Notes in Topology

Herman Jaramillo

July 18, 2015
## Contents

1 Introduction ........................................... 5

2 Metric Spaces ....................................... 7
    2.1 Definition ........................................ 7
    2.2 Examples .......................................... 8
        2.2.1 The Discrete Metric .......................... 8
        2.2.2 The Absolute Difference Metric ......... 9
        2.2.3 The Norm Induced Metric ................. 10
        2.2.5 The Space of Sequences .................... 15
        2.2.6 The Product Space .......................... 16
        2.2.7 The Space of Functions .................... 18
        2.2.8 Bounded Metrics ............................ 24
    2.3 Topological Definitions .......................... 27
    2.4 Examples: ......................................... 36
        2.4.5 Continuity .................................. 47
        2.4.7 Limit, Convergence, Cauchy Sequence, and Complete-
             ness ........................................... 48
        2.4.14 Compactness ................................ 62

3 Basic Topological Concepts ......................... 79
    3.1 Introduction ...................................... 79
    3.2 Basic Definitions ................................ 79
        3.2.5 Examples ................................... 86
        3.2.6 Examples/Counter-examples based on the discrete topol-
            ogy ........................................... 88
    3.3 Base of a topology ................................ 89
        3.3.2 Examples ................................... 91
    3.4 The Concept of Continuity and Homomorphism .... 93
## CONTENTS

3.4.3 Examples: .......................................................... 94
3.5 Topology Sizes ......................................................... 96
3.6 Derived Topologies ................................................... 98
   3.6.1 The Induced Topology: Subspaces of Topological Spaces 98
   3.6.2 The Intersection Topology ..................................... 98
   3.6.3 Topologizing on sets ........................................... 101
   3.6.4 The Product Topology ......................................... 101
   3.6.10 The Quotient Topology ....................................... 119

A About the Largest Rational smaller than $\sqrt{2}$ .......................... 137

B Important Inequalities .................................................... 141
   B.1 Convex Functions .................................................. 141
       B.1.1 Jensen’s Inequality .......................................... 143
       B.1.2 Weighted geometric/arithmetic mean inequality ....... 145
       B.1.3 Young’s inequality .......................................... 146
       B.1.4 Hölder’s inequality ......................................... 146
       B.1.5 Minkowski’s inequality .................................... 150

C Notes in Real Analysis .................................................. 155
   C.1 Basic Definitions and Theorems on Bounded Sets .......... 155
   C.2 Completeness of The Reals $\mathbb{R}$ ............................. 157
       C.2.1 Historical Introduction ..................................... 157
       C.2.2 The Bolzano-Weristrass Theorem ......................... 158
       C.2.3 Completeness of the Reals ................................. 160

D Partitions, Equivalent Relations, and Functions ................. 163
   D.1 Equivalence Relations ........................................... 163
       D.1.3 Examples: .................................................... 165
   D.2 Partitions .......................................................... 167
       D.2.2 Examples: .................................................... 168
   D.3 Functions .......................................................... 173

E Derivation of the some twisted surfaces in Cartesian Coor- 
   dinates: ................................................................. 179
Chapter 1

Introduction

Motivation:
Prerequisites: set theory. Before topology courses in analysis should be taken. Real analysis, functional analysis . . .

The first chapter works on metric spaces. The second chapter develops topological spaces.

More:
Conventions: The convention \( a := b \) means that \( a = b \) by definition. When we say \( O_\alpha \), we mean \( O_{\alpha \in A} \), where \( A \) is an arbitrary index set. It could be the natural numbers or it could be any number in the interval \([0, 1]\) for example. When we say \( O_i \), we mean \( O_{i \in I} \) where \( I \) is a countable set (finite in most of the cases). For example the integers or natural numbers.

The Complement of a set. It is common to call the complement \( A \) as of a set as \( X \setminus A \). When we do not need to make explicit that the universe is \( X \), we can simplify this notation with \( A^c \).

Credits: Latex, Asymptote, Tikz, Grapher (Imaginery) Stephan Klaus (Solid Möbius strips as algebraic surfaces, Books: Munkres,
Chapter 2

Metric Spaces

The topic of metric spaces belong to the branch of mathematical analysis, however we believe it is convenient to present an introduction in metric spaces since it helps to provide a pool of examples for topological concepts as well as a reference to appreciate the generality of topology. The concept of “metric space” was introduced by Maurice Fréchet\footnote{https://en.wikipedia.org/wiki/Maurice_Ren%C3%A9_Fr%C3%A9chet} in his Ph.D. thesis in 1906. However Fréchet called the metric spaces as E spaces. The name metric spaces is due to Hursdoff in 1914 \footnote{Hausdorff, F. (1914). Grundzüge der Mengenlehre. Verlag von Veit and Company, Leipzig.}

2.1 Definition

The center concept of metric spaces is a definition of a “metric” or distance. This is precisely why they are called metric spaces. The definition of metric space follows.

**Definition 2.1.1.** A **metric space** is an ordered pair \((M, d)\) where for \(M\) is a set and \(d\) is a **metric** or **distance** on \(M\) defined as a function

\[
d : M \times M \to \mathbb{R}
\]

satisfying the following properties:

For \(x, y, z \in M\),

(i) **No negativity**: \(d(x, y) \geq 0\)
(ii) **Identity of Indiscernible Points:** \(d(x, y) = 0 \iff x = y\)

(iii) **Symmetry:** \(d(x, y) = d(y, x)\)

(iv) **Triangle inequality:** \(d(x, y) \leq d(x, y) + d(y, z)\).

It is easy to show that the first property follows from the other three, however the non-negativity property is very important and so it is usually included in the definition of metric spaces. It is common, to simplify notation, to overload the symbol \(M\) both for metric spaces and for the set \(M\).

The concept of metric space was introduced by Maurice René Fréchet in 1906.

### 2.2 Examples

We provide a pool of examples which will be useful not only to illustrate the concept of metric and metric spaces, but to serve as a reference when visiting the topic of topology. We do not prove the validity of all the examples as metrics and those leave those unproven to the reader as exercises.

There are two sources to generate examples. The sets and the metric. That is, for a given metric we can use it under different contexts. Of course different metrics provide a new category of examples. We use the metric as a way to classify the examples, and show how the example is extended to other spaces based on the domain of definition. We start with a metric that can be defined in any set whatsoever, then we introduce a sequence of examples from the real numbers in one dimension, to the real numbers in any dimension as well as the complex number, then the complex or reals in any nu

#### 2.2.1 The Discrete Metric

Define the distance called **discrete metric** between two points by the formula

\[
d(x, y) := \begin{cases} 
0 & \text{if } x = y \\
1 & \text{if } x \neq y.
\end{cases}
\]

Note the similarity of this definition with the Kronecker delta, hence sometimes the notation \(d(x, y) = \delta(x, y)\) is used for this metric. It should be obvious that \(d(x, y)\) as defined here satisfies the axioms of a metric.
This is the most general metric since any set can be provided with this metric, even if the elements are not numbers or functions. Any set can be endowed with a discrete metric. However it is not of much use except for counter-examples where it proves to be very handy.

Except for this example, the examples below have some kind of structure or flow through them. We can think of the idea of degrees of freedom. Starting with one degree of freedom, the measure, distance, or size are unanimously defined by the absolute value. However as the degrees of freedom go from one to several dimensions the concept of metric could get more involved. Do we say that a box (just three dimensions here) is larger than the other because the surface area?, volume?, length of its largest size?, diagonal, etc. What if we go to a infinite dimensions. We note that there could be many ways to measure objects of such a vast collection of degrees of freedom, and offer here just a few examples which are commonly used in the literature. At the end of the section 2.2.4.1 we bring again this discussion and see up to which level the metrics defined there respond our questions and what questions are left unsolved.

### 2.2.2 The Absolute Difference Metric

The absolute difference metric is used only in subsets of real numbers, since the absolute value is defined only there. That is for any two \( x, y \) real numbers the absolute difference metric is given by

\[
d(x, y) := |x - y|.
\]

The first two axioms of a metric are trivially satisfied by this distance. Let us show the triangular inequality. We first show that

\[
|x + y| \leq |x| + |y|
\]

\[
\begin{align*}
|x + y| &= \sqrt{(x + y)^2} \\
&= \sqrt{x^2 + 2xy + y^2} \\
&\leq \sqrt{x^2 + 2|x|y + y^2} \\
&= \sqrt{(|x| + |y|)^2} \\
&= |x| + |y|
\end{align*}
\]

(2.2.2)
Now, directly from this, we get $|x-y| = |(x-z)+(z-y)| \leq |x-z| + |z-y|$. We can consider $x, y$ to be complex numbers and develop a similar proof (left to the reader).

### 2.2.3 The Norm Induced Metric

This is perhaps the most general and used metric. The normed is defined in the most general global context in Normed Vector Spaces and in particular in Banach Spaces.

**Definition 2.2.4.** Given a field $\mathbb{F}$ (complex or real numbers, for example) and two vectors $x, y \in V$ where $V$ is some vector space and a scalar $\alpha \in \mathbb{F}$, the following properties constitute the definition of a norm $\|\cdot\|$:

(i) **No negativity** $\|x\| \geq 0$

(ii) **Identity of Indiscernible Points:** $\|x\| = 0 \iff x = 0$

(iii) **Absolute Homogeneity** $\|\alpha x\| = |\alpha|\|x\|$

(iv) **Triangle inequality:** $\|x + y\| \leq \|x\| + \|y\|$.

From the first property we have $\|0\| = 0$ and $\|-x\| = \|x\|$, so by the triangle inequality $\|x\| \geq 0$. That is again, as in the definition of metric, the first property is a consequence of the last three. Still, this property is usually included in the definition of norm.

The metric induced by the norm is

$$d(x, y) := \|x - y\|.$$  

It is clear that this metric satisfies the axioms attached to its definition, however the opposite is not always true. That is, can we say that given a metric, this metric induces a norm? In other words, if we define the norm as $\|x\|$ as $d(x, 0)$ (the natural definition to think of) then this really a norm? The answer is no and here is where the discrete topology comes handy. Let us see:

Assume that we define a norm as $\|x\| = d(x, 0)$, where $d$ is the discrete metric. Let us examine the absolute homogeneity property (iii). By using $\alpha \neq 1$ and $x \neq 0$,
\[ \|\alpha x\| = d(\alpha x, 0) = 1 \]

so

\[ 1 = \|\alpha x\| = |\alpha|\|x\| = \alpha. \]

So every norm is a metric but every metric is no necessary a norm. Another example of a metric which is not induced from the norm is presented in section 2.2.8.1.

In this context the metric spaces are more general than the normed spaces, and we can say that the normed spaces are a proper subset of the metric spaces.

Let us see some examples:

### 2.2.4.1 The \( \mathbb{F}^n \) space

Here \( \mathbb{F} = \mathbb{R} \) or \( \mathbb{F} = \mathbb{C} \). Think about two objects \( x = (x_1, x_2, \ldots, x_n) \) and \( y = (y_1, y_2, \ldots, y_n) \) in \( \mathbb{F}^n \). We can define here the following metrics

- **Taxi cab** or **Manhattan** norm. Think of the following norm:

\[
\|x\| := \sum_{i=1}^{n} |x_i|
\]

The proof can be done by mathematical induction where the induction step is the same as that in 2.2.2. However we see that this is particular case of the distance formula defined in equation 2.2.3 for the case \( p = 1 \). So a proof of the more general should be sufficient for any particular case of it.

The name of this norm is clear since a taxi can not cut through the corners. The corresponding metric is

\[
d(x, y) := \|x - y\| = \sum_{i=1}^{n} |x_i - y_i|
\]
• **The Euclidean** or **Pythagoric** norm. This is computed by the usual distance

\[ \|x\| := \sqrt{\sum_{i=1}^{n} x_i^2} \]

from which the distance formula follows as

\[ d(x, y) := \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2} \]

The previous two norms can be seen as a special case of a more general norm known as the \( \ell^p \) norm. This is

• **The \( \ell^p \) norm** is defined as

\[ \|x\|_p := \left( \sum_{i=1}^{n} |x_i|^p \right)^{1/p} \quad (2.2.3) \]

where in the first case \( p = 1 \) and in the second \( p = 2 \). It can be shown in general that for any real \( 1 \leq p \leq \infty \), \( \|x\|_p \) is indeed a norm.

Proving that the \( \ell^p \) is a norm for \( \mathbb{R}^n \) as well as for sequence spaces is not easy. The difficulty is centered at the triangular inequality. The triangular inequality for norms in the \( \ell_p \) space is known as the Minkowski inequality. This is proven in Appendix B, and more specifically in section B.1.5.

The Minkowski inequality generalizes all the way to the continuous where instead of sums we have integrals.

What happens if \( p < 1 \)? We could set \( p < 1 \) and still get a norm. For example for \( p = 1/2 \) we have

\[ \|x\|_{1/2} = \left( \sum_{i=1}^{n} |x_i|^{1/2} \right)^2. \]
2.2. EXAMPLES

The unit ball in \( \mathbb{R}^2 \), that is the set of points \( \mathbf{x} \in \mathbb{R}^2 \), such that \( \| \mathbf{x} \|_{1/2} = 1 \) is shown in Figure 2.3 and it is non-convex. That is, you can find a segment that joints two points on the inside of the curve which is no totally contained in the curve. Non-convex norms are not well suited for inversion problems due non-uniqueness. A distance to a convex boundary is unique, a distance to a non-convex boundary could have infinite number of solutions. The best example is to think about measuring distance from the outside of a circle to its boundary, (one solution) versus measuring distance from the center of the circle to the boundary (non-countable infinite many solutions). We do not provide examples for non-convex norms, but this document by Keith Conrad has plenty of material on the topic.

The corresponding metric for the \( \ell^p \) norm is

\[
d_p(\mathbf{x}, \mathbf{y}) := \left( \sum_{i=1}^{n} |x_i - y_i|^p \right)^{1/p} \quad (2.2.4)
\]

We are tempted to take the limit as \( p \to \infty \). Indeed this limit exists and it can be shown that

\[
\lim_{p \to \infty} \| \mathbf{x} \|_p = \max_{i=1}^{n} |x_i|, \quad (2.2.5)
\]

from which we can define the \( \ell^\infty \) norm as

- **The \( \ell^\infty \) norm**, is defined as

\[
\| \mathbf{x} \|_\infty := \max_{i=1}^{n} |x_i|,
\]

It is easy to show that this is a norm. In fact the triangular inequality (the most complicated in most of the cases and particularly here) is proven in the following line:

\[
\| \mathbf{x} + \mathbf{y} \|_\infty = \max_{i=1}^{n} |x_i + y_i| \leq \max_{i=1}^{n} |x_i| + \max_{i=1}^{n} |y_i| = \| \mathbf{x} \|_\infty + \| \mathbf{y} \|_\infty.
\]

From this norm the distance function is

\[
d_\infty(\mathbf{x}, \mathbf{y}) := \max_{i=1}^{n} |x_i - y_i|.
\]

\(^3\)http://www.math.uconn.edu/~kconrad/blurbs/analysis/lpspace.pdf
2.2.4.1.1 Discussion: We started a discussion at the end of the section 2.2.1 about how to measure objects with several degrees of freedom. A box (in three dimensions) is a good way to put the questions that we formulated in concrete and relate measures in a box with the metrics defined here. Let us do a list of them.

1. The size of the largest side. This is given by \( \| \mathbf{x} \|_\infty = \max_{i=1}^{3} \left| x_i \right| \).

2. The perimeter or sum of lengths of its sides. This is \( P = 4 \sum_{i=1}^{3} |x_i| = 4 \| \mathbf{x} \|_1 \). The 4 is a scaling factor (because each side should be counted 4 times) in no way changes the meaning of a norm. Any norm is defined uniquely up to a non-zero scaling value. If \( \mathbf{x} = 0 \), then \( s \mathbf{x} = 0 \) for any scaling value \( s \). In addition \( \| \alpha \, s \, \mathbf{x} \| = |\alpha| \, \| s \mathbf{x} \| \), and \( \| s \mathbf{x} + s \mathbf{y} \| \leq \| s \mathbf{x} + s \mathbf{y} \| \).

3. The diagonal is computed as \( \| \mathbf{x} \|_2 = \sqrt{\sum_{i=1}^{3} x_i^2} \).

What about areas and volumes? Well, the volume is computed using the determinant of the vectors that characterize the three sides of the box converging at one point that we can call the origin. Unfortunately the determinant is not a norm. We can have non-zero matrices with zero determinant. The areas are related to minor determinants. It seems that from the formal mathematical point of view the metric does not involve multiplying mixed components of the vectors. It is obvious that the metric was not created to solve all questions about sizes of different objects under different dimensionalities, which is the topic of measure theory\(^4\) (theory of integration). However, it is useful to understand how metrics relate to day-to-day objects. Would it be useful to have some spaces of matrices such that the minors (sub-determinants of lower degree) are related to the matrix norms? If this is possible the norm equations would be really much more complicated than the simple equations shown here where there is a great deal of interactions between \( x_i \) and \( x_j \) components for \( i \neq j \). Observe that this interaction does not happen in the norms considered here. However, we should remember that a metric is a “distance” not a “volume” or an “area” and this is why it is not designed to solve problems of areas, volumes, or any other measures which correspond to the measure theory branch. It is then natural to think that whenever variables are exponentiated to a power \( n \), later the \( n^{th} \) root is

\(^4\)http://en.wikipedia.org/wiki/Measure_%28mathematics%29
taken to retrieve the dimensional character of length. It is like dimensional analysis. Of course this does not happens always. The discrete measure has nothing to do with any dimensions. It is defined dimensionless regardless the objects it deals with. We show later another norm which is dimensionless.

Finally, note that for \( n \geq 3 \) the \( \ell^p \) norm does not represents the length of the diagonal of a hyper-box but some distance that is squeezing toward the maximum norm \( \max_i |x_i| \), as the dimension increases toward infinity as indicated in equation 2.2.5.

2.2.5 The Space of Sequences

The Sequence space\(^5\) is defined as the set of elements \((x_i)_{i\in\mathbb{N}}\) such that \( x_i \in \mathbb{F} \), and \( \mathbb{F} \) is the field of complex or real numbers. We clarify the notation \((x_i) :\)

\[
(x_i) := (x_1, x_2, \cdots, x_n, \cdots),
\]

where we use parenthesis instead of curly braces “\(\{\}\)” to stress that the order is important. For example the sequence

\[
\left(\frac{1}{n}\right) = \left(1, \frac{1}{2}, \frac{1}{3}, \cdots, \frac{1}{i}, \cdots\right)
\] (2.2.6)

for which its tail converges to 0 as \( n \to \infty \).

A particular type of sequence space is the series. A series

\[
\sum_{i=1}^{\infty} x_i
\]

can be seen as a sequence \((S_n)\) where \( S_n = \sum_{i=1}^{n} x_i \).

We observe that the space of sequences can be seen as a special case of a space of functions from the natural to the numbers \( \mathbb{F} \). That is we could also define a sequence as

\[
(x_i) : \mathbb{N} \to \mathbb{F}
\]

\[
i \mapsto x(i) = x_i
\]

We discuss the space of functions in section 2.2.7.

\(^5\)http://en.wikipedia.org/wiki/Sequence\_space
In the same fashion that we define norms for \( \mathbb{R}^n \) we can define norms for sequences. That is, we can define the \( \ell^p \) norm for a sequence \((x_i)\) as
\[
\| (x_i) \|_p := \left( \sum_{i=1}^{\infty} |x_i|^p \right)^{1/p}, \quad 1 < p < \infty
\]
whenever the norm converges (that is, when this is a finite number). The crux of proving the norm axioms rests in the triangular inequality, which is the Minkowski inequality shown in Appendix B and particularly in section B.1.5.

The metric corresponding to the \( \ell^p \) norm is
\[
d[(x)_i, (y)_i] := \left( \sum_{i=1}^{\infty} |x_i - y_i|^p \right)^{1/p}, \quad 1 < p < \infty
\]

A sequence space for which \( \| . \|_p \) is bounded for each element is known as an \( \ell^p \) space.

While in the finite case we defined \( \| x \|_\infty \) as the \( \max_{i=1}^n |x_i| \), in the sequence space we need to be more careful. For example the sequence \( S = (1 - 1/n) \) does not have a maximum, but is has a supremum \(^6\). That is a number \( s = \sup \{ (s_n) \} \) such that for any \( \epsilon > 0 \), there exists a member \( s_n \), with \( s - \epsilon > s_n \). This supremum is, for this particular example \( s = 1 \). We define then
\[
\| (x_i) \|_\infty := \sup \{ |x_i| \}
\]
hence we can define the distance as
\[
d[(x)_i, (y)_i] := \sup \{ |x_i - y_i| \}
\]

In the following examples we extend from the finite dimensional spaces to the infinite dimensional spaces.

2.2.6 The Product Space

Given two spaces \( A_1 \) with norm \( \| . \|_1 \) and \( A_2 \) wit norm \( \| . \|_2 \) we can define the product space \( A_1 \times A_2 \) as
\[
A_1 \times A_2 := \{ (x_1, x_2) : x_1 \in A_1, \ x_2 \in A_2 \}.
\]

\(^6\)see Appendix C section C.1.2 for more information about the supremum
and a norm
\[ \| (x_1, x_2) \| := \max\{ \| x_1 \|, \| x_2 \| \}. \]

It is easy to prove that this is a well defined norm. If \( A_1 = A_2 = A \), we write \( A^2 \). This is the case of \( \mathbb{R}^2 \) and \( \mathbb{C}^2 \).

Likewise the product distance can be defined as
\[ d[(x_1, x_2), (y_1, y_2)] := \max\{ \| x_1 - y_1 \|, \| x_2 - y_2 \| \}. \]

The previous definitions can be extended in the same way to the product of a finite number of sets \( \prod_{i=1}^{n} A_i \), and we can see \( \mathbb{R}^n \) and \( \mathbb{C}^n \) as special cases of this product. More generally if \( A \) is a index with a countable number of elements we defined
\[ \prod_{i \in A} A_i = \{ (x_i) : i \in A \} \]

Here \((x_i)\) is a sequence where each \( x_i \) is an element of \( A_i \). In this case we cannot guarantee that there would be a maximum, but we can find a supremum. That is, we define
\[ \| (x_i) \| = \sup_{i \in A} (\| x_i \|_i). \]

and the induced metric is
\[ d[(x_i), (y_i)] = \sup_{i \in A} (\| x_i - y_i \|_i). \]

If all spaces \( A_i \) are equal, let us say \( A_i = A \) we can use the notation \( A^A \). In particular if \( A = \mathbb{N} \) we can call the product space as \( A^\mathbb{N} \). Here are a few spaces of interest, which are all product spaces:

- \( \mathbb{R}^\mathbb{N} \): The space of real sequences.
- \( \mathbb{C}^\mathbb{N} \): The space of complex sequences.
- \( \mathbb{R}^\mathbb{Z} \): The space of real bilateral sequences. That is here
  \[ (x_i) = (\cdots x_{-i}, x_{-i+1}, \cdots, x_{-1}, x_0, x_1, \cdots, x_i, \cdots). \]
- \( \mathbb{C}^\mathbb{Z} \): The space of complex bilateral sequences. This is a space useful in the formulation of the discrete Fourier transform.
2.2.7 The Space of Functions

The function space is a set of functions from a set $X$ to a set $Y$ which have a specific structure. For example, they form a vector space, or they could be a Hilbert space, etc. There are many types of function spaces as shown in the Wikipedia page above. They can be classified according to degree of smoothness (how many continuous derivatives can the function hold in its entire domain), by its integrability (the $L_p$ space, where $L$ stands for Lebesgue), or its boundedness. These are not the only categories, but they are the most commonly used. Much of the understanding of these spaces belongs to the branch of Functional Analysis and no proofs of the properties of these functions are provided here. The space of functions is too broad and we only consider a few for the list of examples below.

2.2.7.1 The Space of Bounded Functions $B(A)$

Assume a domain $A$ and the space of all bounded functions from $A$ to the real line $\mathbb{R}$. Since the functions are bounded they have a supremum and so we can define the supremum norm:

$$\|f(t)\| := \sup_{t \in [a,b]} |f(t)|$$

from which the inherited metric is

$$d[f(t), g(t)] := \sup_{t \in [a,b]} |f(t) - g(t)|.$$  

2.2.7.2 Continuous Functions in an Interval $C^0[a, b]$  

Let us think of the space of real continuous functions defined in an interval $[a, b]$. That is, the space $C^0([a, b])$ such that if $f \in C^0([a, b])$ then $f : [a, b] \to \mathbb{R}$ is continuous. We use the “0” in the upper index to indicate that we do not impose the existence of the derivative. This is a vector space and since the space is bounded (the proof of this is Functional Analysis texts) so we can assign the supremum norm. That is we

$$\|f(t)\| := \sup_{t \in [a,b]} |f(t)|$$

---

7 http://en.wikipedia.org/wiki/Function_space
8 http://en.wikipedia.org/wiki/Functional_analysis
2.2. EXAMPLES

Figure 2.1: We show the function $f_n(x) = x^n$, for $n = 1, 5, 10 \cdots 100$, $x \in [0,1]$, and see how the limit (thick red line and dot) approaches the discontinuous function $f(x) = 0$, $0 \leq x < 1$, and $f(1) = 1$.

Note that this is similar to the example of infinite sequences where now the index, instead of running through the natural numbers $\mathbb{N}$ is running continuously along the interval $[0,1]$. In fact since any closed bound set in the real space has a maximum and a minimum we can change the word “sup” above by the word “max”.

This norm is very useful since it is associated with the uniform convergence. That is we say that a sequence $f_1, f_2, \cdots f_n, \cdots$ converges uniformly to $f$ if and only if $\| f_n - f \| \to 0$, as $n \to \infty$. The norm here is the sup norm shown above. This is convergence on the large or global convergence, also known as strong which is different from the weak convergence or point-to-point which says that a sequence of functions $f_1, f_2, \cdots, f_n, \cdots$ converges at a point $x$, if $\lim_{n \to \infty} f_n(x) = f(x)$. The importance of uniform convergence could not be underestimated. The commutation of limit operators, derivative, and integral operators, infinite sums with integrals and derivatives, etc. is guaranteed by uniform convergence. The conservation of continuity is also guaranteed by uniform convergence. This is not the case for point-to-point convergence as shown in Figure 2.1 where the point to point convergence of continuous function breaks the continuity at $x = 1$. 
2.2.7.3 Space of Continuous Functions with $k$ Continuous Derivatives $C^k[a, b]$

By definition $f : [a, b] \to \mathbb{F}$ belongs to $C^k[a, b]$ if $f$ has up to $k$ derivatives and the $k^{th}$ derivative is continuous. It is clear that if $p < q$, $C^q[a, b] \subset C^p[a, b]$, and

$$C^\infty[a, b] = \bigcap_{i=0}^{\infty} C^i[a, b] \tag{2.2.7}$$

It is interesting to observe that as we add constraints to the space, the metric in the original space might need changes to adjust to the new constraints.

For example, the space $C^1[a, b]$ is the space of functions with one continuous derivative. We can show examples where the sup norm will not suffice to characterize this space correctly. For example define the sequence (series)

$$f_n(x) = \sum_{i=1}^{n} \frac{\sin nx}{n^2}.$$ 

In the supremum metric this series is uniformly convergent. However its derivative is

$$f'_n(x) = \sum_{i=1}^{n} \frac{\cos nx}{n}.$$ 

which is not uniformly convergent. For example, consider $x = 0$. Here the $f'_n(x)$ does not converges since the harmonic series $\sum 1/n$ is divergent. In fact this series converges in the complex plane as long as Re($z$) is not a multiple of $2\pi$ and it does to the function

$$f'(z) = -\frac{\log(2 - 2\cos z)}{2}.$$ 

This can be shown by doing contour integration in the upper half of the complex plane along a small path $z : z = x + i\epsilon, 0 < x < \pi/2$ and at the end letting $\epsilon$ approach zero. For details see this discussion.

It is then necessary to limit the scope of functions in such a way that the derivatives could be bounded and we could find their supremum. A norm that do this would be

[^9]: http://math.stackexchange.com/questions/321736/the-fourier-series-sum-n-1-infty-1-n-cos-nx/1316593#1316593
2.2. EXAMPLES

\[ \|f\| := \sup_{x \in [a,b]} |f(x)| + \sup_{x \in [a,b]} |f'(x)|. \]

That this is a norm is consequence to the fact that if \( \| \cdot \|_1 \) is a norm and \( \| \cdot \|_2 \) is a norm, then \( \| \cdot \|_1 + \| \cdot \|_2 \) is a norm as well, which is easy to prove.

It can be shown that this is a norm. In general, if \( f \) can take up to \( k \) derivatives and it is continuous in its \( k^{th} \) derivative we say that \( f \in C^k[a,b] \), where the norm in this space is defined as

\[ \|f\| := \sum_{i=0}^{k} \sup_{x \in [a,b]} |f^{(i)}(x)| \tag{2.2.8} \]

where this is also a norm, by applying many times the property of the sums of two norms. The more conditions we impose, the smaller is the space with \( C^\infty[a,b] \) shown in equation 2.2.7 being the smallest. We have the proper chain of subsets

\[ C^\infty[a,b] \subsetneq \cdots \subsetneq C^k[a,b] \subsetneq \cdots \subsetneq C^0[a,b] = C[a,b] \subsetneq B[a,b]. \]

That \( C^\infty[a,b] \) is not empty is easy to see. For example all polynomials are \( C^\infty[a,b] \), analytic functions in complex variables have an infinite number of continuous derivatives. Here \( f(z) = \exp(\lambda z), g(z) = \cos \omega z \) are two of infinite non-countable collection of \( C^\infty[a,b] \) functions. Still the size of the space of functions is so large that most continuous functions are nowhere differentiable \(^{10}\) and we add that most functions are no continuous. We could say, statistically, that we work with the zero percent of functions out there, still we get some great benefits from the little functions we use.

We ask what is the norm imposed in \( C^\infty[a,b] \)? If \( k \) is finite, the definition is given by equation 2.2.8 but we can not define the norm for \( C^\infty[a,b] \) with an equation such as

\[ \|f\| = \sum_{i=0}^{\infty} \sup_{x \in [a,b]} |f^{(i)}(x)| \]

\(^{10}\)http://homepages.math.uic.edu/~marker/math414/fs.pdf
CHAPTER 2. METRIC SPACES

In most cases this number does not even exists. We leave this as an open question for now and return to this point later in the text when we talk about Fréchet Spaces. There are even spaces which extend to fractional derivatives but we will not consider those here.

The metric associated with this norm is then

\[
d(f, g) = \sum_{k=0}^{k} \sup_{x \in [a,b]} |f^{(k)}(x) - g^{(k)}(x)|.
\]

2.2.7.4 Lebesgue Spaces

The study of Lebesgue Integration\(^{11}\) belongs to the area of measure theory\(^{12}\) and it is out of the scope of this document. We provide the main ideas of Lebesgue integration and the comparison between Lebesgue and Riemann integration in a nutshell.

- Lebesgue integration is an extension of Riemann integration. Every function which is Riemann integrable is Lebesgue integrable, but the opposite is no necessarily true. The characteristic function or indicator function\(^{13}\) of the rational numbers in the interval \([0, 1]\), that is

\[
\chi_{[0,1]}(Q) = \begin{cases} 
1 & x \in \mathbb{Q} \\
0 & x \not\in \mathbb{Q}
\end{cases}
\]

is an example of a function which is Lebesgue integrable but not Riemann integrable.

- Lebesgue integration is not concerned with the number of discontinuities unless it is of measure other than 0. Riemann integration fails to account for a countable infinite number of discontinuities such as those in the example above.

\(^{11}\)http://en.wikipedia.org/wiki/Lebesgue_integration

\(^{12}\)http://en.wikipedia.org/wiki/Measure_%28mathematics%29

\(^{13}\) Also known as indicator function. Given a non-empty set \(A\), the indicator function \(\chi_{A}\) is defined as

\[
\chi_{A}(x) = \chi_{x \in A} = 1(x) := \begin{cases} 
1 & x \in A \\
0 & x \not\in A
\end{cases}
\]

The symbol “\(\chi\)” is the first letter of the word “characteristic” in Greek.
2.2. EXAMPLES

Figure 2.2: The Riemann idea of integral in blue shows vertical rectangular slices along a curve based on a partition of the $x$ axis. The Lebesgue idea of integral in red shows horizontal slices along the curve based on a partition of the $y$ axis. Figure taken from Wikipedia.

- Lebesgue integration considers partitions along the $y$ axis. In this way, it measures the sets in the $x$ axis and based on this, computes the partial areas before going to the limit. In this way, instead of thinking in the Riemannian idea of $\sum f(x_i) \Delta x_i$, it thinks of $\sum \mu[f^{-1}(x_i)] \Delta y_i$, where $\mu[f^{-1}(\xi_i)]$ is the measure of the inverse image of a point $\xi \in [y_{i-1}, y_i]$ of the partition $y_0, y_1, \cdots, y_n$ of the $y$ axis. Figure 2.2 (from Wikipedia) illustrates this point.

- Lebesgue integral is the same for a function $f$ that for any function which is different to $f$ only on a set of measure 0. In this way an equivalence relation is defined. $f$ is equivalent to $g$ ($f \sim g$) if the measure of the set where the two functions are different is zero. That is if

$$\mu\{x : f(x) \neq g(x)\} = 0,$$

and we say that $f$ almost everywhere equal to $g$.

The existence of such integrals is, as many things in analysis, the guarantee that its evaluation is finite. The definition, properties, and ways to compute it are outside of the scope of this document. The reader who does not want to extend his/her knowledge base to Lebesgue integration can think of
what we will write next as Riemann integrals and still walk in a safe ground, except that when the words “almost everywhere” are used in Lebesgue, they should be replaced by “everywhere” under Riemann.

As in the $C^k[a,b]$ spaces, and to simplify, we choose a closed interval $[a, b]$ but the set could be generalized. The $L^p[a,b]$ space of $p^{th}$-power integral functions, is defined as the set of function such that

$$
\|f\|_{L^p([a,b])} := \left(\int_a^b |f(x)|^p \, dx\right)^{1/p} < \infty
$$

for $1 < p < \infty$. Observe the similarity with the space of sequences $\ell^p$ discussed in section 2.2.5. Actually the $\ell^p$ norm can be obtained from the $L_p$ norm when going from the continuous to the discrete measure. For example, replacing the continuous function with a comb of delta functions. As in the discrete case we can take the limit as $p \to \infty$ and if the functions are bounded we have

$$
\|f\|_{L^\infty([a,b])} := \sup_{x \in [a,b]} |f(x)|.
$$

That this is a norm is shown by verifying the axioms of norm which is an easy task except for the triangular inequality which happens to be the Minkowski inequality shown in Appendix B, section B.1.5.2.

The norms above induce the metric

$$
d_p(f,g) := \left(\int_a^b |f(x) - g(x)|^p \, dx\right)^{1/p}
$$

and

$$
d_\infty(f,g) := \sup_{x \in [a,b]} |f(x) - g(x)|.
$$

2.2.8 Bounded Metrics

Bounded metrics will prove to be useful when using topologies on infinite product spaces. Any metric space is bounded under a bounded metric, and this is a useful thing.
2.2. EXAMPLES

2.2.8.1 Shrink Metric

Let \( x, y \) two points of a metric space \((X, d(.,.))\). Then

\[
d_s(x, y) := \frac{d(x, y)}{1 + d(x, y)},
\]

is a metric.

**proof:** The first three axioms are a direct consequence of the distance definition on the right side of the expression. Let us prove the triangular inequality.

We use the auxiliary function

\[
f(t) = \frac{t}{1 + t},
\]

with derivative \( f'(t) = 1/(1 + t^2) \), so \( f \) is monotonic increasing, and so then, since \( d(x, y) < d(x, z) + d(z, y) \) we have

\[
\frac{d(x, y)}{1 + d(x, y)} \leq \frac{d(x, z) + d(z, y)}{1 + d(x, z) + d(z, y)}
\]

\[
= \frac{d(x, y)}{1 + d(x, y) + d(z, y)} + \frac{d(z, y)}{1 + d(x, z) + d(z, y)}
\]

\[
\leq \frac{d(x, y)}{1 + d(x, y)} + \frac{d(z, y)}{1 + d(z, y)}
\]

That is

\[
d_s(x, y) \leq d_s(x, z) + d_s(z, y),
\]  \( (2.2.10) \)

and \( d_s \) is a metric.

In particular

\[
d(x, y) = \frac{\|x - y\|}{1 + \|x - y\|}
\]  \( (2.2.11) \)

is a metric. Note that this is another example of a metric which is not induced by the norm, still it could be defined from the norm. That is,

\[
\|x\|_s = \frac{\|x\|}{1 + \|x\|}
\]

is not a norm (show this).
2.2.8.2 Shrink Infinite Product Spaces

Next example is a generalization of the previous one, when we consider an infinite number of spaces with possibly different metrics.

If \( x = (x_i) \) and \( y = (y_i) \) are points of a some space \( A_i \), with metric \( d_i \), then

\[
d(x, y) = \sum_{i=0}^{\infty} \frac{1}{2^i} \frac{d_i[(x_i), (y_i)]}{1 + d_i[(x_i), (y_i)]}
\]

is a metric in the product space \( \prod_{i=1}^{\infty} A_i \).

**proof:** We first note that the summation index \( i \) starts in zero and no in one. Many texts use one. Let us see why 0 might be a “better” choice. Assume that all points are separated by 1, that is \( d_i[(x_i), (y_i)] = 1, \forall i \). Then the sum with \( i \) starting at zero is

\[
\frac{1}{2} + \frac{1}{4} + \cdots \frac{1}{2^n} + \cdots = 1.
\]

If instead \( i \) starts in 1, the sum would be

\[
\frac{1}{4} + \cdots \frac{1}{2^n} + \cdots = \frac{1}{2}.
\]

Do we want to bound this by 1? Probably there is no need, the function \( d(\ldots) \) is still a metric, but being all the distances 1, the product distance is nicer being 1.

We only prove the triangular inequality since the other axioms are a direct inheritance from the metrics under the summation sign.

Based on how we solved the previous problem let us consider only one arbitrary term of this sum and worked on it until getting what we need. Consider third point \( z = (z_i) \).

Repeated application of the equation [2.2.10] into the infinite sum yields

\[
\sum_{i=0}^{\infty} \frac{1}{2^i} \frac{d_i[(x_i), (y_i)]}{1 + d_i[(x_i), (y_i)]} \leq \sum_{i=0}^{\infty} \frac{1}{2^i} \frac{d_i[(x_i), (z_i)]}{1 + d_i[(x_i), (z_i)]} + \sum_{i=0}^{\infty} \frac{1}{2^i} \frac{d_i[(z_i), (y_i)]}{1 + d_i[(z_i), (y_i)]}
\]

or

\[
d(x, y) \leq d(x, z) + d(z, y).
\]
2.3 Topological Definitions

The concept of ball is in the center of the topology of a metric space. Open sets, neighborhoods, limits, continuity, etc., can be defined in terms of balls. Starting with the definition of a ball we present a collections of basic definitions built upon the concept of a ball.

Definition 2.3.1. Given a metric space \((X, d)\), a set \(A \subset X\), a set \(O \subset X\), a point \(x \in X\), a point \(x_0 \in X\), and a real number \(r\) we provide the following set of definitions:

- A ball centered at \(x_0\) and radius \(r\) is the set \(B(x_0, r)\) defined as
  \[ B(x_0, r) := \{ x : d(x_0, x) < r \} \]

- A sphere centered at \(x_0\) and radius \(r\) is defined as the set
  \[ S(x_0, r) := \{ x : d(x_0, x) = r \} \]

- A closed ball centered at \(x_0\) and radius \(r\) is the set \(\overline{B}(x_0, r)\) defined as
  \[ \overline{B}(x_0, r) := \{ x : d(x_0, x) \leq r \} \]

- A set \(A\) is open if for any point \(x \in A\), there is a ball such that \(B(x, r) \subset A\). According to Gregory Moore\(^{15}\) Lebesgue introduced the word “open” in his doctoral dissertation in 1902.

- A neighborhood of a point \(x\) is any open set \(O\), such that \(x \in O\). A set \(C\) is closed if its complement \((X \setminus C)\) is open. Gregory Moore also claims that Cantor provided the first definition of a closed set in 1884. His original definition was that a set is closed if it contains all its limit points. We provide the definition of limit point in \(\text{2.4.2}\). It can be proved that Cantor’s original definition coincides with the definition provided here.

\(^{14}\) Clearly from the definitions above \(S(x_0, r) = \overline{B}(x_0, r) \setminus B(x_0, r)\), and \(S(x_0, r) = \partial B(x_0, r)\). Spheres and balls could be very different from those in the Euclidean space. For example, in the discrete metric, balls (closed and open) are the entire space \(X\) and spheres are the whole space except the center \(x_0\). In the normalize metric \(\text{2.2.8.1}\) unit balls are the whole space \(X_0\), and unit spheres are empty sets \(\emptyset\).

\(^{15}\) http://www.sciencedirect.com/science/article/pii/S0315086008000050
A point \( x_0 \) is in the **interior** of a set if there is a neighborhood \( O \) such that \( x \in O \).

- The **interior** of a set \( A \) is the collection of all its interior points. This is noted as \( \text{Int}A \) or \( \overset{\circ}{A} \).

- The **exterior** of a set is the interior of its complement. That is

\[
\text{Ext}A = \text{Int}(A^c)
\]

- A point is called a **boundary** point of \( A \) if it is not in the interior, neither in the exterior of \( A \), word frontier is used and notations for this set are \( bd(A), fr(A) \), and \( \partial A \). We used \( \partial A \) to note the boundary of a set. In symbols:

\[
x \in \partial A \iff x \notin \text{Int}A \text{ and } x \notin \text{Ext}A
\]

In other words we could have said \( X = \text{Int}A \cup \text{Ext}A \cup \partial A \).

A good exercise is to visit all examples on the previous section and rehearse each definition here on each of the examples. We will visit a few examples, but before doing so we present a few properties as theorems which sometimes are used as definitions.

The geometry of a ball is translation invariant. The radius of the ball will scale its size without changing its shape. To know how a ball looks like we can centered it at 0 and use a radius \( r = 1 \). A ball centered at 0 in \( \mathbb{R}^2 \) with the \( \ell^p \) norm is defined as

\[
B(0, 1) = \{(x, y) : |x|^p + |y|^p < 1\}
\]

Note that for \( p = 2 \) (the Euclidean distance) a ball is a circle \(^{16}\) If \( p \neq 2 \) new shapes arise.

Figure\[\text{2.3}\] illustrates a set of different unit balls under different \( \ell^p \) metrics. The color curves are the spheres \( S(0, 1) \) and the white regions bounded by the spheres are the balls.

Let us forget for the moment about the most interior curve corresponding to \( p = 1/2 \). The smallest ball, corresponding to \( p = 1 \), is a 45 degree tilted square. The largest ball corresponds to the lines \( \|x\|_\infty = 1, 0 \leq x \leq 1, \)

\(^{16}\) a ball for \( \mathbb{R}^3 \) and an open interval for \( \mathbb{R} \)
2.3. **TOPOLOGICAL DEFINITIONS**

Figure 2.3: Set of unit balls are shown for the values of $p = 0.5, 1.0, 1.5, \cdots 4.0$. The colorbar labeled by $p$ helps to identify the correct balls. Note that the most inside figure corresponds to the value $p = 1/2$.

which is the infinite norm. We see visually from the unit ball figure how the infinite norm is the limit of the $\ell^p$ norms as $p \to \infty$. The space of unit balls as function of $p$ is a continuous deformation starting at a tilted (45 degrees tilted) square and immediately after that converting it to a smooth until a circle (ball or sphere), and from there starts flattening the top and bottom into the non tilted square containing the starting square at $p = 1$. On the other direction $p < 1$, the figure should shrink by kissing the coordinate axis closer and closer until it merges on them.

The next proposition is a theorem for metric spaces but it is an axiom for topology as we will see in the next chapter.

**Theorem 2.3.1.** In any given metric space $(X,d)$ the following statements are true:

(i) An arbitrary union of open sets is open.

(ii) A finite intersection of open sets is open.

*Proof.*
(i) Let us assume that \( A = \bigcup_{\alpha \in A} O_\alpha \), where \( O_\alpha \) is open, and \( \alpha \) is an index over a set \( A \).

Let us pick \( x \in A \), then there \( x \in O_\alpha \) for some \( \alpha \in A \), and since \( O_\alpha \) is open, there is ball \( B(x, r) \subset O_\alpha \subset A \). That is \( A \) is open.

(ii) Let us assume \( A = \bigcap_{i=1}^{n} O_i \), for some finite number \( n \). Pick \( x \in A \), then \( x \in A_i \), for each \( A_i \). Hence there is a sequence of balls \( B(x, r_i) \) with radius \( r_i \) for each \( A_i \). We pick \( r = \min_{i=1}^{n} r_i \), and sure enough, \( B(x, r) \subset A_i \) for all \( i = 1 \cdots n \). Hence \( B(x, r) \subset A = \bigcap_{i=1}^{n} A_i \).

By taking complements in the previous theorem we find the following corollary:

**Corollary 2.4.** In any given metric space \((X, d)\) the following statements are true:

(i) An arbitrary intersection of closed sets is closed.

(ii) A finite union of closed sets is closed.

We provide an discussion in the next chapter about why the topology axioms are not defined in terms of closed sets instead of open sets. We now show that given a set \( A \subset X \), all sets \( O \setminus A \) where \( O \) are open sets in \( X \), are open in \( A \).

**Theorem 2.4.1.** Let \( A \subset X \), and \( O \) open in \( A \). Hence the set \( A \setminus O \) is open in \( A \).

**Proof.** Let us assume \( x \in A \setminus O \). Then \( x \in A \) and \( x \in O \) and since \( O \) is open in \( X \) there is a ball \( B(x, r) \subset X \). Let us now define \( B_A(x, r) = B(x, r) \cap A \). By definition this is the set of values in \( A \) closer to \( x \) than \( r \), so \( B_A(x, r) \) is an open ball in \( A \) having \( x \). Then \( A \setminus O \) is open in \( A \).

This property shown here as a theorem will be presented as a definition for the induced topology in the next chapter.

We now define convexity.

**Definition 2.4.1.** A set \( A \) is said **convex** if for any two points in the set \( x \in A \), \( y \in A \), the segment that joints them, that is the set of points \( \{ x + ty : t \in [0, 1] \} \) is totally included in \( A \).
2.3. **TOPOLOGICAL DEFINITIONS**

For an example look at Figure 2.3. All unit balls for \( p \geq 1 \) are convex. The unit ball for \( p = 1/2 \) is not convex, and in general any unit ball with \( p < 1 \) is not convex.

**Definition 2.4.2.** We provide the following set of definitions:

- The **closure** of a set \( A \) noted as \( \overline{A} \) is defined as the smallest closed set which contains \( A \), that is
  \[
  \overline{A} = \bigcap \{ C \mid A \subset C \text{ and } A \text{ closed} \}.
  \]

- A set \( A \subset X \) is called **dense** or **everywhere dense** if \( \overline{A} = X \).

- A set \( A \subset X \) is called **nowhere dense** if \( X \setminus \overline{A} \) is dense.

- Let \( O \) be a neighborhood of a point \( x \). A **deleted neighborhood** of a point \( x \) is a neighborhood of \( x \) without the point \( x \). That is, \( O \setminus x \) is a deleted neighborhood of \( x \).

- A point \( x \) is called an **accumulation point** or **limit point** if every deleted neighborhood of \( x \) contains a point of \( A \). It can then be proved that any neighborhood of \( A \) will have infinite many points of \( A \), by choosing smaller and smaller neighborhoods around it. That is, \( x \) is a limit point of a set \( A \) if for any \( \epsilon > 0 \), there is some point \( y \in A \) other than \( x \), such that \( d(x, y) < \epsilon \).

- The set \( P' \) of limit points of a set \( P \) is called the **derived set**.

- A point \( x \in A \subset X \) is called **isolated** for \( A \) if the singleton \( \{x\} \) is open in \( A \).

- A set \( P \) is know as a **perfect set** if \( P = P' \). This is equivalent to say that it is closed and it contains no isolated points.

- A space \( X \) is said to be **separable** if has a countable dense subset.
According to [17] The name “limit point” was first published by Cantor in 1872 (“Grenzpunkt”), but was invented by Weierstrass. However, according to [18] It was Eduard Heine who introduced the concept of accumulation point. In addition, according to S. M. Srivastava it was also Cantor who precised the notions of accumulation points, derived sets, dense sets, and nowhere dense sets among others.

There are a couple of equivalent definitions of closed sets and isolated points. That is

**Theorem 2.4.2.** A set \( A \) is closed if and only if it contains all its limit points.

**Proof.** Let us assume that \( A \) is closed and \( x \) is a limit point of \( A \). Since \( A \) is closed its complement is open. If \( x \notin A \), then \( x \in X \setminus A \), which is open and so there is a ball \( B(x, r) \), such that \( B(x, r) \subset X \setminus A \), this means that \( x \) is not a limit point for \( A \), which contradicts the hypothesis. So \( x \in A \).

On the other hand if all limit points of \( A \) are included in the set, the complement of \( A \) is open, since for any \( x \in X \setminus A \), \( x \) is not an accumulation point so there is a ball \( B(x, r) \subset X \setminus A \). Hence \( A \) is closed. \( \square \)

Saying that a point is isolated if an only its singleton \( \{x\} \) is an open in \( A \subset X \) set does not seem to provide much insight. The open sets in \( A \) are the intersection of opens in \( X \) with \( A \) (see Theorem 2.4.1). In this sense the we can find an open ball \( B(x, r) \) of \( X \) such that \( B(x, r) \cap A = \{x\} \). We show that the set op isolated points is precisely the set of no limit points.

**Theorem 2.4.3.** A point \( x \in A \) is isolated if and only if it is not a limit point of \( A \). That is, the set of isolated points is the complement of the set of limit points.

**Proof.** Let us assume \( x \) is an isolated point of \( A \), so there is a ball in \( X \), such that \( B(x, r) \cap \{x\} = \{x\} \). This is equivalent to say that \( x \) is not a limit point, since limit points (say \( x \)) are those where each ball around them, no matter the size, contains points of \( A \), other than \( x \). \( \square \)

We formulate equivalence definitions which characterized the isolated points more clearly.

---

2.3. **TOPOLOGICAL DEFINITIONS**

On the other hand we show that open sets do not have any limit point. $O$ is open then its complement is closed and has all the accumulation points. Hence no accumulation points of $O$

**Theorem 2.4.4.** A closed set $A$ contains all the points of its boundary, and an open set $O \subset X$ has none of them.

*Proof.*

- Let us assume $A$ is a closed set and $x \in \partial A$. By definition any ball $B(x, r)$ has points of $A$ (as well as from $\text{Ext} A$) other than $x$, so $x$ is an accumulation point for $A$ and since $A$ is closed, from Theorem 2.4.2 $x \in A$.

- If $O$ is open, its complement is close. Both, a set and its complement share the boundary, so by the previous item this boundary should be in the complement of $O$. So, no points of the boundary are in $O$.

We now show that the closure of a set is the smallest closed set that contains it and that the interior of a set is the largest open contained on it.

**Theorem 2.4.5.** Given a set $A \subset X$:

- The closure of $A$, $\bar{A}$ is the smallest set $B \subset X$, which is closed and such that $A \subset B$.

- The interior of $A$, $\interior{A}$ is the largest set $B \subset A$, which that is open.

*Proof.*

- We show that that if $B$ is a closed set such that $A \subset B$, then $\bar{A} \subset B$.

  Pick $x \in \bar{A}$. If $x \in A$, since $A \subset B$, then $x \in B$. Let us assume $x \notin A$. Still $x$ is a limit point of $A$, since $x \in \bar{A}$, and since $A \subset B$, and $B$ is closed, $x$ is a limit point of $B$. We showed in Theorem 2.4.2 that a closed set contains all its limit points, then $x \in B$. Hence $\bar{A} \subset B$.

- We show that if $B$ is any open set such $B \subset A$, then $B \subset \interior{A}$. Let us assume $x \in B$, since $B$ is open there is a ball $B(x, r) \subset B$, but since $B \subset A$, $B(x, r) \subset A$. This means, by definition, that $x$ is an interior point of $A$, and so $x \in \interior{A}$. Then $B \subset \interior{A}$. 

\[\square\]
CHAPTER 2. METRIC SPACES

The definition of nowhere dense as saying that $A$ is nowhere dense in $X$, if $X \setminus \bar{A}$ is dense in $X$ is correct but very abstract. It can be shown that this definition is equivalent to say that the interior of the closure is empty. That is, $\text{Int}(\bar{A}) = \emptyset$. This is the same as saying that all points are isolated or that there are no accumulation points. A classical example of nowhere dense sets is the Cantor set explained in detail in section 2.4.4.2 below. Let us prove these statements.

**Theorem 2.4.6.** Given a metric space $X$, the boundary of $A$ is $\partial A = \bar{A} \setminus \text{˚}A$.

**Proof.**
- $\subset$: Let us assume $x \in \partial A$, so, for any ball $B(x, r)$ there are points $y \neq x$, with $y \in A \cap B(x, r)$. So $x$ is a limit point of $A$, and then $x \in \bar{A}$. Since an open set can not have boundary points (theorem 2.4.4) we find that $x \in \bar{A} \setminus \text{˚}A$, and so $\partial A \subset \bar{A} \setminus \text{˚}A$.

- $\supset$: On the other hand if $x \in \bar{A} \setminus \text{˚}A$, pick a ball $B(x, r)$. Since $x \in \bar{A}$, there is at least a $y \in B(x, r)$ with $y \in A$. Since $x \notin \text{˚}A$, $x \in X \setminus \bar{A}$, which is closed, so any ball, an in particular $B(x, r)$ has points of $X \setminus \bar{A}$. If $B(x, r) \subset A$, then $x \in \bar{A}$, so there is a $y$ such that $y \in B(x, r)$ and $y \notin A$. So $x \in \partial A$.

\[ \square \]

**Theorem 2.4.7.** A set $A$ is dense in $X$ if an only if for any $x \in X$, and $B(x, r) \subset X$, there is a $y \in B(x, r) \cap A$.

**Proof.** Let us assume $A$ is dense in $X$, then $\bar{A} = X$. Pick $x \in X$ then $x \in \bar{A}$, and so since a closed set contains all the points of its boundary (Theorem 2.4.4) then $x \in \partial A$ and so any ball around $x$ has points $y$ such that $y \in B(x, r) \cap A$. On the other hand, if for any ball $B(x, r) \subset X$, there is a $y$ such that $y \in B(x, r) \cap A$, then $x \in \bar{A}$, and so $X = \bar{A}$, so $A$ is dense in $X$.

\[ \square \]

**Theorem 2.4.8.** A set $A$ is nowhere dense if an only its interior is empty.

**Proof.**
- $\implies$: Let us assume that $A$ is nowhere dense. Then from definition $X \setminus \bar{A}$ is dense in $X$. This means that for any ball $B(x, r) \subset X \setminus \bar{A}$, there is at least some $y$ such that $y \in (X \setminus \bar{A}) \cap B(x, r)$. This means that each ball is such that $B(x, r) \not\subset \bar{X}$, and so $B(x, r) \not\subset X$, or $x \notin \bar{X}$. 

\[ \square \]
Figure 2.4: The yellow regions represent the set $A$ is not connected. The green regions are the open sets $U$ and $V$, such that they intersect $A$ and the union of those intersections reconstruct $A$ as a whole. The intersections of the four sets needs to be empty. In the lower part of the figure, in an extreme case of two kissing intervals in the real numbers $\mathbb{R}$. The intervals $(a, b)$ and $(b, c)$ form our set $A$. They are disconnected at $a$, since the two open sets (the green intervals) intersect $A$ at non-empty sets but disjoint. Think what would happen if for example we choose the interval $(a, b]$ instead?

- “$\leftarrow$”: Let us assume that $\hat{A} = \emptyset$. So for any $x \in X$, and any $r > 0$, $B(x, r) \cap (X \setminus A)$ not empty. This means that $X \setminus A$ is dense in $X$. That is $A$ is nowhere dense in $X$.

Next we define the concepts of connected and disconnected sets.

**Definition 2.4.3.** We say that a set $A$ is **disconnected** if it is possible to find a pair of open sets $U$ and $V$ such that

\[
U \cap A \neq \emptyset, \quad V \cap A \neq \emptyset,
\]

\[
(U \cap A) \cap (V \cap A) = \emptyset \quad \text{and,}
\]

\[
(U \cap A) \cup (V \cap A) = A.
\]

If no such couple of sets $U$ and $V$ exist, then we say that $A$ is connected.

Figure 2.4 illustrates this definition.

Many of the definitions here are repeated in the next chapter. We provide a list of examples that are valid here as well as for topological spaces in the next chapter.
2.4.4 Examples:

2.4.4.1 General Examples

- Closed, Open, Interior, Exterior, Boundary, Interior Point, Boundary Point:
  - In the real numbers $\mathbb{R}$ any closed interval $[a, b]$ is closed. A finite union of closed sets is closed (this last statement is in any topology true by knowing that closed are complements of open). Any open interval $(a, b)$, including $a = \infty$ or/and $b = \infty$ is open.
  - In the complex space $\mathbb{C}$ or $\mathbb{R}^2$, the open discs are open, the closed discs are closed. That is, by definition
    \[ N_r(z_0) = B(z_0, r) = \{ z : |z - z_0| < r \} \text{ is open} \]
    \[ \overline{N_r(z_0)} = B(z_0, r) = \{ z : |z - z_0| \leq r \} \text{ is closed} \]

We could say that closed an open sets fill up the space of possibilities, but that is no true. There are sets which are not close neither open. Think of an open disc. By adding a point to the discs boundary, the disc stops from being open because of that point (there is not a ball around that point which is contained totally inside the disc). Still, it is not closed either. Most of the disc has an open boundary. The interior is the same open disc (the interior of an open set is the same set), the exterior is the whole space after removing the closure of the set, and the boundary is the sphere

\[ \partial A = \{ z : |z - z_0| = r \} \]

Figure 2.5 illustrates the concept of close, open, and none in the complex plane. In addition this figure show sketches of the boundary, the interior and the exterior of the set, and a interior point $z_i$ with a neighborhood totally inside $A$ and a boundary point $z_b$ where any neighborhood will have points inside and outside of $A$.

- Dense set: Every closed set $A$ is dense into itself, since $\bar{A} = A$.
  - The rational numbers $\mathbb{Q}$ are a dense set of the real line, its closure are the real numbers that is $\bar{\mathbb{Q}} = \mathbb{R}$. This is true also for the irrational numbers, but the rational numbers are infinite countable,
2.3. TOPOLOGICAL DEFINITIONS

Figure 2.5: Illustration of open (left), closed (center) and none of the above right. We show a disc $A$ with not boundary as an open set. A disc with boundary as a closed set. We added a point to the third open disc to make it no open, nor closed. The boundary on the disc in the center coincides with the definition of boundary $\partial A$. The interior of $A$ is the open disc on the left. The exterior of $A$ is the green rectangle which is the complement of the closure of $A$. That is, $X \setminus \bar{A}$. The point $z_i$ in the first disc is an interior point and the point $z_b$ in the last disc is a boundary point.

while the irrational are infinite uncountable. The proof of this is based on the fact that each real number can be approximated as much as we want with a sequence of rational numbers. So for any ball in the real numbers, no matter how small, you will always find a rational. It is in this way that we say that the rational numbers are dense. They are everywhere in the real line, but they do not fill the real line.

The density property is very important in the theory of approximations. If we can approximate an element of a set $A$ with other objects of a set $B \supset A$ as close as we want, then we say that $A = \bar{B}$, and that means that $B$ is dense in $A$. Density is a required property if we want to assure good approximations with any degree of accuracy as we want. A famous example is the approximation of $\pi$ using rational numbers (since rationals are dense in the real line). Here is a Website\(^{19}\) that illustrates this. Another famous example is the approximation of continuous functions using polynomials. This is:

- The polynomials defined on a closed real interval $[a, b]$ are dense

\(^{19}\)http://www.isi.edu/ johnh/BLOG/1999/0728_RATIONAL_PI/
in the space of continuous functions, defined in the same interval. This is a statement of the Stone-Weierstrass Theorem\(^{20}\).

- The Trigonometrical Polynomials\(^{21}\) are dense also in the continuous functions. The Fourier series show how well we can approximate functions using trigonometrical polynomials.

It is interesting to observe that \(\partial \bar{A} \subset \partial A\), and \(\partial \dot{A} \subset \partial A\), and the three sets \(\partial \bar{A}, \partial A\), and \(\partial \dot{A}\) are not necessarily equal. Here is an example. Take \(A\) as the set of rational numbers. That is: \(A = \mathbb{Q}\). We know that \(\bar{\mathbb{Q}} = \mathbb{R}\), and the boundary of \(\mathbb{R}\) is the empty set \(\emptyset\) or if you want to include infinity, it consists only of \(\pm \infty\), but the boundary of \(A\) are the real numbers themselves. This makes \(\partial \bar{\mathbb{Q}} \subset \partial \mathbb{Q} = \mathbb{R}\). On the other hand, the interior of \(\mathbb{Q}\) is empty, since there is no way to get a ball around the rationals without having an irrational squeeze inside it. By vacuity, the boundary of the empty set is the empty set which is a subset of any set. That is, \(\emptyset = \partial \bar{\mathbb{Q}} \subset \partial \mathbb{Q} = \mathbb{R}\).

- **Limit or Cluster Point:** A cluster point is such that any ball (no matter how small) around it, not including itself, will have points from the set. It is a limit point. We say that a sequence \(\{x_i\} \subset X\) converge to \(x \in X\) if for every open set \(O\) containing \(x\) there exists an \(N \in \mathbb{N}\), such that \(\{x_i\}_{i>N} \subset O\). The point \(x\) is called a limit point.

The examples below illustrate this. Think about the sequence \(1/n\). For this sequence 0 is a cluster point. Any ball around 0 contains points of these sequence, still 0 is not in the sequence. Adding the cluster points to a set will close it. For example the interval \([0, 1)\) is not closed, but if you define a sequence \(1 - 1/n\) this sequence has a cluster point which is 1, and so adding 1 will close the set. All irrational numbers are cluster points of rational sequences. That is, why adding (joining) the irrationals to the rationals we get closure to the real line.

- **Perfect sets.** Perfect sets in \(\mathbb{R}\) are closed and have no isolated points. The most common example is the closed interval \([a, b]\). Not every closed set is perfect. For example the integers \(\mathbb{Z}\) is closed but every integer is isolated. Perhaps the most famous example of a perfect set is the


\(^{21}\)http://en.wikipedia.org/wiki/Trigonometric_polynomial
2.3. **TOPOLOGICAL DEFINITIONS**

Cantor set shown below where each point is a limit point yet, it is not an interval.

- **Separable spaces.**
  - We know that each point of the discrete metric space is isolated. So any discrete metric space is separable as long as it has a countable number of elements.
  - In the real line \( \mathbb{R} \) the set of rational numbers \( \mathbb{Q} \) is countable and dense. Then \( \mathbb{R} \) is separable.
  - The space \( \ell^\infty \) is not separable. We show this. Each number in the interval \([0, 1]\) has a binary representation \(0.b_1b_2\cdots b_n\cdots\), which we can associate with a sequence \((b_1, b_2, \cdots, b_n, \cdots)\) in \( \ell^2 \). Let us call \( A = \{(b_1, b_2, \cdots, b_n, \cdots) : b_n = 0 \text{ or } b_n = 1\} \). There are uncountable many elements on \( A \). (as many as numbers in the interval \([0, 1]\). Now \( d_\infty(x, y) = 1 \), for \( x \neq y, x, y \in A \). Choose a ball \( B(x, r) \) for each element \( x \in A \), and \( 0 < r < 1/2 \). Not two balls intersect and they are uncountable many balls. There is no way to find a set \( M \), such that \( \overline{M} = X \), since any element of \( M \) should intersect unless one ball and there are uncountable many balls. So \( M \) is not countable. That is \( \ell^\infty \) is not separable.
  - The space \( \ell^p \), \( 1 \leq p < \infty \) is separable. Let us show this. Define
    \[
    M := \{y = (q_1, q_2, \cdots, q_n, 0, 0, \cdots) : n \in \mathbb{N}, q_n \in \mathbb{Q}\}.
    \]
  
  \( M \) is countable, since \( n < \infty \). We show that \( M \) is dense in \( \ell^p \). Let \( x \in \ell^p \) arbitrary. We know that the tail \( p \)-series of any element \( x = (\xi_i) \in \ell^p \) decreases to 0. That is, for any \( \epsilon > 0 \), there is an element \( N \in \mathbb{N} \), such that
  \[
  \sum_{i=N+1}^{\infty} |\xi|^p < \frac{\epsilon^p}{2}.
  \]
  Since the rationals are dense in \( \mathbb{R} \), for each \( \xi_i \) there is a rational \( q_i \) close to it. So we can build a sequence of rationals \( y = (q_i) \), such that
  \[
  \sum_{j=1}^{n} |\xi_i - q_i| < \frac{\epsilon^p}{2}.
  \]
Then
\[ [d(x, y)]^p = \sum_{i=1}^{N} |\xi_i - q_i|^p + \sum_{i=N+1}^{\infty} ||\xi_i||^p < \epsilon^p \]
and \( d(x, y) < \epsilon \) as desired. So \( M \) is dense in \( \ell^\infty \).

The idea of separable spaces is interesting. It indicates that we can approximate any element of the space as much as we want with a sequence with a finite or countable number of elements.

### 2.4.4.2 The Cantor Set

We make a stop in the Cantor set due to its historical value and its practical value as an example and counter-example in many situations. It illustrates some puzzling phenomena of numbers which we demonstrate below.

#### 2.4.4.2.1 History

According to [Gregory Moore](http://www.sciencedirect.com/science/article/pii/S0315086008000050) “In an unpublished letter of 21 June 1882 the Swedish analyst Gösta Mittag-Leffler asked Cantor if he could prove that a nowhere dense set \( P^\alpha \) of real numbers had an empty derive set for some \( \alpha \).” Cantor defined \( P^{(n)} \) as the derived set of \( P^{(n-1)} \), and he took the limit as \( n \to \infty \) and called it \( P^\infty \). For example let \( A = \mathbb{Q} \cap [0, 1] \), then \( A' = P = [0, 1] \), and after this \( P^{(n)} = [0, 1] \), all the way to \( n = \infty \). This was the first encounter of Cantor with transfinite iterations. We observe that if \( P' \) is the derived set, \( P'' \) is the derived of the derived set, and so \( P^{(n)} \) is the \( n^{th} \) derived set. Interestingly the word “derivative” is closed to the word “derive” and the notation for multiple derivatives of a function is used as well for iterated derived sets.

Cantor responded in the 25th of June by providing a counter-example, that is by proving that the statement is false. He found an example of a set which is nowhere dense, but for which the derived set is not empty. In fact it is the complete whole set.

The counter-example is what we know now as the Canto set. According to [Wikipedia](http://en.wikipedia.org/wiki/Cantor_set) The Cantor set was discovered by Henry John Stephen Smith in 1874.

This example provides lots of insights into topological, numerical, and set theory concepts. For example we visit the following properties: measure,
2.3. **TOPOLOGICAL DEFINITIONS**

cardinality, closeness, closure, interior, perfection, nowhere dense, etc. The Cantor set also has properties of interest for fractal theory but we consider those outside of the scope of these notes except that self-similarity which could be appreciated in the displayed Figure 2.6. Self-similarity is one of the axioms of fractals.

2.4.4.2.2 **Geometrical Definition:** There are many representations of the Cantor set. A geometrical approach indicates that the Cantor set is the points lying on a straight segment constructed in an infinite iterative process by deleting each time the middle third of each subsegment left in the previous iteration. That is, assume that the segment is \([0, 1]\), delete the second half interval \((1/3, 2/3)\) to be left with \([1, 1/3] \cup [2/3, 1]\). Then delete the second half on each of these two to find \([1, 1/9] \cup [2/9, 1/3] \cup [2/3, 7/9] \cup [8/9, 1]\). Find the pattern and take the limit. The pattern is as follows.

- The number of subintervals is \(2^n\), starting at \(n = 1\).
- At the \(n^{th}\) subdivision there are \(2^n/2\) intervals on the left of \(2/3\) and \(2^n/2\) intervals on the right of \(2/3\). The \(2^n/2\) intervals on the left, are a copy of the previous set, scaled by \(1/3\). The \(2^n/2\) intervals on the right, are the same as those in the left, but shifted by \(2/3\).
- The equation is

\[
C_n := \frac{C_{n-1}}{3} \cup \left(\frac{2}{3} + \frac{C_{n-1}}{3}\right),
\]

with \(C_0 = [0, 1]\), and taking the limit as \(n \to \infty\)

\[
C := \lim_{n \to \infty} C_n = \bigcap_{n=1}^{\infty} C_n = \bigcap_{n=0}^{\infty} \frac{C_{n-1}}{3} \cup \left(\frac{2}{3} + \frac{C_{n-1}}{3}\right). \quad (2.4.12)
\]

Figure 2.6 shows a diagram of the construction Cantor set up to 6 iterations. That is up to \(C_5\).

This Figure was taken from the blog [stackexchange](http://tex.stackexchange.com/questions/31999/drawing-cantor-set) which provides help of drawing under \LaTeX{} development.

---

\(^{24}\)http://tex.stackexchange.com/questions/31999/drawing-cantor-set
CHAPTER 2. METRIC SPACES

Figure 2.6: Five steps of the Cantor set construction. The TeX code for this Figure is provided by Alain Matthes.

2.4.4.2.3 Numerical Definition: This definition can be derived from the geometrical approach. A numeric approach indicates that the Cantor set $C \subset [0,1]$ is constructed as all ternary (base 3) representations without the digit “1” on them. For example, $0.202002 \in C$, but $0.0212 \notin C$. More precisely, a number $x \in C$ has the ternary representation $x = 0.t_1t_2 \cdots t_n \cdots$ where each $t_n, n = 1 \cdots \infty$ is either 0 or 2. We provide the construction of the Cantor set next.

The idea behind the construction of the function is to find a representation of the Cantor set $C$ in terms of numbers. That is, we can write any number in the interval $[0,1]$ in the form $0.t_1t_2 \cdots t_n \cdots$, with $t_n$ equal to 0, 1 or 2.

Refer to the Figure 2.6 as we will follow the steps from top to bottom in that figure.

- Start with the interval $[0,1]$ (top solid line). If we want to remove the middle third, we can erase the subinterval $(1/3, 2/3)$ (second line). Now the number $1/3$ in ternary (base 3) representation is 0.1. So the number 1 should not appear as the first digit.

- Now the second deletion (third line) happens on the next digit. Let us see. At this point we have two subintervals, $[0, 1/3] \cup [2/3, 1]$, which do not have a digit 1 as the $t_1$ member of the ternary representation. The first interval has a representation $0.0t_2t_3 \cdots$ and the second interval is represented by numbers $0.2t_2t_3 \cdots$. We want to work in $t_2$. If $t_2 = 1$, then we are in the middle of the first and second intervals. After the deletion we should have 4 subintervals (line 4 of Figure 2.6). We

https://en.wikipedia.org/wiki/Ternary_numeral_system
want the first subinterval to have \( t_2 = 0 \), the second subinterval \( t_2 = 2 \),
the third subinterval \( t_2 = 0 \), and the fourth subinterval \( t_2 = 2 \).

- The pattern should be clear after the previous steps.
  - After the first iteration we set \( t_1 \) to be either 0 or 2 (for left and right intervals).
  - After the second iteration we set \( t_2 \), to be 0, 2, 0, 2 in that order for the fourth subintervals of line 3 of Figure 2.6.
  - After \( n \) iterations we should have \( 2^n \) subintervals. The \( t_n \) digit for each interval is set to 0 for odd intervals and 2 for even intervals. At infinitum we get a number \( x \in C = 0.t_1t_2\cdots t_n \cdots \) where there the digit “1” was removed from the set of numbers \( \{t_i\}_{i=1}^{\infty} \).

We find a collection of interesting properties of this set:

### 2.4.4.2.4 Properties

- **Measure:** We now measure the Cantor set. Starting at measure of 1, corresponding to the original interval \( [0, 1] \), we will count how much remove at each iteration and subtract that result form 1. In the first iteration we remove \( 1/3 \), on the second \( 2/9 \) and in the \( n^{\text{th}} \) iteration \( 2^{n-1}/3^n \), that is, that in the limit we remove

\[
\sum_{n=1}^{\infty} \frac{2^{n-1}}{3^n} = \frac{1}{3} \sum_{n=1}^{\infty} \frac{2^{n-1}}{3^{n-1}} = \frac{1}{3} \sum_{n=0}^{\infty} \frac{2^n}{3^n} = \frac{1}{3} \left( \frac{1}{1-2/3} \right) = 1.
\]

So we are left with a set of measure 0. That is, the Cantor set has measure 0.

- **Cardinality:** It is amazing that the Cantor set is uncountable. That is, it has as many numbers as the real line, still its measure is zero. This is counter-intuitive as many things that happen with the infinity. For example we say how can be that the set of natural numbers \( \mathbb{N} \) being infinite has measure zero? Then we say, well... it is countable, the same for integers and even for rational \( \mathbb{Q} \) when we know that the rationals are dense in the real line, but \( \cdots \) uncountable? That is, the Cantor set is a counter-example of sets with measure zero which are uncountable.
The way to show this is by exhibiting a function $f$ that is surjective in $[0, 1]$. That is, such that for each number in $y \in [0, 1]$ there is a number $x \in C$ such that $f(x) = y$. The trick here is to think that all the interval $[0, 1]$ can have a binary representation $0.b_1b_2 \cdots b_n \cdots$ where each $b_n$, $n = 1 \cdots \infty$ is either 0 or 1. Let us construct $f$ as follows: Pick any $x \in C$, then change all $t_i = 2$, in the ternary representation of $x$ by 1. So now $x$ has now a binary representation $x = 0.b_1b_2 \cdots b_n \cdots$ where $b_n = 0$ or $b_n = 1$. The question now is. Is $f$ surjective? That is, is any number $y$ in the interval $y \in [0, 1]$, an image of $f$ for some number $x \in C$? Pick an arbitrary number $y \in [0, 1]$. Then the number has a binary representation $y = 0.b_1b_2 \cdots b_n \cdots$, where $b_i$ is either 1 or 0. Then replace all the $b_i = 1$ by $b_i = 2$, and call this number $x$. So for each $y \in [0, 1]$ there is an $x \in C$, such that $f(x) = y$. Then the cardinal of $C$ has to be equal or larger than the cardinal of the interval $[0, 1]$. It can not be larger since $C \subset [0, 1]$ so it is equal.

The function can be explicitly constructed by following the ideas of the previous paragraph. That is, any binary number $y$ in the interval $[0, 1]$ has the representation

$$y = 0.b_1b_2 \cdots b_n \cdots = \sum_{n=1}^{\infty} b_n 2^{-n} = \sum_{n=1}^{\infty} \frac{2b_n}{2} 2^{-n}.$$

To find the pre-image $y$ by using the same digits $b_n$, under a ternary representation we now note that $t_n = 2b_n$ is either 0 or 2, so we write

$$y = \sum_{n=1}^{\infty} t_n 2^{-n},$$

with $t_n = 2b_n$, that is:

$$f : C \rightarrow [0, 1]$$

$$x = \sum_{n=1}^{\infty} t_n 3^{-n} \mapsto y = f(x) = \sum_{n=1}^{\infty} \frac{t_n}{2} 2^{-n}$$

with $t_n = 2b_n$, is surjective from $C$ to the interval $[0, 1]$.

If we think, that after every division of the unit interval, we get two new intervals, and we can make as many division as the natural numbers,
then in the limit where the thickness of each subinterval becomes 0, we find the number of points $2^{\aleph_0}$ of no thickness, which is the cardinal of the real numbers, where $\aleph_0$ is the cardinal of the natural numbers $\mathbb{N}$.

- **Closeness:** Please observe equation [2.4.12] which defines precisely the Cantor set. Each $C_n$ is a finite union of closed sets, and from the corollary [2.4] then the union of the $C_n$’s is itself closed. Then we have an infinite intersection of closed sets, which, again by corollary [2.4] is closed. So the closure of the Cantor set is itself. That is $\overline{C} = C$.

- **Empty Interior:** We show that the Cantor set has an empty interior. Let us assume that the interior is not empty and there exists $x$ such that $x \in C$, so there is ball $B(x, r) \subset C$. This is, for all $y$ such that $|x - y| < r$, $y \in C$. Since $x \in C$, it has a ternary representation $x = 0.t_1t_2t_3\ldots$, where $t_n$ is either 0 or 2. Let us pick $n$ on this representation such that $(1/3)^n < r$, and choose $y = x + (1/3)^n$. Then $|x - y| = (1/3)^k < r$. Clearly the representation of $x$ in the $n^{th}$ digit has a 1, by simple summation. Hence $y \notin C$, but we said that $y \in B(x, r) \subset C$ which is a contradiction. Hence the interior of the cantor set is empty. That is $\mathring{C} = \emptyset$.

- **Connectedness:** The Cantor set is totally disconnected. That is, we show that given any two distinct points $x$ and $y$ in the cantor set $C$, there is a point $z \notin C$ that lies between $x$ and $y$. This means that there is no any open interval $(a, b)$ no matter how tiny such that for which $(a, b) \subset C$. Since $x \neq y$, then $|x - y| > 0$, and so we can find a number $N$ such that $|x - y| > 1/3^N$. Why this number? Recall that after $N$ subdivisions we end up with $2^N$ subsets (partitions of $C_N$) each of size $(1/3)^N$. We know that $C_N \subseteq C$, and, so there is no way that $x$ and $y$ are in the same subset of $C_N$. They should be in different subsets of $C_N$. All those subsets are separated (disconnected), that is there must be an open interval $(a, b)$ between them which which is not contained in $C$, otherwise the union of this interval with its immediate neighbours would create a closed interval with length larger than $(1/3)^N$, which is not possible.

- **Perfection:** The Cantor set is perfect. We already showed that it is closed. The second condition for perfection is that every point should be an accumulation point. We show this by using the bisection algorithm.
Take $x \in C$. Given any $\epsilon > 0$ we show that there is an $N$ such that there is a $y \in C$ with $|x - y| < \epsilon$.

Think about $C_1$. Either $x$ is in the first third (left) part of the interval $[0, 1]$ or in the third (right) part of the interval $[0, 1]$. In either case, $x$ is within a distance of $1/3$ of the Cantor set points. The next bisection occurs for $C_2$. Now we have four intervals, and $x$ should be on any of those intervals. Now $x$ is within a distance of $1/3^2$ of another $y$ Cantor point. Keep going and we can find that after some $N$ we have $1/3^N < \epsilon$. That is what we require for $|x - y| < \epsilon$, with $y \in C$. So the Cantor set is perfect.

**Density:** The Cantor set is nowhere dense. The reason for this is because its interior is empty. That is, for any $x \in C$, there is not a ball (interval) $B(x, r) \subset C$. We showed this in the previous item.

**Compactness:** At this moment we have not defined compactness. We show in the section [2.4.14](#) that the Cantor set is compact.

The Cantor set can be generalized to several dimensions and several shapes. For example in 2D, think of a square $[0, 1] \times [0, 1]$, make a partition with parallel lines through the axis at $1/3, 2/3$, creating 9 squares with sides of length $1/3$, and remove the middle square $[1/3, 2/3] \times [1/3, 2/3]$, then keep removing the middle squares of the remaining squares, in this fashion. After $n$ iterations we would need to remove $2^{2n}$ squares, of side lengths $(1/3)^n$ each. Actually we can see the 2D Cantor set as the Cartesian product of 1D Cantor sets, and get the square of the Cantor set set $C^2$. In general we can divide a box $[0, 1]^n$, in $\mathbb{R}^n$, in the same fashion and the result would be the Cantor set to the power $n$, $C^n$. An interesting example of a Cantor choosing the geometry of a triangle instead of a box is known as the Sierpiński triangle shown below.

![Sierpiński Triangle](image)

where we start with an equilateral triangle, divided in three equilateral triangles with vertices at the middle points of the original triangle and remove
2.3. TOPOLOGICAL DEFINITIONS

the middle triangle. Then we do the same to all remaining triangles and keep looping up to infinity. The Figure above shows up to the third iteration.

2.4.5 Continuity

We start with a definition of continuity which coincides the definition in analysis (or Calculus) courses. Then we provide a Theorem which is the definition of continuity in Topology.

Definition 2.4.6. Let \((X_1, d_1)\) and \((X_2, d_2)\) be metric spaces. A function \(T : X \to Y\) is said to be continuous at a point \(x_0 \in X\) if for every \(\epsilon > 0\) there is a \(\delta > 0\) such that

\[
d_2[T(x), T(x_0)] < \epsilon, \quad \forall x \quad \text{provided that} \quad d_1(x, x_0) < \delta.
\]

\(T\) is continuous if it is continuous at every point of \(X\).

Next theorem shows that a function is continuous if and only if returns open sets from the range into open sets into the domain. This is the definition of continuity in topology.

Theorem 2.4.9 (Continuity). A function \(T\) from a metric space \((X_1, d_1)\) into a metric space \((X_2, d_2)\) is continuous if and only if the inverse image of any open subset of \(Y\) is an open subset in \(X\).

Proof.

- \(\implies\): Let us assume an open set \(V \in f(X_1)\). Choose a point \(y \in V\), that is \(y = f(x_0)\), for some \(x_0 \in U \subset X_1\). Since \(V\) is open there is a ball \(B(y, \epsilon) \subset V\). From the definition of continuity this means that there is \(\delta > 0\) such that \(B(x_0, \delta) \subset X_1\). If for each ball in \(V\) we return a ball in \(X_1\), then the arbitrary union of these balls is an open set in \(X_1\).

- \(\iff\): Let us now assume that for every open set \(V\) in \(f(X_1)\), the inverse image is an open set in \(X_1\). So pick \(\epsilon > 0\), and \(y \in V\). The \(V \cap B(y, \epsilon) \subset X_2\) is an open set in \(X_2\). From the hypothesis this means that there exits \(\delta > 0\), and some \(x\), with \(f(x) = y\), such the set \(f(x) \in B(y, \epsilon)\) provided that \(B(x, \delta) \subset X_1\). That is, for any \(\epsilon > 0\), there exists \(\delta > 0\), such that if \(B(x, \delta) \subset X_1\), then \(B(x, \epsilon) \subset V\). This is precisely the definition of continuity.

\(\blacksquare\)
2.4.7 Limit, Convergence, Cauchy Sequence, and Completeness

The concept of limit arrives from the need to fit “data” with better measurements. In the field of mathematics the irrational numbers were seen as limits of rational sequences. Fractions with infinite number of digit representations in the decimal expansion could motivate the definition of limit clearly.

For example, let us assume that we want to evaluate numerically $1/3$. We use elementary division and find that

$$\frac{1}{3} = 0.33333 \cdots$$

In the real world (of computers) we need to stop the division somewhere. If we want to improve precision we should admit an error or tolerance level that we can live with. The error could be noted as $\epsilon$ for the lowercase Greek letter $e$. This error can be measured with the absolute value, and we want to say that

$$|\frac{1}{3} - 0.333| < \epsilon.$$  

The problem here is that we are trying to find the error between something that we know and something that we do not know. We do not know what exactly is $1/3$ (in the world of decimals). But we can define a sequence that approaches this number. This sequence is the sequence $(x_n)$ where

$$x_n = \sum_{i=1}^{n} 3 \times 10^{-i}$$

This sequence produces the numbers $0.3, 0.33, 0.333 \cdots$. One thing we know well is that two consecutive numbers of this sequence get very small as $n$ goes large, that is

$$|x_{n-1} - x_n| = 3 \times 10^{-n}$$

Even more, given any two numbers $n > m$ which most of the time we want to be large, we have that

$$|x_n - x_m| = 3 \sum_{i=m}^{n} 10^{-i} = 3 \times 10^{-m} \sum_{i=1}^{n-m} 10^{-i}.$$  

\[26\] In engineering, data are understood as measurements. In mathematics we and in this context we understand data as some information that we already know.
but we know that \(\sum_{i=0}^{n-m} 10^{-i} < 2\). So we can say that
\[|x_n - x_m| = 6 \times 10^{-m}\]
This means that we have the error \(\epsilon\) under control by just finding a big number \(N\) and let \(m > N\) and \(n > N\). This motivates the definition of a Cauchy sequence

**Definition 2.4.8.** Given a metric space \((X, d)\), and a sequence \((x_n)\) with \(x_n \in X\), we say that the sequence is a **Cauchy sequence** if given \(\epsilon > 0\), there is an \(N \in \mathbb{N}\) such that for any \(m > N\), and \(n > N\), \(d(x_n, x_m) < \epsilon\).

This is just a definition and it says nothing about convergence of existing of a limit point. It is enough to check that the condition \(d(x_{n+1} - x_n) < \epsilon\) is satisfied, since we can write (assuming \(n < m\))
\[d(x_n, x_m) \leq d(x_n, x_{n+1}) + d(x_{n+1}, x_{n+2}) + \cdots + d(x_{m-1}, x_m)\]
by repeating the triangular inequality \(m-n\) times. We can redefine the error as \(\epsilon/(m-n)\), and so obtain \(d(x_n, x_m) < \epsilon\) as desired. In general is much easier to check the inequality of a member of the sequence with its immediate neighbour.

We wish to have a point \(x\) such that the sequence \((x_n)\) converges. That is, we want to say

**Definition 2.4.9.** Given a metric space \((X, d)\), we say that \(x\) is the **limit of a sequence** \((x_n)\), with \(x_n \in X\), if given any \(\epsilon > 0\), there is a number \(N \in \mathbb{N}\), such that for any \(n > N\), \(d(x_n, x) < \epsilon\). We write \(\lim_{n \to \infty} x_n = x\).

Now, we are saying that there is a point \(x\) which is a limit point and that we would like to find. Let us provide an example. Choose \(X\) to be the interval \(X = (0, 1)\) in the real numbers, and \(x_n = 1/n\). Then we know that \(\lim_{n \to \infty} x_n = 0\) (choose \(N > 1/\epsilon\)). However \(0 \notin X\). So if our space is \(X\) we do not know the number \(0\). We say that the function does not converges in \(X\). However if we add \(0\) to \(X\) then we have now \(Y = [0, 1) = X \cup \{0\}\), and the function converges in \(Y\). Adding points so that the limit exists is called completion.

The history of limits is quite interesting. The book [Practical Analysis in One Variable](http://www.stat.colostate.edu/estep/education/M618/notesonreals.pdf) and in particular chapter 9 [Sequences and Limits](http://www.stat.colostate.edu/estep/education/M618/notesonreals.pdf) from Donald Estep offers a good history on the matters discussed here.
We know that the differential and integral Calculus are based on top of the concept of limits. Newton understood the concept but could never get it in a formal presentation. More than one hundred years passed until Cauchy came with the definition above.

The decimal representation of fractional numbers is attributed to Simon Stevin. We well know now that 1/2 = 0.5 and fractions have a finite number of decimal digits, or an infinite periodic number of decimal digits. We could think about the rational numbers in these two different forms and say that in a way the rational numbers, such as 1/3 in our example, are a completion of those rational numbers with finite number of digits. That is, the periodic representations are also part of the rational numbers. However the history is more interesting than this. J. J. O’Connor and E.F. Robertson offer a good introduction to the history of the real numbers. We extract a few interesting items related to the completion of the real line from the rational numbers.

• Hankel (student of Weirstrass) in 1867 addressed the question whether there were other “number system” which had essentially the same rules as the real numbers.

• Two years after Hankel’s announcement M´eray published an article where he considered that Cauchy sequences of rational numbers converge to what he called “fictitious limit” (which we know today as irrational numbers).

• Heine, three years after M´eray, published a similar result. He defined sequences of rational numbers to be equivalent if they converge to the same (limit) real number. This created a partition of the real numbers in equivalence classes. Each equivalence class is a real number and they complete the real line

• Cantor, in 1872 published the same idea of Heine.

\[\text{Note that Donal Estep gives credit for the } \epsilon - \delta \text{ definition of limit to Weirstrass, as well as for the introduction of the notation “lim”.}\]

\[\text{28}\text{https://en.wikipedia.org/?title=Simon\_Stevin}\]

\[\text{29}\text{http://www-history.mcs.st-and.ac.uk/HistTopics/Real\_numbers\_2.html}\]

\[\text{30}\text{If it the limit is rational, we can see it as a constant } (x_n = x), \text{ with } x \in \mathbb{Q}.\]
2.3. TOPOLOGICAL DEFINITIONS

2.4.9.1 Example:

The use of Cauchy sequences is important because a Cauchy sequence contains a collection of elements that we should know, regardless their limit (which might be unknown to us). So when we compare elements, we are comparing objects that are known to us.

We provide three ways to compute $\sqrt{2}$ using rational sequences approximations. The first form illustrates directly how the Cauchy sequence appears in a simple algorithm. The bisection algorithm is easy to understand, easy to implement, but more costly than the Newton-Raphson method explained in the second example. The second example provides a recursive formula which is convenient from the computational point of view but it is not convenient from the theoretical (proof building) point of view. Still we proof that the sequence generated is Cauchy as a good exercise of a more complicated problems. The third example provides an analytic expansion which provides an easy path to proof that the sequence is a Cauchy sequence.

- The bisection algorithm: The idea for the bisection algorithm is easily explained with this example. Think of the function $f(x) := x^2 - 2$. We want to solve the equation $f(x) = 0$ for some $x \in [a, b]$ where $[a, b]$ is some interval such that $f(a) < 0$, and $f(b) > 0$. Since the function is continuous there has to be a point in the interval such that $f(x) = 0$. We can then think about the interval $[1, 2]$. Since $f(1) < 0$ and $f(2) > 0$. The algorithm is as follows:

  - Pick $x_0 = a$, $x_1 = b$, $x_2 = (x_0 + x_1)/2$ (the midpoint of the interval).
  - Choose $x_3$ as follows:

$$x_3 = \begin{cases} \frac{x_0+x_2}{2} & \text{if } f(x_2) > 0 \\ \frac{x_1+x_2}{2} & \text{if } f(x_2) < 0 \end{cases}$$

  - In general, for $n \geq 3$ choose

$$x_n = \begin{cases} \frac{x_{n-2}+x_{n-1}}{2} & \text{if } f(x_{n-1}) > 0 \\ \frac{x_{n-3}+x_{n-1}}{2} & \text{if } f(x_{n-1}) < 0 \end{cases}$$

A sequence of points created with this algorithm is

$$(1, 2, 1.5, 1.25, 1.375, 1.406, 1.4375 \cdots).$$
It takes up to 10 terms of the sequence to get closer than $1.41406265$ to $\sqrt{2}$, after $x_2 = 1.5$. It takes only 3 iterations of the Newton’s method to get there (this is shown in Figure 2.7).

In the first step, after the first two points where created the interval is reduce to $1/2$ of the original interval. Then after every step it is reduced each time by $1/2$. So after $n$ points are selected we can say that $|x_{n+1} - x_n| \leq 2^{-n}$. Clearly this is saying that the series is Cauchy, So if the error is $\epsilon = 2^{-n}$, we can find choose any $N > |\log_2(\epsilon)|$, and this will guarantee that our error $\epsilon$ is a good bound for the level of accuracy required.

- **The Newton’s method:** This method, also called the [Newton-Raphson method](https://en.wikipedia.org/wiki/Newton%27s_method), is designed to compute roots of a function based on the iteration

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}, \quad (2.4.13)$$

provided the derivative exists and is non-zero in the search interval. We know $\sqrt{2}$ is irrational. We can use Newton’s method to solve the equation $f(x) := x^2 - 2 = 0$. That is,

$$x_{n+1} = x_n - \frac{x_n^2 - 2}{2x_n} = \frac{x_n}{2} + \frac{1}{x_n} \quad (2.4.14)$$

An easy way to interpret this formula is to think that $f(x) = 0$, (the root we want to find) and, and $x_{n+1} = x$ (we stop close enough to the root) write the equation as

$$x_{n+1} = x_n + \frac{f(x) - f(x_n)}{f'(x_n)}$$

from which

$$f'(x_n) = \frac{f(x) - f(x_n)}{(x - x_n)}$$

which is the approximation for the derivative at point $x_n$. Another way (and the most commonly interpretation) is to write this as

$$f(x) = f(x_n) + f'(x_n)(x - x_n),$$

which geometrically is the segment tangent at $f$ at $x = x_n$. This is illustrated in Figure 2.7.
which with the initial guess of \( x_0 = 1 \), produces the sequence

\[(1, 3/2, 17/12, \cdots),\]

or in decimals \((1, 1.5, 1.4166\ldots, \cdots)\).

Actually we can rewrite equation 2.4.14 in two different ways and interpret it.

\[
x_{n+1} = \frac{x_n^2 + 2}{2x_n} = \frac{1}{2} \left( x_n + \frac{2}{x_n} \right).
\]

The first interpretation, for the expression in the middle, is that for \( x_n \approx \sqrt{2} \), we have in the numerator twice the square of the denominator. So to recover the root, we should divide by 2 and then by \( x_n \) to eliminate the power of two in the numerator. This is precisely what is written in the equation. The second interpretation is that each term is the average (the \( 1/2 \) in the last expression) of the previous iteration and the inverse of the first iteration weighted by 2, which is the squared value. So if the first term \( x_n \) overestimates the value, the second balance this value by having it in the denominator, then the average is taken. In general change 2 by \( a > 0 \) and this works the same for any positive root. This method is called the Babylonian or Heron’s method. \(^{34}\)

We would like to check that the sequence is Cauchy. Note that since the initial value is \( x_0 = 1 \), then from

\[
x_{n+1} = \frac{1}{x_n} + \frac{x_n}{2},
\]

all the values of the sequence are bounded below by 1. That is \( x_n \geq 1 \), \( \forall n \in \mathbb{N} \). We search for an upper bound. For this we show that the sequence \((x_n)\) is decreasing, after the first evaluation and so we can start at the first value \( x_1 = 3/2 \).

\[
x_{n+1}^2 = \frac{1}{4} \left( x_n + \frac{2}{x_n} \right)^2 = \frac{1}{4} \left( x_n - \frac{2}{x_n} \right)^2 + 2 \geq 2, \quad (2.4.15)
\]

\(^{34}\)https://en.wikipedia.org/wiki/Methods_of_computing_square_roots
that is, for \( n = 0, 1, 2, \cdots \) we have \( 2 - x_{n+1}^2 < 0 \), and then

\[
x_{n+1} - x_n = \frac{2 - x_n^2}{2x_n} \leq 0 \quad n \geq 1.
\] (2.4.16)

Then, after passing the first element \( x_0 = 1 \), \((x_n)\) is decreasing and we can locate all \((x_n)\) in the interval \([1, 3/2]\).

If after some \( n \), \( x_n = 2 \), then any new term on the iteration will be 2 and that is the limit. Let us then assume that we do not get in any finite step \( x_n = 2 \), so the sequence is strictly decreasing and we have an infinite number of values in the finite interval \([1, 3/2]\) which by the monotone convergence Theorem C.2.1 has a limit point.

We further show this without recurring to the monotone convergence theorem. From equation 2.4.15

\[
x_{n+1}^2 - 2 = \frac{(x_n^2 - 2)^2}{4x_n^2} \quad \text{hence} \quad 0 \leq x_{n+1}^2 - 2 \leq \frac{1}{4}(x_n^2 - 2)^2.
\]

In the last equation, we replace

\[
x_n^2 - 2 \leq \frac{1}{4}(x_{n-1}^2 - 2)^2
\]

and so

\[
(x_n^2 - 2)^2 \leq \left(\frac{1}{4}\right)^2 (x_{n-1}^2 - 2)^4
\]

and keep doing recursively until

\[
0 \leq x_{n+1}^2 - 2 \leq \left(\frac{1}{4}\right)^{n+2} (x_0^2 - 2)^{2(n+2)} = \left(\frac{1}{4}\right)^{n+2}
\]

since \( x_0 = 1 \). Now from equation 2.4.16

\[
|x_{n+1} - x_n| = \frac{2 - x_n^2}{2x_n} \leq |2 - x_n^2| \leq \left(\frac{1}{4}\right)^{n+2}
\]

From which \((x_n)\) is a Cauchy sequence. Figure 2.7 illustrates the Newton’s method for this example. Figure 2.8 shows the points selected my the Newton’s method.
Figure 2.7: The Newton-Raphson method on \( y = x^2 - 2 \). The method converges so fast that after two iterations we could not distinguish more roots at the resolution of the plot. Here is the description of the path (red) of search: Start at \( x_0 = 1 \), find \( f(x_0) = -1 \), trace a tangent to \( f \) until hits the x-axis at \( x_1 = 1.5 \). Find \( f(x_1) = 0.25 \), trace a tangent from \( x_1 \) to the curve \( f \) until hitting the x axis at \( x_2 = 1.4167 \), then find \( f(x_2) = 0.00703889 \). This is so small that after this points in the plot would look as one. It is clear that the Newton-Raphson method converges faster than the bisection method. The bisection method choose the middle of the current interval each time regardless the shape of the function \( f \). The Newton-Raphson method, uses tangents along the curve for \( f \) to get faster to the solution. It is a non linear sequence of linear events. In the bottom of the figure the points \( y_0, y_1, \ldots, y_5 \) show the x-coordinates of the bisection algorithm.
Figure 2.8: Computation of $\sqrt{2}$. After the 4\textsuperscript{th} iteration the function converges within 5 decimal places. The green point shows the initial guess.

- **The Taylor Series approach:** A way to get a rational sequence that converges to $\sqrt{2}$, and such that we can check the Cauchy sequence criteria easily is through a Taylor series expansion. That is, let us assume $f(x) = \sqrt{1+x}$, and let us expand this series at $x = 1$. Not that $f(1) = \sqrt{2}$. That is, from the binomial expansion

$$(1 + x)^{1/2} = 1 + \frac{x}{2} - \frac{x^2}{2.4} + \frac{1.3x^3}{2.4.6} \cdots (-1)^{k+1} \frac{(2k-3)!!}{(2k)!!} + \cdots$$

where the double factorial $n!!$ is defined as

$$n!! = n(n-2)(n-4) \cdots 1$$

where the “1” at the end is either 1! for $n$ odd or 0!, for $n$ even. In the particular case of $x = 1$, we have

$$\sqrt{2} = 1 + \frac{1}{2} + \sum_{k=2}^{\infty} (-1)^{k+1} \frac{(2k-3)!!}{(2k)!!}.$$
We show that the sequence
\[ x_n = \sum_{k=0}^{n} (-1)^{k+1} \frac{(2k-3)!!}{(2k)!!}. \]
is Cauchy. Here
\[ |x_{n+1} - x_n| = \left| \frac{[2(n+1) - 3]!!}{2(n+1)!!} \right| = \frac{(2n-1)!!}{(2n+2)!!} \leq \frac{(2n-1)!!}{(2n+1)!!}. \]
That is,
\[ |x_{n+1} - x_n| \leq \frac{1}{2n+1} \to 0 \quad \text{when} \quad n \to \infty. \]

We do not compare how fast the Taylor series (binomial expansion) approximation is compared to Newton-Raphson. We leave this to the reader. Appendix A shows how to build a family of monotonically increasing sequences which converges to \( \sqrt{2} \). The ideas exposed above, are part of the history of the completion of the real line and should motivate the following definition.

**Definition 2.4.10.** A metric space \((X, d)\) is **complete** if any Cauchy sequence \((x_n)\), with \(x_n \in X\), converges in \(X\). That is if \(\lim_{n \to \infty} x_n = x\), with \(x \in X\).

The concept of complete spaces is very important in analysis and topology. In fact the following definition introduces Banach Spaces\(^{35}\).

**Definition 2.4.11.** A Banach Space is a complete normed vector space.

In fact the real numbers \(\mathbb{R}\) form a Banach space. Why is this? We have mentioned a few names in the construction of the real line but we are missing still some important names. Richard Dedekind worked also in the problem of filling up the real line. He introduced the concept of Dedekind cut\(^{36}\). The way Dedekind though about filling up the real line was by separating the rational numbers in two sets. All rationals smaller or equal than a given number \(x\), and all rationals larger than a given number \(x\). This \(x\) was the

---

\(^{35}\)https://en.wikipedia.org/?title=Banach_space

\(^{36}\)https://en.wikipedia.org/wiki/Dedekind_cut
“cut”. While for a rational $x$ number there is no question that including $x$ all the real line consists of those numbers under $x$ and those above $x$. For an irrational $x$ there were issues. It is easy to proof that $\sqrt{2}$ is not rational, so if we pick this number as the cut, we can find a rational under it, but between that rational and $\sqrt{2}$ there is always another. So we can never find the largest rational smaller or equal than $\sqrt{2}$. Appendix A builds an infinite family of examples that proof the statement of non existence of a maximum rational smaller that $\sqrt{2}$. The Website in the Dedekind cut above suggests a particular example of the general proof in Appendix A. So we need an axiom that assumes that there is a least (least) upper bound for all the rationals smaller or equal than $\sqrt{2}$. That lowest upper bound (LUB) is exactly the irrational $\sqrt{2}$ which is the Dedekind cut. This motivates the introduction of the Axiom of Completness of the real numbers, which states that every real set with an upper bound must have a least upper bound (LUB). This is a way to fill the gaps in the real line created by the existence of irrational numbers. For Dedekind, the existence of the Dedekind cuts is itself the Axiom of Completeness. In the sense of Cauchy we still should be able to show that every Cauchy sequence in the real numbers has a limit point. The completion of the real line and complex plane is very important since the completion of many spaces resting on those (such as for example $\mathbb{R}^n$, $\mathbb{C}^n$, $\ell_p$, etc.) depend on these. The actual proof that $\mathbb{R}$ (and $\mathbb{C}$) is complete is presented in Appendix C Theorem C.2.5.

Before making new definitions let us illustrate the concept of completeness with examples.

2.4.11.1 Examples

It is interesting that for completeness it is easier to find counterexamples than examples. An examples requires a proof. This is because completeness requires that ALL Cauchy sequences convergence, and this is a bit of a burden. It is only necessary to find one that does not converge to prove that the space is not complete. Hence, most of the examples below are counterexamples.

- Let us assume the real numbers $\mathbb{R}$.
  - The reals themselves form a complete set. See theorem C.2.5.

\[37\text{https://en.wikipedia.org/wiki/Completeness_of_the_real_numbers}\]
2.3. TOPOLOGICAL DEFINITIONS

- Any closed and bounded set in the reals is complete. We proof this after we cover the concept of compactness.

- \((0, 1]\) is not complete in \(\mathbb{R}\). The sequence \((1/n)\) is Cauchy but it converges to 0 which is not even in \((0, 1]\).

- The set \(\mathbb{Q}\) is not complete in \(\mathbb{R}\). Three different examples (counterexamples) of this are shown in 2.4.9.1 where we show Cauchy sequences of rational numbers which converge to \(\sqrt{2}\), which is no rational.

- Completeness plays an important role in spaces with infinite many dimensions.

- We already showed an example of this. Figure 2.1 shows a sequence of continuous functions \(f_n(x) = x^n\), in the interval \([0, 1]\). Let us consider the space of continuous functions in the interval \([0, 1]\) which are absolutely integrable. We show that this space is not complete. The sequence \(f_n\) is Cauchy since for \(n > m > N\) (for some \(N\))

\[
\|f_n(x) - f_m(x)\|_1 = \int_0^1 (x^m - x^n)dx = \frac{x^{m+1}}{m} - \frac{x^{n+1}}{n} < \frac{1}{m}
\]

and so pick \(N > 1/\epsilon\), and then \(\|f_n - f_m\|_1 \leq \epsilon\) for all \(m, n > N > 1/\epsilon\). Then the sequence \((f_n)\) is Cauchy in the space \(L_1[0, 1]\), still

- Another example in the same space, is the sequence of functions:

\[
f_n(x) = \begin{cases} 
0 & x \leq \frac{1}{2} \\
(n(x - \frac{1}{2})) & \frac{1}{2} < x \leq \frac{1}{2} + \frac{1}{n} \\
1 & x < \frac{1}{2} + \frac{1}{n}
\end{cases}
\]  

(2.4.17)

Let us verify that \(f_n\) is continuous. The function is defined in three segments. The first segment is 0, the third segment is 1. Those two segments, parallel to the x-axis are continuous. The segment in the middle is a linear segment with slope \(n\). Let us now see that the segments are welded together. The only possible discontinuities are at \(x = 1/2\) where, from the left we have 0, and
Figure 2.9: We have two functions \( f_n \) (black) and \( f_m \) (blue) defined by equation \ref{eq:2.4.17} for \( m > n \).

From the right we have 0. The other point \( 1/2 + 1/n \). Here from the left we have \( n(1/2 + 1/n - 1/n) = 1 \), and from the right we have 1. So each \( f_n \) is continuous. Figure \ref{fig:2.9} illustrates two of these functions.

Without doing any computation, it should be clear that the yellow area (which represents \( \|f_n - f_m\|_1 \)), goes to zero after \( m \) and \( n \) grow large, and that in the limit the function goes to a step function with the step at \( 1/2 \), which is discontinuous. We leave the details about filling the proof to the reader.

Note that we defined continuity without recurring to the concept of limit. This is because while in Calculus the concept of limit is at the center of all the developments, in topology it is the concept of open set (or an open ball) which is in the center of all the developments. We defined a limit point \( x \) of a set \( A \), as one such that for any ball \( B(x, r) \), there are points \( y \neq x \), such that \( y \in B(x, r) \cap A \). We can borrow from Calculus the definition on limit as follows

**Definition 2.4.12.** Given a function \( f : X \to Y \), where \((X, d_X)\), and \((Y, d_Y)\) are metric spaces, we say that

\[
\lim_{x \to x_0} f(x) = y,
\]

if given \( \epsilon > 0 \), there exists a \( \delta > 0 \), such that if \( d_X(x, x_0) < \epsilon \), then
d_Y(f(x),y) < \delta. Here y is the limit of the function f(x) as x approaches x_0.

Again, the symbol \(\epsilon\) corresponds to the Greek symbol for “e” which means error. The symbol \(\delta\) can be seen as the Greek symbol for “d”, and could be associated with “distance” in the domain space.

We now define boundedness

**Definition 2.4.13.** A set \(A \subset X\) is said **bounded** if the set of all distances \[D = \{d(x,y) : x,y \in X\}\]
is bounded in the real numbers. The \(\text{sup}(D) = \delta(A)\) is known as the **diameter** of the set. A sequence that is bounded as a set is defined as a **bounded sequence**.

We show the boundedness, limit Theorem. That is,

**Theorem 2.4.10.** Let \((X,d)\) be a metric space then:

- A convergent sequence in \(X\) is bounded, and its limit is unique.
- If \(x_n \to x\) and \(y_n \to y\), then \(d(x_n,y_n) \to d(x,y)\).

**Proof.** Let \((X,d)\) be a metric space.

- Let us assume that a sequence \((x_n)\) is converging to a point \(x\). That is, there is an \(N \in \mathbb{N}\) such that for each \(n > N, d(x_n, x) < \epsilon\). Hence \(\forall n > N\), we have that \(x_n \in B(x, \epsilon)\). We want to squeeze \(x_1, x_2, \ldots, x_{N-1}\) in that ball, or make it larger, but finite, if necessary. That is, pick

  \[r = \max\{\max\{|x - x_i|\}, \epsilon\}\]

then, by construction, the ball \(x_n \in B(x, r)\) for \(n = 1, 2, \ldots, \infty\), so the sequence \((x_n)\) is bounded.

- Let us now show that if the limit exists, is unique. By contradiction let us assume that there are two limits \(\alpha\) and \(\beta\). Such that \(\alpha \neq \beta\). Then from the definition of limit there are numbers \(N_1\) and \(N_2\) such that

  \[d(x_n, \alpha) < \frac{\epsilon}{2} \quad n > N_1\]
  \[d(x_n, \alpha) < \frac{\epsilon}{2} \quad n > N_2\]
So for any \( n > \max\{N_1, N_2\} \) we should have that (using the triangular inequality)
\[
d(\alpha, \beta) \leq d(x_n, \alpha) + d(x_n, \beta) < \epsilon
\]
and since \( \epsilon \) is as small as we want, \( d(\alpha, \beta) = 0 \) and so \( \alpha = \beta \). That is the limit is unique.

• On the second item. If \( x_n \to x \) there is a number \( N_1 \in \mathbb{N} \) such that for \( n > N_1 \), \( d(x, x_n) < \epsilon/2 \), likewise there is a number \( N_2 \in \mathbb{N} \) such that for any \( n > N_2 \), \( d(y, y_n) < \epsilon/2 \). Then from the triangular inequality and for any \( n > \max\{N_1, N_2\} \)
\[
d(x, y) < d(x, x_n) + d(x_n, y) \\
d(x_n, y) < d(x_n, y_n) + d(y_n, y)
\]
so
\[
d(x, y) \leq d(x, x_n) + d(x_n, y_n) + d(y_n, y).
\]
Then
\[
d(x, y) - d(x_n, y_n) \leq d(x, x_n) + d(y_n, y).
\]
If in the previous inequality we swap \( x_n \) with \( x \) and \( y_n \) with \( y \) and since the metric is symmetric we find:
\[
d(x_n, y_n) - d(x, y) \leq d(x, x_n) + d(y_n, y).
\]
Hence
\[
|d(x_n, y_n) - d(x, y)| \leq d(x, x_n) + d(y_n, y) < \epsilon
\]
Then we say that \( d(x_n, y_n) \to d(x, y) \).

2.4.14 Compactness

We present the definition of compactness and important theorems relating closed sets, bounded sets, and convergence in these sets. The main theorems here are the Heine-Borel Theorem and the Bolzano-Weirstrass theorem.
2.4.14.1 Historical Introduction

The term “compact” is one of the most abstract concepts of topology. To motivate the introduction of this new concept we provide some history which is a short summary of the article “A pedagogical history of compactness” by Manya Raman-Sundström.

The development of compact spaces started in the last quarter of the nineteen century. There where two main currents of development on the concept of compact spaces.

- A current based on analysis and more specifically in sequences of real numbers and functions. The main originators here are Bolzano and Weirstrass.

- A current oriented toward the topological concepts of open covering of sets led by Heine, Borel, Lebesgue, and later by the Russian mathematicians Alexandroff and Urysohn.

The concept was born initially from the need to find convergence of sequences of real numbers. The need to find minimum and maximum points in continuous functions. According to Raman-Sundström, Fréchet proposed the theorem about a continuous function reaching a maximum and a minimum in a closed interval, but he attributed this result to Weirstrass. The first appearance of compactness is due to Fréchet who defined the term “compact”, today known as sequential compactness in his 1906 thesis. The concept of limit point compactness was developed by Bolzano and Weirstrass. It basically indicates that a set $X$ of real numbers (now this transcended to general topological spaces) is said to be limit point compact if every infinite subset of $X$ has a limit point in $X$.

While we do not know exactly the reasons for Fréchet to introduce the word “compact” there is evidence that some mathematicians of his time did not like the choice. Schöflies suggested the name “lückenlos” (without gaps) or “abschliessbar” (closable). The precise intuition behind the term compactness was not yet clear. Even more, at the end of his life, Fréchet could not remember why he chose that term. Here is a quote from Raman-Sundström paper, which is a translation to English from an original Fréchet's publication: “Doubtless I wanted to avoid a solid dense core with a single thread

---

39 https://en.wikipedia.org/wiki/Limit_point_compact
going off to infinity being called compact. This is a hypothesis because I have completely forgotten the reasons of my choice!"

On the topological side, Borel proved the following lemma in his 1894 thesis:

**Lemma 2.4.11.** If on a line one has an infinite number of subintervals, such that every point of the line is interior to at least one of the intervals then one can determine effectively a bounded number of intervals from among the given intervals that have the same property (every point of the line is interior to at least one of them).

Borel method was similar to the approach by Heine in 1872 to prove that a continuous function on a closed interval was uniformly continuous. A theorem already proved by Dirichlet in 1852 using more explicit coverings and subcoverings that Heine. Dirichlet does not get credit for this because he waited until 1904 to publish his proof.

Perhaps the most important result in this field is the Heine-Borel theorem that we announce and prove later on a broader context.

**Theorem 2.4.12 (Heine-Borel).** A subset of $\mathbb{R}$ is compact if and only if it is closed and bounded.

The theorem was initially known as the Borel’s theorem, but according to Raman-Sundström, a student of Weirstress (Schönfiles) noticed the connection between Borel’s theorem and Heine’s theorem after which it is known as the Heine-Borel theorem. Still Heine did not prove the theorem, while Cousin generalized Borel lemma in 1985 to arbitrary covers and got no credit. In fact the credit for Cousin’s theorem, was given to Lebesgue, who said he was aware of the result and published his proof in 1904. It is clear then that there is some debate about who did what about the covering of closed and bounded sets in $\mathbb{R}$, which today is known as the Borel-Lebesgue property and which was used by Fréchet to introduce the first definition of compact.

The concept of compactness was extended from the set of reals to the set of functions. That is, functional analysis was an extension of real analysis to the space of functions and compactness was an important property to extend. We already showed examples in the space of functions where the completeness fails (see examples 2.4.11.1). The problem there is that continuity was seen as a point-to-point continuity instead of uniform continuity. A sequence of functions $(f_n)$ converges uniformly or is said to be equicontinuous
if for any given $\epsilon > 0$, there exists $\delta > 0$, such that $|x - y| < \delta$ implies $|f_n(x) - f_n(y)| < \epsilon$. This is continuity in the large, not point-to-point. The functions all converge simultaneously in a global way and not locally at each point. The extension of Bolzano-Weierstrass theorem to the space of functions is known as the **Arzelà-Ascoli theorem**, which states in modern language:

**Theorem 2.4.15 (Arzelà-Ascoli).** Any bounded equicontinuous sequence of functions in $C^0[a, b]$ has a uniformly convergent subsequence.

The extension of the Heine-Borel theorem to the space $C^0[a, b]$ says that a subset of $C^0[a, b]$ is compact if and only if it is closed, bounded and equicontinuous. Ascoli proved the sufficient condition in 1884, while Arzelà the necessity in 1889.

The development of the notion of compactness as well as the setting of the basis for modern analysis and topology is centered in the heads of many mathematicians. Here are a few listed by countries:

- **France**: Hadamard (Fréchet advisor), Lebesgue, and Fréchet.
- **Russia**: Alexandroff and Urysohn.
- **Germany**: Hausdorff, Hilbert, Schönflies, and Cantor.
- **Netherlands**: Brouwer.
- **U.S.**: Chittenden, Hendrick, and Moore.

While Fréchet introduced the term “compact” his definitions are what we know today as **sequential compactness** and **countable compactness**. The Russian school in head of Alexandroff and Urysohn appear to be the first to state the most general form of compactness in terms of open covers. Fréchet was in close contact with Alexandroff and Urysohn and was not unaware of the possibility of using neighborhoods to characterize compactness.

In 1914, Hausdorff published a paper where he introduced the name “metric spaces” referring to Fréchet $E$ spaces, and gave the name “compact” a set $E$ from which every subset has a limit point in $E$. This is what we call today limit-point compactness and is equivalent to countable compactness prevailed during the rapid development of point-set topology in the 1920s.

We present first the definition of compact as known today, which is of a topological character, then we will introduce other definitions related which
are equivalent in metric spaces but not in general topological spaces. These definitions are directly related to the work of Bolzano-Weirstrass on analysis and convergence. They are **sequentially compact**, **limit point compact**, and **countable compact**.

Let us start with the definition of cover.

**Definition 2.4.16 (cover).** Assume \((X, d)\) is a metric space. Let \(\{O_\alpha\}_{\alpha \in A}\) be a collection of open subsets of \(X\) such that \(\bigcup_{\alpha \in A} O_\alpha = X\). Then we say that the family \(O_\alpha\) is an open cover of \(X\). A collection \(\{O_\beta\}_{\beta \in B}\), where \(B \subset A\) such that \(O_{\beta \in B} = X\) is said to be an open subcover of \(X\).

**Definition 2.4.17 (compact).** A metric space \(X\) is said to be compact if every open cover has a finite subcover.

Recall from Theorem 2.4.1 that if \(A\) is a subset of the \((X, d)\) space, then this induces a new class of open sets space which is the intersection of all open sets of \(X\) with \(A\). Then we can think of a covering and subcovering in the same terms defined above, this time applied to \(A\). That is, a family \(\{O_\alpha \cap A\}\) is an open cover of \(A\) if \(A = \bigcup_{\alpha \in A}(O_\alpha \cap A)\), and a family \(\{O_\beta \cap A\}\) is an open cover of \(A\) if \(A = \bigcup_{\beta \in B}(O_\beta \cap A)\), with \(B \subset A\). We say that the set \(A\) is compact if for every cover there is a finite subcover.

We want to draw a picture to explain the concept, however we warn that this figure is impossible to draw with a clear meaning. If we draw a finite cover of a set, then any subcover is finite and regardless of the set (compact or not) the figure would imply compactness. We need to draw an arbitrary infinite set of sets covering a set, and this is simply impossible. As we use suspensive dots to mean that sometimes go to infinity like for example \(\mathbb{N} = \{1, 2, \ldots\}\) we can imagine those dots in Figure 2.10 which represents an oversimplification of the concept topologically compact. Note that some of the balls fall out of the boundary. We should not be concerned with the left overs outside of our compact set, since it is the intersection of balls with the set what generates the open sets in the induced topology as indicated above.

We provide next a few examples to get familiarity this concept.
2.3. **TOPOLOGICAL DEFINITIONS**

Figure 2.10: Imagine that the set on the left (gray) is covered with arbitrary many open neighborhoods (center). If from those arbitrary many open neighborhoods we can select a finite (right) collection which still covers the set, we say that the set is compact.

### 2.4.17.1 Examples

- Let us consider the interval \( A = (0, 1) \) in the metric space of the real numbers \( \mathbb{R} \). Then let us define the family of open sets

\[
O_i = \left( \frac{1}{i+1}, 1 - \frac{1}{i+2} \right)
\]

Here the first few \((1/2, 2/3), (1/3, 3/4), (1/4, 4/5) \cdots\). It should be clear that

\[
(0, 1) = \bigcup_{i=1}^{\infty} O_i.
\]

So the family \( \{O_i\} \) is a cover of the set \( A = (0, 1) \). However we can not find a finite subcover which covers \( A \), since this will imply to stop the union at some, possible large, \( i = N \). Then the whole subinterval \((0, 1/(N+1))\) would be excluded on the lower end. Similarly, a piece of interval will be chopped off the upper end.

- The Heine-Borel theorem, which we prove below says that the compact sets in \( \mathbb{R} \) are those, and only those, which are closed and bounded. Hence for example \([a, b]\) is compact. Let us prove, as an exercise, that \( A = [a, b] \) is compact.
The difficulty on showing compactness rest on the fact that the cover should be arbitrary and not just any particular cover.

Let \( \mathcal{O} \) be an arbitrary open cover for \( A \). We define

\[
P = \{ x \in [a, b] : [a, x] \text{ is covered by a finite number of } O \in \mathcal{O} \}.
\]

The singleton \( O = \{ a \} \in P \) since \( a \) should belong to at least one set of \( \mathcal{O} \). \( P \) is bounded above by \( b \), since no open set in \( \mathcal{O} \) will cover above \( b \). From the Axiom of Completeness of the real numbers \( P \) has a supremum \( s \). Appendix \( \mathcal{C} \) shows the definition of supremum and infimum (sup and inf) as well as the formulation of the axiom of completeness.

We first show that \( [a, s] \) can be covered by finitely many sets in \( \mathcal{O} \). This is trivial when \( s = a \), so assume \( s > a \). Let \( O_s \in \mathcal{O} \) be a set containing \( s \). Then there is an \( \epsilon \in (a, s) \) such that \((s - \epsilon, s] \subseteq O_s \). By assumption, there is a finite subcover of \([a, s - \epsilon/2] \). By adding \( O_s \) to that finite subcover, we get a finite subcover of \([a, s] \), shown below.

We now show that \( s = b \). Suppose \( s < b \) and let \( O_s \in \mathcal{O} \) be a set containing \( s \). Then there is an \( \epsilon > 0 \) such that \([s, s + \epsilon] \subseteq O_s \). So taking a finite subcover of \([a, s] \) and adding the set \( O_s \) gives us a finite subcover of \([a, s + \epsilon/2] \), contradicting the fact that \( s \) is the supremum of \( P \). Hence \([a, b] \) is compact.

This compactness of the interval could be generalized to any number of finite dimensions. That is to \( \mathbb{R}^n \). A generalization of an interval to \( n \) dimensions is called a cell or closed box. That is, given two points \( a, b \in \mathbb{R}^n \), such that \( a_i < b_i \), for \( i = 1, 2, \ldots, n \), the \( n \)-cell is the set of all points \( x \), such that \( a_i \leq x \leq b_i \). The proof of this is shown in theorem \( 2.4.19 \). This is a special case of the compactness of the product space, which is known as Tychonoff’s theorem.\footnote{https://en.wikipedia.org/?title=Tychonoff%27s_theorem}

We show a few theorems out of the definition of a compact set.

**Theorem 2.4.13.** Any closed subset \( A \) of a compact space \( X \) is compact.
2.3. TOPOLOGICAL DEFINITIONS

Proof. Let $A \subset X$ be a closed subset of a compact space $X$. Let $\{O_\alpha\}$ an arbitrary cover of $A$. Together with this cover we claim that $\bigcup (U_\alpha \cap A^c) = X$ is a cover of $X$. Since $X$ is compact, we find from here a finite subcover such that $A = \bigcup_{i=1}^n \{O_{\alpha_i}\}$. Note that $A^c$ would not be in this cover for obvious reasons. However, we needed $A$ to be closed, otherwise $A^c$ would not be open.

In what follows, we will need the Hausdorff property which we define next.

Definition 2.4.18. If for a given space $X$, we choose two arbitrary distinct points $x \neq y, x, y \in X$ and prove that there are two disjoint open sets $U$ and $V$ such that $x \in U$, and $y \in V$, we say that the space $X$ has the Hausdorff property.

It is obvious that a metric space $X$ satisfies the Hausdorff property, since for any two $x \neq y$, we can pick $B(x, d(x, y)/2) \cap B(y, d(x, y)/2) = \emptyset$.

Theorem 2.4.14. Any compact subset $A$ of a metric space $X$ is closed.

Proof. Let $A$ be a compact subset of a metric space $X$. We show that $A$ is closed by proving that $A^c$ is open. Let $x \in A^c$, since $X$ is Hausdorff, for any $y \neq x$, there are two disjoint open subsets $U$ and $V$ of $X$ such that $x \in U$, and $y \in V$. Let us now build the union of all open sets with $y \in A$. Since $A$ is compact, from that cover we find a finite subcover $\{U_i\}$ of $A$. For each open $V_i \subset A$, there is an open $U_i$, such that $U_i \cap V_i = \emptyset$. Let us construct now two sets $V = \bigcup V_i = A$, and $U = \bigcap U_i$. $U$ is open since it is a finite intersection of open sets. We chose $x \in A^c$, and $x \in U_i$, for each $i$, so $x \in U$. This completes the proof.

We now show that continuous functions map compact sets into compact sets, but before we verify the following statement:

Lemma 2.4.15. Let $f : X \to Y$ a function. If $Y = \bigcup_{\alpha} O_\alpha$, then

$$X = f^{-1}(Y) = \bigcup_{\alpha} f^{-1}(O_\alpha).$$

Proof. Choose $x \in X$. Then $f(x) = y \in Y$, and since $Y = \bigcup_{\alpha} O_\alpha, x \in O_\alpha$ for some $\alpha$. That is $x = f^{-1}(y) \in f^{-1}(O_\alpha)$, from we say that $X \subset \bigcup_{\alpha} f^{-1}(O_\alpha)$. By the definition of function $U_\alpha f^{-1}(O_\alpha) \subset X$. So the equality is proven.
CHAPTER 2. METRIC SPACES

Theorem 2.4.16. Let \((X, d_X)\) a compact space and \(f\) a continuous function from \(X\) onto another (or same) space \((Y, d_Y)\). Then \(Y\) is compact.

Proof. Let \(\{V_\alpha\}\) be an open cover of \(Y\). Since \(f\) is continuous each \(U_\alpha = f^{-1}(V_\alpha)\) is open in \(X\) and from lemma 2.4.15 above \(X = \bigcup_\alpha U_\alpha\). Now since \(X\) is compact, from this union we can choose a finite collection \(\{U_i\}\) that covers \(X\). Now since \(f(\bigcup U_i) = \bigcup f(U_i)\), we have that \(Y = f(X) = f(\bigcup U_i) = \bigcup V_i\), where the indices \(i\) run over a finite collection. Then \(Y\) is compact.

We now define an interesting concept that is closed related to compactness.

Definition 2.4.19 (Finite Intersection Property (FIP)). We say that \(X\) satisfies the finite intersection property for closed sets if any collection \(\{O_\alpha\}\) of closed sets in \(X\) with finite intersection \(\cap_{i=1}^n O_\alpha \neq \emptyset\), then the complete intersection is also non-empty. That is, \(\cap_\alpha O_\alpha \neq \emptyset\).

We show next that the FIP is another way to characterize compactness. It is in a way a “complementary” formulation.

Theorem 2.4.17. Let \(X\) be a metric space\(^{42}\). Then \(X\) is compact if and only if \(X\) satisfies the FIP.

Proof. Choose a fixed \(J\) in the index set for a cover of \(X\).

\[
\begin{align*}
X \text{ is compact } & \iff \text{ for any cover } \{A_\alpha\} \text{ of } X \text{ there is a subcover } \{A_{\alpha_i}\} \text{ of } X \\
& \iff X = \bigcup_\alpha A_\alpha \text{ arbitrary } \implies X = \bigcup_{\alpha_i} A_{\alpha_i} \text{ finite} \\
& \iff X = A_J \cup (\cup_{\alpha \neq J} A_\alpha) \implies X = A_J \cup (\cup_{\alpha_i} A_{\alpha_i} \neq J) \\
& \iff A_J^c \subset (\cup_{\alpha \neq J} A_\alpha)^c \implies A_J^c \subset (\cup_{\alpha_i} A_{\alpha_i} \neq J)^c \\
& \iff A_J^c \subset \cap_{\alpha \neq J} A_\alpha^c \implies A_J^c \subset \cap_{\alpha_i} A_{\alpha_i}^c \\
& \iff \cap_{\alpha_i} A_{\alpha_i} \neq \emptyset \implies \cap_{\alpha} A_\alpha \neq \emptyset \\
& \overset{\text{def}}{=} X \text{ satisfies FIP}. \\
\end{align*}
\]

\(^{41}\)from the concept of complement of a set
\(^{42}\)this is true for general topological spaces and the proof here would work the same there.
In the last step we use the counter-reciprocal property \( p \implies q \iff \neg q \implies \neg p \). Also we used the de Morgan’s laws and other basic set properties. The arbitrariness is not loss when saying that an arbitrary family of sets is an arbitrary family of complements of these sets. The equivalence works the same moving both ways of the “\( \iff \)” sign.

In this proof we see how working with complements and intersections is equivalent to working with sets and their unions, and how the words finite and arbitrary change roles. This duality is an important fact in general topology which we will discuss in detail in the next chapter just after the definition of topology 3.2.1.

As a direct consequence of this theorem we state the following:

**Corollary.** If \( \{ A_i \} \) is a sequence of non-empty compact sets such that \( A_{n+1} \subset A_n \), then \( \bigcap A_i \neq \emptyset \).

It is interesting that this exactly how Fréchet defined E-classes in his Ph.D. thesis. Here is Raman-Sundrstöm translation of Fréchet’s definition for compact:

**Definition 2.4.20 (Fréchet E-class compact).** A set \( E \) is called compact if, whenever \( E_n \) is a sequence of none-empty, closed subsets of \( E \) such that \( E_{n+1} \) is a subset of \( E_n \) for each \( n \), there is at least one element that belongs to all of the \( E_n \)’s.

We provide a few examples.

### 2.4.20.1 Examples

The following examples assume that we are working in the set of real numbers \( \mathbb{R} \).

- **Nested tails** Let us define

  \[
  E_n = \{ x : x \geq n \}
  \]

  For example \( E_0 = [0, \infty) \), \( E_{10} = [10, \infty) \).

  \[
  \begin{array}{ccccccccccc}
  0 & 1 & 2 & 3 & 4 & 5 & \cdots & n & \cdots & \infty
  \end{array}
  \]

  Clearly \( E_{n+1} \subset E_n \), still \( \bigcap_{n=1}^{\infty} E_n = \emptyset \). What fails here is that each \( E_n \) is non-bounded, and then it is non compact.
• **Nested closed intervals**: Let us define \( E_n = [-1/n, 1/n] \). Both hypothesis are honored. \( E_{n+1} \subset E_n \), and each \( E_n \) is compact. We see that \( \bigcap_{n=1}^{\infty} E_n = \{0\} \).

\[
\begin{array}{ccccccccc}
& -1 & -\frac{1}{2} & -\frac{1}{3} & -\frac{1}{4} & 0 & \frac{1}{4} & \frac{1}{3} & \frac{1}{2} & 1 \\
\hline
\end{array}
\]

Let us see that if we remove the number 0, from the sets \( E_n \), that is we define \( E_n = [-1/n, 0) \cup (0, 1/n] \), they still satisfy the nested hypothesis, but they fail to be compact, and so the intersection is empty as we could see (since we removed the element 0, which is the only possible intersection of the original sets \( E_n \)).

\[
\begin{array}{ccccccccc}
& -1 & -\frac{1}{2} & -\frac{1}{3} & -\frac{1}{4} & 0 & \frac{1}{4} & \frac{1}{3} & \frac{1}{2} & -1 \\
\hline
\end{array}
\]

This example could be misleading and we explain why in the next two examples.

• Think of the nested closed intervals \( E_n = [-1, 1 + 1/n] \). It is clear that \( E_n \) is closed and bounded, and \( E_{n+1} \subset E_n \). So the hypothesis are satisfied. We find that

\[
\bigcap_{n=1}^{\infty} E_n = [-1, 1].
\]

\[
\begin{array}{ccccccccc}
& -1 & 0 & \frac{15}{8} & \frac{3}{2} & \frac{11}{8} & \frac{3}{2} & 1 \\
\hline
\end{array}
\]

We see here that 1 (the red tick) is an accumulation (limit) point from the sequence on its right.

• Let us remove the number 0 from the previous sequences of intervals \( E_n \). Then we now define

\( E_n = [-1, 0) \cup (0, 1 + 1/n] \).

Each \( E_n \) is bounded, \( E_{n+1} \subset E_n \), and

\[
\bigcap_{n=1}^{\infty} E_n = [-1, 0) \cup (0, 1].
\]
So the intersection of all $E_n$ sets is not empty (the highlighted yellow and green segments). This shows that the counter-reciprocal is not necessarily true. That is, if the sets $E_n$ are not compact, no necessarily the finite total intersection is empty. That is, the total intersection and compactness are not equivalent.

- In general if we have sets of the form $E_n = [a_n, b_n]$ such that $a_n < b_n$, and $a_n$ is an increasing sequence, while $b_n$ is a decreasing sequence, then we should have that $\bigcap_{n=1}^{\infty} E_n = [a, b]$ for some $a, b$.

While this is guaranteed by the use of the previous corollary, we can see this from basic principles discussed in Appendix C. That is, since the sequence $a_n$ is bounded above by $b_1$ it has a sup. The monotone convergence theorem [C.2.1] indicates that there is a limit point for the sequence of $\{a_n\}$. Now that limit point is $a = \sup \{a_n\}$, and since any $a_n < b_n$, we have any $b_n$ will be an upper bound for the set of $\{a_n\}$, and it could not be smaller than the LUB, so $a \leq b_n$, for all $n$. Now the sequence $(b_n)$ has a lower bound $(a)$, and so it has a limit $b \geq a$ from the monotone convergence theorem [C.2.1]. The intersection then is the interval $[a, b]$. If $a = b$ it is just a point.

We now consider the case of $n$-cells.

**Theorem 2.4.18.** Let $\{B_n\}$ a sequence of $n$-cells (closed boxes) such that $B_{n+1} \subset B_n$. Then $\cap_n B_n \neq \emptyset$.

*Proof.* The proof of this theorem is done by considering every dimension as a one dimensional problem, and using the last example iteratively. Along each dimension $i$ the intersection is non-empty, and a subinterval $[a_i, b_i]$ exists. The point in the intersection is constructed as the n-cell $[a, b]$.

We now used Theorem 2.4.18 to prove that every cell (closed) in $\mathbb{R}^n$ is compact.

**Theorem 2.4.19.** Every closed box is compact.
Figure 2.11: Imagine that the blue dots are a just a few of infinite number of open neighborhoods that cover the box. Two dimensional bisection is done until the yellow box where we show that the infinite union of open hypothesis breaks.

**Proof.** Let us assume that the theorem is no true and start with a closed box $B$, which is covered by a family of open sets $\{O_{a}\}$ from which no finite subcover can be found.

Figure 2.11 is an sketch of the proof’s path. We will do multi-dimensional bisection and assume the $\ell_2$ (Euclidean) norm. Let us assume that the box is denoted by the interval $[a, b]$, with $a = (a_1, a_2, \ldots, a_n)$, $b = (b_1, b_2, \ldots, b_m)$. We call $a_1 = a$, $b_1 = b$. The diagonal size is $\delta = \sqrt{\sum_{i=1}^{n}(b_i - a_i)^2}$. We divide this box into new boxes by cutting it with hyperplanes that parallel to their sides that go through the coordinates $(a_i + b_i)/2$. In the case of a 2D box, Figure 2.11 shows 4 new boxes. In the n-dimensional box we should get $2^n$ new boxes. At least one of the new $2^n$ boxes should have an infinite number of subsets of the cover of the initial box, otherwise we get a finite cover which contradicts the hypothesis and we are done. Let us indicate this
2.3. **TOPOLOGICAL DEFINITIONS**

box as \([a_2, b_2]\). The diagonal of the box has the size \(\delta_2 = \delta/2\). Let us choose then a new box with infinite number of open sets covering it. We apply the same principle to this new box (n-dimensional bisection) and divide into \(2^n\) new boxes by using hyperplanes parallel to the original box side, through the new midpoint coordinates \((a_{i2} + b_{i2})/2\). The diagonal of this box has side \(\delta/2^n\). In general we can go as far as we want. After \(n\) of these steps we have \(2^n\) new the boxes \(\{[a_i, b_i]\}_{i=1,2,...,n}\), with diagonal sizes \(\delta/2^n\).

Since, the cover is open, and there is no empty intersection, no matter how large is \(n\) we still have at least one point \(x\) in the box \([a_n, b_n]\). Now, for the open set in the infinite cover that holds \(x\), call it \(O_\alpha\), draw a ball of diameter \(r > \delta/2^n\), call this ball \(B(x, r)\). There has to be an \(n\) large enough for this ball to wrap the tiny box \([a_n, b_n]\], as illustrated in the Figure 2.11. Now we have

\[
[a_n, b_n] \subset B(a_n, r) \subset O_\alpha,
\]

Then only one \(O_\alpha\) set from the original cover contains the whole box \([a_n, b_n]\), this contradicts the hypothesis that each box in a chain contains an infinite number of sets covering it.

We now show that an infinite set in a compact space needs to have an accumulation point.

**Theorem 2.4.20.** Any infinite subset \(A\) of a compact set \(K\) has a point of accumulation in \(K\).

**Proof.** If any \(x \in K\) is not an accumulation point, then there exists a neighborhood of \(x\), \(O_x\) such that the only point of \(A\) in the neighborhood is \(x\). No finite subcollection which has \(O_x\) as a set cannot cover \(A\), so \(K\) is no compact. This contradicts the hypothesis.

We are now ready to proof the most important theorem of this section. In the following theorem, the equivalence between (i) and (ii) is known as the **Heine-Borel** theorem, and the equivalence between (i) and (iii) is known as the **Bolzano-Weirstrass** theorem.

**Theorem 2.4.21.** Let \(A\) be a subset of \(\mathbb{R}^n\). Then the following three properties are equivalent:

(i) \(A\) is closed and bounded.
(ii) A is compact.

(iii) Every infinite subset of A has a point of accumulation in A.

**Proof.** We refer to the item (iii) above as **sequentially compact**.

(a) (i) $\Rightarrow$ (ii): If A is closed and bounded it can be enclosed by a box which is compact by theorem 2.4.19. Now, from theorem 2.4.14.

(b) (ii) $\Rightarrow$ (iii): This is proved in theorem 2.4.20.

(c) (iii) $\Rightarrow$ (i): This piece has two parts. The easy one is to prove that A is bounded and we prove it first, then we prove that A is closed.

- **A is bounded:** Let us assume that A is not bounded. Let choose an infinite subset of A, $\{x_k\}$, such that $\|x\|_k > k$. Then $\{x_k\}$ does not have no accumulation points in $\mathbb{R}^n$, and since $A \subset R^n$, it does not have accumulation points in A. This contradicts the hypothesis.

- **A is closed:** Let us assume the contrary. From using theorem 2.4.2 we see that there should be a point $x \in \mathbb{R}^n$, which is an accumulation point for A, but $x \notin A$. That is, there is a sequence $(x_k)$, with $x_k \in A$, such that $x$ is a limit point of $(x_k)$. If we show that $(x_k)$ has not an accumulation point (other than $x$, of course) in A then we break the hypothesis and the proof will be done.

Let us then choose $y \in \mathbb{R}^n$ such that $y \neq x$. Then

$$\|y - x_k\| = \|y - x_0 + x_0 - x_k\| \geq \|y - x_0\| - \|x_0 - x_k\|$$

since $\|x_0 - x_k\| \to 0$, as $k \to \infty$, we can choose $\|x_0 - x_k\| < \frac{1}{2}\|y - x_0\|$, and then

$$\|y - x_k\| \geq \frac{1}{2}\|y - x_0\|.$$ 

So, no $y$ other than $x_0$ could be an accumulation point for the sequence $(x_k)$. This breaks the hypothesis and we assert the validity of the theorem.

\[\square\]
2.4.20.2 Examples

The previous theorem can provide a method to build many examples of compact sets.

- Any closed interval \([a, b]\) in \(\mathbb{R}\), or closed box \([a, b]\) in \(\mathbb{R}\) is compact. Any finite union of closed intervals \(\bigcup_i [a_i, b_i]\) in \(\mathbb{R}\), or closed boxes \(\bigcup_i [a_i, b_i]\) in \(\mathbb{R}^n\) is compact.

- Any unit closed ball \(\bar{B}(x, 1)\) in a finite dimensional space is compact.

- A closed unit ball in \(\ell_\infty\) is no sequentially compact. We show this. The unit ball \(\bar{B}(0, 1)\) is given by the sequences with all components \(|x_i| \leq 1\). In particular select from this ball the points such that each component \((x_n)_i\) is the Kronecker delta \((x_n)_i = \delta_{n1} = 1\) for \(n = 1\), and \(0\) for \(n \neq 1\). The distance between any two different elements of this sequences is \(d(x_n, x_m) = 1\), using the \(\|\cdot\|_\infty\) that is the maximum of the difference which is 1. Note that a closed ball is closed and bounded. Since the dimension of the space is infinity we do not have the item (iii) in theorem 2.4.21 (sequentially compact) satisfied.

- The Cantor set \(C\) is compact. The cantor set belongs to \(\mathbb{R}\), it is bounded, and it is closed, then by theorem 2.4.21 it is compact.

We might come back to metric spaces and define Fréchet derivatives but ... this is good for now.
Chapter 3

Basic Topological Concepts

3.1 Introduction

This chapter helps by providing the basic concepts in which the idea of continuity is rooted. Concepts such a topology, an open set, a closed set, a compact set, a connected set, a convex set and more are explained here.

3.2 Basic Definitions

The following definition took more than one hundred years to be formalized, and the work of a good group of mathematicians such as Cantor, Weirstrass, Bolzano, Frèchet, Dedekind, Borel, Peano, Jordan, Lebesgue, Hausdorff, Riez, Bairy, Kuratowski, Aleksandrov, Uryshon, Newman, the French Nicolas Bourbaky school (particularly Andrè Weil), and Sierpiński among others. The definition as we know it today was formulated by the American mathematician John L. Kelly’s in 1955. The history of the origin of this definition including the origin of many of the definitions in this section is found in the article by Gregory Moore’s “The emergence of open sets, closed sets, and limit points in analysis and topology” published in Elsevier’s Historia Mathematica Journal in 2008.

Definition 3.2.1. A topological space is a pair \((X, \mathcal{T})\) where \(X\) is a set and \(\mathcal{T}\) is a family of subsets of \(X\) (called the topology of \(X\)) whose elements are called open sets such that:

\[\text{http://www.sciencedirect.com/science/article/pii/S0315086008000050}\]
(i) \( \emptyset, X \in \mathcal{T} \).

(ii) Arbitrary union of open is open. If an arbitrary family of open sets 
\( \{O\}_{\alpha \in \mathcal{A}} \subset \mathcal{T} \), where \( \mathcal{A} \) is some index set, then its union is also an open set. In other words, \( \bigcup_{\alpha \in \mathcal{A}} O \in \mathcal{T} \).

(iii) Finite intersection of open is open. If \( \{O_i\}_{i=1}^k \subset \mathcal{T} \), then \( \bigcap_{i=1}^k O_i \in \mathcal{T} \).

Before showing examples of topologies let illustrate why for union we want to consider an arbitrary family of sets and for intersection only a finite family. Or why topology is defined in terms of open sets and not in terms of closed sets.

Consider \( n \in \mathbb{N} \), and define the sets (intervals) in \( \mathbb{R} \).

\[
A_n = (1/n, 1 - 1/n) = \{ x \in \mathbb{R} : 1/n < x < 1 - 1/n \},
\]

Then

\[
\bigcup_{n \in \mathbb{N}} A_n = (0, 1) , \quad \bigcap_{n \in \mathbb{N}} A_n = \emptyset
\]

Now, for a second example:

\[
B_n = (-1/n, 1/n) = \{ x \in \mathbb{R} : -1/n < x < 1 - 1/n \}.
\]

Then we have that

\[
\bigcup_{n \in \mathbb{N}} B_n = (-1, 1) , \quad \bigcap_{n \in \mathbb{N}} B_n = \{0\}
\]

In the first example we show that infinite union of open sets is open, while infinite intersection is the empty set (that is open and closed by vacuity but not an interesting set). In the second example we showed that infinite union of open sets is open but the infinite intersection is closed (just the singleton \( \{0\} \)). Preservation of a property at \( \infty \) is desired\(^2\) In this case the property is open sets. However, we could choose another set of examples that motivates another definition of topology based only on closed sets. This is.

\(^2\)As is desired, for example, the preservation of continuity of infinite sequences, sums of functions, and its derivatives and integrals by enforcing uniform continuity.
Consider $n \in \mathbb{N}$, and define the sets (intervals) in $\mathbb{R}$.

\[ A_n = \left[\frac{1}{n}, 1 - \frac{1}{n}\right] = \{x \in \mathbb{R} : 1/n \leq x \leq 1 - 1/n\}, \]

Then

\[ \bigcup_{n \in \mathbb{N}} A_n = (0, 1), \quad \bigcap_{n \in \mathbb{N}} A_n = \emptyset \]

Now, for a second example:

\[ B_n = \left[-\frac{1}{n}, \frac{1}{n}\right] = \{x \in \mathbb{R} : -1/n \leq x \leq 1 - 1/n\}. \quad (3.2.1) \]

Then we have that

\[ \bigcup_{n \in \mathbb{N}} B_n = (-1, 1), \quad \bigcap_{n \in \mathbb{N}} B_n = \{0\} \]

In the first example we show that infinite union of closed sets is open, while infinite intersection is the empty set (that is open and closed by vacuity but not an interesting set). In the second example we showed that infinite union of closed sets is open and the infinite intersection of closed sets is closed (just the singleton $\{0\}$). In this case we are preserving the “closeness” of the sets. So we could define topology instead in terms of closed sets by imposing finite union and arbitrary intersection. It is a choice between working with some sets or their complements. There is a complete duality between open and closed, and finite intersection and infinite union.

It is interesting that the question made by many of us “Why is a topology make up of ‘open’ sets?” is a blog in the mathoverflow website \(^3\) which started in the year 2010, and has grown to many pages of discussions.

Here are a few comments from that blog that we find amusing:

- “Topoloy is the art of reasoning about imprecise measurements” By Dan Piponi. Dan compares open sets with rulers of different sizes.
- “It’s ironic that for this question about open sets is now closed” By Deane Yang.

\[^3\)http://mathoverflow.net/questions/19152/why-is-a-topology-made-up-of-open-sets
Toby Bartles quotes Einstein with his famous “Everything should be made as simple as possible, but not simpler”. This is an interesting quote. Think about the open interval \((1/n, 1)\). On the left end, we can always get closer to 0, but never get there. It is like, a world record or the 100 meter race. It can approach 0 minutes, but will never make it (because speed is finite).

A quote from Poincaré “Mathematics is the art of giving the same name to different things”. This is quite profound. This means “abstraction”. Topology gives the name of “continuity” to many things according to which spaces we are working with. We believe that “abstraction” is the intersection of all intelligent points of view. It should be minimal but contain the essence of the concept.

Deane Yang also quotes “To me, the concept of open set is a distillation and abstraction of the properties of open intervals (on the real line) that are critical to defining the working with a continuous function. In my opinion students should never be introduced to the abstract concepts of topology (notably, the concepts of open sets and compactness) unless they have already mastered analysis of functions on the real line and finite dimensional vector spaces and understand thoroughly the role of open sets and compactness in those settings”. We agree with Deane’s ideas. Topology represents the abstraction of many years of thinking of great mathematicians. In order to appreciate is value history should be honored and this history is embedded on the development of analysis and differential geometry among other fields. There is a principle of causality here. People developed analysis and they made further generalizations to build abstract axioms of topology. In a way education should provide at least a little bit of background to introduce the definition of topology. Start a course in topology with the definition that we present here without trying to explain the work and history behind it make the process anti-causal and dry. The student need to understand the past to appreciate the present and avoid the same mistakes from the past. Also, learning the effort of so many good mathematicians over so many years, gives us a sense of relief that this is no black magic that was written in a teaching board, and make of this a human product, still made but very smart people or geniuses if you prefer that word.
3.2. BASIC DEFINITIONS

• However we also agree with Dan McLaury when he says “It’s very hard to appreciate compactness as a concept in its own right in a context where it just means ‘closed and bounded’”.

• Steve Hurder provides a reference by Gregory Moore. Here is a link for that reference: [The emergence of open sets, closed sets, and limit points in analysis and topology][4] Again, history is very important to appreciate the present, and avoid to repeat errors from the past.

To simplify notation will omit the symbol $\mathcal{T}$ and refer to the topological space as the set $X$, unless the $\mathcal{T}$ needs to be explicitly used.

We now introduced many definitions which are similar to those in the context of metric spaces, but now in the more general context of topological spaces.

Definition 3.2.2. • Neighborhood : If $x \in X$, then any open set $A$ such that $x \in A$ is said to be a neighborhood of $x$.

• closed : We say that $A$ is closed, if its complement is open.

• closure : The closure of a set $A$ is noted as $\bar{A}$ and is defined as the smallest closet set which contains $A$, that is

$$\bar{A} = \bigcap\{ C \mid A \subset C \text{ and } A \text{ closed} \}.$$  

Hence $\bar{A}$ is the smallest closed set containing $A$.

• dense : A set $A \subset X$ is called dense if $\bar{A} = X$.

• nowhere dense : A set $A \subset X$ is called nowhere dense if $X \setminus \bar{A}$ is dense.

• Deleted neighborhood : A deleted neighborhood if $x$, is neighborhood of $x$ without the point $x$.

• Accumulation points : A point $x \in X$ is called an accumulation point (or limit point ) of $A$, if every deleted neighborhood of $x$ contains infinitely many points of $A$ . Actually we could make a weaker statement, and say that if it contains at least one point of $A$, then we can proof that there could be infinite by making the neighborhood as small as we want around the point.

CHAPTER 3. BASIC TOPOLOGICAL CONCEPTS

• **Interior point**: A point \( x \in A \) is called an interior point of \( A \) if there exists a neighborhood of \( x \) contained in \( A \).

• **Interior of a set**: The set of all interior points is called the interior of \( A \) and denoted by \( \text{Int}A \) or sometimes by \( \mathring{A} \).

• **Boundary point**: A point \( x \) is called a boundary point if it does not belong to the interior of \( A \) neither to the interior of its complement. That is if \( x \notin \mathring{A} \) and \( x \notin \text{Int}(X \setminus A) \). We use the notation \( \partial A \) for the boundary of a set.

• **Exterior of set**: The exterior of a set is the interior of its complement. That is \( \text{Int}(X \setminus A) \). We note the exterior as \( \text{Ext}(A) \).

We present a few theorems that brings better understanding to the concepts of interior, exterior, boundary, closed, and open sets.

**Theorem 3.2.1.** The interior, the boundary and, the exterior fill the whole space \( X \).

*Proof.* This is seen directly from the definition. \( \square \)

**Theorem 3.2.2.** Let us assume \( x \in \partial A \). Then any neighborhood of \( A \) has points from \( A \) and from \( \text{Ext}(A) \).

*Proof.* We show first that the intersection \( A \cap O_x \), for any neighborhood \( O_x \) is non-empty. Then that the intersection of \( \text{Ext}(A) \cap O_x \neq \emptyset \) for any neighborhood \( O_x \) of \( x \).

(i) Let us pick \( x \in \partial A \). If there is an open neighborhood \( O_x \), such that \( O_x \cap A = \emptyset \), then \( x \in \text{Int}(X \setminus A) \), that is \( x \in \text{Ext}(A) \). In fact \( A \subset X \setminus O_x \), which is a closed set, and since since \( \bar{A} \subset X \setminus O_x \), (being by definition the smallest closed containing \( A \)) then \( \bar{A} \subset X \setminus O_x \), so \( x \notin \mathring{A} \). This implies that \( x \in \text{Ext}(A) \) and so \( x \notin \partial A \). The contradiction implies that \( A \cap O_x \neq \emptyset \).

(ii) If \( \text{Ext}(A) \cap O_x = \emptyset \), then \( y \notin \text{Ext}(A) \), for every \( y \in O_x \). This implies that \( O_x \subset A \), which makes \( x \) an interior point of \( A \). But this contradicts that \( x \in \partial A \). We conclude that \( \text{Ext}(A) \cap O_x \neq \emptyset \), for all neighborhoods \( O_x \) of \( x \). \( \square \)
3.2. BASIC DEFINITIONS

We now proof that the closure of a set is the disjoint union of its interior with its boundary. That is,

**Theorem 3.2.3.** \( \bar{A} = \hat{A} \cup \partial A \).

**Proof.**

(i) “\( \subset \):” Let us assume \( x \in \hat{A} \). If \( x \in \hat{A} \) there is nothing to show. So let us assume that \( x \notin \hat{A} \), then by theorem [3.2.1] \( x \in \partial A \) or \( x \in \text{Ext}(A) \). Let us assume that \( x \in \text{Ext}(A) \). Then since the exterior is open there is a neighborhood \( O_x \) of \( x \) such that \( O_x \subset \text{Ext}(A) \), Now \( \text{Ext}(A) = \text{Int}(X \setminus A) \subset X \setminus A \), and \( O_x \cap A = \emptyset \), and \( A \subset O_x^c \), and since \( O_x^c \) is closed and \( \hat{A} \) is the smallest closed having \( A \), we have that \( \hat{A} \subset O_x^c = X \setminus O_x \). So \( x \notin \hat{A} \) which contradicts the hypothesis. Then \( x \in \partial A \).

(ii) “\( \supset \):” It is clear that \( \hat{A} \subset A \subset \bar{A} \). We show that if \( x \in \partial A \), then \( x \in \bar{A} \).

Let us assume \( x \in \partial A \). If \( x \not\in \bar{A} \), then \( x \in \hat{A}^c \subset A^c \). So since \( \text{Int}(\hat{A}^c \subset \text{Int}(A^c) = \text{Ext}(A) \), which is open, then we find a neighborhood of \( x \), \( O_x \subset \text{Ext}(A) \). But, since by hypothesis \( x \in \partial A \), this violates theorem [3.2.2]. So, \( x \in \bar{A} \), and we are done.

We have then the following corollary.

**Corollary.** \( \partial A = \bar{A} \setminus \hat{A} \).

**Proof.** Since \( \bar{A} \) and \( \hat{A} \) are disjoint, by the previous theorem the equality applies.

**Definition 3.2.3 (separable).** A space \( X \) is said to be separable if it has a finite or countable dense subset

**Definition 3.2.4 (isolated).** A point \( x \in X \) is called isolated if the singleton \( \{x\} \) is open.

It is interesting that in topology a set is open if it belongs to the topology. Nothing is said about balls, metric or norms. Still we show next that an open set in topology has properties similars to those in metric spaces.

**Theorem 3.2.4.** A set \( O \) is open if an only every point of \( O \) is an interior point.
CHAPTER 3. BASIC TOPOLOGICAL CONCEPTS

Proof.

• “$\implies$”: Let us assume that $O$ is open. Choose any point $x \in O$. Then since $O$ is a neighborhood of $x$, this is an interior point.

• “$\impliedby$”: Let us assume that every point is interior. Then the set $O = \bigcup O_x$, where $O_x \subset O$ is a neighborhood of $x$, is open by the union axiom of a topology.

Let us see a few examples.

3.2.5 Examples

• The real numbers $\mathbb{R}$ with all the open intervals $(a, b)$ or arbitrary union of them, or finite intersection of them. In general in $\mathbb{R}^n$ the most basic open set is a ball. A ball centered at $x_0$ and with radius $r$ is defined as

$$B(x_0, r) = \{x : \|x - x_0\| < r\}$$

• The complex space $\mathbb{C}$. In the complex space ball is sometimes known as a $\rho$-neighborhood. $N_\rho(z_0)$, of a point $z_0$, is defined as an open disc centered at $z_0$ and with radius $\rho$. That is,

$$N_\rho(z_0) = \{z : |z - z_0| < \rho\},$$

Note that a $\rho$-neighborhood is a ball, while the term neighborhood of $z_0$ could apply to any open containing a point $z_0$, no necessarily a ball. In the real line balls are named as intervals and in the complex plane balls are known as open discs.

The set $S = \{z : \rho_1 < |z - z_0| < \rho_2\}$ is called an open annulus. A point $z$ is called a boundary point of $S$ if every neighborhood $N_\delta(z)$ has a non-empty intersection with $S$. That is if

$$N_\delta(z) \cap S \neq \emptyset, \quad \forall \delta > 0.$$  

• Pick any discrete set $\{x_i\}$. where the distance between two points is defined by the formula[2.2.1]
Note that the set \( \{x_i\} \) is a topological space if we consider each point \( x_i \) as an open set. This does not seem to be a useful example, however it comes handy when looking for counter-examples. This is known as the **discrete topology** and the distance \( d \) as the discrete metric.

Particular examples are the sets of integers \( \mathbb{Z} \) and natural numbers \( \mathbb{N} \).

Note that there is no need for the set to be discrete. For example the discrete topology could be set on the real numbers \( \mathbb{R} \). Still the metric \( d(x, y) = 0 \) if \( x \neq y \), and \( d(x, y) = 1 \), for \( x = y \), works fine and any single real number consider as a singleton \( \{x\} \) is an open set. This would be the largest topology in the real numbers. On the other extreme is the smallest topology associated with a set \( X \).

Topology is defined beyond metric spaces, and we do not need a metric to define a topology. Given any set \( X \) we can define the **discrete topology** by saying that all its subsets are open. There is no need of metric for this.

- **The indiscrete or trivial topology** consisting only of the two sets \( \{\emptyset, X\} \). The previous two topologies are pretty uninteresting but useful for counter-examples.

- After the trivial topology, the smallest topology is that of a singleton. That is if \( X = \{x\} \), then the topology \( (X, T) \), for this set is

\[
\{\emptyset, \{x\}\}.
\]

where every element of the topology is open and closed at the same time.

- The binary set \( X = \{x_1, x_2\} \) based discrete topology. That is, here

\[
T_s = \{\emptyset, X, \{0\}, \{1\}\}.
\]

where all sets are open, and all sets are closed.

- **The Sierpiński space:** It is based on a binary set \( X = \{0, 1\} \), where the topology is defined as

\[
T_s = \{\emptyset, X, \{1\}\}. \quad (3.2.2)
\]
where \( \{0\} \) is a closed set and \( \{1\} \) is an open set. In this way this is not a discrete topology (where all subsets are open) and in fact it is the smallest topology which is non-trivial, and no discrete.

Wikipedia\(^5\) presents a comprehensive description of this space and its properties.

The Sierpinsky space is a good source for counter-examples.

- **Counter examples:** Examples of spaces that are no topological.
  - Think about \( \mathbb{R} \) with \( X \) being the finite open sets \((a, b)\). That is \(|a - b| = r < \infty\). The following union of them \( \cup_{a \in \mathbb{R}} (a, a + 1) \) will generate the whole real line, so it is not “open” because its length is not finite.
  - Take all intervals of the form \((a, \infty)\), with \( a \in \mathbb{R} \). Then we can have intersections of the form \((a, b)\), with \( b < \infty \), so the reals with \( X \) defined with the property above is not a topological space.

In the last few lines we defined more than ten new concepts making the text dense. Let us use some examples to illustrates those concepts.

Section 2.4.4 shows a list of examples in metric spaces which are valid for topological spaces. It is important to see that in metric space the concept of open, which is the gene of the topology is based on a metric, while in topology a set is open if it belongs to the topology, whether there is a metric or not.

### 3.2.6 Examples/Counter-examples based on the discrete topology

- **Open and Closed** : Each set of the discrete topology space is open and closed at the same time. Every element of any topology is open by definition. Now, take that element away from the set for a moment. The other elements (the complement) are a union of each of them which are open. So the complement of that element is open. The element itself is closed, since its complement is open. So each element of the discrete topology is open and closed at the same time. This is a counter-example that illustrates that open and closed are not exclusive.

\(^5\)http://en.wikipedia.org/wiki/Sierpi%C5%84ski_space
Then you can use single elements to build any set of the topology with open and closed sets simultaneously.

- **Interior, Exterior, Balls, Neighborhoods and, Cluster points:**
  By definition of a ball centered at $x_0$, and based on the discrete metric \(d_{2.2.1}\), if we pick \(0 < r < 1\) then the ball is just the set \(\{x_0\}\), and if \(r \geq 1\), the ball is the whole space \(X\). There are no cluster points in the discrete topology. The only ball around a point \(x\) that takes points other than \(x\) itself is the whole space \(X\), which is a ball of radius 1. The interior of the discrete topology consist of all its points. Each point as a set is open, and a neighborhood of radius \(r < 1\), consist of the point itself. There is no exterior since the closure of the discrete topology is the discrete topology itself, and its complement is empty (with respect to itself).

- **Density:** The only dense set in the discrete topology is the set \(X\) itself. No proper subset of the discrete topology can be closed to get \(X\), since all subsets of the discrete topology are already closed. This provides counter-examples of topologies without proper dense sets.

- **Isolated points and separable:** For the discrete topology on the set \(X\), each singleton \(\{x\}\) is isolated, since it is open. The discrete topology is separable if the number of elements of \(X\) is finite or countable. Hence, for example, the real numbers with the discrete topology does not form a separable space.

### 3.3 Base of a topology

As in linear algebra we want to be able to represent a topology by using the last amount of elements to simplify the problem. The number of subsets of a set could be excessively large. For a finite set with \(n\) elements this is \(2^n\), for the natural numbers is \(2^{\aleph_0} = \aleph_1\), which is the cardianl of the continuum (the cardinal of the real numbers). Even much larger is if instead of using the natural numbers we use the real numbers we would talk about \(2^{\aleph_1} = \aleph_2\), which is an even larger cardinal.

**Definition 3.3.1.** Let \(\mathcal{B}\) be a collection of open sets in a topological space \((X, \mathcal{T})\), such that every open set can be written as a union of the open sets in \(\mathcal{B}\). Then \(\mathcal{B}\) is a base (or basis) for the topological space \(X\).
CHAPTER 3. BASIC TOPOLOGICAL CONCEPTS

Figure 3.1: $x \in B_1 \cap B_2$, and so there is a base (yellow) $B_3$ with $x \in B_3$.

It is much like in linear algebra saying that any vector of a vector space could be written as a linear combination of vectors of the base. Linear combinations are in linear algebra like unions in a topology. In fact the whole space $X$ can be represented as a union of its bases, since $X$ is open, then for any $x \in X$, $x \in B_i \in \mathcal{B}$, and so $X = \cup B_i$ with $B_i \in \mathcal{B}$. That the union of the bases is contained in $X$ is a trivial matter since each base set $B_i$ is a subset of $X$. However, while in linear algebra the cardinality of the basis is the same for all bases, in topology this is not the case and there could be bases with much larger open neighborhoods than others. Moreover the representation of a vector in term of its basis is unique, but the representation of an open sets in terms of union sets from the basis is not unique.

The following proposition is a direct consequence of the definition of a base $\mathcal{B}$ of a topology.

**Theorem 3.3.1.** Given a set $X$, a basis for a topology on $X$ is a collection $\mathcal{B}$ of subsets of $X$ such that

(i) For each $x \in X$, $x \in B$, for at least one $B \in \mathcal{B}$.

(ii) If $x \in B_1 \cap B_2$, with $B_i \in \mathcal{B}$, $i = 1, 2$, then there is a basis element $B_3$ such that $x \in B_3 \subset B_1 \cap B_2$.

Figure 3.1 illustrates the second part of this theorem.
3.3. BASE OF A TOPOLOGY

Proof. Let the basis be defined by the set \( B = \{ B_\alpha : \alpha \in A \} \)

(i) We showed above that \( X = \bigcup_{\alpha \in A} B_\alpha \), so clearly \( x \in X \) implies that \( x \in B_\alpha \) for some \( \alpha \in A \).

(ii) Assume \( x \in B_1 \cap B_2 \), with \( B_i \in B, i = 1, 2 \). The intersection \( B_1 \cup B_2 \) is open (since it is finite intersection). Then from the definition there should be an open set \( B_3 \), such that \( B_1 \cup B_2 \subset B_3 \).

If a topological space has a countable base it is called a **second countable**. A **base of neighborhoods of a point** \( x \) is a collection \( \{ B_\alpha \in A \} \) such that \( x \in \bigcap B_\alpha \). If any point of a topological space has a countable base of neighborhoods, then the topology is called **first countable**.

In contrast with linear algebra where all basis have the same cardinality, in topology this is not the case. Even worse, a topology \( T \) is a basis for itself. That is, \( T \) is a basis for \( T \). The smallest possible cardinality of a base is called the **weight** of the topological space.

Finally before showing a few examples, we assert that every topological space \( X \) with a countable basis is separable. To see this, pick a point in each set of the countable basis of the space \( X \). The set of such points is countable. Now, since each set of the basis can adjust to any open set of \( X \) no matter its size, any point chosen can be approximated to any object in \( X \) with any degree of precision, and so the picked points are dense in \( X \). So \( X \) is separable.

Let us now list a few examples.

3.3.2 Examples

- Let us consider the real numbers \( \mathbb{R} \). Think about all the open intervals of the form \((a, b)\), where \( a \) and \( b \) could be \(-\infty\) and \( +\infty\). This form a basis for the topology of open sets in the reals. Every open set can be seen as a union of open sets of this kind. However the cardinal of this set is too large. We could consider only open sets with center in the rational numbers. That is also a base for the topology but still has a very large size. Then we consider all the open sets with center at the rational numbers and radius a rational number. It can be shown that any open in \( \mathbb{R} \) can be written as a union of sets of this
kind, and this is a countable number of sets, hence the real numbers with the usual (open set) topology is second countable.

- The previous item could be extended to $\mathbb{R}^n$ using the same arguments. This time we want to locate centers of balls in points with rational coordinates and/or rational radius. In particular for the complex space $\mathbb{C}$ we can think about open discs with rational centers and rational radius.

- Let us consider the discrete topology on a set $X$. If $X$ is finite or countable, then it is second countable and first countable, otherwise it is only first countable. Moreover if $X$ is finite with $n$ elements. The total number of open sets is $2^N$, but we only need $N$ singletons $\{x_i\}$ to generate all the $2^N - 1$ non-empty open subsets of $X$ and add to this the empty set. So the base is certainly much smaller than the topology size.

We now define the term subbasis and show that indeed it generates at topology.

**Definition 3.3.3 (subbasis).** A **subbasis** for a topology $X$ is a collection of subsets $S = \{O_\alpha\}$ of $X$ such that $X = \bigcup \alpha O_\alpha$. The **topology generated by the subbasis** $S$ is defined to be the collection $T$ of all unions of finite intersection of elements of $S$.

To verify that indeed $S$ generates the topology $T$ we can proof that the finite intersection of elements of $S$ produce the basis $B$ of $T$, because then, by definition [3.3.1], their arbitrary unions produce the topology $T$.

We use definition [3.3.1]. First, given $x \in X$, then since $X = \bigcup \alpha O_\alpha$, then $x \in O_\alpha$ for some $\alpha \in B$. Now let us assume that

$$B_1 = \bigcup_{i_1} O_{i_1} = \bigcup_{i_2} O_{i_2}$$

are two arbitrary elements of $B$. Now

$$B_1 \cap B_2 = (\bigcup_{i_1} O_{i_1}) \cap (\bigcup_{i_2} O_{i_2}).$$

where $i_1, i_2 \in \mathbb{N}$ can take only a finite number of values, and so the intersection is still finite. So $B_1 \cap B_2 \in B$. 

3.4 The Concept of Continuity and Homomorphism

The concept of continuity in analysis, calculus, and complex variable is a particular case of that of continuity in topology. Topology carries a great deal of generality. It does not need a metric to define continuity, it has its own internal rules. The amazing thing is that with these minimal rules metric and other spaces inherit the topological continuity of open sets. Nowhere else the definition of continuity is more general than in topology.

We have the following definitions:

Definition 3.4.1.

- Let \((X, \mathcal{T})\) and \((Y, \mathcal{S})\) be topological spaces. A map \(f : X \to Y\) is said to be **continuous** if it returns open sets in the range to open sets in the domain. That is if for any open set \(O \in \mathcal{S}\), \(f^{-1}(O) \in \mathcal{T}\).

- \(f\) is an **open map** if it sends open sets from \(\mathcal{T}\) to open sets in \(\mathcal{S}\). That is, if for each \(O \in \mathcal{T}\) (that is, open), \(f(O) \in \mathcal{S}\) is open.

- \(f\) is an **closed map** if it sends closed sets from \(\mathcal{T}\) to closed sets in \(\mathcal{S}\). That is, if for each \(C \in \mathcal{T}\) (that is, closed), \(f(C) \in \mathcal{S}\) is closed.

- \(f\) is **continuous at the point** \(x\) if for any neighborhood \(A\) of \(f(x)\) in \(Y\) the preimage \(f^{-1}(A)\) contains a neighborhood of \(x\).

- A function \(f\) from a topological space to \(\mathbb{R}\) is said to be **upper semicontinuous** if \(f^{-1}(-\infty, c) \in \mathcal{T}\) for all \(c \in \mathbb{R}\), and **lower semicontinuous** if \(f^{-1}(c, \infty) \in \mathcal{T}\) for all \(c \in \mathbb{R}\).

It can be shown that continuity at every point of \(f\) is equivalent to continuity of the map.

Provided a set of maps \(\mathcal{F} = \{f_{a \in A}\}\) from \(X\) to \(Y\), and assuming that \(Y\) is a topology, we can construct a topology in \(X\) by defining as open all the inverse images of maps \(f_a\) from \(Y\). These inverse images, with their arbitrary union and finite intersections provide the weakest topology under the collection of maps \(\{f_{a \in A}\}\). If the set \(\mathcal{F}\) consist of a single map \(f\) the induced topology in \(X\) is know as the **pullback topology**. It is precisely this property which allows as to define the quotient topology in section 3.6.10.
Category Theory\(^6\) study the morphisms which are equivalence relations between different spaces. In the case of topological spaces we have the following definition

**Definition 3.4.2.** A map \( f : X \to Y \) between topological spaces is a homeomorphism if it is continuous and bijective with continuous inverse.

If there is a homeomorphism \( X \to Y \), then \( X \) and \( Y \) are said to be homeomorphic or topologically equivalent.

### 3.4.3 Examples:

Continuity is a concept which any student of topology should be aware of. Before the first topology course, a student should have taken at least one analysis course. All examples in real and complex analysis are valid in topological spaces. However continuity in topology goes beyond continuity in analysis. We provide here a few examples which are not found in regular analysis courses and seem counter-intuitive.

(i) Every function between two discrete topologies \( X \) and \( Y \) is

- continuous,
- open,
- closed.

This is obvious since all the subsets of \( X \) and subsets of \( Y \) are open, and all subsets of \( Y \) are closed. In any case any function returns open sets in the range to open sets in the domain, sends open sets in the domain to open sets in the range, and sends closed sets in the domain to closed sets in the range.

(ii) The Sierpiński space is defined in \(3.2.2\). We recall that \( S \) provides the set \( S = \{0, 1\} \) with a topology where the open sets given by

\[
\mathcal{T}_S = \{\emptyset, S, \{1\}\}.
\]

Note that \( \{0\} \) is a closed set. This set is useful to build counter-examples.

\(^6\)http://en.wikipedia.org/wiki/Category-theory
Functions on the Sierpiński space $S$. Let $X$ be an arbitrary set. Given a binary property we can partition any set according to that property. That is, whether a given element satisfies or not the property. This introduces in a natural way the characteristic function of the set. That is

$$\chi_U(x) = \begin{cases} 1 & x \in U \\ 0 & x \notin U. \end{cases}$$

Now let the property above be “open” in a topology $(X, T)$. Then, the characteristic function from the set $X$ with topology $T$, is continuous. The reason is that $\{1\} \in S$ and $\chi_U^{-1}(1) = U$, which is precisely open in $T$. So we have a continuous function, not in the discrete topology, such that only takes the values 0 and 1. Then we can characterize the topology $(X, T)$ as the set of continuous functions from $X$ to the Sierpiński space $S$, and say that the topology $T$ is homeomorphic to the space of continuous functions from $X$ to the Sierpiński space $S$, or $C(X, S) \cong T$.

(iii) Let us consider the constant map $f(x) = 1$, in the Sierpiński space $S$. $f$ is continuous since $\{1\}$ is an open set in $S$, and the inverse $f^{-1}(1) = S$ which is open. $f$ is an open map since for any open set in $S$, the image is $\{1\}$ which is open in $S$. Finally, $f$ is not closed since the image of $\{0\}$ which is closed in $S$ is $\{1\}$ which is open in $S$. On the other hand the mapping $f(x) = 0$, is closed and not open. Note that still the mapping $f(x) = 0$ is continuous since the only open having 0 as an element is $S$ itself and $f^{-1}(S) = S$.

(iv) Let us consider the set $X = \{0, 1\}$, and two topologies. The discrete topology $X_1 = (X, T_D)$, and the indiscrete topology $X_2 = (X, T_I)$. Then consider the identity function

$$i : X_2 \rightarrow X_1$$

$$x \mapsto x$$

This function is not continuous since the element $\{0\}$ is open in $X_1$, but $i^{-1}(\{0\}) = \{0\}$, which is not open in $X_2$.

This is an example where even the identity function is not continuous. If we reverse the order of $X_1$ and $X_1$ this would be an example where
the identity is continuous but it is not an open map neither a closed map.

Finally we prove an important theorem of continuity which will be used often in the book. The composition of functions is continuous.

**Theorem 3.4.1 (Continuity of function composition).** Let $X, Y$ and $Z$ be three topological spaces and $f : X \rightarrow Y$, $g : Y \rightarrow Z$ continuous. Then $g \circ f : X \rightarrow Z$ is continuous.

**Proof.** It is interesting to observe that the proof of this theorem is purely topological which is much more general than the proof provided in Calculus, still the proof here is straightforward.

Let us assume that $O_Z$ is an open set in $Z$, we show that $(g \circ f)^{-1}(O_Z)$ is open in $X$. By definition $(g \circ f)^{-1}(O_Z) = f^{-1}(g^{-1}(Z))$. Since $g$ is continuous $g^{-1}(Z)$ is open in $Y$, and since $f$ is continuous $f^{-1}(g^{-1}(Z))$ is open in $X$. \( \square \)

### 3.5 Topology Sizes

Assume two topologies $\mathcal{T} \subset \mathcal{S}$. We say that $\mathcal{S}$ is

- **stronger** than $\mathcal{T}$ and
- **$\mathcal{T}$ is weaker** than $\mathcal{S}$

If $\mathcal{T} \subset \mathcal{S}$ we say that $\mathcal{S}$ is **compatible** with $\mathcal{T}$, or either $\mathcal{T}$ is compatible with $\mathcal{S}$. Then both the empty set $\emptyset$ and the space $X$ are compatible with all topologies generated by $X$.

The word **finer** is used as synonymous of stronger, while the word **coarser** is used as synonymous of weaker. Given any set $X$ the weakest topology is the indiscrete or trivial topology given by $\{\emptyset, X\}$, and the strongest is the discrete topology where each element of $X$ is an open set. Any other topology associated with a set $X$ is in between these two. As an example consider the binary set $\{x_1, x_2\}$ with the discrete topology. The open sets are $\{x_1\}, \{x_2\}, \{x_1, x_2\}, \emptyset$. We can chain the topologies from the weaker to the strongest as
3.5. **TOPOLOGY SIZES**

![Diagram of topologies]

We see two chains, from which we can construct the following topologies:

(i) \(\{\emptyset, \{x_1\}\}\)

The indiscrete (trivial), discrete, weakest, and strongest for \(X = \{x_1\}\).
Not the weakest, nor the strongest for \(X = \{x_1, x_2\}\).

(ii) \(\{\emptyset, \{x_2\}\}\):

The indiscrete (trivial), discrete, weakest, and strongest for \(X = \{x_2\}\).
Not the weakest, nor the strongest for \(X = \{x_1, x_2\}\).

(iii) \(\emptyset, \{x_1, x_2\}\) :

The weakest for \(X = \{x_1, x_2\}\).

(iv) \(\emptyset, \{x_1\}, \{x_2\}, \{x_1, x_2\}\) :

The discrete and strongest for \(X = \{x_1, x_2\}\).

In general for a finite set \(X = \{x_i\}_{i=1}^n\) we can construct \(2^n\) open sets of the discrete topology corresponding to the \(2^n\) subsets of \(X\).

Counting topologies in finite sets is an interesting problem and it belongs to combinatorial theory. For example, we found that 4 topologies could be extracted from a set of 2 elements. However the topologies (i) and (ii) in the above list are equivalent. That is the topological spaces \(\{\emptyset, \{x_1\}\}\) and \(\{\emptyset, \{x_2\}\}\) are homomorphic. There is a bijective to-one function between them and they are open sets in the discrete metric. The name \(x_1\) or \(x_2\) is irrelevant, so there are actually only 3 distinct topologies in the set \(x_1, x_2\). A more interesting exercise is to find the discrete topologies of \(X = \{x_1, x_2, x_3\}\), then count the total number of them, and the total number of distinct (no counting homomorphisms) topologies. This problem is solved in [Wikipedia](http://en.wikipedia.org/wiki/Finite_topological_space) where the general counting problem is addressed with the notation \(T(n)\).

\[7\text{http://en.wikipedia.org/wiki/Finite_topological_space}\]
3.6 Derived Topologies

We now show how to construct new topologies from old topologies and sets. In what follows we define the following concepts:

- The Induced Topology
- The Intersection Topology
- Topologizing of sets
- The Product Topology
- The Quotient Topology

3.6.1 The Induced Topology: Subspaces of Topological Spaces

Let us assume that the old topology is $X$. For any subset $Y \subset X$, we can build a new topology by defining as open sets the intersection of $Y$ with the open sets of $X$. That is the induced topology is defined as

$$\mathcal{T}_Y := \{O \cap Y | O \in \mathcal{T}\}.$$ 

It is easy to show that $\mathcal{T}_Y$ satisfies the conditions of a topology space. As an example consider $\mathbb{R}^3$. Open balls in $\mathbb{R}^3$ are spheres, take the subset with coordinates $(x, y, 0)$ corresponding to the horizontal plane. The intersection of this plane with the spheres are discs, which are the natural neighborhoods in the plane. The subset with coordinates $(x, 0, 0)$ corresponding to the $x$-axis intersects the spheres in open intervals. It is interesting that the set does not have to be a subspace from the linear algebra point of view. Any arbitrary set, a surface, curve, a set of random points, any set would work.

3.6.2 The Intersection Topology

Let us assume a family of topologies $\{\mathcal{T}_\alpha : \alpha \in \mathcal{A}\}$. Then the intersection of the family is a topology.

We claim that:

$$\bigcap_\alpha \mathcal{T}_\alpha = \{O : \forall \alpha \in \mathcal{A}, O \in \mathcal{T}_\alpha\}$$
is a topology. This is true because any element is open in all topologies. So it is any arbitrary union of them or finite intersection of them.

The union is not necessarily a topology. Let us see what conditions could we impose in the union of topologies to make it a topology. Let us assume that topologies $T_1$ and $T_2$ corresponding to the same set $X$. Let $T = T_1 \cup T_2$. Trivially $\emptyset \in T$, and since $X \in T_1$ and $X \in T_2$ then $X \in T$. Let us review the three axioms of a topology.

(i) **Arbitrary union of open sets:** We need to have that if $\alpha \in A$, where $A$ is an arbitrary index set, then

$$\bigcup_{\alpha \in A} O_\alpha$$

is open for $O_\alpha \in T$.

Since this union is of sets of the union topology, some sets are in $T_1$ and some in $T_2$. Let us divide the sets into the two categories as follows:

$$\bigcup_{\alpha \in A} O_\alpha = (\bigcup_{\alpha \in A_1}) \bigcup (\bigcup_{\alpha \in A_2}) = U_1 \cup U_2.$$

where $A \supset A_1 \cup A_2$, and $A_i$ is the sets of indices participating from the topology $T_i$, $i = 1, 2$.

At this point we have no way to tell that $U_1 \cup U_2$ is open in $T$. Yes, they are open in $T_1$ and $T_2$, but is their union an open set of $T$? The answer is no necessarily, here is a counter example: Define $X = \{a, b, c\}$, $T_1 = \{\emptyset, X, \{a, c\}\}$ and $T_2 = \{\emptyset, X, \{a, b\}\}$. Then

$$T = T_1 \cup T_2 = \{\emptyset, X, \{a, b\}, \{a, c\}\}$$

Now, this is not a topology since

$$\{a, b\} \cup \{a, c\} = \{a, b, c\}$$

is not an open set of the union $T$.

(ii) **Finite intersection of open sets:** Here we want to know if under which circumstances is

$$\bigcap_{i \in I} O_i$$
for \( I \) a finite index set and \( O_i \) is an open set of \( \mathcal{T} \). Again, this operation can be divided in two sets

\[
\bigcap_{i \in I} O_i = \left( \bigcap_{i \in I_1} O_i \right) \cap \left( \bigcap_{i \in I_2} O_i \right) = U_1 \cap U_2.
\]

where \( I = I_1 \cup I_2 \), and \( U_i \) is open in \( \mathcal{T}_i \) for \( i = 1,2 \). However can we say that \( U_1 \cap U_2 \) is open in the union topology? Clearly the answer is “no necessarily” and we show this by recycling the previous counter-example, this time for intersection. In this case choose the two open sets \( \{a, b\} \in \mathcal{T}_1 \) and \( \{a, c\} \in \mathcal{T}_2 \), then the intersection is \( \{a\} \), which is not an open set of the union set \( T \) in equation 3.6.3.

In preparation for the discussion on the product topology we introduce another counter-example that shows that the union of two product topologies is no necessarily a topology.

**Example 3.6.2.1.** Assume two topological spaces \( X_1 \) and \( X_2 \). Define the following topologies:

\[
\mathcal{T}_{X_1} = \{ O_1 \times X_2 : O_1 \text{ is open of } X_1 \}
\]

\[
\mathcal{T}_{X_2} = \{ X_1 \times O_2 : O_2 \text{ is open of } X_2 \}
\]

It should be clear that \( \mathcal{T}_{X_1} \) as well as \( \mathcal{T}_{X_2} \) are topologies. However the union is not because \((O_1 \times X_2) \cap (X_1 \times O_2) \notin \mathcal{T}_{X_1} \cup \mathcal{T}_{X_2}\). In fact assume that \( X_1 = X_2 \) and take a proper set open set in both topologies \( O \in \mathcal{T}_i \), \( i = 1, 2 \). Then \( O \times X_2 \) is not even equal \( X_1 \times O \). Since \( O \times X_2 \) allows the whole space as a second component, while \( X_1 \times O \) allows the whole space in the first component and \( O \subsetneq X \) is a proper set of \( X \). In fact \((O_1 \times X_2) \cap (X_1 \times O_2) = (O_1 \times O_2) \) which could not be in \( \mathcal{T}_{X_1} \cup \mathcal{T}_{X_2} \).

Where is the flow in the following reasoning? Consider the induced topology of the reals in the interval \((0, 1)\). All sets, of the form \((0, 1 - 1/n)\), for \( n = 1, 2, \cdots \), are open. An arbitrary union of them is not. That is

\[
\bigcup_{i=1}^{\infty}(0, 1 - 1/n) = (0, 1]
\]

which is no open.

The intersection topology is the weakest of all topologies of the family.
3.6. DERIVED TOPOLOGIES

3.6.3 Topologizing on sets

We want to know how to build a topology given that we have a family of sets. Let us then assume that we have a family

\[ X_A = \{ A_\alpha : \alpha \in A \} \]

of sets of \( X \). To build the topology \( (X_A, \mathcal{T}) \) we start with the empty set \( \emptyset \), the whole set \( X \), all finite intersections of \( A_\alpha \) and arbitrary unions of \( A_\alpha \). This generates a set call subbasis for \( \mathcal{T}_A \). By construction of \( \mathcal{T}_A \) is a topology. In fact it is the smallest containing the sets in \( X_A \), since the intersection reduces it to the minimum.

3.6.4 The Product Topology

In basic set theory, there is a definition of Cartesian product in a finite collection of sets. For example let us assume two sets \( A_1 \) and \( A_2 \). The cartesian product of \( A_1 \) and \( A_2 \). That is the set

\[ \prod_{i=1}^{2} A_i = \{ (x, y) : x \in A_1, y \in A_2 \}. \]

We can think of \( \mathbb{R}^2 \) as the cartesian product of \( \mathbb{R} \) with itself. The same definition can be extended to a finite family of sets where we can write

\[ \prod_{i=1}^{n} A_i = \{ (x_1, x_2, \cdots, x_n) : x_i \in A_i \} \]

We can extend this to infinite countable sets as follows

\[ \prod_{i=1}^{\infty} A_i = \{ (x_1, x_2, \cdots) : x_i \in A_i \}. \]

In fact, any element of this product is an infinite sequence. That is, we can see sequences as elements of a countable infinite product of sets. We want to generalize this definition to the product of an arbitrary family of sets where the index \( \alpha \) now can run along a continuum or along an infinite non-countable set. In this case we can not index using integer numbers but we can say that an element of the form \( (x_\alpha) \) is an element where the component \( x_\beta \) corresponds to a set \( A_\beta \), with \( \beta \in \mathcal{A} \). We can not write the sequence,
Figure 3.2: An element \((\alpha^2)\) of the product space of reals where the index \(\alpha\) is the interval \([0, 1]\) (in green). Think that \(A_\alpha\) is a set of the reals numbers \((\mathbb{R})\). We can pick from the set of those reals, the square of the index. The result is the red line in the figure.

but we can think about a function of the index set \(\mathcal{A}\) to the product objects \((x_\alpha)\) such that there is an association of each \(\beta \in \mathcal{A}\) with the component \(x_\beta\) in the set of \((x_\alpha)\)s, and write

\[
\prod_{\alpha \in \mathcal{A}} A_\alpha = \{(x_\alpha) : \alpha \in A_\alpha\}.
\]

For example, consider \(\mathcal{A} = [0, 1]\). We can think of

\[
\prod_{\alpha \in [0,1]} A_\alpha = \{(x_\alpha) : \alpha \in [0, 1]\}.
\]

One element of this set is the function

\[
f : [0, 1] \to \mathbb{R}
\]

\[
\alpha \mapsto x_\alpha = \alpha^2
\]

where we used the \(x\)-axis for the indices and the \(y\)-axis for the element of the product space. Figure 3.2 illustrates this particular element of the product set. Note that if we used instead the function \(f(\alpha) = \alpha\) we get the same
red line in the $y$-axis but the density distribution of the points is different. The blue line (the curve of the function) provides a hint about the density of the points along the vertical. This is just an interpretation. The concept of arbitrary product spaces is much more general. Instead of the nice function ($\alpha^2$) we could have a random set of numbers, or the domain does not have to be in a nice interval such as [0, 1] but any subset of the real line. We say that the representation of the product as shown in equation 3.6.4 is a functional representation of the product.

The inverse operation of a product is a projection. That is a projection takes one component on the product coordinates. For example we can see the $x$-axis in $\mathbb{R}^2$ as the projection of $\mathbb{R}^2$ into is first coordinate. More explicitly if $P_x(x, y) = x$. Similarly the $y$-axis would be $P_y(x, y) = y$. Note that projections are not one-to-one, and so they can not be inverted, however they are the right inverses of the Cartesian products, in the sense that if you start with a set $A_1$ and extend it via a Cartesian product $\prod_{i=1}^2 A_i$, then when you project into the first set you get back to $A_1$.

However when talking about topologies we need to be a bit more careful. We could say that the unions of products of open sets on each of the participating spaces are open and in this way we can build the product topology. However we want to refer the reader to example 3.6.2.1 where the union of just two topological spaces is no necessarily a topology. There are two main issues to consider:

(i) The cardinality of family $\{X_{\alpha \in A}\}$, that is, the size of the index set $A$. Mainly if it is of finite size, or not. Here is a bit of history from the Handbook of the History of General Topology, Volume 1 by C. E. Aull. Details, more history, and references are found there.

The product topology spaces are known today as Tychonoff spaces. Tychonoff in 1930 gave credit to Usysohn for the concept which was originally introduced by him (Usysohn) in 1925 with the name complete regular, as the topological product of spaces. Tukey in 1940 suggested renaming them as Tychonoff spaces. The concept of product topologies on families with index sets $A$ of any size was introduced by Tietze in 1923 with the name box product, instead of topological product. It is not just a name. If the number of sets in the family

---

8 In the literature is common to see $p_x$, or $\pi_x$ instead of $P_x$ for projection.

9https://books.google.com/books?isbn=0792344790
indexed by \( A \) is finite, the box product coincides with the topological product originally proposed by Usysohn. If the number of sets is infinity, the box product is not the best option and here the topological product (which is not a direct abstraction from the finite case) is more adequate.

(ii) The kind of topology that we want to provide in the new product space. For example, we could think about the discrete or indiscrete but these are not interesting. We would like to see more significant topologies obtained from the product of spaces. That is, we would like to inherit structure from their product components (the projections), such as for example the open sets, basis. This would mean that the projections should be continuous.

We consider these cases by dividing them into two sections. The finite case and the infinite case.

3.6.4.1 Finite Products

Let us, for simplicity start with an index set with only two elements, \( A = \{1, 2\} \). Then our family only has the two sets: \( X_1 = X \) and \( X_2 = Y \). We ask what are the open sets in the product topology? We already know how to find products of sets, but we want these sets to inherit some topology from their component spaces. It seems natural to think that the open sets in the product space are products of open sets in the component spaces \( X \) and \( Y \). That is, we would like to consider the product of the of all open sets \( O_X \subset X \), and \( O_Y \subset Y \) and show that this new product form a basis for a topological space that we call product topology.

That is, let us define the product topology as follows. Consider the family product

\[
S = \{O_X \times O_Y\}
\]

where \( O_X \) and \( O_Y \) as indicated above, are arbitrary open sets of \( X \) and \( Y \) respectively.

Let us us consider as an example the Sierpiński spaces topologies on the sets \( X = \{0, 1\} \), and \( Y = \{a, b\} \),

\[
\mathcal{T}_1 = \{\emptyset, \{1\}, \{0, 1\}\}
\]

\[
\mathcal{T}_2 = \{\emptyset, \{a\}, \{a, b\}\}.
\]
Then we construct

\[ T_1 \times T_2 = \{ \emptyset, \{1\} \times \{a\}, \{0,1\} \times \{a\}, \{0,1\} \times \{a,b\} \} \]

We include the empty set \( \emptyset \) and \( X \times Y \) in the collection of open sets in the product topology. Pick an arbitrary union of open sets in the product \( O_{X_\alpha} \times O_{Y_\beta} \), that is

\[ \bigcup_{\alpha,\beta} O_{(X \times Y)} \overset{def}{=} \bigcup_{\alpha,\beta} (O_{X_\alpha} \times O_{Y_\beta}) = \left( \bigcup_{\alpha,\beta} O_{X_\alpha} \right) \times \left( \bigcup_{\alpha,\beta} O_{Y_\beta} \right) \]

Then an arbitrary union of open sets in the product defined here produced also open sets in the product topology. Now for finite intersection of open sets we see that:

\[ \bigcap_{\alpha,\beta} O_{X \times Y} \overset{def}{=} \bigcap_{\alpha,\beta} O_{X_\alpha} \times O_{Y_\beta} = \left( \bigcap_{\alpha,\beta} O_{X_\alpha} \right) \times \left( \bigcap_{\alpha,\beta} O_{Y_\beta} \right) \]

and the finite intersection property is passed from the components to the product.

We show that \( S \) is a basis for the product topology. More formally let us define the new topology

**Definition 3.6.5.** Let \( X \) and \( Y \) be topological spaces. The **product topology** is the topology having as basis the collection \( B \) of all the sets of the form \( O_{X_\alpha} \times O_{Y_\beta} \), where \( O_{X_\alpha} \) is an open set of \( X \) and \( O_{Y_\beta} \) is an open set of \( Y \).

Let us now prove that indeed the set \( B \) defined above is a basis. Let us prove that indeed the product topology for this particular is a topology.

**Theorem 3.6.6.** The collection \( B \) in the above definition is a basis.

**Proof.** We use theorem [3.3.1]. That is

(i) Let us choose \((x, y) \in X \times Y\). Since \( x \in X \), and \( y \in Y \), then \( x \in O_x \) for some open set in \( X \) and \( y \in O_y \) for some open set in \( Y \), by definition of the basis \( B_i \), \( i = 1, 2 \) in the spaces \( X \) and \( Y \), respectively. Then \((x, y)\) is in the open set \( O_x \times O_y \) of the topology \( X \times Y \).
(ii) Let us now assume that \((x, y) \in B_1 \cap B_2\), with \(B_i \in \mathcal{B}_i\), \(i = 1, 2\). From the definition of the product basis, \(x \in B_{1x}\) and \(y \in B_{1y}\), for \(i = 1, 2\) which are basis members in \(X\) and \(Y\) respectively. Hence, there is a \(B_{3x} \subset B_{1x} \cap B_{2x}\), and \(B_{3y} \subset B_{1y} \cap B_{2y}\), such that \(x \in B_{3x}\) and \(y \in B_{3y}\), since both \(X\) and \(Y\) are topologies and \(B_{1x}\) and \(B_{2x}\) are in a basis for \(X\) and likewise \(B_{1y}\) and \(B_{2y}\) are in a basis for \(Y\). Hence \((x, y) \in B_3 = B_{3x} \times B_{3y} \in \mathcal{B}\). Then \(\mathcal{B}\) is a basis for the product space \(X \times Y\).

While the product topology is closed under arbitrary union of sets of \(\mathcal{B}\), the basis \(\mathcal{B}\) is no necessarily closed under the union operation. Let us construct a simple example. Consider the real intervals \((0, 0)\) and \((2, 2)\) and the product \(B_1 = (0, 0) \times (2, 2)\), now consider the intervals \((1, 1)\) and the product \(B_2 = (1, 1) \times (3, 3)\). The union \(B_1 \cup B_2\) can not be generated from a product of open sets in \(\mathbb{R}\). Figure 3.3 illustrates this situation. The difficulty here rises from the fact that the product of the union is not equal to the union of the product. That is the operations product and union do not necessarily commute. What we see in Figure 3.3 is the union of two products. That is \([(0, 2) \times (0, 2)] \cup [(1, 3) \times (1, 3)]\). The product of the union is \([(0, 2) \cup (1, 3)] \times [(0, 2) \cup (1, 3)] = (0, 3) \times (0, 3)\) which is the smallest square containing both, the blue and the green squares. As a review of set theory note that

\[
[(0, 2) \cup (1, 3)] \times [(0, 2) \cup (1, 3)] = [(0, 2) \times (0, 2)] \cup [(0, 2) \times (1, 3)]
\cup [(1, 3) \times (0, 2)] \cup [(1, 3) \times (1, 3)],
\]
and clearly only the first and fourth sets in the right hand side correspond to the union of the product. In general the union of the product is a subset of the product of the union. That is,

\[
\bigcup_{\alpha} \prod_{\beta} O_{\alpha\beta} \subset \prod_{\beta} \bigcup_{\alpha} O_{\alpha\beta}
\]

and this comes naturally from the distributive law of products against unions, exactly as it does in the distribution of multiplication against sums by saying that \(\sum_{i} \prod_{j} a_{ij} \neq \prod_{j} \sum_{i} a_{ij}\), where in the left we have only \(n\) terms of \(j\)-products and on the right we have \(nm\) terms of \(j\)-products. The same rules of distribution between products and intersections apply.

However we need to be careful. We can write

\[
\left( \bigcup_{\alpha} U_{\alpha} \right) \times \left( \bigcup_{\beta} V_{\beta} \right) = \bigcup_{\alpha,\beta} (U_{\alpha} \times V_{\beta}).
\]

and more generally

\[
\bigcup_{\alpha} \prod_{\beta} O_{\alpha\beta} = \prod_{\beta} \bigcup_{\alpha} O_{\alpha\beta}
\]

The difference between this equation and (3.6.5) is the coupling of indices. Here the union is take over the whole set \(\alpha_{\beta}\) which is a set, of a set of indices. The product also is evaluated over those indices, but then every open set \(O_{\alpha_{\beta}}\) takes only one of the indices \(\alpha_{\beta}\). That is, by having an two indeces \(\alpha_{\beta} \neq \alpha_{\gamma}\), the corresponding open sets \(O_{\alpha_{\beta}}\) and \(O_{\alpha_{\gamma}}\) belong at different spaces and their intersection is empty. Here we do not let the indices run freely but they should be coupled in corresponding to the set. Again, these matters are easier to understand with sums and products which are everyday objects. What we are saying is that

\[
\sum_{i,j} a_{i}b_{j} = \sum_{i} a_{i} \sum_{j} a_{j},
\]

or in general

\[
\sum_{i_k} \prod_{k} a_{i_k} = \prod_{k} \sum_{i_k} a_{i_k}.
\]

Next, we illustrate a few examples about the new open sets created while doing products between different spaces.
CHAPTER 3. BASIC TOPOLOGICAL CONCEPTS

Figure 3.4: Open sets in rectangular (left) and polar (right) coordinates. The yellow regions represent a typical base open set on each coordinate system, constructed by the Cartesian product $U \times V$, where $U$ and $V$ are open sets in each of the component coordinates.

3.6.6.0.1 Examples

- **Open sets in $\mathbb{R}^2$.** We think naturally of open sets in $\mathbb{R}^2$ as discs. However the box topology (and hence the name) produces rectangular open sets instead of circular open sets. Take two intervals $(a_i, b_i)$ in the real line $i = 1, 2$, and in $\mathbb{R}^2$ the product $(a_1, b_1) \times (a_2, b_2)$. Figure 3.4 (left frame) illustrates this.

- **Open sets in polar coordinates in $\mathbb{R}^2$.** We could think that the reason we do not get discs as the open sets in $\mathbb{R}^2$ is because we are using rectangular coordinates which generate box type open sets. Let us then think in the space $\mathbb{R}^2 - \{0\}$, where we remove the origin because it is a singularity in the polar coordinate transformation. Let us call the new coordinates as $r$, and $\theta$. If we allow open sets in the $r$ coordinate as $(a_r, b_r)$ and open sets in the $\theta$ coordinate as $(a_\theta, b_\theta)$, then the product $(a_r, b_r) \times (a_\theta, b_\theta)$, is not a disc either. It is a “piece of cake”. Figure 3.4 (right frame) illustrates this.

- **Open sets in spherical coordinates:** The equations that define the transformation between Cartesian and spherical-cylindrical coordinates
3.6. DERIVED TOPOLOGIES

are

\[
\begin{align*}
  x &= r \sin \theta \cos \varphi \\
  y &= r \sin \theta \sin \varphi \\
  z &= r \cos \theta.
\end{align*}
\]

Here \(0 < r < \infty\) is the radius, \(0 \leq \varphi < 2\pi\) is the azimuthal angle, and \(0 \leq \theta \pi\) is the polar angle. The origin \(O = (0, 0, 0)\) is a singular point the transformation so the transformation is well defined in \(\mathbb{R}^2 - \{O\}\).

We consider the product of the three (closed) intervals \(C = [r, r + dr] \times [\varphi, \varphi + d\varphi] \times [\theta, \theta + d\theta]\). The open set is the interior of this interval that is \(O = \mathcal{C}\).

Figure 3.5 illustrates a typical open set in the spherical coordinates. The open set is enclosed by 6 surfaces. Two spheres, one with radius \(r\), and the other with radius \(r + dr\) (gray color) and 4 planes. Two polar planes (no colored except for the blue shade between the two spheres) with corresponding to \(\theta = \text{constant}\) and \(\theta = \text{constant} + d\theta\) and two azimuthal planes (pink), a plane \(\varphi = \text{constant}\) and the other plane \(\varphi = \text{constant} + d\varphi\).

- **A Cylinder:** A Cylinder can be seen as the Cartesian product of a line and a circle. Think, for example in the interval \(A = (0, a)\), and a disc

\[
D = \{(x, z) : x^2 + z^2 < 1, x, z \in \mathbb{R}\}.
\]

Then the product \(A \times D\) is the an open cylinder in \(\mathbb{R}^3\). The product of an open set in \(A\), \(O_A\) with an open set in \(D\), \(O_D\) will be an open cylinder, with the shape of the open set \(O_D\). Figure 3.6 illustrates the product \(A \times D\).

- **The Torus:** Think about the cylinder in Figure 3.6 and bend it as shown Figure 3.7. That is set the interval to \(A = [0, 2\pi)\) and

\[
T : A \to B \subset \mathbb{R}^2 \\
  t \mapsto (R \cos t, R \sin t),
\]

with \(R > r\), and \(r\) is the radius of the disc \(D\). We can think about the Torus as the product \(\mathcal{T} = B \times D\), which lives in \(\mathbb{R}^4\), but we can see
110

CHAPTER 3. BASIC TOPOLOGICAL CONCEPTS

Figure 3.5: An open set and its closure (blue) in spherical coordinates. The open set is the interior surrounded by six surfaces. Two spheres (gray), two polar planes which are not colored, except for the blue sides in between the two spheres, and two azimuthal planes (pink). These are in the front and back of the prism-like volume from the observers point of view.

$T$ as a 1D manifold, and $a$ by changing the coordinates from Cartesian to polar with $R$ fixed, we see the product $A \times D$ as a subset of $\mathbb{R}^4$, with the coordinate $R$ fixed. In this sense it lives in $\mathbb{R}^3$. That is, in polar coordinates we have $B$ is the product $[0,2\pi) \times \{R\}$, and so $T = A \times D = [0,2\pi) \times D \times \{R\}$, which is basically a 3D object, with a label $R$.

Finally Figure 3.8 shows two typical open sets of the cylinder product topology and the torus product topology where the open set is not a disc in $\mathbb{R}^2$ but a more general shape which is not even convex.

3.6.6.1 Infinite Products

3.6.6.1.1 The Box Topology The natural generalization of the product topology above for infinite number of spaces is known as the box topology defined next.
3.6. DERIVED TOPOLOGIES

Figure 3.6: Here $A$ is an interval, $D$ is a disk, and the cylinder is the product $A \times D$.

Figure 3.7: The torus can be thought of as the bending of a cylinder.

Definition 3.6.7. Given $X$ such that

$$X := \prod_{\alpha \in \mathcal{A}} X_{\alpha}$$

on the Cartesian product of the topological spaces $X_{\alpha}$. The box topology is generated by the base

$$B = \left\{ \prod_{\alpha \in \mathcal{A}} O_{\alpha} : O_{\alpha} \text{ is open in } X_{\alpha} \right\}.$$

If the cardinal of $\mathcal{A}$ is finite, then this definition reduces to the previous definition on finite collection of component spaces.

While this topology is natural to follow it has a few disadvantages with respect to the product topology shown below.

- **Fine size topology:** Consider, for example, the set $\mathbb{R}_d$ with the discrete topology. This set would have $2^{\aleph_0}$ open sets. Now, when we
consider an infinite box topology on this set we would generate $2^{2^{\aleph_0}}$ open sets. In general the box topology has “to many” open sets. It is a very fine topology. It is easier to work with coarse topologies.

- **Good for counterexamples:** A counterexample for lost of compactness is shown below. Then a counterexample for loss of continuity after it.

- **Loss of sequentially compactness:** We showed in section 2.4.20.2 that the Cantor set $C$ is compact, and from theorem 2.4.21 it is sequentially compact. In the $\mathbb{R}$ topology with the absolute value metric (distance) the Cantor is compact, but we can look at it in a different way.

The Cantor set is topologically homeomorphic to the countable many copies of the set $X = \{0,1\}$, where each copy carries the discrete topology. This is the space of sequences of two digits, that we can write as

$$2^{\aleph_0} = \{(x_n) : x_n \in \{0,1\} \text{ for } n \in \mathbb{N}\}. \quad (3.6.6)$$

However we are now working in $\mathbb{R}^\infty$ which is an infinite dimensional space. Here we can find that a sequence $(x_n)$ where the components of $(x_n)$ are defined as $(x_n)_i = \delta_{in}$, that is 1 for $i = n$, and 0 for $i \neq n$. The distance between any two points in this sequence is always 1. So the sequence is not sequentially compact.
• **Loss of continuity:** For example, let us recall example defined in equation 3.2.1 where we define a sequence of open sets $B_n = (-1/n, 1/n)$. We know that

\[
\bigcap_{n=1}^{\infty} B_n = \{0\}.
\]

So we have an infinite number of open sets which intersection is a closed set. This generate problems in the topology as shown in the following example. Define the following function

\[
f : \mathbb{R} \rightarrow \mathbb{R}^\infty \\
x \mapsto (x, x, x, \cdots x, \cdots)
\]

All finite components of the function are continuous. The problem happens at infinity when the finite intersection breaks since the product lengths of the component spaces $B_n$, shrink at the rate $1/n$ to 0 at $\infty$. The set $O = \prod_{n=1}^{\infty} B_n$ is by definition an open set in the box topology. Still $f^{-1}(O) = \{0\}$, since by letting $n \rightarrow \infty$, the element 0 is the only one surviving in the last component. But the singleton $\{0\}$ is not open in $\mathbb{R}$. So $f$ is not continuous.

The key point here is the pass to infinity. If only a finite set of $B_n$ would be in the product, this would not be a problem. Another important point is that the projection functions should be continuous, which they are for finite products. This motivations for the product topology defined next.

### 3.6.7.1 The Product Topology

To fix the problem of the box topology, the product topology assumes that only finite number of product spaces $X_i$ are proper subsets of $X_i$. That is,

**Definition 3.6.8 (Product Topology).** Let $\{X_\alpha\}_{\alpha \in \mathcal{A}}$ be a collection of topological spaces. We define the **product topology** on $\prod_\alpha X_\alpha$ to be the topology generated by the basis

\[
\mathcal{B} = \left\{ \prod_\alpha O_\alpha : O_\alpha \subset X_\alpha \text{ is open and } O_\alpha = X_\alpha \text{ for all but finitely many } \alpha \right\}.
\]
CHAPTER 3. BASIC TOPOLOGICAL CONCEPTS

Note here that we impose that only a finite number of spaces $X_i$ will have proper open sets. That is, open sets other than the whole space itself. This makes a big difference with the box topology where no restrictions on open sets of the components spaces are imposed.

Let us extend the definition of projection from the finite product to the infinite product.

**Definition 3.6.9.** Let $\{X_\alpha\}$ be a collection of sets indexed by a set $A$. If $\alpha \in A$, then the $\alpha$-projection map is defined as

$$P_\beta : \prod_{\alpha \in A} X_\alpha \to X_\beta$$

$$(x_\alpha) \mapsto x_\beta$$

That is, the projection $P_\beta$ drops all the components except by the component (coordinate) $x_\beta$ from the point $(x_\alpha)$. From the definition of the product topology every projection $P_\beta$ is continuous, since for any open set $O_\beta$ in $X_\beta$, the inverse projection $P^{-1}_\beta$ produces the set

$$P^{-1}(O_\beta) = \prod_{\alpha \neq \beta} X_\alpha \times O_\beta.$$

which is open in the product topology.

Let us now prove that the collection of inverse projections of open sets in the product topology are a subbasis for the product topology, that is

**Theorem 3.6.1.** The collection

$$S = \bigcup_{\alpha} \{P^{-1}(O_\beta) : O_\beta \text{ open in } X_\alpha\}$$

is a subbasis for the product topology $\prod X_\alpha$.

Let us call $T_X$ the topology for the product $\prod X_\alpha$, and $T_S$ the topology generated by $S$. We show that $T_S \subset T_X$. Let us choose an open set $B \in T_S$. The inverse projection of any open in $T_X$ (since we assume continuity of the projections). An arbitrary union of these open sets is open in $T_X$ is in $T_X$, and then a finite intersection of any open in $T_X$ is a set of $T_X$, then $B \in T_X$. Figure 3.9 shows the case of two inverse projections in $\mathbb{R}^2$. Each inverse projection is a stripe of infinite length in one direction and finite length in the other. The intersection of both inverse projections provide an open set.
in the product topology (a box). Since only a finite number of intersections is allowed (and hence the important on defining the product topology based on this finite criterion) we see that this intersection provides a set in the product topology.

Let us now show that $\mathcal{T}_X \subset \mathcal{T}_S$. Assume an element $B \in \mathcal{T}_X$. Then it is a product of the form $B = \prod_\alpha Y_\alpha$ such that $Y_\alpha = X_\alpha$ for all but a finite number of $\alpha$’s. The finite number of proper spaces $Y_\alpha$, can be seen as projections of the product topology $\mathcal{T}_X$. That is

$$P_\alpha(B) = Y_\alpha,$$

and

$$P^{-1}_\alpha(Y_\alpha) = \prod_{\alpha < \alpha_i} X_\alpha \times Y_\alpha \times \prod_{\alpha > \alpha_i} X_\alpha. \quad (3.6.8)$$

We show that

$$\bigcap_{\alpha_i} P^{-1}_\alpha(Y_\alpha) = B$$

and so $B \in \mathcal{T}_S$. Before going through these steps look again at Figure 3.9. It provides a graphical view of why this equality should be true.
(i) “⊂”: Pick \( x \in \bigcap_{\alpha_i} P_{\alpha_i}^{-1}(Y_{\alpha_i}) \). Then \( x \in P_{\alpha_i}^{-1}(Y_{\alpha_i}) \) for all \( \alpha_i \), and from equation [3.6.8] we find that by writing \( x = (x_\alpha) \), \( x_\alpha \in Y_{\alpha_i} \), for the finite amount of \( \alpha_i \)’s and \( x_\alpha \in X_\alpha \) for the rest of spaces \( X_\alpha \). This means that \( x \in B \).

(ii) “⊃”: It is clear that \( B \subset P_{\alpha_i}^{-1}(Y_{\alpha_i}) \), since here there is only one of the \( Y_{\alpha_i} \) which is proper and all the other spaces \( X_\alpha \) are the whole space. In the case of \( B \) not only includes this \( Y_{\alpha_i} \) as the proper sets but others. Then \( B \subset \bigcap_{\alpha_i} P_{\alpha_i}^{-1}(Y_{\alpha_i}) \).

As discussed in section [3.4] provided the family \( \mathcal{F} = \{ P_{\alpha \in A} \} \) where \( P_\alpha \) are the projections from the product space into their component spaces, the topology induced by forcing continuity on the projections is the coarsest topology for that set of functions. This is a general statement. The least we could do is to ensure continuity and that will induce a topological structure in the domain. In different words, we say that in the product topology \( P^{-1} \prod_{\alpha} X_\alpha = \bigcap_{\alpha} P_{\alpha}^{-1} X_\alpha \), and so continuity of the components is equivalent to say that the sets \( P^{-1} X_\alpha \) are open, which is equivalent to say that their finite intersection is open. That is, we say that the projection operator \( P^{-1} \) is continuous if and only if the componets of it are continuuos. Note that the finiteness on the number of the proper components \( Y_{\alpha_i} \) is necessary, because otherwise we can not guarantee the equality above (based only on finitely many intersections) as it is demonstrated in the counter-example under equation [3.6.7]. More generally we provide the following theorem:

**Theorem 3.6.2.** Let \( A \) be a space and \( \{ X_\alpha \}_{\alpha \in A} \) a family of spaces. Let us assume that for each \( \alpha \in A \), there is a continuous mapping \( f_\alpha : A \to X_\alpha \). Then there is a unique continuous map \( f : A \to \prod_{\alpha} X_\alpha \) such that \( P_\alpha \circ f = f_\alpha \) for all \( \alpha \in A \). Conversely, if \( f : A \to \prod_{\alpha} X_\alpha \) is continuous, then each of the maps \( f_\alpha = P_\alpha \circ f \) is continuous.

**Proof.** We do this in parts.

(i) Let us first verify that \( f_\alpha = P_\alpha \circ f \) is a well defined function. We think of the function \( f \) as a vector field function. The definition is given by

\[
\begin{align*}
f : A &\to \prod_{\alpha} X_\alpha \\
x &\mapsto (f_\alpha(x)),
\end{align*}
\]
That is, the function $f$ maps $x$ into an $\alpha$-tuple of the product space $\prod_\alpha X_\alpha$. Now, by the definition of the projection we see that $P_\alpha \circ f(x) = f_\alpha(x)$.

(ii) We now prove the continuity of $f$. Let us choose and open set in the product topology. This is $O = \prod_\alpha Y_\alpha$, where $Y_\alpha = X_\alpha$ for all but a finite number of indices $\alpha$. Then as shown above $f^{-1}(O) = f^{-1}(\prod_\alpha Y_\alpha) = \bigcap_\alpha f_\alpha^{-1}(Y_\alpha)$. Since

$$f_\alpha^{-1}(Y_\alpha) = \left(\prod_{\beta < \alpha} X_\beta\right) Y_\alpha \left(\prod_{\beta > \alpha} X_\beta\right)$$

then it is clear that the intersection in terms of proper sets is finite and open, then $f^{-1}(O)$ is open and $f$ is continuous.

(iii) If $f$ is continuous then each component $f_\alpha = P_\alpha \circ f$ is continuous because of the continuity of the function composition theorem 3.4.1 and the continuity of each projection function $P_\alpha$. In the new topology $U$ is open (by the definition of topology), and so since $f^{-1}(U)$ is open, $f$ is continuous also in this new topology $\mathcal{T}_Y'$, which is larger than $\mathcal{T}_Y$. This is a contradiction since by definition $\mathcal{T}_Y$ is the largest topology that makes $f$ continuous.

\[\square\]

As a direct application of the continuity of projections let us prove that a polynomial $p : \mathbb{R}^n \to \mathbb{R}$ is continuous.

We now proof that the sum of continuous functions is continuous. Let us start by showing that the shift by the first component operator is continuous.

$$S_x : \mathbb{R}^2 \to \mathbb{R}^2$$

$$(x, y) \mapsto (x + x, y + x)$$

Pick $(x, y) \in \mathbb{R}$ then $S_x(x) = (2x, x + y)$. An arbitrary open ball around $S_x(x)$ is $B[(2x, x + y), r]$, Now $S^{-1}[B(2x, x + y), r]] = B[(x, y), r]$, so $S_x$ is continuous.

Define

$$h : A \to \mathbb{R}^2$$

$$x \mapsto (f(x), g(x)) \quad (3.6.9)$$

$$x \mapsto (f(x), g(x)) \quad (3.6.10)$$
Let us now assume that \( f : A \to \mathbb{R} \), and \( g : A \to \mathbb{R} \) are continuous. We show that

\[
\begin{align*}
f + g : A &\to \mathbb{R} \\
x &\mapsto f(x) + g(x)
\end{align*}
\]

We observe that

\[
S_x[h(x)] = [2f(x), f(x) + g(x)]
\]

Now:

- Since \( f(x) \) and \( g(x) \) are continuous then \( h(x) \) is continuous.
- The function composition is continuous, that is \( S_x \circ h \) is continuous.
- Since \( S_x \circ h \) is continuous its components \( 2f(x) \) and \( f(x) + g(x) \) are continuous.

So \( f + g \) is continuous.

Similarly the multiplication \( kf : \mathbb{R} \to \mathbb{R} \) for any constant \( k \) is continuous, for if the constant is \( k = 0 \), then \( kf = 0 \), and the zero function is continuous (any open set in the range consist of the singleton \( \{0\} \), and \( (kf)^{-1}(0) = \mathbb{R} \). If \( k \neq 0 \), then

\[
kf : \mathbb{R}^2 \to \mathbb{R}^2 \\
(x, y) \mapsto (kx, ky)
\]

is such that an open ball in the range is of the form \( B[(kx, ky), r] \) and its inverse image \( (kf)^{-1}(B[(kx, ky), r]) \) is \( B[(x, y), r] \). So \( kf \) is continuous.

Let us now proof that the product function \( fg \) is continuous. We define the auxiliary function

\[
p_x : \mathbb{R}^2 \to \mathbb{R}^2 \\
(x, y) \mapsto (x^2, xy),
\]

which scales every point by \( x \). Pick a ball in the range. That is \( B[(x^2, xy), r] \). We consider two cases:

(i) \( x = 0 \). Then \( p_0 = 0 \), and as showing above with the case of multiplication by \( k \), the only open ball in the range is the singleton \( \{0\} \), from which inverse image is the whole space \( \mathbb{R}^2 \) which is open.
(ii) \( x \neq 0 \). Then for any ball in the range \( B[(x^2, xy), r] \), there corresponds a ball in the domain \( B[(x, y), r] = p_x^{-1}(B[(x^2, xy), r]) \). So in any case the product is continuous.

We use the same definition of \( h \) in equation 3.6.9 and then the composed function \( (p_x \circ h)(x) = p_x((h(x)) = [f(x)^2, f(x)g(x)] \), is continuous and each of its components is continuous so in particular \( f(x)g(x) \) is continuous. The proofs provided here are valid for a finite number of iterations (using induction) and replacing \( A \) by \( \mathbb{R}^n \), and \( \mathbb{R}^2 \) by \( \mathbb{R}^n \) as appropriate. That is, instead of the definitions above we provide the extensions:

\[
\begin{align*}
S_{x_1} &: \mathbb{R}^n \to \mathbb{R}^n \\
(x_i) &\mapsto (x_i + x_1), \\
h &: \mathbb{R}^n \to \mathbb{R}^n \\
(x_i) &\mapsto (f_i((x_i))), \\
kf &: \mathbb{R}^n \to \mathbb{R}^n \\
(x_i) &\mapsto k((x_i)), \\
p_{x_1} &: \mathbb{R}^n \to \mathbb{R}^n \\
(x_i) &\mapsto x_1((x_i))
\end{align*}
\]

and by applying the chain rule as done in the 2D case, we find that

\[
\sum c_if_i : \mathbb{R}^n \to \mathbb{R} \\
(x_j) \mapsto \sum_i c_if_i(x_j),
\]

and

\[
\prod f_i : \mathbb{R}^n \to \mathbb{R} \\
(x_j) \mapsto \prod_i f_i(x_j)
\]

are continuous, and in particular a polynomial from \( \mathbb{R}^n \) to \( \mathbb{R} \) which is a combinations of a finite number of sums and products.

**3.6.10 The Quotient Topology**

By “quotient” we mean several things here. We introduce in this section the following concepts:

- Quotient map,
CHAPTER 3. BASIC TOPOLOGICAL CONCEPTS

- Quotient topology, and
- Quotient space.

We provide examples and important theorems.

Given a topological space $X$ we can use continuity on a function $f$ to create a topology in a set $Y$ with the following rule.

3.6.10.1 Quotient Map and Quotient Topology

**Definition 3.6.11 (Quotient Topology).** Let $X$ be a topological space, $Y$ a set, and $f$ a mapping of $X$ onto. Then the largest topology $T_Y$ for $Y$ such that $f$ is continuous is called **quotient topology** for $Y$. Another name for the quotient topology is **identification topology**. The map $f$ is called the **quotient map**.

The following theorem is considered by many mathematicians as the definition of quotient space. We show that the sets that conform to the topology $T_Y$ are precisely those $O \subset Y$, such that $f^{-1}(O)$ is open in $X$.

**Theorem 3.6.3.** Let $f : X \to Y$ be a surjective function with $X$ a topological space and $Y$ a set. The quotient topology $T_Y$ satisfies the following equivalence relation:

$$O \text{ open in } Y \iff f^{-1}(O) \text{ is open in } X.$$  \hspace{1cm} (3.6.11)

**Proof.**

- “$\Rightarrow$” : By definition of quotient topology $f$ is continuous. That is, for any open set $O \in Y$, $f^{-1}(O)$ is open in $X$.

- “$\Leftarrow$” : We assume that $f^{-1}(U)$ is open in $X$. Let us also assume that $U$ is not open in $Y$ and get to a contradiction. If $U$ is not open in $Y$, that is, if $U \notin T_Y$, we can add $U$ to build a new topology $T_Y' = T_Y \cup \{U\} \nsubseteq T_Y$. In this new topology $T_Y'$, $f$ is continuous, because the inverse image of $U$ is open in $X$. This contradicts that $T_Y$ is the largest topology that makes $f$ continuous surjective function from $X$ onto $Y$.

So $U$ is open in $T_Y$.

The equivalence is proved.
We verify that indeed the quotient topology is a topology. Given an arbitrary family \( \{O_\alpha\}_\alpha \), we have that
\[
f^{-1}\left(\bigcup_{\alpha} O_\alpha\right) = \bigcup_{\alpha} f^{-1}(O_\alpha),
\]
and since \( \bigcup_\alpha f^{-1}(O_\alpha) \) is an open set in \( X \), then \( f^{-1}(\bigcup_\alpha O_\alpha) \) is an open set in \( X \) from which \( \bigcup_\alpha O_\alpha \) is an open set in \( Y \). The proof for finite intersection is the same changing \( \bigcup \) for \( \bigcap \). The empty set is open by vacuity and \( X = f^{-1}(Y) \), so \( Y \) is in the quotient topology.

Since \( f^{-1}(Y \setminus O) = f^{-1}(Y) \setminus f^{-1}(O) = X \setminus f^{-1}(O) \), then it is equivalent to say that \( Y \) has a quotient topology if
\[
C \text{ closed in } Y \iff f^{-1}(C) \text{ is closed in } X.
\]

The attribute “quotient map” for \( f \) is stronger than continuity. Continuity is only the “\( \Rightarrow \)” in theorem 3.6.3. We are tempted to say that the other implication means that a quotient map sends open to open and closed to closed but that is short from true. We will explain this. For that we need to develope a few concepts.

**Definition 3.6.12 (saturated).** Let \( X \) be a topological space and let \( R \) be an equivalence relation. Then \( C \subset X \) is called saturated with respecto to \( R \) if it is a union of equivalence classes.

**Definition 3.6.13 (fiber).** Let \( f : X \to Y \) be a map. The fiber of an element \( y \in Y \), is defined as
\[
f^{-1}\{y\} = \{x \in X : f(x) = y\}.
\]

We know this also with the names inverse image, or preimage or level set. For example in functions from \( \mathbb{R}^2 \to \mathbb{R} \), they are level curves (or contours), and in functions from \( \mathbb{R}^3 \to \mathbb{R} \) they are called level surfaces.

The set of fibers forms a partition of \( X \), and so any union of them forms saturated sets. For example, let us assume that we have the function \( f : \mathbb{R}^3 \to \mathbb{R} \) defined by \( f(x_1, x_2, x_3) = x_1^2 + x_2^2 + x_3^2 \). The level surfaces are spheres. Any union of those spheres forms a saturated set in \( \mathbb{R}^3 \) under \( f \). In particular any ball centered at the origin form a saturated set in \( \mathbb{R}^3 \) under \( f \).
The set of fibers forms a partition \( P \) of \( X \). Any union \( C = \cup_{\alpha} P_\alpha \) with \( P_\alpha \in \mathcal{P} \) is a saturated set for that partition. We know that given any partition \( \mathcal{P} \) we can define a projection function \( p : X \rightarrow \mathcal{P} \), which identifies any element \( x \in X \) with its class \([x] \in \mathcal{P} \). In general this is not a one-to-one function but it is surjective. For any \( y \in \mathcal{P} \), we have \( p^{-1}(y) = [y] \).

In general:

**Theorem 3.6.4 (saturated 1).** A set \( C \subset X \), is saturated with respect to the surjective map \( p : X \rightarrow Y \), if for any \( A = p^{-1}\{y\} \), such that \( A \cap C \neq \emptyset \) then \( A \subset C \).

This is, sometimes (Munkres) considered to be the definition of saturated.

**Proof.** We know that \( A = p^{-1}(y) \) is a fiber and \( X \) is a partition of fibers defined by the function \( f \).

(i) \( \implies \) : The set of fibers \( p^{-1}(y) \), for any \( y \in Y \), forms a partition of \( X \), and \( C \) is a union of partitions. \( A \) belongs to the set of partitions of \( X \), that is \( A \) is a fiber for some \( y \in Y \). So either \( A \subset C \) or \( A \setminus C = \emptyset \). So if \( A \setminus C \neq \emptyset \) then \( A \subset C \).

(ii) \( \impliedby \) : Let us assume any \( A_y = p^{-1}(y) \), such that \( A_y \cap C \neq \emptyset \). Let us call the set of \( y \)s, \( \mathcal{Y} \). Then \( A_{y \in \mathcal{Y}} \subset C \). Hence any union of fibers that has not empty intersection with \( C \), is contained in \( C \). That is we have \( \cup_{y \in \mathcal{Y}} A_y \subset C \). We now show that \( C \subset \cup_{y \in \mathcal{Y}} A_y \). Let us assume \( x \in C \). Then for any \( y \in \mathcal{Y} \), either \( x \in f^{-1}(y) \), or \( x \notin f^{-1}(y) \). If \( x \in f^{-1}(y) \) then \( x \in A_y \), and so \( x \in \cup_{y \in \mathcal{Y}} A_y \). If \( x \notin f^{-1}(y) \), then \( x \in A_y \), with \( y \notin \mathcal{Y} \), that is \( f^{-1}(y) \cap C = \emptyset \). But we assumed above \( x \in f^{-1}(y) \). so \( x \notin C \), which contradicts our hypothesis \( x \in C \).

\( \square \)

Another characterization for saturated sets is:

**Theorem 3.6.5 (saturated 2).** Let \( X \) and \( Y \) be sets and \( f : X \rightarrow Y \) be a function. A subset \( C \subset X \) is saturated with respect to \( f \) if and only if \( a \in C \) and \( f(x) = f(a) \implies x \in C \).

**Proof.**
3.6. DERIVED TOPOLOGIES

(i) “⇒” : Let us assume that \( C \) is saturated in \( X \) with respect to \( f \). Call \( y = f(x) = f(a) \). This means that \( x \in f^{-1}(y) = A_y \). We have also that \( a \in f^{-1}(y) = A_y \). If \( x \neq a \), then \( a \in C \cap A_y \), so \( A_y \) is not empty and so \( A_y \subset C \), by theorem 3.6.4, so \( x \in C \). If, on the other hand, \( x = a \), then by hypothesis \( a = x \in C \).

(ii) “⇐” : Let us assume that if \( a \in C \) and \( f(x) = f(a) \Rightarrow x \in C \), then prove that \( C \) is saturated.

Let us assume \( a \in C \), and that for each \( x \) such that \( f(x) = f(a) \), \( x \in C \). Then the fiber \( f^{-1}(y) = A_y \), for \( y = f(a) \) is such that \( A_y \subset C \). So \( C \) is the union of fibers and hence by definition is saturated.

For our final theorem on saturation, we show that \( C \) is saturated if and only \( p^{-1}p(C) = C \).

**Theorem 3.6.6 (saturated 3).** \( C \) is saturated \( \iff p^{-1}p(C) = C \).

**Proof.**

(i) “⇒” We already know from Theorem [D.3.1] that \( C \subset p^{-1}p(C) \). Let us now show that \( p^{-1}p(C) \subset C \). Let us assume \( x \in p^{-1}p(C) \). That is, there is \( y \in p(C) \), such \( y = p(x) \) for some \( x \in X \). If \( x \) is the only element such that \( y = p(x) \), that is if \( p^{-1}p(C) = \{ x \} \). Then \( C = \{ x \} \), and \( x \in C \). If there is another element \( a \neq x \), such that \( y = p(a) = p(x) \), then from theorem [3.6.5] we see that \( x \in C \). So \( p^{-1}p(C) = C \).

(ii) “⇐” The set \( p^{-1}p(C) \) is a set of \( p^{-1}B \), with \( B = p(C) \). We assume \( C \neq \emptyset \). We know that the union \( \cup\{p^{-1}(y) : y \in B\} \) is a union of fibers which are equivalence classes. By definition \( C \) is saturated if it is a union of fibers. So \( C \) is saturated.

As an interesting corollary we say that

**Corollary.** If \( p : X \rightarrow Y \) is a one-to-one function then any non-empty subset \( C \subset Y \) is saturated.

**Proof.** If \( p \) is one-to-one, then \( p^{-1}p(C) = C \) for any \( C \subset X \), and from the previous theorem \( C \) is saturated.
We now return to the equivalence 3.6.11. The forward implication “⇒” indicates that the quotient topology $\mathcal{T}_Y$, provides continuous mappings on the set $X$. Now the reverse implication “⇐” seems like the quotient topology send open sets in $X$ to open sets in $Y$. This is no true in general but we show that if $O$ is a saturated set in $X$ then $f(O)$ is a saturated set in $Y$, so the open map is valid only for saturated sets.

Let $U$ be an open saturated set in $X$ under the mapping $f$, and call $O = f(U)$. We show that $O$ is open in $Y$. Since $U$ is saturated in $f$, $U = f^{-1}(f(U))$. The reverse implication in equivalence 3.6.11 says that if $f^{-1}(O)$ is open $X$, then $O$ is open in $Y$. This proofs that $f$ is an open map under saturated sets. Similarly we can show that $f$ is a closed map under saturated sets.

We have then that if $f$ is a surjective continuous map that is either open or closed, then $f$ is a quotient map, however the opposite is not true. We show this in the following list of examples.

### 3.6.13.1 Examples: Saturated Sets and Quotient Maps

- Let us consider the function

  \[
  f : \mathbb{R}^3 \rightarrow \mathbb{R}^+ \cup \{0\}
  \]

  \[
  (x, y, z) \mapsto x^2 + y^2 + z^2.
  \]

  $f$ is continuous in the Euclidean metric topology. It is also surjective on the positive line. The fibers of $f$ at each point $r^2$ ($r \in \mathbb{R}$) are spheres of radius $r$. The saturated sets are any union of those spheres (fibers). That is the saturated sets are , balls, spherical shells, or simply a set of disjoint spheres of several radius, all centered at the origin. $f$ is also an open map. Sets other than those above are not saturated.

- Let $X = [0, 1] \cup [2, 3]$, and $Y = [0, 2]$. Consider the following function

  \[
  f : X \rightarrow Y \quad (3.6.12)
  \]

  \[
  x \mapsto \begin{cases} 
  x & \text{for } x \in [0, 1] \\
  x - 1 & \text{for } x \in [2, 3]
  \end{cases}
  \]

  Figure 3.10 shows a figure of this function.
It is interesting that while this function looks discontinuous it is actually continuous in the topological spaces considered here. The topology induced from $\mathbb{R}$ to the sets $X$ and $Y$ is the intersection of open sets in $\mathbb{R}$ with the sets $X$ and $Y$. Hence intervals such as $[0,1]$ or $[0,2]$ which are closed in the $\mathbb{R}$ topology are open in the induced topology. For example $[0,1] = (-1,2) \cap [0,1]$. The function $f$ returns open sets in the $Y$ range (observe that the function is surjective) to open sets in the domain $X$. In addition $f$ is closed since it send close sets into close sets, but it is not open. See that $[0,1]$ is open in $X$ and its image is $[0,1] \subset Y$. No the inverval $[0,1] \subset Y$ is not open since its complement $(1,2]$ is an open set in the induced topology on $Y$. We then have that $f$ is a quotient map.

We remove the point 1 from the set $X$ and define the new set as $A$. That is, let us define $A = X \setminus \{1\} = [0,1) \cup [2,3]$. Define a new map $q(x) = p(x)$, $x \in A$ (the restriction map). The new map $q$ is continuous, and surjective. Now, the set $[2,3]$ is open in $A$ and in addition it is saturated with respect to $q$, since $q^{-1}q([2,3]) = [2,3]$. Still $q[2,3] = [1,2]$, which is not open in $Y$. So the map is not a quotient map.

- Let us consider the set $X = \{0,1\}$, with the discrete topogy as $X_1$ and with the indiscrete topology as $X_2$. Then the identity function

$$i : X_1 \rightarrow X_2$$

$$x \mapsto x.$$ 

Is continuous, and surjective, but it is not open, neither closed. Any set $C \in X$ is saturated, since $i^{-1}i(C) = C$, so it is not a quotient map.
Changing the order, that is defined instead \( i : X_2 \to X_1 \) will make the function \( i \) discontinuous.

### 3.6.13.2 Quotient Space

Appendix D reviews the ideas of equivalence relations and partitions that are needed to understand this section. An equivalent relation determines a partition and vice-versa. The space \( X/\sim \) of partitions is defined as

\[
X/\sim = \{[x] : x \in X\}.
\]

We can define a mapping:

\[
q : X \to X/\sim \quad x \mapsto [x].
\]

The mapping \( q \) is surjective. Then we have the following definition

**Definition 3.6.14 (quotient space).** The mapping \( q \) defined between a space and its equivalence classes (partitions) induces a quotient topology. The name of the space \( Y = X/\sim \) which induces the quotient topology is the quotient space. Another name for the quotient space is the identification space or decomposition space.

In other words, the quotient topology on \( X/\sim \) is defined by

\[
\mathcal{T}_{X/\sim} = \{O \subset X/\sim : q^{-1}(O) \text{ is open in } X\}.
\]

In this way, a typical open set in \( \mathcal{T}_{X/\sim} \) is a collection of equivalence classes whose union is an open set in \( X \).

To better understand this concept we show a few examples.

### 3.6.14.1 Examples:

- **From a segment to a circle.** Think of the interval \( X = [0,1] \) and make the following partition.

\[
[x] = \begin{cases} 
  x & \text{if } x \neq 0 \text{ and } x \neq 1 \\
  \{0,1\} & \text{if } x = 0 \text{ or } x = 1.
\end{cases}
\]

If \( x \in (0,1) \) then an open in \( X \) is the same as an open in \( X/\sim \). The interesting case is what happens at the edges. Think first about the
point 0. This point maps to the class \([0] = \{0, 1\}\). The open around \([0]\) in the \(X/\sim\) space will be the union of two open sets in the interval \([0, 1]\). Since both points 0 and 1 map into the same class \([0]\), the open at 0 is given by \([0, \epsilon]\), for some \(\epsilon > 0\), and the open at the other end (1) is given by \((1 - \delta, 1]\). So the open sets in the new topology \(X/\sim\), are all internal open sets of \((0, 1)\), and the union \([0, \epsilon) \cup (1 - \delta, 1]\). Figure 3.11 explains the quotient mapping and the open sets in the new partition.

![Figure 3.11: Illustration of how a segment can be seen under the quotient topology as a glued circle.](image)

The segment is made into a circle with the points 0 and 1 glued at the top. The open intervals at the end of the segment (green) join into an open interval in the circle from \(\epsilon\) to \(\delta\). The interior open interval (blue) is shown in the bottom of the circle as an open interval in the new quotient topology.

To be precise the mapping can be written as

\[
\begin{align*}
f : X & \rightarrow X/\sim \\
t & \rightarrow r(\cos 2\pi t, \sin 2\pi t),
\end{align*}
\]

with \(r = (2\pi)^{-1}\). We see that 0 and 1 map both to the point \((0, 1)\), \(\epsilon\) goes to \(r(\cos 2\pi \epsilon, \sin 2\pi \epsilon)\), and likewise \(\delta\) goes to \(r(\cos 2\pi \delta, \sin 2\pi \delta)\). The mapping is one-to-one except at the point 0 and 1 where the circle is closed.

However topology is not about being precise. The exact mapping defined above is required for differential geometry or analysis but in topology all we need is to say that there is some mapping between \(X\) and \(X/\sim\) that defines a quotient topology. Any distorted circle closed loop is fine as long as there is only one intersection at the end points. Some-
thing like a number 8 will not work, but ellipses or even star shapes forms will work.

- **From a rectangle to a cylinder:** Let us go one dimension further. Instead of the interval $[0, 1]$, let us consider a rectangle $[a, b] \times [c, d]$. We can think about bending the rectangle so that the edges defined by the segments $(a, c) - (a, d)$ and $(b, c) - (b, d)$ are glued together into the single vertical line. More specifically we create the following classes, assuming $x = (x, y)$

$$
[x] = \begin{cases} 
\{x\} & \text{if } x \in (a, b) \times [c, d] \\
\{(a, y), (b, y) : c \leq y \leq d\} & \text{otherwise}
\end{cases}
$$

The mapping

$$
f : X \rightarrow X/\sim \\
x \mapsto [x]
$$

Figure 3.6.14.1 illustrates this mapping.

![Figure 3.6.14.1](image)

Figure 3.12: On the left we show a rectangle $[a, b] \times [c, d]$ (not a soccer field). The rectangle can be bent into a cylinder.

To be precise the mapping $f$ is defined as

$$
f : X \rightarrow X/\sim \\
(x, y) \mapsto r \left( \cos 2\pi \left( \frac{x - a}{b - a} \right), \sin 2\pi \left( \frac{x - a}{b - a} \right), y - c \right)
$$

The radius $r$ is such that $b - a = 2\pi r$. So $r = 2\pi / (b - a)$. In the figure $x = (a, (c + d)/2$, and $x' = (b, c + d)/2$. Then the arguments of the angles are $0$ and $2\pi$, which map into the same value $r(1, 0, (c + d)/2)$. More generally both vertical sides of the initial rectangle are mapped to $r(1, 0, y)$. As in the 1D case, an open set that was two "half"-open sets in the topology $X$ becomes a "whole"-open set in the quotient topology for $X/\sim$. 
3.6. DERIVED TOPOLOGIES

Figure 3.13: Think of a rectangle with two arrows (red) at the edges pointing to the same direction. Then glue the edges together in such a way that the arrows coincide. We left a small opening to suggest an instance of time before the gluing occurs.

Again, topology is not about being precise. Distorted cylinders are fine as long as we do not get the surface crossed onto itself.

A common way to display the gluing of objects is using arrows. We can think of a rectangle where the two edges have arrow heads pointing to the same direction. When gluing the edges together the arrows should coincide. Figure 3.13 illustrates this situation.

• The Möbius strip. Consider again a rectangle \([a, b] \times [c, d]\), with \((b - a) > (d - c)\). This time we will glue the rectangle differently. Draw two arrows, each at the side (left and right) edges of the rectangle, but with the arrow heads pointing in opposite directions. Then glue the sides so that the arrows match. Figure 3.14 sketches this.

Although topology is not about being precise we would like to find a parametrization of the Mobius strip out of curiosity. Start with the rectangle and define on it a grid. That is a Cartesian coordinate system. We think about the length as \(\theta \in [0, 2\pi r]\), where \(r\) will be explained below, and the distance to the center parameter to be \(d = [-w/2, w/2]\), where \(w\) is the width (height) of the rectangle. We consider then the origin at the left middle point as shown in Figure 3.15 indicated by the
CHAPTER 3. BASIC TOPOLOGICAL CONCEPTS

Figure 3.14: The transformation from a rectangle to the Möbius strip.

Figure 3.15: Sketch to find a parametric representation of the Möbius strip.

black dot. We refer to this figure for the derivation of the parametric
equation of the Möbius surface.

The red line in the center of the rectangle should be leveled with the
zero curve \((z = 0)\) in the Möbius strip. That is for \(z = 0\) we should
have in the \((x, y)\) coordinates \(r(\cos \theta, \sin \theta)\), where \(r\) is the radius of
the circle (red) in the figure. Note that we assume that the center of
the coordinate system for the strip is the center of the leveled (red)
circle. The black point which in the rectangle is \((0, 0)\), in the strip is
\((1, 0, 0)\). Let \(d\) be the (vertical) distance between the center line and
any arbitrary point in the rectangle. Each vertical segment (grid line)
in the rectangle will map to an straight segment in the strip but the
slope will change with respect to the vertical continuously between 0 slope (for the left edge of the rectangle in blue, see also the blue line in the strip) an π, after a complete turn of the circle. We can call this angle as the polar angle \( \varphi \). The polar angle is a function of the azimuthal angle in a way that it is 0 at \( \theta = 0 \), and \( \pi \) at \( \theta = 2\pi \). In other words we say that \( \varphi = \theta/2 \). With respect to the red circle at \( z = 0 \), any point on the surface has a distance \( d \), and a dip (polar) angle \( \varphi \).

We project this point into the radial (horizontal ) direction \( \rho \) and find that this horizontal projection is \( d\rho \cos \varphi = d(\cos \theta \cos \varphi, \sin \theta \cos \varphi) \), while the vertical projection is \( z = d \sin \varphi \).

Then with respect to the origin (center of red circle) we have:

\[
\begin{align*}
 x &= r \cos \theta + d \cos \theta \cos \frac{\theta}{2} \\
y &= r \sin \theta + d \sin \theta \cos \frac{\theta}{2} \\
z &= d \sin \frac{\theta}{2},
\end{align*}
\]

or factoring

\[
\begin{align*}
 x &= \cos \theta \left( r + d \cos \frac{\theta}{2} \right) \\
y &= \sin \theta \left( r + d \cos \frac{\theta}{2} \right) \\
z &= d \sin \frac{\theta}{2}.
\end{align*}
\]

These equations show a precise relation between the points in the rectangle, and the points in the quotient space \( X/\sim \), for the Möbius strip where the partition is defined by the equivalence classes, with \( x = (x, y) \):

\[
[x] = \begin{cases} 
(x, -y) & (2, 2\pi r) \times [-w/2, w/2] \\
[0, -y] \cup [2\pi r, -y] & [0, 2\pi r] \times [-w/2, w/2].
\end{cases}
\]

- Figure 3.7 shows how a Torus can be constructed with a cylinder. This is very similar to the first example above when we use a segment to
build a circle. Now imagine that the segment is thick like a tube. Instead of joining the two end points of the segment we are joining to disks, or cross sections of the tubes.

As before we want to obtain the parametric equation of the torus and the quotient space representation. We will use for this Figure 3.16

\[
\begin{align*}
x &= R \cos \theta + r \cos \varphi \cos \theta \\
y &= R \sin \theta + r \cos \varphi \sin \theta \\
z &= r \sin \varphi.
\end{align*}
\]

which after factoring is

\[
\begin{align*}
x &= (R + r \cos \varphi) \cos \theta \\
y &= (R + r \cos \varphi) \sin \theta \\
z &= r \sin \varphi.
\end{align*}
\] (3.6.14)

The equation in Cartesian coordinates is easy to derive by observing selecting any point \((x, y, z)\) in the surface of the torus, and dropping
3.6. DERIVED TOPOLOGIES

a perpendicular to the horizontal plane, we can build a right triangle
where the vertical leg has the size $z$, and the horizontal leg has a size $d$ which is easy to compute. We can form another triangle from the
point $(x, y, z)$ to the origin $(0, 0, 0)$, and to the base of the perpendicular
above, so the horizontal leg of this larger triangle is $R+d$. From the two
triangles and the Pythagoras theorem we can eliminate $d$ as follows.

\[
(R + d)^2 + z^2 = x^2 + y^2 + z^2 \quad \quad (3.6.15)
\]
\[
d^2 + z^2 = r^2.
\]

Figure 3.17 illustrates the cross section discussed here, with all variables
involved.

From the first equation, we have that

\[
(R + d)^2 = x^2 + y^2,
\]

that is

\[
d = \sqrt{x^2 + y^2} - R.
\]

The result can be positive or negative depending on which side of the
circle $x^2 + y^2 = R^2$ the point $(x, y, z)$ is located. However this does not
matter because the quantity will be squared. We replace this in the
second equation

Actually, this equation could have been derived without any geometrical insight and just an easy exercise on algebra and trigonometry. If
we square the first and second parametric equations in 3.6.14 and add them we find

\[ x^2 + y^2 = (R + r \cos \varphi)^2. \]

from which

\[ \sqrt{x^2 + y^2} - R = r \cos \varphi, \]

squaring both sides we find

\[ \left( \sqrt{x^2 + y^2} - R \right)^2 = r^2 \cos^2 \varphi = r^2(1 - \sin^2 \varphi) \]

but \( r \sin \varphi = z \), so we have

\[ \left( \sqrt{x^2 + y^2} - R \right)^2 = r^2 - z^2 \]

or

\[ \left( \sqrt{x^2 + y^2} - R \right)^2 + z^2 = r^2, \quad (3.6.16) \]

which is equation 3.6.13. This equation can be squared to get rid of the square root to get the four order polynomial equation:

\[ (x^2 + y^2 + z^2 + R^2 - r^2)^2 = 4R^2(x^2 + y^2). \quad (3.6.17) \]

If we compare equations 3.6.13 with 3.6.14 there is a great deal of symmetry between the two equations. If we match the following symbols

<table>
<thead>
<tr>
<th>equation 3.6.13</th>
<th>equation 3.6.14</th>
</tr>
</thead>
<tbody>
<tr>
<td>( r )</td>
<td>( R )</td>
</tr>
<tr>
<td>( d )</td>
<td>( r )</td>
</tr>
<tr>
<td>( \theta )</td>
<td>( \varphi )</td>
</tr>
<tr>
<td>( \theta/2 )</td>
<td>( \theta )</td>
</tr>
</tbody>
</table>

the equations are the same. Since the angles are eliminated we can, in exactly the same way, arrive to the corresponding to equation 3.6.15 in terms of \( r \) and \( d \), instead of \( R \) and \( r \). That is,
\[
\left( \sqrt{x^2 + y^2} - r \right)^2 + z^2 = d^2. \tag{3.6.18}
\]

The problem this time is that \( d \) is not a constant, and it depends on \( d(x, y, z) \). The parameter \( d \) can be eliminated from equations \( 3.6.13 \).

We leave this algebraic work to the reader.

It is interesting to observe that we found two different ways to address the torus as a topological entity.

(i) In section \( 3.6.4.1 \) we described the torus as the product of two sets using the product topology. The two sets where two circles, the horizontal circle with radius \( R \) and the vertical circle with radius \( r \).

(ii) Here we found the torus in terms of the bending of a rectangular surface.

In differential geometry the torus is constructed out of a circle of radius \( r \) in the vertical plane which is rotated 360 degrees.

Let us now leave algebra and geometry and return to the topological matters. We want to construct the quotient map from the rectangle to the torus.

• We now show how circle can be glued into a sphere. Start with a circle, and push the inside at the same time that the circumference is contracted into a point.
Appendix A

About the Largest Rational smaller than $\sqrt{2}$

We prove that there is not such a thing as the largest rational which is smaller than $\sqrt{2}$. As a byproduct of this proof we show how to build an infinite family of examples of monotonic increasing sequences toward $\sqrt{2}$.

Define

$$A = \mathbb{Q} \cap [0, \sqrt{2}]$$  \hspace{1cm} (A.0.1)

We prove that for any rational $p \in A$, there is a larger rational $q > p$, such that $q < \sqrt{2}$.

Let us define $q$ to be a rational function of $p$. Rational function of $p$ is

$$q = f(p) = \frac{ap + b}{cp + d}$$  \hspace{1cm} (A.0.2)

where the undetermined positive rational coefficients $a, b, c, d$ should be found.

The function $f(x)$ should satisfy two basic conditions:

(i) It should be monotonically increasing towards $\sqrt{2}$.

(ii) At some point $x < f(x)$ (that is $p < q$)

In order to satisfy the first condition we need:

$$\lim_{x \to \sqrt{2}} f(x) = \frac{a\sqrt{2} + b}{c\sqrt{2} + d} = \sqrt{2}$$  \hspace{1cm} (A.0.3)
APPENDIX A. ABOUT THE LARGEST RATIONAL SMALLER THAN $\sqrt{2}$

That is,

$$a\sqrt{2} + b = 2c + d\sqrt{2}.$$  \hfill (A.0.4)

From this equation

$$a = d$$
$$b = 2c$$ \hfill (A.0.5)

since $a, b, c, d$ are rationals.

From monotonicity:

$$f'(x) = \frac{(cx + d)a - c(ax + b)}{(cx + d)^2} = \frac{ad - bc}{(cx + d)^2}$$ \hfill (A.0.6)

and since $f'(x) > 0$, so that $f$ is monotonically increasing, we have that

$$ad > bc$$ \hfill (A.0.7)

or since $a = d$ and $b = 2c$,

$$a^2 > 2c^2.$$ \hfill (A.0.8)

For example $a = 2, b = 2, c = 1, d = 2$, which are used in the [Wikipedia](https://en.wikipedia.org/wiki/Dedekind_cut) website satisfy the constraints [A.0.5] and [A.0.8] found here. We show that the general expression

$$q = \frac{ap + 2c}{cp + a}$$ \hfill (A.0.9)

where $a$ and $c$ satisfy [A.0.8] provides $q > p$.

Let us show that $q^2 - 2 < 0$ (that is $q < \sqrt{2}$) and that $q > p$. Obviously $q$ is rational from the construction, given that $a$ and $c$ are rational numbers
as well as \( p \). Now,

\[
q^2 - 2 = \frac{(ap + 2c)^2}{(cp + a)^2} - 2
\]

\[
= \frac{a^2p^2 + 4apc + 4c^2}{(cp + a)^2} - 2
\]

\[
= \frac{a^2p^2 + 4apc + 4c^2 - 2c^2p^2 - 4c^2ap - 2a^2}{(cp + a)^2}
\]

\[
= \frac{a^2p^2 + 4c^2 - 2c^2p^2 - 2a^2}{(cp + a)^2}
\]

\[
= \frac{p^2(a^2 - 2c^2) - 2(a^2 - 2c^2)}{(cp + a)^2}
\]

\[
= \frac{(p^2 - 2)(a^2 - 2c^2)}{(cp + a)^2}
\]

(A.0.10)

and since \( a^2 > 2c^2 \) (from condition A.0.8) and \( p < \sqrt{2} \) we conclude that

\[
q^2 - 2 < 0,
\]

(A.0.11)

that is \( q < \sqrt{2} \).

Now we show that \( q > p \). From A.0.9

\[
q = p + \frac{ap + 2c}{cp + a} - p
\]

\[
= p + \frac{2ap + 2c - cp^2 - ap}{cp + a}
\]

\[
= p + \frac{c(2 - p^2)}{cp + a}
\]

(A.0.12)

and since \( c > 0, a > 0, \) and \( p > 0, \) and \( 2 - p^2 > 0 \), we find that \( q > p \). In a way this proof the infinite density of the rational numbers. They can squeeze toward \( \sqrt{2} \) as much as we want, never to “touch” \( \sqrt{2} \).
APPENDIX A. ABOUT THE LARGEST RATIONAL SMALLER THAN $\sqrt{2}$
Appendix B

Important Inequalities

We proof a few inequalities that are used in this document.

B.1 Convex Functions

Let $I \subset \mathbb{R}$ be an interval. A function $f : I \to \mathbb{R}$ is called convex if

$$f(\lambda x + (1 - \lambda)y) \leq \lambda f(x) + (1 - \lambda)f(y).$$  \hfill (B.1.1)

for all $x, y \in I$ and all $\lambda \in [0, 1]$.

It is interesting to observe that if the function $f$ is linear then

$$f(\lambda x + (1 - \lambda)y) = \lambda f(x) + (1 - \lambda)f(y).$$

which is always convex. But a concave (or convex down) function satisfies the inequality $f(\lambda x + (1 - \lambda)y) \geq \lambda f(x) + (1 - \lambda)f(y)$. So a linear function is simultaneously convex and concave. A function $f$ is convex if and only its negative $-f$ is concave.

Figure [B.1] shows an illustration of a convex function.

The interesting functions in this area are functions which satisfy strictly inequality. That is, they have curvature. and we prove first that for a convex function, the second derivative is always greater or equal to 0 (equal if it is linear), or in symbols $f''(z) \geq 0$, $\forall z \in I$. The proof of this is easy. Pick

$$x = z + h \quad y = z - h \quad \lambda = 1/2$$

in equation [B.1.1] and find

$$f((z + h)/2 + (z - h)/2) \leq f(z + h)/2 + f(z - h)/2$$
that is
\[ f(z + h) - 2f(z) + f(z - h) \geq 0 \]
and dividing by \( 2h \)
\[ \frac{f(z + h) - 2f(z) + f(z - h)}{2h} \geq 0 \]
we take the limit as \( h \to 0 \) and find
\[ f''(z) \geq 0. \]

Since the second derivative of \(-\log x\) is \(1/x^2\) then \(-\log x\) is convex in the interval \((0, \infty)\). We now prove the reverse. That is if \( f''(x) \geq 0 \) for all \( x \in I \) where \( I \) is an interval over the real numbers, then \( f \) is convex in \( I \). Let us first prove that if \( f'(x) \) is monotonically increasing. From the mean value theorem, pick \( x < y \) in \( I \) we know that there is a \( c \in [x, y] \) such that
\[ f''(c) = \frac{f'(y) - f'(x)}{y - x} \]
and since \( f''(c) \geq 0 \), and \( y > x \) then
\[ f'(y) \geq f'(x). \]
So \( f' \) is monotonically increasing. Now, let us choose \( s = \lambda x + (1 - \lambda)y \). Then \( x < s < y \). By the mean value theorem exists \( c_1, x < c_1 < s \) and, \( c_2, \)
$s < c_2 < y$ such that
\[ f'(c_1) = \frac{f(s) - f(x)}{s - x} \quad f'(c_2) = \frac{f(y) - f(s)}{y - s}, \]
since $f'$ is monotonically increasing and $c_1 \leq c_2$ then
\[ \frac{f(s) - f(x)}{s - x} \leq \frac{f(y) - f(s)}{y - s} \]
\[ \frac{f[\lambda x + (1 - \lambda)y] - f(x)}{\lambda x + (1 - \lambda)y - x} \leq \frac{f(y) - f(\lambda x + (1 - \lambda)y)}{y - (\lambda x + (1 - \lambda)y)} \]
\[ \frac{f[\lambda x + (1 - \lambda)y] - f(x)}{(1 - \lambda)(y - x)} \leq \frac{f(y) - f(\lambda x + (1 - \lambda)y)}{\lambda(y - x)} \]
\[ \lambda f[\lambda x + (1 - \lambda)y] - \lambda f(x) \leq (1 - \lambda)f(y) - f[\lambda x + (1 - \lambda)y] + \lambda f[\lambda x + (1 - \lambda)y] \]
\[ f[\lambda x + (1 - \lambda)y] \leq \lambda f(x) + (1 - \lambda)f(y). \]

So $f$ is convex.

Next, we show a variation of [Jensen’s inequality][1] for a discrete set.

### B.1.1 Jensen’s Inequality

Suppose $I \in \mathbb{R}$ is an interval and $h : I \rightarrow \mathbb{R}$ is convex. Then for all $n \in \mathbb{N}$, all $\lambda_i \geq 0$, $i = 1, \cdots, n$ and $\sum_{j=1}^{n} \lambda_j = 1$, and all $z_i \in I$,
\[ h \left( \sum_{j=1}^{n} \lambda_j z_j \right) \leq \sum_{j=1}^{n} \lambda_j h(z_j) \]  
(B.1.2)

That is, the function is “under-linear” for more than just two points. It is like the definition of convexity extended to many points.

We use induction. For $n = 1$ it is trivial that $h(z_1) \leq h(z_1)$. Let us assume that up to some $n \geq 2$ the inequality is valid. Then we want to evaluate
\[ h \left( \sum_{j=1}^{n+1} \lambda_j z_j \right). \]

The idea is to split the sum in two sets of points each with less than $n + 1$ elements. By assumption
\[ \sum_{i=1}^{n+1} \lambda_i = 1 \]

If any of the $\lambda_i$ is equal to 1, all others have to be equal to 0 because $\lambda_i \geq 0$, $i = 1, \ldots, n + 1$. So, necessarily $\lambda_i < 1$, for all $i$’s. Let us separate $\lambda_1$ from the rest of $\lambda_i$’s and write

$$\sum_{j=1}^{n+1} \lambda_j z_j = \lambda_1 z_1 + \sum_{j=2}^{n+1} \lambda_j z_j$$

Since $z_1$ has already a coefficient $\lambda_1$, we want to write the second sum as some number $u$ with coefficient $1 - \lambda_1$. That is, we call

$$(1 - \lambda_1)u = \sum_{j=2}^{n+1} \lambda_j z_j,$$

or define

$$u = \frac{\sum_{j=2}^{n+1} \lambda_j z_j}{1 - \lambda_1} = \sum_{j=2}^{n+1} \frac{\lambda_j z_j}{1 - \lambda_1}.$$ 

We obtained what we wanted. That is,

$$\sum_{j=1}^{n+1} \lambda_j z_j = \lambda_1 z_1 + (1 - \lambda)u$$

and we can apply the convexity of $h$ to these two terms to obtain

$$h\left(\sum_{j=1}^{n+1} \lambda_j z_j\right) = h(\lambda_1 z_1 + (1 - \lambda_1)u) \leq \lambda_1 h(z_1) + (1 - \lambda_1)h(u).$$

Now we should expand $h(u)$. We note that the coefficients of $u$ add to 1, because

$$\sum_{i=1}^{n+1} \lambda_i = 1 \quad \Rightarrow \quad \lambda_1 = 1 - \sum_{i=2}^{n+1} \lambda_i \quad \Rightarrow \quad \sum_{i=2}^{n+1} \lambda_i = 1 - \lambda_1$$

and so

$$\sum_{i=2}^{n+1} \frac{\lambda_i}{1 - \lambda_1} = \frac{1 - \lambda_1}{1 - \lambda_1} = 1,$$

and then by the induction hypothesis

$$h(u) \leq \sum_{i=2}^{n+1} \frac{\lambda_i}{1 - \lambda_1} h(z_i).$$
and finally
\[ h \left( \sum_{j=1}^{n+1} \lambda_j z_j \right) = h(\lambda_1 z_1) + (1 - \lambda_1)u \leq \lambda_1 h(z_1) + \frac{1 - \lambda_1}{\lambda_1} \sum_{j=2}^{n+1} \lambda_i h(z_i), \]
that is
\[ h \left( \sum_{j=1}^{n+1} \lambda_j z_j \right) \leq \sum_{j=1}^{n+1} \lambda_i h(z_i), \]
which proves the theorem.

The Jensen’s inequality is useful to prove other inequalities. This is the case of the following inequality.

B.1.2 Weighted geometric/arithmetic mean inequality

Suppose \( \sum_{j=1}^{n} \lambda_j a_j \) is a convex combination of non-negative numbers \( a_1, \cdots, a_n \). Then
\[ a_1^{\lambda_1} a_2^{\lambda_2} \cdots a_n^{\lambda_n} \leq \sum_{j=1}^{n} \lambda_j a_j. \quad (B.1.3) \]

What this is telling us is that the weighted geometric mean is smaller than the weighted arithmetic mean. We assume that \( 0^0 = 0 \) by definition, so if one of the \( a_i = 0 \), then the inequality is obvious and so we assume that all \( a_i > 0 \).
We use Jensen’s inequality applied to the convex function \( h(x) = -\log x \), on the interval \( I = (0, \infty) \). That is
\[ -\log \left( \sum_{j=1}^{n} \lambda_j a_j \right) \leq - \sum_{j=1}^{n} \lambda_j \log a_j = - \log \left( a_1^{\lambda_1} \cdots a_n^{\lambda_n} \right) \]
and so, multiplying by \(-1\) and taking the exponential function we observe the inequality. In particular when each \( \lambda_i = 1/n \), we find
\[ (a_1 a_2 \cdots a_n)^{\frac{1}{n}} \leq \frac{1}{n} \sum_{j=1}^{n} a_j \quad (B.1.4) \]
which is the classical geometric/arithmetic mean inequality.

\(^2\text{by convex combination we mean a weighted average where the weights } \lambda_j \text{ add to 1.}\)
B.1.3 Young’s inequality

We define Hölder’s conjugates to any pair of numbers \( p, q \in [1, \infty] \) such that \( 1/p + 1/q = 1 \). Young’s inequality is a direct consequence of the Weighted geometric/arithmetic mean inequality.

Let us assume \( p, q \) to be Hölder’s conjugates. Then, for any nonnegative numbers \( a, b \)

\[
ab \leq \frac{a^p}{p} + \frac{b^q}{q}.
\]  

(B.1.5)

We choose, in equation B.1.3, \( \lambda_1 = 1/p \) and \( \lambda_2 = 1/q \), and \( a_1 = a^p \), and \( a_2 = b^q \). and equation B.1.5 follows.

B.1.4 Hölder’s inequality

Hölder’s inequality is a generalization of the Cauchy-Schwartz inequality \(^3\). According to Wikipedia \(^4\) Hölder’s inequality was discovered first by Rogers, L. J. and published in 1888, while Hölder’s publication dates 1889.

We proof the Hölder’s inequality for both, the discrete and continuous case. That is, for sums and integrals.

B.1.4.1 Discrete

We prove Hölder’s inequality for finite dimensional spaces. However please observe that changing \( n \) by \( \infty \) does not change the proof below, then Hölder’s inequality is also valid in the spaces of sequences \( \ell_p \).

For \( \mathbf{x}, \mathbf{y} \in \mathbb{C}^n \) and \( 1 \leq p \leq \infty \)

\[
|\mathbf{x} \cdot \mathbf{y}| := \sum_{j=1}^{n} |x_j y_j| \leq \|\mathbf{x}\|_p \|\mathbf{y}\|_q, \quad \text{where} \quad \frac{1}{p} + \frac{1}{q} = 1.
\]  

(B.1.6)

For any \( 1 \leq p \leq \infty \) the norm \( \|\cdot\|_p \) is defined as

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

\(^3\)https://en.wikipedia.org/wiki/Cauchy%E2%80%93Schwarz_inequality

\(^4\)https://en.wikipedia.org/wiki/H%C3%B6lder%27s_inequality
and

\[ \|x\|_\infty = \max_{i=1}^n |x_i|. \]

When \( p = 2 \), by default nothing is written as sub-index. That is \( \|\cdot\|_2 = \|\cdot\| \).

Let us first prove this inequalities for the extrema values \( p = 1 \) and \( p = \infty \).

If \( p = 1 \), then \( q = \infty \) we have

\[ \sum_{j=1}^n |x_j y_j| \leq \max_{i=1}^n |y_i| \sum_{j=1}^n |x_j| = \|x\|_1 \|y\|_\infty. \]

The case \( p = \infty \) and \( q = 1 \) is the equivalent to this (just change \( x \) by \( y \) and being those vectors arbitrary we do not have to repeat that proof here.

Let us then assume that \( 1 < p < \infty \). If one of or both vectors \( x \) and \( y \) are zero the inequality turns into the equality \( 0 \leq 0 \). So we assume that both \( x \) and \( y \) are non-zero vectors.

For any \( a_j, b_j \geq 0 \) we use Young’s inequality \[ \text{B.1.5} \]. That is,

\[ a_j b_j \leq \frac{a_j^p}{p} + \frac{b_j^q}{q}. \]

We set

\[ a_j = \frac{|x_j|}{\|x\|_p} \quad b_j = \frac{|y_j|}{\|y\|_q} \]

and so

\[ \left( \frac{|x_j|^p}{\|x\|_p^p} \right)^{\frac{1}{p}} \left( \frac{|y_j|^q}{\|y\|_q^q} \right)^{\frac{1}{q}} \leq \frac{1}{p} \|x\|_p^p + \frac{1}{q} \|y\|_q^q \]

We add over \( j = 1, \cdots n \), and find

\[ \sum_{i=1}^n \left( \frac{|x_i|^p}{\|x\|_p^p} \right)^{\frac{1}{p}} \left( \frac{|y_i|^q}{\|y\|_q^q} \right)^{\frac{1}{q}} \leq \sum_{i=1}^n \frac{1}{p} \|x\|_p^p + \frac{1}{q} \|y\|_q^q. \]

We write the left sum as

\[ \sum_{i=1}^n \left( \frac{|x_i|^p}{\|x\|_p^p} \right)^{\frac{1}{p}} \left( \frac{|y_i|^q}{\|y\|_q^q} \right)^{\frac{1}{q}} = \frac{1}{\|x\|_p \|y\|_q} \sum_{i=1}^n |x_i| |y_i|. \]
and the right side as

\[ \sum_{i=1}^{n} \frac{1}{p} \frac{|x_j|^p}{\|x\|_p^p} + \frac{1}{q} \frac{|y_j|^q}{\|y\|_q^q} = \frac{1}{p} \sum_{j=1}^{n} |x_j|^p + \frac{1}{q} \sum_{j=1}^{n} |y_j|^q \frac{1}{p} + \frac{1}{q} = 1 \]

and so, we showed that

\[ \frac{1}{\|x\|_p \|y\|_q} \sum_{i=1}^{n} |x_i||y_j| \leq 1, \]

and from here

\[ \sum_{i=1}^{n} |x_j||y_j| \leq \|x\|_p \|y\|_q \]

which is the Hölder’s inequality. If \( p = q = 2 \) we find the famous

B.1.4.1.1 Cauchy-Schwartz inequality

\[ |x \cdot y| := \sum_{i=1}^{n} |x_i||y_j| \leq \|x\| \|y\| \]

B.1.4.2 Continuous

In the continuous spaces we define the inner product of two functions, which are integrable in some domain as

\[ f \cdot g := \int f g d\mu \]

where \( \mu \) is the Lebesgue measure. The norm is defined as

\[ \|f\|_p := \left( \int |f|^p d\mu \right)^{1/p}, \]

whenever the integral exists. Hölder’s inequality dictates that

\[ |f \cdot g| \leq \|f\|_p \|g\|_q. \]

\(^5\) we ignore the integration domain since it is not relevant for the proofs, as long as the functions have integrals in that domain.
Note that this is a general inequality that also covers the discrete case above, and that for \( p = q = 1/2 \) is the Cauchy’s inequality. In our particular case of integrals we have that

\[
|f \cdot g| = \left| \int f \, g \, d\mu \right| \leq \int |f \, g| \, d\mu
\]

so it suffices to prove

\[
\int |f \, g| \, d\mu \leq \left( \int |f|^p \, d\mu \right)^{1/p} \left( \int |g|^q \, d\mu \right)^{1/q}
\]

(B.1.7)

which will do next.

**B.1.4.2.1 Proof:** If the norm of any \( f \) or \( g \) is zero, then from the definition of norm, the corresponding \( f \) or \( g \) is zero, and so the inequality is the trivial equality \( 0 = 0 \). So we assume that \( f, g \neq 0 \). Let us do the special case \( p = 1, q = \infty \). This is

\[
\int |f \, g| \, d\mu \leq \int |f| \, |g| \, d\mu
\]

\[
\leq \left( \int |f| \, d\mu \right) \left( \sup \int |g| \, d\mu \right)
\]

\[
\leq \|f\|_1 \|g\|_{\infty}
\]

where “sup” means the supremum value over the domain of integration up to a set of measure zero. That is, we ignore the values of \( g(x) \) in any set of measure 0, and take the sup of the left values. We assume that the function is bounded on this set. Hence we will assume in what follows that \( p > 1 \) and \( q > 1 \) (the argument for \( q = 1 \) is symmetric to the argument for \( p = 1 \) and so we could rephrase it in the same way).

It is easier to prove the inequality for normalized functions. That is, since we can not assume that \( \|f\|_p \neq 0 \), and \( \|g\|_q \neq 0 \), we could divide by these scalars and prove that

\[
|f \cdot g| \leq 1
\]

where \( \|f\|_p = \|g\|_q = 1 \). Similarly to what we did above in the discrete case, and using Young’s inequality by choosing

\[
a = |f| \quad \text{and} \quad b = |g|
\]
APPENDIX B. IMPORTANT INEQUALITIES

we find that everywhere except for a set of measure zero.

\[ |f||g| = ab \leq \frac{a^p}{p} + \frac{b^q}{q} = \frac{|f|^p}{p} + \frac{|g|^q}{q} \]

and after integrating both sides

\[ |f \cdot g| \leq \frac{1}{p} + \frac{1}{q} = 1 \]

That is, since \(f\) and \(g\) where normalized we have that would not \(f\) and \(g\) been normalized, we get.

\[ \frac{f}{\|f\|_p} \cdot \frac{g}{\|g\|_q} \leq 1 \]

or

\[ \|f \cdot g\| \leq \|f\|_p \|g\|_q. \]

B.1.5 Minkowski’s inequality

We proof the Minkowski’s inequality for both, the discrete and continuous case. That is, for sums and integrals.

B.1.5.1 Discrete

We prove Minkowski’s inequality for finite dimensional spaces. However please observe that changing \(n\) by \(\infty\) does not change the proof below, then Minkowski’s inequality is also valid in the spaces of sequences \(\ell_p\).

Let \(\mathbf{x}, \mathbf{y} \in \mathbb{C}^n\) and \(1 \leq p \leq \infty\). Then

\[ \|\mathbf{x} + \mathbf{y}\|_p \leq \|\mathbf{x}\|_p + \|\mathbf{y}\|_p. \]  

(B.1.8)

Let us start with \(p = 1\) here

\[ \|\mathbf{x} + \mathbf{y}\|_1 = \sum_{i=1}^{n} |x_i + y_i| \leq \sum_{i=1}^{n} |x_i| + \sum_{i=1}^{n} |y_i| = \|\mathbf{x}\|_1 + \|\mathbf{y}\|_1. \]

and if \(p = \infty\)

\[ \|\mathbf{x} + \mathbf{y}\|_\infty = \max_{i=1}^{n} |x_i + y_i| \leq \max_{i=1}^{n} |x_i| + \max_{i=1}^{n} |y_i| = \|\mathbf{x}\|_\infty + \|\mathbf{y}\|_\infty. \]
B.1. CONVEX FUNCTIONS

Let us now assume that $1 < p < \infty$. We write

$$
\|x + y\|_p^p = \sum_{j=1}^{n} |x_j + y_j|^p \\
= \sum_{j=1}^{n} |x_j + y_j||x_j + y_j|^{p-1} \\
\leq \sum_{j=1}^{n} |x_j||x_j + y_j|^{p-1} + \sum_{j=1}^{n} |y_j||x_j + y_j|^{p-1}.
$$

(B.1.9)

We apply the Hölder’s inequality to each of the terms above. That is

$$
\sum_{j=1}^{n} |x_j||x_j + y_j|^{p-1} \leq \|x\|_p \left( \sum_{j=1}^{n} |x_j + y_j|^{(p-1)q} \right)^{1/q}
$$

and from

$$
\frac{1}{p} + \frac{1}{q} = 1 \quad \Rightarrow \quad (p - 1)q = p,
$$

we see that

$$
\sum_{j=1}^{n} |x_j||x_j + y_j|^{p-1} \leq \|x\|_p \|x + y\|_p^{p/q} = \|x\|_p \|x + y\|_p^{p-1} \quad \text{(B.1.10)}
$$

Similarly

$$
\sum_{j=1}^{n} |y_j||x_j + y_j|^{p-1} \leq \|y\| \|x + y\|_p^{p/q} \|y\| \|x + y\|_p^{p-1}.
$$

(B.1.11)

Combining B.1.9, B.1.10, and B.1.11, we find

$$
\|x + y\|_p^p \leq (\|x\|_p + \|y\|_p)\|x + y\|_p^{p-1}
$$

which after dividing by $\|x + y\|_p^{p-1}$ is

$$
\|x + y\|_p \leq \|x\|_p + \|y\|_p.
$$

This is the triangular inequality for vectors in $\mathbb{R}^n$. Together with the homogeneity and no-negativity properties of $\|\cdot\|_p$, this proves that the $p$-norm is actually a norm in the space $\mathbb{R}$. 
Appendix B. Important Inequalities

B.1.5.2 Continuous

We show that

$$\|f + g\|_p \leq \|f\|_p + \|g\|_p.$$  

We start by showing that $\|\cdot\|_p$ is closed under the sum. That is, if $f$ and $g$ have a $\|\cdot\|_p$ which is finite, their sum $f + g$ has a $\|\cdot\|_p$ which is also finite. We should found a bound $M$ such that

$$\|f + g\| \leq M(\|f\| + \|g\|).$$

For this we use Jensen’s inequality [B.1.2] for the convex function $h(x) = x^p$ (where we note that $p > 1$, otherwise the function is no convex), with $\lambda_1 = \lambda_2 = 1/2$, $z_1 = f$, and $z_2 = g$. That is

$$\left|\frac{f}{2} + \frac{g}{2}\right|^p \leq \frac{1}{2}|f|^p + \frac{1}{2}|g|^p.$$

and so

$$|f + g|^p \leq 2^{p-1}(|f|^p + |g|^p).$$

We can integrate both sides and find that $|f + g|$ is bounded by the norms of $f$ and $g$ added and multiplied by a constant. Taking a power $1/p$ will not change finiteness conditions.

We now prove the Minkowski inequality.

$$\|f + g\|_p^p := \int |f + g|^p d\mu$$

$$= \int |f + g||f + g|^{p-1} d\mu$$

$$\leq \int (|f| + |g|)|f + g|^{p-1} d\mu$$

$$= \int |f||f + g|^{p-1} d\mu + \int |g||f + g|^{p-1} d\mu.$$

We now use Hölder’s inequality [B.1.7] From the first term, mapping $|f| \rightarrow f$, and $|f + g|^{p-1} \rightarrow g$, we find

$$\int |f||f + g|^{p-1} d\mu \leq \left(\int |f|^p d\mu\right)^{1/p} \left(\int |f + g|^{q(p-1)} d\mu\right)^{1/q}.$$
B.1. CONVEX FUNCTIONS

Likewise for the second term, mapping \(|g| \rightarrow f\) and \(|f + g|^{p-1} \rightarrow g\), we find

\[
\int |g| |f + g|^{p-1} \, d\mu \leq \left( \int |g|^p d\mu \right)^{1/p} \left( \int |f + g|^{q(p-1)} d\mu \right)^{1/q}
\]

Then after adding we find

\[
\|f + g\|_p^p \leq \left( \left( \int |f|^p d\mu \right)^{1/p} + \left( \int |g|^p d\mu \right)^{1/p} \right) \left( \int |f + g|^{q(p-1)} d\mu \right)^{1/q}.
\]

Since \(p, q\) are Hölder’s conjugate,

\[
\frac{1}{p} + \frac{1}{q} = 1 \quad \Rightarrow \quad \frac{1}{q} = 1 - \frac{1}{p} = \frac{p - 1}{p} \quad \Rightarrow \quad q(p - 1) = p
\]

Then we write

\[
\|f + g\|_p^p \leq \left( \left( \int |f|^p d\mu \right)^{1/p} + \left( \int |g|^p d\mu \right)^{1/p} \right) \left( \int |f + g|^{p-1} d\mu \right)^{1-1/p}
\]

\[
= \left( \|f\|_p + \|g\|_q \right) \frac{\|f + g\|_p^p}{\|f + g\|_p},
\]

from which Minkowski inequality

\[
\|f + g\|_p \leq \|f\|_p + \|g\|_p
\]

falls.
Appendix C

Notes in Real Analysis

The main objective of this appendix is to prove the (Cauchy) completeness of the real numbers set \( \mathbb{R} \). For that we need to define a few terms related to bounds of a set, supremum, infimum, minimum, maximum, etc. In the way we prove the Bolzano-Weirstrass theorem, which is fundamental for analysis, and introduce two axioms. The axiom of completeness and the Arquemedian property of the real numbers.

C.1 Basic Definitions and Theorems on Bounded Sets

We start with the definitions of upper and lower bounds. Since we are only interested on the real numbers we assume, for the following definitions, that the context is that of real numbers. The definition could be more general but we want to simplify.

**Definition C.1.1.** A real number \( \alpha < \infty \) is called an upper bound for a set \( B \), if \( \alpha \geq x \ \forall x \in B \). A real number \( \beta > -\infty \) is called an lower bound for a set \( B \), if \( \beta \leq x \ \forall x \in B \).

Next we show the definitions of supremum, infimum, maximum, and minimum.

**Definition C.1.2.** Assume that the set \( B \subset R \) is bounded above. The supremum is the least upper bound. It is call lub or sup(A).
Definition C.1.3. Assume that the set \( B \subset \mathbb{R} \) is bounded below. The infimum is the greatest lower bound. It is called glb or \( \inf(A) \).

Examples:

- For \( A = (a, b) \), \( \sup(A) = b \) \( \inf(A) = a \). Any number \( x > b \) is an upper bound of \( A \); any number \( x < a \) is a lower bound of \( A \).

- For \( B = \{ x \in \mathbb{I} : x < \sqrt{2} \} \), \( \sup(B) = \sqrt{2} \). The set \( B \) does not have a lower bound, so it does not have an infimum. Any number \( x \geq \sqrt{2} \) is an upper bound of \( B \).

- For \( C = \{ x \in \mathbb{R} : x \leq \sqrt{2} \} \), \( \sup(C) = \sqrt{2} \). The set \( C \) does not have a lower bound, so it does not have an infimum. Any number \( x \geq \sqrt{2} \) is an upper bound for \( C \); the set \( C \) does not have a lower bound.

- For \( D = \{ x \in \mathbb{Q} : x > \sqrt{2} \} \), \( \inf(D) = \sqrt{2} \). The set \( D \) does not have an upper bound, so it does not have an supremum. Any number \( x \leq \sqrt{2} \) is a lower bound of \( D \).

Definition C.1.4. A set \( B \) is said to be bounded above if it has an upper bound, \( A \) set \( B \) is said to be bounded below if it has a lower bound, \( A \) set \( B \) is said to be bounded if it has both an upper and lower bounds.

We now define maximum and minimum.

Definition C.1.5. A set \( B \) has a maximum element \( a = \max(A) \) if \( a \in A \), and \( a = \sup(A) \). A set \( B \) has a minimum element \( a = \min(A) \) if \( a \in A \), and \( a = \inf(A) \).

In the examples above. \( A \) does not have a minimum, neither a maximum, \( B \) does not have a minimum, neither a maximum, \( C \) has a maximum \( \max C = \sqrt{2} \), but it does not have a minimum, and finally \( D \) does not have a minimum, neither a maximum.

We now present the axiom of completeness of the real line.

Axiom 1 (Completeness). Every nonempty set \( B \subset \mathbb{R} \) bounded above has a supremum.

We now show a few theorems related to the concepts above.
Theorem C.1.6. Assume that $A$ is bounded above. The supremum $s = \sup A$ is such that for any $\epsilon > 0$, there is an $x \in A$, with $x > s - \epsilon$.

Proof. Assume the contrary. That is, assume that for all $x \in A$, $x \leq s - \epsilon$. Since $\epsilon > 0$, the number $s - \epsilon < s$ is a bound of $A$. This is impossible since $s$ is the least upper bound.

Theorem C.1.7. Assume that $A$ is bounded below. The infimum $i = \inf A$ is such that for any $\epsilon > 0$, there is an $x \in A$, with $x < s + \epsilon$.

The proof is similar to the previous for the supremum. It is left to the reader.

Next theorem should be intuitively clear by doing a mirror reflection from a set in the real line with respect to 0.

Theorem C.1.8. If $A$ is bounded above, then $\sup(A) = -\inf(-A)$, and $\inf(A) = -\sup(-A)$. If there is a max and min they satisfy similar properties.

The proofs of this is simple and we leave it to the reader. This last properties are important we we want to proof things only, let us say, on the sup and not on the inf. Then when we need to prove something on the inf, we reverse the set and prove what we want on the sup, and then reverse back to get what we want. A good exercise is to prove Theorem C.1.7 by using the reflection properties above.

C.2 Completeness of The Reals $\mathbb{R}$

C.2.1 Historical Introduction

In 1817 Bolzano proved that a bounded infinite sequence has to have a limit (accumulation point). The proof of this is based on the method of bisection explained in example 2.4.9.1. It can be seen in multiple dimensions by applying the one dimension theorem to each dimension and selecting from there a common subsequence. Think without loss in generality that your sequence is bounded about a closed interval $[a, b]$. If you step in the middle of this interval $((a + b)/2)$, and look at both halves, at least one half has to have an infinite number of points. You choose that half, and do the same. That is, step into the middle of the half interval with infinite number of points, and
again, one of the two new half intervals of the current interval should have at least infinite number of points. If you keep in this process your intervals of search are reducing at the speed $2^{-n}$. So as the thickness of the interval converges to zero exponentially to 0, the points squeezed there should converge to a limit. That is, for any $\epsilon > 0$, after some $N > |\log_2 \epsilon|$ and for any $n > N$, you are within the limits of your error $\epsilon$. Fifty years after Bolzano’s proof, Weirstrass retook the problem. This became the famous [Bolzano-Weierstrass theorem](https://en.wikipedia.org/wiki/Bolzano%E2%80%93Weierstrass_theorem) The Bolzano-Weirstrass theorem could be used to show the completeness of the real line. For more information about the construction of the real numbers $\mathbb{R}$, see section 2.4.7.

### C.2.2 The Bolzano-Weirstrass Theorem

The Bolzano-Weirstrass theorem indicates that given an infinite sequence of points in a bounded interval, there is a subsequence of those points which converges to some limit.

The theorem indicates that we have an infinite sequence of points in a bounded set $A \subset \mathbb{R}^n$ (or $A \subset \mathbb{C}^n$) then there should be a subsequence convergent to some limit. The idea is that since a cluster of infinite points could have many accumulation (limit) points, a method to find one such point should be shown. One method to find a convergence subsequence is that proposed by Bolzano based on bisection. Weirstrass propose a different algorithm selecting an increasing or decreasing sequence within the set. This motivates the introduction of the following theorem known as the monotone convergence Theorem

**Theorem C.2.1.** Assume a real sequence $(x_n)$, which is bounded and monotone then there is a point $x$ such that $\lim_{n \to \infty} x_n = x$

**Proof.** Let us assume first that the sequence is increasing. Since the set $A = \{x_n\}$ is bounded, by the Completeness Axiom it has a sup. We show that the sup is the limit. Let $s = \sup \{x_n\}$. Pick $\epsilon > 0$. By Theorem C.1.6 there is an $x_N \in A$, such that $x_N > s - \epsilon$, and since the sequence is monotonically increasing for any $n > N$, $x_n > s - \epsilon$ as well. Rearranging $s - x_n < \epsilon$, and since $s \geq x_n$, for all $x_n \in A$ (definition of sup) then $s - x_n = |s - x_n|$ and we proved that

$$|x_n - s| < \epsilon \quad \forall n > N$$

---

which according to the definition of the limit of a sequence $\lim_{n \to \infty} x_n$ says that $x = s = \lim_{n \to \infty} x_n$. Instead of the sequence being increasing it could be decreasing. We can reverse the sequence by multiplying every number by (-1) and then apply to theorem to the new increasing sequence. Reverse the sign and that is the limit of the original sequence.

We now show the Bolzano-Weierstrass theorem.

**Theorem C.2.2 (Bolzano-Weierstrass).** Given an infinite bounded sequence $(x_i) \in \mathbb{R}^n$ (or $\mathbb{C}^n$), there is a convergent subsequence.

**Proof.** Let us remark again that Bolzano provided a proof of this by using the bisection method which selects a special subsequence. The proof presented here finds a monotone subsequence. There could be actually at least two monotone subsequences each of them converging to a limit. The proof of this is shown in Theorem [C.2.5] which provides a less graphic and more formal methodology. Here we offer a proof of finding only one subsequence that converges since that is all we need.

Let us start with $n = 1$, that is the real line $\mathbb{R}$. The idea here is to construct a sequence that is monotonic and use the monotone convergence Theorem [C.2.1].

Figure [C.1] illustrates the methodology of the proof. Let us assume that after some $n_1$ all terms of the sequence are smaller than $x_{n_1}$. In the example, the red point at $n_1 = 1$, then find the next (smaller) index $n_2$ after which every point in the sequence is smaller (in the Figure this is red point at $n_2 = 5$). Keep selecting the next highest point in the tail. This yields the peak points at $\{x_{n_1}, x_{n_2}, x_{n_3}, x_{n_5}, \ldots \}$. In the Figure this corresponds to the points $x_i$ with $i = 1, 5, 9, 14, 17$.

We consider the following cases for the subsequence of $x_{n_i}$’s which we know, by construction, should be decreasing.

- **The sequence is infinite.** Then from Theorem [C.2.1] it converges.

- **The sequence is finite (including empty).** Let $N$ be the last red point. If there is no such a point, set $N = 0$. Then $n_1 = N + 1$ is not a peak so there is some $n_2 > N$ such that $x_{n_1} \leq x_{n_2}$. Again, since $n_2 > N$ is not a peak, there is an $n_3$, such that $x_{n_2} \geq x_{n_3}$. We keep iterating in this way and find a subsequence $(x_{n_i})$ which is increasing monotonically. Then we apply Theorem [C.2.1] and find that the subsequence $(x_{n_i})$ so chosen converges.
Figure C.1: The red points represent the peaks from the path ahead.

Now for $\mathbb{R}^n$. The first component $(x_1_i)$ forms itself a sequence, which by the result in one dimension has a subsequence which is convergent. Select this subsequence. From the selected subsequence pick the second components, that his $(x_2_i)$. This, again, is a sequence on the reals and so there is a subsequence of it that converges. Since the number of dimensions is finite, this process can be carried on $n$-times and then find a subsquence $(x_j)$ which converges. For the purposes of this proof we can see $\mathbb{C}^n$, as $\mathbb{R}^{2n}$, by allocating the real components of $\mathbb{C}$ in the odd components of the sequence of vectors and the imaginary components of $\mathbb{C}$ in the even components of the sequence of vectors.

C.2.3 Completeness of the Reals

We now state the axiom of the Archimedian property of the real line and use it to proof an important theorem that is used often to proof that some quantities are 0.

Axiom 2 (Arquemedian). The Arquemedian property of the real line indicates that given any positive number $x$, and any positive number $y$, there exits $n \in \mathbb{N}$, such that $nx > y$.

In words, this says: give me a stick and I can measure the world.

The first corollary from this axiom is:

Theorem C.2.3. The set of natural numbers $\mathbb{N}$ is unbounded.
Proof. Assume that it is bounded by a bound \( x \). Then provide a stick of measure 1. From the Arquemedian property there is a number \((n)(1) = n > x\). So this contradicts that \( x \) is a bound for the natural numbers.

A more important property is the following which will be used to proof that some numbers are zero.

**Theorem C.2.4.** If for all \( \epsilon > 0 \), we have that \(|x - y| < \epsilon\). Then \( x = y \).

Proof. Let us assume \( x \neq y \), and call the difference \(|x - y| = w\). From the Arquemedian principle, there exists \( n \in \mathbb{N} \) such that \( wN > 1 \), or in other words \(|x - y| > 1/N\). We are free to pick \( \epsilon > 0 \), but once we pick \( \epsilon \) we can find \( N \) such that \( \epsilon < 1/N \), (from Theorem C.2.3) then

\[
|x - y| > \epsilon
\]

which contradicts the hypothesis. Then \( x = y \).

We now proceed to proof the completeness of the reals in the Cauchy sense. That is,

**Theorem C.2.5.** The space \( \mathbb{R} \) is complete.

Proof. This means that given any sequence \((x_n)\), which is Cauchy, that is for any \( \epsilon > 0 \), there exists an \( N \in \mathbb{N} \), such that for any \( m, n > N \), \(|x_n - x_m| < \epsilon\).

We first prove that if the sequence is Cauchy then it is bounded and from the Bolzano-Weierstrass theorem it has a limit point. The limit point should be unique.

We know that given \( \epsilon = 1 > 0 \) and for \( n \geq N \)

\[
|x_n| \leq |x_n - X_N + X_N| \leq |X_N - x_n| + |x_N| < 1 + |x_N|,
\]

and so the sequence \((x_n)\) is bounded by

\[
B = \max\{|x_1|, \cdots |x_{N-1}|\} + |x_N| + 1 , \quad \forall n \geq 1.
\]

Then the sequence \((x_n)\) is bounded both from above and below and so we can find the inf and the sup of this sequence and of any subsequence of the sequence. Let us call form the following new subsequences:

\[
y_n = \inf\{x_m : m \geq n\} \quad z_n = \sup\{x_m : m \geq n\}
\]
This construction is like the construction done to proof the Bolzano-Weirstrass theorem, so look at Figure [C.1]. The red points represents what we call here the subsequence $z_n$. The sequence $(y_n)$ is monotone increasing ($y_n \leq y_{n+1}$) while $(z_n)$ is monotone decreasing ($z_{n+1} \leq z_n$), and $y_n \leq x_n \leq z_n$ for each $n \geq 1$. Thus from the monotone convergence Theorem [C.2.1] we have

\[
\lim_{n \to \infty} y_n = L = \sup y_n : n \geq 1 \quad , \quad \lim_{n \to \infty} z_n = U = \inf z_n : n \geq 1 ,
\]

with $L \leq U$. From the definition of a Cauchy sequence, then provided $\epsilon > 0$ we can find $N_1, N_2 \in \mathbb{N}$ such that for any $n \geq N_1, N_2$, $|y_n - y_{N_1}| < \epsilon/4$, and $|z_n - z_{N_2}| < \epsilon/3$. From theorems of sup and inf [C.1.6] and [C.1.7] we know that there are $N_3, N_4 \in \mathbb{N}$ such that for all $n > N_3, N_4$ $|L - y_n| < \epsilon/3$ and $|U - z_n| < \epsilon/4$. Recall that the sequence $y_n$ is increasing and the sequence $z_n$ is decreasing. Take $N = \max\{N_i\}, i = 1, \ldots, 4$, and $n, m > M$. We put all this together:

\[
0 \leq U - L \\
\leq |U - z_n + z_n - y_m + y_m - L| \\
\leq |U - z_n| + |z_n - y_m| + |y_m - L| \\
< \epsilon/3 + \epsilon/3 + \epsilon/3 \\
= \epsilon
\]

We now use Theorem [C.2.4] and conclude that $U = L$, so the limit is unique. In conclusion: each Cauchy sequence converges in $\mathbb{R}$ and the real line is complete. \qed
Appendix D

Partitions, Equivalent Relations, and Functions

D.1 Equivalence Relations

Given a set we might present some properties that separates the set into disjoint sets each sharing the same property. This is very convenient to simplify big sets.

For example, when we represent a sequence of functions by the formula $f_n(x) = (-1)^ng(x)$, $n \in \mathbb{Z}$ and knowing $g(x)$, we can easily evaluate $f_n(x)$ without having to think of all the possible integer numbers. We just need to know if $n$ is even or $n$ is odd. In the first case $f_n(x) = g(x)$, and in the second $f_n(x) = -g(x)$. So by partitioning the integers into two sets, we simplify the problem, since for the purpose of the evaluation $f_n(x)$ we only require two know if $n$ is even or odd. The property “even” can serve to define a relation in the integer numbers as follows. We say $x$ is related to $y$, and write $xRy$, or $xEy$, or more commonly $x \sim y$ if $x \in \mathbb{Z}$ and the remainder of dividing $x - y$ and 2 is zero. That is, if $x - y = 2n$, for some $n$. It is then obvious that by dividing sets according to properties of interest we simply problems. Sorting groups of large numbers help on searching algorithms. A dictionary is seen as sorted according to the alphabet, and we could, for a given purpose consider all words that start with “a” to be “equivalent”. That is “airplane” and “arm” would be seen as equivalent from the purpose of the first letter.
The operation of dividing the integers into even and odd is a particular case of the modulus operation. We fix an integer \( n > 0 \) and choose two arbitrary \( x > y \) integers. Then the remainder of dividing \( x - y \) to \( n \) is one of the numbers \( 0, 1, \ldots n - 1 \). We can say that \( x \sim y \), if the remainder of dividing \( x - y \) to \( n \) is any of those \( n \) possible remainders. This splits the whole set of integers \( \mathbb{Z} \) into \( n \) sets equally distributed. For example if \( n = 3 \), we have the following three sets:

\[
[0] = \{\ldots, -6, -3, 0, 3, 6, \ldots\} \\
[1] = \{\ldots, -5, -2, 1, 4, 7, \ldots\} \\
[2] = \{\ldots, -4, -1, 2, 5, 8, \ldots\}.
\]

Clearly \( [0] \cup [1] \cup [2] = \mathbb{Z} \), and the mutual intersection of any two sets is empty. We see that \( -3 \sim 0 \) since they both belong to the same class \([0]\). The name of the class could be different. Any representative of the class can take the name, but usually the smallest positive representative takes the name of the class. This name is called the canonical representation of the class. In the particular case of integers we say that \( x \sim y \) using the following notation \( x \equiv y \mod n \), which means that the remainder of dividing \( x \) by \( n \) is the same as the remainder of dividing \( y \) by \( n \). We say that \( x \) is congruent to \( y \) modulus \( n \).

We now introduce a few formal definitions.

**Definition D.1.1 (Relation).** A relation on a set \( X \) is a subset \( C \) of the Cartesian product \( X \times X \).

A particular case for a relation is a function \( f : X \to X \). A function is a subset of relation such that for each in the domain, there is one, and only one element \( f(x) \).

**Definition D.1.2.** Let us assume that two elements \( a \) and \( b \) of a set \( X \) are related according to some rule \( \sim \). We say that \( \sim \) is:

- **reflexive** if \( a \sim a \).
- **symmetric** if \( a \sim b \Rightarrow b \sim a \).
- **transitive** if \( a \sim b \) and \( b \sim c \) implies \( a \sim c \).
D.1. EQUIVALENCE RELATIONS

If the set of elements satisfy the three properties above, they are said to be a class, and we say that the relation is an equivalent relation. Any element of the class can represent the name of the class. That is,

\[ [a] = \{ x \in X | a \sim x \} \]

D.1.3 Examples:

The following examples are easy to verify and we will only make observations where needed.

- The simplest example is equality “=” . This is an equivalence relation.
- The congruence relation shown above \( a \equiv b \mod n \), is an equivalence relation.
- The “less than ” relation “<” is not an equivalence relation, because it fails to be both reflexive and symmetric. Similarly, the “ less or equal than” relation “\( \leq \)” is not an equivalence relation.
- The proper subset “\( \subset \)” relation is not an equivalence relation. As in the relation “\( < \)” it is not reflexive, neither symmetric. Similarly The inclusion “\( \subseteq \)” is not an equivalence relation.
- In a tree all nodes are connected. If a node is above another node we call it ancestor, if a node is under another node we call it descendant, otherwise if a node is at the same level we call it sibling. The relation \( A \) is ancestor of \( B \) is not an equivalence relation. In the same way the relation \( B \) is descendant of \( A \) is not an equivalence relation. The relation \( A \) is a sibling of \( B \) is an equivalence relation.
- If \( X \) and \( Z \) are two sets, \( f : X \rightarrow Z \), and we define \( x \sim y \) to mean \( f(x) = f(y) \), then \( \sim \) is an equivalence relation.
- Define a relation \( \sim \) between two persons \( a \) and \( b \) to have the same birthday. This is an equivalence relation. All people with the same birthday share the same class. There would be 366 (assuming February 29) classes.
Think about the rational numbers. Two rational numbers $a/b$ and $c/d$ are equal if $ad = bc$. We can write the set

$$S = \{(x, y) \in \mathbb{Z}^2 : y \neq 0\},$$

and define the relation $(a, b) \sim (c, d)$ if $ad = bc$. Then the property $ad = bc$ defines a relation between $(a, b)$ and $(c, d)$. This is an equivalence relation in the set $S$. For example since $1/2 = 2/4$ we say that $(1, 2) \sim (2, 4)$.

A similar relation happens with triangles. This time define the set in $\mathbb{R}^3$

$$T = \{(x, y, z) : x > 0, y > 0, z > 0\}$$

such that $(a, b, c) \sim (e, f, g)$ if the ratios $a/e = b/f = c/g$ are the same. Geometrically this indicates that two triangles are similar. This defines an equivalence relation. Another equivalence relation between triangles is triangle congruence (lengths and angles are equal correspondingly).

Another way to see the rationals as equivalence classes is that proposed by Heine in 1872 as indicated in section 2.4.7. That is, given two real sequences $(x_n)$ and $(y_n)$, they are related if they have the same limit. That is, $(x_n) \sim (y_n)$ if $\lim_{n \to \infty} x_n = \lim_{n \to \infty} y_n$. This limit defines an equivalence relation and the rationals, as define by Heine, are equivalence classes. Note that Cantor also used Heine’s idea of rationals as elements of equivalence classes to provide a construction of the real line.

In logic, the equivalence “$\iff$” is an equivalence relation.

In linear algebra, two matrices $A$ and $B$ are similar if there is a non-singular matrix $M$ such that $A = MAM^{-1}$. The linear algebra similarity is an equivalence relation.

In normed spaces. The norm of a vector defines an equivalence relation. That is is $\|x\| = \|y\|$, we say that $x \sim y$. The classes are those of vectors with the same norm. We can say that all sticks with the same length belong to the same class. Instead of norm we can use the angle as the attribute and say that all parallel lines belong to the same class.
Figure D.1: The classes \([x]\) (blue) and \([y]\) (green).

**Theorem D.1.1.** Two equivalence classes \([x]\) and \([y]\) are either disjoint or equal.

**Proof.** We show this by contradiction. We use Figure D.1 for reference. Assume that we have intersection and that there is \([x]\) and \([y]\) are different. That is that \(x\) is not related to \(y\). Since there is intersection there is a \(z\) such that \(z \sim x\), and \(z \sim y\). From the symmetry we have that \(x \sim z\) and from transitivity that \(x \sim y\). This is a contradiction because we assume that \(x\) and \(y\) are not related. So the only option is that the sets are equal or disjoint. 

\[\square\]

### D.2 Partitions

Since \(X\) is our universal set each class \([x]\) is a subset of \(X\). On the other hand, for any \(x \in X\), there should be a class such that \(x \in [x]\). That is \(X = \bigcup_{x \in X} [x]\). This motivates the following definition.

**Definition D.2.1.** A partition of a set \(X\) is a collection of disjoint nonempty subsets of \(A\) whose unions is all of \(A\). That is: The partition \(\{P_\alpha\}_{\alpha \in A}\) satisfies:

\[X = \bigcup_{\alpha} P_\alpha \quad P_\alpha \neq \emptyset \quad P_\alpha \cap P_\beta = \emptyset \quad \text{for } \alpha \neq \beta.\]

Then an equivalence relation on \(X\) determines a partition on \(X\). Actually, the implication in the other direction is also true. That is:
Theorem D.2.1 (Fundamental Theorem on Equivalence Relations). An equivalence relation \( \sim \) on a non-empty set \( X \) determines a partition of \( X \) and conversely, a partition of \( X \) induces an equivalence relation \( \sim \) on \( X \).

Proof. We only need to proof the implication “\( \Leftarrow \)” since the direct implication “\( \Rightarrow \)” was shown above.

Assume that \( P = \{X_\alpha\}_{\alpha \in A} \) is a partition of \( X \). We say that \( x \sim y \) if there is an \( \alpha \in A \) such that \( x, y \in X_\alpha \). We show that \( \sim \) is an equivalence relation.

(i) For any \( x \in X \), \( x \in X_\alpha \) for some \( \alpha \) by the definition of partition. Now, \( x = x \), so \( x \sim x \).

(ii) Let us assume that \( x \sim y \), so \( x, y \in X_\alpha \), or what is the same \( y, x \in X_\alpha \), so \( y \sim x \).

(iii) Let us assume \( x \sim y \) and \( y \sim z \). So \( x, y \in X_\alpha \) for some \( \alpha \), and \( y, z \in X_\beta \) for some \( \beta \). Since \( P \) is a partition and \( y \in X_\alpha \cap X_\beta \), then \( X_\alpha = X_\beta \).

\( \Box \)

So any partition defines an equivalence relation. This provides a method to build new equivalence relations from partitions.

D.2.2 Examples:
- In a set the “\( = \)” equivalence relation creates a partition of the set in singletons. That is any class has exactly one element. \([x] = \{x\}\).
- We can see how the congruence relation between integers splits sets with a simple example. Think of \( X = \{0, 1, \cdots, 10\} \). We can not divide 11 into 3 exactly but we can split \( A \) in groups of 3 and the remainder will be naturally assigned to groups as follows. Think of three bins and start by putting the numbers in order into each bin. That is, if the bins are labeled as \( A, B, \) and \( C \), we put 0 into \( A \), 1 into \( B \), 2 into \( C \) and then return to \( A \) with 3, and so until run out of numbers.

\[
\begin{align*}
A & \leftarrow 0\ 3\ 6\ 9 \\
B & \leftarrow 1\ 4\ 7\ 10 \\
C & \leftarrow 2\ 5\ 8
\end{align*}
\]
D.2. PARTITIONS

Here $X = [0] \cup [1] \cup [2]$. The classes are divided in the most equal numerical distribution having whole number of elements. Classes $[0] = \{0, 3, 6, 9\}$ and $[1] = \{1, 4, 7, 10\}$ have 4 elements each and class $[2] = \{2, 5, 8\}$ has 3 elements. Element $b$ is in the class $[a]$ if and only if $b \equiv a \mod 3$. That is the remainder after dividing by 3 is zero.

• Let us assume a set $X_n = \{a_i\}_{i=1}^n$ and formulate the following questions:

  (i) How many relations can we obtain from $X$?
  (ii) How many of those relations are equivalence relations?
  (iii) How many of those relations are functions?
  (iv) Build the explicit list of equivalence relations for the set $X_3 = \{a_1, a_2, a_3\}$.

We solve this example, which most of it belongs to the field of combinatorial theory.

Solution

(i) The number of related couples $(a_i, a_j)$ is given by the number of choices of $i = 1, \ldots, n$, and $j = 1, \ldots, n$. That is the product $n \times n = n^2$. The number of relations is the number of subsets of the related couples, that is $2^{n^2} - 1$, where we eliminate the empty set $\emptyset$.

(ii) The number of equivalence relations is the same number of partitions of set set $X_n$, from Fundamental Theorem of Equivalence Relations [D.2.1]. The number of partitions of a set is known as the Bell number $B_n$.

There are many interesting representations of the Bell numbers in terms of recursion formulas, generating functions, closed formulas, integration on the complex plane, etc. We derive a recursion formula using induction. For $n = 0$ the empty set has only 1 partition. That is $B_0 = 1$. With $n = 1$, $X_n$ has only one element $X_1 = \{x_1\}$ and $B_1 = 1$. Now $X_2$ has two elements. What we need to know is how this new element $x_2$ changes the picture.

$$X_2 = \{x_1, x_2\} = X_1 \cup \{x_2\}.$$

\footnote{https://en.wikipedia.org/wiki/Bell_number}
The new element \( x_2 \) can go in any old subset of the old partition of \( X_1 \). That is for each of the \( B_i \) partitions \( i < 2 \), of \( X_1 \), \( x_2 \) can go into it as follows:

(i) Join the \( \emptyset \). That is
\[
\{x_2\} \cup \emptyset = \{x_2\},
\]
We count this as \( 1B_0 = \binom{1}{0}B_0 \).

(ii) join \( x_1 \). That is
\[
\{x_1, x_2\}.
\]
We count this as \( 1B_1 = \binom{1}{1}B_1 \).

\[
B_2 = \binom{1}{0}B_0 + \binom{1}{1}B_1 = 2.
\]

Explicitly we have
\[
P_1 = \{\{x_1\}, \{x_2\}\}
\]
\[
P_2 = \{\{x_1, x_2\}\}.
\]

We now assume that we know \( B_n \) for the set \( X_n \) and construct the set \( X_{n+1} \) where the new element \( x_{n+1} \) should be added to the set \( X_n \). We ask, how this new element \( x_{n+1} \) will shake things up.

Given a partition of \( X_{n+1} \), the element \( x_{n+1} \) should be present in one (and only one) block of this partition. Let us take that block and assume that the block has \( k \) elements (no counting the \( x_{n+1} \) new element). How many of these block prototypes would we have? We can pick any \( k \) out of \( n \) elements and this will represent the number of blocks with \( k + 1 \) (the 1 is for \( x_{n+1} \)) elements where \( x_{n+1} \) is present. This number is \( \binom{n}{k} \). We still have to count partitions. Every block of this type corresponds to a partition with \( n + 1 - (k + 1) = n - k \) elements. That is for every block of this we have \( B_{n-k} \) partitions. The total number of partitions is
\[
\sum_{k=0}^{n} \binom{n}{k}B_{n-k}.
\]
Note that if $k = 0$, this means that $x_{n+1}$ goes into a singleton \{x_{n+1}\}, where $\binom{n}{0} = 1$ and we have $B_n$ partitions all without $x_{n+1}$. This happened just before adding $x_{n+1}$. On the other extreme is $k = n$. This means that the block is the whole set $X_n$ and there is no much to play with other than adding the element $x_{n+1}$ to the whole set $X_n$ making it $X_n^{n+1}$. Note that, since $\binom{n}{k} = \binom{n}{n-k}$ we can write

$$\sum_{k=0}^n \binom{n}{k} B_{n-k} = \sum_{k=0}^n \binom{n}{n-k} B_{n-k} = \sum_{k=0}^n \binom{n}{k} B_k.$$ 

The first few 6 Bell numbers are: 1, 1, 2, 5, 15, 52.

(iii) To count the total number of functions we can list the values explicitly as

$$f = (f(x_1), f(x_2), \ldots, f(x_n)) \in X_n \times X_n \times X_n = X_n^n$$

Since any $f(x_i)$ can take any value in $X_n$, we have a total of $n^n$. Note that $n^n \ll 2^{n^2}$ for $n > 4$.

(iv) To find the equivalence relations is easier by doing the partitions of the set \{a_1, a_2, a_3\}. That is, explicitly

$$\emptyset$$

$$\{\{a_1\}\}, \{\{a_2\}\}, \{\{a_3\}\}$$

$$\{\{a_1\}, \{a_2, a_3\}\}$$

$$\{\{a_1, a_2\}, \{a_3\}\}$$

$$\{a_1, a_2, a_3\}.$$ 

We then write as couples the relations

$$R_1 = \{(a_1, a_1), (a_2, a_2), (a_3, a_3)\}$$

$$R_2 = \{(a_1, a_1), (a_2, a_3), (a_3, a_2), (a_2, a_2), (a_3, a_3)\}$$

$$R_3 = \{(a_1, a_2), (a_2, a_1), (a_3, a_3), (a_2, a_2), (a_1, a_1)\}$$

$$R_4 = \{(a_1, a_2), (a_2, a_1), (a_1, a_3), (a_3, a_1), (a_1, a_1), (a_2, a_2),$$

$$\quad (a_3, a_3), (a_2, a_3), (a_3, a_2)\}.$$
It is easy to show that each relation $R_i$, with $i = 1, 2, 3, 4$ is an equivalence relation. Figure IV sketches the relations $R_i$ above.

**Definition D.2.3 (Quotient set).** Given an equivalence relation $\sim$, on a set $X$, the set of all classes

$$Q = X/\sim \overset{\text{def}}{=} \{[x] : x \in X\}$$

is called a quotient set.

where the symbol $X/\sim$ means quotient of $X$ with respect to the relation determined by $\sim$. The word “quotient” means that the space is divided in pieces. These spaces are used in algebra, and topology. From the example above illustrated in Figure IV we see four quotient sets

$$X_3/R_1 = \{\{a_1\}, \{a_2\}, \{a_3\}\}$$

$$X_3/R_2 = \{\{a_1\}, \{a_2, a_3\}\}$$

$$X_3/R_3 = \{\{a_1, a_2\}, \{a_3\}\}$$

$$X_3/R_4 = \{a_1, a_2, a_3\}.$$  

**Definition D.2.4 (Quotient map).** The map $q : X \to X/\sim$ defined by letting $q(x)$ be the equivalence class that contains $x$ is called the quotient map or natural map. Other names given to the quotient map are projection map or factorization map.

For illustration we show in Figure D.2.2 the corresponding quotient maps to those equivalence relations in Figure IV. In white background are the singletons, blue the sets of two elements, and in green the set of three elements. See that somehow the name project map fits well by noticing that components are beige projected in their representative classes.
D.3 Functions

We used functions above and want to introduce a more formal definition of function here together with a few related attributes.

Definition D.3.1 (function). Let $X$ and $Y$ be two non-empty sets. A subset $f$ of $X \times Y$ is called a function from $X$ to $Y$ if and only if to each $x \in X$, there exists a unique $y \in Y$ such that $(x, y) \in f$. $X$ is called the domain of $f$, $Y$ is called the codomain of $f$. The unique element $y$ assigned to $x$ is called the image; we write $y = f(x)$. The element $x$ is called pre-image or inverse image of $y$; we call it $x = f^{-1}(y)$. The function $f$ can be seen as its graph, that is as the points $\{(x, y) = f(x)\}$, with $x \in X$. The range of $f$ is the set of all images. We write the range as $f(X) = \{y \in Y : y = f(x)\}$. Given a set $A \subset X$, we call $f(A) = \{f(x) : x \in A\}$ the image of $A$ under $f$. If $B \subset Y$, then the $f^{-1}(B) = \{x \in X : f(x) \in B\}$ is called the inverse image of $B$ under $f$.

Other terms used for functions are mappings, transformations or operators. We note the function as $f : X \rightarrow Y$. It follows from the definition that

(i) For each $x \in X$ there is a $y \in Y$ such that $(x, y) \in f$. That is, each $x$ in $X$ has an image in $Y$.

(ii) Given $(x, y_1) \in f$ and $(x, y_2) \in f$, then $y_1 = y_2$. For each $x \in X$ there can be at most an image point in $Y$.

We introduce a few important functions.
Definition D.3.2.

- **Equal functions**: Two functions \( f : X \to Y \), and \( g : X \to Y \) are equal if and only if \( f(x) = g(x) \) for all \( x \in X \). We write \( f = g \).

- **Constant function**: A function \( f : X \to Y \) is called constant if for some \( y_0 \in Y \), \( f(x) = y \), for all \( x \in X \).

- **Identity**: The identity function is only defined in the same space. It is the diagonal of \( X \times X \). That is the set \((x, x)\), \( x \in X \).

- **Inclusion**: If \( B \subset A \) then the function

\[
i : B \to A
\]

\[
x \mapsto x
\]

is called the inclusion function from \( B \) to \( A \). Note that \( A = B \) the inclusion function is the identity.

- **Restriction**: If \( f : X \to Y \), and \( A \subset X \), then the mapping

\[
g : A \to Y
\]

\[
x \mapsto f(x)
\]

is called a **restriction of \( f \) to \( A \)**, and it is **denote as \( f/A \) or \( f_A \)**. 
**See that** \( f/A = f \cap (A \times Y) \).

- **Extension**: In the previous definition \( f \) is called an **extension of \( g \)**.

- **Many-one**: The function \( f \) is said to be **many-one** if and only if different elements in \( X \) have the same image in \( Y \). **one-to-one**: The function \( f \) is said to be **one-to-one** if and only

\[
f(x_1) = f(x_2) \implies x_1 = x_2 \text{ or } x_1 \neq x_2 \implies f(x_1) \neq f(x_2).
\]

One-to-one mappings are also called **injections**, and the function \( f \) is said to be **injective**.

- **into**: The function \( f \) is said **into** if there exists \( y \in Y \), such that \( f(x) \neq y \), for all \( x \in X \). That is, if \( f \) does not cover the whole space \( Y \), or the range \( f(X) \subset Y \).
D.3. FUNCTIONS

• **onto**: The function \( f \) is said to be **onto** if \( f(X) = Y \). That is, if for any \( y \in Y \), there is an \( x \in X \), such that \( f(x) = y \). We also say that the function \( f \) is **surjective**.

For the following theorem we provide easy counter-examples illustrated in Figure D.3.

**Theorem D.3.1.** Let \( f : X \rightarrow Y \) and suppose that \( \{A_\alpha\}_{\alpha \in A} \) and \( \{B_\beta\}_{\beta \in B} \) are families over arbitrary index sets \( A \) and \( B \). Then

(i) \( B_\beta = f(A_\alpha) \implies A_\alpha \subset f^{-1}(B_\beta) \). A counter-example is shown as \( f_1 \) in Figure D.3. An all-to-one function. Take \( A_\alpha = \{a_1\} \), then \( f(a_1) = a_2 \) but \( f^{-1}(a_2) = \{a_1, a_2, a_3\} \). Here the preimage is strictly larger than the original set.

Now, assuming \( B_\beta = f(A_\alpha) \), we have then that the result here suggests \( A_\alpha \subset f^{-1}(f(A_\alpha)) \).

The condition to make the sets equal is a one-to-one map. That is, if \( f \) is one-to-one then \( B_\beta = f(A_\alpha) \implies A_\alpha = f^{-1}(B_\beta) \). Again, function \( f_1 \) in Figure D.3 shows why the one-to-one condition is required.

This suggests that if \( f \) is one-to-one \( A_\alpha = f^{-1}(f(A_\alpha)) \).

(ii) \( A_\alpha = f^{-1}(B_\beta) \implies f(A_\alpha) \subset B_\beta \). This suggests \( f(f^{-1}(B_\beta)) \subset B_\beta \). The strict inclusion is explained by the first example of Figure D.3. Here we need the function to be onto to guarantee that the sets are equal.

That is, if \( f \) is onto, then \( A_\alpha = f^{-1}(B_\beta) \implies f(A_\alpha) = B_\beta \). Hence if \( f \) is onto then \( f(f^{-1}(B_\beta)) = B_\beta \).

For a counter-example when \( f \) is not onto, look again at the Figure D.3 for the function \( f_1 \). Assume \( B_\beta = \{a_2\} \). Then \( f_1^{-1}(B_\beta) = \{a_1, a_2, a_3\} \), and \( f(f_1^{-1}(B_\beta)) = \{a_1, a_2, a_3\} \notin B_\beta = \{a_2\} \).

(iii) \( A_\alpha \subset A_\gamma \implies f(A_\alpha) \subset f(A_\gamma) \).

(iv) \( B_\beta \subset B_\phi \implies f^{-1}(B_\beta) \subset f^{-1}(B_\phi) \).

We make a summary of the theorem here:

\[
A_\alpha \subset f^{-1}(f(A_\alpha)) \quad \text{equality if } f \text{ is one-to-one}
\]
\[
f(f^{-1}(B_\beta)) \subset B_\beta \quad \text{equality if } f \text{ is onto}
\]
\[
B_\alpha \subset A_\gamma \implies f(A_\alpha) \subset f(A_\gamma)
\]
\[
B_\beta \subset B_\phi \implies f^{-1}(B_\beta) \subset f^{-1}(B_\phi)
\]
Appendix D. Partitions, Equivalent Relations, and Functions

Figure D.4: Counter-example corresponding to theorem [D.3.1]

The proof of these statements is straightforward and left to the reader.

We now state a theorem for functions using unions and intersections of sets.

Theorem D.3.2.

(i) \( f(\bigcup \alpha A_\alpha) = \bigcup \alpha f(A_\alpha) \)

(ii) \( f(\bigcap \alpha A_\alpha) \subset \bigcap \alpha f(A_\alpha) \), equality is achieved when \( f \) is one-to-one.

(iii) \( f^{-1}(\bigcup \beta B_\beta) = \bigcup \beta f^{-1}(B_\beta) \)

(iv) \( f^{-1}(\bigcap \beta B_\beta) = \bigcap \beta f^{-1}(B_\beta) \)

Proof.

(i) \( y \in f(\bigcup \alpha A_\alpha) \iff \) there is an \( x \in \bigcup \alpha A_\alpha \), such that \( f(x) = y \)

\( \iff \) there is an \( \alpha \) such that \( x \in A_\alpha \) and \( f(x) = y \)

\( \iff \) \( y \in f(A_\alpha) \), for some \( \alpha \)

\( \iff \) \( y \in \bigcup \alpha A_\alpha \).

So the equality is proved.

(ii) Let us discuss first when the equality is no true. We should look for counter examples in into functions. For example consider \( f(x) = x^2 \), and two sets. \( A_1 = [-1,0] \), and \( A_2 = [0,1] \), then \( A_1 \cap A_2 = \{0\} \), and
so \( f(A_1 \cap A_2) = \{ f(0) \} = \{ 0 \} \). On the other hand \( f[-1, 0] = f[0, 1] = [0, 1] \), so \( f(A_1) \cap f(A_2) = [0, 1] \), and since \( \{ 0 \} \not\subset [0, 1] \) we show that the equality does not happen for this example.

Let \( y \in f(\cap_{\alpha} A_{\alpha}) \), then there exists \( x \in \cap_{\alpha} A_{\alpha} \), such that \( y = f(x) \). So for all possible \( \alpha \) indices \( x \in A_{\alpha} \). Then by definition \( y = f(x) \) is in \( f(A_{\alpha}) \), and so \( y \in \cap_{\alpha} f(A_{\alpha}) \).

(iii) The proof of this item is mixed of (i) and (iv) which we present next. We leave it to the reader.

(iv) \( x \in f^{-1}(\cap_{\beta} B_{\beta}) \iff \) there is a \( y \in \cap_{\beta} B_{\beta} \) such that \( y = f(x) \)

\[ \iff y \in B_{\beta} \text{ for all } \beta, \text{ and } y = f(x) \]

\[ \iff x \in f^{-1}B_{\beta} \text{ for all } \beta \]

\[ \iff x \in \cap_{\beta} f^{-1}(B_{\beta}), \]

and the equality is proved.

The final theorem on this appendix shows how functions behave under set subtraction.

**Theorem D.3.3.** Let \( A_1 \) and \( A_2 \) two subsets of \( X \) and \( B_1 \) and \( B_2 \) two subsets of \( Y \). Then

(i) \( f^{-1}(B_1 \setminus B_2) = f^{-1}(B_1) \setminus f^{-1}(B_2) \).

(ii) \( f(A_1 \setminus A_2) \supset f(A_1) \setminus f(A_2) \).

**Proof.**

(i) \( x \in f^{-1}(B_1 \setminus B_2) \iff \) there is a \( y \in B_1 \setminus B_2 \) such that \( y = f(x) \)

\[ \iff \text{there is } y \in B_1, y \not\in B_2 \text{ such that } y = f(x) \]

\[ \iff x \in f^{-1}(B_1) \text{ and } x \not\in f^{-1}(B_2) \]

\[ \iff x \in f^{-1}(B_1) \setminus f^{-1}(B_2). \]

and the equality is proved.

(ii) We first show a counter-example when the equality fails. Look back to the example in Figure [D.3] Consider \( A_1 = \{ a_1, a_2, a_3 \} \) and \( A_2 = \{ a_1 \} \). Then \( A_1 \setminus A_2 = \{ a_2, a_3 \} \), \( f_1(A_1 \setminus A_2) = \{ a_2 \} \). Now \( f(A_1) \setminus f(A_2) = \{ a_1 \} \setminus \{ a_2 \} = \emptyset \). So here \( \{ a_2 \} = f_1(A_1 \setminus A_2) \not= f_1(A_1) \setminus f_1(A_2) = \emptyset. \)
Let us assume \( y \in f(A_1) \setminus f(A_2) \). That is \( y \in f(A_1) \) and \( y \notin f(A_2) \). This means that there is an \( x \in A_1 \), and \( x \notin A_2 \), such that \( y = f(x) \). That is, there is an \( x \in A_1 \setminus A_2 \), such that \( y = f(x) \), or in other words \( x \in A_1 \setminus A_2 \), with \( y = f(x) \). So \( y \in f(A_1 \setminus A_2) \), and this proves \( f(A_1) \setminus f(A_2) \subset f(A_1 \setminus A_2) \). 
\( \square \)
Appendix E

Derivation of the some twisted surfaces in Cartesian Coordinates:

Let us consider a point \( (x, y) \in \mathbb{R}^2 \) a distance \( d \) from the origin and an angle \( \phi \in [0, 2\pi) \) with respect to the \( x \) axis. That is,

\[
d = x^2 + y^2, \quad x = d \cos \phi \quad \text{and} \quad y = d \sin \phi.
\]

We think about points \( (x, y) \) as lying on the horizontal plane. Let us now consider a closed path (or a segment as a degenerated case), in a vertical plane with coordinates \( (t, z) \). We want to allow rotations of this path with respect to the \( z \) axis by an angle \( \phi \) and around the a fixed point in the \( (x, y) \) plane \((z = 0)\) by an (twisting) angle \( \psi \). This is achieved by a multiplication of the point \((t, z)\) by a rotation matrix

\[
\begin{pmatrix}
c & s \\
-s & c
\end{pmatrix}
\]

with \( c = \cos \psi \), and \( s = \sin \psi \). The closed path is defined by an implicit equation \( f(t, z) = 0 \). We want to find the surface of the twist-revolution generated by this closed path.

Let us start by saying that the fixed point of rotation in intersection of the vertical plane, and the horizontal plane is a constant distance \( d \) and \( z = 0 \). We then want to locate the vertical plane at the distance \( d \), so instead of \( t \)
we talk about \( t - d \) (shifting by \( d \)). Then after rotation in the vertical plane (twisting), we find the new coordinates

\[
\begin{align*}
t' &= (t - d)c + sz \\
z' &= -(t - d)s + cz,
\end{align*}
\]

and the function becomes

\[
f(t', z') = f[(t - d)c + sz, -(t - d)s + cz].
\]

If we want to find \( f \) in terms of \((x, y, z)\) we should eliminate \( \phi, \psi, t \) (recall \( c = \sin \psi \), and \( s = \sin \psi \)).

To do this we observe that the point \( t \) is in the cylindrical radial direction \( \rho = (x, y) \), with the property that

\[
\begin{align*}
x &= t \cos \phi \\
y &= t \sin \phi
\end{align*}
\]

and from this \( x^2 + y^2 = t^2 \). In addition to the two equations \( E.0.4 \) we need one more to eliminate the three free parameters. We force the twist angle \( \psi \) to be a multiple of half \( \phi \). The Möbius strip requires \( \psi = \phi/2 \), but we can have a number \( k \) of complete twists by having \( \psi = k\phi/2 \). So the third equation is

\[
\psi = \frac{k\phi}{2}.
\]

Equations \( E.0.4 \) and \( E.0.4 \) form a system of three equations for \((x, y, z)\) in terms of \( t, \phi, \psi \). Elimination using these three equations of the \( \phi, \psi \) and \( t \) parameters provides a twisted surface of revolution in Cartesian coordinates.

Let us consider a few cases:

(i) \( k = 0 \). The simplest case. This is the case of a torus or a topological equivalent.

This means that \( \psi = 0 \). Then \( c = 1 \) and \( s = 0 \), and the transformed coordinates become:

\[
\begin{align*}
t' &= (t - d)c + sz = t - d \\
z' &= (t - d)s + cz = z
\end{align*}
\]
The function becomes \( f(t - d, z) = f(\sqrt{x^2 + y^2} - d, z) \). If the vertical curve is a circle, that is

\[
 f(t', z') \equiv (t')^2 + (z')^2 - r^2 = 0,
\]
we have that

\[
 \left( \sqrt{x^2 + y^2} - d \right)^2 + z^2 = r^2.
\]

This is precisely the Cartesian equation of the torus \( ?? \), with \( d = R \). A torus with elliptical cross-section is then given by the equation

\[
 \frac{(\sqrt{x^2 + y^2} - d)^2}{a} + \frac{z^2}{b} = 1,
\]
where \( a \) and \( b \) are the semi-axis lengths. When the semi-axis lengths are \( a = b = r \) we get back to the circular cross-section torus.

In general a deformed torus is given by the Cartesian equation

\[
 f(\sqrt{x^2 + y^2} - d, z) = 0,
\]
defined above, with \( f(t, z) \) being a close loop.

(ii) \( k = 2 \). It is interesting that the double twist is easier to derive than the single twist, since if \( k = 2 \), then \( \phi = \psi \), and then we can use the three equations

\[
\begin{align*}
 c &= \frac{x}{t} \\
 s &= \frac{y}{t} \\
 t^2 &= x^2 + y^2
\end{align*}
\]

To further simplify function \( \text{E.0.2} \) we find that

\[
 t' = (t - d)c + sz = \frac{x(t - d)}{t} + \frac{yz}{t} = \frac{xt - dx + yz}{t} \quad \text{(E.0.5)}
\]
and

\[
 z' = -(t - d)s + cz = -\frac{y(t - d)}{t} + \frac{xz}{t} = \frac{-yt + dy + xz}{t} \quad \text{(E.0.6)}
\]
where we leave $t$ unevaluated for simplification on the writing.

So the general equation here is

$$f(t', z') = f\left(\frac{xt - dx + yz}{t}, \frac{-yt + dy + xz}{t}\right)$$

Let us think of a circular cross section. That is,

$$f(t', z') \equiv (t')^2 + (z')^2 - r^2 = 0,$$

then

$$(xt - dx + yz)^2 + (-yt + dy + xz)^2 = t^2 r^2.$$ 

Squaring we find

$$x^2 t^2 + d^2 x^2 + y^2 z^2 - 2x^2 td + 2xyzt - 2dxyz$$

$$+ y^2 t^2 + d^2 y^2 + x^2 z^2 - 2y^2 td - 2xyzl + 2dxyz = t^2 r^2$$

and collecting terms

$$t^2(x^2 + y^2 - r^2) + d^2(x^2 + y^2) + z^2(x^2 + y^2) = -2td(x^2 + y^2)$$

Since $t^2 = x^2 + y^2$ we can factor this on the left and write

$$(x^2 + y^2)(x^2 + y^2 - r^2 + d^2 + z^2) = 2\sqrt{x^2 + y^2}d(x^2 + y^2)$$

and squaring this

$$(x^2 + y^2 + z^2 + d^2 - r^2)^2 = 4(x^2 + y^2)d^2$$

Which is a 4 order polynomial equation for a twisted cylinder and connected on its ends in a circular path. Well, this is equation 3.6.16 of the torus with $d = R$ and. Since the rotation of circle will leave the circle invariant this happens to be the equation of a torus. So, in order to see the M"obius strip-like shape we should use an elongated ellipse. That is we choose the equation
\[ f(t', z') \equiv \frac{(t')^2}{a} + \frac{(z')^2}{b} - 1 = 0. \] (E.0.7)

where \( b \gg a \). Using the transformations E.0.5 and E.0.6 we find

\[
\frac{(xt - dx - yz)^2}{t^2a} + \frac{(-yt + dy + xz)^2}{t^2b} = 1
\]

we multiply by \( t^2ab \) to find

\[ b(xt - dx - yz)^2 + a(-yt + dy + xz)^2 = t^2ab \]

Expanding the square factors and collecting terms we find

\[ t^2(bx^2 + ay^2 - ab) + 2xyzd(a + b) = -2t[bx(dx + yz) + ay(yd + xz)] \]

We should square this equation to make \( t \) into \( t^2 \) and get rid of the square root. Then

\[
[(x^2 + y^2)(bx^2 + ay^2 - ab) + 2xyz(a + b)]^2 = 4(x^2 + y^2)[bx(dx + yz) + ay(yd + xz)]^2
\] (E.0.8)

For graphing purpose let us choose

\[
\begin{align*}
  d &= 1 \\
  a &= 0.04 \\
  b &= 0.38
\end{align*}
\]

Figure E.1 includes the surface for equation E.0.8.
APPENDIX E. DERIVATION OF THE SOME TWISTED SURFACES IN CARTESIAN COORDINATES:

Figure E.1: Double twist on a Mobius for equation \([E.0.8]\) and parameters \(d = 1.0, a = 0.04, b = 0.38\). Figure made with GRAFER. Since the \(b/a\) ratio is large it looks more like a Möbius strip rather than a Möbius solid

(iii) \(k = 1\). Here only one twist is applied. If \(k = 1\) then \(\psi = 2\phi\), and using the formulas for \(\cos \phi/2\) and \(\sin \phi/2\) we find

\[
\begin{align*}
    c^2 &= \frac{1 + \cos \phi}{2} \\
    s^2 &= \frac{1 - \cos \phi}{2},
\end{align*}
\]

and from here

\[
    c^2 s^2 = \frac{1 - \cos^2 \phi}{4} \implies cs = \frac{\sin \phi}{2}.
\]

It is convenient to multiply and divide each fraction in the above equations by \(t\) for conversion to the \((x, y, z)\) coordinates. That is, we get

\[
\begin{align*}
    c^2 &= \frac{t + x}{2t} \\
    s^2 &= \frac{t - x}{2t} \\
    cs &= \frac{y}{2t}.
\end{align*}
\] (E.0.10)
Before using these results, let us rewrite the coordinate transformation between \((t, z)\) and \((t', z')\), 

\[
\begin{align*}
  t' &= (t - d)c + sz \\
  z' &= -(t - d)s + cz,
\end{align*}
\]

and the twisting ellipse is

\[
\frac{[(t - d)c + sz]^2}{a} + \frac{[(t - d)s - cz]^2}{b} - 1 = 0.
\]

We want simplify the twisting ellipse by expanding and multiplying by \(ab\). This is,

\[
\begin{align*}
  b(t - d)^2c^2 + 2bsct - d^2s^2 + as^2(t - d)^2 \\
  -2ascz(t - d) + ac^2z^2 - ab &= 0.
\end{align*}
\]

We further expand the binomial expressions and collect terms in \(t\).

\[
\begin{align*}
  (a + b)t^3 - [2(a + b)d - (a - b)x]t^2 + \\
  [(a + b)(z^2 + d^2) - 2d(a - b)(x + yz) - 2ab]t \\
  +(a - b)(xz^2 + 2dyz - d^2x) = 0 \quad \text{(E.0.11)}
\end{align*}
\]

We now move odd powers of \(t\) to the left and even to the right.

This is

\[
\begin{align*}
  (a + b)t^3 + [(a + b)(z^2 + d^2) - 2d(a - b)(x + yz) - 2ab]t \\
  = [2(a + b)d - (a - b)x]t^2 - (a - b)(xz^2 + 2dyz - d^2x)
\end{align*}
\]

Factor a \(t\) on the left

\[
\begin{align*}
  t[(a + b)t^2 + [(a + b)(z^2 + d^2) - 2d(a - b)(x + yz) - 2ab]] \\
  = [2(a + b)d - (a - b)x]t - (a - b)(xz^2 + 2dyz - d^2x)
\end{align*}
\]
and square both sides

\[ t^2[(a + b)t^2 + [(a + b)(z^2 + d^2) - 2d(a - b)(x + yz) - 2ab)]^2 \]
\[ = [2(a + b)d - (a - b)x]t^2 - (a - b)(xz^2 + 2dyz - d^2x)]^2 \]

Finally replace \( t^2 \) by \( x^2 + y^2 \).

\[ (x^2 + y^2)[(a + b)(x^2 + y^2 + z^2 + d^2) - 2d(a - b)(x + yz) - 2ab)]^2 \]
\[ = [2(a + b)d - (a - b)x][x^2 + y^2) - (a - b)(xz^2 + 2dyz - d^2x)]^2 \]

Which is a six order polynomial in \((x, y)\) and fourth order polynomial in \(z\).

Figure E.2 shows four surfaces corresponding to different chosen parameters \(a, b, c\). The parameters used are shown in the caption for each figure. In none of the figures there is a clear twist, we believe this is due to the smoothness of the surfaces and the fatness form rotating an ellipse instead of a straight segment.

(iv) The final case that we consider is that of a segment instead of an ellipse.
$a = 0.3, b = 0.3, \text{ and } d = 1.0$

$\text{Figure E.2: Tests for equation E.0.12. In the top left we have a torus where } a = b, \text{ and } R = 1, \text{ in the top right a torus where } d \text{ is too small to create the hole in the middle of the torus. In the bottom left we see a Möbious volumetric surface, for one twist. Due to the fatness and smoothness on the graphic we can not see the twist. Last, in the bottom right the separation between } a \text{ and } b \text{ opened up the torus into a bean shape figure.}$

$\text{a} = 0.03, b = 0.03 \text{ and } d = 0.08$

$a = 0.23, b = 0.28, \text{ and } d = 1.0$

$a = 0.08, b = 0.6, \text{ and } d = 0.9$
## Index

<table>
<thead>
<tr>
<th>Term</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>$C^0[0, 1]$ space</td>
<td>18</td>
</tr>
<tr>
<td>$L^p[a, b]$ spaces</td>
<td>23</td>
</tr>
<tr>
<td>$\ell^p$</td>
<td>16</td>
</tr>
<tr>
<td>$\ell^p$ distance</td>
<td>16</td>
</tr>
<tr>
<td>$\ell^p$ norm for sequences</td>
<td>16</td>
</tr>
<tr>
<td>$\ell^p$ norms</td>
<td>12</td>
</tr>
<tr>
<td>$\ell^\infty$</td>
<td>13</td>
</tr>
<tr>
<td>$\rho$-neighborhood</td>
<td>86</td>
</tr>
<tr>
<td>absolute difference metric</td>
<td>9</td>
</tr>
<tr>
<td>absolute homogeneity</td>
<td>10</td>
</tr>
<tr>
<td>accumulation point</td>
<td>83</td>
</tr>
<tr>
<td>accumulation point in metric spaces</td>
<td>31</td>
</tr>
<tr>
<td>almost everywhere</td>
<td>22</td>
</tr>
<tr>
<td>Arquemedian property</td>
<td>158</td>
</tr>
<tr>
<td>Arzelà-Ascoli theorem</td>
<td>65</td>
</tr>
<tr>
<td>Axiom of Completeness</td>
<td>58</td>
</tr>
<tr>
<td>Axiom of Completeness in metric spaces</td>
<td>68</td>
</tr>
<tr>
<td>ball in metric spaces</td>
<td>26</td>
</tr>
<tr>
<td>Banach Space</td>
<td>57</td>
</tr>
<tr>
<td>base</td>
<td>89</td>
</tr>
<tr>
<td>base for the topology</td>
<td>90</td>
</tr>
<tr>
<td>base of a topology</td>
<td>89</td>
</tr>
<tr>
<td>base of neighborhoods</td>
<td>91</td>
</tr>
<tr>
<td>Basic Topology</td>
<td>79</td>
</tr>
<tr>
<td>basis</td>
<td>89</td>
</tr>
<tr>
<td>Bell number</td>
<td>167</td>
</tr>
<tr>
<td>bisection algorithm</td>
<td>51</td>
</tr>
<tr>
<td>Bolzano-Weirstrass</td>
<td>75</td>
</tr>
<tr>
<td>Bolzano-Weirstrass Theorem</td>
<td>157</td>
</tr>
<tr>
<td>boundary in metric spaces</td>
<td>27</td>
</tr>
<tr>
<td>boundary point</td>
<td>84</td>
</tr>
<tr>
<td>bounded</td>
<td>61</td>
</tr>
<tr>
<td>bounded above</td>
<td>154</td>
</tr>
<tr>
<td>bounded below</td>
<td>154</td>
</tr>
<tr>
<td>bounded metrics</td>
<td>24</td>
</tr>
<tr>
<td>bounded metrics in infinite product space</td>
<td>25</td>
</tr>
<tr>
<td>bounded sequence</td>
<td>61</td>
</tr>
<tr>
<td>box product</td>
<td>103</td>
</tr>
<tr>
<td>box topology</td>
<td>111</td>
</tr>
<tr>
<td>canonical representation</td>
<td>162</td>
</tr>
<tr>
<td>Cantor set</td>
<td>40</td>
</tr>
<tr>
<td>Cantor set connectedness</td>
<td>45</td>
</tr>
<tr>
<td>Cardinality of the Cantor Set</td>
<td>43</td>
</tr>
<tr>
<td>Cartesian product</td>
<td>101</td>
</tr>
<tr>
<td>category theory</td>
<td>94</td>
</tr>
<tr>
<td>Cauchy sequence</td>
<td>49</td>
</tr>
<tr>
<td>Cauchy-Schwartz inequality</td>
<td>12</td>
</tr>
<tr>
<td>characteristic function</td>
<td>22</td>
</tr>
<tr>
<td>class</td>
<td>162</td>
</tr>
<tr>
<td>closed</td>
<td>83</td>
</tr>
<tr>
<td>closed ball in metric spaces</td>
<td>26</td>
</tr>
<tr>
<td>closed box</td>
<td>68</td>
</tr>
<tr>
<td>closed in metric spaces</td>
<td>27</td>
</tr>
<tr>
<td>closed map</td>
<td>93</td>
</tr>
<tr>
<td>closure</td>
<td>83</td>
</tr>
<tr>
<td>closure in metric spaces</td>
<td>31</td>
</tr>
<tr>
<td>closure of the Cantor set</td>
<td>45</td>
</tr>
</tbody>
</table>
INDEX

coarser, 96

codomain, 171

compact metric space, 66

compatible topology, 96

complete, 57

complete regular, 103

congruent, 162

constant function, 172

continuity, 47

continuous, 47, 93

continuous at a point, 93

continuous Hölder’s inequality, 146

continuous Minkowski’s inequality, 150

contours, 121

converge, 38

convex functions, 139

convex set, 30

counting finite topologies, 97

decomposition space, 126

Dedekind cut, 57

deleed neighborhood, 31, 83

dense, 83

dense in metric spaces, 31

density of the Cantor set, 46

derived set, 31

diameter, 61

disconnected, 35

Discrete Hölder’s inequality, 144

discrete metric, 8, 87

discrete Minkowski inequality, 148

discrete topology, 87

distance, 7

domain, 171

equal functions, 172

equicontinuous, 65

equivalent relation, 163

Euclidean norm, 12

every where dense, 31

extension, 172

exterior in metric spaces, 27

factorization map, 170

fiber, 121

finer, 96

finite intersection property (FIP), 70

first countable, 91

function, 162, 171

Functional Analysis, 18

functional representation of the product, 103

graph, 171

Hölder’s conjugaes, 144

Hölder’s inequality, 12, 144

Hausdorff property, 69

Heine-Borel, 64, 75

homeomorphic, 94

homeomorphism, 94

identification space, 126

identification topology, 120

identity, 172

identity of indiscernible points of the norm, 10

identity of indiscernible points, 8

image, 171

image of $A$ under $f$, 171

inclusion function, 172

indicator function, 22

indiscrete topology, 87

induced topology, 98

infimum, 154

injection, 172

<table>
<thead>
<tr>
<th>Term</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>injective</td>
<td>172</td>
</tr>
<tr>
<td>interior</td>
<td>84</td>
</tr>
<tr>
<td>interior in metric spaces</td>
<td>27</td>
</tr>
<tr>
<td>interior of the Cantor set</td>
<td>45</td>
</tr>
<tr>
<td>interior point</td>
<td>84</td>
</tr>
<tr>
<td>intersection topology</td>
<td>98</td>
</tr>
<tr>
<td>interval</td>
<td>86</td>
</tr>
<tr>
<td>into</td>
<td>172</td>
</tr>
<tr>
<td>inverse image</td>
<td>121</td>
</tr>
<tr>
<td>inverse image of B under f</td>
<td>171</td>
</tr>
<tr>
<td>isolated</td>
<td>85</td>
</tr>
<tr>
<td>isolated point in metric spaces</td>
<td>31</td>
</tr>
<tr>
<td>Jensen’s inequality</td>
<td>141</td>
</tr>
<tr>
<td>Lebesgue Integration</td>
<td>21</td>
</tr>
<tr>
<td>Lebesgue Spaces</td>
<td>21</td>
</tr>
<tr>
<td>level curves</td>
<td>121</td>
</tr>
<tr>
<td>level set</td>
<td>121</td>
</tr>
<tr>
<td>level surfaces</td>
<td>121</td>
</tr>
<tr>
<td>limit</td>
<td>61</td>
</tr>
<tr>
<td>limit of a sequence</td>
<td>49</td>
</tr>
<tr>
<td>limit point</td>
<td>38</td>
</tr>
<tr>
<td>limit points in metric spaces</td>
<td>31</td>
</tr>
<tr>
<td>lower bound</td>
<td>153</td>
</tr>
<tr>
<td>Manhattan distance</td>
<td>11</td>
</tr>
<tr>
<td>many-one</td>
<td>172</td>
</tr>
<tr>
<td>mappings</td>
<td>171</td>
</tr>
<tr>
<td>maximum</td>
<td>154</td>
</tr>
<tr>
<td>Measure of the Cantor Set</td>
<td>43</td>
</tr>
<tr>
<td>metric</td>
<td>7</td>
</tr>
<tr>
<td>metric space</td>
<td>7</td>
</tr>
<tr>
<td>minimum</td>
<td>154</td>
</tr>
<tr>
<td>Minkowski inequality</td>
<td>12</td>
</tr>
<tr>
<td>modulus</td>
<td>162</td>
</tr>
<tr>
<td>monotone convergence Theorem</td>
<td>156</td>
</tr>
<tr>
<td>n-cell</td>
<td>68</td>
</tr>
<tr>
<td>natural map</td>
<td>170</td>
</tr>
<tr>
<td>neighborhood</td>
<td>83</td>
</tr>
<tr>
<td>neighborhood in metric spaces</td>
<td>27</td>
</tr>
<tr>
<td>nested closed intervals</td>
<td>72</td>
</tr>
<tr>
<td>nested tails</td>
<td>71</td>
</tr>
<tr>
<td>Newton’s method</td>
<td>52</td>
</tr>
<tr>
<td>Newton-Raphson method</td>
<td>52</td>
</tr>
<tr>
<td>no negativity</td>
<td>7</td>
</tr>
<tr>
<td>no negativity of the norm</td>
<td>10</td>
</tr>
<tr>
<td>nowhere dense</td>
<td>83</td>
</tr>
<tr>
<td>nowhere dense in metric spaces</td>
<td>31</td>
</tr>
<tr>
<td>one-to-one</td>
<td>172</td>
</tr>
<tr>
<td>onto</td>
<td>173</td>
</tr>
<tr>
<td>open</td>
<td>85</td>
</tr>
<tr>
<td>open annulus</td>
<td>86</td>
</tr>
<tr>
<td>open cover</td>
<td>66</td>
</tr>
<tr>
<td>open disc</td>
<td>86</td>
</tr>
<tr>
<td>open in metric spaces</td>
<td>27</td>
</tr>
<tr>
<td>open map</td>
<td>93</td>
</tr>
<tr>
<td>open set</td>
<td>79</td>
</tr>
<tr>
<td>open subcover</td>
<td>66</td>
</tr>
<tr>
<td>operators</td>
<td>171</td>
</tr>
<tr>
<td>partition</td>
<td>165</td>
</tr>
<tr>
<td>perfect set</td>
<td>31</td>
</tr>
<tr>
<td>perfection of the Cantor set</td>
<td>45</td>
</tr>
<tr>
<td>point-to-point convergence</td>
<td>19</td>
</tr>
<tr>
<td>pre-image</td>
<td>171</td>
</tr>
<tr>
<td>premiage</td>
<td>121</td>
</tr>
<tr>
<td>Product of Normed Spaces</td>
<td>16</td>
</tr>
<tr>
<td>product topology</td>
<td>104</td>
</tr>
<tr>
<td>product topology on X × Y</td>
<td>105</td>
</tr>
<tr>
<td>projection map</td>
<td>170</td>
</tr>
<tr>
<td>pullback topology</td>
<td>93</td>
</tr>
<tr>
<td>Pythagoric</td>
<td>12</td>
</tr>
<tr>
<td>quotient map</td>
<td>120</td>
</tr>
<tr>
<td>quotient set</td>
<td>170</td>
</tr>
<tr>
<td>Term</td>
<td>Page</td>
</tr>
<tr>
<td>-------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>quotient space</td>
<td>126</td>
</tr>
<tr>
<td>quotient topology</td>
<td>120</td>
</tr>
<tr>
<td>range</td>
<td>171</td>
</tr>
<tr>
<td>reflexive</td>
<td>162</td>
</tr>
<tr>
<td>relation</td>
<td>162</td>
</tr>
<tr>
<td>restriction function</td>
<td>172</td>
</tr>
<tr>
<td>Riemann Integration</td>
<td>21</td>
</tr>
<tr>
<td>saturated set</td>
<td>122</td>
</tr>
<tr>
<td>saturated with respect to $f$</td>
<td>122</td>
</tr>
<tr>
<td>second countable</td>
<td>91</td>
</tr>
<tr>
<td>separable</td>
<td>85</td>
</tr>
<tr>
<td>separable in metric space</td>
<td>31</td>
</tr>
<tr>
<td>sequential compactness</td>
<td>63</td>
</tr>
<tr>
<td>sequentially compact</td>
<td>76</td>
</tr>
<tr>
<td>Sierpiński space</td>
<td>87</td>
</tr>
<tr>
<td>Space of Bounded Functions $B(A)$</td>
<td>18</td>
</tr>
<tr>
<td>space of functions</td>
<td>18</td>
</tr>
<tr>
<td>sphere in metric spaces</td>
<td>26</td>
</tr>
<tr>
<td>Stone Weirstrass Theorem</td>
<td>38</td>
</tr>
<tr>
<td>strong convergence</td>
<td>19</td>
</tr>
<tr>
<td>stronger</td>
<td>96</td>
</tr>
<tr>
<td>subbasis</td>
<td>92</td>
</tr>
<tr>
<td>supremum</td>
<td>16</td>
</tr>
<tr>
<td>surjective</td>
<td>173</td>
</tr>
<tr>
<td>symmetric</td>
<td>162</td>
</tr>
<tr>
<td>symmetry</td>
<td>8</td>
</tr>
<tr>
<td>taxi cab distance</td>
<td>11</td>
</tr>
<tr>
<td>the sequence space</td>
<td>15</td>
</tr>
<tr>
<td>Topological Definitions</td>
<td>79</td>
</tr>
<tr>
<td>Topological Space</td>
<td>79</td>
</tr>
<tr>
<td>topologically equivalent</td>
<td>94</td>
</tr>
<tr>
<td>topology</td>
<td>79</td>
</tr>
<tr>
<td>topology generated by the subbasis</td>
<td>92</td>
</tr>
<tr>
<td>totally disconnected</td>
<td>45</td>
</tr>
<tr>
<td>transformations</td>
<td>171</td>
</tr>
<tr>
<td>transitive</td>
<td>162</td>
</tr>
<tr>
<td>triangle inequality</td>
<td>8</td>
</tr>
<tr>
<td>triangle inequality of the norm</td>
<td>10</td>
</tr>
<tr>
<td>Trigonometrical Polynomials</td>
<td>38</td>
</tr>
<tr>
<td>trivial topology</td>
<td>87</td>
</tr>
<tr>
<td>Tychonoff spaces</td>
<td>103</td>
</tr>
<tr>
<td>uniform convergence</td>
<td>19</td>
</tr>
<tr>
<td>upper bound</td>
<td>153</td>
</tr>
<tr>
<td>upper semicontinuous</td>
<td>93</td>
</tr>
<tr>
<td>weak convergence</td>
<td>19</td>
</tr>
<tr>
<td>weaker</td>
<td>96</td>
</tr>
<tr>
<td>weight</td>
<td>91</td>
</tr>
<tr>
<td>Weighted geometric/arithmetic mean</td>
<td>143</td>
</tr>
<tr>
<td>inequality</td>
<td></td>
</tr>
<tr>
<td>Young’s inequality</td>
<td>144</td>
</tr>
</tbody>
</table>