $y = f(x)$. By definition of derivative we can write $f(t) - f(x)$ and $g(s) - g(y)$ as

$$f(t) - f(x) = (t-x)[f'(x) + u(t)] \tag{18}$$

$$g(s) - g(y) = (s-y)[g'(y) + v(x)], \tag{19}$$

where u and v converge to zero as t and s converge to x and y, respectively. We set $s = f(t)$. We have

$$\begin{aligned} h(t) - h(x) &= g(f(t)) - g(f(x)) \\ &= [f(t) - f(x)][g'(y) + v(s)] \\ &= (t-x)[f'(x) + u(t)][g'(y) + v(s)], \end{aligned}$$

where we have used first Equation (19) and then Equation (18).

When $t \neq x$, we can devide by $t - x$ and obtain

$$\frac{h(t) - h(x)}{t - x} = [g'(y) + v(s)][f'(x) + u(t)].$$

Taking the limit $t \longrightarrow x$, and considering that $s \longrightarrow y$ by continuity of f, we obtain the result. □

Definition 5.26. Let $f : X \longrightarrow \mathbb{R}$, where X is a metric space. Then, we say that f has a *local maximum* at $q \in X$ f there exists a $\delta > 0$ such that $f(x) \leq f(q)$ for all $x \in B(q, \delta)$. Similarly one defines a *local minimum*.

Theorem 5.27. *Let f be defined on $[a,b]$. If f has a local maximum at $x \in (a,b)$ and $f'(x)$ exists, then $f'(x) = 0$. Similar result holds for a local minimum.*

Proof. We can find δ such that the definition of local maximum is satisfied, and such that $B(x, \delta) \subset (a,b)$.

We consider two cases. First, let $x - \delta < t < x$, in which case we have

$$\frac{f(t) - f(x)}{t - x} \geq 0,$$

since both numerator and denominator are negative. Then, taking the limit $t \longrightarrow x$, we find $f'(x) \geq 0$.

For the second case, when $x < t < x + \delta$, we have

$$\frac{f(t) - f(x)}{t - x} \leq 0,$$

since the numerator is negative, and the denominator is positive. Then, taking the limit $t \longrightarrow x$, we find $f'(x) \leq 0$. The combination of the two inequalities gives $f'(x) = 0$. □

Theorem 5.28 (Mean Value Theorem)**.** *Let f and g be continuous functions on $[a,b]$, which are differentiable in (a,b). Then, there is a point $x \in (a,b)$ at which*

$$[f(b) - f(a)]g'(x) = [g(b) - g(a)]f'(x).$$

Proof. We start by defining the auxiliary function

$$h(t) = [f(b) - f(a)]g(t) - [g(b) - g(a)]f(t),$$

for all $t \in [a, b]$. Observe that h is continuous on $[a, b]$ and differentiable on (a, b), just like f and g. Moreover, $h(a) = f(b)g(a) - f(a)g(b) = h(b)$. Note that if $h'(x) = 0$ for some $x \in (a, b)$, then the result would follow, since $h'(x) = [f(b) - f(a)]g'(x) - [g(b) - g(a)]f'(x)$. We will show that this happens.

If h is the constant function, then there would be nothing to prove, since $h'(x) = 0$ would be true for all x. Therefore, we can assume that h is not the constant function. Suppose that there is a point t such that $h(t) > h(a) = h(b)$. Since h is defined over a compact, it has a maximum. Let x be a point where h attains its maximum. The point x is neither a nor b, since if it was, then $h(t)$ could not be larger than $h(a)$ and $h(b)$ against our assumption. We then have $x \in (a, b)$ and by Theorem 5.27 it follows that $h'(x) = 0$. If t is such that $h(t) < h(a) = h(b)$, we can proceed analogously with the minimum. The theorem is proved. □

As a particular case, we obtain the following (more common) form of the mean value theorem, which is just the same as above, where $g(x) = x$.

Theorem 5.29. *Let f be real and continuous oveer $[a, b]$, and differentiable over (a, b). Then there is a point $x \in (a, b)$ at which*

$$f(b) - f(a) = (b - a)f'(x).$$

Theorem 5.30. *Let f be differentiable in (a, b). Then*

- *If $f'(x) \geq 0$ for all $x \in (a, b)$, f is monotonically increasing.*
- *If $f'(x) = 0$ for all $x \in (a, b)$, then f is constant.*
- *If $f'(x) \leq 0$ for all $x \in (a, b)$, then f is monotonically decreasing.*

Exercise 5.31. Prove Theorem 5.30 using Theorem 5.29.

The following result gives a sort of analogue of the intermediate value theorem for the derivative of a function.

Theorem 5.32. *Let f bea real differentiable function on $[a, b]$. Let $f'(a) < \lambda < f'(b)$ for some $\lambda \in \mathbb{R}$. Then there is a point $x \in (a, b)$ such that $f'(x) = \lambda$. Similarly, the same conclusion holds when $f'(a) > \lambda > f'(b)$.*

Proof. Set $g(t) = f(t) - \lambda t$. Then we have that $g'(a) < 0$, which implies that g is decreasing around a, and therefore we can find $t_1 > a$ such that $g(t_1) < g(a)$. Similarly, $g'(b) > 0$, and we can find $t_2 < b$ such that $g(t_2) < g(b)$. We know that g has to attain its minimum in $[a, b]$ since f is continuous (and therefore g as well) over a compact. From the inequalities for $g(t_1)$ and $g(t_2)$ we find that the minimum has to be in (a, b). By Theorem 5.27 $g'(x) = 0$ at the minimum, implying that $g'(x) = f'(x) - \lambda = 0$. This completes the proof. □

Corollary 5.33. *If f is differentiable on $[a, b]$, then f' does not have any discontinuities of the first kind on $[a, b]$.*

Proof. First, observe that by definition, a diascontinuity of the first kind can only happen in (a, b). Then, suppose by way of contradiction that the point $x \in (a, b)$ is a discontinuity of the first kind. Let $l^- := f'(x^-)$ and $l^+ := f'(x^+)$. Without loss of generality let $l^- < l^+$. We can choose $\epsilon > 0$ such that $l^- + \epsilon < l^+ - \epsilon$. By definition of left and right limit, we can find δ such that for $x - \delta < z < x$

we have $|f'(z) - l^-| < \epsilon$, and for $x < z < x + \delta$ we have $|f'(z) - l^+| < \epsilon$. Then, in the interval $[x - \delta', x + \delta']$ where $\delta' < \delta$, we see that any value $l^- + \epsilon < \lambda < l^+ - \epsilon$ is not attained by f', against the result of Theorem 5.32. □

Definition 5.34. Let f be a function defined on $E \subset \mathbb{R}$, such that for each interval $(a, +\infty)$, we have $(a, +\infty) \cap E \neq \emptyset$. Then, we say that $\lim_{x \to +\infty} f(x) = +\infty$ if for any $M \in \mathbb{R}$ we can find $\gamma \in \mathbb{R}$ such that for all $x > \gamma$ it holds $f(x) > M$. We say that $\lim_{x \to +\infty} f(x) = L$ if for all $\epsilon > 0$ we can find $\gamma > 0$ such that for all $x > \gamma$ we have $|f(x) - L| < \epsilon$. Similar definitions can be posed for $x \to -\infty$.

Theorem 5.35 (de L'Hôpital's rule)**.** *Let f and g be differentiable in (a, b), with $g'(x) \neq 0$ for all $x \in (a, b)$ where $-\infty \leq a < b \leq +\infty$. Suppose that*

$$\lim_{x \to a} \frac{f'(x)}{g'(x)} = L, \tag{20}$$

and either of the following holds:

$$f(x), g(x) \longrightarrow 0, \tag{21}$$

$$g(x) \longrightarrow +\infty, \tag{22}$$

as $x \longrightarrow a$. Then, $\lim_{x \to a} \frac{f(x)}{g(x)} = L$. Similar statements also hold when $x \to b$, or when $g(x) \to -\infty$.

Proof. First, assume that $-\infty \leq L < +\infty$. We choose a number q such that $L < q$ and we pick r such that $L < r < q$. Equation 20 implies that we can find $c \in (a, b)$ such that $\frac{f'(x)}{g'(x)} < r$ whenever $a < x < c$. For $a < x < y < c$, by Theorem 5.28 we find $t \in (x, y)$ such that

$$\frac{f(x) - f(y)}{g(x) - g(y)} = \frac{f'(t)}{g'(t)} < r \tag{23}$$

. Then, if condition (22) holds we can take $x \to a$ and get $\frac{f(y)}{g(y)} \leq r < q$ for all $a < y < c$. If condition (22) holds, we can keep y fixed and take a point $c_1 \in (a, y)$ wuch that $g(x) > g(y)$ (the limit goes to $+\infty$) and $g(x) > 0$ for all $a < x < c_1$ (the limit goes to $+\infty$). Then, multiplying Equation (23) by $\frac{g(x)-g(y)}{g(x)}$ we get

$$\frac{f(x)}{g(x)} < r - r\frac{g(y)}{g(x)} + \frac{f(y)}{g(x)}, \tag{24}$$

whenever $a < x < c_1$. Taking the limit as $x \to a$ we find $\frac{f(x)}{g(x)} < q$ eventually, say for $a < x < c_2$. So, we can find some value c^* such that whenever $a < x < c^*$ we have $\frac{f(x)}{g(x)} < q$ with any choice of $q > L$. Similarly, we can show that whenever $-\infty < L \leq +\infty$ and chosen any p such that $p < L$, we can find some c^* such that $a < x < c^*$ gives $\frac{f(x)}{g(x)} > p$. The two results prove the theorem. □

Definition 5.36. If f is differentiable on $[a, b]$, and f' is itself differentiable, we can take the derivative of f' and call it *second derivative*, also denoted by $f''(x)$ or $\frac{d^2}{dx^2}(f)$ or $D_x^2(f)$ and so on. If the result is differentiable, one can continue applying this procedure and take further derivatives. We denote them by $f^{(n)}$ and similar symbols as above.

Theorem 5.37 (Taylor's Theorem)**.** *Let f be a real function on $[a, b]$, and let n be a natural number. Suppose that $f^{(n-1)}$ is continuous on $[a, b]$, and that $f^{(n)}(t)$ exists at all $t \in (a, b)$. Let*

α, β be in $[a,b]$ with $\alpha \neq \beta$, and define the quantity

$$P(t) = \sum_{k=0}^{n-1} \frac{f^{(k)}(\alpha)}{l!}(t-\alpha)^k. \tag{25}$$

Then, there exists a point x between α and β such that

$$f(\beta) = P(\beta) + \frac{f^{(n)}(x)}{n!}(\beta-\alpha)^n. \tag{26}$$

Proof. We define M as $M := \frac{f(\beta)-P(\beta)}{(\beta-\alpha)^n}$, which is well defined since $\beta \neq \alpha$. We define also the auxiliary function

$$g(t) = f(t) - P(t) - M(t-\alpha)^n,$$

for $t \in [a,b]$. Observe that $g^{(n)} = f^{(n)} - n!M$, since P has only powers up to $n-1$ and therefore has vanishing n^{th} derivative. If we find a point x in (α, β) such that $g^{(n)}(x) = 0$, we would have at such a point that $f^{(n)}(x) = n!M$, which would complete the proof.

Observe that by construction we have that $P^{(k)}(\alpha) = f^{(k)}(\alpha)$ for all $k = 0, 1, \ldots, n-1$. Therefore, $g^{(k)}(\alpha) = 0$ for all $k = 0, 1, \ldots n-1$, where the 0^{th} derivative just represents g with no differentiation at all. By the choice of M, it also holds that $g(\beta) = 0$. Applying the Mean Value Theorem 5.29, there must be a point between α and β, say x_1, such that $g'(x_1) = 0$. Then, we can apply again Theorem 5.29 to α and x_1, since $g'(\alpha) = g'(x_1) = 0$, finding some $x_2 \in (\alpha, x_1)$ where $g''(x_2) = 0$. So proceeding, we find for each $k = 0, 1, \ldots, n$ some x_k such that $g^{(k)}(x_k) = 0$. In particular, $x = x_n$ is the value we were looking for at which $g^{(n)}(x) = 0$. This completes the proof. □

6. Riemann-Stieltjes Integral

We begin this section with a definition.

Definition 6.1. Let $[a,b]$ be an interval. A *partition* P of $[a,b]$ is a finite collection of points $x_0, \ldots, x_n$ such that

$$a = x_0 \leq x_1 \leq \cdots \leq x_{n-1} \leq x_n = b.$$

Given a partition P, we write $\Delta x_i = x_i - x_{i-1}$ for all $i = 1, \ldots, n$.

Suppose f is a bounded real function defined on a closed interval $[a,b]$. For example a continuous function would be bounded, but we do not require such strong assumptions. For an arbitrary partition P of $[a,b]$ we define

$$\begin{aligned} M_i &= \sup_{[x_{i-1},x_i]} f(x) \\ m_i &= \inf_{[x_{i-1},x_i]} f(x) \\ U(P,f) &= \sum_{i=1}^{n} M_i \Delta x_i \\ L(P,f) &= \sum_{i=1}^{n} m_i \Delta x_i. \end{aligned}$$

Additionally, we define the sets $U(f) := \{U(P, f) \mid P \subset [a, b] \text{ is a partition}\}$ and $L(f) := \{L(P, f) \mid P \subset [a, b] \text{ is a partition}\}$. Lastly, we define

$$\overline{\int_a^b} f dx = \inf U(f), \tag{27}$$

$$\underline{\int_a^b} f dx = \sup L(f),. \tag{28}$$

We call $\overline{\int_a^b} f dx$ *upper Riemann integral*, and $\underline{\int_a^b} f dx$ *lower Riemann integral.* Osberve that since f is assumed to be bounded, the sets $U(f)$ and $L(f)$ are both bounded above and below, and therefore the upper and lower Riemann integrals are finite numbers.

Definition 6.2. Let f be a bounded function such that

$$\overline{\int_a^b} f dx = \underline{\int_a^b} f dx. \tag{29}$$

Then f is said to be Riemann integrable, and we define its *Riemann integral* over $[a, b]$ to be $\int_a^b f(x)dx := \overline{\int_a^b} f dx = \underline{\int_a^b} f dx$. If f is Riemann integrable, we also write $f \in \mathcal{R}$, having denoted by $\mathcal{R}$ the collection of Riemann integrable functions.

We now consider a generalization of the notion of Riemann integrability, and Riemann integral.

Definition 6.3. Let α be a monotonically increasing function on $[a, b]$. In particular, α is bounded since for all x we have $\alpha(a) \leq \alpha(x) \leq \alpha(b)$ by monotonicity. Given a partition P of $[a, b]$, we define

$$\Delta\alpha_i = \alpha(x_i) - \alpha(x_{i-1}).$$

Observe that when α is just the identity map, we recover Δx_i as defined above.

Suppose f is a bounded real function defined on a closed interval $[a, b]$. For example a continuous function would be bounded, but we do not require such strong assumptions. For an arbitrary partition P of $[a, b]$ we define

$$U(P, f, \alpha) = \sum_{i=1}^{n} M_i \Delta\alpha_i$$

$$L(P, f, \alpha) = \sum_{i=1}^{n} m_i \Delta\alpha_i.$$

Additionally, we define the sets $U(f, \alpha) := \{U(P, f, \alpha) \mid P \subset [a, b] \text{ is a partition}\}$ and $L(f, \alpha) := \{L(P, f, \alpha) \mid P \subset [a, b] \text{ is a partition}\}$. Lastly, we define

$$\overline{\int_a^b} f d\alpha = \inf U(f, \alpha), \tag{30}$$

$$\underline{\int_a^b} f d\alpha = \sup L(f, \alpha),. \tag{31}$$

Definition 6.4. When $\overline{\int_a^b} f d\alpha = \underline{\int_a^b} f d\alpha$ we say that f is *Riemann-Stieltjes integrable with respect to* α, and we set $\int_a^b f d\alpha := \overline{\int_a^b} f d\alpha = \underline{\int_a^b} f d\alpha$, which we call the *Riemann-Stieltjes integral* of f over $[a, b]$. If f is Riemann-Stieltjes integrable, we also write $f \in \mathcal{R}(\alpha)$, having denoted by $\mathcal{R}(\alpha)$ the set of Riemann-Stieltjes integrable functions with respect to α.

Definition 6.5. Given two partitions of $[a,b]$, P and P^*, we say that P^* is a *refinement* of P if $P^* \supset P$. Given two partitions P_1 and P_2, we say that P^* is their *common refinement* if $P^* = P_1 \bigcup P_2$.

Theorem 6.6. *Let P^* be a refinement of P. Then, we have*

$$L(P,f,\alpha) \leq L(P^*,f,\alpha) \tag{32}$$

$$U(P^*,f,\alpha) \leq U(P,f,\alpha). \tag{33}$$

Proof. We will prove (32), as the proof of (33) is substantially the same.

It is clear that if we prove the result for P^* consisting of only an extra point with respect to P, the result would hold for any refinement, since we can apply the procedure multiple times until we find the inequality for P and an arbitrary refinement P^*.

So, suppose that P^* is obtained from P by adding one extra point x^*, which we assume to satisfy $x_{i-1} < x^* < x_i$. Let $w_1 := \inf_{x\in[x_{i-1},x^*]} f(x)$ and $w_2 := \sup_{x\in[x^*,x_i]} f(x)$. It is clear that if $m_i := \inf_{x\in[x_{i-1},x_i]} f(x)$, we have $w_1, w_2 \geq m_i$. Then, we compute

$$\begin{aligned}
L(P^*,f,\alpha) - L(P,f,\alpha) &= \sum_{j=1}^{i-2} m_j\Delta\alpha_j + w_1(\alpha(x^*) - \alpha(x_{i-1})) + w_2(\alpha(x_i) - \alpha(x^*)) + \sum_{j=i}^{n} m_j\Delta\alpha_j \\
&\quad -\sum_{j=1}^{i-2} m_j\Delta\alpha_j - m_i(\alpha(x_i) - \alpha(x_{i-1})) - \sum_{j=i}^{n} m_j\Delta\alpha_j \\
&= w_1(\alpha(x^*) - \alpha(x_{i-1})) + w_2(\alpha(x_i) - \alpha(x^*)) - m_i(\alpha(x_i) - \alpha(x_{i-1})) \\
&= w_1(\alpha(x^*) - \alpha(x_{i-1})) + w_2(\alpha(x_i) - \alpha(x^*)) \\
&\quad -m_i(\alpha(x_i) - \alpha(x^*) + \alpha(x^*) - \alpha(x_{i-1})) \\
&= (w_1 - m_i)[\alpha(x^*) - \alpha(x_{i-1})] + (w_2 - m_i)[\alpha(x_i) - \alpha(x^*)] \\
&\geq 0.
\end{aligned}$$

The proof is complete. ☐

Theorem 6.7. $\underline{\int_a^b} f d\alpha \leq \overline{\int_a^b} f d\alpha$.

Proof. Let P_1 and P_2 be arbitrarry partitions. Let $P^* = P_1 \bigcup P_2$ be their common refinement. Then, applying Theorem 6.6 we find that

$$L(P_1,f,\alpha) \leq L(P^*,f,\alpha) \leq U(P^*,f,\alpha) \leq U(P_2,f,\alpha),$$

where the central inequality holds because the lower sum of a partition is always smaller than the upper sum for the same partition by definition. It follows that

$$L(P_1,f,\alpha) \leq U(P_2,f,\alpha),$$

for any P_1 and P_2. Therefore, we also have that

$$\underline{\int_a^b} f d\alpha = \sup_{P_1} L(P_1,f,\alpha \leq U(P_2,f,\alpha),$$

which in turn implies that

$$\underline{\int_a^b} f d\alpha \leq \inf_{P_2} U(P_2,f,\alpha) = \overline{\int_a^b} f d\alpha.$$

☐

The following result gives us a directly applicable criterion to determine whether a function is integrable or not. We will use it several times in the rest of the course.

Theorem 6.8. *We have that $f \in \mathcal{R}(\alpha)$ if and only if for every $\epsilon > 0$ there exists a partition P such that*

$$U(P, f, \alpha) - L(P, f, \alpha) < \epsilon. \tag{34}$$

Proof. ($\Longleftarrow$) For every partition P, by definition and Theorem 6.7, we have that

$$L(P, f, \alpha) \leq \underline{\int_a^b} f d\alpha \leq \overline{\int_a^b} f d\alpha \leq U(P, f, \alpha).$$

Therefore, if P is such that (34) holds, we have that

$$0 \leq \overline{\int_a^b} f d\alpha - \underline{\int_a^b} f d\alpha < \epsilon.$$

Since this can be done for every $\epsilon > 0$, we obtain that $\overline{int_a^b} f d\alpha - \underline{int_a^b} f d\alpha = 0$ and therefore $f \in \mathcal{R}(\alpha)$.

($\Longrightarrow$) For a given $\epsilon > 0$, by definition of infimum and supremum, we have that we can find two partitions P_1 and P_2 that satisfy

$$\begin{aligned} U(P_2, f, \alpha) &< \int_a^b f d\alpha + \frac{\epsilon}{2}, \\ \int_a^b f d\alpha &< L(P_1, f, \alpha) + \frac{\epsilon}{2}. \end{aligned}$$

Let $P = P_1 \bigcup P_2$ be their common refinement, then we obtain the inequalities

$$\begin{aligned} U(P, f, \alpha) &\leq U(P_2, f, \alpha) \\ &< \int_a^b f d\alpha + \frac{\epsilon}{2} \\ &< L(P_1, f, \alpha) + \epsilon \\ &\leq L(P, f, \alpha) + \epsilon. \end{aligned}$$

This completes the proof, as the partition P is the one we were looking for. □

Theorem 6.9. *The following facts hold.*

(a) *If* (34) *holds for some P and $\epsilon > 0$, then it also holds for the same ϵ for all refinements of P.*

(b) *If* (34) *holds for $P = \{x_0, \ldots, x_n\}$ and if s_i, t_i are arbitrary elements in $[x_{i-1}, x_i]$, then*

$$\sum_{i=1}^{n} |f(s_i) - f(t_i)| \Delta\alpha_i < \epsilon.$$

(c) *If $f \in \mathcal{R}(\alpha)$ and the hypotheses of (b) hold, then*

$$|\sum_{i=1}^{n} f(t_i) \Delta\alpha_i - \int_a^b f d\alpha| < \epsilon.$$

Proof. (a) follows directly from Theorem 6.6. To prove (b), one just needs to notice that $f(s_i), f(t_i) \in [m_i, M_i]$ for all $i = 1, \dots, n$, and therefore $|f(s_i) - f(t_i)| \leq M_i - m_i$. Then, one gets

$$\begin{aligned} \sum_{i=1}^{n} |f(s_i) - f(t_i)|\Delta\alpha_i \\ &\leq \sum_{i=1}^{n}[M_i - m_i]\Delta\alpha_i \\ &= \sum_{i=1}^{n} M_i\Delta\alpha_i - \sum_{i=1}^{n} m_i\Delta\alpha_i \\ &= U(P, f, \alpha) - L(P, f, \alpha) \\ &< \epsilon. \end{aligned}$$

To prove (c), observe that the following inequalities follow directly from the definition

$$\begin{array}{ccc} L(P, f, \alpha) \leq & \sum_i f(t_i)\Delta\alpha_i & \leq U(P, f, \alpha) \\ L(P, f, \alpha) \leq & \int_a^b f d\alpha & \leq U(P, f, \alpha). \end{array}$$

Then, $|\sum_{i=1}^{n} f(t_i)\Delta\alpha_i - \int_a^b f d\alpha| \leq U(P, f, \alpha) - L(P, f, \alpha) < \epsilon$, which completes the proof. □

The following result gives our first relatively large class of functions that Riemann-Stieltjes integrable.

Theorem 6.10. *If f is continuous on $[a, b]$, then $f \in \mathcal{R}(\alpha)$.*

Proof. The idea of the proof is to use the criterion of Theorem 6.8 to show that f is integrable. To this purpose, let $\epsilon > 0$ be given and fixed. Choose $\eta > 0$ such that $\eta\alpha(b) - \alpha(a) < \epsilon$. Since f is continuous over a compact, it is uniformly continuous by the Heie-Cantor Theorem. It follows that we can find $\delta > 0$ such that whenever $|x - y| < \delta$ ($x, y \in [a, b]$) it follows that $|f(x) - f(y)| < \eta$. Let us choose a partition P such that $\Delta x_i < \delta$. Then, by uniform continuity we have that for each $[x_{i-1}, x_i]$ we have $M_i - m_i \leq \eta$. As a consequence we have

$$\begin{aligned} U(P, f, \alpha) - L(P, f, \alpha) &= \sum_{i=1}^{n}[M_i - m_i]\Delta\alpha_i \\ &\leq \eta\sum_{i=1}^{n} \Delta\alpha_i \\ &= \eta[\alpha(b) - \alpha(a)] \\ &< \epsilon. \end{aligned}$$

This completes the proof. □

Theorem 6.11. *Let f be monotonic on $[a, b]$, and let α be continuous on $[a, b]$ (in addition to being monotonic). Then $f \in \mathcal{R}(\alpha)$.*

Proof. Let $\epsilon > 0$ be fixed. Let $n \in \mathbb{N}$ be arbitrarily chosen. We set $x_0 = a$ and $x_n = b$. Since α is continuous, by the intermediate value theorem we can find a point $x_1 \in (a, b) = (x_0, x_n)$ such that $\alpha(x_1) = \alpha(x_0) + \frac{\alpha(b)-\alpha(a)}{n}$. Then, applying again the intermediate value theorem, we can find a point $x_2 \in (x_1, x_n)$ such that $\alpha(x_2) = \alpha(x_1) + \frac{\alpha(b)-\alpha(a)}{n}$. So proceeding, we find $x_1, \dots, x_{n-1}$ with the property that for all $i = 1, \dots, n$ we have $\Delta\alpha_i = \frac{\alpha(b)-\alpha(a)}{n}$. This gives us a partition

$P = \{x_0, x_1, \ldots, x_n\}$ of $[a, b]$. Suppose, without loss of generality, that f is increasing. Then, in each interval $[x_{i-1}, x_i]$ we have $M_i = f(x_i)$ and $m_i = f(x_{i-1})$. We obtain

$$\begin{aligned} U(P, f, \alpha) - L(P, f, \alpha) &= \frac{\alpha(b) - \alpha(a)}{n} \sum_{i=1}^{n} [f(x_i) - f(x_{i-1})] \\ &= \frac{\alpha(b) - \alpha(a)}{n} [f(b) - f(a)]. \end{aligned}$$

Since n is arbitrary, we can choose n large enough such that $\frac{\alpha(b)-\alpha(a)}{n}[f(b) - f(a)] < \epsilon$, and the proof is complete. □

Theorem 6.12. *Suppose that f is bounded on $[a, b]$, and has a finite set of points of discontinuity in $[a, b]$. Suppose that α is continuous at all points where f is not continuous. Then $f \in \mathcal{R}(\alpha)$.*

Proof. Let $\epsilon > 0$ be fixed. We define $M = \sup_{x \in [a,b]} |f(x)|$. Let $E = \{e_1, \ldots, e_q\}$ be the set of points of discontinuity of f. Since α is continuous at each point, we can find intervals $[u_i, v_i]$ such that $|\alpha(t) - \alpha(e_i)| < \frac{\epsilon}{2}$ for each $t \in [u_i, v_i] \cap [a, b]$, and $e_i \in (u_i, v_i)$. Therefore, in particular, we have that $\alpha(v_i) - \alpha(u_i) = |\alpha(v_i) - \alpha(u_i)| \leq |\alpha(v_i) - \alpha(e_i)| + |\alpha(e_i) - \alpha(u_i)| < \epsilon$.

We put $K = [a, b] - [\bigcup_i (u_i, v_i)]$. Since $\bigcup_i (u_i, v_i)$ is open, we have that K is a closed set inside the compact $[a, b]$, and it is therefore compact as well. Since the points of discontinuity of f are inside the sets (u_i, v_i), it follows that f is continuous over K, and it is therefore uniformly continuous by the Heine-Cantor Theorem. We can therefore find $\delta > 0$ such that $|f(s) - f(t)| < \epsilon$ whenever $s, t \in K$ and $|s - t| < \delta$.

Let P be a partition constructed as follows. Each point u_i and v_i is part of the partition, but no partition point fall inside (u_i, v_i). The other points are taken in K in such a way that $\Delta x_j < \delta$ for all j.

We can therefore subdivide the sum for $U(P, f, \alpha) - L(P, f, \alpha)$ in terms corresponding to those j's such that $x_{j-1} = u_i$ for some i, and those j's corresponding to points that are not u_i's, for any of the i's. We define

$$\begin{aligned} A &= \{j \mid x_{j-1} = u_i \ \text{ for some } i\}, \\ B &= \{j \mid x_{j-1} \neq u_i \ \text{ for all } i\}. \end{aligned}$$

Observe that by definition of M we have that $|f(s) - f(t)| \leq 2M$ for all $t, s \in [a, b]$, while for $j \in B$ we have that $M_j - m_j \leq \epsilon$. When $j \in A$ we have by construction that $\Delta\alpha_j < \epsilon$. Then we find

$$\begin{aligned} U(P, f, \alpha) - L(P, f, \alpha) &= \sum_{j \in A} [M_j - m_j] \Delta\alpha_j + \sum_{j \in B} [M_j - m_j] \Delta\alpha_j \\ &\leq \epsilon \sum_{j \in A} [M_j - m_j] + \epsilon \sum_{j \in B} \Delta\alpha_j \\ &\leq \epsilon[2M + \alpha(b) - \alpha(a)]. \end{aligned}$$

Since ϵ is arbitrary, we can choose it in such a way to make the whole product as small as we want. This completes the proof. □

Theorem 6.13. *Suppose that $f \in \mathcal{R}(\alpha)$ on $[a, b]$, with $m \leq f(x) \leq M$ for all $x \in [a, b]$ and ϕ continuous on $[m, M]$. Set $h(x) = \phi \circ f(x)$. Then, $h \in \mathcal{R}(\alpha)$.*

Proof. Let $\epsilon > 0$ be fixed. Since ϕ is uniformly continuous on $[m, M]$ we can find $\delta > 0$ such that $|\phi(s) - \phi(t)| < \epsilon$ whenever $|s - t| \leq \delta$. Additionally, we can assume $\delta < \epsilon$ too, upon possibly taking a smaller δ than the one found by definition of uniform continuity.

Since $f \in \mathcal{R}(\alpha)$, we can find a partition P of $[a, b]$ such that

$$U(P, f, \alpha) - L(P, f, \alpha) < \delta^2.$$

We take M_i, m_i to be supremum and infimum of f on each interval $[x_{i-1}, x_i]$ determined by the partition, and we let M_i^*, m_i^* indicate their counterparts for h. We divide the numbers $i = 1, \dots, n$ in two classes. We set $i \in A$ if $M_i - m_i < \delta$, while we set $i \in B$ if $M_i - m_i \geq \delta$.

When $i \in A$, by the choice of δ and the definition of M_i^*, m_i^*, we find that $M_i^* - m_i^* = \sup_{t\in[x_{i-1},x_i]} f(t) - \inf_{t\in[x_{i-1},x_i]} f(t)| \leq \epsilon$. When $i \in B$, we have $M_i^* - m_i^* \leq 2\sup_{x\in[m,M]} |\phi(x)|$. By the choice of partition we have

$$\begin{aligned}
\delta \sum_{i\in B} \Delta\alpha_i &\leq \sum_{i\in B} (M_i - m_i)\Delta\alpha_i \\
&\leq \sum_{i=1}^{n} (M_i - m_i)\Delta\alpha_i \\
&= U(P, f, \alpha) - L(P, f, \alpha) \\
&< \delta^2,
\end{aligned}$$

from whichwe deduce $\sum_{i\in B} \Delta\alpha_i < \delta$. We therefore obtain

$$\begin{aligned}
U(P, h, \alpha) - L(P, h, \alpha) &= \sum_{i=1}^{n} (M_i^* - m_i^*)\Delta\alpha_i \\
&= \sum_{i\in B} (M_i^* - m_i^*)\Delta\alpha_i + \sum_{i\in B} (M_i^* - m_i^*)\Delta\alpha_i \\
&\leq \epsilon\Delta\alpha_i + 2K \sum_{i\in B} \Delta\alpha_i \\
&\leq \epsilon[\alpha(b) - \alpha(a)] + 2K\delta \\
&< \epsilon[\alpha(b) - \alpha(a) + 2K].
\end{aligned}$$

Since ϵ was arbitrary and $\alpha(b) - \alpha(a) + 2K$ is a constant independent of the partition P chosen, we can make $\epsilon[\alpha(b) - \alpha(a) + 2K]$ arbitrarily small and the proof is complete. □

The following result, proof of which is left to the reader, lists some useful properties of the Riemann-Stieltjies integral.

Theorem 6.14. *The following facts hold.*

(1) *If $f_1, f_2 \in \mathcal{R}(\alpha)$, then $f_1 + f_2, cf_1 \in \mathcal{R}(\alpha)$, where $c \in \mathbb{R}$ is arbitrary. Moreover,*

$$\int_a^b (f_1 + f_2) d\alpha = \int_a^b f_1 d\alpha + \int_a^b f_2 d\alpha$$

and

$$\int_a^b c f_1 d\alpha = c \int_a^b f_1 d\alpha.$$

(2) *If $f_1(x) \leq f_2(x)$ for all $x \in [a, b]$, where $f_1, f_2 \in \mathcal{R}(\alpha)$, then*

$$\int_a^b f_1 d\alpha \leq \int_a^b f_2 d\alpha.$$

(3) *If* $f \in \mathcal{R}(\alpha)$ *and* $c \in (a,b)$*, then*

$$\int_a^b f d\alpha = \int_a^c f d\alpha + \int_c^b f d\alpha.$$

(4) *If* $f \in \mathcal{R}(\alpha)$ *and* $|f(x)| \leq M$ *for all* $x \in [a,b]$*, then*

$$|\int_a^b f d\alpha| \leq M(b-a).$$

(5) *If* $f \in \mathcal{R}(\alpha_1)$ *and* $f \in \mathcal{R}(\alpha_2)$*, then* $f \in \mathcal{R}(a\alpha_1 + b\alpha_2)$ *for any* $a, b \in \mathbb{R}$*, and*

$$\int_a^b f d(\alpha_1 + \alpha_2) = a\int_a^b f d\alpha_1 + b\int_a^b f d\alpha_2.$$

Exercise 6.15. Prove Theorem 6.14.

Theorem 6.16. *If* $f, g \in \mathcal{R}(\alpha)$ *then*

(1) $fg \in \mathcal{R}(\alpha)$.
(2) $|f| \in \mathcal{R}(\alpha)$ *and* $|\int_a^b f d\alpha| \leq \int_a^b |f| d\alpha$.

Proof. To prove the first statement, consider the function $\phi(t) = t^2$, which is continuous. If $q(x)$ is an integrable function, then $h(x) = q(x)^2$ is integrable by Theorem 6.13, being the composition of q and ϕ. So, the functions $(f+g)^2$ and $(f-g)^2$ are integrable since $q = f+g$ and $q = f-g$ are integrable (linear combinations of integrable functions). Since we have

$$fg = \frac{1}{4}[(f+g)^2 - (f-g)^2],$$

we obtain that fg is integrable as well.

To prove the second part, observe that $|f(x)| = (\phi \circ f)(x)$, where $\phi(t) = |t|$ is continuous. Then, again by Theorem 6.13 we find that $|f(x)|$ is integrable. Lastly, we can choose $c = \pm 1$ such that $\int f d\alpha = c \int f d\alpha$, depending on the sign of $\int f d\alpha$. Then, one has

$$\begin{aligned} |\int f d\alpha| &= c\int f d\alpha \\ &= \int c f d\alpha \\ &\leq \int |f| d\alpha, \end{aligned}$$

where we have used the fact that $cf(x) \leq |f(x)|$ for all x, along with Theorem 6.14. □

Definition 6.17. The unit step function I is defined as

$$I(x) = \begin{cases} 0 & x \leq 0 \\ 1 & x > 0. \end{cases}$$

Theorem 6.18. *Let* $a < s < b$ *and* $f \in \mathcal{R}(\alpha)$ *on* $[a,b]$*, with* f *continuous at* s *and* $\alpha(x) = I(x-s)$*. Then, we have*

$$\int_a^b f d\alpha = f(s).$$

Proof. We consider a family of partitions $P_x = \{a, s, x, b\}$, where a, b, s are fixed, and we let x take arbitrary values between s and b. Observe that for any $t \leq s$, one has $\alpha(t) = I(t-s) = 0$, while for any $t > s$ one has $\alpha(t) = I(t-s) = 1$.

For the upper sum we compute, for any arbitrary $b > x > s$:

$$\begin{aligned} U(P_x, f, \alpha) &= M_1[\alpha(s) - \alpha(a)] + M_2[\alpha(x) - \alpha(s)] + M_3[\alpha(b) - \alpha(x)] \\ &= M_1[0-0] + M_2[1-0] + M_3[1-1] \\ &= M_2. \end{aligned}$$

Similarly, one has $L(P_x, f, \alpha) = m_2$. As $x \longrightarrow s$ we have $M_2, m_2 \longrightarrow f(s)$ by continuity of f at s. Therefore, $\int_a^b f d\alpha = f(s)$ as stated. □

Theorem 6.19. *Suppose that $c_n \geq 0$, for $n \in \mathbb{N}$, is such that $\sum c_n$ converges, and s_n is a sequence of distinct points in (a, b). Set $\alpha(x) = \sum_{n=1}^{\infty} c_n I(x - s_n)$. Let f be continuous on $[a, b]$. Then*

$$\int_a^b f d\alpha = \sum_{n=1}^{\infty} c_n f(s_n).$$

Proof. Since for all $x \in [a, b]$ $c_n I(x - s_n) \leq c_n$, we find that (comparison test) for any choice of x, the series defining $\alpha(x)$ converges, so that α is a well defined function. Moreover, since $I(t_1) \leq I(t_2)$ for all $t_1 < t_2$, and $c_n, I(x - s_n) \geq 0$ for all n, it follows that $\alpha(x)$ is a monotonic function. Additionally, $\alpha(a) = 0$ and $\alpha(b) = \sum_n c_n$, which means that α is bounded. Therefore, we can define the Riemann-Stieltjies integral with respect to α.

Let $\epsilon > 0$ be given, and choose N such that $\sum_{n=N+1}^{\infty} c_n < \epsilon$, which has to exist because $\sum c_n$ is convergent. For all x, define

$$\begin{aligned} \alpha_1(x) &= \sum_{n=1}^{N} c_n I(x - s_n), \\ \alpha_2(x) &= \sum_{n=N+1}^{\infty} c_n I(x - s_n). \end{aligned}$$

Therefore one has $\alpha(x) = \alpha_1(x) + \alpha_2(x)$. Applying Theorem 6.14 and Theorem 6.18, it immediately follows that

$$\int_a^b f d\alpha_1 = \sum_{n=1}^{N} c_n f(s_n).$$

Now, observe that $\alpha_2(b) - \alpha_2(a) = \sum_{n=N+1}^{\infty} c_n - 0 < \epsilon$, which gives us

$$\begin{aligned} |\int_a^b f d\alpha_2| &\leq M \int_a^b d\alpha_2 \\ &\leq M\epsilon, \end{aligned}$$

where we have set $M := \sup_{x\in[a,b]} |f(x)|$, which is finite since $|f(x)|$ is a continuous function over a compact. Putting these all together we find that

$$\begin{aligned} |\int_a^b f d\alpha - \sum_{n=1}^N c_n f(s_n)| &= |\int_a^b f d(\alpha_1+\alpha_2) - \sum_{n=1}^N c_n f(s_n)| \\ &= |\int_a^b f d\alpha_1 + \int_a^b f d\alpha_2 - \sum_{n=1}^N c_n f(s_n)| \\ &= |\int_a^b f d\alpha_2| \\ &\leq M\epsilon. \end{aligned}$$

Therefore, we have that $\int_a^b f d\alpha - \sum_{n=1}^\infty c_n f(s_n) = 0$ by taking the limit $N \longrightarrow +\infty$. □

Exercise 6.20. Show that given $\epsilon > 0$, the series $\sum c_n$ as in the proof of the previous theorem has to admit an N such that $\sum_{n=N+1}^\infty c_n < \epsilon$. (Hint: Think of the Cauchy criterion of convergence)

In the following, recall that we call $\mathcal{R}$ the set of Riemann integrable functions, which is the same as the Riemann-Stieltjes integrable functions with respect to $\alpha(x) := x$.

Theorem 6.21. *Let α be a monotonically increasing function such that $\alpha' \in \mathcal{R}$. Let f be a bounded real function on $[a,b]$. Then $f \in \mathcal{R}(\alpha)$ if and only if $f\alpha' \in \mathcal{R}$. Moreover, in that case we have*

$$\int_a^b f d\alpha = \int_a^b f(x)\alpha'(x)dx.$$

Proof. Let $\epsilon > 0$ be fixed, and by applying the integrability criterion to α' take a partition $P = \{x_0, \ldots, x_n\}$ of $[a,b]$ such that $U(P,\alpha') - L(P,\alpha') < \epsilon$. By the Mean Value Theorem, for each $i = 1, \ldots, n$, we find points t_i in each interval $[x_{i-1}, x_i]$ such that

$$\Delta\alpha_i = \alpha'(t_i)\Delta x_i.$$

For $s_i \in [x_{i-1}, x_i]$ we have

$$\begin{aligned} \sum_{i=1}^n |\alpha'(s_i) - \alpha'(t_i)|\Delta x_i &\leq \sum_{i=1}^n (M_i(\alpha') - m_i(\alpha'))\Delta x_i \\ &= U(P,\alpha') - L(P,\alpha') \\ &< \epsilon, \end{aligned}$$

where we have used $M_i(\alpha')$ and $m_i(\alpha')$ to indicate the supremum and infimum, respectively, of α' in $[x_{i-1}, x_i]$. Let $M := \sup_{[a,b]} f$. From $\Delta\alpha_i = \alpha'(t_i)\Delta x_i$, we have

$$\sum_{i=1} f(s_i)\Delta\alpha' = \sum_{i=1}^n f(s_i)\alpha'(t_i)\Delta x_i.$$

It therefore follows that

$$\begin{aligned} |\sum_{i=1}^n f(s_i)\Delta\alpha_i - \sum_{i=1}^n f(s_i)\alpha'(s_i)\Delta x_i| &= |\sum_{i=1}^n f(s_i)\alpha'(t_i)\Delta x_i - \sum_{i=1}^n f(s_i)\alpha'(s_i)\Delta x_i| \\ &= \sum_{i=1}^n |f(s_i)[\alpha'(t_i) - \alpha'(s_i)]\Delta x_i| \\ &\leq M\sum_{i=1}^n |\alpha'(t_i) - \alpha'(s_i)|\Delta x_i \\ &< M\epsilon. \end{aligned}$$

Therefore, we obtain that

$$\sum_{i=1}^n f(s_i)\Delta\alpha_i \leq U(P, f\alpha') + M\epsilon,$$

for any choice of points s_i. Therefore, we obtain that

$$U(P, f, \alpha) \leq U(P, f\alpha') + M\epsilon. \tag{35}$$

The same reasoning also shows that

$$U(P, f\alpha') \leq U(P, f, \alpha) + M\epsilon. \tag{36}$$

Combining them, we have $|U(P, f\alpha') - U(P, f, \alpha)| \leq M\epsilon$. Since if P is replaced by a refinement the inequality will still hold (since both (35) and (36) still hold true), we find that

$$|\overline{\int_a^b} f d\alpha - \overline{\int_a^b} f(x)\alpha'(x)dx| \leq M\epsilon.$$

Since ϵ is arbitrary and the upper integrals do not depend on the choice of it, we obtain that $\overline{\int_a^b} f d\alpha = \overline{\int_a^b} f(x)\alpha'(x)dx$. For the lower integrals one can follow the same procedure. This completes the proof, since applying the integrability criterion for f with respect to α is the same as applying the integrability criterion for $f\alpha'$ for the Riemann integral. □

We now proceed to show the change of variable rule.

Theorem 6.22. *Suppose ϕ is a strictly increasing continuous function that maps $[A, B]$ onto $[a, b]$. Suppose that α is a monotonically increasing function on $[a, b]$ and let $f \in \mathcal{R}(\alpha)$. Define the functions*

$$\beta(y) := \alpha(\phi(y)) \tag{37}$$

$$g(y) := f(\phi(y)). \tag{38}$$

Then, $g \in \mathcal{R}(\beta)$ and $\int_A^B g d\beta = \int_a^b f d\alpha$.

Proof. Observe that any partition $P = \{x_0, \ldots, x_n\}$ of $[a, b]$ induces a partition $Q = \{y_0, \ldots, y_n\}$ of $[A, B]$, where each y_i is chosen in such a way that $\phi(y_i) = x_i$, where such y_i exists since ϕ is onto. Notice that Q is a partition because ϕ is strictly monotonic, so $y_0 < y_1 < \cdots < y_n$. Moreover, all partitions of $[A, B]$ are obtained in this way, since we can take $x_i = \phi(y_i)$ for any partition $Q = \{y_0, \ldots, y_n\}$ and obtain a partition P which admits Q as its induced partitions. This construction gives us a bijective correspondence $\Phi : \mathcal{P}([a, b]) \longrightarrow \mathcal{P}([A, B])$ between the class of partitions of $[a, b]$, and that of the partitions of $[A, B]$, as $\Phi(P) = Q$ (and therefore such that $P = \Phi^{-1}(Q)$. In the rest of this proof, P and Q are related through the correspondence Φ just described. Moreover, by definition of g, we have the equality of sets $f([x_{i-1}, x_i]) = g([y_{i-1}, y_i])$. It follows that

$$\begin{aligned} U(Q, g, \beta) &= U(P, f, \alpha) \\ L(Q, g, \beta) &= L(P, f, \alpha). \end{aligned}$$

Since $f \in \mathcal{R}(\alpha)$, given $\epsilon > 0$ we can find P such that $L(P, f, \alpha) - U(P, f, \alpha) < \epsilon$. Therefore, also $L(Q, g, \beta) - U(Q, g, \beta) < \epsilon$ and $g \in \mathcal{R}(\beta)$. In addition, we also have that $\overline{\int_a^b} f d\alpha = \overline{\int_A^B} g d\beta$, both being the suprema over P and Q, respectively, of the sets $U(P, f, \alpha)$ and $U(Q, g, \beta)$. Similarly, $\underline{\int_a^b} f d\alpha = \underline{\int_A^B} g d\beta$. It follows that $\int_A^B g d\beta = \int_a^b f d\alpha$. □

We now consider the relation between integration and differentiation.

Theorem 6.23. *Let $f \in \mathcal{R}$ on $[a, b]$. For every $x \in [a, b]$ define*

$$F(x) = \int_a^x f(t)dt.$$

Then, F is uniformly continuous (and therefore continuous) over $[a, b]$. Moreover, if f is continuous at $x_0 \in [a, b]$, then F is differentiable at x_0 and $F'(x_0) = f(x_0)$.

Proof. By assumption, f is bounded, since $\mathcal{R}$ is defined as a subset of bounded functions. Suppose that $|f(t)| \leq M$ for all $t \in [a, b]$. Consider x, y such that $a \leq x < y \leq b$. Then, we have

$$\begin{aligned} |F(y) - F(x)| &= |\int_a^y f(t)dt - \int_a^x f(t)dt| \\ &= |\int_a^x f(t)dt + \int_x^y f(t)dt - \int_a^x f(t)dt| \\ &= |\int_x^y f(t)dt| \\ &\leq M(y - x). \end{aligned}$$

So, for any chosen $\epsilon > 0$, whenever $y \in (x - \frac{\epsilon}{M}, x + \frac{\epsilon}{M})$ it follows that $|F(y) - F(x)| < \epsilon$. This means that F is continuous at x for any $x \in [a, b]$. In addition, the value $\delta = \frac{\epsilon}{M}$ does not depend on x, but only on ϵ. The function F is therefore uniformly continuous.

Assume now that x_0 is a point of continuity for f. Fix $\epsilon > 0$, and from the fact that $f(t) \longrightarrow f(x_0)$ as $t \longrightarrow x_0$, choose $\delta > 0$ such that

$$|f(t) - f(x_0)| < \epsilon,$$

whenever $|t-x_0|<\delta$, with $t\in[a,b]$. For $s,t\in[a,b]$ with $x_0-\delta<s\le x_0\le t<x+\delta$ we have that

$$\begin{aligned}
&|\frac{F(t)-F(s)}{t-s}-f(x_0)| \\
&= |\frac{F(t)-F(s)}{t-s}-\frac{f(x_0)(t-s)}{t-s}| \\
&= |\frac{F(t)-F(s)}{t-s}-\frac{\int_s^t f(x_0)du}{t-s}| \\
&= |\frac{1}{t-s}[F(t)-F(s)-\int_s^t f(x_0)du]| \\
&= \frac{1}{t-s}|\int_a^t f(u)du-\int_a^s f(u)du-\int_s^t f(x_0)du| \\
&= \frac{1}{t-s}|\int_a^s f(u)du+\int_s^t f(u)du-\int_a^s f(u)du-\int_s^t f(x_0)du| \\
&= \frac{1}{t-s}|\int_s^t [f(u)-f(x_0)]du| \\
&\le \frac{1}{t-s}\epsilon(t-s) \\
&= \epsilon.
\end{aligned}$$

This shows that $F'(x_0)=f(x_0)$, which completes the proof. □

We are now proving the following result, which as its name suggest, is quite important.

Theorem 6.24 (Fundamental Theorem of Calculus)**.** *If $f\in\mathcal{R}$ on $[a,b]$ and if $f=F'$ on $[a,b]$ for some differentiable function F, then*

$$\int_a^b f(x)dx = F(b)-F(a).$$

Proof. Let $\epsilon>0$ be given. Let $P=\{x_0,\dots,x_n\}$ be a partition of $[a,b]$ such that $U(P,f)-L(P,f)<\epsilon$, whose existence is guaranteed by the fact that $f\in\mathcal{R}$. For each subinterval $[x_{i-1},x_i]$ by the Mean Value Theorem we can find points t_i such that $F(x_i)-F(x_{i-1})=F'(t_i)\Delta x_i$. Therefore, we have

$$\begin{aligned}
\sum_{i=1}^n f(t_i)\Delta x_i &= \sum_{i=1}^n F'(t_i)\Delta x_i \\
&= \sum_{i=1}^n F(x_i)-F(x_{i-1}) \\
&= F(b)-F(a).
\end{aligned}$$

Applying Theorem 6.9 (c), we find that

$$\begin{aligned}
|F(b)-F(a)-\int_a^b f(x)dx| &= |\sum_{i=1}^n f(t_i)\Delta x_i-\int_a^b f(x)dx| \\
&\le \epsilon.
\end{aligned}$$

Since the inequality holds for all $\epsilon>0$, while $F(b)-F(a)$ and $\int_a^b f(x)dx$ are independent of it, it follows that $\int_a^b f(x)dx=F(b)-F(a)$. □

The following result is very useful in practice for computations.

Theorem 6.25 (Integration by Parts)**.** *Suppose that F and G are differentiable functions over $[a,b]$, such that $F' = f \in \mathcal{R}$ and $G' = g \in \mathcal{R}$. Then,*

$$\int_a^b F(x)g(x)dx = F(b)G(b) - F(a)G(a) - \int_a^b f(x)G(x)dx.$$

Proof. Since F and G are differentiable, it follows that they are continuous and therefore integrable. We set $H(x) = F(x)G(x)$. Since $H' = F'G + FG' = fG + Fg$, it follows that $H' \in \mathcal{R}$, because F, G, f, g all are in $\mathcal{R}$. Applying Theorem 6.24 to H' we find that

$$\int_a^b H'(x)dx = H(b) - H(a).$$

From $H' = F'G + FG' = fG + Fg$ we therefore find

$$\begin{aligned}\int_a^b f(x)G(x)dx + \int_a^b F(x)g(x)dx &= \int_a^b [f(x)G(x) + F(x)g(x)]dx \\ &= \int_a^b H'(x)dx \\ &= H(b) - H(a) \\ &= F(b)G(b) - F(a)G(a).\end{aligned}$$

This completes the proof. □

Let $f_1, \dots, f_n$ be n real valued functions defined on $[a,b]$. Define $\mathbf{f} = (f_1, \dots, f_n)$. Then, $\mathbf{f} : [a,b] \longrightarrow \mathbb{R}^n$. In this situation, if α is a monotonic bounded function, we can define $\int_a^b \mathbf{f} d\alpha = (\int_a^b f_1 d\alpha, \dots, \int_a^b f_n d\alpha)$, whenever each $f_i \in \mathcal{R}(\alpha)$ in which case we say that $\mathbf{f} \in \mathcal{R}(\alpha)$.

Exercise 6.26. State and prove the Fundamental Theorem of Calculus for vector functions.

Most of the results studied thus far immediately translate to analogous results in the vector valued setup, mutatis mutandis. We consider a result that presents some differences with respect to the case of $n = 1$.

Theorem 6.27. *Let $\mathbf{f} \in \mathcal{R}(\alpha)$. Then also $\|\mathbf{f}\| \in \mathcal{R}(\alpha)$, and*

$$\| \int_a^b \mathbf{f} d\alpha \| \leq \int_a^b \|f\| d\alpha,$$

where $\|(y_1, \dots, y_n)\| = \sqrt{y_1^2 + \cdots + y_n^2}$ is the Euclidean norm of $\mathbb{R}^n$.

Proof. Let $f_1, \dots, f_n$ be the components of $\mathbf{f}$. Since $\mathbf{f} \in \mathcal{R}(\alpha)$, we have that $f_i \in \mathcal{R}(\alpha)$ for all $i = 1, \dots, n$. Moreover, from previous results we know that $f_i^2 \in \mathcal{R}(\alpha)$, as well as $\sum_i f_i^2 \in \mathcal{R}(\alpha)$. Since the function $h(t) = \sqrt{t}$ is continuous. Then, by Theorem 6.13 it follows that $\|f\| \in \mathcal{R}(\alpha)$. We now need to prove the inequality. Let $\mathbf{y} := (y_1, \dots, y_n)$ where $y_i := \int_a^b f_i d\alpha$ for all $i = 1, \dots, n$. By definition, we have $\mathbf{y} = \int_a^b \mathbf{f} d\alpha$. Then we have

$$\begin{aligned}
\|\mathbf{y}\|^2 &= \sum_i y_i^2 \\
&= \sum_i y_i \int_a^b f_i d\alpha \\
&= \sum_i \int_a^b (y_i f_i) d\alpha \\
&= \int_a^b (\sum_i y_i f_i) d\alpha.
\end{aligned}$$

By the Cauchy-Schwarz inequality $\sum_i y_i f_i(t) \leq \|\mathbf{y}\| \|f(t)\|$ for all $t \in [a, b]$. Therefore, Theorem 6.14 shows that

$$\begin{aligned}
\|\mathbf{y}\|^2 &= \int_a^b (\sum_i y_i f_i) d\alpha \\
&\leq \int_a^b \|\mathbf{y}\| \|f\| d\alpha \\
&= \|\mathbf{y}\| \int_a^b \|f\| d\alpha,
\end{aligned}$$

which completes the proof. □

Definition 6.28. A continuous function $\gamma : [a, b] \longrightarrow \mathbb{R}^k$ is called a *curve*. Moreover, if γ is one-to-one, then it is called an *arc*, while if $\gamma(a) = \gamma(b)$ it is called a *closed curve* or a *loop*.

Remark 6.29. Different curves might have the same associated set of points in $\mathbb{R}^k$ (i.e. the image of the curves). This is because curves are functions, rather than just sets of points. Our definition entails not only the path, but also some parametrization of it according to some parameter $\in [a.b]$ which we can think of as being time.

We define now, for a given γ, a correspondence $\Lambda_\gamma : \mathcal{P}([a, b]) \longrightarrow \mathbb{R}_0^+$ from the class of partitions on $[a, b]$ to the nonnegative real numbers defined as follows. For a partition $P = \{x_0, \ldots, x_n\}$ of $[a, b]$

$$\Lambda_\gamma(P) = \sum_{i=1}^n \|\gamma(x_i) - \gamma(x_{i-1})\|.$$

The number $\Lambda_\gamma(P)$ is therefore the length of a polygonal curve obtained by selecting n points on the curve γ. The selected points, $\{\gamma(x_i)\}$ are the vertices of the polygon. The idea is that such length is just an approximation to the actual length of the curve γ, and therefore as our partition becomes finer and finer we obtain approximations of increased accuracy. According to this reasoning, the following seems an intuitively meaningful definition.

Definition 6.30. Let γ be a curve. Then, we set the length of γ, which we indicate by the symbol $\Lambda(\gamma)$ as $\Lambda(\gamma) := \sup_{P \in \mathcal{P}([a,b])} \Lambda_\gamma(P)$. If $\Lambda(\gamma) < +\infty$, we say that γ is a *rectifieble curve*.

We prove now that when a curve is continuously differentiable, we can calculate the length of γ by performing a Riemann integral.

Theorem 6.31. *Let γ be a curve which is differentiable and such that γ' is continuous. Then, γ is recifiable, and*

$$\Lambda(\gamma) = \int_a^b \|\gamma'(t)\|dt.$$

Proof. Given a partition $P = \{x_0, \dots, x_n\}$, we have that

$$\begin{aligned}\|\gamma(x_i) - \gamma(x_{i-1})\| &= \|\int_{x_{i-1}}^{x_i} \gamma'(t)dt\| \\ &\leq \int_{x_{i-1}}^{x_i} \|\gamma'(t)\|dt,\end{aligned}$$

where we have used the Fundamental Theorem of Calculus for vector functions and Theorem 6.27. It follows that $\Lambda_\gamma(P) \leq \int_a^b \|\gamma'(t)\|dt$. Thefore, the supremum as P varies among all partition is also bounded by $\int_a^b \|\gamma'(t)\|dt$, and therefore the curve γ is recifiable. So, we have found that $\Lambda(P) \leq \int_a^b \|\gamma'(t)\|dt$. By showing the converse inequality we complete the proof.

Let $\epsilon > 0$ be given. Since γ' is continuous over a compact, it is uniformly continuous. There exists, therefore, a $\delta > 0$ such that

$$\|\gamma'(s) - \gamma'(t)\| < \epsilon,$$

whenever $|s - t| < \delta$. Let $P = \{x_0, \dots, x_n\}$ be a partition of $[a, b]$ where $\Delta x_i < \delta$ for all i. Then, by taking $s = x_i$ in the previous inequality we obtain that $\|\gamma'(t)\| < \|\gamma'(x_i)\| + \epsilon$, whenever $t \in [x_{i-1}, x_i]$. Therefore, we have

$$\begin{aligned}\int_{x_{i-1}}^{x_i} \|\gamma'(t)\|dt &\leq \int_{x_{i-1}}^{x_i} \|\gamma'(x_i)\|dt + \int_{x_{i-1}}^{x_i} \epsilon dt \\ &= \|\gamma'(x_i)\|\Delta x_i + \epsilon \Delta x_i \\ &= \|\int_a^b [\gamma'(t) - \gamma'(t) + \gamma'(x_i)]dt\| + \epsilon \Delta x_i \\ &\leq \|\int_a^b \gamma'(t)dt\| + \|\int_a^b [\gamma'(x_i) - \gamma'(t)]dt\| + \epsilon \Delta x_i \\ &\leq \|\gamma(x_i) - \gamma(x_{i-1})\| + 2\epsilon \Delta x_i.\end{aligned}$$

Adding all the terms for $i = 1, \dots, n$ together we obtain

$$\begin{aligned}\int_a^b \|\gamma'(t)\|dt &\leq \Lambda_\gamma(P) + 2\epsilon(b - a) \\ &\leq \Lambda(\gamma) + 2\epsilon(b - a).\end{aligned}$$

Since $\epsilon > 0$ was arbitrary, it follows that

$$\int_a^b \|\gamma'(t)\|dt \leq \Lambda(\gamma),$$

which completes the proof. □

7. Sequences and series of functions

We consider now sequences of functions and introduce and discuss their convergence. This is a fundamental problem in all analysis and its applications.

Definition 7.1. Let $\{f_n\}$ be a sequence of real or complex functions defined on a set E. Suppose that $f_n(x)$ is a convergent numerical sequence for all $x \in E$. Then, one can define a function obtained by setting $f(x) := \lim_n f_n(x)$. In this stuation we say that f is the *limit function*, and we also say that convergence happens *pointwise*. Similarly, if the series $\sum_n f_n(x)$ is convergent for all x, then we can define a function obtained by setting $f(x) := \sum_n f(x)$ for all $x \in E$. The function so obtained is called the *sum* of f_n.

The definition of convergence and sum raises the question of whether this is a "good" definition, in the sense that we are interested in knowing whether it has some interesting properties. For instance, one immediate problem one might ask is whether if all functions f_n are continuous, f is continuous as well when it is defined as the limit or the sum of $\{f_n\}$. In general, this is not true, as the following exercise shows.

Exercise 7.2. Let $f_n(x) = \frac{x^2}{(1+x^2)^n}$. Show that the function $f(x) = \sum_n f_n(x)$ is well defined for all x, and show that the function f is not continuous at $x = 0$.

In general, the limit function is not continuous, the sum is not continuous, the integral of the limit does not coincide with the limit of the integrals and so on. In other words, this definition is not really well behaved. This brings us to the following stronger definition.

Definition 7.3. We say that a sequence of functions $\{f_n\}$ *converges uniformly* on E to a function f if for every $\epsilon > 0$ there exists an integer ν such that for any $n > \nu$

$$|f_n(x) - f(x)| < \epsilon,$$

for all $x \in E$. We say that the series $\sum_n f_n$ converges uniformly if the sequence of partial sums convergence uniformly.

Observe that the fundamental difference with our original definition (pointwise) is that while such ν exists for each x, but might depend on x when convergence is pointwise, in the case of uniform convergence ν is independent of x. We have a Cauchy criterion for uniform convergence.

Theorem 7.4. *The sequence of functions $\{f_n\}$ converges uniformly on E if and only if for every $\epsilon > 0$ there exists an integer ν such that for all $n, m > \nu$ we have*

$$|f_n(x) - f_m(x)| < \epsilon, \tag{39}$$

for all $x \in E$.

Proof. Suppose that $\{f_n\}$ converges uniformly to f. Then, given $\epsilon > 0$, we can find ν such that $|f_n(x) - f(x)| < \frac{\epsilon}{2}$ whenever $n > \nu$, for all $x \in E$. It follows that if $n, m > \nu$, then for all $x \in E$ we have

$$\begin{aligned} |f_n(x) - f_m(x)| &\leq |f_n(x) - f(x)| + |f(x) - f_m(x)| \\ &< \epsilon. \end{aligned}$$

Conversely, if 39 holds, from the Cauchy criterion for numerical sequences we find that for all $x \in E$ the sequence $f_n(x)$ converges to some value, which we call $f(x)$ the limit. Then, we have defined a function f to which f_n converges pointwise. We need to show that the convergence is uniform. Fixed $\epsilon > 0$ arbitrary, we find ν such that 39:

$$|f_n(x) - f_m(x)| < \epsilon$$

Then, by taking the limit $n \longrightarrow +\infty$.

$$|f_n(x) - f(x)| \leq \epsilon,$$

which is equivalent to the condition for uniform convergence. ☐

Exercise 7.5. Show the following. Suppose that $lim_n f_n(x) = (x)$ for all $x \in E$. Then, set $M_n = \sup_{x \in E} |f_n(x) - f(x)|$. Then, f_n converges uniformly to f if and only if $M_n \longrightarrow 0$ for $n \longrightarrow +\infty$.

Theorem 7.6. *Suppose $\{f_n\}$ is a sequence of functions defined on E such that*

$$|f_n(x)| \leq M_n,$$

for all $x \in E$ and all $n \in \mathbb{N}$. Then $\sum_n f_n$ converges uniformly if $\sum_n M_n$ converges.

Proof. Showing that $\sum_n f_n$ is uniformly convergent accounts to showing that the sequence of partial sums $\sum_{k=1}^n f_k$ is uniformly convergent. Applying Cauchy criterion for the uniform convergence to the partial sums, we need to find for every $\epsilon > 0$ a ν such that $n, m > \nu$ implies

$$|\sum_{k=1}^{n} f_k(x) - \sum_{q=1}^{m} f_q(x)| < \epsilon,$$

for all $x \in E$. Then, since $\sum_n M_n$ is a convergent numerical series, we have (by the Cauchy criterion for series of numbers) that we can find ν such that $n, m > \nu$ implies $|\sum_{k=m}^n M_k| < \epsilon$. Then, for $n > m > \nu$ we have

$$\begin{aligned} |\sum_{k=1}^{n} f_k(x) - \sum_{q=1}^{m} f_q(x)| &= |\sum_{k=m}^{n} f_k(x)| \\ &\leq |\sum_{k=m}^{n} M_k| \\ &= \sum_{k=m}^{n} M_k \\ &< \epsilon. \end{aligned}$$

This completes the proof. ☐

Theorem 7.7. *Let $\{f_n\}$ be a sequence of continuous functions uniformly convergent on E to f. Then f is continuous.*

Proof. To prove continuity of f, we need to show that for every $x \in E'$, given $\epsilon > 0$, we can find a $\delta > 0$ such that $|f(t) - f(x)| < \epsilon$ for all $|t - x| < \delta$. Then, let us fix $x \in E'$ arbitrary, and take an $\epsilon > 0$. Since f_n converges uniformly to f, we can find ν such that whenever $n > \nu$ we have

$$|f_n(z) - f(z)| < \frac{\epsilon}{3},$$

for all $z \in E$. Moreover, for fixed $n > \nu$, since f_n is continuous, we can find $\delta > 0$ such that

$$|f_n(t) - f_n(x)| < \frac{\epsilon}{3},$$

whenever $|t-x|<\delta$. Then, with such choice of n and δ, whenever $|t-x|<\delta$ holds we find

$$\begin{aligned} |f(t)-f(x)| &= |f(t)-f_n(t)+f_n(t)-f_n(x)+f_n(x)-f(x)| \\ &\leq |f(t)-f_n(t)|+|f_n(t)-f_n(x)|+|f_n(x)-f(x)| \\ &< \frac{\epsilon}{3}+\frac{\epsilon}{3}+\frac{\epsilon}{3} \\ &= \epsilon. \end{aligned}$$

This shows continuity of f at x. Since x was arbitrary, it follows that f is continuous. □

We consider now a case when pointwise convergence can be strengthened to uniform convergence.

Theorem 7.8. *Let K be compact. Suppose that $\{f_n\}$ is a sequence of continuous functions converging to a continuous function f pointwise. Suppose that $f_{n+1}(x) \leq f_n(x)$ for all $n \in \mathbb{N}$ and all $x \in K$. Then f_n converges to f uniformly.*

Proof. We define the sequence $g_n = |f_n - f| = f_n - f$ and show that g_n converges to 0 uniformly, where 0 here means the constant function with value zero (uniform convergence is a property that refers to functions!). Observe that we can remove the absolute value beacause f_n is a decreasing sequence, so $f_n(x) \geq f(x)$ for each x. We have that $g_n \longrightarrow 0$ pointwise (again zero here is the zero function), and $g_n \geq g_{n+1}$ by assumption on f_n and f. Let $\epsilon > 0$ be given. Let K_n be the set of $x \in K$ such that $g_n \geq \epsilon$. Since g_n is continuous, K_n is closed since it is the preimage of a closed set under a continuous map. From Theorem 3.37 we have that K_n is compact. Since $g_n(x) \geq g_{n+1}$, it follows that $K_{n+1} \subset K_n$ since if $x \in K_{n+1}$ then $g_n(x) \geq g_{n+1}(x) \geq \epsilon$ shows that $x \in K_n$ as well.

Observe that if we can show that $K_n = \emptyset$ for some n, then we would complete the proof, since it would follow that $g_n(x) < \epsilon$ for all $x \in K$, which implies uniform convergence as follows: Using $g_{n+1}(x) \leq g_n(x)$, for all $m > n$ we would have that $g_m \leq g_n(x) < \epsilon$ as well.

Suppose by way of contradiction that $K_n \neq \emptyset$ for all $n \in \mathbb{N}$. Notice that since $K_{n+1} \subset K_n$, any finite collection of nonempty sets K_n has nontrivial intersection. Therefore, $\{K_n\}$ satisfies the hypotheses of Theorem 3.39, which gives us $\bigcap K_n \neq \emptyset$. Fix $x \in K$. Since $g_n(x) \longrightarrow 0$, it follows that $g_m(x) < \epsilon$ for m large enough, and therefore such x is not in K_m, which also implies $x \notin \bigcap K_n$. This reasoning can be applied to each $x \in K$, which shows that $\bigcap K_n$ is the empty set. This contradiction shows that one of the K_n must be the empty set. As observed above, this completes the proof. □

Definition 7.9. Let X be a metric space, and let $\mathcal{C}(X)$ denote the set of all real (or complex) valued functions which are continuous and bounded on X. We define a *norm* on f, which is given by

$$\|f\| := \sup_{x \in X} |f(x)|.$$

Exercise 7.10. Show the following properties of $\|f\|$:

- $\|f\| \geq 0$, and $\|f\| = 0$ if and only if $f(x) = 0$ for all $x \in X$.
- If $k \in \mathbb{R}$ (or $\mathbb{C}$), we have that $\|kf\| = |k|\|f\|$.
- It holds: $\|f + h\| \leq \|f\| + \|h\|$.
- By defining $d(f, h) := \|f - g\|$ we introduce a metric over $\mathcal{C}(X)$.
- A sequence $\{f_n\}$ converges to f in $\mathcal{C}(X)$ if and only if f_n converges uniformly to f.

Theorem 7.11. *The space $\mathcal{C}(X)$ with the metric d defined above is a complete metric space.*

Proof. Let $\{f_n\}$ be a Cauchy sequence in $\mathcal{C}(X)$. Then, this means that for all $\epsilon > 0$ there exists $\nu \in \mathbb{N}$ such that for $n, m > \nu$ implies $\|f_n - f_m\| < \epsilon$. Since convergence in the norm is the same as uniform convergence by Exercise 7.10, Theorem 7.4 shows that there exists a function f to which $\{f_n\}$ converges uniformly. Since each f_n is continuous (because all elements of $\mathcal{C}(X)$ are so) Theorem 7.7 shows that f is continuous as well. To complete the proof, we need to show that f is also bounded, which would imply that $f \in \mathcal{C}(X)$. it is clear that f is bounded since we can find (by convergence of $\{f_n\}$ to f) a natural number n such that $\|f - f_n(x)\| < 1$, which implies by definition of $\| \bullet \|$ that $|f(x) - f_n(x)| < 1$ for all $x \in X$. □

We now consider the relation between uniform convergence and integrability.

Theorem 7.12. *Let α be monotonically increasing on $[a, b]$. Suppose $f_n \in \mathcal{R}(\alpha)$ for all $n \in \mathbb{N}$, and $f_n \longrightarrow f$ uniformly. Then $f \in \mathcal{R}(\alpha)$. Moreover, we have*

$$\int_a^b f d\alpha \quad = \quad \lim_n \int_a^b f_n d\alpha.$$

In particular this means that the limit on the right hand side exists.

Proof. We define $\epsilon_n := \sup_{x\in[a,b]} |f_n(x) - f(x)|$. It follows that for all $x \in [a, b]$ we have $f_n(x) - \epsilon_n \leq f(x) \leq f_n(x) + \epsilon_n$. These inequalities imply that for each $n \in \mathbb{N}$ we have

$$\int_a^b f d\alpha - \epsilon_n(\alpha(b) - \alpha(a)) = \int_a^b (f_n - \epsilon_n) d\alpha \leq \underline{\int_a^b} f d\alpha \tag{40}$$

$$\leq \overline{\int_a^b} f d\alpha \leq \int_a^b (f_n + \epsilon_n) d\alpha = \int_a^b f d\alpha + \epsilon_n(\alpha(b) - \alpha(a)). \tag{41}$$

We therefore obtain:

$$0 \leq \overline{\int_a^b} f d\alpha - \underline{\int_a^b} f d\alpha \leq 2\epsilon_n(\alpha(b) - \alpha(a)). \tag{42}$$

Since f_n converges to f, $\epsilon_n \longrightarrow 0$ as $n \longrightarrow +\infty$. Therefore, $\overline{\int_a^b} f d\alpha - \underline{\int_a^b} f d\alpha$ can be made smaller than any $\epsilon > 0$. This means that $\overline{\int_a^b} f d\alpha = \underline{\int_a^b} f d\alpha$, and therefore $f \in \mathcal{R}(\alpha)$. Applying again (40) we find that

$$|\int_a^b f_n d\alpha - \int_a^b f d\alpha| \quad \leq \quad \epsilon_n(\alpha(b) - \alpha(a)).$$

As $\epsilon_n \longrightarrow 0$, we see that $\int_a^b f d\alpha = \lim_n \int_a^b f_n d\alpha$ as stated. □

In particular we obtain the following.

Corollary 7.13. *If $f_n \in \mathcal{R}(\alpha)$ and if $f(x) = \sum_{n=1}^{\infty} f_n(x)$ where the series converges uniformly over $[a, b]$, we have that $f \in \mathcal{R}(\alpha)$ and*

$$\int_a^b f d\alpha \quad = \quad \sum_{n=1}^{\infty} \int_a^b f_n d\alpha.$$

We will now consider the relation between uniform convergence and differentiation.

Theorem 7.14. *Suppose that* $\{f_n\}$ *is a sequence of differentiable functions on* $[a,b]$ *such that* $\{f(x_0)\}$ *converging for some* x_0*. Suppose that* $\{f'_n\}$ *converges uniformly on* $[a,b]$*. Then,* $\{f_n\}$ *converges uniformly on* $[a,b]$ *to a function* f*, and*

$$f'(x) = \lim_n f'_n(x),$$

for all $x \in [a,b]$*.*

Proof. Let $\epsilon > 0$ be given. For $\nu \in \mathbb{N}$ large enough, we have that $n, m > \nu$ implies

$$|f_n(x_0) - f_m(x_0)| < \frac{\epsilon}{2},$$

and

$$|f'_n(t) - f'_m(t)| < \frac{\epsilon}{2(b-a)},$$

for all $t \in [a,b]$. We now apply the Mean Value Theorem to the function $f_n - f_m$. This gives us that for any pair of points $x, t \in [a,b]$ we can find $c \in (x,t)$ such that

$$\begin{aligned} |f_n(x) - f_m(x) - f_n(t) + f_m(t)| &= |x-t||f'_n(c) - f'_m(c)| \\ &\leq |x-t|\frac{\epsilon}{2(b-a)} \\ &\leq \frac{\epsilon}{2}, \end{aligned} \tag{43}$$

whenever $n, m > \nu$. It follows that

$$\begin{aligned} |f_n(x) - f_m(x)| &\leq |f_n(x) - f_m(x) - f_n(x_0) + f_m(x_0)| + |f_n(x_0) - f_m(x_0)| \\ &< \frac{\epsilon}{2} + \frac{\epsilon}{2} \\ &= \epsilon, \end{aligned}$$

for all $x \in [a,b]$, from which we deduce that f_n converges uniformly on $[a,b]$. Let f be the function to which f_n converges. We need to show that $f'(x) = \lim_n f'_n(x)$ for all x. Let $x \in [a,b]$ be fixed and arbitrary. We define for $t \neq x$ and $t \in [a,b]$

$$\phi_n(t) := \frac{f_n(t) - f_n(x)}{t - x},$$

and

$$\phi(t) := \frac{f(t) - f(x)}{t - x}.$$

Then, we have

$$\lim_{t \to x} \phi_n(t) = f'_n(x), \tag{44}$$

by definition. From 43 we have

$$|\phi_n(t) - \phi_m(t)| \leq \frac{\epsilon}{2(b-a)}, \tag{45}$$

which shows uniform convergence of ϕ_n for $t \neq x$. Since $\{f_n\}$ converges to $\{f\}$, we obtain that $\phi_n(t) \longrightarrow \phi(t)$ uniformly, for all $t \neq x$. By Theorem 7.7, we obtain that $\lim_{t\to x} \lim_{n\to\infty} f_n(t) = \lim_{n\to\infty} \lim_{t\to x} f_n(t)$. So, applying this result to ϕ_n along with (44) and (45) completes the proof. $\square$

Theorem 7.15. *There exists a real continuous function that is nowhere differentiable.*

Proof. We define $\phi(x) = |x|$ for $x \in [-1, 1]$, and extend its definition to whole $\mathbb{R}$ by requiring the periodicity condition $\phi(x+2) = \phi(x)$. By definition, for all $s, t \in \mathbb{R}$ we have $|\phi(s) - \phi(t)| \leq |s - t|$. This also implies that ϕ is continuous. Define now

$$f(x) \quad := \quad \sum_{n=0}^{\infty} (\frac{3}{4})^n \phi(4^n x).$$

Since ϕ is bounded by 1, we have that $(\frac{3}{4})^n \phi(4^n x) \leq (\frac{3}{4})^n$. The geometric series $\sum (\frac{3}{4})^n$ is convergent, and we can therefore apply Theorem 7.6 to determine that $\sum_{n=0}^{\infty} (\frac{3}{4})^n \phi(4^n x)$ converges uniformly to f. Moreover, since each term in the series is continuous, and the convergence is uniform, it follows that f is continuous by Theorem 7.7.

Let now x be fixed, and let m be an integer. Define δ_m as follows. If $[4^m x, 4^m x + \frac{1}{2}]$ does not contain any integer, set $\delta_m = \frac{1}{2} 4^{-m}$. If $(4^m x - \frac{1}{2}, 4^m x)$ does not contain any integer, then set $\delta_m = -\frac{1}{2} 4^{-m}$. Observe that exactly one of the two options holds true for each m, since the size of the interval $(4^m x - \frac{1}{2}, 4^m x + \frac{1}{2}]$ has size 1, and it therefore contains exactly one integer which will lie in one of the two cases. Define

$$\gamma_n := \frac{\phi(4^n(x + \delta_m)) - \phi(4^n x)}{\delta_m}.$$

If $n > m$, it follows that $4^n(x + \delta_m)$ differs from $4^n x$ by a multiple of 2, and therefore, by the periodicity of ϕ we have that $\gamma_n = 0$. When $n \leq m$, we can apply the inequality $|\phi(s) - \phi(t)| \leq |s - t|$ to determine that $|\gamma_n| \leq 4^n$. Observe also that $|\gamma_m| = 4^m$. We then have

$$\begin{aligned} |\frac{f(x + \delta_m) - f(x)}{\delta_m}| &= |\sum_{n=0}^{\infty} (\frac{3}{4})^n \gamma_n| \\ &= |\sum_{n=0}^{m} (\frac{3}{4})^n \gamma_n| \\ &\geq 3^m - \sum_{n=0}^{m-1} 3^n \\ &= \frac{1}{2}(3^m + 1). \end{aligned}$$

This shows that f is not differentiable at x, and since x was arbitrary, it completes the proof. □

Definition 7.16. Let $\{f_n\}$ be a family of functions defined on some set E. We say that $\{f_n\}$ is *pointwise bounded* if there is a function $\phi : E \longrightarrow \mathbb{R}^+$ such that

$$|f_n(x)| \quad < \quad \phi(x),$$

for all $x \in E$ and all $n \in \mathbb{N}$. We say that $\{f_n\}$ is *uniformly bounded* if in the previous situation $\phi(x) = M$ for some number M. In other words, if it holds

$$|f_n(x)| \quad < \quad M,$$

for all $x \in E$ and $n \in \mathbb{N}$.

Two natural questions related to boundedness of sequences of functions regard convergence of subsequences. One might ask whether any sequence admits a pointwise converging subsequence. In addition, one might also ask whether a converging sequence admits a uniformly convergent subsequence. The following result will prove useful.

Theorem 7.17. *If $\{f_n\}$ is a pointwise bounded sequence of functions on a countable set E, then $\{f_n\}$ admits a subsequence $\{f_{n_k}\}$ such that $f_{n_k}(x)$ converges for all $x \in E$.*

Proof. Since E is countable, we can arrange the elements of E in a sequence $E = \{x_1, x_2, \ldots\}$. Since $f_n(x_1)$ is bounded by hypothesis, we can find a convergent subsequence $f_{n_k}(x_1)$ which we denote for ease of notation $f_{n_k}(x_1) = f_{1,k}(x_1)$. Moreover, we put $S_1 := f_{1,k}$. We now construct a sequence of sets S_n inductively. The base of induction is already given by S_1 contructed above. Assume that $S_n = \{f_{n,1}, \ldots, f_{n,m}, \ldots\}$ is constructed for some n. Since the initial sequence $\{f_n\}$ is bounded pointwise, we have that $f_{n,k}(x_{n+1})$ is bounded as $k \longrightarrow \infty$ (and n is kept fixed). Then, we can extrapolate a convergent subsequence which we denote by $f_{n+1,k}(x_{n+1})$. Then we set $S_{n+1} = \{f_{n+1,1}, \ldots, f_{n+1,k}, \ldots\}$. Moreover, given $f_{n+1,1}$, there are only finitely many functions in S_n that could correspond to functions $f_{n,k}$ in S_n that were on the left of $f_{n+1,1}$ in S_n. We remove them. For $f_{n+1,2}$ we perform the same procedure and continue in this way for all $f_{n+1,k}$ so that we prune down S_{n+1} to a subsequence of functions were the order of the functions did not change between S_n and S_{n+1}, and still $f_{n+1,k}(x_{n+1})$ is still a convergent subsequence. By induction, we construct S_n for all $n \in \mathbb{N}$. The sequence of sets $\{S_n\}$ so constructed satisfies the following properties.

(1) S_{n+1} is a subsequence of S_n for al $n \in \mathbb{N}$.
(2) For each $n \in \mathbb{N}$ we have that $\{f_{n,k}(x_n)\}$ is convergent as $k \longrightarrow \infty$.
(3) In passing from S_n to S_{n+1} the order of the functions is not changed.

Let us now consider the "diagonal" sequence $S = \{f_{n,n}\}$. By property (3), we have that only finitely many values of n can exist such that $f_{n,n}$ is not in one of the S_n. By property (2), it follows that for any $x_i \in E$, the sequence $f_{n,n}(x_i)$ is convergent. This completes the proof. □

Both questions asked before, however, turn out to have negative answers for more general sets E (i.e. uncountable). We therefore need some stronger condition on boundedness of sequences of functions.

Definition 7.18. A family $\mathcal{F}$ of real (or complex) valued functions defined on a subset E of a metric space X is said to be *equicontinuous* if for every $\epsilon > 0$ there exists a $\delta > 0$ such that

$$|f(x) - f(y)| \quad < \quad \epsilon,$$

for all $x, y \in E$ with $d(x, y) < \delta$ and for all $f \in \mathcal{F}$.

Remark 7.19. Observe that if $\mathcal{F}$ is an equicontinuous family of functions and $f \in \mathcal{F}$, then f is uniformly continuous. So, uniform continuity says that δ does not depend on x and y, and equicontinuity says that in addition to that, δ does not even depend on which $f \in \mathcal{F}$ is taken.

The following result shows a case in which equicontinuous families naturally arise.

Theorem 7.20. *Let K be a compact in a metric space, and let $\{f_n\} \subset \mathcal{C}(K)$ be a uniformly convergent sequence. Then $\{fn\}$ is equicontinuous.*

Proof. Let $\epsilon > 0$ be given. From uniform convergence and Cauchy criterion, we have that for all $n, m > \nu$, for some $\nu \in \mathbb{N}$, it holds

$$\|f_n - f_m\| \quad < \quad \epsilon.$$

For $i = 1, \ldots, \nu + 1$, f_i is a continuous function on a compact, and therefore uniformly continuous. We can therefore find a $\delta_i > 0$ such that

$$|f_i(x) - f_i(y) \quad < \quad \epsilon,$$

whenever $d(x,y) < \delta_i$. Upon choosing $\delta \le \min\{\delta_i\}$ we find that $|f_i(x) - f_i(y) < \epsilon$ holds for all $i = 1, \ldots, \nu+1$ whenever $d(x,y) < \delta$. For $n > \nu$ and $d(x,y) < \delta$ we have

$$\begin{aligned} |f_n(x) - f_n(y)| &\le |f_n(x) - f_{\nu+1}(x)| + |f_{\nu+1}(x) - f_{\nu+1}(y)| + |f_{\nu+1}(y) - f_n(y)| \\ &< 3\epsilon. \end{aligned}$$

So, for all $n \in \mathbb{N}$, we have that $|f_n(x) - f_n(y)| < 3\epsilon$ whenever $d(x,y) < \delta$. This completes the proof. □

Exercise 7.21. Let K be compact in a metric space. Show that there is a countable E which is dense in K.

Theorem 7.22 (Ascoli-Arzelá)**.** *Let K be compact, and $f_n \in \mathcal{C}(X)$ be an equicontinuous sequence of pointwise bounded functions. Then,*

(i) *$\{f_n\}$ is uniformly bounded on K;*
(ii) *$\{f_n\}$ contains a uniformly convergent subsequence.*

Proof. We prove first that $\{f_n\}$ is uniformly bounded on K. Using equicontinuity for $\{f_n\}$, given $\epsilon > 0$ we can find $\delta > 0$ such that

$$|f_n(x) - f_n(y)| < \epsilon,$$

for all $n \in \mathbb{N}$, whenever $d(x,y) < \delta$. We can cover K with finitely many balls $B(x_i, \delta)$, centered at $x_i \in \mathbb{K}$, with $i = 1, \ldots, k$, since K is compact. Since $\{f_n\}$ is pointwise bounded, we can find M_i, for each $i = 1, \ldots, k$, such that

$$|f_n(x_i)| < M_i,$$

for all $n \in \mathbb{N}$. Let us choose $M \ge \max_i\{M_i\}$. Then, for $x \in K$, we can find x_i such that $d(x, x_i) < \delta$, from which we obtain that $|f_n(x) - f_n(x_i) < \epsilon$, and therefore $|f_n(x) < M + \epsilon$ for all $n \in \mathbb{N}$. Since x was arbitrary, this completes the proof of (i).

We now prove (ii). Be Exercise 7.21, we can find a countable set E dense in K. Applying Theorem 7.17, we can find a subsequence $\{f_{n_k}\}$ such that $\{f_{n_k}(x)\}$ converges for each $x \in E$. To get rid of the double index, we define a sequence $g_k := f_{n_k}$. By proving that g_i converges uniformly, we complete the proof of (ii). Let $\epsilon > 0$ be given, and choose $\delta > 0$ as in the proof of (i). Since E is dense in K, we can cover K by balls $B(x,\delta)$ where $x \in E$. Due to compactness of K, we can extract a finite subcovering $B(x_r, \delta)$, with $r = 1, \ldots, s$. By construction of g_i, we have that there exists $\nu \in \mathbb{N}$ such that

$$|g_i(x_r) - g_j(x_r)| < \epsilon,$$

for all $i, j > \nu$ and all $r = 1, \ldots, s$. For an arbitrary $x \in K$, there exists r such that $x \in B(x_r, \delta)$, and therefore

$$|g_i(x) - g_i(x_r)| < \epsilon,$$

for all $i \in \mathbb{N}$. Then we have

$$\begin{aligned} |g_i(x) - g_j(x)| &\le |g_i(x) - g_i(x_r)| + |g_i(x_r) - g_j(x_r)| + |g_j(x_r) - g_j(x)| \\ &< 3\epsilon, \end{aligned}$$

whenever $i, j > \nu$. This shows uniform convergence, and completes the proof. □

Theorem 7.23 (Weierstrass)**.** *Let f be a real or complex continuous function on $[a, b]$. Then there exists a sequence of polynomials P_n such that*

$$\lim P_n(x) = f(x),$$

uniformly on $[a, b]$. When f is real, P_n may be taken to be real.

Proof. The two cases are substantially the same, so we will just consider the real case. Since the interval $[a, b]$ can be transformed affinely into $[0, 1]$, we can assume that $[a, b] = [0, 1]$. Moreover, we can assume that $f(0) = f(1) = 0$, since if this is not the case, we can consider the function

$$g(x) = f(x) - f(0) - x[f(1) - f(0)]$$

satisfies the property that $g(0) = g(1) = 0$, and if we find P_n that converges to g uniformly on $[0, 1]$, then $f(x) = g(x) + f(0) + x[f(1) - f(0)]$ is the unoform limit of the sequence of polynomials $\hat{P}_n(x) = P_n(x) + f(0) + x[f(1) - f(0)]$. We extend f to the whole $\mathbb{R}$ by setting $f(x) = 0$ for all $x \in \mathbb{R} - [0, 1]$. Since $f(0) = f(1) = 0$, it follows that f so defined is uniformly continuous over the whole $\mathbb{R}$. Set

$$Q_n(x) = c_n(1 - x^2)^n,$$

for $n \in \mathbb{N}$, where c_n is given by $c_n = \frac{1}{\int_{-1}^{1} Q_n(x)dx}$. We have

$$\begin{aligned} \int_{-1}^{1} (1 - x^2)^n dx &= 2\int_0^1 (1 - x^2)^n dx \\ &\geq \int_0^{\frac{1}{\sqrt{n}}} (1 - x^2)^n dx \\ &\geq \int_0^{\frac{1}{\sqrt{n}}} (1 - nx^2) dx \\ &= \frac{4}{3\sqrt{n}} \\ &> \frac{1}{\sqrt{n}}, \end{aligned}$$

where in the first inequality we have used the fact that $Q_n(x) = (1 - x^2)^n$ is even and we can therefore calculate the integral on half of the domain, then we used the fact that the integral of a nonnegative function is monotonic with respect to the size of the interval (so, smaller interval makes the integral smaller), and for the third inequality we have used the fact that the function $h(x) = (1 - x^2)^n - (1 - nx^2)$ has nonnegative derivative on $[0, \frac{1}{\sqrt{n}}]$ and satisfies $h(0) = 0$ (therefore it grows from zero, and therefore it is positive in the given interval). Since $\int_{-1}^{1}(1 - x^2)^n dx > \frac{1}{\sqrt{n}}$, by definition of c_n it follows that $c_n < \sqrt{n}$.

For any given $\delta > 0$, by definition of $Q_n(x)$, we have that for all x such that $\delta \leq |x| \leq 1$, the inequality

$$Q_n(x) \leq \sqrt{n}(1 - x^2)^n.$$

Since the right hand side converges to zero, as n goes to $+\infty$, we find that $Q_n(x) \longrightarrow 0$ uniformly on $[-1, -\delta] \bigcup [\delta, 1]$.

We define $P_n(x) = \int_{-1}^{1} f(x+t)Q_n(t)dt$. Since f is zero outside the interval $[0,1]$, it follows that $f(x+t) = 0$ when $t \notin [-x, 1-x]$. Therefore, we find that

$$\begin{aligned} P_n(x) &= \int_{-x}^{1-x} f(x+t)Q_n(t)dt \\ &= \int_0^1 f(s)Q_n(s-x)ds, \end{aligned}$$

where in the last line we have used the change of variable $s = x + t$ to transform the integral. The last integral is indeed a polynomial in x, since x appears inside Q_n which is a polynomial, and the variable s is integrated over. By uniform continuity of f, given $\frac{\epsilon}{2}$, we can find $\delta > 0$ such that

$$|f(x) - f(y)| < \frac{\epsilon}{2},$$

whenever $|x - y| < \delta$. Put $M = \sup_{x\in\mathbb{R}} |f(x)|$. The latter is finite because $|f|$ has a supremum on $[0,1]$ being continuous over a compact, and we have extended f by giving it value zero outside of $[0,1]$. Then, we have

$$\begin{aligned} |P_n(x) - f(x)| &= |\int_{-1}^{1} f(x+t)Q_n(t)dt - f(x)\int_{-1}^{1} Q_n(x)dx| \\ &= |\int_{-1}^{1} [f(x+t) - f(x)]Q_n(t)dt| \\ &\leq \int_{-1}^{1} |[f(x+t) - f(x)]Q_n(t)|dt \\ &= \int_{-1}^{1} |f(x+t) - f(x)|Q_n(t)dt \\ &= \int_{-1}^{-\delta} |f(x+t) - f(x)|Q_n(t)dt + \int_{-\delta}^{\delta} |f(x+t) - f(x)|Q_n(t)dt \\ &\quad + \int_{\delta}^{1} |f(x+t) - f(x)|Q_n(t)dt \\ &\leq 2M\int_{-1}^{-\delta} \sqrt{n}(1-\delta^2)^n dt + \frac{\epsilon}{2} + 2M\int_{\delta}^{1} \sqrt{n}(1-\delta^2)^n dt \\ &\leq 4M\sqrt{n}(1-\delta^2)^n + \frac{\epsilon}{2}, \end{aligned}$$

which can uniformly be made smaller than ϵ for a choice of n large enough (since $\sqrt{n}(1-\delta^2)^n \longrightarrow 0$). This completes the proof. □

Corollary 7.24. *For every interval $[-a, a]$, there is a sequence of real polynomials P_n such that $P_n(0) = 0$ for all n, and $P_n \longrightarrow |x|$ uniformly.*

Proof. Since $f(x) = |x|$ is continuous on any interval $[-a, a]$, Theorem 7.23 gives that there is a sequence of polynomials P_n^* which converges to f uniformly. Therefore, we have that $P_n^*(0) \longrightarrow 0$. Setting $P_n(x) := P_n^*(x) - P_n(0)$ gives the result, since $P_n(0) = 0$, and $\lim_n P_n(x) = \lim_n[P_n^*(x) - P_n(0)] = \lim_n P_n^*(x) = f(x)$, where the convergence is uniform. □

Definition 7.25. Let $\mathcal{A}$ be a vector space over $\mathbb{R}$. Then, $\mathcal{A}$ is said to be an *algebra* if the following conditions are satisfied.

- There is an associative operation $\cdot : \mathcal{A} \times \mathcal{A} \longrightarrow \mathcal{A}$.

- Right and lef distributivity with respect to + hold: $(x+y)\cdot z = x\cdot z + y\cdot z$ and $x\cdot(y+z) = x\cdot y + x\cdot z$, for all $x, y, z \in \mathcal{A}$.
- Compatibility with scalars holds: $(ax)\cdot(by) = (ab)(x\cdot y)$, where $a, b \in \mathbb{R}$ and $x, y \in \mathcal{A}$.

Lemma 7.26. *The set of real valued functions over $E \subset \mathbb{R}$ is an algebra.*

Exercise 7.27. Prove Lemma 7.26.

Definition 7.28. In general, when we have a vector space of functions which is closed under multiplication, we will say that this is an algebra of functions.

Exercise 7.29. Show that the set of polynomials is an algebra of functions. Does the same hold true for the set of polynomials of degree less than or equal to n?

Definition 7.30. Let $\mathcal{A}$ be an algebra of real functions over E. If for any sequence of functions $\{f_n\} \subset \mathcal{A}$ which converges uniformly to f on E it follows that $f \in \mathcal{A}$, then $\mathcal{A}$ is called a *uniformly closed algebra.*

Given an algebra of functions $\mathcal{A}$, let $\mathcal{B}$ be the set of functions consisting of uniform limits of sequences in $\mathcal{A}$. Then $\mathcal{B}$ is called the *uniform closure of* $\mathcal{A}$.

Theorem 7.31. *Let $\mathcal{B}$ be the uniform closure of an algebra $\mathcal{A}$ of bounded functions. Then, $\mathcal{B}$ is uniformly closed.*

Proof. The proof is left to the reader as an exercise. □

Definition 7.32. Let $\mathcal{A}$ be a family of functions on a set E. Then, $\mathcal{A}$ is said to *separate points* on E if for every pair of points $x_1, x_2 \in E$, we can find a function $f \in \mathcal{A}$ such that $f(x_1) \neq f(x_2)$.

Moreover, if for each $x \in E$ we can find an element $g \in \mathcal{A}$ such that $g(x) \neq 0$, we say that $\mathcal{A}$ *vanishes at no point* of E.

Theorem 7.33. *Let $\mathcal{A}$ be an algebra of functions which separates points in E, and vanishes at no point of E. Suppose $x_1, x_2 \in E$ are distinct, and let c_1 and c_2 be constants. Then, $\mathcal{A}$ contains a function f such that $f(x_1) = c_1$ and $f(x_2) = c_2$.*

Proof. By assumption, we can find a function g such that $g(x_1) \neq g(x_2)$ (since $\mathcal{A}$ separates points), and we can find functions h and k such that $h(x_1) \neq 0$ and $g(x_2) \neq 0$ (since $\mathcal{A}$ vanishes at no point). Define now

$$\begin{aligned} u &= gk - g(x_1)k, \\ v &= gh - g(x_2)h. \end{aligned}$$

Since $\mathcal{A}$ is an algebra, $u, v \in \mathcal{A}$ as well. Moreover, $u(x_1) = v(x_2) = 0$, and $u(x_2) \neq 0$ and $v(x_1) \neq 0$. It follows that

$$f = \frac{c_1 v}{v(x_1)} + \frac{c_2 u}{u(x_2)}$$

is the function we wanted to construct. □

We now proceed to proving the main result of this section, which is a generalization of Weierstrass' Thoerem 7.23

Theorem 7.34 (Stone-Weierstrass)**.** *Let $\mathcal{A}$ be an algebra of real continuous functions on a compact K. Suppose that $\mathcal{A}$ separates points on K and vanishes at no point of K. Then, the uniform closure $\mathcal{B}$ of $\mathcal{A}$ consists of all continuous functions on K. Therefore, $\mathcal{A}$ is dense in $C(K)$.*

Proof. The proof will consists of four steps.

Step 1. We shall prove that if $f \in \mathcal{B}$, then $|f| \in \mathcal{B}$ as well. To this purpose, let $M := \sup_{x\in K} |f(x)|$. Observe that since $\mathcal{B}$ is the uniform closure of $\mathcal{A}$, there exists a sequence of continuous functions in $\mathcal{A}$ which converges uniformly to f. Since each function in the sequence is continuous, and therefore it is bounded because K is compact, we have that f is also bounded, and therefore M is not infinite. Now, let $\epsilon > 0$ be given. By Corollary 7.24, there exists a sequence of polynomials uniformly convergent to the function $|x|$ on $[-M, M]$. Then, in correspondence of the given $\epsilon > 0$ we can find a polynomial $P(y) = \sum_{i=0}^{k} a_i y^i$ such that

$$|P(y) - |y|| \quad < \quad \epsilon,$$

for all $y \in [-M, M]$. We now consider the function $g(x) = \sum_{i=0}^{k} a_i f(x)^i$, which is obtained from P by replacing $y = f(x)$. Then, we have that for all $x \in K$, $f(x) \in [-M, M]$ and

$$|g(x) - |f(x)|| \quad < \quad \epsilon.$$

Moreover, since $\mathcal{B}$ is an algebra, it follows that $g(x) = \sum_{i=0}^{k} a_i f(x)^i$ is in $\mathcal{B}$ as well, since it is obtained by using only algebra operations (linear combinations and products of algebra elements). This shows that $|f|$ is in the uniform closure of $\mathcal{B}$. However, since $\mathcal{B}$ is uniformly closed, it follows that $|f|$ is in $\mathcal{B}$ as well, completing the first step.

Step 2. For f, g functions, we define the new functions

$$\max(f,g)(x) := \Big\{ f(x) \quad f(x) \geq g(x) g(x) \quad g(x) \geq f(x),$$

and

$$\min(f,g)(x) := \Big\{ f(x) \quad f(x) \leq g(x) g(x) \quad g(x) \leq f(x).$$

We want to prove that if $f, g \in \mathcal{B}$, then $\max(f,g), \min(f,g) \in \mathcal{B}$. This follows directly by the following facs:

$$\begin{aligned} \max(f,g)(x) &= \frac{f(x)+g(x)}{2} + \frac{|f(x)-g(x)|}{2}, \\ \min(f,g)(x) &= \frac{f(x)+g(x)}{2} - \frac{|f(x)-g(x)|}{2}, \end{aligned}$$

and an application of Step 1. Then, following the same process one can show that given $f_1, \ldots, f_n$ functinos in $\mathcal{B}$, we have that $\max(f_1, \ldots, f_n), \min(f_1, \ldots, f_n) \in \mathcal{B}$.

Step 3. We want to prove that given a real continuous function f on K, and given a point $x \in K$, we can find a function $g_x \in \mathcal{B}$ such that $g_x(x) = f(x)$ and

$$g_x(t) > f(t) - \epsilon,$$

for all $t \in K$. Observe that $\mathcal{A}$ satisfies the hypotheses of Theorem 7.33. Then, for each $y \in K$ we can find a function $h_y \in \mathcal{A} \subset \mathcal{B}$ such that $h_y(x) = f(x)$ and $h_y(y) = f(y)$. Since h_y is continuous (any function in $\mathcal{A}$ is) we can find a neighborhood J_y of y such that $h_y(t) > f(t) - \epsilon$ for all $t \in J_y$. Since K is compact, we can obtain a finite subcovering of the J_y, which we denote $J_1, \ldots, J_n$, referring to $y_1, \ldots, y_n$. Then, the function we are seeking is obtained by putting $g = \max(h_1, \ldots, h_n)$, where $h_i := h_{y_i}$. Step 2 guarantees that $g \in \mathcal{B}$.

Step 4. We want to prove that given a real continuous f on K, and given $\epsilon > 0$, we can find a function $h \in \mathcal{B}$ such that

$$|h(x) - f(x)| \quad < \quad \epsilon,$$

for all $x \in K$. Observe that proving this, we would have shown that f is in the uniform closure of $\mathcal{B}$ which, in turn, shows that $f \in \mathcal{B}$ since $\mathcal{B}$ is uniformly closed. In other words, this completes the proof of the theorem.

For each $x \in K$, let g_x denote the function obtained from Step 3. Then, by continuity of g_x, we can find a neighborhood V_x such that

$$g_x(t) \quad < \quad f(t) + \epsilon, \tag{46}$$

for all $t \in V_x$. We can then cover K with V_x neighborhoods, and extract a finite subcovering $V_1, \dots, V_m$, corresponding to points $x_1, \dots, x_m$. Defining $h := \min(g_1, \dots, g_m)$ gives the function with the required properties and, applying Step 2 we see that $h \in \mathcal{B}$ as required. This completes the proof. □

The complex version of the Stone-Weierstrass theorem needs an extra assumption. We need to require that $\mathcal{A}$ be self-adjoint, meaning that whenever $f \in \mathcal{A}$, also its complex conjugate $\bar{f} \in \mathcal{A}$, which is defined by $\bar{f}(x) = \bar{f(x)}$.

Theorem 7.35 (Complex Stone-Weierstrass)**.** *Suppose $\mathcal{A}$ is a self-adjoint algebra of continuous complex functions over a compact K, such that $\mathcal{A}$ separates points on K and does not vanish at any point of K. Then, the uniform closure $\mathcal{B}$ of $\mathcal{A}$ consists of all complex continuous functions on K. Therefore, $\mathcal{A}$ is dense in $C(K)$ (complex continuous functions with uniform norm).*

Proof. Let $\mathcal{A}_{\mathbb{R}}$ denote the set of all real functions in $\mathcal{A}$. Observe that if $f \in \mathcal{A}$, with $f = u + iv$, we have that $\bar{f} \in \mathcal{A}$ by self-adjoint assumption and then $u = \frac{1}{2}[f + \bar{f}] \in \mathcal{A}$ as well, meaning that $u \in \mathcal{A}_{\mathbb{R}}$. If $x_1 \neq x_2$ are two points in K, we can find $f \in \mathcal{A}$ such that $f(x_1) = 0$ and $f(x_2) = 1$ (by Theorem 7.33, complex version). Therefore, $u(x_1) = 0$ and $u(x_2) = 1$, which shows that $\mathcal{A}_{\mathbb{R}}$ separates points on K. If $x \in K$ we can find $g \in \mathcal{A}$ such that $g(x) \neq 0$. There exists a number $\lambda \in \mathbb{C}$ such that $\lambda g(x) > 0$ and therefore the function $f = \lambda g$, has a real part u such that $u(x) > 0$, showing that $\mathcal{A}_{\mathbb{R}}$ does not vanish at any point of K. Therefore, $\mathcal{A}_{\mathbb{R}}$ satisfies the hypotheses of Theorem 7.34, which implies that any function which is real and continuous lies in the uniform closure of $\mathcal{A}_{\mathbb{R}}$, and therefore lies in $\mathcal{B}$. Then, if f is a complex continuous function, we can write it as $f = u + iv$ where u and v are real and continuous and therefore in $\mathcal{B}$. Therefore, $f \in \mathcal{B}$ as well. □

8. Lebesgue Theory

In this section we intrduce the theory of integration of Lebesgue. This is not a complete account, but it introduces some of the main topics and definitions.

Definition 8.1. Let $\mathcal{R}$ be a family of sets. Then $\mathcal{R}$ is said to be a *ring* if $A, B \in \mathcal{R}$ implies that $A \cup B, A - B \in \mathcal{R}$. Observe that since $A \cap B = A - (A - B)$, one also finds that $A \cap B \in \mathcal{R}$. A ring $\mathcal{R}$ is said to be a *σ-ring* if $\bigcup_{n=1}^{\infty} A_n \in \mathcal{R}$ whenever $A_1, \dots, A_n, \dots$ are elements of $\mathcal{R}$. Once again, since $\bigcap_{n=1}^{\infty} A_n = A_1 - \bigcup_{n=1}^{\infty}(A_1 - A_n)$, we see that $\bigcap_{n=1}^{\infty} A_n \in \mathcal{R}$ as well in a σ-ring.

Definition 8.2. Let $\mathcal{R}$ be a σ-ring. We say that $\phi : \mathcal{R} \longrightarrow [-\infty, +\infty]$ is an additive set function if $\phi(A \cup B) = \phi(A) + \phi(B)$ whenever $A \cap B = \emptyset$. We say that ϕ is countably additive if whenever$A_i \cap A_j = \emptyset$ for all $i, j \in \mathbb{N}$ $(i \neq j)$, we have $\phi(\bigcup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} \phi(A_n)$.

Remark 8.3. We will make the assumption that ϕ either cotnains $+\infty$ or $-\infty$ alone, but not both, int its image. So, summing $\phi(A)$ and $\phi(B)$ always makes sense.

The following are useful properties.

Proposition 8.4. *Let ϕ be additive on $\mathcal{R}$. Then*

1. $\phi(\emptyset) = 0$.
2. $\phi(A_1 \cup \cdots \cup A_n) = \phi(A_1) + \cdots + \phi(A_n)$, *whenever* $A_i \cap A_j = \emptyset$ *for* $i, j = 1, \ldots, n$, $i \neq j$.
3. *For* $B \subset A$, *and* $|\phi(B)| < +\infty$, *we have* $\phi(A - B) = \phi(A) - \phi(B)$.
4. $\phi(A \cup B) + \phi(A \cap B) = \phi(A) + \phi(B)$. *This is usually called "modularity".*
5. *If* $\phi(A) \geq 0$ *for all* $A \in \mathcal{R}$, *then* $\phi(B) \leq \phi(BA$ *whenever* $B \subset A$.

Proof. 1. Observe that $\phi(\emptyset) = \phi(\emptyset \cup \emptyset) = \phi(\emptyset) + \phi(\emptyset)$. Subtracting $\phi(\emptyset)$ from both sides shows that $\phi(\emptyset) = 0$.

2. This is just the consecutive application of the additive property to A_1 and $A_2 \cup \cdots \cup A_n$, then to A_2 and $A_3 \cup \cdots \cup A_n$ and so on.

3. Observe that $(A - B) \cap B = \emptyset$. Therefore, using the additivity condition we have $\phi(A) = \phi((A - B) \cup B) = \phi(A - B) + \phi(B)$, which gives the result.

4. We have that $A = (A \cap B) \cup (A - B)$, $B = (A \cap B) \cup (B - A)$ and $A \cup B = (A \cap B) \cup (A - B) \cup (B - A)$, and all unions are disjoint. Therefore, we have that

$$\begin{aligned}\phi(A) + \phi(B) &= \phi(A \cap B) + \phi(A - B) + \phi(A \cap B) + \phi(B - A)\\ \phi(A \cup B) &= \phi(A \cap B) + \phi(A - B) + \phi(B - A).\end{aligned}$$

Comparing the two equalities above, we find that $\phi(A \cup B) + \phi(A \cap B) = \phi(A) + \phi(B)$.

5. Since $(A - B) \cap B = \emptyset$, we have $\phi(A) = \phi((A - B) \cup B) = \phi(A - B) + \phi(B) \geq \phi(B)$. □

Theorem 8.5. *Suppose $\mathcal{R}$ is a σ-ring, and let ϕ be countably additive. Let $A_1 \subset A_2 \subset \cdots \subset A_n \subset \cdots$ be a chain of elements in $\mathcal{R}$. Then, we have*

$$\phi(\bigcup_{n=1}^{\infty} A_n) = \lim \phi(A_n).$$

Proof. Set $B_1 = A_1$, and $B_n = A_n - A_{n-1}$ for all $n \geq 2$. Then we have that $B_i \cap B_j = \emptyset$ whenever $i \neq j$ and $A_n = B_1 \cup \cdots \cup B_n$, so that $\bigcup_{n=1}^{\infty} B_n = \bigcup_{n=1}^{\infty} A_n$. By the countably additivity of ϕ we have that

$$\begin{aligned}\phi(A_n) &= \phi(B_1 \cup \cdots \cup B_n)\\ &= \sum_{i=1}^{n} \phi(B_i),\end{aligned}$$

and also

$$\phi(A) = \sum_{n=1}^{\infty} \phi(B_n).$$

Since $\sum_{n=1}^{\infty} \phi(B_n) = \lim \sum_{i=1}^{n} \phi(B_i) = \lim \phi(A_n)$, the proof is complete. □

We now define the Lebesgue measure.

We will say that $I \subset \mathbb{R}^p$ is a *p-interval*, or a *box*, if I is of the form $I = [a_1, b_1] \times \cdots [a_p, b_p]$. We also allow the intervals whose product gives I to be open or semi-open, so each a_i or b_i might be included or excluded. We allow the case where $a_i = b_i$ for some $i = 1, \ldots, p$.

A set A which is the finite union of boxes is called an *elementary set*. For a box I we define

$$m(I) = \prod_{i=1}^{p} (b_i - a_i).$$

If $A = I_1 \cup \cdots \cup I_n$ is an elementary set with $I_s \cap I_t = \emptyset$ for $s \neq t$, we define

$$m(A) \quad = \quad m(I_1) + \cdots + m(I_n).$$

Define now $\mathcal{E}$ as the class of all elementary sets in $\mathbb{R}^p$.

Proposition 8.6. *We have the following properties:*

1. *$\mathcal{E}$ is a ring, but it is not a σ-ring.*
2. *If $A \in \mathcal{E}$, then A can be written as a finite union of boxes that are mutually disjoint. So, m can be defined on the whole $\mathcal{E}$.*
3. *The definition of m does not depend on the decomposition given in 2. In other words, if A is written as finite union of disjoint boxes in two different ways, the value of m does not change.*
4. *m is additive on $\mathcal{E}$.*

Proof. 1. If $A = \bigcup_{k=1}^n I_k$ and $B = \bigcup_{\ell=1}^m J_\ell$, then $A \cup B = \bigcup_{k=1}^n I_k \bigcup \bigcup_{\ell=1}^m J_\ell$ is the union of $n+m$ boxes, and therefore is in $\mathcal{E}$. We want to show now that $A - B \in \mathcal{E}$ as well. First, note that

$$A - B \quad = \quad (A_1 - B) \cup \cdots \cup (A_n - B).$$

So, we can reduce our proof to showing that $A_i - B \in \mathcal{E}$ for one of the i's. In other words, without loss of generality we can take $n = 1$. Then, observe that $A - B = (A - B_1) \cap \cdots \cap (A - B_m)$. Now, by direct inspection it is clear that $A - B_k$ is a finite union of boxes for all k. We just need to show that the finite intersection of finite union of boxes is again a finite union of boxes. Again without loss of generality we can assume that $m = 2$, and we reduce the step to a single intersection, since one can then itereate the process. This follows now from the distributive property of union and intersection, along the fact that finite intersection of boxes is again a finite union of boxes. To show that $\mathcal{E}$ is not a σ-ring, consider for instance (when $p = 1$) the family of elements of $\mathcal{E}$ given by $A_n = [n, n+1]$. Its union is not an elementary set, as it is not bounded, and therefore cannot be written as a finite union of elementary sets. Similar considerations hold for other values of p.

2. Let $A \in \mathcal{E}$. Say, $A = I_1 \cup \cdots \cup I_n$, where each I_k is a box. We consider I_1 and $I_2 \cup \cdots \cup I_n$. If $I_1 \cap (I_2 \cup \cdots \cup I_n) = \emptyset$, then we consider I_2 and $I_3 \cup \cdots \cup I_n$. If the intersection is not empty, as in the proof of part 1, we can find finitely many boxes $J_1^1, \cdots, J_1^{k_1}$ such that $\bigcup_{i=1}^{k_1} J_1^i = I_1 - (I_2 \cup \cdots \cup I_n)$, and $\bigcup_{i=1}^{k_1} J_1^i \bigcap \bigcup_{j=2}^n I_j = \emptyset$. Then, we write replace I_1 with $J_1^1, \cdots J_1^{k_1}$. We move on to applying the same procedure to I_2 and $I_3 \cup \cdots \cup I_n$, for which we find a finite set of boxes whose union gives $I_2 - (I_3 \cup \cdots \cup I_n)$, and which does not intersect neither $I_3 \cup \cdots \cup I_n$ nor I_1. This process ends in finitely many steps (n at most), after which we are left with boxes J_s^r which are pairwise disjoint, and whose union gives A.

3. Let $A = I_1 \cup \cdots \cup I_n = J_1 \cup \cdots \cup J_m$ be two different decompositions of A into disjoint boxes. Then, we want to show that $m(I_1) + \cdots + m(I_n) = m(J_1) + \cdots + m(J_m)$. We first show that if I is a box, and $I = K_1 \cup \cdots \cup K_r$ is a decomposition of I into subboxes that are pairwise disjoint which are obtained by cutting I horizontally and perpendicularly to the directions of $\mathbb{R}^p$ and go from one side of A to the other, then $m(I) = m(K_1) + \cdots + m(K_r)$. We proceed by induction on the dimension p of $\mathbb{R}^p$. For $p = 1$, this is true because I is a 1-dimensional interval, and $K_1, K_2, \cdots, K_r$ are disjoint intervals which can be rearranged to be consecutive. Then, we have $m(K_1) + \cdots + m(K_r) = b_1 - a_1 + b_2 - a_2 + \cdots + b_r - a_r$ where we must have $b_1 = a_2, b_2 = a_3$ and so on. Therefore, $m(K_1) + \cdots + m(K_r) = b_r - a_1 = b - a$, where $I = [a, b]$. This completes the proof of the base case for induction. Suppose now that the the claims holds true for all values between 1 and some $p \geq 1$. We want to show that it also holds for $p + 1$. Since the cuts that give

$K_1, \cdots, K_r$ from A are all from side to side (and parallel to the cartesian coordinates of $\mathbb{R}^{p+1}$), then we can consider each K_s as consisting of several adjecent boxes which are not stacked on top of each other along the direction $p+1$. We can therefore apply the inductive hypothesis to each K_s to find that $m(K_s) = m(K_s^1) + \cdots + m(K_s^d)$. Then, we can again apply the inductive hypothesis on each slice to show that $m(I) = m(K_1) + \cdots + m(K_r)$, which completes the proof of the claim. To finish the proof of 3, now observe that we can always subdivide each $I_1, \cdots, I_n$ and $J_1, \cdots, J_m$ in such a way that they consist of a disjoint union of boxes obtained as in the claimed proved above. The proof of 3 is therefore complete.

4. It s enough to show the result for two disjoint boxes, as general elements of $\mathcal{E}$ are simply finite union of them. However, the case of two disjoint boxes just reduces to 3, where $A = I_1 \cup I_2$, and I_1 and I_2 are the given disjoint boxes. This completes the proof of 4, and the proposition as well. □

Definition 8.7. A nonnegative additive set function ϕ on $\mathcal{E}$ is said to be regular if it has the following property:

- For any $A \in \mathcal{E}$ and every $\epsilon > 0$, there exists sets $F, G \in \mathcal{E}$, such that F is closed, G is open, $F \subset A \subset G$ and

$$\phi(G) - \epsilon \leq \phi(A) \leq \phi(F) + \epsilon.$$

We have already come across an example of regular function on $\mathcal{E}$, as the following proposition shows.

Proposition 8.8. *The set function m is regular.*

Proof. First assume that A is a box. Then, one sees that $m(A^\circ) = m(A) = m(\bar{A})$ by definition, and therefore one can take $F = \bar{A}$ and $G = A^\circ$. The general case follows from this special case by taking $A \in \mathcal{E}$ and decompose it into finite unions of boxes. Then one can apply the definition of m as the sum of m evaluated on each box. □

We will show that any regular function μ on $\mathcal{E}$ can be extended to a countably additive set function on a σ-ring that contains $\mathcal{E}$.

Definition 8.9. Let μ be regular and finite on $\mathcal{E}$. Let $E \subset \mathbb{R}^p$, let $A = \{A_1, \ldots, A_n, \ldots\}$ be a countable covering of E with open elementary sets A_n, and let $\mathcal{A}$ be the collection of such families A. Define

$$\mu^*(E) := \inf_{A \in \mathcal{A}} \sum_{n=1}^{\infty} \mu(A_n).$$

The function μ^* is called *outer measure* of E.

Theorem 8.10. *If μ is as in Definition 8.9. Then the following two facts hold.*

(a) *For every $A \in \mathcal{E}$, $\mu^*(A) = \mu(A)$.*
(b) *If $E = \bigcup_{n=1}^{\infty} E_n$, then $\mu^*(E) \leq \mu^*(E_n)$. (This is called subadditivity property)*

Proof. (a) Let $A \in \mathcal{E}$, and $\epsilon > 0$ be given. Since μ is regular, we can find an open elementary set G with $A \subset G$ and

$$\mu(G) \quad \leq \quad \mu(A) + \epsilon.$$

Since $\mu^*(A)$ is defined as the infimum of a sum over all collections of elementary sets covering A, we have that $\mu^*(A) \leq \mu(G)$, and therefore $\mu^*(A) \leq \mu(A) + \epsilon$. Since ϵ is arbitrary, it follows that $\mu^*(A) \leq \mu(A)$.

Conversely, by definition of μ^*, for an arbitrary choice of $\epsilon > 0$, we can find a sequence of open elementary sets A_n covering A such that

$$\sum_{n=1}^{\infty} \mu(A_n) \leq \mu^*(A) + \epsilon.$$

Since μ is regular, we can find a closed elementary set $F \subset A$ such that $\mu(F) \geq \mu(A) - \epsilon$. Since A is a finite union of boxes (because A is an elementary set), it follows that A is bounded. Therefore, $F \subset A$ is closed and bounded, which means that F is compact. We can find finitely many of the opens in the collection $\{A_n\}$ that cover F. If N is the largest integer in the finite subcollection, then $F \subset A_1 \cup \cdots \cup A_N$. Then, it follows that

$$\begin{aligned} \mu(A) &\leq \mu(F) + \epsilon \\ &\leq \mu(A_1 \cup \cdots \cup A_n) + \epsilon \\ &\leq \sum_{n=1}^{N} \mu(A_n) + \epsilon \\ &\leq \sum_{n=1}^{\infty} \mu(A_n) + \epsilon \\ &\leq \mu^*(A) + 2\epsilon. \end{aligned}$$

Since ϵ was arbitrary, we find that $\mu(A) \leq \mu^*(A)$. The equality is now proved.

(b) Let $E = \bigcup_{n=1}^{\infty} E_n$ and $\epsilon > 0$ be given. Observe that if one of the E_n satisfies $\mu^*(E_n) = \infty$, then there is nothing to prove. Therefore, we will assume that $\mu^*(E_n) < \infty$ for all $n \in \mathbb{N}$. By the definition of μ^*, we can find for each $n \in \mathbb{N}$ a sequence of open elementary sets A_{nk} such that

$$\sum_{k=1}^{\infty} \mu(A_{nk}) \leq \mu^*(E_n) + \frac{\epsilon}{2^n}.$$

Then, one has

$$\begin{aligned} \mu^*(E) &\leq \sum_{n=1}^{\infty} \sum_{k=1}^{\infty} \mu(A_{nk}) \\ &\leq \sum_{n=1}^{\infty} \mu(E_n) + \sum_{n=1}^{\infty} \frac{\epsilon}{2^n} \\ &\leq \sum_{n=1}^{\infty} \mu(E_n) + \epsilon, \end{aligned}$$

where in the first inequality we have used the fact that $\{A_{nk}\}_{n,k\in\mathbb{N}}$ is a countable family of open elementary sets that covers E, and thefore $\mu^*(E)$ by definition is less than or equal to the sum of all $\mu(A_{nk})$. Since ϵ was arbitrary, the inequality follows. □

Definition 8.11. For any $A, B \subset \mathbb{R}^p$ we define the *symmetric difference* of A and B as

$$S(A, B) := (A - B) \cup (B - A). \tag{47}$$

We also define

$$d(A, B) := \mu^*(S(A, B)). \tag{48}$$

If $\lim d(A_n, A) = 0$, we write $A_n \longrightarrow A$. If $A_n \longrightarrow A$ for a sequence of elementary sets A_n, we say that A is *finitely μ-measurable*. We denote by $\mathcal{M}_F(\mu)$ the class of finitely μ-measurable sets. If a set

X is the union of countably many finitely μ-measurable sets, we will say that X is μ-measurable. We denote by $\mathcal{M}(\mu)$ the class of μ-measurable sets.

Exercise 8.12. Prove the following proposition.

Proposition 8.13. *The following properties of $S(A, B)$ and $d(A, B)$ hold.*

- $S(A, B) = S(B, A)$, $S(A, A)$.
- $S(A, B) \subset S(A, C) \cup S(C, B)$.
- $S(A_1 \cup A_2, B_1 \cup B_2), S(A_1 \cap A_2, B_1 \cap B_2), S(A_1 - A_2, B_1 - B_2) \subset S(A_1, B_1) \cup S(A_2, B_2)$.
- $d(A, B) = d(B, A)$
- $d(A_1 \cup A_2, B_1 \cup B_2), d(A_1 \cap A_2, B_1 \cap B_2), d(A_1 - A_2, B_1 - B_2) \leq d(A_1, B_1) + d(A_2, B_2)$.
- *Define the congruence relation $A \cong B$ when $d(A, B) = 0$. Prove that on the set of equivalence classes d is a metric.*

Proposition 8.14. *We have $|\mu^*(A) - \mu^*(B)| \leq d(A, B)$, whenever one among $\mu^*(A)$ and $\mu^*(B)$ is finite.*

Proof. Suppose that $\mu^*(B)$ is finite, and $\mu^*(A) \geq \mu^*(B)$. Then, applying Proposition 8.13 we find

$$d(A, \emptyset) \quad \leq \quad d(A, B) + d(B, \emptyset).$$

Since $d(A, \emptyset) = \mu^*(A)$ and $d(B, \emptyset) = \mu^*(B)$, we have the result. □

Theorem 8.15. *$\mathcal{M}(\mu)$ is a σ-ring, and μ^* is countably additive on $\mathcal{M}(\mu)$.*

Proof. We first show that $\mathcal{M}_F(\mu)$ is a ring. Let $A, B \in \mathcal{M}_F(\mu)$. By definition of $\mathcal{M}_F(\mu)$, we can find two sequences of elementary sets A_n, B_n such that $A_n \longrightarrow A$ and $B_n \longrightarrow B$. Using Proposition 8.13, we find that $A_n \cup B_n \longrightarrow A \cup B$, since $d(A_n \cup B_n, A \cup B) \leq d(A_n, A) + d(B_n, B) \longrightarrow 0$. Therefore, $A \cup B \in \mathcal{M}_F(\mu)$. Similarly, we also have that $A_n - B_n \longrightarrow A - B$. These two facts together imply that $\mathcal{M}_F(\mu)$ is a ring, since from Proposition 8.6 we know that each $A_n \cup B_n$ and $A_n - B_n$ is an elemntary set. Moreover, we have that $A_n \cap B_n \longrightarrow A \cap B$. Since each A_n satisfies $\mu^*(A_n) = \mu(A_n) < +\infty$, using Proposition 8.14 we see that $|\mu^*(A_n) - \mu^*(A)| \leq d(A_n, A) \longrightarrow 0$, which implies that $\mu^*(A_n) \longrightarrow \mu^*(A)$, and therefore $\mu^*(A) < +\infty$. Similar results hold for B_n and B, as well as $A_n \cup B_n$ and $A \cup B$, $A_n \cap B_n$ and $A \cap B$.

Applying modularity to each A_n and B_n, we find that

$$\mu(A_n) + \mu(B_n) \quad = \quad \mu(A_n \cup B_n) + \mu(A_n \cap B_n).$$

From $\mu^*(A_n) \longrightarrow \mu^*(A)$, $\mu^*(B_n) \longrightarrow \mu^*(B)$, $\mu^*(A_n \cup B_n) \longrightarrow \mu^*(A \cup B)$, and $\mu^*(A_n \cap B_n) \longrightarrow \mu^*(A \cap B)$, it follows that

$$\mu^*(A) + \mu^*(B) \quad = \quad \mu^*(A \cup B) + \mu^*(A \cap B).$$

When $A \cap B = \emptyset$, then $\mu^*(A \cap B) = 0$, and we obtain (finite, not countable yet) additivity of μ^* on $\mathcal{M}_F(\mu)$.

Let us now consider $A \in \mathcal{M}(\mu)$. We want to show that A can be written as the countable union of sets in $\mathcal{M}_F(\mu)$. In fact, by definition of $\mathcal{M}(\mu)$, we can write $A = \bigcup_{n \in \mathbb{N}} A'_n$, where the sets A'_n need not necessarily be pairwise disjoint. We set $A_1 := A'_1$, and $A_n := A'_n - (A'_1 \cup \cdots \cup A'_{n-1})$ for each $n \in \mathbb{N}$. Then, by construction $\bigcup_{n \in \mathbb{N}} A_n = \bigcup_{n \in \mathbb{N}} A'_n = A$, and the A_n are pairwise disjoint. By additivity of μ^* on $\mathcal{M}_F(\mu)$, which we proved at the beginning of the proof, we have that for each

$n \in \mathbb{N}$ it holds

$$\begin{aligned} \mu^*(A) &\geq \mu^*(\bigcup_{k=1}^{n} A_k) \\ &= \mu^*(A_1) + \cdots + \mu^*(A_n), \end{aligned}$$

which implies $\mu^*(A) \geq \sum_{n=1}^{\infty} \mu^*(A)$. Also, applying Theorem 8.10, we obtain that $\mu^*(A) \leq \sum_{n=1}^{\infty} \mu^*(A_n)$, from which we get the equality $\mu^*(A) \leq \sum_{n=1}^{\infty} \mu^*(A_n)$. Suppose that $\mu^*(A) < +\infty$. Then, the equality $\mu^*(A) \geq \sum_{n=1}^{\infty} \mu^*(A)$ shows that $\sum_{n=k}^{\infty} \mu^*(A_n) \longrightarrow 0$, as $k \longrightarrow \infty$. Then, if we set $B_n := A_1 \cup \cdots \cup A_n$, we have found that

$$\begin{aligned} d(A, B_k) &= \mu^*(\bigcup_{n=k+1}^{\infty} A_n) \\ &= \sum_{n=k+1}^{\infty} \mu^*(A_n) \longrightarrow 0. \end{aligned}$$

It follows that $B_n \longrightarrow A$. Since each $B_n \in \mathcal{M}_F(\mu)$ by construction, it follows that $A \in \mathcal{M}_F(\mu)$. We have therefore proved that when $A \in \mathcal{M}(\mu)$ and $\mu^*(A) < +\infty$, then $A \in \mathcal{M}_F(\mu)$.

The last fact allows us to prove that μ^* is countably additive on $\mathcal{M}(\mu)$. In fact, suppose that $A = \bigcup_{n\in\mathbb{N}} A_n$ is the disjoint union of elements of $\mathcal{M}_F(\mu)$. If $\mu^*(A)$ is finite, we have already shown that $\mu^*(A) = \sum_{n=1}^{\infty} \mu^*(A_n)$. When $\mu^*(A)$ is infinite, then the equality just follows from Theorem 8.10.

We want to show now that $\mathcal{M}(\mu)$ is a σ-ring. Suppose $A_n = \bigcup_{k\in\mathbb{N}} A_{n,k}$ where $A_{n,k} \in \mathcal{M}_F(\mu)$ for each $n, k \in \mathbb{N}$. Then, the sets $A_{n,k}$ are countably many, and obviously $\bigcup_{n\in\mathbb{N}} A_n = \bigcup_{n,k\in\mathbb{N}} A_{n,k}$. This shows that $\bigcup_{n\in\mathbb{N}} A_n \in \mathcal{M}(\mu)$ as well. Let now $A, B \in \mathcal{M}(\mu)$, and therefore assume that $A = \bigcup_{n\in\mathbb{N}} A_n$ and $B = \bigcup_{n\in\mathbb{N}} B_n$. Then, we have

$$A_n \cap B = \bigcup_{k\in\mathbb{N}} (A_n \cap B_k),$$

where each $A_n \cap B_k \in \mathcal{M}_k(\mu)$, which implies that $A_n \cap B \in \mathcal{M}(\mu)$. Since $\mu^*(A_n \cap B) \leq \mu^*(A_n) < +\infty$, it follows that $A_n \cap B \in \mathcal{M}_F(\mu)$. Therefore, $A_n - B \in \mathcal{M}_F(\mu)$, which in turn implies that $A - B \in \mathcal{M}(\mu)$, since $A - B = \bigcup_{n\in\mathbb{N}}(A_n - B)$. This completes the proof. □

In what follows, for simplicity, we will write $\mu(A)$ instead of $\mu^*(A)$ for elements $A \in \mathcal{M}(\mu)$.

Definition 8.16. The function μ is called a *measure*. When $\mu = m$, this is called *Lebesgue measure*.

Remark 8.17. The following facts are useful.

- Any open set A satisfies $A \in \mathcal{M}(\mu)$. This is due to the fact that any open set is a countable union of open boxes, since $\mathbb{R}^p$ has a countable basis (for the topology) consisting of open boxes. Also, since $\mathcal{M}(\mu)$ is an algebra, by taking complements we have that any closed set C satisfies $C \in \mathcal{M}(\mu)$, being the complement of an open set by definition.
- If $A \in \mathcal{M}(\mu)$ and $\epsilon > 0$, then there exist a closed set F and and open set G such that $F \subset A \subset G$ with $\mu(G - A) < \epsilon$ and $\mu(A - F) < \epsilon$. To show this fact, first observe that upon passing to complements, the first fact implies the second. Therefore, we just need to show that we can find an open set G such that $\mu(G - A) < \epsilon$. However, this holds because μ (actually μ^* according to the definitions above) is defined through coverings of open elementary sets.

- We define the *Borel sets* as the collection $\mathcal{B}$ of all sets obtained by applying countable operations of unions and intersections, and taking complements, to open sets. One can show that $\mathcal{B}$ is a σ-ring, and it is the smallest σ-ring which contains all open sets.
- If $A \in \mathcal{M}(\mu)$, then there exist Borel sets F and G such that $F \subset A \subset G$ and $\mu(G - A) = \mu(A - F) = 0$. This follows directly from the second properties listed in this remark. Moreover, since $A = F \cup (A - F)$, it follows that any μ-measurable set A is the disjoint union of a Borel set and a set of measure zero.
- The sets of measure zero form a σ-ring.
- For the Lebesgue measure m, any countable set has measure zero.

We now are in the position of introducing measure spaces.

Definition 8.18. Let X be an arbitrary set. Then X is said to be a *measure space* if there exists a σ-ring $\mathcal{M}$ of subsets of X (called the measurable sets) and a nonnegative countably additive set function μ on $\mathcal{M}$ (called the measure). If, in addition, $X \in \mathcal{M}$, we say that X is a *measurable space*.

The construction that gave us the Lebesgue measure shows an example of measure space.

Definition 8.19. Let f be a function defined on a measurable space X, with values in the extended real numbers $\mathbb{R} \cup \{\pm\infty\}$. Then, the function f is said to be measurable if if the set $\{x \in X \mid f(x) > a\}$ is measurable for each $a \in \mathbb{R}$.

Theorem 8.20. *The following conditions are equivalent.*

(i) $\{x \in X \mid f(x) > a\}$ *is measurable for every* $a \in \mathbb{R}$.
(ii) $\{x \in X \mid f(x) \geq a\}$ *is measurable for every* $a \in \mathbb{R}$.
(iii) $\{x \in X \mid f(x) < a\}$ *is measurable for every* $a \in \mathbb{R}$.
(iv) $\{x \in X \mid f(x) \leq a\}$ *is measurable for every* $a \in \mathbb{R}$.

Proof. We prove (i) $\implies$ (ii). First, observe that we can write

$$\{x \in X \mid f(x) \geq a\} = \bigcap_{n \in \mathbb{N}} \{x \in X \mid f(x) > a - \frac{1}{n}\}.$$

Since $\{x \in X \mid f(x) \geq a\}$ can be written as the countable intersection of measurable sets, it is measurable, since $\mathcal{M}$ is a σ-algebra. Next, to show that (ii) implies (iii) we have

$$\{x \in X \mid f(x) < a\} = X - \{x \in X \mid f(x) \geq a\},$$

showing that $\{x \in X \mid f(x) < a\}$ is the complement of a measurable set, hence measurable. To show (iii) $\implies$ (iv) note that

$$\{x \in X \mid f(x) \leq a\} = \bigcap_{n \in \mathbb{N}} \{x \in X \mid f(x) < a + \frac{1}{n}\}.$$

Lastly, to show that (iv) implies (i) we observe that

$$\{x \in X \mid f(x) > a\} = X - \{x \in X \mid f(x) \leq a\}.$$

This completes the proof. □

Theorem 8.21. *If* f *is measurable, then* $|f|$ *is measurable too.*

Proof. Suppose that $a \geq 0$. By definition of absolute value we have that

$$\{x \in X \mid |f(x)| < a\} \quad = \quad \{x \in X \mid f(x) < a\} \cap \{x \in X \mid f(x) > -a\},$$

where both sets in the intersection are measurable because f is measurable. If $a < 0$ then $\{x \in X \mid |f(x)| < a\} = \{x \in X \mid f(x) < a\} = \emptyset$, which is measurable too. □

Theorem 8.22. *Let $\{f_n\}$ be a sequence of measurable functions. Then, the functions*

$$g(x) = \sup_n f_n(x),$$
$$h(x) = \limsup_n f_n(x),$$

are both measurable functions. Similar results hold for inf *and* lim inf*.*

Proof. We have

$$\{x \in X \mid g(x) > a\} \quad = \quad \bigcup_{n \in \mathbb{N}} \{x \in X \mid f_n(x) > a\}.$$

Since each set in the union is measurable, $\{x \in X \mid g(x) > a\}$ is measurable and g is measurable.

To prove the result for h, observe that

$$h(x) \quad = \quad \inf g_m(x),$$

where $g_m(x) := \sup_{n \geq m} f_n(x)$. So, using the fact that sup of measurable functions is measurable (proved in the first part of the proof), and the analogous version for the infimum, one obtains the result for h as well. □

The following now follows easily.

Corollary 8.23. *If f and g are measurable, then the functions $\max(f, g)$ and $\min(f, g)$ are measurable. In particular, the functions $f^+ := \max(f, 0)$ and $f^- := \min(f, 0)$ are meaurable. Also, the limit of a convergent sequence of measurable functions is measurable.*

Theorem 8.24. *Let f and g be meaasurable real-valued functions defined on the measurable space X. Let $F : \mathbb{R}^2 \longrightarrow \mathbb{R}$ be a continuous function. Define the function $h : X \longrightarrow \mathbb{R}$ by*

$$h(x) \quad := \quad F(f(x), g(x)).$$

Then, h is measurable.

Proof. Let $G = \{(u, v) \in \mathbb{R}^2 \mid F(u, v) > a\}$. By construction, $G = F^{-1}(a, +\infty)$, and it is therefore open since F is continuous. Since the topology of $\mathbb{R}^2$ has a countable basis consisting of rectangles of type $(a_n, b_n) \times (c_n, d_n)$, it follows that $G = \bigcup_{n \in \mathbb{N}} (a_n, b_n) \times (c_n, d_n)$, for some choices of $a_, b_n, c_n, d_n$. Observe that

$$\{x \in X \mid f(x) \in (a_n, b_n)\} \quad = \quad \{x \in X \mid f(x) > a_n\} \cap \{x \in X \mid f(x) < b_n\},$$

which means that the set $\{x \in X \mid f(x) \in (a_n, b_n)\}$ is measurable because f is measurable. Similarly, $\{x \in X \mid g(x) \in (c_n, d_n)\}$ is measurable as well. From

$$\begin{aligned}
\{x \in X \mid F(f(x), g(x)) > a\} &= \{x \in X \mid (f(x), g(x)) \in G\} \\
&= \bigcup_{n \in \mathbb{N}} \{x \in X \mid (f(x), g(x)) \in (a_n, b_n) \times (c_n, d_n)\} \\
&= \bigcup_{n \in \mathbb{N}} (\{x \in X \mid f(x) \in (a_n, b_n)\} \cap \{x \in X \mid g(x) \in (c_n, d_n)\}),
\end{aligned}$$

it follows that $\{x \in X \mid F(f(x), g(x)) > a\}$ is measurable, because each set $(\{x \in X \mid f(x) \in (a_n, b_n)\} \cap \{x \in X \mid g(x) \in (c_n, d_n)\})$ is measurable. Since a was arbitrary, we are done. $\square$

Remark 8.25. As an application of Theorem 8.24 we have that the operations $+$ and $\cdot$ are measurable, when composed with measurable functions. More specifically, if f and g are measurable, the same is true for $f+g$ and fg.

Remark 8.26. Measurability of functions is a concept only depending on the σ-ring on X.

We now introduce the fundamental concept of *simple function*, which is instrumental in defining integration for abstract measure spaces. Simple functions are a generalization of the notion of step function.

Definition 8.27. We say that a function $s : X \longrightarrow \mathbb{R}$ is a simple function if its range consists of finitely many points. For a subseet $E \subset X$ we define the characteristic function of E, χ_E as

$$\chi_E(x) \quad = \quad \begin{cases} 1 & x \in E \\ 0 & x \notin E \end{cases} \quad .$$

We call the set E the *support* of χ_E.

Proposition 8.28. *Let s be a simple function. Then, $s = \sum_{i=1}^n c_i \chi_{E_i}$ for some sets E_i. Moreover, s is measurable if and only if each E_i is measurable.*

Proof. If s is a simple function, and $c_1, \dots, c_n$ (pairwise distinct) are the values s attains. We define the preimages $E_i = s^{-1}(\{c_i\})$. Then, s can be written as a linear combination of characteristic functions χ_{E_i} as $s = \sum_i c_i \chi_{E_i}$. In fact, note that $E_i \cap E_j = \emptyset$ for each $i \neq j$ since $c_i \neq c_j$. Then, for $x \in X$, there is a single E_{i^*} such that $x \in E_{i^*}$. Then, $\chi_{E_{i^*}}(x) = 1$ and $\chi_{E_j} = 0$ for all $j \neq i^*$. So, $\sum_j c_j \chi_{E_j}(x) = c_{i^*}\chi_{E_{i^*}}(x) = c_{i^*} = s(x)$. This shows the first part of the proposition.

For the second part, let $E \subset X$, then

$$\{x \in X \mid \chi_E(x) > a\} \quad = \quad \begin{cases} E & a < 1 \\ \emptyset & a \geq 1 \end{cases} \quad .$$

Since $\emptyset$ is always measurable, we have that χ_E is a measurable if and only if E is measurable. Assume now that $s = \sum_{i=1}^n c_i \chi_{E_i}$. If E_i is measurable for each E_i, as we have just shown each χ_{E_i} is measurable, and applying Theorem 8.24 we have that s is measurable as well. Conversely, if s is measurable, without loss of generality we can assume that the numbers c_i are ordered as $c_1 < c_2 < \cdots < c_n$. We have then notice that

$$E_i \quad = \quad \{x \in X \mid f(x) > c_{i-1}\} \cap \{x \in X \mid f(x) < c_{i+1}\},$$

where $\{x \in X \mid f(x) > c_{i-1}\} = \emptyset$ when $i = 1$ and $\{x \in X \mid f(x) < c_{i+1}\} = \emptyset$ when $i = n$. Then, each E_i is measurable, and the proof is complete. $\square$

The result that allows us to relate simple functions to more general functions is the following approximation result.

Theorem 8.29. *Let f be a real function on X. Then, there exists a sequence of simple functions s_n which converges pointwise to f, i.e. $s_n(x) \longrightarrow f(x)$ for all $x \in X$. If f is measurable, then s_n can be chosen to be a sequence of measurable functions. If $f \geq 0$, then s_n can be chosen to be monotonically increasing.*

Proof. Since $f = f^+ - f^-$, where $f^+, f^- \geq 0$, we can prove the result for nonnegative functions, because if $s_n \longrightarrow f^+$ pointwise and $s'_n \longrightarrow f^-$ pointwise, we have that $s_n - s'_n \longrightarrow f^+ - f^-$ pointwise. Suppose therefore that $f \geq 0$. For each $n \in \mathbb{N}$, and for $i = 1, 2, \ldots, n2^n$, define the sets

$$\begin{aligned} E_{n,i} &= \{x \in X \mid \frac{i-1}{2^n} \leq f(x) < \frac{i}{2^n}\} \\ F &= \{x \in X \mid f(x) \geq n\}. \end{aligned}$$

We define the simple function s_n as

$$s_n := \sum_{i=1}^{n2^n} \frac{i-1}{2^n} \chi_{E_{n,i}} + n\chi_{F_n}.$$

Let $x \in X$ and fix $\epsilon > 0$ arbitrary. Then, since $f(x)$ is real, there exists an n_0 large enough such that $f(x) < n_0$ Take $\nu \in \mathbb{N}$ such that $\nu > n_0$ and such that $\frac{1}{2^\nu} < \epsilon$. Then, for any $n > \nu$, $x \in E_{n,i}$ since $f(x) < n$ and it follows that $f(x) \in [\frac{i^*-1}{2^n}, \frac{i^*}{2^n}]$ for some i^*. Therefore

$$\begin{aligned} |s_n(x) - f(x)| &= |\sum_{i=1}^{n2^n} \frac{i-1}{2^n} \chi_{E_{n,i}}(x) + n\chi_{F_n}(x) - f(x)| \\ &= |\frac{i^*-1}{2^n} - f(x)| \\ &\leq |\frac{i^*-1}{2^n} - \frac{i^*}{2^n}| \\ &= \frac{1}{2^n} < \epsilon. \end{aligned}$$

Pointwise convergence follows. When f is measurable, the sets F_n are all measurable by definition. Moreover, notice that

$$E_{n,i} = \{x \in X \mid f(x) \geq \frac{i-1}{2^n}\} \cap \{x \in X \mid f(x) < \frac{i}{2^n}\},$$

and therefore each $E_{n,i}$ is measurable as well. By Proposition 8.28 we have that each s_n is measurable. Since each s_n is monotonic increasing by construction, the proof is complete. □

Remark 8.30. We note that when f is bounded, convergence in Theorem 8.29 is uniform.

We are now in the position to define integration for a measure μ.

Definition 8.31. Let $s = \sum_{i=1}^n c_i \chi_{E_i}$ be a measurable simple function, and therefore by Proposition 8.28 we have that each E_i is measurable, where each $c_i \geq 0$. Then, for a measurable set $E \in \mathcal{M}$ we define the integral of s as

$$\int_E s d\mu = \sum_{i=1} c_i \mu(E \cap E_i).$$

If f is a measurable nonnegative function, we define its integral over E as

$$\int_E f d\mu = \sup_s \int_E s d\mu,$$

as s varies among the simple functions s such that $0 \leq s \leq f$. We call $\int_E f d\mu$ the *Lebesgue integral* of f over E with respect to μ.

Remark 8.32. We observe that the Lebesgue integral of a function can be infinite.

Moreover, the definition of integral for a simple function is well posed, as one has

$$\sup_\sigma \int_E \sigma d\mu \quad = \quad \sum_{i=1} c_i \mu(E \cap E_i),$$

where the supremum on the left-hand-side is taken over all step functions σ such that $0 \leq \sigma \leq s$.

Definition 8.33. Let f be measurable, and consider the two nonnegative functions f^+ and f^-. Then, if at least one of the integrals of f^+ and f^- is finite, we set

$$\int_E f d\mu \quad = \quad \int_E f^+ d\mu - \int_E f^- d\mu.$$

When both $\int_E f^+ d\mu$ and $\int_E f^- d\mu$ are finite, the value $\int_E f d\mu$ is finite as well, and we say that f is *Lebesgue integrable*, or *summable*. In this case we write $f \in \mathcal{L}_E(\mu)$ or simply $f \in \mathcal{L}(\mu)$ if the set E is understood.

Proposition 8.34. *We have the following properties of the Lebesgue integral.*

(a) *If f is measurable and bounded on E, and $\mu(E) < +\infty$, then $f \in \mathcal{L}(\mu)$.*
(b) *If $a \leq f(x) \leq b$ for all $x \in E$, and $\mu(E) < +\infty$, then*

$$a\mu(E) \leq \int_E f d\mu \leq b\mu(E).$$

(c) *If $f, g \in \mathcal{L}(\mu)$ and $f \leq g$, then*

$$\int_E f d\mu \quad \leq \quad \int_E g d\mu.$$

(d) *If $f, g \in \mathcal{L}(\mu)$, then $cf \in \mathcal{L}(\mu)$ for all $c \in \mathbb{R}$. Moreover,*

$$\int_E cf d\mu \quad = \quad c \int_E f d\mu.$$

(e) *If $\mu(E) = 0$ and f is measurable, then $\int_E f d\mu = 0$.*
(f) *If $f \in \mathcal{L}(\mu)$ on E, and $A \in \mathcal{M}$ with $A \subset E$, then $f \in \mathcal{L}(\mu)$ on A.*

Exercise 8.35. Prove Proposition 8.34.

Theorem 8.36. *Let f be measurable and nonnegative on X. For $A \in \mathcal{M}$ define*

$$\phi(A) := \int_A f d\mu.$$

Then ϕ is countably additive on $\mathcal{M}$. The same holds if $f \in \mathcal{L}(\mu)$ on X.

Proof. For $f \in \mathcal{L}(\mu)$, writing $f = f^+ - f^-$, where both f^+ and f^- are nonnegative, it follows that ϕ is countably additive on $\mathcal{M}$ if we prove the result for measurable nonnegative functions. We will therefore only prove the case where f is measurable and nonnegative.

Suppose $A = \bigcup_{n \in \mathcal{N}} A_n$ where $A_i \cap A_j = \emptyset$ whenever $i \neq j$. We need to show that $\phi(A) = \sum_{n \in \mathbb{N}} \phi(A_n)$. When f is a characteristic function, since its integral is just defined through μ, which is countably additive, it immediately follows that ϕ is countably additive. When f is simple, the result also immediately follows because it is a linear combination of characteristic functions.

Let now f be not simple. Whenever $0 \leq s \leq f$, we have that

$$\begin{aligned}\int_A s d\mu &= \sum_{n=1}^{+\infty} \int_{A_n} s d\mu \\ &\leq \sum_{n=1}^{+\infty} \int_{A_n} f d\mu \\ &= \sum_{n \in \mathbb{N}} \phi(A_n).\end{aligned}$$

Therefore, by definition of Lebesgue integral we have that $\phi(A) \leq \sum_{n\in\mathbb{N}} \phi(A_n)$. The case when $\phi(A_n) = +\infty$ for one of the sets A_n gives both sides of the equality to be infinite, since $\phi(A) = \int_A f d\mu \geq \int_{A_n} f d\mu = \phi(A_n)$. We therefore consider the case where $\phi(A_n) < +\infty$ for all $n \in \mathbb{N}$. Let $\epsilon > 0$ be arbitrary. By definition of Lebesgue integral, we can find a simple function s such that

$$\begin{aligned}\int_{A_1} s d\mu &\geq \int_{A_1} f d\mu - \epsilon, \\ \int_{A_2} s d\mu &\geq \int_{A_2} f d\mu - \epsilon.\end{aligned}$$

Then, we have

$$\begin{aligned}\phi(A_1 \cup A_2) &= \int_{A_1 \cup A_2} s d\mu \\ &\geq \int_{A_1} s d\mu + \int_{A_2} s d\mu \\ &\geq \phi(A_1) + \phi(A_2) - 2\epsilon.\end{aligned}$$

Therefore, $\phi(A_1 \cup A_2) \geq \phi(A_1) + \phi(A_2)$. Inductively, for every $n \in \mathbb{N}$ we have that $\phi(A_1 \cup \cdots \cup A_n) \geq \phi(A_1) + \cdots + \phi(A_n)$. Also, since $A \supset A_1 \cup + \cdots \cup A_n$, it follows that $\phi(A) \geq \phi(A_1 \cup + \cdots \cup A_n) \geq \phi(A_1) + \cdots + \phi(A_n)$. It now follows that

$$\phi(A) \geq \sum_{n \in \mathbb{N}} \phi(A_n).$$

The result now follows. □

Corollary 8.37. *If $A, B \in \mathcal{M}$ with $B \subset A$ and $\mu(A - B) = 0$, then $\int_A f d\mu = \int_B f d\mu$.*

Proof. Applying Theorem 8.36 to the case where $A_1 = A$ and $A_2 = B$, along with Proposition 8.34 (e), the result immediately follows. □

Corollary 8.37 shows that sets of measure zero can be neglected when discussing integration. Therefore, two functions that are different from each others on a set of measure zero are substantially the same object when integration is involved. This motivates the following definition.

Definition 8.38. Suppose that the set $\{x \in E \mid f(x) \neq g(x)\}$ has measure zero, i.e. $\mu(\{x \in E \mid f(x) \neq g(x)\}) = 0$. Then, we write $f \sim g$. A direct inspection shows that $\sim$ is an equivalence relation.

From Corollary 8.37 we have that if $f \sim g$, then $\int_E f d\mu = \int_E g d\mu$, when the integrals exist. More generally, whenever $A \subset E$ we have $\int_A f d\mu = \int_A g d\mu$.

When a function f satisfies a property for all $x \in E - A$, where A is a set of measure zero, then we say that f satisfies a property *almost everywhere* on E. The following is easy to verify.

Proposition 8.39. *Let $f \in \mathcal{L}(\mu)$ on E. Then f is finite almost everywhere.*

As a consequence of the proposition one may assume that a function is finite, since the set of values for which this does not happen has zero measure.

Theorem 8.40. *Let $f \in \mathcal{L}(\mu)$ on E. Then $|f| \in \mathcal{L}(\mu)$ as well, and*

$$|\int_E f d\mu| \quad \leq \quad \int_E |f| d\mu. \tag{49}$$

Proof. We decompose E as $E = A \cup B$, where $A = \{x \in E \mid f(x) \geq 0\}$ and $B = E - A$. Theorem 8.36 gives

$$\begin{aligned} \int_E |f| d\mu &= \int_A |f| d\mu + \int_B |f| d\mu \\ &= \int_A f^+ d\mu + \int_B f^- d\mu \\ &< +\infty, \end{aligned}$$

and therefore $|f| \in \mathcal{L}(\mu)$. Moreover, from the fact that $f \leq |f|$ and $-f \leq |f|$ we find that $\int_E f d\mu \leq \int_E |f| d\mu$ and $-\int_E f d\mu \leq \int_E |f| d\mu$. These last two facts together give (49). □

Theorem 8.41. *Suppose f is measurable, and $|f| \leq g$ with $g \in \mathcal{L}(\mu)$. Then $f \in \mathcal{L}(\mu)$.*

Proof. Since $|f| \leq g$, we have that both $f^+ \leq g$ and $f^- \leq g$, from which $\int_E f^+ d\mu \leq \int_E g d\mu < +\infty$ and $\int_E f^- d\mu \leq \int_E g d\mu < +\infty$. This implies readily that $\int_E f d\mu < +\infty$, and $f \in \mathcal{L}(\mu)$. □

Theorem 8.42 (Lebesgue Monotone Convergence Theorem)**.** *Let $E \in \mathcal{M}$, and let f_n be a sequence of measurable functions such that $0 \leq f_1(x) \leq f_2(x) \leq \cdots \leq f_n(x) \leq \cdots$ for all $x \in E$. Define f as the pointwise limit of f_n. Then we have that*

$$\int_E f_n d\mu \quad \longrightarrow \quad \int_E f d\mu.$$

Proof. Observe that the definition of f makes sense since for each $x \in E$, the sequence $f_n(x)$ is monotonic increasing, and it therefore admits a limit. Moreover, applying Theorem 8.22 we know that the resulting function is measurable. For any simple function s such that $0 \leq s \leq f_n$, it follows also that $0 \leq s \leq f_{n+1}$, and therefore $\int_E f_n d\mu \leq \int_E f_{n+1} d\mu$. Since the sequence of integrals is monotonic increasing, the sequence admits a limit. We set $\lim \int_E f_n d\mu = \alpha$. Since for each $n \in \mathbb{N}$ we have that $f_n \leq f$, Proposition 8.34 implies that $\int_E f d\mu \geq \int_E f_n d\mu$ for all $n \in \mathbb{N}$. It follows that

$$\alpha \quad \leq \quad \int_E f d\mu.$$

We choose c such that $0 < c < 1$, and we let s be a simple measurable function such that $0 \leq s \leq f$. We define the set

$$E_n \quad := \quad \{x \in E \mid f_n(x) \geq cs(x)\}.$$

Since f_n is monotonic, it follows that $E_1 \subset E_2 \subset \cdots \subset E_n \subset \cdots$. Since f is the pointwise limit of f_n, we also have that

$$E \quad = \quad \bigcup_{n \in \mathbb{N}} E_n.$$

Since $E_n \subset E$, we have for each $n \in \mathbb{N}$ the inequalities

$$\begin{aligned}\int_E f_n d\mu &\geq \int_{E_n} f_n d\mu \\ &\geq c\int_{E_n} s d\mu.\end{aligned}$$

Taking limits, we find that $\lim \int_E f_n d\mu \geq \lim c \int_{E_n} s d\mu$. Applying Theorem 8.36 and Theorem 8.5, $\lim c \int_{E_n} s d\mu = c \int_E s d\mu$, while by definition we have that $\lim \int_E f_n d\mu = \alpha$. Hence we have

$$\alpha \geq c\int_E s d\mu.$$

Taking the limit $c \longrightarrow 1$, we have

$$\alpha \geq \int_E s d\mu,$$

and since $\int_E f d\mu$ is the supremum of all integrals $\int_E s d\mu$ as s varies among the measurable simple functions such that $0 \leq s \leq f$, we find that $\alpha \geq \int_E f d\mu$, which completes the proof. □

Theorem 8.43. *Let $f = f_1 + f_2$, where $f_i \in \mathcal{L}(\mu)$ for $i = 1, 2$. Then $f \in \mathcal{L}(\mu)$, and*

$$\int_E f d\mu = \int_E f_1 d\mu + \int_E f_2 d\mu. \tag{50}$$

Proof. We first assume that $f_i \geq 0$, $i = 1, 2$. If both f_1 and f_2 are simple functions, then by definition of integral we immediately find that (50) holds. If f_1 and f_2 are not simple, we can choose by Theorem 8.29 monotonically increasing sequences of simple functions $\{s'_n\}$ and $\{s''_n\}$ converging to f_1 and f_2, respectively. We set $s_n = s'_n + s''_n$, and applying (50) for simple functions we find that $\int_E s_n d\mu = \int_E s'_n d\mu + \int_E s''_n d\mu$. Taking the limit $n \longrightarrow +\infty$, and applying Theorem 8.42, we find that

$$\begin{aligned}\int_E f d\mu &= \int_E (\lim_n s_n) d\mu \\ &= \lim_n \int_E s_n d\mu \\ &= \lim_n \int_E s'_n d\mu + \lim_n \int_E s''_n d\mu \\ &= \int_E (\lim_n s'_n) d\mu + \int_E (\lim_n s''_n) d\mu \\ &= \int_E f_1 d\mu + \int_E f_2 d\mu.\end{aligned}$$

Suppose now that $f_1 \geq 0$ and $f_2 \leq 0$. We consider the two disjoint sets

$$\begin{aligned}A &:= \{x \in E \mid f(x) \geq 0, \}, \\ B &:= \{x \in E \mid f(x) < 0, \}.\end{aligned}$$

On A, we have that $f_1(x) = f(x) - f_2(x)$, where $f_1(x)$, $f(x)$ and $f_2(x)$ are all nonnegative. Therefore, we can apply the first part of the proof to obtain

$$\int_A f_1 d\mu = \int_A f d\mu - \int_A f_2 d\mu.$$

On B, we have $-f_2(x) = -f(x) + f_1(x)$, where $f_1(x)$, $-f(x)$ and $f_2(x)$ are all positive. Therefore, we can again apply the first part of the proof to obtain

$$-\int_B f_2 d\mu \;=\; -\int_B f d\mu + \int_A f_1 d\mu.$$

In both cases we obtain therefore that

$$\int_C f d\mu \;=\; \int_C f_1 d\mu + \int_C f_2 d\mu,$$

where $C = A, B$. Adding the two results gives (50). The general case is treated similarly, by decomposing E into four sets upon which the signs of f_1 and f_2 are constant. Then (50) holds on each of the sets by considering the previously treated cases, and then summing all the equations the result follows. □

Theorem 8.44. *Let $E \in \mathcal{M}$. If $\{f_n\}$ is a sequence of measurable nonnegative functions and*

$$f(x) \;=\; \sum_{n=1}^{\infty} f_n(x).$$

Then, we have

$$\int_E f d\mu \;=\; \sum_{n=1}^{\infty} int_E f_n d\mu.$$

Proof. The sequence of partial sums $s_n = \sum_{k=1}^n f_k$ is an increasing sequence and f is the pointwise limit of s_n by definition. We can therefore apply Theorem 8.42. □

The following result is usually called Fatou's Lemma.

Theorem 8.45 (Fatou's Lemma)**.** *Let $E \in \mathcal{M}$. Let $\{f_n\}$ be a sequence of nonnegative measurable functions, and set*

$$f(x) = \liminf f_n(x).$$

Then we have

$$\int_E f d\mu \;\leq\; \liminf \int_E f_n d\mu.$$

Proof. By definition of limit inferior, we have that $f(x) = \lim g_n(x)$, where $g_n(x)$ is defined as

$$g_n(x) \;=\; \inf_{i \geq n} f_i(x).$$

By construction, we have that $g_n(x) \leq g_{n+1}(x)$, and $g_n(x) \geq 0$ for all $n \in \mathbb{N}$ and all $x \in E$. Applying Theorem 8.42 we have that

$$\int_E f d\mu \;=\; \lim \int_E g_n d\mu.$$

Since $g_n(x) \leq f_n(x)$ holds for all $n \in \mathbb{N}$ and $x \in E$, we also have that

$$\int_E g_n d\mu \;\leq\; \int_E f_n d\mu,$$

for all $n \in \mathbb{N}$, from which

$$\begin{aligned} \lim \int_E g_n d\mu &= \liminf \int_E g_n d\mu \\ &\leq \liminf \int_E f_n d\mu. \end{aligned}$$

This completes the proof. □

Theorem 8.46 (Lebesgue Dominated Convergence Theorem)**.** *Let $E \in \mathcal{M}$, and let $\{f_n\}$ be a sequence of measurable functions such that $f_n(x) \longrightarrow f(x)$. Suppose there exists a function $g \in \mathcal{L}(\mu)$ such that*

$$|f_n(x) \quad \leq \quad g(x).$$

Then, we have

$$\lim \int_E f_n d\mu \quad = \quad \int_E f d\mu.$$

Proof. Applying Theorem 8.41, it follows immediately that $f_n \in \mathcal{L}(\mu)$ for all $n \in \mathbb{N}$, and also $f \in \mathcal{L}(\mu)$. For each n we also have that $f_n + g \geq 0$, and applying Fatou's Lemma (Theorem 8.45) we obtain that

$$\int_E (f+g) d\mu \quad \leq \quad \liminf \int_E (f_n + g) d\mu.$$

The previous inequality gives (observe that g does not depend on n)

$$\int_E f d\mu \quad \leq \quad \liminf \int_E f_n d\mu.$$

Analogously, since $g - f_n \geq 0$, we have

$$\int_E (g - f) d\mu \quad \leq \quad \liminf \int_E (g - f_n) d\mu,$$

and therefore

$$-\int_E f d\mu \quad \leq \quad \liminf [-\int_E f_n d\mu].$$

The latter is equivalent to

$$\int_E f d\mu \quad \geq \quad \limsup \int_E f_n d\mu.$$

It therefore follows that $\limsup \int_E f_n d\mu = \liminf \int_E f_n d\mu = \lim \int_E f_n d\mu$, and

$$\lim \int_E f_n d\mu \quad = \quad \int_E f d\mu$$

now follows. □

As a particular case we also obtain the following.

Corollary 8.47. *Suppose $\mu(E) < +\infty$, f_n is uniformly bounded, and $f_n(x) \longrightarrow f(x)$. Then*

$$\lim \int_E f_n d\mu \quad = \quad \int_E f d\mu.$$

We will now denote by $d\mu$ the Lebesgue measure on the interval $[a, b]$. We denote by $\int_a^b f d\mu$ the Lebesgue integral, and $\int_a^b f dx$ the Riemann integral. Our next objective is to compare the two notions.

Theorem 8.48. *The following facts hold.*

- *Let $f \in \mathcal{R}$ on $[a, b]$. Then $f \in \mathcal{L}$ on $[a, b]$, and*

$$\int_a^b f d\mu \quad = \quad \int_a^b f dx.$$

- *Let f be a bounded function on $[a, b]$, then $f \in \mathcal{R}$ if and only if f is continuous almost everywhere.*

References

[1] Tom M Apostol, *Mathematical analysis*, 1974.
[2] Andrew Browder, *Mathematical analysis: an introduction*, Springer Science & Business Media, 2012.
[3] Andrei Nikolaevich Kolmogorov and Sergei Vasilevich Fomin, *Introductory real analysis*, Courier Corporation, 1975.
[4] Walter Rudin, *Principles of mathematical analysis*, 1953.
[5] James Stewart, *Single variable calculus: Early transcendentals*, Cengage Learning, 2015.

Department of Mathematics and Statistics, Idaho State University, Physical Science Complex — 921 S. 8th Ave., Stop 8085 — Pocatello, ID 83209
E-mail address: emanuelezappala@isu.edu