A First Course in Topology

I. Bokor

December, 2014
The important thing to remember about mathematics is not to be frightened.

Richard Dawkins

Preface

Topology is one of the newer branches of mathematics, originating at about the close of the nineteenth century and beginning of the twentieth. The word topology only came to be commonly used later. The first tracts on topology included Poincaré’s *Analysis situs* (“Analysis of Place”) and Hausdorff’s *Mengenlehre* (“Set Theory”).

The twentieth century saw a rapid expansion in the study of topology. Several branches of topology emerged as disciplines in their own right, including point-set topology, algebraic topology, differential topology and geometric topology. Category theory and homological algebra both grew out of (algebraic) topology. Hardly any branch of mathematics remains untouched by topology.

Today, topology is more broadly applied than ever before. In the late twentieth century and in the twenty-first century, topology, category theory and homological algebra have become indispensable to current theoretical physics.

What Is Topology and Why Study It?

One way to regard topology is as the attempt to understand continuity in its broadest possible context. This is the perspective we adopt.

More specifically, we seek to answer a concrete question.

What additional structure must sets support in order to be able to speak of the continuity of functions between them?

Moreover, this structure should characterise continuity in the sense that two such structures on the space are “essentially the same” if and only if they lead to the same collection of continuous functions.

While this is already enough to justify the study of topology, it turns out that there is much more to topology, and its applicability to many other parts of mathematics, where questions of continuity do not arise, is truly remarkable.

We cannot possibly delve into these matters in a first introduction to topology, but it is important for the reader to be aware that this course is a very first and small sample of a vast part of modern mathematics.

Outline of This Course

Before embarking on our investigations, we review the mathematical background and fix notation. Our investigation of continuity begins with a closer examination of the notion of continuity familiar from calculus, the study of functions between subsets of various Euclidean spaces, $\mathbb{R}^n \ (n \geq 1)$.

We readily see that the definition of continuity met in calculus, depends only on a notion of *distance* as a measure of *closeness*, and on no other property of subsets of $\mathbb{R}^n$. 


By characterising “distance” we are led to the notion of a *metric*, which provides a measure of “distance” between elements. *Metric spaces* — sets equipped with a metric — and maps between them, generalise this aspect of subsets of $\mathbb{R}^n$ and we can define continuity of functions between them by simply rewriting the definition from calculus.

Viewing familiar sets and functions as metric spaces not only provides a new perspective, but also yields deeper insight into their structure and properties: Familiar theorems are specific instances of more general theorems. Moreover, these more general theorems are often simpler to prove than the already familiar special cases.

Our brief, preliminary study of metric spaces shows that metric spaces are not what we are seeking in our study of continuity. We provide examples of two essentially different metrics on the same space which give rise to identical sets of continuous functions.

A closer analysis of the relationship between a given metric on a set and the collection of functions continuous with respect to that metric, shows that it is not the metric itself which is significant, but only those subsets which are *open* with respect to the metric. Such sets generalise the notion of (unions of) open intervals of the real line, $\mathbb{R}$.

The characterisation of continuity in terms of open sets proves to be the characterisation of continuity we seek, for we show that for any set with two metrics the following two statements are equivalent.

(a) The two metrics determine the same classes of continuous functions.

(b) A subset of the given set is open with respect to one of the metrics if and only if it is open with respect to the other.

Thus, characterising continuous functions is equivalent to characterising open sets.

We formulate an axiomatic characterisation of the open subsets of a metric space, which does not explicitly mention the metric, and turn this into a definition: We call the collection of open sets a *topology* and define a *topological space* to be a set with a distinguished set of subsets satisfying the axioms characterising open sets in a metric space.

It is easy to see that two topologies on the same underlying set are “essentially the same” if and only if the same functions are continuous with respect to one of these topologies as are continuous with respect to the other.

A topology on a set therefore provides the characterisation of continuity we are seeking.

Since understanding continuity is the same as understanding topological spaces, we turn to investigating topological spaces in general. Several questions arise immediately.

1. *When do two topological spaces count as being “essentially the same”?*

2. *How can new topological spaces be constructed from given ones?*

3. *Which sets can be rendered topological spaces?*

4. *Can topological spaces be classified?*

These and similar questions motivate the further development of this course.

Since we are studying topological spaces in order to understand continuity, we consider two topological spaces to be “essentially the same” if and only if the cannot be distinguished by appealing to continuous functions alone. This occurs and only if there is a continuous bijection between them whose inverse is also continuous. Such functions are called *homeomorphisms* and so we consider two topological spaces “essentially the same” if and only if they are *homeomorphic*.

We can construct topological spaces from given ones in several ways.

For example, the subsets of a topological space inherit a *natural* topology.
Again, given two topological spaces, the Cartesian product of their underlying sets can be endowed with a *natural* topology.

Many problems in topology are concerned with whether properties of given topological spaces, or of functions defined in terms of them, are preserved by such constructions. Tychonov’s Theorem is an example of an affirmative answer. It states that if each member of a collection of topological spaces enjoys the property of *compactness*, then the *product* of that collection is also compact.

It may come as a surprise to learn that each and every set admits a topology. Moreover, if a set has at least two elements, it admits non-equivalent topologies, two of which are *natural*. One of these is the *discrete topology* and any function whatsoever defined on a space endowed with the discrete topology is automatically continuous.

This fact leads to a noteworthy situation. Plainly, topological spaces are special cases of sets and continuous functions are special cases of functions. But, we can also think of sets as a special case of topological spaces and every function as a special case of continuous functions!

An important part of understanding topological spaces is to classify them (up to homeomorphism). An ideal classification would consist of a finite list of *invariants* with the property that two spaces are homeomorphic if and only if all of their invariants agree. A topological invariant is anything which remains constant for all topological spaces homeomorphic to each other.

The idea is an old one, familiar to students of Euclidean geometry, where figures are classified up to *congruence*. Recall that two plane figures are said to be congruent if and only if one can be made to cover the other precisely by means of *rigid motions* alone — translations, reflections and rotations.

Thinking of geometry as the study of those properties which are invariant under rigid motions, we can think of topology as the study of those properties which are invariant under continuous motions. Since every rigid motion is continuous, topology is readily seen as a generalisation of geometry.

In the case of triangles a complete set of invariants is given by three positive numbers, namely the lengths of the sides: Two triangles are congruent if and only if their sides are of the same length.

No such complete classification of topological spaces is possible, the reasons being connected with *foundational problems*: a classification would imply solving some problems known to be unsolvable. (The details are beyond the scope of this course.) Because of the impossibility of this global classification and because of the importance and significance of several large classes of topological spaces, much effort has been expended on *restricted* classification problems, for example, classifying topological spaces up to weaker equivalence, such as *homotopy*, or classifying special classes of topological spaces, such as *manifolds*, which are studied in differential geometry and find application in such other disciplines, as physics.

Applications of topology to functional analysis, theoretical physics and differential geometry frequently depend upon two such invariants, namely *connectedness* and *compactness*. (The latter may be fruitfully considered as a “finiteness” condition.) These notions are introduced and investigated. The power of these notions is illustrated by proving two results familiar from calculus, namely the *Intermediate Value Theorem* and the *Extreme Value Theorem*. We also prove an apparently purely algebraic result about the field complex numbers, the *Fundamental Theorem of Algebra*.

Because many applications of topology involve metric spaces, with methods of approximation playing a central rôle, we investigate metric spaces more closely, especially in terms of convergence. This leads to the notion of *completeness*.

We also investigate compactness in the case of metric spaces, where it has different, but equivalent formulations.
Being only introductory, this course cannot aspire to present topology in its entirety. Hence its focus is on core aspects of topology — those parts which find application in all branches. It covers what is to be expected in a standard course in point-set topology: metric spaces, topological spaces, compactness, completeness and connectedness. No particular branch or application of topology is granted privileged status, so that functional analysts, real analysts, algebraic topologists, algebraic geometers and differential geometers will all find ample grounds to decry omissions.

The course is presented in the modern manner. The guiding light is provided by category theory, which arose from topology, so that commutative diagrams and universal properties appear wherever they simplify the presentation.

The only pre-requisites for the course are acquaintance with the notion of continuity from multivariate calculus and enough mathematical maturity for an elementary course on linear algebra.

The exercises form an integral part of the course. Some of the problems are intended to provide familiarity, others fill in details omitted from proofs in lectures and still others are applications to other branches of mathematics. The reader should definitely attempt them.
Contents

Preface
  What Is Topology and Why Study It?  i
  Outline of This Course  i

Chapter 1. Notation; Sets and Functions; Universal Properties  1
  1. Logical Notation  1
  2. Sets  1
  3. Functions  4
  4. Commutative Diagrams  12
  5. Equivalence Relations and Partitions  13
  6. Universal Properties  15
  7. Exercises  21

Chapter 2. Continuity in $\mathbb{R}^n$ and Metric Spaces  23
  1. Exercises  27

Chapter 3. Metric Preserving Functions and Isometry  31
  1. Exercises  34

Chapter 4. The Cartesian Product of Metric Spaces  35
  1. The Metric $\rho \times \sigma$  35
  2. The Metric $\rho_{X \times Y}$  37
  3. The Metric $\rho + \sigma$  38
  4. Comparison of the Three Metrics  39
  5. Exercises  41

Chapter 5. Open Sets in a Metric Space  43
  1. Exercises  47

Chapter 6. Topological Spaces  49
  1. Exercises  55

Chapter 7. Generating Topologies  57
  1. Exercises  62

Chapter 8. Constructing Topological Spaces from Given Ones  63
  1. Topological Sub-Spaces  63
  2. The Product of Topological Spaces  64
  3. The Quotient of a Topological Space  66
  4. Exercises  67

Chapter 9. Homeomorphism  69
1. Exercises 71

Chapter 10. Closed Sets 73
  1. Exercises 78

Chapter 11. Connectedness 81
  1. Exercises 87

Chapter 12. Separation 89
  1. Exercises 94

Chapter 13. Compactness 95
  1. The Fundamental Theorem of Algebra 101
  2. Exercises 102

Chapter 14. Locally Compact Spaces 105
  1. Compactification and Calculus 108
  2. Exercises 110

Chapter 15. Sequences, Convergence and Completeness 113
  1. A Fixed-Point Theorem 121
  2. Exercises 124

Chapter 16. Compact Metric Spaces 127
  1. Exercises 133

Chapter A. A Construction of the Real Numbers 135
  1. Dedekind’s Comments 135
  2. The Construction 136

Chapter B. The Greek Alphabet 143
Chapter 1

Notation; Sets and Functions; Universal Properties

Our first aim is to determine the additional structure sets may support in order for us to be able to speak sensibly about the continuity of functions. Accordingly, we begin by examining sets and functions, using the occasion to fix notation and to introduce universal properties, as they play a significant rôle not only in these notes.

1. Logical Notation

It is sometimes convenient to use logical notation. We list the notation we use.

\( P \implies Q \) for “if \( P \), then \( Q \)”

or “\( Q \) whenever \( P \)”;

or “\( P \) only if \( Q \)”;

\( P \iff Q \) for “\( P \) if and only if \( Q \)”

or \( P \) and \( Q \) are logically equivalent;

\( P :\iff Q \) for “\( P \) is defined to be equivalent to \( Q \)”;

\( \forall \) for “For every . . . ”;

\( \exists \) for “There is at least one . . . ”;

\( \exists! \) for “There is a unique . . . ”,

or “There is one and only one . . . ”.

2. Sets

A set is almost any reasonable collection of things. We do not attempt a more formal definition in this course. The things in the collection are called the elements of the set in question. We write

\[ x \in A \]

to denote that \( x \) is an element of the set \( A \) and

\[ x \notin A \]

to denote that \( x \) is not an element of the set \( A \).

We do not exclude the possibility that \( x \) be a set in its own right, except that \( x \) cannot be \( A \):

We explicitly exclude \( A \in A \).
Two sets are considered to be the same when they comprise precisely the same elements, in other words, when every element of the first set is also an element of the second and vice versa. Formally,

**Definition 1.1.** Given sets $A$ and $B$, $A$ is a *subset* of $B$ if and only if $x \in B$ whenever $x \in A$. We write $A \subseteq B$ whenever this is the case. $A = B$ if and only if $x \in A$ when and only when $x \in B$. In other words, $A = B$ if and only if $A \subseteq B$ and $B \subseteq A$.

$B$ is called a *proper* subset of $A$ if $B$ is a subset of $A$, but $B \neq A$. In such a case we write $B \subset A$.

Using our notational conventions, given two sets $A$ and $B$,$A \subseteq B :\iff (x \in A \implies (x \in B))$

$A = B :\iff ((x \in A) \iff (x \in B))$.

One of the axioms of set theory is that any collection of elements of a set is again a set. When we wish to describe a set, we can do so by listing all of its elements. Thus, if the set $A$ has precisely $a, b$ and $c$ as its elements, then we write $A = \{a, b, c\}$.

**Example 1.2.** By Definition 1.1, $\{a, b\}$, $\{a, b, b, b\}$ and $\{a, a, a, a, a, a, a, b\}$ are all the same set.

Another way of describing a set is by prescribing a number of *conditions* for membership of the set. In this case we write $\{x \mid P(x), Q(x), \ldots \}$ to denote that the set in question consists of all those $x$ for which $P(x), Q(x), \ldots$ all hold.

**Definition 1.3.** The *power set* of the set $A$, written $\mathcal{P}(A)$, comprises all the subsets of $A$. In symbols $B \in \mathcal{P}(A)$ if and only if $B \subseteq A$.

**Remark 1.4.** One of the axioms of set theory is that the collection of all subsets of any set is again a set, that is, the power set of a set is a set.

We regard forming the power set of sets as an *operation* on sets. There are several other important operations on sets.

**Definition 1.5.** The *union* of two sets $A$ and $B$, $A \cup B$, consists of all those objects which are in one or other (or both) of $A$ and $B$. In symbols, $x \in A \cup B :\iff x \in A$ or $x \in B$.

**Definition 1.6.** The *intersection* of the sets $A$ and $B$, $A \cap B$ consists of all those objects which are elements of both. In symbols, $x \in A \cap B :\iff x \in A$ and $x \in B$.

**Definition 1.7.** Those elements of $A$ that are not also elements of $B$ form a set in their own right, which we denote by $A \setminus B$, so that $A \setminus B := \{x \in A \mid x \notin B\}$. 
2. Sets

Here := has been used to signify that the expression on the left hand side is defined to be equal to the expression on the right hand side.

Definition 1.8. Given two sets $A$ and $B$, their (Cartesian) product, $A \times B$, consists of all ordered pairs, with the first of each pair an element of $A$ and the second of $B$. Symbolically, $A \times B := \{(x, y) \mid x \in A, y \in B\}$.

We can extend unions, intersections and cartesian products to larger collections of sets than merely pairs.

Definition 1.9. An indexed family of sets, with indexing set $\Lambda$ consists of a collection of sets, containing one set, say $A_\lambda$, for each element, $\lambda$, of the indexing set $\Lambda$. We write $\{A_\lambda \mid \lambda \in \Lambda\}$.

Definition 1.10. Given the indexed family of sets, $\{A_\lambda \mid \lambda \in \Lambda\}$, their union, intersection and (Cartesian) product are defined respectively by

$\bigcup_{\lambda \in \Lambda} A_\lambda := \{x \mid \text{there is a } \lambda \in \Lambda \text{ with } x \in A_\lambda\}$

$\bigcap_{\lambda \in \Lambda} A_\lambda := \{x \mid x \in A_\lambda \text{ for every } \lambda \in \Lambda\}$

$\prod_{\lambda \in \Lambda} A_\lambda := \{(x_\lambda)_{\lambda \in \Lambda} \mid x_\lambda \in A_\lambda \text{ for all } \lambda \in \Lambda\}$.

Here $(x_\lambda)_{\lambda \in \Lambda}$ denotes a generalised sequence: For each $\lambda \in \Lambda$, we choose an element, $x_\lambda$, from the set $X_\lambda$. Ordered pairs arise when $\Lambda = \{1, 2\}$ and sequences when $\Lambda = \mathbb{N}$.

The axioms of set theory assert that the union, intersection and cartesian product of any indexed set of sets is again a set.

Another axiom we use is the Axiom of Choice, on which there is extensive literature, and which has numerous equivalent formulations, one of which is the general version Tychonov’s Theorem (13.16). We record here a more familiar version

The Axiom of Choice. Given a non-empty set of non-empty sets, $\{A_\lambda \mid \lambda \in \Lambda\}$ there is a set $A$, such that for each $\lambda \in \Lambda$ there is an $x_\lambda \in A_\lambda$ with $A \cap A_\lambda = \{x_\lambda\}$

In other words, $A$ contains exactly one element from each of the given sets.

Several sets occur with such frequency that special notation has been introduced for them. These include the sets $\mathbb{N}$, $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$ and $\mathbb{C}$ consisting respectively of all natural numbers, all integers, all rational numbers, all real numbers and all complex numbers. Explicitly,

$\mathbb{N} := \{0, 1, 2, 3, \ldots\}$

$\mathbb{Z} := \{\ldots - 3, -2 - 1, 0, 1, 2, 3, \ldots\}$

$\mathbb{Q} := \{x \in \mathbb{R} \mid x = \frac{p}{q} \text{ for some } p, q \in \mathbb{Z}, \text{ with } q \neq 0\}$

$= \{x \in \mathbb{R} \mid x = \frac{p}{q} \text{ for some } p \in \mathbb{Z}, q \in \mathbb{N} \setminus \{0\}\}$

$\mathbb{C} := \{x + iy \mid x, y \in \mathbb{R} \text{ where } i^2 = -1\}$

Observe that $\mathbb{N} \subset \mathbb{Z} \subset \mathbb{Q} \subset \mathbb{R} \subset \mathbb{C}$.

One of the axioms of set theory is that there is a set with no elements. We write $\emptyset$ for this set and call it the empty set,
Note that it is a subset of every set, that is, if \( X \) is any set, then \( \emptyset \subseteq X \). It is therefore unique.

Another set playing an important rôle is the set \( \mathbb{2} \). It contains precisely two distinct elements, which we denote by 0 and 1, so that
\[
\mathbb{2} := \{0, 1\}
\]

### 3. Functions

To compare sets, we have the notion of a *function* or *map* or *mapping*.

**Definition 1.11.** A function, map, or mapping consists of three separate data, namely
1. a set, called the domain, on which the function is defined,
2. a set, called the co-domain, in which the function takes its values, and
3. the assignment to each element of the domain of a uniquely determined element of the co-domain.

We depict functions diagrammatically by
\[
f: X \rightarrow Y,
\]
or
\[
X \xrightarrow{f} Y
\]
Here \( X \) is the domain, \( Y \) is the co-domain and \( f \) is the name of the function. We write \( X = \text{dom}(f) \) and \( Y = \text{codom}(f) \) to indicate that \( X \) is the domain and \( Y \) the co-domain of \( f \).

We sometimes denote the function by \( f \) alone, but only when there is no danger of confusion.

If we wish to express explicitly that the function, \( f: X \rightarrow Y \), assigns the element \( y \in Y \) to the element \( x \in X \), then we write
\[
f: x \mapsto y
\]
or, equivalently, \( y = f(x) \), the latter notation being certainly familiar to the reader.

Sometimes the two parts are combined as
\[
f: X \rightarrow Y, \quad x \mapsto y
\]
or as
\[
f: X \rightarrow Y
\]
\[
x \mapsto y.
\]

**Definition 1.12.** If \( f \) assigns \( y \in Y \) to \( x \in X \), then we say that \( y \) is the image of \( x \) under \( f \) or just the image of \( x \).

The *image* of the function \( f: X \rightarrow Y \), is the subset, \( \text{im}(f) \), of \( Y \) given by
\[
\text{im}(f) := \{ y \in Y \mid y = f(x) \text{ for some } x \in X \}
\]
\[
= \{ f(x) \mid x \in X \}
\]
In other words, the image of a function is the set of “values” \( f \) actually takes.

Two functions \( f \) and \( g \) are equal, that is \( f = g \), if and only if
\[
\begin{align*}
\text{(i) } & \text{ dom}(f) = \text{dom}(g) \\
\text{(ii) } & \text{codom}(f) = \text{codom}(g) \\
\text{(iii) } & f(x) = g(x) \text{ for every } x \in \text{dom}(f).
\end{align*}
\]

**Observation 1.13.** To be the same, two functions must share both domain and co-domain, as well as agreeing everywhere.
3. Functions

Observation 1.14. By definition $\text{im}(f) \subseteq \text{codom}(f)$, that is, the image of a function is always a subset of its codomain. But the image need not be all of the co-domain.

Example 1.15. For $f: \mathbb{R} \to \mathbb{R}$ defined by $f(x) := 1$ for every $x \in \mathbb{R}$, that is $f: \mathbb{R} \to \mathbb{R}, \ x \mapsto 1$,
we have $\text{im}(f) = \{1\} \neq \mathbb{R} = \text{codom}(f)$.

Remark 1.16. It is essential to understand that a function is not just a formula. There are functions defined on the same domain, taking the same values everywhere, so that they have the same image, while not being the same function. The only difference between them is the values they do not take!
We shall meet an example soon.
It may seem peculiarly pedantic to distinguish such functions, but there are important algebraic and geometric examples which make this distinction indispensable. These notes adhere to the practice of specifying functions in the formally correct manner, so that it may become ordinary matter of course for the reader do so as well.

It should also be noted that a function given by a formula can be expressed in terms of different formulæ. Indeed, it is often a significant fact that two formulæ define the same function.

Example 1.17. An example familiar to the reader is provided by the function in Example 1.15, $\mathbb{R} \to \mathbb{R}, \ x \mapsto 1$,
which, as we know from trigonometry, is the same function as
$\mathbb{R} \to \mathbb{R}, \ x \mapsto \cos^2 x + \sin^2 x$.
That these two formulæ define the same function is Pythagoras’ Theorem.

Definition 1.18. Given sets $X$ and $Y$, we write $\mathcal{F}(X,Y)$, $[X,Y]$, or sometimes $Y^X$, for the collection of all functions $f: X \to Y$.

Remark 1.19. It follows from the axioms of set theory that the collection of all functions from any set to any set is again a set.

A function can sometimes be represented conveniently by means of its graph, as is familiar from calculus.

Definition 1.20. The graph, $\text{Gr}(f)$, of the function $f: X \to Y$ is
$\text{Gr}(f) := \{(x,y) \in X \times Y \mid y = f(x)\}$.

Definition 1.21. Given a function $f: X \to Y$ and subsets $A$ of $X$ and $B$ of $Y$, we define
$f(A) := \{y \in Y \mid y = f(x) \text{ for some } x \in A\}$
$= \{f(x) \mid x \in A\}$
$f^{-1}(B) := \{x \in X \mid f(x) \in B\}$.
Then $f(A)$ is called the image of $A$ under $f$ and $f^{-1}(B)$ is called the inverse image of $B$ under $f$, or the pre-image of $B$ under $f$.

When the subset, $B$, in question is a singleton, say $B = \{b\}$, we usually write $f^{-1}(b)$ instead of $f^{-1}(\{b\})$, unless there is danger of ambiguity.

For $y \in Y$, the subset $f^{-1}(y)$ of $X$ is called the fibre of $f$ over $y$. 
Given $A, B \subseteq X$ and $G, H \subseteq Y$ the relationships between

\[ f(A \cup B) \quad \text{and} \quad f(A \cap B), \quad \text{and} \quad f^{-1}(G \cup H) \quad \text{and} \quad f^{-1}(G \cap H) \]

on the one hand, and

\[ f(A) \cup f(B) \quad \text{and} \quad f(A) \cap f(B), \quad \text{and} \quad f^{-1}(G) \cup f^{-1}(H) \quad \text{and} \quad f^{-1}(G) \cap f^{-1}(H) \]

on the other, are important. These are investigated in the exercises, which the reader is expected to attempt.

Given the centrality of functions to mathematics, it is important to know what functions we have. Given a set, there is always at least one function defined on that set.

**Definition 1.22.** The *identity function* on the set $X$ is the function

\[ id_X: X \longrightarrow X, \quad x \mapsto x, \]

which we may also write as $id_X(x) := x$.

Notice that both the domain and co-domain must be precisely $X$ for this definition to specify the identity function.

**Remark 1.23.** While it may seem to be trivial, the identity function is the single most important function in mathematics.

**Definition 1.24.** If $X$ is a subset of $Y$, then the *inclusion map*, is

\[ i_X: X \longrightarrow Y, \quad x \mapsto x, \]

which we may also write as $i_X(x) := x$. It is often denoted simply by $i$ when the context makes the domain and co-domain clear.

This provides a characterisation subsets using functions.

**Theorem 1.25.** The set $X$ is a subset of the set $Y$ if and only if

\[ X \longrightarrow Y, \quad x \mapsto x \]

is a function.

**Proof.** Suppose that $X$ is a subset of $Y$. Then each element of $X$ is also an element of $Y$, whence, since each $x$ uniquely determines itself,

\[ X \longrightarrow Y, \quad x \mapsto x \]

is, indeed, a function.

If $X \longrightarrow Y, \quad x \mapsto x$ is a function, then each $x \in X$ must also lie in $Y$, whence $X \subseteq Y$. \qed
Observation 1.26. Using equational notation, we have \( i_Y^X(x) = x \).
This is the same formula as the one for the identity function \( id_X \).
But in the case of \( i_Y^X \), the \( x \) on the left of the equality sign is viewed as an element of the set \( X \), whereas on the right hand side it is viewed as an element of the set \( Y \).

This provides an example of two functions with the same domain and taking the same values without being the same function.

Example 1.27. Let \( X \) be a proper subset of \( Y \), \( X \subset Y \), then the two functions
\[
\begin{align*}
id_X : X & \longrightarrow X, \quad x \longmapsto x \\
i_Y^X : X & \longrightarrow Y, \quad x \longmapsto x
\end{align*}
\]
have the same domain and are given by the same formula, whence they have precisely the same image. Yet they are different functions, even though the only difference between them lies in the values they do not take.

We have just characterised subsets of a fixed set \( X \) in terms of functions with co-domain \( X \), assigning a specific function to each subset of \( X \). We can also characterise subsets of \( X \) in terms of functions whose domain is \( X \).

Definition 1.28. Given a subset \( A \) of the set \( X \), we define its characteristic function, \( \chi_A \), by
\[
\chi_A : X \longrightarrow \mathbb{2}, \quad x \longmapsto \begin{cases} 1 & \text{if } x \in A \\ 0 & \text{if } x \not\in A \end{cases}
\]
A little thought shows that these characteristic functions determine all subsets of \( X \) and distinguish between them. We shall express this more formally later.

Definition 1.29. Given a function \( f : X \longrightarrow Y \) and subset \( A \) of \( X \), the restriction of \( f \) to \( A \), is the function
\[
f |_A : A \longrightarrow Y, \quad x \longmapsto f(x)
\]
Note that unless of course \( A = X \), this is not the same function as \( f \), even though the two functions agree everywhere they are both defined.

Functions can sometimes be composed.

Definition 1.30. Given functions \( f : X \longrightarrow Y \) and \( g : Y \longrightarrow Z \) their composition, denoted by \( g \circ f \), is the function defined by
\[
g \circ f : X \longrightarrow Z, \quad x \longmapsto g(f(x)),
\]
as long as \( \text{codom}(f) = \text{dom}(g) = Y \).

In such a case,
\[
\begin{align*}
\text{dom}(g \circ f) &= \text{dom}(f) \\
\text{codom}(g \circ f) &= \text{codom}(g) \\
\text{im}(g \circ f) &\subseteq \text{im}(g)
\end{align*}
\]
Equality need not hold in (\( \ast \)). To see this consider the functions
\[
\begin{align*}
f : \mathbb{R} & \longrightarrow \mathbb{R}, \quad x \longmapsto 1 \\
g : \mathbb{R} & \longrightarrow \mathbb{R}, \quad y \longmapsto y.
\end{align*}
\]
Clearly \( \text{im}(g \circ f) = \{1\} \neq \mathbb{R} = \text{im}(g) \).
Observation 1.31. Given $A \subseteq X$ and the function $f: X \to Y$, the restriction $f \mid_A: A \to Y$ is, in fact, the composition of $f$ with the inclusion of $A$ into $X$:

$$f \mid_A = f \circ i_A$$

Composition of functions is associative.

Theorem 1.32. Given functions $h: W \to X, g: X \to Y$ and $f: Y \to Z$,

$$(f \circ g) \circ h: W \to Z \quad \text{and} \quad f \circ (g \circ h): W \to Z$$

are the same function.

Proof. Let $w \in W$,

$$((f \circ g) \circ h)(w) := (f \circ g)(h(w)) := f(g(h(w))) =: f((g \circ h)(w)) =: f((f \circ (g \circ h))(w))$$

Thus, the domains and co-domains agree. Since for all $w \in W$,

$$(f \circ g) \circ h = f \circ (g \circ h).$$

The identity functions act as neutral elements with respect to the composition of functions.

Theorem 1.33. Let $f: X \to Y$ be any function. Then

$$i_Y \circ f = f = f \circ i_X$$

Proof. The result follows directly from the definitions.

Theorem 1.36. If $f: X \to Y$ has both a left and a right inverse, then these must be the same, and then $f$ is invertible with uniquely determined inverse.
Attempting to decide, from first principles, whether the function $f$ has an inverse, would involve all functions $Y 	o X$ and see which, if any, satisfy the conditions in the definition. It would be preferable to be able to determine from intrinsic properties of $f$ whether it admits an inverse. Such an intrinsic criterion is available, as we show. To do so, we introduce some important properties of functions.

**Definition 1.37.** The function $f : X \to Y$ is

(i) 1-1 or injective or mono if and only if $x = x'$ whenever $f(x) = f(x')$;
(ii) onto or surjective or epi if and only if given any $y \in Y$ there is an $x \in X$ with $f(x) = y$ — in other words $\text{im}(f) = \text{codom}(f)$;
(iii) 1-1 and onto or bijective or iso if and only if it is both 1-1 and onto.

Thus a function is injective if and only if it distinguishes different elements of its domain: different elements of its domain are mapped to different elements of its co-domain.

Similarly, a function is surjective if and only if its image coincides with its co-domain.

We can formulate injectivity and subjectivity in terms of solving equations or in terms of the fibres of a function.

**Lemma 1.38.** Let $f : X \to Y$ be a function.

(a) The following statements are equivalent.

(i) $f$ is injective.
(ii) For each $y \in Y$, the equation $y = f(x)$ has at most one solution $x \in X$.
(iii) Each fibre, $f^{-1}(y)$ contains at most one element.

(b) The following statements are equivalent.

(i) $f$ is surjective.
(ii) For each $y \in Y$, the equation $y = f(x)$ has at least one solution $x \in X$.
(iii) Each fibre, $f^{-1}(y)$ contains at least one element.

**Proof.** Immediate from the definitions. □

**Example 1.39.** Let $\mathbb{R}_0^+ := \{ x \in \mathbb{R} \mid x \geq 0 \}$.

(i) $f : \mathbb{R} \to \mathbb{R}$, $x \mapsto x^2$ is neither injective nor surjective, as $f(1) = f(-1)$ and there is no $x$ with $f(x) = -4$.
(ii) $g : \mathbb{R} \to \mathbb{R}_0^+$, $x \mapsto x^2$ is not injective, but it is surjective, as $f(1) = f(-1)$ and every non-negative real number can be written as the square of a real number.
(iii) $h : \mathbb{R}_0^+ \to \mathbb{R}$, $x \mapsto x^2$ is injective, but it is not surjective, as $f(x) = f(x')$ if and only if $x^2 = x'^2$ if and only if $x' = \pm x$ if and only if $x' = x$ since, by definition, $x, x' \geq 0$. On the other hand, there is no $x$ with $f(x) = -4$.
(iv) $k : \mathbb{R}_0^+ \to \mathbb{R}_0^+$, $x \mapsto x^2$ is both injective and surjective, as is clear from parts (ii) and (iii).

The differences between these functions are illustrated by their graphs.
THEOREM 1.40. Take a non-empty set $X$. The function $f: X \rightarrow Y$ has

(i) a left inverse if and only if it is injective (or 1–1),
(ii) a right inverse if and only if it is surjective (or onto) and
(iii) an inverse if and only if it is bijective.

PROOF. (i) Suppose that $f: X \rightarrow Y$ has $g: Y \rightarrow X$ as left inverse.

Given $x, x' \in X$ with $f(x) = f(x')$,

\[
\begin{align*}
  x &=: \text{id}_X(x) \\
  &= (g \circ f)(x) \quad \text{as } g \text{ is left inverse to } f \\
  &= g(f(x)) \\
  &= g(f(x')) \quad \text{as } f(x) = f(x') \\
  &=: (g \circ f)(x') \\
  &= \text{id}_X(x') \quad \text{as } g \text{ is left inverse to } f \\
  &= x',
\end{align*}
\]

showing that $f$ is injective.

Conversely, suppose that $f: X \rightarrow Y$ is injective. Since $X \neq \emptyset$, we may choose $x_0 \in X$. Define $g: Y \rightarrow X$ by

\[
g(y) := \begin{cases} 
  x & \text{if } y = f(x) \\
  x_0 & \text{otherwise.}
\end{cases}
\]

It is immediate that $g \circ f = \text{id}_X$, so it only remains to show that $g$ is, in fact, a function.

The domain and co-domain of $g$ are $Y$ and $X$ respectively, so the only possible difficulty is that $g$ might assign more than one element of $X$, say $x$ and $x'$, to some element $y$ of $Y$.

By the definition of $g$, this could only happen when $y \in \text{im}(f)$ and then we would have $y = f(x) = f(x')$. Since $f$ is injective, this forces $x' = x$, showing that $g$ is a function.

(ii) Let $g: Y \rightarrow X$ be right inverse to $f: X \rightarrow Y$.

Take $y \in Y$ and put $x := g(y)$. Then, since

\[
\begin{align*}
f(x) &= f(g(y)) \\
  &=: (f \circ g)(y) \\
  &= \text{id}_Y(y) \quad \text{as } g \text{ is right inverse to } f \\
  &= y,
\end{align*}
\]

$f$ is surjective.

For the converse, suppose that $f: X \rightarrow Y$ is surjective. For each $y \in Y$, choose an $x_y \in X$ with $f(x_y) = y$. This is always possible is because $f$ is surjective. Define

\[
g: Y \rightarrow X, \quad y \mapsto x_y
\]

This $g$ is obviously a function.

For $y \in Y$, $(f \circ g)(y) = f(g(y)) = f(x_y) = y$, by the definition of $g$, so that $f \circ g = \text{id}_Y$. 
(iii) This follows from Theorem 1.36 and parts (i) and (ii) here.

\[ \text{□} \]

**Notation.** We write \( X \cong Y \) to denote that there is a bijection between \( X \) and \( Y \). We sometimes write \( f: X \overset{\cong}{\longrightarrow} Y \) when \( f \) is bijective.

We can express whether \( f \) is injective or surjective in terms of its fibres.

**Lemma 1.41.** *The function \( f: X \longrightarrow Y \) is*

(i) surjective if and only if each fibre \( f^{-1}(y) \) \((y \in Y)\) is non-empty;

(ii) injective if and only if each fibre \( f^{-1}(y) \) \((y \in Y)\) contains at most one element.

**Proof.** The statements simple restate the definitions. \[ \text{□} \]

**Observation 1.42.** Recall from Example 1.27 that, if \( X \subset Y \), then

\[ \text{id}_X: X \longrightarrow X, \quad x \mapsto x \]

\[ i_X^Y: X \longrightarrow Y, \quad x \mapsto x \]

are different functions, even though they share a common domain and agree at every point.

Theorem 1.40 provides a reason, other than formal definition. For while \( \text{id}_X \) is invertible — it is its own inverse — \( i_X^Y \) cannot be invertible, since, failing to be surjective, it cannot have a right inverse.

A number of significant mathematical results essentially reduce to asserting that there is a bijection between two sets (although the bijection may also be required to satisfy additional conditions). Our next theorem provides an illustration. This theorem is the more precise formulation promised earlier of the statement that the characteristic functions determine and distinguish the subsets of a given set.

**Theorem 1.43.** *The function which assigns to each subset of a fixed set its characteristic function is bijective. Formally, for any set \( X \)

\[ \chi: \wp(X) \longrightarrow 2^X, \quad A \mapsto \chi_A \]

is a bijection.*

**Proof.** The two things to check are that \( \chi \) is, indeed, a function, and that it is bijective. The details are left to the reader as an exercise. \[ \text{□} \]

**Observation 1.44.** While only some functions are surjective or injective, every function has a *mono-epi factorisation.** that is, every function can be written as a surjective function followed by an injective one.

Given the function \( f: X \longrightarrow Y \), put \( R := \text{im}(f) \).

Since, by definition, \( R \subset Y \), we have the function \( i_R^Y: R \longrightarrow Y \), which we abbreviate to \( i \).

Defining \( \hat{f}: X \longrightarrow R \), \( x \mapsto f(x) \), it is easy to see that

(i) \( i \) is injective, that is, a mono,

(ii) \( \hat{f} \) is surjective, that is, an epi, and

(iii) \( f = i \circ \hat{f} \).

This factorisation justifies the common practice of being sloppy with the composition of functions, allowing functions \( f \) and \( g \) to be composed as long as \( \text{im}(f) \subset \text{dom}(g) \) even if \( \text{codom}(f) \neq \text{dom}(g) \).
The discussion above accounts for this, for if \( \text{im}(f) \subseteq \text{dom}(g) \), and we abbreviate the function 
\[ i_{\text{im}(f)}^{\text{dom}(g)} : \text{im}(f) \to \text{dom}(g) \] to \( i \), we can compose \( g \) with \( i \circ \hat{f} \), to form
\[ g \circ i \circ \hat{f} : \text{dom}(f) \to \text{codom}(g) \]
It is this function which is commonly abbreviated to \( g \circ f \), just as we often write \( f \mid_A \) instead of \( f \circ i_A^X : A \to \text{codom}(f) \) when \( A \subseteq X = \text{dom}(f) \).

4. Commutative Diagrams

We often depict functions using diagrams, with each set represent by a vertex and each function \( X \to Y \) represented by an arrow from the vertex representing \( X \) to the vertex representing \( Y \).

**Definition 1.45.** A diagram commutes if and only if for any two vertices in the diagram any two paths from one to the other represent the same function.

**Example 1.46.** The diagram

\[
\begin{array}{ccc}
A & \xrightarrow{f} & B \\
\downarrow{h} & & \downarrow{g} \\
C & & \end{array}
\]

commutes when \( h = g \circ f \), in other words, when the composition \( g \circ f \) coincides with \( h \). Similarly, the diagram

\[
\begin{array}{ccc}
A & \xrightarrow{f} & B \\
\downarrow{j} & & \downarrow{g} \\
C & \xrightarrow{k} & D \\
\end{array}
\]

commutes when \( k \circ j = g \circ f \), that is when the two compositions \( g \circ f \) and \( k \circ j \) coincide.

We can use commutative diagrams to express the fact that the composition of functions is associative. If \( f : W \to X \), \( g : X \to Y \) and \( h : Y \to Z \) are functions then
\[ (h \circ g) \circ f = h \circ (g \circ h) \]
if and only if the diagram

\[
\begin{array}{ccc}
W & \xrightarrow{f} & X \\
\downarrow{g \circ f} & & \downarrow{h \circ g} \\
Y & \xrightarrow{h} & Z \\
\end{array}
\]

commutes, or, equivalently, that

\[
\begin{array}{cc}
W & \xrightarrow{f} & X \\
\downarrow{g \circ f} & & \downarrow{h \circ g} \\
Y & \xrightarrow{g} & Z \\
\end{array}
\]

commutes.
5. EQUIVALENCE RELATIONS AND PARTITIONS

For many purposes, two distinct objects can be indistinguishable, or the differences irrelevant: we treat them as equivalent. We next define this notion formally.

A relation between the elements of the set $X$ and those of $Y$ can be represented by the subset of $X \times Y$ comprising those pairs $(x, y)$ ($x \in X$, $y \in Y$) such that $x$ stands in the relation $R$ to $y$. We often write $xRy$ to denote this.

An example is provided by the telephone book. Here we regard $X$ as the set of all subscribers, and $Y$ as all telephone numbers.

If $Y$ happens to coincide with $X$, we speak of a binary relation on $X$.

**Definition 1.47.** An equivalence relation on $X$ is a binary relation $\sim$ on $X$, which is reflexive, symmetric and transitive. That is to say, given any $x, y, z \in X$,

- **Reflexivity:** $x \sim x$
- **Symmetry:** $x \sim y$ if and only if $y \sim x$.
- **Transitivity:** If $x \sim y$ and $y \sim z$, then $x \sim z$.

Given an equivalence relation $\sim$ on $X$, and $x \in X$, we define the equivalence class of $x$ by $[x] := \{ t \in X \mid x \sim t \}$.

Any $z \in [x]$ a representative of $[x]$.

Finally, quotient set, denoted by $X/\sim$, is the set of all such equivalence classes, so that $X/\sim := \{ [x] \mid x \in X \}$.

We then have a function, called the natural map, the quotient function or the quotient map $\eta: X \rightarrow X/\sim$, $x \mapsto [x]$.

**Example 1.48.** Each positive integer, $n$ can be used to define and equivalence relation on $\mathbb{Z}$:

The integers $x$ and $y$ are congruent modulo $n$, written

$$x \equiv y \pmod{n}$$

if and only if $x - y$ is divisible by $n$, or equivalently, if and only if there is an integer, $k$ with

$$x - y = kn.$$

The set of all equivalence classes is usually referred to as the set of integers modulo $n$ and denoted by $\mathbb{Z}/n\mathbb{Z}$.

The reader has met this example before, when studying modular arithmetic.

**Example 1.49.** The function $f: X \rightarrow Y$ induces a natural equivalence relation, $\sim$, on $X$, namely,

$$x \sim x' \text{ if and only if } f(x) = f(x')$$

The equivalence classes are precisely the non-empty fibres of $f$.

Another important notion is that of a partition of a set.

**Definition 1.50.** A partition of the set $X$ is a set, $\{X_\lambda \mid \lambda \in \Lambda\}$, of pairwise disjoint non-empty subsets of $X$, whose union is $X$. Thus $\{X_\lambda \mid \lambda \in \Lambda\}$ is a partition of $X$ if and only if

1. $\emptyset \subset X_\lambda \subseteq X$ for each $\lambda \in \Lambda$;
2. $X_\lambda \cap X_\mu = \emptyset$ whenever $\lambda \neq \mu$;
3. $X = \bigcup_{\lambda \in \Lambda} X_\lambda$.

However different they may appear, an equivalence relation on a set is essentially the same as a partition of that set.
Theorem 1.51. There is a one-to-one correspondence between equivalence relations on a set and partitions of that set. In other words, every equivalence relation on the set $X$ determines a unique partition of $X$, and vice versa.

Proof. We outline a proof, leaving it as an exercise for the reader to provide any missing details.

Given the equivalence relation $\sim$ on $X$, the equivalence classes form a partition of $X$, that is every $x \in X$ belongs to some equivalence class, and if $[x] \cap [x'] \neq \emptyset$, then $[x] = [x']$.

If $\{X_\lambda \mid \lambda \in \Lambda\}$ is a partition of $X$, then

$$x \sim x' \text{ if and only if } x, x' \in X_\lambda \text{ for some } \lambda \in \Lambda$$

defines an equivalence relation on $X$.

If we start with an equivalence relation, construct the associated partition, then the associated equivalence relation is the original one.

Finally, if we start with a partition, define the associated equivalence relation, then the associated partition is the original one. $\square$

Example 1.52. The function $f : X \longrightarrow Y$ on the non-empty set $X$, induces a natural partition of $X$, by taking the non-empty fibres of $f$,

$$f^{-1}(y) \ (y \in \text{im}(f))$$

as the partitioning subsets (cf. Example 1.49).

Observation 1.53. The quotient of the set by the equivalence relation illustrates ambiguities which can arise through notational convenience.

Given a function $f : X \longrightarrow Y$ we assign to each $B \subseteq Y$, its pre-image, or inverse image,

$$f^{-1}(B) := \{ x \in X \mid f(x) \in B \}$$

using the convention that when $B$ is the singleton set $\{y\}$, we write $f^{-1}(y)$ in lieu of $f^{-1}(\{y\})$.

Given the equivalence relation, $\sim$, on the set $X$, we have a natural surjective function

$$\eta : X \longrightarrow X/\sim, \quad x \longmapsto [x],$$

with $[x]$ denoting the equivalence class of $x$. Using conventional notation, we then have

$$\eta^{-1}([x]) = [x].$$

Here, the $[x]$ on the left hand side is an element of the set $X/\sim$, whereas the $[x]$ on the right hand side is a subset of $X$. The situation is clarified immediately if we eschew notational convenience and use precise notation, to obtain

$$\eta^{-1}(\{[x]\}) = [x].$$

While this averts potential ambiguity, the precision comes at the cost of readability. As a result, systematic ambiguity has become thoroughly entrenched in mathematics:

(i) whenever the emphasis is on clear delineation, precision has priority;
(ii) once clarity has been established, it is customary to revert to convenience and readability, unless there is genuine danger of serious confusion, trusting the good sense of the reader to discern from the context which interpretation is intended.

We shall use systematic ambiguity judiciously.
6. Universal Properties

We make use of universal properties in this course. A *universal property* is one which holds for all members of a class. These are extremely useful in mathematics, since they may be used in quite general settings and anything enjoying a universal property in a suitable context is uniquely defined by it.

We provide two examples to clarify the terminology and establish the uniqueness in each case, namely, the Cartesian product of two sets and the construction of the quotient set from an equivalence relation on a given set.

6.1. The Cartesian Product of Two Sets. Recall that $X \times Y$, the Cartesian product of the sets $X$ and $Y$, is defined by

$$X \times Y := \{(x, y) \mid x \in X, y \in Y\}.$$ 

In other words, $X \times Y$ is the set of all ordered pairs, the first member of which is an element of $X$ and the second an element of $Y$.

One consequence of this definition is that the Cartesian product of the sets $X$ and $Y$ comes equipped with two *natural maps*, namely, the *canonical projections onto the factors*,

$$pr_X : X \times Y \rightarrow X, \quad (x, y) \mapsto x$$

$$pr_Y : X \times Y \rightarrow Y, \quad (x, y) \mapsto y$$

The Cartesian product, together with the canonical projections, enjoys a universal property.

Given any set and any pair of functions from that set, one into the first factor of the Cartesian product, the other into the second, there is a uniquely determined function from the given set to the Cartesian product such that the given functions agree with the new function composed with the respective canonical projections.

More formally,

**Theorem 1.54.** Given any set $W$ and any pair of functions, $f : W \rightarrow X$ and $g : W \rightarrow Y$, there is a unique function $h : W \rightarrow X \times Y$ such that

$$f = pr_X \circ h$$

$$g = pr_Y \circ h$$

This is expressed diagrammatically by

```
W \xrightarrow{\exists! h} X \times Y \xleftarrow{pr_X} X \xleftarrow{pr_Y} Y
```

or, equivalently, by

```
\begin{tikzcd}
Y & Y \\
& X \times Y \\
W \arrow{r}{\exists! h} \arrow{u}{g} & X \times Y \arrow{u}{pr_Y} \arrow{d}{pr_X} \\
& X \arrow{u}{pr_Y}
\end{tikzcd}
```
We prove the theorem by choosing the obvious candidate for $h$ and then verifying that it satisfies all the requirements. Since this is likely to be the first time the reader has met universal properties, we present a detailed proof.

**Proof of Theorem 1.54.** A function, $h: W \rightarrow X \times Y$, assigns to each $w \in W$ a uniquely determined $h(w) \in X \times Y$.

By the definition of $X \times Y$, $h(w)$ must be of the form $(x_w, y_w)$ for some uniquely determined $x_w \in X$ and uniquely determined $y_w \in Y$: $h(w) = (x_w, y_w)$ for some $x_w \in X, y_w \in Y$.

Applying the canonical projections, we obtain see that
\[
(pr_X \circ h)(w) = pr_X(x_w, y_w) = x_w \\
(pr_Y \circ h)(w) = pr_X(x_w, y_w) = y_w
\]

Hence, the only way for the function $h: W \rightarrow X \times Y$ to satisfy $f = pr_X \circ h$ and $g = pr_Y \circ h$, is to have $x_w = f(w)$ and $y_w = g(w)$. Thus, $h$ must be given by
\[
h(w) := (f(w), g(w))
\]

This is establishes the uniqueness of $h$, and it only remains to verify that $h(w) := (f(w), g(w))$ does, in fact, define a function $h: W \rightarrow X \times Y$.

To achieve this, we must show that when we take an arbitrary element $w$ of $W$ $h$ assigns to it a uniquely determined element of $X \times Y$.

So take any $w \in W$.

Since $f$ is a function, it assigns to $w$ the uniquely determined element $f(w)$ of $X$.

Since $g$ is also a function, it assigns to $w$ the uniquely determined element $g(w)$ of $Y$.

By the definition of $X \times Y$, these determine uniquely the element $(f(w), g(w))$ of $X \times Y$.

Thus $h$ assigns to $w$ the uniquely determined element $(f(w), g(w))$ of $X \times Y$, as required. \(\square\)

**Observation 1.55.** Theorem 1.54 asserts that for given sets $X$ and $Y$, a set $X \times Y$, together with functions $pr_X: X \times Y \rightarrow X$ and $pr_Y: X \times Y \rightarrow Y$, can be found with the prescribed property. In other words, the property is enjoyed not by $X \times Y$ alone, but by $X \times Y$ together with $pr_X: X \times Y \rightarrow X$ and $pr_Y: X \times Y \rightarrow Y$.

**Observation 1.56.** The property is universal in the sense that it applies to each and every set $W$, each and every function $f: W \rightarrow X$ and each and every function $g: W \rightarrow Y$.

We show that the universal property in Theorem 1.54 uniquely determines the Cartesian product of two sets and the canonical projections up to a unique equivalence.

To see this, let $W$, $q_X: W \rightarrow X$, $q_Y: W \rightarrow Y$ also exhibit the same universal property. The following commutative diagram suffices as proof.
To see how this constitutes a proof, we do some “diagram-chasing”, starting at the left-hand copy of \( W \). (The reader is strongly advised to work through this proof carefully, with pencil and paper at hand, checking each step and drawing each diagram.)

Let \( f: W \to X \) be the composition \( \text{id}_X \circ q_X \) and \( g: W \to Y \) be the composition \( \text{id}_Y \circ q_Y \). In other words, we look at the following part of our diagram:

\[
\begin{array}{c}
Y & \xrightarrow{id_Y} & Y & \xrightarrow{id_Y} & Y & \xrightarrow{id_Y} & Y \\
q_Y & & pr_Y & & q_Y & & pr_Y \\
W & \xrightarrow{X \times Y} & \exists! \phi & \xrightarrow{W} & \exists! \phi & \xrightarrow{X \times Y} \\
q_X & & pr_X & & q_X & & pr_X \\
X & \xrightarrow{id_X} & X & \xrightarrow{id_X} & X & \xrightarrow{id_X} & X \\
\end{array}
\]

Then we have two functions defined on \( W \), one taking values in \( X \), the other in \( Y \).

By the universal property of \((X \times Y, \text{pr}_X: X \times Y \to X, \text{pr}_Y: X \times Y \to Y)\), there is a unique function \( \phi: W \to X \times Y \) such that

\[
\begin{align*}
\text{pr}_X \circ \phi &= q_X \\
\text{pr}_Y \circ \phi &= q_Y
\end{align*}
\]

yielding

\[
\begin{array}{c}
Y & \xrightarrow{id_Y} & Y & \xrightarrow{id_Y} & Y & \xrightarrow{id_Y} & Y \\
q_Y & & pr_Y & & q_Y & & pr_Y \\
W & \xrightarrow{\phi} & X \times Y & \xrightarrow{\exists! \psi} & W & \xrightarrow{\phi} & X \times Y \\
q_X & & pr_X & & q_X & & pr_X \\
X & \xrightarrow{id_X} & X & \xrightarrow{id_X} & X & \xrightarrow{id_X} & X \\
\end{array}
\]

Similarly, starting from the left-hand copy of \( X \times Y \) and look at the following part of our diagram:

\[
\begin{array}{c}
Y & \xrightarrow{id_Y} & Y & \xrightarrow{id_Y} & Y & \xrightarrow{id_Y} & Y \\
q_Y & & pr_Y & & q_Y & & pr_Y \\
W & \xrightarrow{\phi} & X \times Y & \xrightarrow{\exists! \psi} & W & \xrightarrow{\phi} & X \times Y \\
q_X & & pr_X & & q_X & & pr_X \\
X & \xrightarrow{id_X} & X & \xrightarrow{id_X} & X & \xrightarrow{id_X} & X \\
\end{array}
\]

we find a unique \( \psi: X \times Y \to W \) such that

\[
\begin{align*}
q_X \circ \psi &= \text{pr}_X \\
q_X \circ \psi &= \text{pr}_Y
\end{align*}
\]
We next show that these two (uniquely determined) functions, \( \varphi \) and \( \psi \) are, in fact, mutually inverse, and so, by Theorem 1.40, each is a bijection.

To do so, we show that \( \psi \circ \varphi = id_W \) and \( \varphi \circ \psi = id_{X \times Y} \). (Actually, we only establish the first of these equalities, since the second follows \textit{mutatis mutandis}.)

We again start from the left-hand copy of \( W \).

This time we let \( f : W \to X \) be the composition \( id_X \circ id_X \circ q_X \) and \( g : W \to Y \) the composition \( id_Y \circ id_Y \circ q_Y \).

By the universal property of \((W, q_X : W \to X, q_Y : W \to Y)\), there is a unique function \( \vartheta : W \to W \) such that
\[
q_X \circ \vartheta = f = id_X \circ id_X \circ q_X
\]
\[
q_Y \circ \vartheta = g = id_Y \circ id_Y \circ q_Y
\]
yielding
\[
Y \xrightarrow{id_Y} Y \xrightarrow{id_Y} Y
\]
\[
W \xrightarrow{q_Y} X \times Y \xrightarrow{\psi} W
\]
\[
X \xrightarrow{id_X} X
\]

Since, plainly, \( id_W \) shares this property with \( \vartheta \), the uniqueness forces
\[
\vartheta = id_W
\]

We now consider the composition
\[
\psi \circ \varphi : W \to W
\]
as depicted by the commutative diagram
By the calculations above,
\[ q_X \circ \psi \circ \varphi = pr_X \circ \varphi = q_X \]
\[ q_Y \circ \psi \circ \varphi = pr_Y \circ \varphi = q_Y \]
Since \( \psi \circ \varphi \) shares this property with \( \vartheta \), uniqueness implies that
\( \psi \circ \varphi = \vartheta = id_W \).

We apply an analogous argument to \( \varphi \circ \psi: X \times Y \longrightarrow X \times Y \) starting with the left-most copy of \( X \times Y \) and conclude that \( \varphi \circ \psi = id_{X \times Y} \).

We include here the relevant diagram, but leave the finer details to the reader.

This completes the proof.

We have executed this last piece of “diagram chasing” in detail because this is a common method of proof in modern topology, algebra and their applications.

Observation 1.57. The last result asserts that there is a bijection between maps from \( W \) to \( X \times Y \) and pairs of maps, the first from \( W \) to \( X \) and the second from \( W \) to \( Y \). Writing \( B^A \) for the set of all functions with domain \( A \) and co-domain, we can formulate this as
\[ (X \times Y)^W \cong (X^W) \times (Y^W) \]

Observation 1.58. This universal property has been described purely in terms of functions between sets, without any mention of the elements of the sets. Hence we can ask whether there are objects of interest displaying the analogous universal property in any situation which can be represented by diagrams similar to the one above.

For example, if we take \( X \) and \( Y \) to be real vector spaces and insist that the arrows in our diagram represent linear transformations (rather than just functions), then the analogous situation consists of being given two real vector spaces \( X \) and \( Y \) and seeking a real vector space \( X \times Y \) and linear
transformations \( pr_X: X \times Y \to X \), \( pr_Y: X \times Y \to Y \) such that given any real vector space and any linear transformations \( f: W \to X \) and \( g: W \to Y \) there is a unique linear transformation \( h: W \to X \times Y \) with \( f = pr_X \circ h \) and \( g = pr_Y \circ h \). It is well known from basic linear algebra, that the direct sum of two vector spaces, together with the projections onto the direct summands, fills the bill.

### 6.2. Quotient Sets.

Given an equivalence relation \( \sim \) on the set \( X \), the quotient set \( X/\sim \) and the quotient map \( \eta: X \to X/\sim \) also enjoy a universal property:

**Theorem 1.59.** Let \( \sim \) be an equivalence relation on the set \( X \).

Given any set \( Y \) and any function \( f: X \to Y \) with the property that \( f(x) = f(x') \) whenever \( x \sim x' \) there is a uniquely determined function \( \tilde{f}: X/\sim \to Y \) such that \( f = \tilde{f} \circ \eta \).

This is expressed by the commutative diagram

\[
\begin{array}{ccc}
X & \xrightarrow{f} & Y \\
\eta \downarrow & & \downarrow \exists \tilde{f} \\
X/\sim & & \\
\end{array}
\]

**Proof.** Take \( f: X \to Y \) with \( f(x) = f(x') \) whenever \( x \sim x' \).

In order for \( \tilde{f}: X/\sim \to Y \) to satisfy \( f(x) = (\tilde{f} \circ \eta)(x) \) for all \( x \in X \), we must have

\[
\tilde{f}([x]) = \tilde{f}(\eta(x)) = (\tilde{f} \circ \eta)(x) = f(x)
\]

This establishes the uniqueness of a \( \tilde{f} \) rendering the required diagram commutative. It only remains to show that

\( \tilde{f}: X/\sim \to Y \), \( [x] \mapsto f(x) \)

is actually a function.

Since the definition of \( \tilde{f}([x]) \) depends on the choice of has representative \( x \in X \), the only way in which \( \tilde{f} \) could fail to be a function is that for some \( x, x' \in X, [x] = [x'] \), but \( f(x) \neq f(x') \).

However \( [x] = [x'] \) if and only if \( x \sim x' \), in which case \( f(x) = f(x') \) by our assumption on \( f \).

**Example 1.60.** We revisit Observation 1.44.

Let \( f: X \to Y \) be a function defined on the non-empty set \( X \).

We saw in Example 1.49 that \( f \) defines the equivalence relation \( \sim \) on \( X \), where \( x \sim x' \) if and only if \( f(x) = f(x') \)

and the equivalence classes are precisely the non-empty fibres, so that \( [x] = f^{-1}(f(x)) \)

Hence, there is a bijection

\( \theta: X/\sim \to \text{im}(f) \), \( [x] \mapsto f(y) \)

between \( X/\sim \) and \( \text{im}(f) \).
If we use this to identify $X/\sim$ and $\text{im}(f)$, then the natural map
\[ \eta: X \to X/\sim \]
becomes (or, more precisely, is identified with) the function
\[ \hat{f}: X \to X/\sim, \quad x \mapsto f(x) \]
introduced in Observation 1.44, and the function
\[ \tilde{f}: X/\sim \to Y \]
obtained by the universal property of the quotient construction becomes (or, more precisely, is identified with)
\[ \iota_{\text{im}(f)}: \text{im}(f) \to Y, \quad y \mapsto y \]
Hence, the mono-epi factorisation of $f: X \to Y$ is an application of the universal property of the quotient construction.

7. Exercises

1.1. Prove that the set $X$ is a subset of the set $Y$ if and only if
\[ X \to Y, \quad x \mapsto x \]
defines a function.

1.2. Prove that $A \neq \emptyset$ is a proper subset of the set $X$ if and only if the function
\[ \chi_A: X \to \mathbb{2}, \quad x \mapsto \begin{cases} 1 & \text{if } x \in A \\ 0 & \text{if } x \notin A \end{cases} \]
is surjective.

1.3. Prove that for any set $X$
\[ \chi: \mathcal{P}(X) \to \mathbb{2}^X, \quad A \mapsto \chi_A \]
is a bijection.

1.4. Take a function $f: X \to Y$, $A \subseteq X$ and $B \subseteq Y$, prove the following statements.
   (i) $A \subseteq f^{-1}(f(A))$.
   (ii) $f(f^{-1}(B)) \subseteq B$.
   (iii) In general, equality need not hold in either (i) or (ii).
   (iv) $G = f^{-1}(f(G))$ for every subset $G$ of $X$ if and only if $f$ is injective (1–1).
   (v) $f(f^{-1}(H)) = H$ for every subset $H$ of $Y$ if and only if $f$ is surjective (onto).

1.5. Take functions $f: X \to Y$ and $g: Y \to Z$. Prove the following statements.
   (a) If $f$ and $g$ are both injective, then so is $g \circ f$.
   (b) If $f$ and $g$ are both surjective, then so is $g \circ f$.
   (c) If $g \circ f$ is injective, then so is $f$, but not necessarily $g$.
   (d) If $g \circ f$ is surjective, then so is $g$, but not necessarily $f$.
   (e) If $f$ and $g$ are bijective, so is $g \circ f$.
   (f) If $g \circ f$ is bijective, then neither $f$ nor $g$ need be bijective.

1.6. Let $A, B, C$ and $D$ be sets. Determine the relationships between
   (i) $(A \times C) \cap (B \times D)$ and $(A \cap B) \times (C \cap D)$;
   (ii) $(A \times C) \cup (B \times D)$ and $(A \cup B) \times (C \cup D)$.

1.7. Given the function $f: X \to Y$ and subsets $G, H$ of $Y$, prove the following statements.
   (i) $f^{-1}(G \cap H) = f^{-1}(G) \cap f^{-1}(H)$.
   (ii) $f^{-1}(G \cup H) = f^{-1}(G) \cup f^{-1}(H)$. 
(iii) \( f^{-1}(G \setminus H) = f^{-1}(G) \setminus f^{-1}(H) \).
(iv) \( f^{-1}(Y \setminus G) = X \setminus f^{-1}(G) \).

1.8. Given the function \( f: X \to Y \) and subsets \( A, B \) of \( X \), find the relationship between the following pairs of subsets of \( Y \).

(i) \( f(A \cap B) \) and \( f(A) \cap f(B) \).
(ii) \( f(A \cup B) \) and \( f(A) \cup f(B) \).
(iii) \( f(A \setminus B) \) and \( f(A) \setminus f(B) \).
(iv) \( f(X \setminus A) \) and \( Y \setminus f(A) \).

1.9. Give an example of a non-constant function \( f: A \to \mathbb{R} \) where \( A \subseteq \mathbb{R} \) with \( \frac{df}{dx} = 0 \) everywhere on \( A \).

1.10. Give an example of a function \( f: A \to \mathbb{R} \) where \( A \subseteq \mathbb{R} \) which satisfies \( \frac{df}{dx} > 0 \) everywhere on \( A \), but is not monotonic increasing.

1.11. Give an example of a function \( f: A \to \mathbb{R} \) where \( A \subseteq \mathbb{R} \) which satisfies \( \frac{df}{dx} > 0 \) everywhere on \( A \), but is not 1–1.

1.12. \( A \) be a non-empty set of real numbers which is bounded above. Let \( b \) be the supremum (least upper bound) of \( A \). Prove that there is a sequence \( (a_n)_{n \in \mathbb{N}} \) in \( A \) with \( \lim_{n \to \infty} a_n = b \).

1.13. Let \( \sim \) be an equivalence relation on the set \( X \) and \( \eta: X \to X/\sim \) the quotient map. Prove that for any subset, \( A \), of \( X \)
\[
\eta^{-1}(\eta(A)) = \bigcup_{x \in A} [x]
\]

1.14. Given the function \( f: X \to Y \), define
\[
m: X \to X \times Y, \quad x \mapsto (x, f(x))
\]
\[
e: X \times Y \to Y, \quad (x, y) \mapsto y
\]
Prove that

(i) \( m \) is a mono;
(ii) \( e \) is an epi;
(iii) \( f = e \circ m \).

Observation 1.61. We saw in Observation 1.44 that every function has a mon-epi factorisation. This exercise shows that every function also has an epi-mono factorisation.

1.15. Prove that the Axiom of Choice is equivalent to the proposition that every surjective function has a right inverse.
This skipping is another important point. It should be done whenever a proof seems too hard or whenever a theorem or a whole paragraph does not appeal to the reader. In most cases he will be able to go on and later he may return to the parts which he skipped.

Emil Artin

Chapter 2

Continuity in $\mathbb{R}^n$ and Metric Spaces

We begin our quest to find the structure sets must support in order to be able to speak sensibly about the continuity of functions by examining the notion of continuity in the study of real-valued functions of a real variable, which is where continuity is first met when studying mathematics.

Definition 2.1. Given a subset $A$ of $\mathbb{R}$, the function $f : A \to \mathbb{R}$ is continuous at $a \in A$ if and only if for every positive real number, $\varepsilon$, there is a positive real number, $\delta$, such that $f(x)$ is within $\varepsilon$ of $f(a)$ whenever $x$ is within $\delta$ of $a$.

$f$ is continuous on $A$, or simply continuous, if and only if $f$ is continuous at every $a \in A$.

In other words, the function $f$ is continuous at $a$ if and only if given any tolerance whatsoever, we can find a deviation such that $f(x)$ is within that tolerance about $f(a)$ as long as $x$ is within the deviation about $a$.

Observation 2.2. This definition depends on having some measure of distance between elements of the set $\mathbb{R}$. The usual measure of distance between real numbers $x$ and $y$ is $|x - y|$, the absolute value of the algebraic difference between them.

Moreover, as this is the only property of the real numbers to play a rôle in Definition 2.1, we can generalise it to functions $f : X \to Y$ between sets $X$ and $Y$ as long as both have a measure of “distance” between their elements.

Our first step is to formalise what we mean by the distance between two elements of a set $X$. In doing so, we let familiar examples guide us.

Example 2.3. For our first example we return to the real numbers $\mathbb{R}$.

Let $X$ be any subset of $\mathbb{R}$ and take as distance between the elements $x_1$ and $x_2$ the absolute value of the difference between them, $|x_1 - x_2|$.

Of course, we could have taken $\mathbb{Q}$, the set of rational numbers, or $\mathbb{C}$, the set of complex numbers, in lieu of $\mathbb{R}$, without needing to modify anything else.

Example 2.4. The second example, generalising our first one, is from multivariate calculus.

We consider functions $f : X \to \mathbb{R}^n$, where $X$ is now taken to be a subset of $\mathbb{R}^n$, for $n, m \in \mathbb{N}$.

To simplify notation we write $x$ for the element $(x_1, \ldots, x_n)$ of $\mathbb{R}^n$.

The definition of the continuity of functions can be formulated in terms of distance.

The function $f : X \to \mathbb{R}^n$ is continuous at $a \in X$ if and only if given any non-zero distance, $\varepsilon$, there is a non-zero distance, $\delta$, so that $f(x)$ is within $\varepsilon$ of $f(a)$ if $x$ is within $\delta$ of $a$.

In multivariate calculus, the distance between two elements of $\mathbb{R}^n$ is taken to be the square root of the sum of the squares of differences between corresponding coordinates, or, equivalently, the
CONTINUITY IN $\mathbb{R}^n$ AND METRIC SPACES

squares of the absolute values of these differences. Writing $d_n(x, y)$ for the distance between $x$ and $y$, in $\mathbb{R}^n$, we have

$$d_n(x, y) := \sqrt{\sum_{j=1}^{n} |x_j - y_j|^2}.$$  

When $n = 1$, this definition coincides with that in Example 2.3, since if $a \in \mathbb{R}$, then $|a| = \sqrt{a^2} = \sqrt{|a|^2}$. Note that, once again, we could have replaced $\mathbb{R}$ by $\mathbb{Q}$ or $\mathbb{C}$ and continued mutatis mutandis.

These examples are sufficient for our purposes. We investigate the properties of $\varrho$, where write $\varrho(x, y)$ for the distance between $x$ and $y$.

- Since $\varrho$ assigns to each pair $(x, y)$ of elements $x, y$ of $X$ a single (non-negative) real number, it is a real-valued function of two variables. That is to say we have a function $\varrho : X \times X \rightarrow \mathbb{R}$.
- The distance between two distinct elements of $X$ is always strictly positive and the distance of an element of $X$ from itself is, of course, 0. That is to say, $\varrho(x, y) \geq 0$, with $\varrho(x, y) = 0$ if and only if $x = y$. We say that $\varrho$ is positive definite.
- The distance between two elements of $X$ is independent of the order in which we write them. That is to say $\varrho(y, x) = \varrho(x, y)$.
  (Heuristically, the distance between two points is independent of the direction in which we measure it.) We say that $\varrho$ is symmetric.
- The distance between any two elements of $X$ cannot exceed the sum of the distances between each of the two and a third one. That is to say, $\varrho(x, z) \leq \varrho(x, y) + \varrho(y, z)$.
  Note that since $\varrho(x, x) = 0$ always, this inequality is satisfied even when the elements $x, y$ and $z$ are not distinct. We say that $\varrho$ satisfies the triangle inequality.

We first show that the triangle inequality holds in Examples 2.3 and 2.4.

Put $a_j := x_j - y_j$ and $b_j := y_j - z_j$. Then

$$a_j + b_j = x_j - y_j + y_j - z_j = x_j - z_j$$

To show that $\varrho(x, z) \leq \varrho(x, y) + \varrho(y, z)$ is now equivalent to showing that

$$\sqrt{\sum_{j=1}^{n} (a_j + b_j)^2} \leq \sqrt{\sum_{j=1}^{n} a_j^2} + \sqrt{\sum_{j=1}^{n} b_j^2}.$$  

Since both sides of this inequality are non-negative, this is equivalent to showing that

$$\sum_{j=1}^{n} (a_j + b_j)^2 \leq \left( \sum_{j=1}^{n} a_j^2 + \sum_{j=1}^{n} b_j^2 \right)^2.$$  

Expanding both sides, we see that this is equivalent to showing that

$$\sum_{j=1}^{n} a_j b_j \leq \left( \sum_{j=1}^{n} a_j^2 \right) \left( \sum_{j=1}^{n} b_j^2 \right) \quad (*)$$
Since
\[ \sum_{j=1}^{n} a_j b_j \leq \sum_{j=1}^{n} |a_j| |b_j| = \sum_{j=1}^{n} |a_j||b_j| \]
it is enough to verify (\ast) when \( a_j, b_j \geq 0 \) for all \( j \).
This, in turn, is equivalent to showing that if \( a_j, b_j \geq 0 \) for all \( 1 \leq j \leq n \), then
\[ \left( \sum_{j=1}^{n} a_j b_j \right)^2 \leq \left( \sum_{j=1}^{n} a_j^2 \right) \left( \sum_{j=1}^{n} b_j^2 \right). \]

We now turn to showing that this last inequality holds.

**Lemma 2.5.** For non-negative real numbers \( x_1, \ldots, x_n, y_1, \ldots, y_n \),
\[ \left( \sum_{j=1}^{n} x_j y_j \right)^2 \leq \left( \sum_{j=1}^{n} x_j^2 \right) \left( \sum_{j=1}^{n} y_j^2 \right). \]

**Proof.** The function
\[ f: \mathbb{R} \rightarrow \mathbb{R}, \quad \lambda \mapsto \sum_{j=1}^{n} (x_j - \lambda y_j)^2. \]
is clearly never negative, since for every \( \lambda \in \mathbb{R} \), \( f(\lambda) \) is the sum of squares of real numbers.
Moreover
\[ f(\lambda) = \sum_{j=1}^{n} x_j^2 - 2\lambda \sum_{j=1}^{n} x_j y_j + \lambda^2 \sum_{j=1}^{n} y_j^2 \]
is a quadratic expression in \( \lambda \).
Since this is never negative, its discriminant cannot be positive.
In other words,
\[ \left( 2 \sum_{j=1}^{n} x_j y_j \right)^2 - 4 \left( \sum_{j=1}^{n} x_j^2 \right) \left( \sum_{j=1}^{n} y_j^2 \right) \leq 0, \]
which is equivalent to
\[ \left( \sum_{j=1}^{n} x_j y_j \right)^2 \leq \left( \sum_{j=1}^{n} x_j^2 \right) \left( \sum_{j=1}^{n} y_j^2 \right). \]
\[ \square \]

Using the above as our prototype, we now define what we mean by a measure of distance in a set, or, equivalently, a metric on a set.

**Definition 2.6.** A **metric on the set** \( X \) is a function
\[ \rho: X \times X \rightarrow \mathbb{R}_0^+ := \{ r \in \mathbb{R} \mid r \geq 0 \} \]
such that for all \( x, y, z \in X \)
\[ (i) \quad \rho(x, y) = 0 \text{ if and only if } x = y; \]
\[ (ii) \quad \rho(y, x) = \rho(x, y); \]
\[ (iii) \quad \rho(x, z) \leq \rho(x, y) + \rho(y, z). \]

A **metric space** consists of a set \( X \) endowed with a metric \( \rho \).
We write \( (X, \rho) \) to denote this.
CONVENTION. When there is no danger of confusion, we abbreviate \((X, \rho)\) to simply \(X\), leaving \(\rho\) understood.

**Example 2.7.** Given \(n \in \{1, 2, \ldots\}\), the function
\[
\epsilon_n: \mathbb{R}^n \times \mathbb{R}^n \longrightarrow \mathbb{R}_0^+, \quad ((x_1, \ldots, x_n), (y_1, \ldots, y_n)) \mapsto \sqrt{\sum_{j=1}^{n} (x_j - y_j)^2},
\]
is a metric, as we have just shown.
It is the *Euclidean metric* on \(\mathbb{R}^n\).

CONVENTION. When we refer to the metric space \(\mathbb{R}^n (n = 1, 2, \ldots)\), its metric is assumed to be the Euclidean metric, \(\epsilon_n\).

If our prototypes were the only examples of metric spaces, there would be little point to going to this much bother. The next example illustrates that metric spaces are genuinely more general than subsets of \(C^n\).

**Example 2.8.** Put
\[
I := \{t \in \mathbb{R} \mid 0 \leq t \leq 1\}.
\]
[\(I\) is the *closed unit interval* (in \(\mathbb{R}\)].

Take \(X\) to be the set of all continuous real-valued functions defined on \(I\), so that
\[
X := \{f: I \longrightarrow \mathbb{R} \mid f \text{ is continuous} \}
\]
and define
\[
\rho: X \times X \longrightarrow \mathbb{R}, \quad (f, g) \mapsto \max_{0 \leq t \leq 1} \{|f(t) - g(t)|\}
\]
It is left to the reader to verify that \((X, \rho)\) is, indeed, a metric space.

Since metric spaces are genuinely more general than subsets of \(\mathbb{R}^n\), it is natural to ask which sets support a metric. The surprising answer is that every set can be endowed with a metric.

**Definition 2.9.** The *discrete metric* on the set \(X\) is the function
\[
d_X: X \times X \longrightarrow \mathbb{R}_0^+, \quad (x, y) \mapsto \begin{cases} 0 & \text{for } x = y \\ 1 & \text{for } x \neq y \end{cases}
\]
We sometimes write \(d\) instead of \(d_X\) when there is no danger of confusion.

**Theorem 2.10.** Let \(X\) be a set. Then \(d_X\) is a metric on \(X\).

**Proof.** As \(d_X\) is positive definite and symmetric by definition, we need only verify the triangle inequality.

Take \(x, y, z\).

If \(z = x\), then \(d_X(x, z) = 0\), whence \(d_X(x, z) \leq d_X(x, y) + d_X(y, z)\) since \(d_X(x, y) + d_X(y, z) \geq 0\).

If \(z \neq x\), then either \(y \neq x\) of \(y \neq z\). It follows that
\[
d_X(x, y) + d_X(y, z) \geq 1 = d_X(x, z)
\]
□

Now that we have defined metric spaces, we recast our definition of continuity in terms of metrics and express it in formal mathematical language.
Definition 2.11. Given metric spaces \((X, \varrho)\) and \((Y, \sigma)\), the function \(f: X \rightarrow Y\) is \textit{continuous} at \(a \in X\) if and only if for every positive real number, \(\varepsilon\), there is a positive real number, \(\delta\), such that \(\sigma(f(x), f(a)) < \varepsilon\) whenever \(\varrho(x, a) < \delta\).

\(f\) is \textit{continuous on} \(X\) or simply \textit{continuous} if and only if is continuous at every \(a \in X\).

The next theorem shows that we always have at least some continuous functions.

Theorem 2.12. Given metric spaces \((X, \varrho), (Y, \sigma)\) and \((Z, \tau)\)

(a) the identity mapping
\[\text{id}_X: X \rightarrow X, \quad x \mapsto x\]
is continuous;
(b) if \(f: X \rightarrow Y\) and \(g: Y \rightarrow Z\) are continuous, then so is
\[g \circ f: X \rightarrow Z, \quad x \mapsto g(f(x))\]
(c) for each \(b \in Y\), the constant function
\[c_b: X \rightarrow Y, \quad x \mapsto b\]
is continuous;
(d) If \(g = \varrho_X\) is the discrete metric on \(X\), then every function \(f: X \rightarrow Y\) is continuous.

Proof. To show that these functions are continuous, it suffices to show that they are continuous at an arbitrary point \(x \in X\).

(a) Taking \(\delta := \varepsilon\) in the definition of continuity shows that \(\text{id}_X: X \rightarrow X\) is continuous at \(x\).

(b) Take \(\varepsilon > 0\).
Since \(g\) is continuous on \(Y\), it is continuous at \(f(x) \in Y\).
Hence, there is a \(\mu > 0\) such that \(\tau(g(f(x)), g(f(x'))) < \varepsilon\) whenever \(\sigma(f(x), f(x')) < \mu\).
But \(f\) is also continuous, so that it is continuous at \(x\).
Hence, there is a \(\delta > 0\) such that \(\sigma(f(x), f(x')) < \mu\) whenever \(\varrho(x, x') < \delta\).
Combining these two conditions we see that if \(g(x, x') < \delta\), then \(\tau(g(f(x)), g(f(x'))) < \varepsilon\), showing that \(g \circ f\) is continuous at \(x\).
Since this holds for arbitrary \(x \in X\), \(g \circ f\) is continuous on \(X\).

(c) Take \(x \in X\) and \(\varepsilon > 0\).
Take any \(\delta > 0\) and \(x, x' \in X\) with \(\varrho(x, x') < \delta\). Then
\[
\sigma(c_b(x), c_b(x')) = \sigma(b, b) = 0 < \varepsilon
\]
showing that \(c_b\) is continuous at \(x\).

(d) Given any \(x \in X\) and \(\varepsilon > 0\), put \(\delta = 1\).
By Definition 2.9, \(d_X(x, x') < 1\) if and only if \(d_X(x, x') = 0\) which is the case if and only if \(x' = x\).
But then \(f(x') = f(x)\), whence \(\sigma(f(x'), f(x)) = 0 < \varepsilon\).

1. Exercises

2.1. Which of the following pairs \((X, \varrho)\) are metric spaces?
2. Continuity in \( \mathbb{R}^N \) and Metric Spaces

(i) \( X := \mathbb{R}^2 = \{ (x, y) \mid x, y \in \mathbb{R} \} \)

\[ \varrho: X \times X \to \mathbb{R}, \quad (x_1, y_1), (x_2, y_2) \mapsto |x_1 - x_2| \]

(ii) \( X := \{ f: [0, 1] \to \mathbb{R} \mid f \text{ is continuous} \} \)

\[ \varrho: X \times X \to \mathbb{R}, \quad (f, g) \mapsto \sup_{0 < t < 1} |f(t) - g(t)| \]

(iii) \( X := \{ f: [0, 1] \to \mathbb{R} \mid f \text{ is continuous} \} \)

\[ \varrho: X \times X \to \mathbb{R}, \quad (f, g) \mapsto \int_0^1 |f(t) - g(t)| \, dt \]

(iv) \( X := \{ f: \mathbb{N} \to \mathbb{R} \mid f \text{ is bounded} \} = \{ (x_n)_{n \in \mathbb{N}} \subseteq \mathbb{R} \mid |x_n| < K \text{ for some } K \in \mathbb{R} \} \)

\[ \varrho: X \times X \to \mathbb{R}, \quad (x_n)_{n \in \mathbb{N}}, (y_n)_{n \in \mathbb{N}} \mapsto \sup_{n \in \mathbb{N}} |x_n - y_n| \]

2.2. Given integers \( x, y \), write \( x \sim y \) whenever \( x - y \) is divisible by 2.

Show that \( \sim \) defines an equivalence relation on \( \mathbb{Z} \).

Write \( \mathbb{Z}/2\mathbb{Z} \) for the set of equivalence classes and call \( \mathbb{Z}/2\mathbb{Z} \) the integers mod 2.

Note that \( \mathbb{Z}/2\mathbb{Z} = \{ [0], [1] \} \). For convenience, we write 0 (resp. 1) for \([0]\) (resp. \([1]\))

Take \( n \in \mathbb{N} \setminus \{ 0 \} \) and put \( X := (\mathbb{Z}/2\mathbb{Z})^n \), so that \( X \) consists of all \( n \)-tuples of elements of \( \mathbb{Z}/2\mathbb{Z} \), or, equivalently, all strings of 0’s and 1’s of length precisely \( n \).

Write the elements of \( X \) as \( \mathbf{x} \) or as \( (x_1, \ldots, x_n) \) and call \( x_j \) the \( j \)-th co-ordinate of \( \mathbf{x} \).

Define \( \varrho: X \times X \to \mathbb{R}^+_0 \) by letting \( \varrho(\mathbf{x}, \mathbf{y}) \) be the number of co-ordinates \( j \) for which \( x_j \neq y_j \).

Show that \( (X, \varrho) \) is a metric space.

(This metric is known as the Hamming metric and is used in coding theory.)

2.3. Put \( X = \mathbb{R}^2 \). Define \( \varrho: X \times X \to \mathbb{R}^+_0 \) by

\[ \varrho(\mathbf{x}, \mathbf{y}) := |x_1 - y_1| + |x_2 - y_2| \]

where \( \mathbf{x} = (x_1, x_2), \mathbf{y} = (y_1, y_2), (\mathbf{x}, \mathbf{y}) \in X \times X \).

Show that \( (X, \varrho) \) is a metric space.

2.4. Let \( (X, \varrho) \) be a metric space.

Show that for all \( x, y, z \in X \)

\[ |\varrho(x, z) - \varrho(y, z)| \leq \varrho(x, y). \]

2.5. Let \( (X, \varrho) \) be a metric space and take \( a \in X \).

Prove that the function

\[ f: X \to \mathbb{R}, \quad x \mapsto \varrho(a, x) \]

is continuous. [This function is often denoted by \( \varrho(a, \_ \_ \_) \).]

2.6. Let \( (X, \varrho) \) be a metric space. Given a non-empty subset \( A \) of \( X \) and \( x \in X \) define

\[ \varrho(A, x) := \inf_{a \in A} \varrho(a, x). \]

Prove that \( \varrho(A, x) = 0 \) whenever \( x \in A \).

Find an example where \( x \notin A \), but \( \varrho(A, x) = 0 \).

Is \( \varrho(A, \_ \_ \_): X \to \mathbb{R}, \quad x \mapsto \varrho(A, x) \) a continuous function?
2.7. Fix a prime number, $p$.

Given the non-zero integer, $m$, let $v_p(m)$ denote the exponent of $p$ in the prime factorisation of $z$, so that $v_p(m) = j$ if and only if (a) $p^j$ divides $m$ and (b) $p^{j+1}$ does not divide $m$.

Given the non-zero rational number, $z$, define

$$v_p(z) = v_p(m) - v_p(n)$$

where $m, n$ are non-zero integers with $z = \frac{m}{n}$.

Define

$$d_p : \mathbb{Q} \times \mathbb{Q} \rightarrow \mathbb{R}_0^+, \ (x, y) \mapsto \begin{cases} 0 & \text{if } x = y \\ p^{-v_p(x-y)} & \text{if } x \neq y. \end{cases}$$

Show that $(\mathbb{Q}, d_p)$ is a metric space.

**Observation 2.13.** The metric we have defined he is the $p$-adic metric on $\mathbb{Q}$. It plays an important rôle in number theory.

2.8. Take $\mathbb{R}$ and $\mathbb{R}^2$ with their respective Euclidean metrics. Prove that each of the following functions is continuous.

(a) $\alpha : \mathbb{R}^2 \rightarrow \mathbb{R}, \ (x, y) \mapsto x + y$

(b) $\mu : \mathbb{R}^2 \rightarrow \mathbb{R}, \ (x, y) \mapsto xy$

(c) $\Delta : \mathbb{R} \rightarrow \mathbb{R}^2, \ x \mapsto (x, x)$

[$\Delta$ is the diagonal map.]

2.9. Given a complex vector space, $V$, the function

$$\| \| : V \rightarrow \mathbb{R}_0^+$$

is a norm on $V$ if and only if given any $x, y \in V$ and $\lambda \in \mathbb{C}$

(i) $\|x\| = 0$ if and only if $x = 0 \in V$

(ii) $\|\lambda x\| = |\lambda| \|x\|$

(iii) $\|x + y\| \leq \|x\| + \|y\|$

A normed vector space is a pair $(V, \| \|)$ where $\| \| : V \rightarrow \mathbb{R}_0^+$ is a norm on the vector space $V$. Prove that if $(V, \| \|)$ is a normed vector space, then

$$\rho : V \times V \rightarrow \mathbb{R}_0^+, \ (x, y) \mapsto \|x - y\|$$

defines a metric on $V$.

2.10. Prove that if $(X, \rho)$ is a metric space, then so is $(X, \overline{\rho})$, with $\overline{\rho} : X \times X \rightarrow \mathbb{R}_0^+$ given by

$$\overline{\rho}(x, y) := \frac{\rho(x, y)}{1 + \rho(x, y)}.$$

Let $(Y, \sigma)$ be a metric space. Prove the following.

(a) $f : X \rightarrow Y$ is continuous with respect to $\overline{\rho}$ if and only if it is continuous with respect to $\rho$.

(b) $g : Y \rightarrow X$ is continuous with respect to $\overline{\rho}$ if and only if it is continuous with respect to $\rho$.

2.11. Prove that

$$\max : \mathbb{R}^2 \rightarrow \mathbb{R}, \ (x, y) \mapsto \max\{x, y\}$$

is a continuous function.
Chapter 3

Metric Preserving Functions and Isometry

Having defined what we mean by a metric space, we next formulate when it is that we consider two metric spaces to be “essentially the same”. Intuitively, this should mean that except for the names given to the spaces, their elements and their metrics, the two are interchangeable as metric spaces. We formulate this intuitive notion more precisely.

**Definition 3.1.** Given metric spaces \((X, \rho)\) and \((Y, \sigma)\), the function \(f : X \rightarrow Y\) preserves the metric if and only if for all \(x, x' \in X\).

\[ \sigma(f(x), f(x')) = \rho(x, x') \]

If, in addition, \(f\) is surjective, then it is an isometry\(^1\).

The metric spaces \((X, \rho)\) and \((Y, \sigma)\) are isometric if and only if there is an isometry between them.

Two important immediate properties of metric preserving functions presented in the next lemma. These justify and explain our considering two metric spaces to be “essentially the same” precisely when they are isometric.

**Lemma 3.2.** Let \((X, \rho)\) and \((Y, \sigma)\) be metric spaces and \(f : X \rightarrow Y\) a function.

(i) If \(f\) is metric preserving, then it is continuous and 1−1.

(ii) \(f\) is an isometry if and only if there is a metric preserving function \(g : Y \rightarrow X\) such that \(g \circ f = \text{id}_X\) and \(f \circ g = \text{id}_Y\).

**Proof.** (i) It is easy to see that the metric preserving function \(f\) is continuous: simply put \(\delta = \varepsilon\) in the definition of continuity.

To see that \(f\) is 1−1, take \(x, x' \in X\). Then

\[ f(x) = f(x') \quad \text{if and only if} \quad \sigma(f(x), f(x')) = 0 \quad \text{as} \ \sigma \ \text{is a metric on} \ Y \]

\[ \quad \text{if and only if} \quad \rho(x, x') = 0 \quad \text{as} \ f \ \text{preserves the metric} \]

\[ \quad \text{if and only if} \quad x = x' \quad \text{as} \ \rho \ \text{is a metric on} \ X. \]

(ii) Suppose that \(f\) is an isometry.

Then it is onto and, by (i), 1−1 as well.

---

\(^1\)There are differences in usage to be found in the literature. Some authors use the word “isometry” for what we have called a “metric preserving function”. It is then customary to distinguish between an isometry into (our notion of metric preserving function) and an isometry onto (our notion of an isometry).
3. METRIC PRESERVING FUNCTIONS AND ISOMETRY

Hence there is a \((1 \rightarrow 1)\) and onto function \(g: Y \rightarrow X\) such that \(g \circ f = \text{id}_X\) and \(f \circ g = \text{id}_Y\).

To see that \(g\) preserves the metric, take \(y, y' \in Y\). Then

\[
\varrho(g(y), g(y')) = \sigma(f(g(y)), f(g(y'))) \quad \text{as } f \text{ preserves the metric}
\]

\[
= \sigma((f \circ g)(y), (f \circ g)(y'))
\]

\[
= \sigma(y, y').
\]

For the converse, let \(g: Y \rightarrow X\) be a metric preserving function with \(g \circ f = \text{id}_X\) and \(f \circ g = \text{id}_Y\).

It follows immediately that \(f\) must then be \(1\rightarrow 1\) and onto.

To see \(f\) preserves the metric, choose \(x, x' \in X\). Then

\[
\sigma(f(x), f(x')) = \varrho(g(f(x)), g(f(x')))
\]

\[
= \varrho((g \circ f)(x), (g \circ f)(x'))
\]

\[
= \varrho(x, x'). \quad \square
\]

Example 3.3. Take \(X = Y = \mathbb{R}^2\), the Cartesian plane and let \(\varrho = \sigma = \varepsilon_2\), the Euclidean (usual) metric. Then the functions

\[
f: \mathbb{R}^2 \rightarrow \mathbb{R}^2, \quad (x, y) \mapsto (y, x)
\]

\[
g: \mathbb{R}^2 \rightarrow \mathbb{R}^2, \quad (x, y) \mapsto \left(\frac{x-y}{\sqrt{2}}, \frac{x+y}{\sqrt{2}}\right)
\]

are both isometries.

Observation 3.4. Notice a similarity between linear transformations of vector spaces and metric preserving functions.

A function inverse to linear transformation is also a linear transformation.

A function inverse to a metric preserving function also preserves the metric.

Observation 3.5. A metric space is a set which admits additional structure. We saw in Chapter 2 that, in fact, any set, \(X\), can be made into a metric space by endowing it with its discrete metric, \(d_X\) (cf. Definition 2.9 and Theorem 2.10)

This suggests the possibility that a set can support more than one “essentially different” metric space structure. The notion of isometry allows us to formulate this precisely and show that a set can, indeed be made into a matrix space in more than one way.

Example 3.6. Take \(\mathbb{R}\) with its euclidean metric \(\epsilon\) as our first metric space, and \(\mathbb{R}\) with its discrete metric, \(d\) as our second metric space.

Claim. There is no metric preserving function \(f: (\mathbb{R}, \epsilon) \rightarrow (\mathbb{R}, d)\).

Proof. Let \(f: (\mathbb{R}, \epsilon) \rightarrow (\mathbb{R}, d)\) be a function.

\[
d(f(0), f(2)) = \begin{cases} 
0 & \text{if } f(0) = f(2) \\
1 & \text{otherwise}
\end{cases}
\]

But \(\epsilon(0, 2) = 2\).

Hence \(d(f(0), f(2)) \neq \epsilon(0, 2)\). \quad \square

Having satisfied ourselves that metric spaces provide a genuine generalisation of familiar objects of study in mathematics and that they abound, we turn to the problem of how to derive new metric spaces from given ones.

Our first observation is that if \((X, \varrho)\) is a metric space and \(A\) is any subset of \(X\), then \(A\) inherits a metric structure from \(X\), namely, we define a metric on \(A\) by restricting the metric on \(X\) to \(A\). Formally,
Definition 3.7. Let \((X, \rho)\) be a metric space and let \(A\) be a subset of \(X\). Then the function
\[
\rho |_A : A \times A \rightarrow \mathbb{R}, \quad (a, b) \mapsto \rho(a, b)
\]
is restriction of \(\rho\) to \(A\) or the induced metric on \(A\).

Lemma 3.8. \((A, \rho |_A)\) is a metric space and the natural map
\[
i_A^X : A \rightarrow X, \quad a \mapsto a
\]
preserves the metric.

Proof. Take \(a, b \in A\). Abbreviating \(i_X^A\) to \(i\),
\[
\rho(i(a), i(b)) = \rho(a, b) \quad \text{by the definition of } i
\]
\[
= \rho |_A (a, b) \quad \text{by the definition of } \rho |_A
\]
\(\Box\)

Lemma 3.8 completely characterises the induced metric on any subset, as the next lemma shows.

Lemma 3.9. Let \((A, \sigma)\) and \((X, \rho)\) be metric spaces, with \(A \subseteq X\).

The natural map \(i_A^X : A \rightarrow X, a \mapsto a\) preserves the metric if and only if \(\sigma = \rho |_A\).

Proof. Exercise. \(\Box\)

Thus the natural notion of subspace in the case of metric spaces is that of a subset with the induced metric. Formally,

Definition 3.10. A subspace of the metric space \((X, \rho)\) is a subset \(A\) of \(X\) endowed with the induced metric.

Subspaces can be characterised by means of a universal property:

Theorem 3.11. Let \((A, \sigma)\) and \((X, \rho)\) be metric spaces with \(A \subseteq X\). Then the following are equivalent.

(i) \((A, \sigma)\) is a metric subspace of \((X, \rho)\).

(ii) Given any metric space \((W, \tau)\), the function \(f : W \rightarrow A\) preserves the metric if and only if \(i_A^X \circ f : W \rightarrow X\) preserves the metric.

Proof. By Lemmas 3.8 and 3.9, \((A, \sigma)\) is a metric subspace of \((X, \rho)\) if and only if \(i_A^X\) preserves the metric.

If \(i_A^X\) preserves the metric and \(f : W \rightarrow A\) is any function whatsoever, then
\[
\rho((i_A^X \circ f)(w), (i_A^X \circ f)(w')) = \sigma(f(w), f(w')),
\]
so that \(f\) preserves the metric if and only if \(i_A^X \circ f\) does.

Conversely, suppose that given any metric space \((W, \tau)\) and any function \(f : W \rightarrow A\) whatsoever, \(f\) preserves the metric if and only if \(i_A^X \circ f\) does.

Consider the special case \(W := A, \tau := \rho_A\) and \(f : A \rightarrow A, a \mapsto a\), so that \(f = id_A\).

Then \(i_A^X \circ f = i_A^X \circ id_A : A \rightarrow X\).

This clearly preserves the metric, since \(\rho_A\) is the induced metric on \(A\).

But \(i_A^X \circ id_A = i_A^X\), so that \(i_A\) does indeed preserve the metric. \(\Box\)
3. METRIC PRESERVING FUNCTIONS AND ISOMETRY

1. Exercises

3.1. Take \( X = Y = \mathbb{R}^2 \) and let \( \varrho = \sigma = \epsilon_2 \) be the usual (Euclidean) metric.

Prove that the functions
\[
\begin{align*}
  f &: \mathbb{R}^2 \rightarrow \mathbb{R}^2, \quad (x, y) \mapsto (y, x) \\
  g &: \mathbb{R}^2 \rightarrow \mathbb{R}^2, \quad (x, y) \mapsto \left( \frac{x - y}{\sqrt{2}}, \frac{x + y}{\sqrt{2}} \right)
\end{align*}
\]
are both isometries.

3.2. Let \((X, \varrho)\) be a metric space.

Take a sub-additive monotonically strictly increasing function \( \alpha : \mathbb{R}^+ \rightarrow \mathbb{R}^+ \) with \( \alpha(0) = 0 \), that is to say, \( \alpha \) satisfies:

\begin{enumerate}[label=(i)]
  \item \( \alpha(x + y) \leq \alpha(x) + \alpha(y) \) for all \( x, y \in \mathbb{R}^+ \) (Sub-Additive)
  \item \( \alpha(x) < \alpha(y) \) whenever \( x < y \) (Monotonically Strictly Increasing)
  \item \( \alpha(0) = 0 \)
\end{enumerate}

Prove that
\[
\alpha \circ \varrho : X \times X \rightarrow \mathbb{R}^+, \quad (x, y) \mapsto \alpha \left( \varrho(x, y) \right)
\]
is a metric on \( X \).

3.3. Let \((A, \sigma)\) and \((X, \varrho)\) be metric spaces, with \( A \subseteq X \).

Prove that
\[
i_A^X : A \rightarrow X, \quad a \mapsto a
\]
preserves the metric if and only if \( \sigma = \varrho |_A \).

3.4. Let \( \epsilon \) be the Euclidean metric on \( \mathbb{R}^n \).

Let \( A = (a_{ij})_{n \times n} \) be a positive definite symmetric matrix with real coefficients.

Given \( x = (x_1, \ldots, x_n), y = (y_1, \ldots, y_n) \in \mathbb{R}^n \), define
\[
\rho(x, y) := \sqrt{\sum_{i,j=1}^{n} a_{ij}(x_i - y_i)(x_j - y_j)}.
\]

(a) Show that \( \rho \) defines a metric on \( \mathbb{R}^n \) such that \((\mathbb{R}^n, \rho)\) and \((\mathbb{R}^n, \epsilon)\) are isometric.

(b) Show that the function \( f : (\mathbb{R}^n, \epsilon) \rightarrow (\mathbb{R}^n, \rho), \quad x \mapsto x \) is continuous.

3.5. Give an example of two metrics \( \varrho \) and \( \sigma \) on the same space \( X \) such that

(a) \((X, \varrho)\) and \((X, \sigma)\) are isometric, but

(b) the function
\[
f : (X, \varrho) \rightarrow (X, \sigma), \quad x \mapsto x
\]
is not continuous.
Chapter 4

The Cartesian Product of Metric Spaces

Given metric spaces, \((X, \varrho)\) and \((Y, \sigma)\), we wish to define a metric on \(X \times Y\), the cartesian product of the sets \(X\) and \(Y\), which reflects the metric space structure of both \((X, \varrho)\) and \((Y, \sigma)\). While the discrete metric turns \(X \times Y\) into a metric space, it retains no information about either \(\varrho\) or \(\sigma\). So it will not do.

We examine ways in which \(X \times Y\) may be endowed with a metric reflecting the metric space structures of \((X, \varrho)\) and \((Y, \sigma)\). Three constructions come readily to mind. We present each and compare them.

1. The Metric \(\varrho \times \sigma\)

The first is modelled on the Euclidean metric on \(\mathbb{R}^2 = \mathbb{R} \times \mathbb{R}\).

Let \(\epsilon\) be the Euclidean metric on \(\mathbb{R}\), so that \(\epsilon(a, b) = |a - b|\)

Let \(\epsilon_2\) be the Euclidean metric on \(\mathbb{R}^2\). Then

\[
\epsilon_2((x_1, y_1), (x_2, y_2)) = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2} = \sqrt{|x_1 - x_2|^2 + |y_1 - y_2|^2} = \sqrt{\varrho(x_1, x_2)^2 + \sigma(y_1, y_2)^2}
\]

Replacing the first copy of \(\mathbb{R}\) in the product by \((X, \varrho)\), the second by \((Y, \sigma)\) motivates our definition

**Definition 4.1.** Given metric spaces, \((X, \varrho)\) and \((Y, \sigma)\), the **product metric on** \(X \times Y\) **is the function**

\[
\varrho \times \sigma : (X \times Y) \times (X \times Y) \to \mathbb{R}^+\]

\[
((x_1, y_1), (x_2, y_2)) \mapsto \sqrt{\varrho(x_1, x_2)^2 + \sigma(y_1, y_2)^2}
\]

**Lemma 4.2.** Let \((X, \varrho)\) and \((Y, \sigma)\) be metric spaces. Then \(\varrho \times \sigma\) is a metric on \(X \times Y\).

**Proof.** Take \(x_1, x_2, x_3 \in X\) and \(y_1, y_2, y_3 \in Y\).

(i) Clearly,

\[
(\varrho \times \sigma)((x_1, y_1), (x_2, y_2)) = 0 \quad \text{if and only if} \quad \sqrt{\varrho(x_1, x_2)^2 + \sigma(y_1, y_2)^2} = 0
\]

\[
\text{if and only if} \quad \varrho(x_1, x_2) = \sigma(y_1, y_2) = 0.
\]

Since both \(\varrho\) and \(\sigma\) are metrics, this is the case if and only if \((x_1, y_1) = (x_2, y_2)\).
Lemma 4.4

We summarise these considerations in a lemma.

We omit indices from our notation unless there is danger of confusion.

Observation 4.3

Since none of the terms can be negative, this inequality is equivalent to

\[ a^2 + d^2 \leq b^2 + e^2 + 2\sqrt{(b^2 + e^2)(c^2 + f^2)} + c^2 + f^2. \]

Since \( a \leq b + c \) and \( d \leq e + f \),

\[ a^2 + d^2 \leq b^2 + c^2 + 2bc + c^2 + f^2 + 2ef. \]

Hence, it is sufficient to show that

\[ bc + ef \leq \sqrt{(b^2 + e^2)(c^2 + f^2)}. \]

Since all the terms are non-negative, this is equivalent to showing that

\[ 2bcef \leq b^2 f^2 + e^2 c^2, \]

or, equivalently, that

\[ b^2 f^2 - 2bcef + e^2 c^2 \geq 0. \]

Since \( b^2 f^2 - 2bcef + e^2 c^2 = (bf - ec)^2 \), the conclusion is immediate. \( \square \)

Observation 4.3. This construction can be extended to any finite family of metric spaces, \( \{ (X_j, \sigma_j) \mid j = 1, \ldots, N \} \). Putting

\[ \prod_{j=1}^{N} X_j := X_1 \times \cdots \times X_N := \{(x_1, \ldots, x_N) \mid x_j \in X_j, j = 1, \ldots, N\} \]

and writing \( (x_j)_{j=1}^{N} \) for \( (x_1, \ldots, x_N) \) where \( x_j \in X_j \) \( (j = 1, \ldots, N) \), we define

\[ \prod_{j=1}^{N} \sigma_j : \left( \prod_{j=1}^{N} X_j \right) \times \left( \prod_{j=1}^{N} X_j \right) \rightarrow \mathbb{R}_0^+, \quad ((x_j)_{j=1}^{N}, (y_j)_{j=1}^{N}) \rightarrow \sqrt{\sum_{j=1}^{N} (\sigma_j(x_j, y_j))^2} \]

An elementary inductive argument shows that \( \left( \prod_{j=1}^{N} X_j, \prod_{j=1}^{N} \sigma_j \right) \) is a metric space.

Convention. We omit indices from our notation unless there is danger of confusion.

We summarise these considerations in a lemma.

Lemma 4.4. Let \( \{ (X_j, \sigma_j) \mid j = 1, \ldots, n \} \) be a finite set of metric spaces. Then \( (\prod X_j, \prod \sigma_j) \) is also a metric space.
Remark 4.5. This construction can be extended to the case of an infinite sequence \((X_n, \varrho_n)_{n \in \mathbb{N}}\) of metric spaces, as long as we ensure that
\[
\sum_{n \in \mathbb{N}} (\varrho_n(x_n, y_n))^2
\]
is finite for every \((x_n)_{n \in \mathbb{N}}, (y_n)_{n \in \mathbb{N}} \in \prod_{n \in \mathbb{N}} X_n\).
When \((X_m, \varrho_m) = (X_n, \varrho_n)\) for all \(m, n\), such spaces are investigated in detail in functional analysis, where they are known as “\(\ell_2\)-spaces”.

2. The Metric \(\varrho_{X \times Y}\)

Our second construction is based on the observation that the larger of two non-negative numbers is non-negative.

Definition 4.6. Given metric spaces \((X, \varrho)\) and \((Y, \sigma)\), the product metric on \(X \times Y\) is the function
\[
\varrho_{X \times Y} : (X \times Y) \times (X \times Y) \longrightarrow \mathbb{R}^+,
\]
\[
((x_1, y_1), (x_2, y_2)) \longmapsto \max\{\varrho(x_1, x_2), \sigma(y_1, y_2)\}.
\]

Lemma 4.7. Given metric spaces, \((X, \varrho)\) and \((Y, \sigma)\), \(\varrho_{X \times Y}\) is a metric on \(X \times Y\).

Proof. Take \(x_1, x_2, x_3 \in X\) and \(y_1, y_2, y_3 \in Y\).
(i) Since \(\varrho(x_1, x_2), \sigma(y_1, y_2) \geq 0\),
\[
\varrho_{X \times Y} ((x_1, y_1), (x_2, y_2)) = \max\{\varrho(x_1, x_2), \sigma(y_1, y_2)\}
\]
with \(\varrho_{X \times Y} ((x_1, y_1), (x_2, y_2)) = 0\) if and only if \(\varrho(x_1, x_2) = 0\) and \(\sigma(y_1, y_2) = 0\), in which case \(x_1 = x_2\) and \(y_1 = y_2\), or, equivalently, \((x_1, y_1) = (x_2, y_2)\).
(ii) It follows directly from Definition 4.6 that \(\varrho_{X \times Y} ((x_1, y_1), (x_2, y_2)) = \varrho_{X \times Y} ((x_2, y_2), (x_1, y_1))\).
(iii) Finally, it follows from \(\varrho(x_1, x_3) \leq \varrho(x_1, x_2) + \varrho(x_2, x_3)\), that
\[
\varrho(x_1, x_3) \leq \max\{\varrho(x_1, x_2), \sigma(y_1, y_2)\} + \varrho(x_2, x_3)
\]
\[
\leq \max\{\varrho(x_1, x_2), \sigma(y_1, y_2)\} + \max\{\varrho(x_2, x_3), \sigma(y_2, y_3)\}.
\]
Similiarly \(\sigma(y_1, y_3) \leq \max\{\varrho(x_1, x_2), \sigma(y_1, y_2)\} + \max\{\varrho(x_2, x_3), \sigma(y_2, y_3)\}\), so that
\[
\varrho_{X \times Y} ((x_1, y_1), (x_3, y_3)) = \max\{\varrho(x_1, x_3), \sigma(y_1, y_3)\}
\]
\[
\leq \max\{\varrho(x_1, x_2), \sigma(y_1, y_2)\} + \max\{\varrho(x_2, x_3), \sigma(y_2, y_3)\}
\]
\[
= \varrho_{X \times Y} ((x_1, y_1), (x_2, y_2)) + \varrho_{X \times Y} ((x_2, y_2), (x_3, y_3)).
\]

This construction can be extended to any finite family, \(\{(X_j, \varrho_j) \mid j = 1, \ldots, N\}\), of metric spaces, by defining
\[
\varrho_{\prod X_j} : \left(\prod_{j=1}^N X_j \times \prod_{j=1}^N X_j\right) \longrightarrow \mathbb{R}^+,
\]
\[
((x_j), (y_j)) \longmapsto \max\{\varrho_j(x_j, y_j) \mid j = 1, \ldots, N\}.
\]
An inductive proof again shows that \(\varrho_{\prod X_j}\) is in fact a metric, which yields our next lemma.

Lemma 4.8. Suppose that \(\{(X_j, \varrho_j) \mid j = 1, \ldots, N\}\) is a finite set of metric spaces. Then \((\prod X_j, \varrho_{\prod X_j})\) is also a metric space.
Remark 4.9. This construction can even be extended to the of an infinite sequence \((X_n, \varrho_n)_{n \in \mathbb{N}}\) of metric spaces, as long as we replace the maximum by the supremum and ensure that

\[
\sup \{ \varrho_n(x_n, y_n) \mid n \in \mathbb{N} \}
\]

is finite for every \((x_n)_{n \in \mathbb{N}}, (y_n)_{n \in \mathbb{N}} \in \prod_{n \in \mathbb{N}} X_n\).
When \((X_m, \varrho_m) = (X_n, \varrho_n)\) for all \(m, n\), such spaces are investigated in functional analysis, where they are known as \(\ell_\infty\)-spaces.

3. The Metric \(\varrho + \sigma\)

Our third construction is based on the observation that the sum of two non-negative numbers is non-negative.

Definition 4.10. Given metric spaces \((X, \varrho)\) and \((Y, \sigma)\), the sum of \(\varrho\) and \(\sigma\) is the function

\[
\varrho + \sigma: (X \times Y) \times (X \times Y) \rightarrow \mathbb{R}_0^+,
\]

\[
((x_1, y_1), (x_2, y_2)) \mapsto \varrho(x_1, x_2) + \sigma(y_1, y_2).
\]

Lemma 4.11. Given metric spaces \((X, \varrho)\) and \((Y, \sigma)\), \(\varrho + \sigma\) is a metric on \(X \times Y\).

Proof. Take \(x_1, x_2, x_3 \in X\) and \(y_1, y_2, y_3 \in Y\).

(i) Since \(\varrho(x_1, x_2), \sigma(y_1, y_2) \geq 0\),

\[
(\varrho + \sigma)((x_1, y_1), (x_2, y_2)) = \varrho(x_1, x_2) + \sigma(y_1, y_2) \geq 0
\]

with \((\varrho + \sigma)((x_1, y_1), (x_2, y_2)) = 0\) if and only if \(\varrho(x_1, x_2) = 0\) and \(\sigma(y_1, y_2) = 0\), in which case \(x_1 = x_2\) and \(y_1 = y_2\), or equivalently, \((x_1, y_1) = (x_2, y_2)\).

(ii) It follows directly from Definition 4.10 that \((\varrho + \sigma)((x_1, y_1), (x_2, y_2)) = (\varrho + \sigma)((x_2, y_2), (x_1, y_1))\).

(iii) Finally,

\[
(\varrho + \sigma)((x_1, y_1), (x_3, y_3)) = \varrho(x_1, x_3) + \sigma(y_1, y_3)
\]

\[
\leq \varrho(x_1, x_2) + \varrho(x_2, x_3) + \sigma(y_1, y_2) + \sigma(y_2, y_3)
\]

\[= (\varrho + \sigma)((x_1, y_1), (x_2, y_2)) + (\varrho + \sigma)((x_2, y_2), (x_3, y_3))
\]

This construction can be extended to any finite family, \(\{(X_j, \varrho_j) \mid j = 1, \ldots, N\}\), of metric spaces by defining

\[
\sum_{j=1}^N \varrho_j: \left( \prod_{j=1}^N X_j \times \prod_{j=1}^N X_j \right) \rightarrow \mathbb{R}_0^+,
\]

\[
((x_j)_{j=1}^N, (y_j)_{j=1}^N) \mapsto \sum_{j=1}^N \varrho_j(x_j, y_j).
\]

An inductive proof shows that \(\sum_{j=1}^N \varrho_j\) is in fact a metric, which yields our next lemma.

Lemma 4.12. Suppose that \(\{(X_j, \varrho_j) \mid j = 1, \ldots, N\}\) is a finite set of metric spaces. Then \((\prod_{j=1}^N X_j, \sum_{j=1}^N \varrho_j)\) is also a metric space.

Remark 4.13. We can extend this construction to the case of an infinite sequence \((X_n, \varrho_n)_{n \in \mathbb{N}}\) of metric spaces, as long as we ensure that

\[
\sum_{j=1}^{\infty} \varrho_j(x_j, y_j)
\]

is finite for every \((x_j)_{j \in \mathbb{N}}, (y_j)_{j \in \mathbb{N}} \in \prod_{n \in \mathbb{N}} X_n\).
When \((X_m, \varrho_m) = (X_n, \varrho_n)\) for all \(m, n\), such spaces are investigated in functional analysis, where they are known as \(\ell_1\)-spaces.
4. Comparison of the Three Metrics

Given the differences between the three constructions, our first observation may come as a surprise: the three metrics constructed on $X \times Y$ give rise to the same families of continuous functions.

**Lemma 4.14.** Let $(W, \varpi), (X, \varrho), (Y, \sigma)$ and $(Z, \tau)$ be metric spaces. Take functions $h: W \to X \times Y$ and $g: X \times Y \to Z$.

1. The following are equivalent
   - (a) $h$ is continuous with respect to $\varrho \times \sigma$
   - (b) $h$ is continuous with respect to $\varrho_{X \times Y}$
   - (c) $h$ is continuous with respect to $\varrho + \sigma$;

2. The following are equivalent
   - (a) $g$ is continuous with respect to $\varrho \times \sigma$
   - (b) $g$ is continuous with respect to $\varrho_{X \times Y}$
   - (c) $g$ is continuous with respect to $\varrho + \sigma$.

**Proof.** The proof can be achieved by "\(\epsilon - \delta\) arguments", and is left as an exercise. \(\square\)

**Observation 4.15.** A more conceptual (and far easier!) proof will be available once we have defined topologies.

**Convention.** When we refer to the product of finitely many metric spaces, we have in mind any one of the three metrics we have just examined, unless we make an explicit statement to the contrary. In light of Lemma 4.14, this does not affect the continuity of any function, and we may choose whichever metric we find most convenient, if we are only concerned with continuity.

**Observation 4.16.** An easy inductive argument shows that the result holds for any finite product of metric spaces. We leave this argument to the reader to complete.

Recall that the Cartesian product of sets comes with natural projections,

$$
pr_j: \prod_{n=1}^{N} X_n \to X_j, \quad (x_n)_{n=1}^{N} \mapsto x_j
$$

for $j \in \{1, \ldots, N\}$.

**Lemma 4.17.** Let \(\{(X, \varrho_n) \mid n = 1, \ldots, N\}\) be a set of $N$ metric spaces.

Endow $X := \prod_{n=1}^{N} X_n$ with one of the metrics constructed above.

Then each natural projection $pr_j: X \to X_j$ is continuous.

**Proof.** As, by Lemma 4.14, it does not matter which of the three above metrics we use, we take the product metric.

Take $a = (a_1, \ldots, a_n) \in X$ and $\varepsilon > 0$.

Put $\delta := \varepsilon$.

Take $x = (x_1, \ldots, x_n) \in X$, with

$$
\varrho_{\prod X_n}(a, x) < \delta
$$

Then, for each $j \in \{1, \ldots, n\}$

$$
g_j(pr_j(a), pr_j(x)) = \varrho_j(a_j, x_j)
\leq \max\{\varrho_n(a_n, x_n) \mid 1 \leq n \leq N\}
= \varrho_{\prod X_n}(a, x)
< \varepsilon
$$
But more is true.

**Theorem 4.18.** Given a metric space \((W, \varsigma)\) and continuous functions \(f_j : W \to X_j\) \((j = 1, \ldots, N)\) there is a unique continuous function

\[
f : W \to \prod_{j=1}^N X_j
\]

with \(f_j = \text{pr}_j \circ f\) for \(j = 1, \ldots, N\).

**Proof.** Put \(X := \prod_{j=1}^N X_j\).

For each \(j \in \{1, \ldots, N\}\), let \(f_j : W \to X_j\) be a continuous function.

By Theorem 1.54, there is a unique function

\[
f : W \to X, \quad w \mapsto (f_1(w), \ldots, f_N(w))
\]

It only remains to establish continuity of \(f\).

Take \(a \in W\) and \(\varepsilon > 0\).

Then, for each \(j \in \{1, \ldots, N\}\) there is a \(\delta_j\) with \(\varsigma_j (f_j(w), f_j(a)) < \varepsilon\) whenever \(\varsigma(w, a) < \delta_j\).

Put \(\delta := \min\{\delta_j \mid j = 1, \ldots, N\}\) and take \(w\) with \(\varsigma(w, a) < \delta\).

Then \(\varsigma(w, a) < \delta_j\) for each \(j\), whence \(\varsigma_j (f_j(w), f_j(a)) < \varepsilon\) for every \(j\). Thus,

\[
\varsigma_{\prod_j X_j}(f(w), f(a)) = \max\{\varsigma_j(f_j(w), f_j(a)) \mid j = 1, \ldots, N\} < \varepsilon.
\]

Recall that two metric spaces are equivalent as metric spaces if and only if they are isometric.

We have constructed three metrics on the cartesian product of two metric spaces which cannot be distinguished from one another by means of their respective classes of continuous functions. So, unless these three are always isometric, the notion of a metric space does not characterise continuity.

**Theorem 4.19.** Take metric spaces \((X, \varrho)\) and \((Y, \sigma)\).

In general, no two of the three canonical metrics on \(X \times Y\), \(\varrho \times \sigma, \varrho_{X \times Y}, \varrho + \sigma\) result in isometric metric spaces.

**Proof.** Exercise.

Recall that if both \(\varrho\) and \(\sigma\) are the discrete metric, then \(\varrho_{X \times Y}\) is also the discrete metric on \(X \times Y\), but neither \(\varrho \times \sigma\) nor \(\varrho + \sigma\) can be the discrete metric if both \(X\) and \(Y\) have at least two elements.

This is the reason for \(\varrho_{X \times Y}\) being the product metric.
5. Exercises

4.1. Let \((W, \varpi)\), \((X, \wp)\) and \((Y, \tau)\) be metric spaces. Take functions \(f: W \to X \times Y\) and \(g: X \times Y \to W\).

(i) Prove that the following are equivalent.

(a) \(f\) is continuous with respect to \(\wp \times \tau\).
(b) \(f\) is continuous with respect to \(\wp + \tau\).
(c) \(f\) is continuous with respect to \(\wp_{X \times Y}\).

(ii) Prove that the following are equivalent.

(a) \(g\) is continuous with respect to \(\wp \times \tau\).
(b) \(g\) is continuous with respect to \(\wp + \tau\).
(c) \(g\) is continuous with respect to \(\wp_{X \times Y}\).

4.2. Let \((X, \wp)\) and \((Y, \sigma)\) be metric spaces. The function \(f: X \to Y\) is uniformly continuous if and only if for every \(\varepsilon > 0\) there is a \(\delta > 0\) such that \(\sigma(f(u), f(v)) < \varepsilon\), whenever \(\wp(u, v) < \delta\).

Prove that for any metric space \((X, \wp)\), the function \(\wp: X \times X \to \mathbb{R}^+_0\), \((u, v) \mapsto \wp(u, v)\) is uniformly continuous with respect to the Euclidean metric on \(\mathbb{R}\) and one of the metrics on \(X \times X\) discussed above.

4.3. Find metric spaces \((X, \wp)\) and \((Y, \tau)\) such that no two of \((X \times Y, \wp_{X \times Y}), (X \times Y, \wp + \tau)\) and \((X \times Y, \wp \times \tau)\) are isometric.

4.4. Let \(d_X\) be the discrete metric on \(X\) and \(d_Y\) the discrete metric on \(Y\). Show that if \(X\) and \(Y\) each have at least two elements, then

(a) \(\wp_{X \times Y} = d_{X \times Y}\);
(b) No two of the metric spaces \((X \times Y, d_X \times d_Y), (X \times Y, \wp_{X \times Y})\) and \((X \times Y, d_X + d_Y)\) are isometric.

4.5. Let \((X, \wp)\) be a metric space.

Given \(a \in X\) and \(r \in \mathbb{R}^+\), define the circle of radius \(r\) with centre \(A\) with respect to \(\wp\) to be

\[S_{\wp}(a; r) := \{x \in X \mid \wp(a, x) = r\}\]

Take \(\mathbb{R}\) with its Euclidean metric \(\epsilon\).

For \(\mathbb{R}^2 = \mathbb{R} \times \mathbb{R}\), draw, in a single diagram, the circle of radius 1 with centre \((0, 0)\) with respect to each of the metrics \(\epsilon \times \epsilon, \wp_{\mathbb{R} \times \mathbb{R}}\) and \(\epsilon + \epsilon\).

4.6. Give an example of two metrics \(\wp\) and \(\sigma\) on the same space \(X\) such that

(a) \((X, \wp)\) and \((X, \sigma)\) are isometric, but
(b) \(f: (X, \wp) \to (X, \sigma), \ x \mapsto x\) is not continuous.
Chapter 5

Open Sets in a Metric Space

Recall that the open intervals
\[ ] a, b [ := \{ t \in \mathbb{R} \mid a < t < b \}\]
with \( a, b \in \mathbb{R} \) and \( a < b \), play an important rôle in calculus.

If open intervals could be characterised purely in terms of distance, we could generalise them to arbitrary metric spaces.

To this end, observe that the mid-point of the (finite) interval \( ] a, b [ \) is
\[ \xi := \frac{a + b}{2} \]
and its length (“diameter”) is \( b - a \).

It is now immediate that \( x \in ] a, b [ \) if and only if \( |x - \xi| < r \), where \( r := \frac{b - a}{2} \).

We have just shown that
\[ ] a, b [ := \{ t \in \mathbb{R} \mid |t - \xi| < r \}, \]
with \( \xi \) and \( r \) as above.

Using \( \epsilon \), the Euclidean metric on \( \mathbb{R} \), we obtain
\[ ] a, b [ := \{ t \in \mathbb{R} \mid \epsilon(t, \xi) < r \}, \]
that is, each finite open interval comprises precisely those real numbers whose distance from a given real number is less than a given positive number.

This characterisation of open intervals provides the basis for our next definition.

**Definition 5.1.** Let \((X, \rho)\) be a metric space.

Given \( \xi \in X \) and \( r > 0 \), the open ball of radius \( r \) about \( \xi \) is
\[ B(\xi; r) := \{ x \in X \mid \rho(x, \xi) < r \}. \]

In case of ambiguity with regard to the metric concerned, we write \( B^\rho(\xi; r) \) or \( B_\xi(\xi; r) \), to indicate that \( \rho \) is the intended metric.

**Observation 5.2.** If the metric space in question is Euclidean 3-space — that is \( \mathbb{R}^3 \) with its Euclidean metric — then \( B(\xi; r) \) is just the ordinary geometric ball whose centre is located at the point \( \xi \) in space and whose radius is \( r \) units, without its boundary sphere. This example explains the term “open ball”.

43
The next lemma establishes a property of open balls which may seem trite, but is important for subsequent investigations, namely that every element of a given open ball is itself the centre of an open ball contained within the original open ball.

**Lemma 5.3.** Let \( B(\xi; r) \) be an open ball in the metric space \((X, \varrho)\).

For each \( x \in B(\xi; r) \), there is an \( s > 0 \) with \( B(x; s) \subseteq B(\xi; r) \).

**Proof.** Take \( x \in B(\xi; r) \) and put \( s := r - \varrho(\xi, x) \). 

Plainly \( s > 0 \) and if \( y \in B(x; s) \), then \( \varrho(x, y) < s \). Thus

\[
\varrho(\xi, y) \leq \varrho(\xi, x) + \varrho(x, y) < \varrho(\xi, x) + s \leq \varrho(\xi, x) + r - \varrho(\xi, x) = r,
\]

showing that \( y \in B(\xi; r) \). \(\Box\)

This property is the basis for extending the notion of open subset from calculus to general metric spaces. Recall from calculus, that a subset of the set of all real numbers is open if and only if it can be written as a union of (finite) open intervals.

**Definition 5.4.** The subset \( A \) of the metric space \((X, \varrho)\) is open in \(X\), or an open subset of \(X\), or open with respect to \(\varrho\), if and only if if and only if it is the union of open balls.

It is common to simply refer to such an \( A \) as an open set, suppressing the ambient metric space when there is no ambiguity. But it is important to bear in mind that being open is not an intrinsic property of a given set: it is a relationship between that set and the ambient space.

By Lemma 5.3 every open ball is an open set. Even more so, the union of an arbitrarily set of open balls forms an open set. In fact, this property characterises open subsets of a metric space.

**Theorem 5.5.** The subset \( A \) of the metric space \((X, \varrho)\) is open if and only if every element of \( A \) is the centre of an open ball contained in \( A \), or, equivalently, for each \( a \in A \) there is an \( r \) with \( B(a; r) \subseteq A \).

**Proof.** Suppose that for each \( a \in A \) there is an \( r \) with \( B(a; r) \subseteq A \). Given \( x \in A \) choose \( r_x > 0 \) with \( B(x; r_x) \subseteq A \). Then

\[
A = \bigcup_{x \in A} \{x\} \\
\subseteq \bigcup_{x \in A} B(x; r_x) \\
\subseteq \bigcup_{x \in A} A \\
= A.
\]

For the converse, let

\[
A = \bigcup_{\lambda \in \Lambda} B(a_\lambda; r_\lambda)
\]

with \( a_\lambda \in A \) and \( r_\lambda > 0 \) for each \( \lambda \in \Lambda \).

Then, given \( x \in A \), there is a \( \lambda_x \in \Lambda \) with \( x \in B(a_\lambda_x; r_\lambda_x) \).

By Lemma 5.3, there is an \( r_x > 0 \) with \( B(x; r_x) \subseteq B(a_\lambda_x; r_\lambda_x) \).

As \( B(a_\lambda_x; r_\lambda_x) \subseteq A \), the result follows. \(\Box\)
CHOOSING

Note also that Choosing

s

In light of Theorem 5.5, in order to show that the same subsets of Choosing

\[ X \]

and every square contains a concentric inscribed circle. These two set-theoretic inequalities restate the fact from plane geometry that every circle contains Choosing

center 

0

is open with respect to the discrete metric as it contains the open (discrete) ball of radius \( \frac{1}{2} \) with centre 0, but no open (Euclidean) ball in \( \mathbb{R} \) contains only one element.

EXAMPLE 5.7. If a set \( X \) admits two metrics, then it is possible for \( A \subseteq X \) to be open with respect to one metric without being open with respect to the other.

An example is provided by \( \mathbb{R} \) with the discrete and the Euclidean metrics. The subset \{0\} of \( \mathbb{R} \) is open with respect to the discrete metric as it contains the open (discrete) ball of radius \( \frac{1}{2} \) with centre 0, but no open (Euclidean) ball in \( \mathbb{R} \) contains only one element.

EXAMPLE 5.8. If \( \varrho \) and \( \sigma \) are two distinct metrics on \( X \), that is, if

\[ f: (X, \varrho) \longrightarrow (X, \sigma), \quad x \mapsto x \]

is not an isometry, then it may still occur that \( A \subseteq X \) is open with respect to \( \varrho \) if and only if it is open with respect to \( \sigma \).

For example, take \( \mathbb{R}^2 \) with \( \varrho = \epsilon \), the Euclidean metric and \( \sigma \) the metric defined by

\[ \sigma((x_1, y_1), (x_2, y_2)) := \max\{|x_1 - x_2|, |y_1 - y_2|\}. \]

In light of Theorem 5.5, in order to show that the same subsets of \( \mathbb{R}^2 \) are open with respect to the two different metrics, it is enough to show that any open ball with respect to one metric is open in the other metric.

This is equivalent to showing that any open ball with respect to one metric with centre \( \xi \) contains an open ball with respect to the other metric with the same centre, \( \xi \).

In other words, we must show that given any \( \xi \in \mathbb{R}^2 \) and \( r > 0 \), we can find \( s, t > 0 \) with

\[
\begin{align*}
\{ x \in \mathbb{R}^2 \mid \varrho(\xi, x) < s \} & \subseteq \{ y \in \mathbb{R}^2 \mid \sigma(\xi, y) < r \} \tag{*} \\
\{ x \in \mathbb{R}^2 \mid \sigma(\xi, x) < t \} & \subseteq \{ y \in \mathbb{R}^2 \mid \varrho(\xi, y) < r \}. \tag{**}
\end{align*}
\]

Writing \((\xi_1, \xi_2)\) for \( \xi \in \mathbb{R}^2 \) and \((x_1, x_2)\) for \( x \in \mathbb{R}^2 \),

\[
\varrho(\xi, x) = \sqrt{\left(\xi_1 - x_1\right)^2 + \left(\xi_2 - x_2\right)^2}
\]

\[
\sigma(\xi, x) = \max\{|\xi_1 - x_1|, |\xi_2 - x_2|\}.
\]

Thus, \( \varrho(\xi, x) < s \) if and only if \( \sqrt{(\xi_1 - x_1)^2 + (\xi_2 - x_2)^2} < s \).

But then, \( \max\{|\xi_1 - x_1|, |\xi_2 - x_2|\} < s \).

Choosing \( s := r \) now establishes \((*)\).

Note also that \( \sigma(\xi, x) < t \) if and only if \( \max\{|\xi_1 - x_1|, |\xi_2 - x_2|\} < t \).

In that case \( \sqrt{(\xi_1 - x_1)^2 + (\xi_2 - x_2)^2} < \sqrt{2}t \).

Choosing \( t := \frac{1}{\sqrt{2}}r \) establishes \((**)\).

These two set-theoretic inequalities restate the fact from plane geometry that every circle contains a concentric inscribed square and every square contains a concentric inscribed circle.
Plainly, the property of having the same open subsets defines an equivalence relation on the set of all metric spaces which have the same underlying set $X$.

To assist understanding this equivalence relation, we examine characteristic properties of the open subsets of a metric space.

**Theorem 5.9.** Let $(X, \varrho)$ be a metric space. Then

(i) $\emptyset$ and $X$ are open subsets of $X$.

(ii) If $\{A_\lambda \mid \lambda \in \Lambda\}$ is a family of open subsets of $X$, then $\bigcup_{\lambda \in \Lambda} A_\lambda$ is open in $X$.

(iii) If $\{A_j \mid j = 1, \ldots, n\}$ is a finite family of open subsets of $X$, then $\bigcap_{j=1}^n A_j$ is open in $X$.

**Proof.** (i) By Definition 5.1, any open ball centred on any element of $X$ is a subset of $X$. Hence, by Definition 5.4, $X$ is an open subset of itself.

If $x$ is an element of the empty set, then $x$ is the centre of an open ball of $(X, \varrho)$ completely contained in the empty set.

is vacuously true.

(ii) Let $\{A_\lambda \mid \lambda \in \Lambda\}$ be a collection of open subsets of $X$.

If $x \in \bigcup A_\lambda$, then $x \in A_\mu$ for some $\mu \in \Lambda$.

Since $A_\mu$ is an open subset of $X$ with respect to $\varrho$, there is an $\varepsilon > 0$, with $B(x; \varepsilon) \subseteq A_\mu \subseteq \bigcup_{\lambda \in \Lambda} A_\lambda$.

Hence, $\bigcup_{\lambda \in \Lambda} A_\lambda$ is an open subset of $X$.

(iii) Take a finite family of open subsets of $X$, say $\{A_j \mid j = 1, \ldots, n\}$.

If $x \in \bigcap A_j$, then for each $j \in \{1, \ldots, n\}$ there is an $\varepsilon_j$ with $B(x; \varepsilon_j) \subseteq A_j$.

Put $\varepsilon := \min\{\varepsilon_j \mid j = 1, \ldots, n\}$.

Then, $B(x; \varepsilon) \subseteq B(x; \varepsilon_j) \subseteq A_j$ for each $j = 1, \ldots, n$.

Hence $B(x; \varepsilon) \subseteq \bigcap A_j$.

Thus $\bigcap A_j$ is an open subset of $X$, proving (iii).

The significance of the open subsets of a metric space to our investigations lies in the fact that continuity of functions between metric spaces can be defined solely in terms of the open subsets of the metric spaces, without explicit mention of the metrics involved.

**Theorem 5.10.** Let $(X, \varrho)$ and $(Y, \sigma)$ be metric spaces.

The function $f : X \to Y$ is continuous if and only if $f^{-1}(G)$ is an open subset of $X$ whenever $G$ is an open subset of $Y$.

**Proof.** Let $f : X \to Y$ be a continuous function and $G$ an open subset of $Y$.

If $f^{-1}(G) = \emptyset$, then, by Theorem 5.9, $f^{-1}(G)$ is an open subset of $X$.

Otherwise, take $x \in f^{-1}(G)$.

Then $f(x) \in G$, with $G$ open in $Y$.

By Theorem 5.5, there is a positive real number, $\varepsilon$, with $B(f(x); \varepsilon) \subseteq G$. 

By the continuity of $f$ at $x$, there is a positive $\delta$ with $\sigma(f(x), f(y)) < \varepsilon$ whenever $\varrho(x, y) < \delta$.
Thus, $f(y) \in B(f(x); \varepsilon) \subseteq G$ whenever $y \in B(x; \delta)$, or, equivalently,
$$B(x; \delta) \subseteq f^{-1}(G)$$
whence $f^{-1}(G)$ is an open subset of $X$.
For the converse, suppose that the inverse image under $f$ of any open subset of $Y$ is an open
subset of $X$.
Take $x \in X$ and choose any $\varepsilon > 0$.
Since $B(f(x); \varepsilon)$ is an open subset of $Y$, it follows by hypothesis that $f^{-1}(B(f(x); \varepsilon))$ is an open
subset of $X$.
Since $x \in f^{-1}(B(f(x); \varepsilon))$ and $f^{-1}(B(f(x); \varepsilon))$ is open, there is a $\delta > 0$ with $B(x; \delta) \subseteq f^{-1}(B(f(x); \varepsilon))$.
Hence $\sigma(f(x), f(y)) < \varepsilon$ whenever $\varrho(x, y) < \delta$, showing that $f$ is continuous at $x$.
Since $x$ is an arbitrary element of $X$, $f$ is continuous. 
\hfill \Box

1. Exercises

5.1. Let $X = \{f: [0, 1] \rightarrow [0, 1] \mid f \text{ is continuous} \}$.
Show that each of the following three functions defines a metric on $X$
\[d_\infty: X \times X \rightarrow \mathbb{R}_0^+, \quad (f, g) \mapsto \max_{x \in [0, 1]} |f(x) - g(x)|\]
\[d_1: X \times X \rightarrow \mathbb{R}_0^+, \quad (f, g) \mapsto \int_0^1 |f(x) - g(x)| \, dx\]
Represent each $f \in X$ by its graph, so that
\[f: [0, 1] \rightarrow [0, 1], \quad x \mapsto x^{\frac{1}{2}}\]
is represented by

Draw the graph representing a function $g \in X$ such that
\begin{enumerate}
  \item $d_\infty(f, g) = \frac{1}{2}$
  \item $d_1(f, g) = \frac{1}{2}$
\end{enumerate}

5.2. Let $(X, \rho)$ be a metric space and take $A \subseteq X$.
Prove that if $A$ is an open subset of $X$, then the subset $G$ of $A$ is open with respect to $\rho \mid_A$ if and
only if it is open with respect to $\rho$.
Find an example to show that the hypothesis that $A$ be an open subset of $X$ is necessary.

5.3. Show that neither the union nor the intersection of two open balls in a metric space need be
an open ball.

5.4. Take $\mathbb{R}$ with its Euclidean metric $\epsilon$.
Put $X := \mathbb{R} \times \mathbb{R}$.
Draw the open ball $B((0, 0); \varrho)$ when
\begin{enumerate}
  \item $\varrho = \epsilon \times \epsilon$
  \item $\varrho = \varrho_{\mathbb{R} \times \mathbb{R}}$
\end{enumerate}
(iii) \( \varrho = \epsilon + \epsilon \)

5.5. Find an example of metric spaces \((X, \varrho)\) and \((Y, \sigma)\), a continuous function \(f : X \to Y\) and an open ball \(B(a; r) \subseteq X\) such that \(f(B(a; r))\) is not an open subset of \(Y\).

5.6. Find an example of a metric space \((X, \varrho)\) and \(\{A_n \mid n \in \mathbb{N}\}\), a family of open subsets of \(X\), such that

\[
A := \bigcap_{n \in \mathbb{N}} A_n
\]

is not an open subset of \(X\).
Topological Spaces

We have seen that the continuity of a function between metric spaces depends only on the open subsets and not on the metric itself. Since different — that is non-isometric — metrics on a given set can induce the same open sets, the metric cannot be recovered from its open sets.

Thus, to study continuity in general, we need only consider open sets. Theorem 5.9 characterises the set of all open sets in a metric space without reference to the metric. This is the basis for our next definition: we turn Theorem 5.9 into a definition. The characterisation of continuity using only open sets then provides the definition of continuity of functions in this more general setting.

**Definition 6.1.** A **topology** on a set $X$ is $T$, a collection of subsets of $X$, such that

- $\emptyset, X \in T$ 
- $\bigcup_{\lambda \in \Lambda} A_{\lambda} \in T$ for any collection, $\{A_{\lambda} \mid \lambda \in \Lambda\}$, of elements of $T$; 
- $\bigcap_{j=1}^{n} A_{j} \in T$ for any finite collection, $\{A_{j} \mid j = 1, \ldots, n\}$, of elements of $T$.

A **topological space** consists of a set $X$ and a topology $T$ on $X$.

We write $(X, T)$ to denote that $T$ is a topology on the set $X$.

We can express TS3 and TS2 by saying that a topology on a set is a (non-empty) collection of subsets of the set that is closed under finite intersections and under arbitrary unions.

**Example 6.2.** By Theorem 5.9, the open subsets of the metric space $(X, \varrho)$ comprise a topology on $X$. This is the topology on $X$ induced by $\varrho$, or the **metric topology on $X$**.

**Convention.** When we view the metric space $(X, \varrho)$ as a topological space, it is the metric topology which is intended, unless otherwise specifically stated.

We next turn Theorem 5.10 into a definition.

**Definition 6.3.** Given topological spaces $(X, T)$ and $(Y, U)$, the function $f : X \to Y$ is **continuous** if and only if $f^{-1}(G) \in T$ whenever $G \in U$.

$f$ is **continuous at $x \in X$** if and only if for every open subset $G$ of $Y$ containing $f(x)$, there is an open subset $H$ of $X$ containing $x$ such that $f(H) \subseteq G$.

**Convention.** Just as we frequently suppress the metric from our notation in the case of metric spaces so we shall write simply $X$ for $(X, T)$ when there is no danger of confusion.
6. TOPOLOGICAL SPACES

DEFINITION 6.4. The Euclidean topology on $\mathbb{R}^n$ ($n = 1, 2, \ldots$) is the topology on $\mathbb{R}^n$ induced by the Euclidean metric, $\epsilon_n$.

CONVENTION. When we refer to $\mathbb{R}^n$ as a topological space, it is with its Euclidean topology, unless otherwise specified.

OBSERVATION 6.5. When we defined continuity for functions between metric spaces, we first defined continuity at a point: we defined a local notion of continuity. There are functions which are continuous at some points but not at others, for example

$$f : \mathbb{R} \rightarrow \mathbb{R}, \quad x \mapsto \begin{cases} x^2 & \text{for } x \in \mathbb{Q} \\ 0 & \text{otherwise} \end{cases}$$

is continuous at $x = 0$, but nowhere else.

We then defined a function to be continuous without restriction, a global notion of continuity, if and only if it is continuous at every point.

In the case of topological spaces, we first defined continuity globally and then defined continuity-at-a-point separately.

In the case of metric spaces, it is true by definition that a function is globally continuous if and only if it is locally continuous everywhere. While this is still true in the case of topological spaces, it is not simply a matter of definition. It requires proof.

THEOREM 6.6. Given topological spaces $(X, T)$ and $(Y, U)$, the function $f : X \rightarrow Y$ is continuous if and only if $f$ is continuous at every $x \in X$.

PROOF. Suppose that $f$ is continuous and take $x \in X$.

Let $G$ be an open subset of $Y$ with $f(x) \in G$.

By the continuity of $f$, $f^{-1}(G)$ is an open subset of $X$.

Since $x \in f^{-1}(G)$, and $f(f^{-1}(G)) \subseteq G$, $f$ is continuous at $x$.

Since $x$ is arbitrary, $f$ is continuous at every $x \in X$.

Conversely, suppose that $f$ is continuous at every $x \in X$ and let $G \subseteq Y$ be open.

If $f^{-1}(G)$ is empty, then it is open in $X$.

Otherwise choose $x \in f^{-1}(G)$.

Since $f(x) \in G$ and $f$ is continuous at $x$, there is an open subset of $X$, $H_x$, with $x \in H_x \subseteq f^{-1}(G)$.

Since

$$f^{-1}(G) = \bigcup_{x \in f^{-1}(G)} \{x\} \subseteq \bigcup_{x \in f^{-1}(G)} H_x \subseteq \bigcup_{x \in f^{-1}(G)} f^{-1}(G) = f^{-1}(G)$$

$f^{-1}(G)$ a union of open subsets of $X$ and is therefore an open subset of $X$. □

REMARK 6.7. The definition of the continuity of $f : X \rightarrow Y$ at $x \in X$ may seem peculiar. It is tempting to define it analogously to continuity in general, so that $f$ is continuous at $x \in X$ if and only if $f^{-1}(G)$ is open in $X$ whenever $G$ is an open subset of $Y$ containing $f(x)$.

The appeal of this legitimate proposal should be resisted, as the next example shows.
Example 6.8. Consider the function
\[ f: \mathbb{R} \to \mathbb{R}, \quad x \mapsto \begin{cases} 
-1 & \text{if } x < -1 \\
0 & \text{if } -1 \leq x \leq 1 \\
1 & \text{if } x > 1
\end{cases} \]

This function is clearly continuous at 0 — at least in the sense of continuity used in calculus and the theory of metric spaces. If our definition of continuity for topological spaces is to be a generalisation of these notions, then this function should still be continuous at 0.

If we take as the definition of continuity at 0 to be that the inverse image under \( f \) of any open subset of \( \mathbb{R} \) containing \( f(0) = 0 \) be an open subset of \( \mathbb{R} \) containing 0, then this \( f \) would fail to be continuous at 0.

To see why this is the case, take the open subset \( G := \left(-\frac{1}{2}, \frac{1}{2}\right] \) of \( \mathbb{R} \) containing \( f(0) = 0 \).

Then \( f^{-1}(G) = [-1, 1] \), which is not an open subset of \( \mathbb{R} \), as no open ball in \( \mathbb{R} \) containing 1 can be a subset of \([-1, 1]\).

But since \( H := \left]-1, 1\right[ \) is an open subset of \( \mathbb{R} \) satisfying \( 0 \in H \subseteq f^{-1}(G) \), \( f \) is continuous at 0 according to Definition 6.3.

We now establish some basic properties of continuous functions between topological spaces.

Theorem 6.9. Let \((X, \mathcal{T}), (Y, \mathcal{U}), (Z, \mathcal{V})\) be topological spaces.

Take continuous functions \( g: X \to Y \), \( f: Y \to Z \).

(i) \( \text{id}_X: X \to X, \quad x \mapsto x \) is continuous, and

(ii) \( f \circ g: X \to Z \) is continuous.

Proof. (i) \( \text{Observe that for every } G \subseteq X, \text{id}_X^{-1}(G) = G. \)

(ii) \( \text{Let } G \text{ be an open subset of } Z. \)

By the continuity of \( f, f^{-1}(G) \) is an open subset of \( Y. \)

By the continuity of \( g, g^{-1} \left(f^{-1}(G)\right) \) is an open subset of \( X. \)

But \( g^{-1} \left(f^{-1}(G)\right) = (f \circ g)^{-1}(G). \)

Our considerations to here raise some immediate questions.

1. Does every set \( X \) admit a topology?
2. Can a set \( X \) admit more than one topology?
3. Can there be topologies on a set \( X \), which do not arise from a metric on \( X \)?

1. Since every set admits a metric, its discrete metric, \( d_X \), and every metric induces a topology, every set admits a topology. The discrete topology on \( X \), is the topology induced by \( d_X \).

This gives rise to the topological space \((X, \mathcal{D}_X)\) with

\[ \mathcal{D}_X := \mathcal{P}(X), \]

2. On the other hand, it is easy to see that

\[ \mathcal{I}_X := \{\emptyset, X\} \]

also defines a topology on \( X \). This is the indiscrete topology on \( X \).

Convention. We omit the subscript \( X \) from \( \mathcal{D}_X \) and \( \mathcal{I}_X \) when there is no danger of confusion.
When $X$ has at most one element, the discrete and the indiscrete topologies agree. Otherwise they do not.

In other words every set $X$ with at least two elements admits more than one topology.

(3) Let $X$ be a set with at least two elements.

Let $\sigma$ be a metric $X$ and $T$ the metric topology on $X$.

Let $a, b$ be two distinct elements of $X$. Put

$$r = \frac{1}{2} \sigma(a, b)$$

(i) By the definition of the metric topology on $X$, $B(a; r)$ is open.

(ii) Since, by definition, $a \in B(a; r)$, $B(a; r) \neq \emptyset$.

(iii) Since $\sigma(a, b) = r$, $b \notin B(a; r)$, whence $B(a; r) \neq X$.

Thus, $B(a; r) \notin \mathcal{I}_X$, showing that the indiscrete topology on a set with at least two elements cannot be induced by a metric.

Hence, topological spaces are genuinely more general than metric spaces.

Since not every topology on a set arises from a metric, the question whether a given topology arises from a metric is a significant one.

**Definition 6.10.** The topology, $T$, on the set $X$ is **metrisable** if and only if there is a metric, $\sigma$, on $X$ such that the topology on $X$ induced by the metric $\sigma$ is the given topology.

**Example 6.11.** The discrete topology on the set $X$ is metrisable, since it is the topology on $X$ induced by the discrete metric on $X$.

**Remark 6.12.** The only example we have provided of a topology which is not metrisable is the indiscrete topology on a set with at least two elements. This is an extreme example, which may strike the reader as artificial. We presented this example not only because its being a topology is clear from the axioms, but also because, as we show below, it is characterised by a universal property, which is a clear indication that it is a significant construction.

The reader can rest assured that there are naturally occurring topologies which are not metrisable. The **Zariski topology** is an example which plays a central rôle in algebraic geometry. It is defined in terms of the zeroes of polynomials, and is investigated in the exercises in Chapter 10.

The fact that a set with at least two elements can support more than one topology raises the question of comparing different topologies on a given set.

Since a topology on the set $X$ consists of a set of subsets of $X$, we can use the inclusion of sets to define a partially order on the various topologies on a fixed set $X$.

**Definition 6.13.** Given two different topologies $T$ and $U$ on the set $X$, $T$ is coarser (or smaller) than $U$ if and only if $T \subseteq U$, that is to say if and only if every subset of $X$ which is open in the sense of $T$ is also open in the sense of $U$.

$U$ is finer (or larger) than $T$ if and only if $T$ is coarser than $U$.

We can also characterise the fact that $T$ is a coarser topology on $X$ than $U$ in terms of the continuity of functions.

**Theorem 6.14.** Let $T$ and $U$ be topologies on the same set $X$.

The following are equivalent.

(i) $T$ is coarser than $U$.

(ii) Given a topological space $(Y, V)$, a function $f : X \rightarrow Y$ is continuous with respect to $T$ only if it is continuous with respect to $U$. 

(iii) Given a topological space \((Y, V)\), a function \(g: Y \to X\) is continuous with respect to \(U\) only if it is continuous with respect to \(T\).

**Proof.** We prove that (i) \(\Rightarrow\) (ii) and that (i) \(\Rightarrow\) (iii), leaving the rest as an exercise.

(i) \(\Rightarrow\) (ii): Let \(f: X \to Y\) be continuous with respect to \(T\).

Take \(G \in V\).

By the continuity of \(f\) with respect to \(T\), \(f^{-1}(G) \in T\).

Since \(T\) is coarser than \(U\), we have \(T \subseteq U\).

Thus, \(f^{-1}(G) \in U\).

(i) \(\Rightarrow\) (iii): Let \(g: Y \to X\) be continuous with respect to \(U\).

Take \(G \in T\).

Since \(T\) is coarser than \(U\), \(G \in U\).

By the continuity of \(f\) with respect to \(U\), \(f^{-1}(G) \in V\). \(\Box\)

**Corollary 6.15.** The indiscrete topology on a set is the coarsest of all possible topologies on a given set, and the discrete topology is the finest of all topologies on that set.

These two extreme topologies on a given set are characterised by universal properties.

**Theorem 6.16.** The topology, \(T\), on the set \(X\) is the discrete topology on \(X\) if and only if given any topological space \((Y, U)\), every function \(g: X \to Y\) is continuous.

The topology, \(T\), on the set \(X\) is the indiscrete topology on \(X\) if and only if given any topological space \((W, S)\), every function \(f: W \to X\) is continuous.

**Proof.** Suppose that \(T\) is the discrete topology on \(X\).

Let \((Y, U)\) be a topological space and \(f: X \to Y\) a function.

Take \(G \in U\).

Since \(X\) has the discrete topology, every subset, of \(X\), in particular \(f^{-1}(G)\) is open.

Hence \(f\) is continuous.

Conversely, suppose that for every topological space \((Y, U)\), every function \(f: X \to Y\) is continuous.

Put \(Y := X\). Take \(U = D\), the discrete topology on \(X\).

By hypothesis, the function

\[ f: (X, T) \to (X, D), \quad x \mapsto x \]

is continuous.

Take \(G \subseteq X\).

By definition, \(G \in D\).

Since \(f\) is continuous, \(f^{-1}(G) \in T\).

But, by the definition of \(f\), \(f^{-1}(G) = G\).

Hence \(T\) is the discrete topology on \(X\).

The remaining assertions may be proved similarly and this is left as an exercise. \(\Box\)

If \(T\) and \(U\) are topologies on the set \(X\), we can determine which is coarser/finer using a test function.
Theorem 6.17. Let $\mathcal{T}$ and $\mathcal{U}$ be topologies on the same set $X$. $\mathcal{T}$ is coarser than $\mathcal{U}$ if and only if the function

$$f: (X, \mathcal{U}) \longrightarrow (X, \mathcal{T})$$

is continuous.

Proof. The conclusion is immediate, since, by the definition of $f$, $f^{-1}(G) = G$. □

Observation 6.18. Definition 6.3 states that a function is continuous if and only if the inverse image (or pre-image) of any open subset of the co-domain is an open subset of the domain. At first glance, this seems to “go the wrong way”, and that it would be more fitting to require that the image of open subsets of the domain be open subsets of the co-domain.

Definition 6.19. Let $(X, \mathcal{T})$ and $(Y, \mathcal{U})$ be topological spaces. The function $f: X \longrightarrow Y$ is an open mapping if and only if $f(G)$ is open in $Y$ whenever $G$ is open in $X$.

We offer two examples to illustrate the difference between opening mappings and continuous ones.

Example 6.20. Take $X := [0, 2]$ and $Y := [0, 1] \cup [2, 3]$ with their Euclidean metrics. Then

$$f: X \longrightarrow Y, \ x \longmapsto \begin{cases} x & \text{if } x \leq 1 \\ x + 1 & \text{if } x > 1 \end{cases}$$

is an open mapping, but fails to be continuous.

Example 6.21. Take $X = Y = \mathbb{R}$ with the Euclidean metric. The function

$$f: X \longrightarrow Y, \ x \longmapsto 0$$

is continuous, but not an open mapping.

It is possible to characterise continuity in terms of “forward images”, using the notion of the closure of a set, introduced in Chapter 10.

We have seen that to specify a topology on the set $X$, we simply specify which subsets of $X$ are to be considered “open” subsets. This leads to the question:

**Given a topological space $(X, \mathcal{T})$, is there a “nearest” open set to $A \subseteq X$?**

The notion of the interior of a set makes this vague question precise and provides a satisfactory answer.

Convention. Recall that $\mathcal{P}(X)$, the set of all subsets of the set $X$, has a partial ordering $\subseteq$ defined by the inclusion of sets. Given $A, B \subseteq X$

$$A \subseteq B \quad \text{if and only if} \quad A \subseteq B$$

Since a topology, $\mathcal{T}$, on $X$ is a subset of $\mathcal{P}(X)$, $\mathcal{T}$ can also be partially ordered in this manner.

When we speak of maximal or minimal subsets of $X$, it is with respect to this partial ordering, unless otherwise specified.

Definition 6.22. Let $(X, \mathcal{T})$ be a topological space.

The interior of $A \subseteq X$, $\overset{\circ}{A}$, is the maximal subset of $A$ which is an open subset of $X$.

In other words, $H$ is the interior of $A$ if and only if

(i) $H \subseteq A$;

(ii) $H$ is an open subset of $X$;

(iii) if $G \subseteq A$ and $G$ is an open subset of $X$, then $G \subseteq H$.

Theorem 6.23. The subset, $A$, of the topological space, $(X, \mathcal{T})$, has a uniquely determined interior, $\overset{\circ}{A}$, and $A$ is an open subset of $X$ if and only if it is its own interior, that is, $A = \overset{\circ}{A}$.  

Proof. Let $\mathfrak{A}$ denote the set of all subsets of $A$ that are open subsets of $X$, so that
\[ \mathfrak{A} := \{ G \subseteq A \mid G \in \mathcal{T} \}, \]
and put
\[ \hat{A} := \bigcup_{G \in \mathfrak{A}} G. \]
Since $\emptyset \in \mathcal{T}$ and $\emptyset \subseteq A$, $\emptyset \in \mathfrak{A}$, whence $\mathfrak{A} \neq \emptyset$.
Thus, $\hat{A} := \bigcup_{G \in \mathfrak{A}} G$ is a subset of $A$, because it is the union of subsets of $A$.
It is also an open subset of $X$, because it is the union of open subsets of $X$.
Take $H \subseteq A$ with $H \in \mathcal{T}$.
Then $H \in \mathfrak{A}$, whence
\[ H \subseteq \bigcup_{G \in \mathfrak{A}} G = \hat{A} \]
showing that $\hat{A}$ is maximal. \qed

1. Exercises

6.1. Let $(X, \mathcal{T})$ be a topological space.
Suppose that $X = A \cup B$ with $A$ and $B$ open subsets of $X$.
Prove that if $(Y, \mathcal{U})$ is any topological space and that if $f \colon A \to Y$ and $g \colon B \to Y$ are any two continuous functions which agree on $A \cap B$, then there is a unique continuous function
\[ h \colon X \to Y, \quad x \mapsto \begin{cases} f(x) & \text{if } x \in A \\ g(x) & \text{if } x \in B \end{cases} \]
This is expressed by the commutative diagram

```
A \cap B \xrightarrow{i^A} A \\
| \downarrow i_A | \\
B \xrightarrow{i_B} X \\
| \downarrow i_B | \\
Y \\
```

where
\[ i^A : A \cap B \to A \]
\[ i^B : A \cap B \to B \]
\[ i_A : A \to X \]
\[ i_B : B \to A \]
are the inclusions of subsets.
6.2. Find an example of topological spaces $(X, \mathcal{T}), (Y, \mathcal{U})$, together with subsets $A, B$ of $X$ and functions $f: A \rightarrow Y, g: B \rightarrow Y$ such that $X = A \cup B$, $f(x) = g(x)$ whenever $x \in A \cap B$, such that the function

$$h: X \rightarrow Y, \quad x \mapsto \begin{cases} f(x) & \text{if } x \in A \\ g(x) & \text{if } x \in B \end{cases}$$

is not continuous.

6.3. Let $(X, \mathcal{T})$ be a topological space and take $A \subseteq X$.
Prove that there is a maximal open subset of $X$, $\text{ext}(A)$, disjoint from $A$.
This is called the exterior of $A$.

6.4. Let $X$ be a set. Show that

$$\mathcal{T} := \{ A \subseteq X \mid A = \emptyset \text{ or } X \setminus A \text{ is finite} \}$$

is a topology on $X$.
When is this topology metrisable?

6.5. Let $A$ and $B$ be non-empty open subsets of the topological space $(X, \mathcal{T})$.
Suppose there is an $x \in A \cup B$ with the property that every open subset of $X$ which contains $x$ meets both $A$ and $B$, that is, given $G \in \mathcal{T}$, if $x \in G$, then $G \cap A \neq \emptyset$ and $G \cap B \neq \emptyset$.
Prove that $A \cap B \neq \emptyset$.

6.6. Let $\mathcal{T}$ and $\mathcal{U}$ be two different topologies on the set $X$.
Prove that $\mathcal{T}$ is coarser than $\mathcal{U}$ if and only if the function

$$f: (X, \mathcal{U}) \rightarrow (X, \mathcal{T}), \quad x \mapsto x$$

is continuous.
Generating Topologies

A topology on the set $X$ as a set of subsets of $X$ satisfying three conditions. This raises the question:

Which sets of subsets of $X$ form topologies on $X$?

We have already met two extreme topologies on $X$: the indiscrete topology, consisting only of $\emptyset$ and $X$, and the discrete topology, consisting of all subsets of $X$. These are distinct whenever $X$ has at least two elements. They are extreme in the sense that the former is the coarsest possible topology on $X$ and the latter is the finest topology possible on $X$.

We now turn to investigating intermediate topologies.

One approach is to specify that some particular subset or subsets of the set $X$ be open, and to seek a (the?) topology on $X$ which achieves this optimally.

While the discrete topology achieves this, it is independent of the specified subset(s), and so says nothing about these sets — it can have superfluous sets.

We therefore seek the coarsest (smallest) topology on $X$ which renders open the set(s) we specified.

**Example 7.1.** If $A \subseteq X$, then it is easy to verify that $T := \{\emptyset, A, X\}$ is a topology on $X$, and this is the coarsest topology on $X$ with respect to which $A$ is open.

If $A = \emptyset$ or $A = X$, then this is just the indiscrete topology on $X$. Otherwise, it is finer.

**Example 7.2.** If we take two different proper subsets $A, B$ of $X$, neither of which is empty, then $T := \{\emptyset, A \cap B, A, B, A \cup B, X\}$ is not a topology on $X$ unless both $A \cap B$ and $A \cup B$ are also elements of $T$ — as is the case, for example, when $A \subseteq B$ or $A = X \setminus B$. But, in any case, it is easy to verify that $T := \{\emptyset, A \cap B, A, B, A \cup B, X\}$ is always a topology on $X$, and clearly the coarsest one containing both $A$ and $B$.

We could continue in the manner of Examples 7.1 and 7.2 as part of an endeavour to list all possible topologies on a given set $X$. But, even when the set $X$ is finite, presenting an exhaustive list is a task best left to incorrigible masochists. We shall avoid this particular war of attrition, discussing instead the underlying principles.

The pertinent notion is that of a topology generated by a collection of subsets of a given set $X$. This is the coarsest (smallest) topology to contain all the elements of the collection in question. This is analogous to the theory of vector spaces, or the theory of groups, where we investigate the smallest — in the sense of set inclusion — vector space, or group, containing some particular set.

The aptness of the analogy is illustrated by Theorem 7.6 below.

**Definition 7.3.** Let $\mathcal{S}$ be a set of subsets of the set $X$. The topology on $X$ generated by $\mathcal{S}$ is the coarsest topology on $X$ with respect to which each element of $\mathcal{S}$ is open.
Before establishing the existence of a topology generated by a given set of subsets of $X$, we make some general observations.

**Theorem 7.4.** Let $\{T_\lambda \mid \lambda \in \Lambda\}$ be a non-empty family of topologies on the set $X$. Then

$$T := \bigcap_{\lambda \in \Lambda} T_\lambda$$

is also a topology on $X$.

**Proof.**

**T1** Since $\emptyset, X \in T_\lambda$ for each $\lambda \in \Lambda$, and since each $T_\lambda$ is a topology on $X$, $\emptyset, X \in \bigcap_{\lambda \in \Lambda} T_\lambda = T$.

**T2** Take $\{G_\gamma \mid \gamma \in \Gamma\} \subseteq T$ and put $G := \bigcup_{\gamma \in \Gamma} G_\gamma$. Since $G_\gamma \in T_\lambda$ for each $\lambda \in \Lambda$, and since each $T_\lambda$ is a topology on $X$, $G \in T_\lambda$ for each $\lambda \in \Lambda$. Thus $G \in \bigcap_{\lambda \in \Lambda} T_\lambda = T$.

**T3** Take $\{G_1, \ldots, G_n\} \subseteq T$ and put $G := \bigcap_{j=1}^n G_j$. Since each $\{G_1, \ldots, G_n\} \in T_\lambda$ for each $\lambda \in \Lambda$, and since each $T_\lambda$ is a topology on $X$, we have $G \in T_\lambda$ for each $\lambda \in \Lambda$. Thus, $G \in \bigcap_{\lambda \in \Lambda} T_\lambda = T$. □

By contrast, the union of even only two topologies on a set $X$ need not be a topology on $X$.

**Example 7.5.** Take $X := \{1, 2, 3\}$.

Then $T := \{\emptyset, \{1\}, X\}$ and $U := \{\emptyset, \{2\}, X\}$ are both topologies on $X$.

Then $T \cup U = \{\emptyset, \{1\}, \{2\}, X\}$ is not a topology on $X$, because while it contains both $\{1\}$ and $\{2\}$, it does not contain their union, $\{1, 2\}$.

**Theorem 7.6.** Let $S$ be a non-empty set of subsets of the set $X$, so that $S \subseteq \mathcal{P}(X)$.

There is a unique coarsest topology on $X$, $T_S$, with respect to which every element of $S$ is open.

In other words,

(i) $T_S$ is a topology on $X$ with $S \subseteq T_S$.

(ii) If $U$ is a topology on $X$ with $T_S \subseteq U$, then $T_S$ is not finer than $U$, that is $T_S \subsetneq U$.

**Proof.** Let $\mathfrak{A}$ denote the set of all topologies on $X$ with respect to which all the elements of $S$ are open, so that

$$\mathfrak{A} := \{T \mid T \text{ is a topology on } X \text{ and } S \subseteq T\}$$

Since $\mathcal{P}(X)$ is the discrete topology on $X$ and $S \subseteq \mathcal{P}(X)$, $\mathfrak{A} \neq \emptyset$.

Put

$$T_S := \bigcap_{T \in \mathfrak{A}} T.$$

By Theorem 7.4, $T_S$ is a topology on $X$. 
By construction, \( S \subseteq T \).
Let \( U \) be a topology on \( X \) with \( T_S \subseteq U \).
Then \( U \in \mathcal{A} \), whence
\[
T_S := \bigcap_{T \in \mathcal{A}} T \subseteq U \]
\( \square \)

**Corollary 7.7.** The topology generated by a set of subsets of \( X \) is the intersection of all topologies on \( X \) with respect to which all the subsets of \( X \) in \( S \) are open.

**Example 7.8.** If \( S := \{\emptyset\} \), then \( T_S \) is the indiscrete topology on \( X \). The same is true when \( S := \{X\} \), or \( S := \{\emptyset, X\} \).

Thus, the same topology can be generated in different ways.

**Example 7.9.** If \( S := \mathcal{P}(X) \), then \( T_S \) is the discrete topology on \( X \).

**Example 7.10.** If \( S := \{A\} \), with \( A \neq \emptyset, X \), then \( T_S = \{\emptyset, A, X\} \).

While Theorem 7.6 guarantees the existence of the topology generated by any collection of subsets, it does not provide a usable general method for constructing it from the given collection of subsets. We now turn to remediying this,

**Definition 7.11.** The collection, \( B \), of subsets of the topological space \((X, T)\) is a base, or basis, for the topology if and only if every non-empty open subset of \( X \) is a union of elements of \( B \).

The collection \( S \) of subsets of the topological space \((X, T)\) is a sub-base, or sub-basis, for the topology if and only if every non-empty open subset of \( X \) is a union of a set of finite intersections of elements of \( S \).

The elements of a base \( B \) are called basic open subsets of \( X \), and the elements of a sub-base \( S \) are called sub-basic open subsets of \( X \).

**Observation 7.12.** \( S \) is a sub-basis for a given topology \( T \) if and only if the set of all finite intersections of elements of \( S \) forms a basis for \( T \).

**Example 7.13.** By the definition of the metric topology on a metric space, the set of open balls of a metric space comprises a basis for its topology.

**Example 7.14.** \( \{[a, b) | a, b \in \mathbb{R}, a < b\} \) is a basis for the Euclidean topology on \( \mathbb{R} \), because these subsets of \( \mathbb{R} \) are precisely the open balls of the Euclidean metric.

On the other hand,
\[
\{ \] \(-\infty, b) | b \in \mathbb{R} \} \cup \{ [a, \infty) | a \in \mathbb{R} \}
\]
is a sub-basis, because \( ] -\infty, b [ \cap a, \infty [ = ]a, b [ \) whenever \( a < b \) and is empty otherwise.

Hence every subset of \( \mathbb{R} \) which is open in the Euclidean sense can be written as a union of sets each of which is the intersection of two suitable sub-basic open sets.

A significant advantage bestowed by knowing basic and/or sub-basic open sets is that these suffice to determine that a given function is continuous.

**Theorem 7.15.** Let \((X, \mathcal{T})\) and \((Y, \mathcal{U})\) be topological spaces and \( B \) (resp. \( S \)) a base (resp. a sub-base) for the topology on \( Y \).

The function \( f : X \longrightarrow Y \) is continuous if and only if \( f^{-1}(G) \) is open in \( X \) whenever \( G \) is a basic (resp. sub-basic) open subset of \( Y \).

**Proof.** We prove the theorem for the case of basic open sets and leave the case of sub-basic open sets as an exercise.
Suppose that \( f \) is continuous.

Let \( G \) be a basic open subset of \( Y \).

Since \( G \) is an open subset of \( Y \) and \( f \) is continuous, \( f^{-1}(G) \) is an open subset of \( X \).

Conversely, suppose that the inverse images of basic open subsets of \( Y \) are open subsets of \( X \).

Let \( G \) be an open subset of \( Y \).

Then \( G = \bigcup_{\lambda \in \Lambda} G_{\lambda} \) for suitable basic open subsets, \( G_{\lambda} \ (\lambda \in \Lambda) \), of \( Y \). Thus

\[
f^{-1}(G) = f^{-1}\left( \bigcup_{\lambda \in \Lambda} G_{\lambda} \right) = \bigcup_{\lambda \in \Lambda} f^{-1}(G_{\lambda})
\]

Since, by hypothesis, each \( f^{-1}(G_{\lambda}) \) is an open subset of \( X \), \( f^{-1}(G) \) is an open subset of \( X \), being a union of open subsets of \( X \).

Observation 7.16. Any collection \( \mathcal{S} \) whatsoever of subsets of the non-empty set \( X \) can serve as a sub-basis for a topology on \( X \). This particular topology is the topology generated by \( \mathcal{S} \). If \( \mathcal{S} \) is closed under finite intersections, then \( \mathcal{S} \) is, in fact, a basis.

A basis need not be closed under finite intersections, as is shown by the example of all open balls in the metric space \((\mathbb{R}^2, \epsilon)\), where \( \epsilon \) is the Euclidean metric. Take \( x := (0,0) \) and \( y := (1,0) \) then \( B(x; 2) \cap B(y; 2) \) is a non-empty open subset of \( \mathbb{R}^2 \), but it is not a Euclidean open ball.

Observation 7.17. An analogy has appeared here between constructions in the theory linear algebra and the theory of topological spaces, with vectors corresponding to open subsets, and vector sub-spaces to topologies.

(1) (a) In linear algebra, the intersection of any collection of sub-spaces of a given vector space is again a sub-space, but the union need not be.

(b) In topology, the intersection of any collection of topologies on a set is again a topology on that set, but the union need not be.

(2) (a) In linear algebra, the vector sub-space generated by a set of vectors is the smallest vector sub-space containing the given vectors. It is the intersection of all vector sub-spaces containing the given vectors.

(b) In topology, the topology generated by a set of subsets is the smallest topology containing the given subsets. It is the intersection of all topologies containing the given subsets.

We have shown how to generate a topology on a set by requiring specific subsets to be open. An alternative is to specify a topology on the set \( X \) is to require that a given collection of functions be continuous.

If \((W, \mathcal{S})\) and \((Y, \mathcal{U})\) are topological spaces and we have functions \( f: W \rightarrow X \) and \( g: X \rightarrow Y \), then endowing \( X \) with the indiscrete topology ensures that \( f \) is continuous, whereas the discrete topology on \( X \) guarantees the continuity of \( g \).

This is an indelicate way of forcing the continuity of \( f \) and \( g \), for the topologies chosen reflect nothing of the nature of \( f, g, W \) or \( Y \).

To ensure that \( g \) be continuous in a manner which reflects the rôle of \( g \) and \((Y, \mathcal{U})\), we take as the topology on \( X \) induced by \( g \) the coarsest topology on \( X \) containing all subsets of \( X \) of the form \( g^{-1}(H) \) with \( H \in \mathcal{U} \). In other words, we take as sub-basis \( \{g^{-1}(H) \mid H \in \mathcal{U}\} \).

Note that

\[
T1: \emptyset = g^{-1}(\emptyset) \text{ and } X = g^{-1}(Y)
\]
Thus, our sub-basis is already a topology, and it is therefore the coarsest topology to render \( g \) continuous. We summarise this in the form of a definition.

**Definition 7.18.** Let \( X \) be a set and \( (Y, \mathcal{U}) \) a topological space.

The topology on \( X \) induced by the function \( g: X \rightarrow Y \) is \( \mathcal{T}^g \), the coarsest topology on \( X \) with respect to which \( g \) is continuous.

Thus, \( G \in \mathcal{T}^g \) if and only if \( G = g^{-1}(H) \) for some \( H \in \mathcal{U} \).

An important case is when \( X \) is a subset of \( Y \).

**Definition 7.19.** Let \( (Y, \mathcal{U}) \) be a topological space. The subspace topology on the subset \( X \) of \( Y \) is the topology generated on \( X \) by the inclusion function \( i_X: X \rightarrow Y \).

In a similar vein, let \( (W, \mathcal{S}) \) be a topological space. Then the function \( f: W \rightarrow X \) is continuous if and only if \( f^{-1}(G) \) is open in \( W \) whenever \( G \) is open in \( X \). So we take as our starting point the set \( \mathcal{T}_f \), of all those subsets of \( X \) whose inverse image under \( f \) is open in \( W \), so that

\[
\mathcal{T}_f = \{ G \subseteq X \mid f^{-1}(G) \in \mathcal{S} \}
\]

As

\[
\begin{align*}
\text{(T1):} & \quad f^{-1}(\emptyset) = \emptyset \text{ and } f^{-1}(X) = W \\
\text{(T2):} & \quad f^{-1}\left( \bigcup_{\lambda \in \Lambda} G_{\lambda} \right) = \bigcup_{\lambda \in \Lambda} f^{-1}(G_{\lambda}) \\
\text{(T3):} & \quad f^{-1}\left( \bigcap_{j=1}^{n} G_j \right) = \bigcap_{j=1}^{n} f^{-1}(G_j)
\end{align*}
\]

\( \mathcal{T}_f \) is a topology on \( X \).

Moreover, it is clear that no strictly finer (larger) topology on \( X \) can render \( f \) continuous, for such a topology on \( X \) must contain at least one subset of \( X \) whose inverse image under \( f \) is not open in \( W \). Our next definition summarises this.

**Definition 7.20.** Let \( X \) be a set and \( (W, \mathcal{S}) \) a topological space. The topology on \( X \) induced by the function \( f: W \rightarrow X \), is \( \mathcal{T}_f \), the finest topology on \( X \) with respect to which \( f \) is continuous.

Thus,

**Lemma 7.21.** Let \( X \) be a set, \( (W, \mathcal{S}) \) a topological space and \( f: W \rightarrow X \) a function. Then \( G \in \mathcal{T}_f \) if and only if \( f^{-1}(G) \in \mathcal{S} \).

While we have restricted our attention to topologies induced by a single function from one space to another above, similar considerations allow us to define topologies generated by families of functions.

**Definition 7.22.** Let \( X \) be a set. Let \( \{(W_{\lambda}, \mathcal{S}_{\lambda}) \mid \lambda \in \Lambda\} \) and \( \{(Y_{\mu}, \mathcal{U}_{\mu}) \mid \mu \in M\} \) be families of topological spaces.

The topology on \( X \) induced by the functions \( f_{\lambda}: W_{\lambda} \rightarrow X \ (\lambda \in \Lambda) \) is \( \mathcal{T}_{(f_{\lambda})} \), the finest topology on \( X \) with respect to which each \( f_{\lambda} \) is continuous.

The topology on \( X \) induced by the functions \( g_{\mu}: Y_{\mu} \rightarrow X \ (\mu \in M) \) is \( \mathcal{T}^{(g_{\mu})} \), the coarsest topology on \( X \) with respect to which each \( g_{\mu} \) is continuous.
1. Exercises

7.1. Consider the set $X$ and a family of topological spaces $\{(Y_\mu, U_\mu) \mid \mu \in M\}$. Let $g_\mu : X \rightarrow Y_\mu$ ($\mu \in M$) be a family of functions, with $M$ containing more than one element.

(a) Show that the set of all subsets, $G$, of $X$ for which there is a $\mu \in M$ and an $H_\mu \in U_\mu$ with $G = g_\mu^{-1}(H_\mu)$ need not be a topology on $X$.

(b) Show that the set of all subsets $G$ of $X$ such that for every $\mu \in M$ there is an $H_\mu \in U_\mu$ with $G = g_\mu^{-1}(H_\mu)$ need not be a topology on $X$.

7.2. Show that

$$\left\{ \{ (x, y) \in \mathbb{R}^2 \mid a \leq x < b, c \leq y < d \} \mid a, b, c, d \in \mathbb{R}, a < b, c < d \right\}$$

is a basis for a topology on $\mathbb{R}^2$.

7.3. Let $S$ be a sub-base for the topology $U$ on the set $Y$.

Let $(X, T)$ be any topological space.

Show that the function $f : (X, T) \rightarrow (Y, U)$ is continuous if and only if $f^{-1}(G) \in T$ whenever $G \in S$.

7.4. Let $(Y, U)$ be a topological space, and $\varphi : X \rightarrow Y$ a bijection.

Show that the topology on $X$ induced by $\varphi : X \rightarrow Y$ coincides with the topology on $X$ induced by $\varphi^{-1} : Y \rightarrow X$.

7.5. Take $\mathbb{R}$ with its Euclidean topology, and let $S^1$ be the circle

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

Determine the topology on $S^1$ induced by the function $f : \mathbb{R} \rightarrow S^1$, $t \mapsto (\cos 2\pi t, \sin 2\pi t)$.

7.6. Let $(X, T)$ and $(Y, U)$ be topological spaces and $S$ a sub-base for the topology on $Y$.

Prove that the function $f : X \rightarrow Y$ is continuous if and only if $f^{-1}(G)$ is open in $X$ whenever $G$ is a sub-basic open subset of $Y$. 
Over the centuries, monumental upheavals in science have emerged time and again from following the leads set out by mathematics.

Brian Greene

Chapter 8

Constructing Topological Spaces from Given Ones

We have seen how to define topologies on a set $X$ to ensure that a given family of functions be continuous; in the former case we take the coarsest (smallest) topology on $X$ which renders each of the functions continuous and in the latter the finest (largest) one.

We illustrate this principle in practice.

1. Topological Sub-Spaces

Our first construction is that of a topology induced by the topological space $(X, T)$ on the subset $A$ of $X$.

By Theorem 1.25, $A$ is a subset of $X$ if and only if

$$i_A^X: A \rightarrow X, \ a \mapsto a$$

is a function. It is natural to apply Definition 7.18 to $i_A^X$ (which we simplify to $i$) in order to define the sub-space topology on $A$.

**Definition 8.1.** Let $(X, T)$ be a topological space and $A$ a subset of $X$.
The sub-space topology on $A$ is the topology on $A$ induced by the inclusion function

$$i: A \rightarrow X, \ a \mapsto a.$$ 

**Theorem 8.2.** Let $(X, T)$ be a topological space and $A$ a subset of $X$.
The subset $G$ of $A$ is open in the sub-space topology on $A$ if and only if $G = A \cap H$ for some open subset $H$ of $X$.

**Proof.** By Definitions 8.1 and 7.18, $G \subseteq A$ is open in the sub-space topology on $A$ if and only if $G = i^{-1}(H)$ for some open subset $H$ of $X$. But

$$i^{-1}(H) = \{x \in A \mid i(x) \in H\}$$

$$= \{x \in A \mid x \in H\}$$

$$= A \cap H.$$ 

\square

**Corollary 8.3.** If $A$ is an open subset of $X$, then $G \subseteq A$ is open in the sub-space topology on $A$ if and only if $G$ is an open subset of $X$. 

63
Proof. If $G \subseteq A$ is an open subset of $X$, then it is open in the sub-space topology since $G = A \cap G$. (Notice that $A$ does not need to be open for this.)

Conversely, let $G \subseteq A$ be open in the sub-space topology. Then $G = A \cap H$ for some open subset $H$ of $X$. As $A$ is an open subset of $X$, $G$ is the intersection of two open subsets of $X$. Thus, $G$ is an open subset of $X$. □

Observation 8.4. In the case of metric spaces, the sub-space metric was obtained by restricting the domain of definition of the metric to the subset in question. Similarly, in the case of general topological spaces the sub-space topology is obtained by restricting the open sets to the subset in question. Thus, when $A$ is a subset of the metric space $(X, \varrho)$, $X$ induces a topology on $A$ in two different ways.

(1) Take the topology on $A$ induced by the metric on $A$, which is the sub-space metric.

(2) Take the topology on $A$ as subspace of $X$, where $X$ has the topology induced by the metric.

It is left as an exercise to show that these topologies on $A$ coincide.

2. The Product of Topological Spaces

The next illustration also mimics a construction we carried out on metric spaces, namely the construction of the cartesian product of two metric spaces.

Definition 8.5. Let $(X, \mathcal{T})$ and $(Y, \mathcal{U})$ be topological spaces. The product topology on $X \times Y = \{ (x, y) | x \in X, y \in Y \}$ is the coarsest topology which renders continuous the projections of the product, $pr_X: X \times Y \to X$, $(x, y) \mapsto x$ and $pr_Y: X \times Y \to Y$, $(x, y) \mapsto y$.

Thus the product topology on $X \times Y$ has as sub-basis

$$\{ pr_X^{-1}(G) | G \in \mathcal{T} \} \cup \{ pr_Y^{-1}(H) | H \in \mathcal{U} \}.$$ 

If $G \subseteq X$, then $pr_X^{-1}(G) = G \times Y$.

Similarly, $pr_Y^{-1}(H) = X \times H$.

Since for $A, B \subseteq X$ and $C, D \subseteq Y$, $(A \times C) \cap (B \times D) = (A \cap B) \times (C \cap D)$,

$$pr_X^{-1}(G) \cap pr_Y^{-1}(H) = G \times H$$

A direct inductive argument now shows that

$$\left( \bigcap_{i=1}^{m} pr_X^{-1}(G_i) \right) \cap \left( \bigcap_{j=1}^{n} pr_Y^{-1}(H_j) \right) = \left( \bigcap_{i=1}^{m} G_i \right) \times \left( \bigcap_{j=1}^{n} (H_j) \right),$$

which is again of the form $G \times H$ with $G$ open in $X$ and $H$ open in $Y$.

We have thus shown that the set of all finite intersections of sub-basic sets for the product topology on $X \times Y$ is the set $\{ G \times H | G \in \mathcal{T}, H \in \mathcal{U} \}$.

Hence $\{ G \times H | G \in \mathcal{T}, H \in \mathcal{U} \}$ is a basis for the product topology on $X \times Y$.

However, in general, it is not itself a topology, since $(A \times C) \cup (B \times D)$ cannot always be written in the form $E \times F$ (cf. Exercise 1.6).

The next theorem summarises the above.
**Theorem 8.6.** Let \((X, T)\) and \((Y, U)\) be topological spaces. Then \(L \subseteq X \times Y\) is open in the product topology on \(X \times Y\) if and only if
\[
L = \bigcup_{\lambda \in \Lambda} (G_\lambda \times H_\lambda)
\]
for some \(\{G_\lambda \mid \lambda \in \Lambda\} \subseteq T\) and \(\{H_\lambda \mid \lambda \in \Lambda\} \subseteq U\).

**Proof.** Exercise. \(\square\)

The product of topological spaces enjoys a universal property.

**Theorem 8.7.** Let \((X, T)\) and \((Y, U)\) be topological spaces. Given any topological space \((W, S)\) and continuous functions \(f: W \rightarrow X\) and \(g: W \rightarrow Y\), there is a unique continuous function \(h: W \rightarrow X \times Y\) with \(f = pr_X \circ h\) and \(g = pr_Y \circ h\).

This is expressed by the following diagram of continuous functions.

\[
\begin{array}{ccc}
W & \xrightarrow{h} & X \times Y \\
\parallel & \downarrow pr_X & \downarrow pr_Y \\
Y & & \\
\end{array}
\]

**Proof.** Let \((W, S)\) be a topological space. Takes functions \(f: W \rightarrow X\) and \(g: W \rightarrow Y\). By Theorem ?? there is a unique function
\[
h: W \rightarrow X \times Y
\]
with \(pr_X \circ h = f\) and \(pr_Y \circ h = g\).

It remains to show that \(h\) is continuous.

By definition, the sets of the form \(pr_X^{-1}(G)\) and \(pr_Y^{-1}(H)\), with \(G\) an open subset of \(X\) and \(H\) an open subset of \(Y\) form a sub-base for the product topology on \(X \times Y\).

Hence, by Theorem 7.15, it is enough to show that for each \(G \in T\) and each \(H \in U\), \(h^{-1}(pr_X^{-1}(G))\) and \(h^{-1}(pr_Y^{-1}(H))\) are open in \(W\).

\[
h^{-1}(pr_X^{-1}(G)) = (h^{-1} \circ pr_X^{-1})(G) = (pr_X \circ h)^{-1}(G) = f^{-1}(G),
\]
which is an open subset of \(W\) since \(f\) is continuous.

A similar argument shows that \(h^{-1}(pr_Y^{-1}(H))\) is also open in \(W\). \(\square\)

The product of any number — finite or infinite — of topological spaces can be defined. It enjoys the universal property of the last theorem.

**Definition 8.8.** The product of the family, \(\{(X_\lambda, T_\lambda) \mid \lambda \in \Lambda\}\), of topological spaces, is the set
\[
\prod_{\lambda \in \Lambda} X_\lambda := \{(x_\lambda)_{\lambda \in \Lambda} \mid x_\lambda \in X_\lambda \text{ for each } \lambda \in \Lambda\}
\]
endowed with the product topology, that is, the topology induced on \(\prod X_\lambda\) the family
\[
\{pr_\mu: \prod X_\lambda \rightarrow X_\mu \mid \mu \in \Lambda\}
\]
of natural projections
\[ pr_\mu : (x_\lambda)_{\lambda \in \Lambda} \mapsto x_\mu. \]

**Lemma 8.9.** Let \( \{(X_\lambda, T_\lambda) \mid \lambda \in \Lambda\} \) be a family of topological spaces.
The open subsets of \( \prod X_\lambda \) in the product topology are of the form
\[ \bigcup_{\alpha \in \Lambda} \left( \prod_{\lambda \in \Lambda} G_{\alpha \lambda} \right) \]
where \( G_{\alpha \lambda} = X_\lambda \) for all but finitely many \( \lambda \)'s.

**Proof.** Exercise. \( \Box \)

**Theorem 8.10.** Let \( \{(X_\lambda, T_\lambda) \mid \lambda \in \Lambda\} \) be a family of topological spaces.
Put \( X := \prod_{\lambda \in \Lambda} \) and endow \( X \) with the product topology \( T \).
Let \( (W, U) \) be a topological space and \( \{f_\lambda : W \to X_\lambda \mid \lambda \in \Lambda\} \) a family of continuous functions.
There is a unique continuous function
\[ f : W \to X \]
such that for each \( \lambda \in \Lambda \)
\[ f_\lambda = pr_\lambda \circ f \]

**Proof.** Exercise. \( \Box \)

**3. The Quotient of a Topological Space**

Our third application is to the quotient of a topological space.
Let \( (X, T) \) be a topological space and \( \sim \) an equivalence relation on \( X \).
In Section 5 of Chapter 1 we constructed the set of all equivalence classes
\[ X/\sim := \{[x] \mid x \in X\} \]
where
\[ [x] := \{t \in X \mid t \sim x\} \]
and the natural projection
\[ \eta : X \to X/\sim, \quad x \mapsto [x] \]

**Definition 8.11.** The quotient topology on \( X/\sim \) is the topology induced by the natural projection
\[ \eta : X \to X/\sim \]

By Lemma 7.21, \( H \subseteq X/\sim \) is open in the quotient topology if and only if \( \eta^{-1}(H) \) is an open subset of \( X \).

**Theorem 8.12.** Let \( (X, T) \) be a topological space and \( \sim \) an equivalence relation on \( X \).
Let \( \eta : X \to X/\sim \) be the natural projection onto the quotient space.
Take \( X/\sim \) with its quotient topology.

For any topological space \( (Y, U) \) and any continuous function \( f : X \to Y \) with \( f(x) = f(x') \)
whenever \( x \sim x' \), there is a unique continuous function
\[ \tilde{f} : X/\sim \to Y \]
with
\[ f = \tilde{f} \circ \eta \]
This is expressed by the commutative diagram of continuous functions

\[
\begin{array}{ccc}
X & \xrightarrow{f} & Y \\
\eta \downarrow & & \downarrow \exists ! \tilde{f} \\
X/\sim & & \\
\end{array}
\]

**Proof.** By the universal property of \( \eta: X \to X/\sim \), there is a unique function \( \tilde{f}: X/\sim \to Y \) with \( f = \tilde{f} \circ \eta \).

It remains only to show that \( \tilde{f} \) is continuous. This is left as an exercise. \( \square \)

### 4. Exercises

8.1. Let \((X, \varrho)\) be a metric space and \( \mathcal{T} \) the topology on \( X \) defined by \( \varrho \).
Take \( A \subseteq X \).
Prove that the topology on \( A \) defined by \( \varrho|_A \) is the sub-space topology.

In other words, the subspace topology coincides with the topology induced by the sub-space metric.

8.2. Prove that the product topology induced on \( \mathbb{R}^2 \) by the Euclidean topology on \( \mathbb{R} \) is the Euclidean topology on \( \mathbb{R}^2 \).

8.3. Let \((X, \varrho)\) and \((Y, \sigma)\) be metric spaces.
Let \( \mathcal{T} \) be the topology on \( X \) induced by \( \varrho \) and \( \mathcal{U} \) the topology on \( Y \) induced by \( \sigma \).
Prove that the product metric on \( X \times Y \) induces the product topology.

8.4. Let \((X, \mathcal{T})\) be a topological space and \( \sim \) an equivalence relation on \( X \).
Put
\[
[x] := \{ t \in X \mid t \sim x \},
\]
\[
X/\sim := \{ [x] \mid x \in X \}
\]
and
\[
\eta: X \to X/\sim, \quad x \mapsto [x].
\]
Endow \( X/\sim \) with the *quotient topology*, that is the topology induced by \( \eta \).

Prove that if \((Y, \mathcal{U})\) is a topological space, then \( f: X \to Y \) with \( f(x') = f(x) \) whenever \( x' \sim x \) is continuous if and only if \( \tilde{f} \) is continuous, where \( \tilde{f} \) is the function \( X/\sim \to Y \) induced by \( f \).

This is expressed by the commutative diagram

\[
\begin{array}{ccc}
X & \xrightarrow{f} & Y \\
\eta \downarrow & & \downarrow \exists ! \tilde{f} \\
X/\sim & & \\
\end{array}
\]
Let \( \{(X_\lambda, \mathcal{U}_\lambda) \mid \lambda \in \Lambda \} \) be a family of topological spaces.

Let \( \prod X_\lambda \) be the (Cartesian) product of the sets \( X_\lambda (\lambda \in \Lambda) \).

For each \( \mu \in \Lambda \) let \( pr_\mu \) be the canonical projection onto the \( \mu \)-th factor — that is

\[
pr_\mu : \prod_{\lambda \in \Lambda} X_\lambda \rightarrow X_\mu, \quad (x_\lambda)_{\lambda \in \Lambda} \mapsto x_\mu.
\]

The product topology on \( \prod X_\lambda \) is the topology induced by \( \{ pr_\mu \mid \mu \in \Lambda \} \).

(i) Prove that \( G \subseteq \prod X_\lambda \) is open in the product topology if and only if \( G \) may be written as

\[
G = \bigcup_{\alpha \in A} G_\alpha
\]

for some suitable collection \( \{ G_\alpha \mid \alpha \in A \} \) of subsets of \( \prod X_\lambda \), with the property that for each \( \alpha \in A \),

\[
G_\alpha = \prod_{\lambda \in \Lambda} G_{\alpha \lambda}
\]

where each \( G_{\alpha \lambda} \) is an open subset of \( X_\lambda \) and \( G_{\alpha \lambda} = X_\lambda \) for all but finitely many \( \lambda \)'s.

(ii) Prove that for any \( \mu \in \Lambda \) the canonical projection \( pr_\mu : \prod X_\lambda \rightarrow X_\mu \) is an open mapping, that is to say, \( pr_\mu(G) \) is an open subset of \( X_\mu \) whenever \( G \) is an open subset of \( \prod X_\lambda \).
For most people, the major hurdle in grasping modern insights into the nature of the universe is that these developments are usually phrased using mathematics.

Brian Greene

Chapter 9

Homeomorphism

We have defined topological spaces and continuous functions between them. We have also seen that there is an ample supply of topological spaces, even more than there are metric spaces. It remains to explain when two topological spaces are “essentially the same”. Intuitively they need to be essentially the same as sets and corresponding functions should either both be continuous or both discontinuous.

The mathematical formulation of this is the notion of homeomorphism.

Definition 9.1. A homeomorphism between the topological spaces \((X, T)\) and \((Y, U)\) is a continuous function \(f : X \rightarrow Y\) for which there is a continuous function \(g : Y \rightarrow X\) such that \(f \circ g = id_Y\) and \(g \circ f = id_X\).

\((X, T)\) and \((Y, U)\) are said to be homeomorphic if there is a homeomorphism between them. In such a case we write \(X \sim Y\).

Observation 9.2. The metric spaces \((X, \rho)\) and \((Y, \sigma)\) can be homeomorphic without being isometric. For example take \(\mathbb{R}\) with the Euclidean metric, \(\epsilon\), so that \(\epsilon(x, y) = |x - y|\) and define \(\bar{\epsilon}(x, y) := \frac{|x - y|}{1 + |x - y|}\). Then, by Exercise 2.10

\[ f : (\mathbb{R}, \bar{\epsilon}) \rightarrow (\mathbb{R}, \epsilon), \quad x \mapsto x \]

\[ g : (\mathbb{R}, \epsilon) \rightarrow (\mathbb{R}, \bar{\epsilon}), \quad x \mapsto x \]

are both continuous and so, being mutually inverse, they are homeomorphisms. But no function \(f : \mathbb{R} \rightarrow \mathbb{R}\) can be an isometry, because \(\bar{\epsilon}(x, y) < 1\) for all \(x, y \in \mathbb{R}\), so that \(\bar{\epsilon}(f(0), f(1)) < 1 = \epsilon(0, 1)\).

Theorem 9.3. Let \(f : X \rightarrow Y\) be a function between the topological spaces \((X, T)\) and \((Y, U)\). Then the following are equivalent.

(i) \(f\) is a homeomorphism.
(ii) \(f\) is bijective, with both \(f\) and \(f^{-1}\) continuous.
(iii) \(f\) is a bijective, continuous open mapping.
(iv) \(f\) is bijective and induces a bijection between \(T\) and \(U\).

Proof. The equivalence of (i) and (ii) is a restatement of Definition 9.1. Recall that the function \(f : X \rightarrow Y\) induces the function \(f^* : \mathcal{O}(Y) \rightarrow \mathcal{O}(X)\), \(B \mapsto f^{-1}(B)\). We observe that the continuity of \(f\) according to Definition 6.3 is equivalent to the condition that \(f^*(H) \in \mathcal{T}\) whenever \(H \in \mathcal{U}\), or that \(f^*\) restricts to a function \(f^*|_U : \mathcal{U} \rightarrow \mathcal{T}\).
This is the induced function in (iv).
The rest is left as an exercise. \[ \square \]

**Observation 9.4.** The condition that $f^{-1}$ be continuous is necessary and does not follow from the other conditions, as shown by Theorem 6.14.

We provide an alternative illustration using an example from calculus.

Take $X := [0, 1] \cup [2, 3]$ and $Y := [0, 2]$, Euclidean metric. Then the function

$$f: X \to Y, \quad x \mapsto \begin{cases} x & \text{if } 0 \leq x < 1 \\ x - 1 & \text{if } 2 \leq x \leq 3 \end{cases}$$

is bijective and continuous, but not a homeomorphism, because $f^{-1}$ is not continuous at $1 \in [0, 2]$.

One way to verify this is to apply the “$\varepsilon - \delta$” definition of continuity. We shall later give two additional proofs of this fact, one using connectedness, the other compactness.

Two spaces which are homeomorphic may admit many different homeomorphisms.

**Example 9.5.** Take $\mathbb{R}^2$ with its Euclidean topology and represent its elements by column vectors. Then multiplication by invertible real $2 \times 2$ matrix defines an automorphism of $\mathbb{R}^2$ — that is, a homeomorphism of $\mathbb{R}^2$ to itself.

By contrast, the product of given two topological spaces $(X, T)$ and $(Y, U)$ is uniquely determined up to unique homeomorphism by its universal property, as expressed in the commutative diagram

\[
\begin{array}{ccc}
W & \xrightarrow{h} & X \times Y \\
\downarrow f & & \downarrow \text{id}_X × \text{id}_Y \\
X & & Y
\end{array}
\]

To show that this determines $(X \times Y, \text{id}_X \times \text{id}_Y)$ uniquely up to unique homeomorphism, let $W$, together with $q_X: W \to X$, $q_Y: W \to Y$, also enjoys the same universal property. The following commutative diagram suffices as proof.

\[
\begin{array}{cccc}
Y & \xrightarrow{id_Y} & Y & \xrightarrow{id_Y} & Y \\
\downarrow q_Y & & \downarrow q_Y & & \downarrow q_Y \\
W & \xrightarrow{\varphi} & X \times Y & \xrightarrow{\psi} & W \\
\downarrow q_X & & \downarrow \text{id}_X & & \downarrow \text{id}_X \\
X & \xrightarrow{id_X} & X & \xrightarrow{id_X} & X
\end{array}
\]

The proof can be achieved by following the steps in the proof of Theorem 1.54, where we showed that the universal property of the Cartesian product with its canonical projections determines it uniquely up to a unique bijection. The only modifications required are the insertion of the adjective “continuous” before the word “function”, and the replacement of “bijection” by “homeomorphism”.

**Remark 9.6.** The last proof illustrates the power of commutative diagrams, for an argument which can be formulated in terms of “diagram chasing” applies whenever the diagram applies.
In our case, we have seen that an argument involving only sets and functions applies *mutatis mutandis*, to topological spaces and continuous functions. It also applies to vector spaces and linear transformations, to groups and homomorphism of groups, rings and homomorphisms of rings, and many other mathematical contexts.

Homeomorphism is the concept of “essentially the same” in topology, just as *congruence* is in geometry and *isomorphism* in linear algebra and in group theory.

One of the tasks of topology is to attempt to classify all topological spaces up to homeomorphism, just as group theory attempts — amongst other things — to classify all groups up to isomorphism.

The classification of all topological spaces up to homeomorphism is, however, known to be impossible if mathematics is consistent! Nevertheless, restricted versions of this problem have been and are being studied with more or less success, where the focus is on special classes of topological spaces of interest in their own right.

One of the highlights of mathematics is the classification of all *compact surfaces*, which was achieved early in the history of topology. (We shall meet the notion of compactness in this course.) The classification of *manifolds* — higher dimensional analogues of surfaces — has been pursued ever since. One of the most famous problems in mathematics, the *Poincaré Conjecture*, is part of this classification problem. It is one of the “Clay Millenium Problems” with a prize of $US1,000,000 for its solution. It was solved in 2003 by G. Perelman, who refused the prize, and also declined to accept the Fields Medal for his achievements.

1. Exercises

9.1. Prove that each of the following functions is a homeomorphism, where each set below is taken with its Euclidean topology.
   (a) $f : \mathbb{R} \rightarrow \mathbb{R}, \ x \mapsto x^3$
   (b) $g : \mathbb{R} \rightarrow \mathbb{R}, \ x \mapsto xe^{x^2}$
   (c) $h : ]0, 2\pi[ \times ]0, \frac{\pi}{2}[ \rightarrow \mathbb{R}^2 \setminus \{(x,0) \mid x \geq 0\}, \ (u,v) \mapsto (\tan v \cos u, \tan v \sin u)$

9.2. Take topological spaces $(X,T)$ and $(Y,U)$ and a function $f : X \rightarrow Y$.

Suppose that $X = A \cup B$ and $Y = C \cup D$ where

$f|_A : A \rightarrow C$ and $f|_B : B \rightarrow D$

are both homeomorphisms.

Decide whether $f$ must be a homeomorphism.

9.3. Give an example of a continuous 1–1 and onto function between topological spaces whose inverse is not continuous.

9.4. Let $\mathbb{R}$ and $\mathbb{R}^2$ have their respective Euclidean topologies. Endow

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

with the subspace topology induced from $\mathbb{R}^2$.

Define the relation $\sim$ on $\mathbb{R}$ by

$a \sim b$ if and only if $a - b \in \mathbb{Z}$.

Prove that $\sim$ is an equivalence relation on $\mathbb{R}$.

Let $[a]$ denote the $\sim$–equivalence class containing $a$. Put

$\mathbb{R}/\mathbb{Z} := \{[a] \mid a \in \mathbb{R}\}$

and endow $\mathbb{R}/\mathbb{Z}$ with the quotient topology — that is, the topology induced by the canonical projection

$\eta : \mathbb{R} \rightarrow \mathbb{R}/\mathbb{Z}, \ a \mapsto [a]$
Prove that $\mathbb{R}/\mathbb{Z}$ is homeomorphic to $S^1 = \{(x,y) \in \mathbb{R}^2 \mid x^2 + y^2 = 1\}$.

9.5. Let $(X, T)$ be a topological space. Let $\mathbb{I} := [0, 1] := \{t \in \mathbb{R} \mid 0 \leq t \leq 1\}$ be endowed with the Euclidean topology.

Prove that for each $\lambda \in [0, 1]$ the function

$$i_\lambda : X \rightarrow X \times \mathbb{I}, \quad x \mapsto (x, \lambda)$$

is a homeomorphism of $X$ onto $\text{im}(i_\lambda)$, where $X \times \mathbb{I}$ is endowed with the product topology.

9.6. Let $f : X \rightarrow Y$ be a function between the topological spaces $(X, T)$ and $(Y, U)$.

Prove that the following are equivalent.

(i) $f$ is a homeomorphism.

(ii) $f$ is bijective, with both $f$ and $f^{-1}$ continuous.

(iii) $f$ is a bijective, continuous open mapping.

(iv) $f$ is bijective and induces a bijection $f_* : T \rightarrow U$, $A \mapsto f(A)$.

(v) $f$ is bijective and induces a bijection $f^* : U \rightarrow T$, $B \mapsto f^{-1}(B)$. 
Chapter 10

Closed Sets

The set of real numbers, \( \mathbb{R} \) and continuous real valued functions defined on subsets of \( \mathbb{R} \) have provided out motivating examples. Our programme has been to examine these and to isolate their “essential properties”. The notion of an open set arose by examining the open intervals in \( \mathbb{R} \).

Another important class of subsets of \( \mathbb{R} \) are the closed intervals \( [a, b] \), where

\[
[a, b] := \{ x \in \mathbb{R} \mid a \leq x \leq b \}
\]

To generalise closed intervals to topological spaces, we need to characterise closed intervals using only open subsets of \( \mathbb{R} \).

Observe that if \( x \notin [a, b] \), then either \( x < a \) or \( x > b \).

Put \( r := \min\{ |a - x|, |x - b| \} \).

Then \( B(x; r) \cap [a, b] = \emptyset \).

Since every element of the complement (in \( \mathbb{R} \)) of \( [a, b] \) is the centre of some open ball (in \( \mathbb{R} \)) disjoint from \( [a, b] \), the complement (in \( \mathbb{R} \)) of \( [a, b] \) is an open subset of \( \mathbb{R} \).

As this property uses open sets without invoking a metric, we use it to define closed subsets of a topological space.

**Definition 10.1.** Let \((X, \mathcal{T})\) be a topological space.

A subset \( A \) of \( X \) is a closed subset of \( X \) if and only if its complement, \( X \setminus A \), is an open subset of \( X \).

Some important properties follow immediately.

**Theorem 10.2.** Let \((X, \mathcal{T})\) be a topological space. \( \mathcal{F} \) the set of all closed subsets of \( X \).

(i) \( \emptyset, X \in \mathcal{F} \);

(ii) if \( \{ F_\lambda \mid \lambda \in \Lambda \} \subseteq \mathcal{F} \), then \( \bigcap_{\lambda \in \Lambda} F_\lambda \in \mathcal{F} \);

(iii) if \( \{ F_j \mid j = 1, \ldots, n \} \subseteq \mathcal{F} \), then \( \bigcup_{j=1}^n F_j \in \mathcal{F} \).
The essence of the proof is provided by the set-theoretic equalities

\[ X \setminus \left( \bigcap_{\lambda \in \Lambda} F_\lambda \right) = \bigcup_{\lambda \in \Lambda} (X \setminus F_\lambda) \]

\[ X \setminus \left( \bigcup_{j=1}^n F_j \right) = \bigcap_{j=1}^n (X \setminus F_j) \]

The details are left as an exercise. □

Theorem 10.2 presents the analogue for closed sets of the properties of open sets and may be used for an alternative definition of a topological space, namely, by specifying the closed, rather than the open, subsets.

The equivalence of these two is an immediate consequence of the set-theoretic equality

\[ X \setminus (X \setminus A) = A \]

for all subsets \( A \) of the set \( X \).

**Theorem 10.3 (or Alternative Definition).** Let \( F \) be a collection of subsets of \( X \) satisfying

(i) \( \emptyset, X \in F \);

(ii) if \( \{ F_\lambda \mid \lambda \in \Lambda \} \subseteq F \), then \( \bigcap_{\lambda \in \Lambda} F_\lambda \in F \);

(iii) if \( \{ F_j \mid j = 1, \ldots, n \} \subseteq F \), then \( \bigcup_{j=1}^n F_j \in F \).

Then \( \{ X \setminus F \mid F \in F \} \) is a topology on \( X \).

Moreover, the closed sets with respect to this topology are precisely the elements of \( F \).

**Proof.** Exercise. □

**Theorem 10.4 (or Alternative Definition).** Let \( (X, T) \) and \( (Y, U) \) be topological spaces. Then the function \( f : X \to Y \) is continuous if and only if \( f^{-1}(F) \) is a closed subset of \( X \) whenever \( F \) is a closed subset of \( Y \).

**Proof.** This follows immediately from the set-theoretical equality \( f^{-1}(Y \setminus F) = X \setminus f^{-1}(F) \). □

In our discussion of open sets, we saw that for every subset, \( A \), of a topological space, \( X \), has a "nearest" open set, namely, its interior, the largest subset of \( A \) which is an open subset of \( X \).

The corresponding notion for closed sets is that of the closure of a set.

**Definition 10.5.** Let \( A \) be a subset of the topological space \( (X, T) \).

The **closure** of \( A \) in \( X \), \( \overline{A} \), is the "smallest" closed subset of \( X \) containing \( A \). In other words,

(i) \( \overline{A} \) is a closed subset of \( X \);

(ii) \( A \subseteq \overline{A} \);

(iii) if \( F \) is a closed subset of \( X \) with \( A \subseteq F \), then \( \overline{A} \subseteq F \).

**Theorem 10.6.** Every subset \( A \) of the topological space \( (X, T) \) has a unique closure, \( \overline{A} \).

**Proof.** Let \( \mathcal{F} \) denote the set of all closed subsets of \( X \) that contain \( A \).

Since \( X \) is a closed subset of itself, \( \mathcal{F} \neq \emptyset \) as \( X \in \mathcal{F} \). Put

\[ \overline{A} := \bigcap_{K \in \mathcal{F}} K, \]

the intersection of all the closed subsets of \( X \) containing \( A \).
Being the intersection of closed subsets of $X$, $A$ is certainly a closed subset of $X$.

If $F$ is a closed subset of $X$, with $A \subseteq F$, then $F \in \mathcal{F}$, whence

$$A \subseteq \bigcap_{K \in \mathcal{F}} K \subseteq F.$$  

\[\square\]

**Corollary 10.7.** A is closed in the topological space $(X, \mathcal{T})$ if and only if $A = \overline{A}$.

**Proof.** Since $A \subseteq \overline{A}$, it is sufficient to show that $A \subseteq \overline{A}$ if and only if $A$ is closed.

If $A$ is closed, then $A \subseteq \overline{A}$, by definition the definition of $\overline{A}$.

Conversely, if $A \subseteq \overline{A}$, then $A = \overline{A}$ and, since $\overline{A}$ is a closed, $A$ is closed. $\square$

**Corollary 10.8.** If $A \subseteq B \subseteq \overline{A}$, then $B = \overline{A}$.

**Observation 10.9.** Notice the similarity with the interior, $\overset{\circ}{A}$, of the subset $A$ of $X$, the maximal open subset of $X$ contained in $A$. We saw that $A$ is an open subset of $X$ if and only if $A = \overset{\circ}{A}$.

**Remark 10.10.** It is important to avoid being misled by the use of the words “open” and “closed” for proper ties of subsets of a topological space. For while a door is either open or closed, but not both, this is not true of subsets of a topological space: these can be either, both or neither.

**Example 10.11.** (i) Let $(X, \mathcal{T})$ be any topological space.

Since both $\emptyset$ and $X$ are mutually complementary open subsets of $X$, they are both also closed subsets of $X$.

Thus, both $X$ and $\emptyset$ are both open and closed subsets of $X$.

(ii) Take the set $X = \{a, b\}$ ($a \neq b$) with its indiscrete topology, so that the only open subsets are $\emptyset$ and $\{a, b\}$.

Then $\{a\}$ is not open.

But neither is its complement, $\{b\}$. Hence $\{a\}$ is not closed.

A more familiar example is provided by $\mathbb{Q}$, the set of all rational numbers, when regarded as subset of $\mathbb{R}$, the set of all real numbers with its Euclidean topology.

Since every open interval of real numbers contains both rational and irrational number, $\mathbb{Q}$ is neither open nor closed as subset of $\mathbb{R}$.

(iii) Take $\mathbb{R}$ with its Euclidean topology.

The subset $\mathbb{R}^+ = \{x \in \mathbb{R} \mid x > 0\}$ is open, since if $a > 0$, then $B(a; a) \subseteq \mathbb{R}^+$.

Its complement is $\mathbb{R}^- = \{y \in \mathbb{R} \mid y \leq 0\}$, which is not open, since, for every $r > 0$, $\frac{1}{2} \in B(0; r) \setminus \mathbb{R}^-$. Hence $\mathbb{R}^+$ is not closed.

This argument also shows that $\mathbb{R}^-$ is a closed subset of $\mathbb{R}$, but not an open one.

A useful and important fact is that continuity can be defined in terms of forming closures of sets.

**Theorem 10.12.** Let $(X, \mathcal{T})$ and $(Y, \mathcal{U})$ be topological spaces.

The function $f : X \rightarrow Y$ is continuous if and only if for every $A \subseteq X$

$$f(\overline{A}) \subseteq \overline{f(A)}.$$  

**Proof.** Exercise. $\square$

**Observation 10.13.** Theorem 10.12 provides the characterisation of the continuity of a function $f : X \rightarrow Y$ “going forward” promised in Example 6.21.
If the set $B$ is the closure of the set $A$, then, heuristically, there is “no space between $A$ and the rest of $B$”. We provide a precise formulation of this.

**Definition 10.14.** Take $B \subseteq X$.
The subset $A$ of the topological space $(X, T)$ is dense in $B$ if and only if $\overline{A} = B$.

When $\overline{A} = X$, we simply say that $A$ is dense.

The next theorem provides necessary and sufficient conditions for a point in a topological space to be in the closure of a given subset of that space.

**Theorem 10.15.** Let $A$ be a subset of the topological space $(X, T)$.

$x \in \overline{A}$ if and only if every open subset of $X$ containing $x$ meets $A$, that is to say, $G \cap A \neq \emptyset$ whenever $x \in G \in T$.

**Proof.** Exercise. \hfill $\Box$

Theorem 10.15 allows us to construct the closure of a set by “enlarging” that set.

Let $A$ be a subset of the topological space $(X, T)$ and take $x \in A$.

If $x \in G \in T$, then $G \cap A \neq \emptyset$.

This leaves two possibilities:

(i) $x \in G \cap A$, or
(ii) $G$ contains an element of $A$ which is not $x$, that is $(G \cap A) \setminus \{x\} \neq \emptyset$.

**Definition 10.16.** The element $x$ of the topological space $(X, T)$ is a point of accumulation of the subset $A$ of $X$ if and only if given any open subset $G$ containing $x$, $G \cap A$ contains at least one element other than $x$, that is, $(G \cap A) \setminus \{x\} \neq \emptyset$.

The derived set of $A$ is the set of all points of accumulation of $A$ and is denoted by $D(A)$ or $A'$.

**Example 10.17.** Take $\mathbb{R}$ with its Euclidean topology.

Put $A := \{\frac{1}{n+1} \mid n \in \mathbb{N}\}$.

Then $A' = \{0\}$.

To see this, let $G$ be an open subset of $\mathbb{R}$ with $0 \in G$.

Then, by the definition of the Euclidean topology, there is an $r > 0$ with $B(0; r) \subseteq G$.

Choose $n \in \mathbb{N}$ with $n > r$. Then, plainly

$$\frac{1}{2n} \in (G \cap A) \setminus \{0\}$$

showing that 0 is a point of accumulation of $A$.

Take $x \in \mathbb{R} \setminus \{0\}$.

If $x < 0$ or $x > 1$, put $r := \min\{|x|, |x - 1|\}$.

Plainly, $B(x; r) \cap A = \emptyset$.

If $x \in A$, say $x = \frac{1}{n+1}$, put

$$G = \begin{cases} \left[\frac{1}{n+2}, \frac{1}{n}\right] & \text{if } n \neq 0 \\ \left[\frac{1}{2}, \frac{3}{2}\right] & \text{if } n = 0 \end{cases}$$

Plainly, $G$ is open and $G \cap A = \{x\}$.

If $0 < x \leq 1$ and $x \notin A$, then, by the Archimedean property of $\mathbb{R}$, there is a unique $n \in \mathbb{N}$ with $n + 1 < \frac{1}{x} < n + 2$.
or, equivalently,
\[
\frac{1}{n+2} < x < \frac{1}{n+1}
\]
Then \( G = \left( \frac{1}{n+2}, \frac{1}{n+1} \right] \) is an open subset of \( \mathbb{R} \) and \( G \cap A = \emptyset \).
Hence, no non-zero real number is a point of accumulation of \( A \).

**Theorem 10.18.** For the subset, \( A \), of the topological space \((X, T)\),
\[
\overline{A} = A \cup A'
\]

**Proof.** Take \( x \in (\overline{A} \setminus A) \)
Take \( G \in T \) with \( x \in G \).
\[
(G \cap A) \setminus \{x\} = G \cap A \quad \text{as } x \not\in A \\
\neq \emptyset \quad \text{as } x \in \overline{A}
\]
Thus, \( x \) is a point of accumulation of \( A \).

There are two equivalent definitions of a continuous function, one in terms of (the inverse images of) open sets and the other in terms of (the inverse images of) closed sets. That these are equivalent is an immediate consequence of the fact that, for any function \( f : X \to Y \),
\[
f^{-1}(Y \setminus B) = X \setminus f^{-1}(B)
\]
By contrast, it is not true in general that the image of every open subset of \( X \) is an open subset of \( Y \) if and only if the image of every closed subset of \( X \) is a closed subset of \( Y \).
Before offering examples, we formulate the latter property in analogy with the notion of an open map (cf. Definition 6.19).

**Definition 10.19.** Let \( X \) and \( Y \) be topological spaces.

The function \( f : X \to Y \) is a closed map if and only if \( f(K) \) is a closed subset of \( Y \) whenever \( K \) is a closed subset of \( X \).

We show that a function can be an open mapping without being a closed mapping, and vice versa.

**Example 10.20.** Take \( X := \left( 0, 1 \right] \) and \( Y := \mathbb{R} \), both with the Euclidean topology and consider
\[
f : X \to Y, \quad t \mapsto t
\]
It is easy to see that the image of every open subset of \( X \) is an open subset of \( Y \), showing that \( f \) is an open mapping.
But while \( \left( 0, 1 \right] \) is a closed subset of \( X \), its image \( f \left( \left( 0, 1 \right] \right) = \left( 0, 1 \right] \) is not a closed subset of \( Y = \mathbb{R} \), since no (Euclidean) open ball centred on 1 is contained entirely in its complement, \( \left( -\infty, 0 \right] \cup \left[ 1, \infty \right) \), which is, therefore, not open in \( \mathbb{R} \).
Thus, \( f \) is not a closed mapping.

**Example 10.21.** Take \( X := [0, 1] \) and \( Y := \mathbb{R} \), both with the Euclidean topology and consider
\[
f : X \to Y, \quad t \mapsto t
\]
Clearly, the image of every closed subset of \( X = [0, 1] \) is a closed subset of \( Y = \mathbb{R} \), showing that \( f \) is a closed mapping.
But while \([0, 1]\) is an open subset of \( X \), its image \( f ([0, 1]) = [0, 1] \) is not an open subset of \( Y = \mathbb{R} \), since no (Euclidean) open ball centred on 1 can be a subset of \([0, 1]\).
Thus, \( f \) is not an open mapping.
Example 10.22. Take $X := ]0, 1]$ and $Y := \mathbb{R}$, both with the Euclidean topology and consider $f : X \to Y, \ t \mapsto t$

Then $]0, 1]$ is both an open and a closed subset of $X$.

Its image, $f([0, 1]) = [0, 1]$ is neither an open, nor a closed subset of $Y = \mathbb{R}$.

We have shown that a function can be an open mapping without being a closed mapping and vice versa, and that, in fact, it may be both or it may be neither.

However, the situation is different when the function in question is bijective.

Lemma 10.23. Let $f : X \to Y$ be a bijection from the topological space $(X, T)$ to the topological space $(Y, U)$.

Then $f$ is an open mapping if and only if it is a closed mapping.

Proof. If $f$ is bijective, then it has an inverse function, $g = f^{-1} : Y \to X$, so that $f = g^{-1}$.

As $f(A) = g^{-1}(A)$ for all $A \subseteq X$, it follows from Theorem 10.3 that $f$ is an open mapping if and only if $g$ is continuous if and only if $f$ is a closed mapping. \qed

1. Exercises

10.1. Let $(X, \mathcal{T})$ be a topological space.

Prove that

(i) $\emptyset$ and $X$ are closed subsets of $X$;

(ii) if $\{F_\lambda \mid \lambda \in \Lambda\}$ is a family of closed subsets of $X$, then $\bigcap_{\lambda \in \Lambda} F_\lambda$ is a closed subset of $X$;

(iii) if $\{F_j \mid j = 1, \ldots, n\}$ is a finite family of closed subsets of $X$, then $\bigcup_{j=1}^n F_j$ is also a closed subset of $X$.

10.2. Let $X$ be a set and let $\mathcal{F}$ be a collection of subsets of $X$ such that

(i) $\emptyset, X \in \mathcal{F}$;

(ii) if $\{F_\lambda \mid \lambda \in \Lambda\} \subseteq \mathcal{F}$, then $\bigcap_{\lambda \in \Lambda} F_\lambda \in \mathcal{F}$;

(iii) if $\{F_j \mid j = 1, \ldots, n\} \subseteq \mathcal{F}$, then $\bigcup_{j=1}^n F_j \in \mathcal{F}$.

Prove that $\{X \setminus F \mid F \in \mathcal{F}\}$ is a topology on $X$, whose closed sets are precisely the elements of $\mathcal{F}$.

10.3. Let $(X, \mathcal{T})$ be a topological space and take $A \subseteq X$.

Let $\overline{A}$ be the closure of $A$ in $X$.

Prove that $x \in \overline{A}$ if and only if every open subset of $X$ containing $x$ meets $A$.

10.4. Take $\mathbb{R}$ with its Euclidean topology. Find the closure of each of the following subsets.

(i) $\left\{ \frac{1}{n+1} \mid n \in \mathbb{N} \right\}$

(ii) $\mathbb{N}$

(iii) $\mathbb{Q}$

10.5. Find an example of a set which is open, not closed, but can be written as the union of closed sets.
10.6. Find an example of a set which is closed, not open, but can be written as the intersection of open sets.

10.7. Let $\mathbb{R}[t_1, \ldots, t_n]$ denote the set of all real polynomials in the indeterminates $t_1, \ldots, t_n$.
Given a subset $S$ of $\mathbb{R}[t_1, \ldots, t_n]$ define

$$V(S) := \{(x_1, \ldots, x_n) \in \mathbb{R}^n \mid f(x_1, \ldots, x_n) = 0 \text{ for every } f(t_1, \ldots, t_n) \in S\}$$

and

$$\mathcal{F} := \{V(S) \mid S \subseteq \mathbb{R}[t_1, \ldots, t_n]\}.$$

Prove that there is a topology on $\mathbb{R}^n$ whose closed sets are precisely the elements of $\mathcal{F}$ and that this topology is not metrisable.

(This topology is called the Zariski topology on $\mathbb{R}^n$ and is used in algebraic geometry.)

10.8. Let $(X, \mathcal{T})$ and $(Y, \mathcal{U})$ be topological spaces.

Prove that the function $f: X \rightarrow Y$ is continuous if and only if given any $A \subseteq X$, $f(A) \subseteq \overline{f(A)}$.

10.9. Let $A$ be a closed subset of the topological space $(X, \mathcal{T})$.

Prove that $F \subseteq A$ is closed in the subspace topology on $A$ if and only if $F$ is a closed subset of $X$. 

Intuitively, a space is connected if it does not “fall apart into two (or more) distinct and separate pieces”. This suggests a preliminary definition: the topological space $X$ is connected if and only if it cannot be written as $A \cup B$ with $A \cap B = \emptyset$.

However, such a definition is of little use, since given any $A \subseteq X$ we always have $X = A \cup (X \setminus A)$ and, of course, $A \cap (X \setminus A) = \emptyset$. In fact, $X = A \cup B$ with $A \cap B = \emptyset$ if and only if $B = X \setminus A$.

In particular, $X = \emptyset \cup X$.

Some additional condition is required to ensure that such trivial decompositions are excluded. Our choice, invited by the dictum “topology is about open sets”, is to require that $A$ and $B$ be non-empty open subsets of $X$.

**Definition 11.1.** A disconnection of the topological space $(X, T)$ consists of two non-empty, disjoint, proper open subsets of $X$ whose union is $X$, that is $A, B \in T$ with

(i) $\emptyset \subset A \subset X$ and $\emptyset \subset B \subset X$

(ii) $A \cap B = \emptyset$

(iii) $A \cup B = X$

$X$ is connected if and only if there is no disconnection of $X$.

The subset $W$ of $X$ is connected if and only if it is connected with respect to its subspace topology.

The next lemma gathers together convenient alternative formulations of connectedness.

**Lemma 11.2.** Given a topological space $(X, T)$ the following are equivalent.

(a) $(X, T)$ is connected.

(b) $\emptyset$ and $X$ are the only two subsets of $X$ which are both open and closed.

(c) Given non-empty open subsets $A, B$ of $X$ with $A \cup B = X$, $A \cap B \neq \emptyset$.

(d) Given disjoint open subsets $A, B$ of $X$ with $A \cup B = X$, either $A = \emptyset$ (and $B = X$) or $A = X$ (and $B = \emptyset$).

(e) Given non-empty closed subsets $A, B$ of $X$ with $A \cup B = X$, $A \cap B \neq \emptyset$.

(f) Given disjoint closed subsets $A, B$ of $X$ with $A \cup B = X$, either $A = \emptyset$ (and $B = X$) or $B = \emptyset$ (and $A = X$).

**Proof.** Since a subset of $X$ is connected if and only if its complement is open, the theorem merely restates the definition. \qed
Corollary 11.3. The subset \( W \) of the topological space \((X,\mathcal{T})\) is connected if and only if given open subsets \( G \) and \( H \) of \( X \) with \( W \subseteq G \cup H \) and \( W \cap G \cap H = \emptyset \), either \( W \cap G = \emptyset \) or \( W \cap H = \emptyset \).

We can also define connectedness in terms of continuous functions into a test space. Our test space for connectedness will be the topological space \( \mathbb{2} \).

Definition 11.4. The topological space \( \mathbb{2} \) is the set \( \{0,1\} \) with its discrete topology.

Observation 11.5. Any other set with precisely two elements and the discrete topology would do, as all such spaces are clearly homeomorphic. Particularly convenient is the 0-sphere \( S^0 := \{-1,1\} = \{x \in \mathbb{R} \mid x^2 = 1\} \).

Theorem 11.6. The topological space \((X,\mathcal{T})\) is connected if and only if every continuous function \( f : X \to \mathbb{2} \) is constant.

Proof. Let \( f : X \to \mathbb{2} \) be a continuous function.

Since \( \mathbb{2} = \{0,1\} = \{0\} \cup \{1\} \), and \( f \) is a function
\[
X = f^{-1}([0]) \cup f^{-1}([1])
\]
and
\[
f^{-1}([0]) \cap f^{-1}([1]) = \emptyset
\]
Since \( f \) is continuous, both \( f^{-1}([0]) \) and \( f^{-1}([1]) \) are open subsets of \( X \).

Thus, by Lemma 11.2, \( X \) is connected only if one of \( f^{-1}([0]) \) and \( f^{-1}([1]) \) is the empty set, that is to say \( X \) is connected only if \( f \) is a constant function.

Conversely, suppose that \( X = A \cup B \) with \( A, B \) disjoint non-empty open subsets of \( X \).

Define
\[
f : X \to \mathbb{2}, \quad x \mapsto \begin{cases} 0 & \text{if } x \in A \\ 1 & \text{if } x \in B. \end{cases}
\]

Since \( A \cap B = \emptyset \) and \( A \cup B = X \), \( f \) is a function.

By definition, it is not constant.

It remains to check that it is continuous.

The open subsets of \( \mathbb{2} \) are \( \emptyset, \{0\}, \{1\}, \) and \( 2 \), whose inverse images under \( f \) are \( \emptyset, A, B \) and \( X \) respectively.

By hypothesis, each of these is an open subset of \( X \), showing that \( f \) is continuous. \( \square \)

Corollary 11.7. The continuous image of a connected set is connected.

Proof. Let \( f : X \to Y \) be a continuous surjection.

Take a continuous function \( \varphi : Y \to \mathbb{2} \).

Then \( \varphi \circ f \) is continuous, and so, by Theorem 11.6, it must be constant.

Since \( f \) is surjective, \( \text{im}(\varphi) = \text{im}(\varphi \circ f) \), whence \( \varphi \) must be constant. \( \square \)

Our motivating examples being subsets of \( \mathbb{R}^n \), we investigate them for connectedness, beginning, as usual, with \( \mathbb{R} \) itself.

Definition 11.8. \( I \subseteq \mathbb{R} \) an interval if and only if \( [a,b] \subseteq I \) for \( a, b \in I \) with \( a \leq b \).

Theorem 11.9. \( I \subseteq \mathbb{R} \) is connected if and only if it is an interval.
PROOF. If $I$ has at most one element, then there is nothing to prove. Otherwise, take $a, b \in I$ with $a < b$ and $c$ with $a < c < b$.

If $c \not\in I$, put $G := ]-\infty, c[ \cap I$ and $H := ]c, \infty[ \cap I$, so that $a \in G$ and $b \in H$. Clearly, $G$ and $H$ are non-empty open subsets of $I$ and $I = G \cup H$. Thus $I$ is not connected, whence every connected subset of $\mathbb{R}$ is an interval.

For the converse, let $I \subseteq \mathbb{R}$ be an interval.

Suppose that $I \subseteq A \cup B$, with $A$ and $B$ open subsets of $\mathbb{R}$ and $A \cap I, B \cap I \neq \emptyset$.

Take $a \in A \cap I, b \in B \cap I$ and assume that $a < b$.

Since $I$ is an interval, $[a, b] \subseteq I$.

Put $c := \sup(A \cap [a, b]) = \sup\{x \in A \mid a \leq x \leq b\}$, and take $\varepsilon > 0$.

By the defining properties of a supremum, there is an $x \in A \cap [a, b]$ with $c - \varepsilon < x \leq c$.

Hence every open subset of $\mathbb{R}$ containing $c$ must meet $A \cap [a, b] \subseteq A \cap I$.

Again by the defining properties of a supremum, there is a $y \in B \cap [a, b] \subseteq B \cap I$ with $c \leq y \leq c + \varepsilon$:

- If $c = b$ then take $y := c$;
- otherwise there is a suitable $y$ between $c$ and $b$.

Hence, every open ball with centre $c$ meets $B$.

Thus every open subset of $\mathbb{R}$ containing $c$ must meet both $A \cap I$ and $B \cap I$.

Since $c \in (A \cup B) \cap I$, either $c \in A \cap I \subseteq A$ or $c \in B \cap I \subseteq B$ and, by hypothesis, both $A$ and $B$ are open.

Thus $A \cap B \cap I \neq \emptyset$, showing that $I$ is connected.

\[ \square \]

COROLLARY 11.10. $\mathbb{R}$ is connected.

PROOF. $\mathbb{R}$ is an interval.

\[ \square \]

COROLLARY 11.11 (The Intermediate Value Theorem of Calculus). Let $I \subseteq \mathbb{R}$ be an interval and $f : I \rightarrow \mathbb{R}$ a continuous function. Take $a, b \in I$ with $a \leq b$ and suppose that $f(a) \leq f(b)$. Then for each $y \in [f(a), f(b)]$ there is at least one $x \in [a, b]$ with $f(x) = y$.

PROOF. By Theorem 11.9, $I \subseteq \mathbb{R}$ is connected since it is an interval.

By Corollary 11.7, $\text{im}(f) = f(I)$ is connected, as $f$ is continuous.

By Theorem 11.9 $f(I)$ is an interval, being a connected subset of $\mathbb{R}$.

\[ \square \]

REMARK 11.12. A subset of a connected set need not be connected.

EXAMPLE 11.13. $\{0, 1\}$ is a subset of $\mathbb{R}$ which not connected, because it is not an interval.

REMARK 11.14. The intersection of two connected sets need not be connected.

EXAMPLE 11.15. Consider the continuous function

\[ \varphi : \mathbb{R} \rightarrow \mathbb{R}^2, \quad t \mapsto (\cos(2\pi t), \sin(2\pi t)). \]

The image of the interval $[0, 1]$ is the upper semi-circle $\{(x, y) \in \mathbb{R}^2 \mid x^2 + y^2 = 1, y \geq 0\}$ and the image of $[1, 2]$ is the lower semi-circle $\{(x, y) \in \mathbb{R}^2 \mid x^2 + y^2 = 1, y \leq 0\}$. Thus, both semi-circles are connected.

Their intersection, $\{(-1, 0), (1, 0)\}$, is homeomorphic with $2$, and hence not connected.

THEOREM 11.16. Let $\{A_\lambda \mid \lambda \in \Lambda\}$ be a family of connected subsets of the topological space $(X, \mathcal{T})$.

If $\bigcap_{\lambda \in \Lambda} A_\lambda \neq \emptyset$, then $A := \bigcup_{\lambda \in \Lambda} A_\lambda$ is also connected.
Proof. Take \( a \in A = \bigcap_{\lambda \in \Lambda} A_\lambda \) and a continuous function \( f : A \to 2 \). Choose \( x \in A \). Then \( x \in A_\mu \) for some \( \mu \in \Lambda \).
But \( a \in A_\mu \) as well.
As \( f \mid_{A_\mu} := f \circ i_\mu^{A_\mu} : A_\mu \to 2 \) is continuous and \( A_\mu \) is connected, it follows from Theorem 11.6 that
\[
f(x) = f \mid_{A_\mu}(x) = f \mid_{A_\mu}(a) = f(a),
\]
Hence \( f \) cannot be surjective, showing that \( A \) is connected. \( \square \)

Corollary 11.17. Every element \( x \) of the topological space \((X, T)\) is contained in a unique maximal connected subset.

Proof. Let \( \mathfrak{A} \) be the collection of all connected subsets of \( X \) which contain \( x \).
Then \( \mathfrak{A} \) is not empty, since \( \{x\} \in \mathfrak{A} \).
Since \( x \in \bigcap_{A \in \mathfrak{A}} A \), it follows from Theorem 11.16 that
\[
C[x] := \bigcup_{A \in \mathfrak{A}} A
\]
is a connected subset of \( X \) and contains \( x \).
It is clearly maximal with this property.
The uniqueness of \( C[x] \) is immediate. \( \square \)

Definition 11.18. The maximal connected subsets of the topological space \((X, T)\) are the connected components of \( X \). \( C[x] \) is the (connected) component of \( X \) containing \( x \).
The subset \( A \) of the topological space \((X, T)\) is totally disconnected if and only if each connected component of \( A \) is a singleton.

Example 11.19. \( \mathbb{Q} \) is a totally disconnected subset of \( \mathbb{R} \).
To see this, take \( x < y \in \mathbb{Q} \) and put
\[
c := x + \frac{1}{\sqrt{2}}(y - x) = \frac{(\sqrt{2} - 1)x + \sqrt{2} y}{\sqrt{2}}
\]
Plainly, \( x < c < y \) and \( c \notin \mathbb{Q} \) and \( A := \mathbb{Q} \cap ]-\infty, c[ \), \( B := \mathbb{Q} \cap ]c, \infty[ \) is a disconnection of \( \mathbb{Q} \), with \( x \in A \) and \( y \in B \).
Hence \( y \notin C[x] \).

Theorem 11.20. Let \( A \) be a connected subset of the topological space \((X, T)\).
If \( A \subseteq G \subseteq \bar{A} \), then \( G \) is also connected.

Proof. Exercise. \( \square \)

Corollary 11.21. Every component of the topological space \((X, T)\) is closed.

Proof. The statement follows from Theorem 11.20, using the maximality of the components. \( \square \)

We next investigate the relationship between the connectedness of a set of topological spaces and the connectedness of their product. The relationship is quite direct.

Theorem 11.22. The product of a set of topological spaces is connected in the product topology if and only if each factor is connected.
11. CONNECTEDNESS

Proof. Suppose that the product of a set of topological spaces is connected.

By the definition of the topological product, the projections onto the individual factors are continuous surjections. Hence, by Corollary 11.7, each factor is connected.

We prove the converse for a finite set of topological spaces, leaving the general result as an exercise.

Take topological spaces \((X_j, T_j) (j = 1, \ldots, n)\), and let \((X, T)\) be the Cartesian product of the sets \(X_j\) with the product topology.

Since there is nothing to prove if at least one of the \(X_j\) is empty, we assume each is non-empty.

Suppose that each \(X_j\) is connected.

Take \(a := (a_1, \ldots, a_n) \in X\), and let \(\varphi: X \rightarrow 2\) be continuous.

Take \(x = (x_1, \ldots, x_n) \in X\).

We show that \(\varphi(x) = \varphi(a)\) by constructing a sequence of elements of \(X\), \((b_k)_{k=1}^{n+1}\), with \(b_1 = a\), \(b_{n+1} = x\) and \(\varphi(b_{k+1}) = \varphi(b_k)\) for \(k = 1, \ldots, n\).

Put \(b_k = (y_{1k}, \ldots, y_{nk})\), where

\[
y_{jk} = \begin{cases} x_j & \text{for } j < k \\ a_j & \text{for } j \geq k. \end{cases}
\]

Then \(b_1 = a\) and \(b_{n+1} = x\).

For \(k = 1, \ldots, n\), put

\[A_k := \{y_{1k}\} \times \cdots \{y_{(k-1)k}\} \times X_k \times \{y_{(k+1)k}\} \times \cdots \{y_{nk}\}.\]

Then, for \(k = 1, \ldots, n, k, A_k \subseteq X\) and \(A_k\) is connected, being homeomorphic with \(X_k\).

Moreover, since \(y_{j(k+1)} = y_{jk}\) for \(j \neq k\), \(b_k, b_{k+1} \in A_k\).

Since \(\varphi\) is continuous on \(A_k\), \(\varphi(b_{k+1}) = \varphi(b_k)\) for \(k = 1, \ldots, n\).

Thus \(\varphi(x) = \varphi(a)\), showing that \(\varphi\) is constant.

\[\square\]

Corollary 11.23. \(\mathbb{R}^n\) is connected whenever \(n \geq 1\).

Proof. \(\mathbb{R}^n = \prod_{j=1}^{n} \mathbb{R}\).

\[\square\]

A stronger form of connectedness is path-connectedness, which is closer to what “the man in the street” might regard as connectedness, for, heuristically, a space is path-connected if and only if “it can be drawn without lifting the pen from the paper”.

Definition 11.24. Let \(X\) be a topological space and take \(a, b \in X\).

A path in \(X\) from \(a\) to \(b\) is a continuous function \(\gamma: [0, 1] \rightarrow X\) with \(\gamma(0) = a\) and \(\gamma(1) = b\).

The topological space \((X, T)\) is path-connected if and only if for all \(a, b \in X\), there is a path, \(\gamma\), in \(X\) from \(a\) to \(b\).

Theorem 11.25. The continuous image of a path-connected space is path-connected.

Proof. Exercise.

\[\square\]

Theorem 11.26. Every path-connected space is connected.

Proof. Let \((X, T)\) be path-connected.

If \(X = \emptyset\), then there is nothing to prove.

Otherwise, choose \(a \in X\).

For each \(x \in X\), choose a path \(\gamma_x: [0, 1] \rightarrow X\) from \(a\) to \(x\).
Since $[0, 1]$ is connected and $\gamma_X$ is continuous, $G_x := \text{im}(\gamma_X) \subseteq X$ is connected.

Plainly, $X = \bigcup_{x \in X} G_x$ is the union of connected subsets.

Since $a \in G_x$ for every $x \in X$, Theorem 11.16 shows that $X$ must be connected. □

The converse of Theorem 11.26 is not true.

**Example 11.27.** Consider the graph $Gr(f)$ of the continuous function

$$f: \] 0, \pi \] \rightarrow \mathbb{R}, \quad t \mapsto \sin \left(\frac{1}{t}\right),$$

so that $Gr(f) = \{(t, \sin (\frac{1}{t})) \mid 0 < t \leq \pi\}$. Then the subspace of $\mathbb{R}^2$

$$X := Gr(f) \cup \{(0, r) \mid 0 \leq r \leq 1\}$$

is connected but not path-connected.

This is often referred to as the topologists’ sine curve. The reader should sketch it.

**Observation 11.28.** It is clear from our discussion, that both connectedness and path-connectedness are (non-numerical) “topological invariants” — if two spaces are homeomorphic, then they are either both (path-)connected, or neither is.

In this way, we can sometimes distinguish between non-homeomorphic spaces, by determining that one is (path-)connected, but the other one not.

**Remark 11.29.** There is, in fact, an entire sequence of connectedness properties, with path-connected spaces referred to as 0-connected spaces.

This arises from the fact that we can reformulate path-connectedness as being able to extend continuous functions.

To see this, notice that the choice of two elements of $X$, say $a, b$, is equivalent to specifying the function

$$f: \{-1, 1\} \rightarrow X, \quad x \mapsto \begin{cases} a & \text{for } x = -1 \\ b & \text{for } x = 1 \end{cases}$$

and we may replace the path

$$\gamma: [0, 1] \rightarrow X$$

with $\gamma(0) = a$ and $\gamma(1) = b$, by the function

$$\varphi: [-1, 1] \rightarrow X, \quad x \mapsto \gamma \left(\frac{x + 1}{2}\right)$$

for which $\varphi(-1) = a$ and $\varphi(1) = b$.

Recall that the $n$-sphere, $S^n$, and the $(n + 1)$-disc, $D^{n+1}$ are the subsets of $\mathbb{R}^{n+1}$ defined by

$$S^n = \{(x_0, \ldots, x_n) \in \mathbb{R}^{n+1} \mid \sum_{j=0}^{n} x_j^2 = 1\}$$

$$D^{n+1} = \{(x_0, \ldots, x_n) \in \mathbb{R}^{n+1} \mid \sum_{j=0}^{n} x_j^2 \leq 1\}$$

Plainly, $S^n \subset D^{n+1}$. In fact, $S^n$ is what we call in ordinary geometry, the boundary of $D^{n+1}$.

Thus, $\{-1, 1\}$ is $S^0$ and $[-1, 1]$ is $D^1$, and the notion of path-connectedness can be expressed by saying that the topological space, $(X, T)$ is path-connected if and only if every continuous function

$$f: S^0 \rightarrow X$$

can be extended to a continuous function $\hat{f}: D^1 \rightarrow X$.

Similarly, we can ask whether a continuous function $S^n \rightarrow X$ can be extended to a continuous function $D^{n+1} \rightarrow X$. 
This leads to the notion of $n$-connectedness, where the topological space, $(X, T)$ is $n$-connected if and only if every continuous function $S^k \rightarrow X$ can be extended to a continuous function $D^{k+1} \rightarrow X$ whenever $0 \leq k \leq n$.

Plainly, every $n$-connected space is $m$-connected whenever $m \leq n$. But the converse is not true. It is also clear that $n$-connectedness is a topological invariant, with the “higher” connectedness property able to distinguish non-homeomorphic spaces, which “lower” levels of connectedness fail to distinguish.

For this reason, the notion of $n$-connectedness is important in advanced topology, especially algebraic and differential geometry, but also in modern mathematical physics.

1. Exercises


11.2. Prove that the continuous image of a connected space is connected.

11.3 (Topologists’ Sine Curve). Show that

$$\{ (x, \sin(\pi x)) \mid 0 < x \leq 1 \} \cup \{ (0, y) \mid -1 \leq y \leq 1 \}$$

is a connected, but not path-connected, sub-space of $\mathbb{R}^2$.

11.4. Show that $\mathbb{Q}$ is a totally disconnected subset of $\mathbb{R}$.

11.5. Suppose that $A$ is a connected subset of the topological space $X$. Prove that if $A \subseteq G \subseteq \overline{A}$, then $G$ is also connected.

11.6. Prove that the product of a set of topological spaces is connected if and only if each factor is connected.

11.7. Give an alternative proof of Theorem 11.26, by showing that if $(X, T)$ is path-connected, then every continuous function $f: X \rightarrow 2$ must be constant.

11.8. Prove that the continuous image of a path-connected space is path-connected.

11.9. Prove that the product of a set of topological spaces is path-connected if and only if each factor is path-connected.
Neglect of mathematics work injury to all knowledge, since he who is ignorant of it cannot know the other sciences or things of this world. And what is worst, those who are thus ignorant are unable to perceive their own ignorance, and so do not seek a remedy.

Roger Bacon

Chapter 12

Separation

We turn to another (non-numerical) topological invariant, namely separation. This is concerned with how well we can distinguish points in a topological space, using only open subsets.

We have seen that every set — even the empty set — admits at least one topology, the discrete topology, and that if a set has at least two elements, then it admits at least one other topology, the indiscrete one. Furthermore, any other topology on the given set is finer than the indiscrete one but coarser than the discrete one.

We investigate the relative coarseness of other possible topologies through separation properties.

One way of thinking about open sets is that they are a qualitative, rather than quantitative, measure of nearness: a measure of nearness without actually assigning a numerical distance between points in a space. Indeed, open sets are sometimes referred to as neighbourhoods.

It is in such a sense that topologies can be said to separate (or not to separate) points:

Given the topological space \((X, T)\) and two distinct elements \(x, y\) of \(X\), is there an open subset \(G\) of \(X\) which contains one but not the other?

If so, then we can distinguish between the elements \(x\) and \(y\) using only the open subsets of the space, that is, the topology of the space separates the points of the space.

There are differing degrees to which a topology on \(X\) can separate the elements of \(X\) and some of these have been classified systematically.

Definition 12.1. The topological space \((X, T)\) is

- \(T_0\) if and only if given any \(x \neq y \in X\) there is either a \(G \in T\) with \(x \in G\) and \(y \notin G\), or a \(G \in T\) with \(y \in G\) and \(x \notin G\);
- \(T_1\) if and only if given any \(x \neq y \in X\) there are \(G, H \in T\) with \(x \in G\) and \(y \notin G\), and \(y \in H\) and \(x \notin H\);
- \(T_2\) if and only if given any \(x \neq y \in X\) there are \(G, H \in T\) with \(x \in G\), \(y \in H\) and \(G \cap H = \emptyset\);
- \(T_3\) if and only if given any closed subset \(F\) of \(X\) and \(x \notin F\) there are \(G, H \in T\) with \(x \in G\), \(F \subseteq H\) and \(G \cap H = \emptyset\);
- \(T_4\) if and only if given any disjoint closed subsets \(F\) and \(K\) of \(X\) there are \(G, H \in T\) with \(F \subseteq G\), \(K \subseteq H\) and \(G \cap H = \emptyset\).

\(T_1\) spaces are sometimes also called Fréchet spaces.
\(T_2\) spaces are frequently also called Hausdorff spaces.

Lemma 12.2. Every \(T_2\) space is a \(T_1\) space and every \(T_1\) space is a \(T_0\) space.
90  12. SEPARATION

Proof. The statements follow directly from the definitions. □

There are no other general relationships between the separation axioms.

Example 12.3. Take a set, $X$, with at least two elements, with its indiscrete topology, $I$.
As the only non-empty open subset of $X$ is $X$ itself, there is no open subset of $X$ which contains
one element of $X$ but not another specified one.
Hence, $(X, I)$ is not a $T_0$ space.
Take $a \in X$. The only closed subset of $X$ not containing $a$ is the empty set, $\emptyset$, which is open.
Since $X$ is also an open subspace of $X$, $x \in X$, $\emptyset \subseteq \emptyset$ and $X \cap \emptyset = \emptyset$,
Hence, $(X, I)$ is a $T_3$ space.
The only distinct subsets of $X$ are the empty set $\emptyset$ and $X$ itself, since both are also open subsets
of $X$, and $\emptyset \subseteq \emptyset, X \subseteq X$ and $\emptyset \cap X = \emptyset$.
Hence, $(X, I)$ is even a $T_4$ space.
This shows that neither a $T_3$ nor a $T_4$ space need be a $T_2$ space.

Example 12.4. For $a \neq b$, endow $X = \{a, b\}$ with the topology $T := \{\emptyset, \{a\}, X\}$.
The only choice of distinct elements of $X$ is $a$ and $b$.
As $G := \{a\}$ is an open subset of $X$ with $a \in G$ and $b \notin G$, $(X, T)$ is a $T_0$ space.
But as the only open subset of $X$ that contains $b$ also contains $a$, $(X, T)$ is not a $T_1$ space.

Example 12.5. Take $X = \mathbb{R}^2$ and let $T$ be the topology generated by
$$\{\mathbb{R}^2 \setminus \{(x, y)\} \mid (x, y) \in \mathbb{R}^2\}$$
so that the non-empty open subsets of $X$ are precisely the subsets whose complement is finite.
Take $(a, b), (c, d) \in \mathbb{R}^2$ with $(a, b) \neq (c, d)$.
Putting $G = \mathbb{R}^2 \setminus \{(c, d)\}$ and $H = \mathbb{R}^2 \setminus \{(a, b)\}$ we see that both $G$ and $H$ are open and
$$\begin{align*}
(a, b) &\in G \quad \text{and} \quad (c, d) \notin G \\
(a, b) &\notin H \quad \text{and} \quad (c, d) \in H
\end{align*}$$
Thus, $(X, T)$ is a $T_1$ space.

However, any two non-empty open subsets of $(X, T)$ have non-empty intersection.
Thus, $(X, T)$ is not a $T_2$ space.

If, in addition to being $T_3$ or $T_4$ space, we require that the space satisfy the $T_1$ axiom, then the
situation is quite different, as we shall show. We first introduce out terminology for such spaces
and investigate the separation properties.

Definition 12.6. The topological space $(X, T)$ is regular if and only if it is both a $T_1$ space and
a $T_3$ space.

It is normal if and only if it is both a $T_1$ space and a $T_4$ space.\textsuperscript{1}

We next formulate useful criteria for determining whether a space is a $T_1$.

Theorem 12.7. The topological space $(X, T)$ is a $T_1$ space if and only if every singleton subset of
$X$ is closed.

Proof. Suppose that every singleton subset is closed.
Take $x, y \in X, x \neq y$. Put $G := X \setminus \{y\}$ and $H := X \setminus \{x\}$.

\textsuperscript{1}Some authors distinguish between “$T_3$” and “regular” in the opposite sense to us, and do the same with “$T_4$”
and “normal”. The other terms are used uniformly.
Since \( \{x\} \) and \( \{y\} \) are closed subsets of \( X \), \( G \) and \( H \) are both open subsets of \( X \).

Moreover, \( x \in G \setminus H \) and \( y \in H \setminus G \), as required.

Conversely, let \((X, T)\) be a \( T_1 \) space and take \( x \in X \).
For each \( y \in X \setminus \{x\} \) choose an open subset \( G_y \) of \( X \) such that \( y \in G_y \) and \( x \not\in G_y \).
Then, since
\[
X \setminus \{x\} = \bigcup_{y \in X \setminus \{x\}} G_y
\]
is a union of open subsets of \( X \), it is itself an open subset of \( X \).

Hence \( \{x\} \) is a closed subset of \( X \).

We have a similarly convenient criterion for a space to be \( T_2 \).

**Theorem 12.8.** The topological space \((X, T)\) is Hausdorff if and only if the diagonal
\[
\Delta(X) := \{(x, x) \mid x \in X\}
\]
is closed in the product topology on \( X \times X \).

**Proof.** Recall that for all points \( x, y \in X \) and subsets \( G, H \subseteq X \),
(i) \( x \neq y \) if and only if \((x, y) \notin \Delta(X)\), and
(ii) \( G \cap H = \emptyset \) if and only if \((G \times H) \cap \Delta(X) = \emptyset \).

Let \((X, T)\) be a Hausdorff space.

We show that the diagonal is closed in \( X \times X \) by showing that its complement is open.

Take \((x, y) \notin \Delta(X)\), so that \( x \neq y \).

Since \( X \) is Hausdorff, there are disjoint open subsets of \( X \), \( G \) and \( H \) with \( x \in G \) and \( y \in H \).

Thus \((x, y) \in G \times H \subseteq (X \times X) \setminus \Delta(A)\), and \( G \times H \) is an open subset of \( X \times X \).

Hence the complement of the diagonal is open in the product topology.

For the converse, suppose the diagonal is closed in the product topology.

Take \( x, y \in X \), with \( x \neq y \).

Then \((x, y) \notin \Delta(X)\).

As \( \Delta(X) \) is closed, there is a basic open set, \( G \times H \), with
\[
(x, y) \in G \times H \quad \text{and} \quad (G \times H) \cap \Delta(X) = \emptyset
\]

Here \( G \) and \( H \) are open subsets of \( X \).

Since \((x, y) \in G \times H \), \( x \in G \) and \( y \in H \).

Since \((G \times H) \cap \Delta(X) = \emptyset \), \( G \cap H = \emptyset \).

Hence \((X, T)\) is a Hausdorff space.

There is also a useful alternative formulation of the \( T_4 \) separation axiom as well.

**Theorem 12.9.** The topological space \((X, T)\) is \( T_4 \) if and only if given a closed subset, \( F \), and an open subset, \( H \), with \( F \subseteq H \), there is an open subset, \( G \), of \( X \) with
\[
F \subseteq G \subseteq \overline{G} \subseteq H.
\]

**Proof.** Exercise.

Most “naturally occurring” topological spaces are Hausdorff spaces. For example, every metric space is Hausdorff.
Lemma 12.10. Every metric space is Hausdorff.

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

Take \(x, y \in X\) with \(x \neq y\).

Put \(r := \frac{1}{2} g(x, y)\).

Then \(r > 0\), \(G := B(x; r)\) and \(H := B(y; r)\) are disjoint open sets with \(x \in G\) and \(y \in H\).

Observation 12.11. The above separation axioms are intrinsic in that they refer only to elements and subsets of the space in question. Certain extrinsic separation properties are important to analysis.

Extrinsic and intrinsic properties linked, for the intrinsic properties of a space clearly affect its extrinsic properties.

Sometimes extrinsic properties can characterise intrinsic ones. For example, we found that the topological space \((X, T)\) is connected — a notion we defined intrinsically — if and only if every continuous function into a test space is constant — an extrinsic criterion.

Our next theorem provides another illustration.

Theorem 12.12 (Urysohn’s Lemma). The topological space \((X, T)\) is a T4 space if and only if given disjoint closed subsets of \(X\), \(A\) and \(B\), there is a continuous function, \(f : X \to [0, 1]\) with

\[
f(x) = \begin{cases} 
0 & \text{for } x \in A \\
1 & \text{for } x \in B
\end{cases}
\]

Proof. Let \((X, T)\) be a topological space.

Suppose that for all non-empty disjoint closed subsets of \(X\), \(A\) and \(B\), there is a continuous function \(f : X \to [0, 1]\) with

\[
f(x) = \begin{cases} 
0 & \text{for } x \in A \\
1 & \text{for } x \in B
\end{cases}
\]

Let \(K, L\) be disjoint closed subsets of \(X\).

Take \(f : X \to [0, 1]\) with \(f(K) = \{0\}\) and \(f(L) = \{1\}\), and put

\[
G = f^{-1}([0, \frac{1}{2}]) \quad H = f^{-1}(\left(\frac{1}{2}, 1\right])
\]

Plainly \(K \subseteq G\), \(L \subseteq H\) and \(G \cap H = \emptyset\).

Since \(f\) is continuous, both \(G\) and \(H\) are open subsets of \(X\).

Hence, \((X, T)\) is a T4 space.

For the converse, let \((X, T)\) be a T4 space and take disjoint closed subsets \(A\) and \(B\) of \(X\).

Since \(A \cap B = \emptyset\), \(A \subseteq X \setminus B\).

Since \(B\) is closed, \(X \setminus B\) is open.

Hence, by Theorem 12.9, there is an open subset of \(X\), \(V\), with

\[
A \subseteq V \subseteq X \setminus B.
\]

Since

(i) \(A, V\) are closed,
(ii) $V_2, X \setminus B$ are open, and
(iii) $A \subseteq V_2, V_\frac{1}{2} \subseteq X \setminus B$.

we can apply Theorem 12.9 again to find open subsets of $X, V_2$ and $V_\frac{1}{2}$, with

$$A \subseteq V_2 \subseteq V_\frac{1}{2} \subseteq V_2 \subseteq V_\frac{1}{2} \subseteq X \setminus B.$$ 

We construct inductively a family of open subsets of $X$,

$$\{ V_{\frac{i}{2^n}} | 1 \leq k \leq 2^n - 1, n \in \mathbb{N} \setminus \{0\} \}$$

such that for each $n \in \mathbb{N} \setminus \{0\}$

$$A \subseteq V_{\frac{i}{2^n}} \subseteq V_{\frac{i}{2^n}} \subseteq \cdots \subseteq V_{\frac{i}{2^n-1}} \subseteq V_{\frac{i}{2^n-1}} \subseteq X \setminus B.$$ 

It follows from the the construction of the sets $V_t$, that for $t_1, t_2$ both of the form $\frac{i}{2^n}$ with $t_1 < t_2$,

$$A \subseteq V_{t_1} \subseteq V_{t_2} \subseteq V_{t_2} \subseteq X \setminus B.$$ 

Define

$$f : X \rightarrow [0,1], \quad x \mapsto \inf \{ t \mid x \in V_t \},$$

where $t$ is of form $\frac{i}{2^n}$ with $0 < i < 2^n$ and we take infimum to be 1 whenever $x \in V_t$ for no such $t$.

Clearly, $f$ is a function with $f(x) = 0$ whenever $x \in A$ and $f(x) = 1$ whenever $x \in B$.

It remains to prove that $f$ is continuous.

Since the sets of the form $[0,b]$ and $[a,1]$ with $0 \leq a, b \leq 1$ form a sub-base for the Euclidean topology on $[0,1]$, it suffices to show that $f^{-1}([b,1])$ and $f^{-1}([0,a])$ are closed subsets of $X$ whenever $a, b \in [0,1]$.

For $b \in [0,1]$,

$$f(x) \geq b \quad \text{if and only if} \quad \inf \{ t \mid x \in V_t \} \geq b$$

if and only if $x \notin V_t$ for all $t < b$

if and only if $x \in (X \setminus V_t)$ for all $t < b$

if and only if $x \in \bigcap_{t < b} (X \setminus V_t)$

Since each $V_t$ is an open subset of $X$,

$$f^{-1}([b,1]) = \bigcap_{t < b} (X \setminus V_t)$$

is closed in $X$, being the intersection of closed sets.

For $a \in [0,1]$,

$$f(x) \leq a \quad \text{if and only if} \quad \inf \{ t \mid x \in V_t \} \leq a$$

if and only if $x \in V_t$ for all $t > a$

if and only if $x \in V_t$ for all $t \geq a$

if and only if $x \in \bigcap_{t \geq a} V_t$

Since each $V_t$ is a closed subset of $X$,

$$f^{-1}([0,a]) = \bigcap_{t \geq a} V_t$$

is closed in $X$, being the intersection of closed subsets of $X$. □
Urysohn’s Lemma shows that we can distinguish disjoint closed subsets of $T_4$ spaces by means of continuous real valued functions on the space. This motivates our next definition.

**Definition 12.13.** Let $X$ be a set and take $x, y \in X, x \neq y$.

The set of functions defined on $X$, $\mathcal{F}$, separates $x$ and $y$ if and only if there is a function $f \in \mathcal{F}$ with $f(x) \neq f(y)$

If $A, B \subseteq X$ with $A \cap B = \emptyset$, then $\mathcal{F}$ separates $A$ and $B$ if and only if there is an $f \in \mathcal{F}$ with $f(A) \cap f(B) = \emptyset$

Given a topological space $(X, \mathcal{T})$, $\mathcal{C}(X)$ is the set of all continuous real valued functions on $X$.

The $T_1$ space $(X, \mathcal{T})$ is completely regular if and only if given any closed subset $F$ of $X$ and an $x \in X \setminus F$ the set of real valued functions on $X$, $\mathcal{C}(X)$, separates $F$ and $x$.

Urysohn’s Lemma states that if $X$ is normal, then $\mathcal{C}(X)$ separates disjoint closed subsets of $X$.

Since in a normal space every singleton set is closed, it follows that every normal space is completely regular.

**1. Exercises**

12.1. Prove that every subspace of a Hausdorff space is again Hausdorff.

12.2. Prove that a product of non-empty topological spaces is Hausdorff if and only if each of the spaces is Hausdorff.

12.3. Prove that the topological space $(X, \mathcal{T})$ is $T_4$ if and only if given any closed subset $F$ of $X$ and any open subset, $H$ of $X$, if $F \subseteq H$, then there is an open subset $G$ of $X$ such that $F \subseteq G \subseteq \overline{G} \subseteq H$.

12.4. Prove that every metric space is normal.
Chapter 13

Compactness

It is difficult to overstate the importance, significance and usefulness of the notion of compactness. It is also one of the most puzzling concepts which on first encounter.

A useful way of thinking about compactness is that it is a (generalised) form of finiteness: Compact sets behave, in significant ways, as if they were finite, without necessarily being finite. We make this more precise.

The theory of topological spaces concerns sets with additional structure — a topology — and functions between sets that are compatible with this additional structure — continuous functions. Thus the theory of topological spaces is a special case of the general theory of sets and functions between them.

On the other hand, every set can be endowed with its discrete topology, and this topology renders continuous every function defined on that space. We can therefore view the theory of sets and functions between them as a special case of the general theory of topological spaces and continuous functions between them!

Thus the theory of topological spaces and continuous functions between them is simultaneously a special case and a generalisation of the theory of sets and functions.

Here we consider topological spaces as generalisations of sets and consider the latter simply as discrete spaces, which we use for motivating examples and inspiration.

This suggests a strategy: Taking properties of sets expressible purely in terms of subsets of a given set and turn these into topological properties by restricting them to open subsets of the given sets.

Finite sets play an important rôle in the theory of sets. To find a counterpart in the theory of topological spaces, we apply our strategy to arrive at the concept of compactness.

Our first observation is that if we restrict attention to Hausdorff (T$_2$), or even T$_1$ spaces, then every finite topological space is discrete. In other words, the theory of finite Hausdorff spaces (or even T$_1$ spaces) and continuous functions between them is nothing but the theory of finite sets and functions between them.

What if the spaces we consider are not finite, or not necessarily T$_1$ spaces?

We reformulate finiteness in terms of a property of subsets.

Note that if the set $X$ is finite and

$$X = \bigcup_{\lambda \in A} G_{\lambda}$$

It is not knowledge, but the act of learning, not the possession of but the act of getting there, which grants the greatest enjoyment.

Karl Friedrich Gauß
where each $G_\lambda \subseteq X$, then there is a finite subset $\{\lambda_1, \ldots, \lambda_n\}$ of $\Lambda$ such that

$$X = \bigcup_{j=1}^n G_{\lambda_j}.$$  

While this is trivially true, it may seem a rather perverse characterisation of finite subsets of a fixed set. But it is one which is readily adapted to the apply to arbitrary topological spaces.

We begin by introducing concepts we use in our generalisation of finiteness.

**Definition 13.1.** Let $(X, \mathcal{T})$ be a topological space.

The collection, $\{G_\lambda \mid \lambda \in \Lambda\}$, of subsets of $X$ is a **covering** of $X$ if and only if $X = \bigcup_{\lambda \in \Lambda} G_{\lambda}$. It is a **finite** if $\Lambda$ is a finite set. $\{G_\lambda \mid \lambda \in \Lambda\}$ is an **open covering** if each $G_\lambda$ is an open subset of $\mathbb{R}$.

If $\{G_\lambda \mid \lambda \in \Lambda\}$ is a covering of $X$, then a **sub-covering** is a covering $\{G_\mu \mid \mu \in M\}$, with $M \subseteq \Lambda$.

Thus, the set $X$ is finite if and only if every covering of $X$ has a finite sub-covering.

We now extend this definition to arbitrary topological spaces, rather than just discrete ones by requiring that the covering sets by open.

**Definition 13.2.** The topological space $(X, \mathcal{T})$ is **compact** if and only if every open covering has a finite sub-covering.

The subset $A$ of the topological space is called **compact** if it is compact with respect to its subspace topology.

**Example 13.3.** Every finite space is compact.

Given that compactness is intended to generalise finiteness, this example is hardly surprising.

Since the set of all real numbers, $\mathbb{R}$, with its Euclidean topology, is probably the most familiar non-trivial topological space, and has served us faithfully as a source of examples and inspiration, it is natural to ask which, if any, subsets of the set of all real numbers are compact with respect to the Euclidean topology.

The first observation is that $\mathbb{R}$ is not, itself, compact.

**Example 13.4.** $\mathbb{R}$, the set of all real numbers together with its Euclidean topology, is not compact.

For neither of the open coverings $\{\mid -n - 1, n + 1\mid n \in \mathbb{N}\}$ $\{\mid n - 1, n + 1\mid n \in \mathbb{Z}\}$ has a finite sub-covering.

This makes the question of which subsets of $\mathbb{R}$ are compact a non-trivial one.

The **Heine-Borel Theorem** asserts that $\mathbb{R}$ as an abundance of compact subsets, and provides a complete characterisation of them.

**Theorem 13.5 (Classical Heine-Borel Theorem).** Given real numbers $a \leq b$, $[a, b]$ is a compact subset of $\mathbb{R}$.

The Heine-Borel Theorem is more customarily seen in a different form, which we quote next.

---

1 We note here that our formulation of the Heine-Borel Theorem is a reformulation. The Heine-Borel Theorem actually states that any open covering of $[a, b]$ has a finite sub-covering. Since this Heine-Borel property serves as our definition of compactness, the two formulations are equivalent.
Theorem 13.6 (Heine-Borel Theorem). A subset of $\mathbb{R}$ is compact if and only if it is closed and bounded.

We defer the proof until the study of compactness in the special case of metric spaces, where it is an easy corollary of a more general theorem, which is no more difficult than a direct proof of the Heine-Borel Theorem. The Heine-Borel Theorem appears as Corollary 16.16.

Before presenting examples of compact spaces beyond those in Theorem 13.5, we turn to general properties of compact spaces, as this simplifies the construction of examples.

We first show that continuous functions preserve compactness.

Theorem 13.7. The continuous image of a compact space is compact.

Proof. Let $(X, T)$ be a compact space and $f : X \rightarrow Y$ a continuous function onto $(Y, U)$.

Let $\{H_\lambda \mid \lambda \in \Lambda\}$ be an open covering of $Y$.

Since $f$ is a function, $\{f^{-1}(H_\lambda) \mid \lambda \in \Lambda\}$ is a covering of $X$.

By the continuity of $f$, each $f^{-1}(H_\lambda)$ is open in $X$.

By the compactness of $X$, there are $\lambda_1, \ldots, \lambda_n \in \Lambda$ with

$$X = \bigcup_{j=1}^n f^{-1}(H_{\lambda_j}).$$

Then

$$Y = f(X) \quad \text{as } f \text{ is surjective}$$

$$= f\left( \bigcup_{j=1}^n f^{-1}(H_{\lambda_j}) \right)$$

$$= \bigcup_{j=1}^n f\left(f^{-1}(H_{\lambda_j})\right)$$

$$= \bigcup_{j=1}^n H_{\lambda_j} \quad \text{as } f \text{ is surjective}$$

□

This theorem, together with the Heine-Borel Theorem, has important consequences, one of which is the Extreme Value Theorem of Calculus.

Corollary 13.8. Every real valued function defined on a compact space has a maximum and a minimum.

Proof. Let $X$ be compact and $f : X \rightarrow \mathbb{R}$ continuous.

By Theorem 13.7, $\text{im}(f)$ is a compact subset of $\mathbb{R}$.

By Theorem 13.6, $\text{im}(f)$ is a closed and bounded subset of $\mathbb{R}$.

Since $\text{im}(f)$ is a bounded subset of $\mathbb{R}$, it has both a supremum and an infimum.

Since these belong to the closure of the closed subset, $\text{im}(f)$, of $\mathbb{R}$, they must be, respectively, the maximum and the minimum. □

Corollary 13.9 (The Extreme Value Theorem). Take real numbers $a, b$ with $a < b$.

The every continuous function $f : [a, b] \rightarrow \mathbb{R}$ has both a maximum and a minimum.

Proof. By Theorem 13.5, $[a, b]$ is compact.

The result now follows directly from Corollary 13.8. □
Corollary 13.10. *Any space homeomorphic with a compact space is compact.*

Corollary 13.10 shows that compactness is a topological property — given two homeomorphic spaces, either both have, or neither has, this property. In particular, the compactness of a metric space does not depend on the metric itself. This fact is important in numerous applications of topology.

We next show that the property compactness is inherited by closed subsets of compact spaces.

**Theorem 13.11.** *Every closed subset of a compact space is compact.*

**Proof.** Let \( A \) be a closed subset of the compact space \((X, \mathcal{T})\).

Let \( \{G_\lambda \mid \lambda \in \Lambda \} \) be a family of open subsets of \( X \) with \( A \subseteq \bigcup G_\lambda \).

Since \( A \) is closed, \( G := X \setminus A \) is open and

\[
X \subseteq G \cup \bigcup_{\lambda \in \Lambda} G_\lambda.
\]

By the compactness of \( X \) there are \( \lambda_1, \ldots, \lambda_n \in \Lambda \) with

\[
X \subseteq G \cup \left( \bigcup_{j=1}^{n} G_{\lambda_j} \right)
\]

As \( G \cap A = \emptyset \), \( A \subseteq \bigcup_{j=1}^{n} G_{\lambda_j} \).

\( \square \)

The converse of Theorem 13.11 is false in general

**Example 13.12.** Let \( A \) be a non-empty finite proper subset of any set \( X \) endowed with the indiscrete topology.

Then \( A \) is compact, because it is finite.

But \( A \) is not closed, because the only two closed subsets are \( \emptyset \) and \( X \).

There is, however, a partial converse to Theorem 13.11.

**Theorem 13.13.** *A compact subset of a Hausdorff space is closed.*

**Proof.** Let \( A \) be a compact subset of the compact Hausdorff space \((X, \mathcal{T})\).

We use Theorem 13.11 to show that \( A \) is closed by showing by showing that each \( x \in X \setminus A \) is contained in an open subset \( G \) of \( X \) with \( G \cap A = \emptyset \).

Take \( x \in X \setminus A \).

Since \( (X, \mathcal{T}) \) is Hausdorff, for each \( a \in A \), there are disjoint open subsets of \( X \), \( H_a \) and \( G_a \), with \( a \in H_a \) and \( x \in G_a \).

Plainly,

\[
\{ H_a \mid a \in A \}
\]

is an open covering of \( A \).

As \( A \) is compact, this open covering has a finite open sub-covering,

\[
\{ H_{a_j} \mid j = 1, \ldots, n \}
\]

Put

\[
G := \bigcap_{j=1}^{n} G_{a_j}
\]
Then $G$ is an open subset of $X$, as it is the intersection of finitely many open subsets of $X$.
Moreover,

$$A \cap G = \emptyset$$

For if $y \in A$, then $y \in H_{a_j}$ for some $j \in \{1, \ldots, n\}$ and so $y \notin G_{a_j}$.
Since $G \subseteq G_{a_j}$, $y \notin G$. 

**Example 13.14.** By Theorem 13.13, if $a, b$ are real numbers with $a < b$, then neither $]a,b[$ nor $]a,b]$ can be compact, for neither is a closed subset of the Hausdorff space $\mathbb{R}$.

We have seen that, in general, a continuous bijection need not be a homeomorphism. We provide a sufficient (but not necessary!) condition for a continuous bijection to be a homeomorphism.

**Theorem 13.15.** A continuous bijection from a compact space onto a Hausdorff space is a homeomorphism.

**Proof.** Let $f: X \to Y$ be a continuous bijection of the compact space $X$ onto the Hausdorff space $Y$.

Since $f$ is a continuous bijection, we only need to show that its inverse function, $f^{-1}$, is also continuous.

By Theorem 10.4, this is equivalent to showing that $(f^{-1})^{-1}(A)$ is closed in $A$ whenever $A$ is closed in $X$.

But $(f^{-1})^{-1}(A) = f(A)$ for every $A \subseteq X$.

Hence is enough to show that $f$ is a closed mapping.

Let $A$ be a closed subset of the compact space $X$.

By Theorem 13.11, $A$ is compact.

By Theorem 13.7, $f(A)$ a compact subset of the Hausdorff space $Y$.

By Theorem 13.13, $f(A)$ is closed in $Y$. 

We next examine the relationship between the compactness of a product of spaces and the compactness of the individual factors.

Tychonov’s Theorem states that the product of any set of compact spaces is compact. We prove this only for the case of two spaces. A simple induction then establishes it for any finite set of compact spaces. The technique of proof is typical for arguments using compactness.

However, this proof does not apply to the general case, when the collection of compact spaces is not presumed to be finite.

**Theorem 13.16.** $(X, \mathcal{T})$ and $(Y, \mathcal{U})$ are compact if and only if $X \times Y$ is compact.

**Proof.** If $X \times Y$ is compact, then so must $X$ and $Y$ be compact, because the natural projections, $pr_X: X \times Y \to X$ and $pr_Y: X \times Y \to Y$ are continuous surjections.

For the converse, let $(X, \mathcal{T})$ and $(Y, \mathcal{U})$ be compact.

Since any open set is a union of basic open sets, it is sufficient to consider open coverings of $X \times Y$ by of basic open sets.

Moreover, since there are only two factors, we can take as basic open sets for the product topology on $X \times Y$, those of the form $G \times H$, with $G$ an open subset of $X$ and $H$ an open subset of $Y$.

Let $\{G_\lambda \times H_\lambda \mid \lambda \in A\}$ be an open covering of $X \times Y$ by basic open sets.

Choose $x \in X$.

Since $\{G_\lambda \times H_\lambda \mid \lambda \in A\}$ is an open covering of $X \times Y$, it is, a fortiori, an open covering of $\{x\} \times Y$. 

Since \( \{x\} \times Y \) is homeomorphic to \( Y \), it is a compact subset of \( X \times Y \).
Thus, there is a finite subset, \( \{G_j^x \times H_j^x \mid j = 1, \ldots, n_x\} \), of \( \{G_\lambda \times H_\lambda \mid \lambda \in \Lambda\} \) which is an open covering of \( \{x\} \times Y \).
Put
\[
G^x := \bigcap_{j=1}^{n_x} G_j^x.
\]
As \( x \in G^n \), we obtain the open covering of \( X \)
\[
\{G^x \mid x \in X\}
\]
Since \( X \) is compact, there are \( x_1, \ldots, x_m \in X \) with \( \{G^{x_i} \mid i = 1, \ldots, m\} \) is an open covering of \( X \).
We simplify notation, writing
\[
\begin{align*}
n_i & \text{ for } n_{x_i}, \\
G_{ij} & \text{ for } G_{x_i}^j, \text{ and} \\
H_{ij} & \text{ for } H_{x_i}^j.
\end{align*}
\]
Then \( \{G_{ij} \times H_{ij} \mid j = 1, \ldots, n_i, i = 1, \ldots, m\} \) is a finite subset of \( \{G_\lambda \times H_\lambda \mid \lambda \in \Lambda\} \).
We show that it covers \( X \times Y \).
Take \( (x, y) \in X \times Y \).
Then \( x \in G_{i_0j} \) for some \( i_0 \in \{1, \ldots, m\} \) and every \( j \in \{1, \ldots, n_{i_0}\} \).
Since \( \{H_{i_0j} \mid j = 1, \ldots, n_{i_0}\} \) is an open covering of \( Y \), there is a \( j_0 \in \{1, \ldots, n_{i_0}\} \) with \( y \in H_{i_0j_0} \).
Thus \( (x, y) \in G_{i_0j_0} \times H_{i_0j_0} \).
Hence, \( \{G_{ij} \times H_{ij} \mid j = 1, \ldots, n_i, i = 1, \ldots, m\} \) is a covering of \( X \times Y \).  

**Corollary 13.17.** The product of finitely many compact spaces is compact.

**Corollary 13.18 ((Generalised) Heine-Borel Theorem).** A subset of \( \mathbb{R}^n \) is compact if and only if it is closed and bounded.

**Proof.** Let \( A \) be a compact subset of \( \mathbb{R}^n \).
Since \( \mathbb{R}^n \) is a Hausdorff space, it follows by Theorem 13.13, that \( A \) is closed.
As \( pr_j : \mathbb{R}^n \rightarrow \mathbb{R} \), the projection onto the \( j \)-th factor \( \mathbb{R} \), is continuous, it follows by Theorem 13.7 that \( pr_j(A) \) is a compact subset of \( \mathbb{R} \).
By the Heine-Borel Theorem \( pr_j(A) \) is closed and bounded.
Thus \( pr_j(A) \subseteq I_j \) for some closed finite interval \( I_j \).
Then \( A \subseteq I_1 \times \cdots \times I_n \), showing that \( A \) is bounded.
Conversely, if \( A \) is closed and bounded, then it is contained in some \( I_1 \times \cdots \times I_n \) where each \( I_j \) is a closed finite interval in \( \mathbb{R} \).
By the Heine-Borel Theorem, each \( I_j \) is compact.
By Theorem 13.16, \( I_1 \times \cdots \times I_n \) is also compact.
Since \( A \) is a closed subset of a compact set, it is compact by Theorem 13.11.

**Observation 13.19.** Tychonov’s Theorem states that the product of any set (not just a finite set) of compact spaces is again compact. We have proved the finite case, but our proof does not extend to the case of infinitely many factors in the product.
The usual proof of Tychonov’s Theorem in the general case uses the Axiom of Choice, which states that given any non-empty set, $\mathcal{S}$, of non-empty sets, there is a set which contains one element from each of the sets in $\mathcal{S}$. This axiom is independent of the other usual axioms for set theory.

There are numerous equivalent formulations of the Axiom of Choice, including Zorn’s Lemma and the Well-Ordering Principle. In fact, Tychonov’s Theorem is also equivalent to it.

1. The Fundamental Theorem of Algebra

As an application of compactness, we prove the Fundamental Theorem of Algebra, which states that every non-constant polynomial in one indeterminate with complex coefficients is a product of linear terms.

Remark 13.20. It is truly remarkable that while the Fundamental Theorem of Algebra, as stated above, is purely algebraic, with no apparent connection with topology, all known proofs of it depend crucially on an argument using compactness, which is a topological property.

Observation 13.21. The above statement of the Fundamental Theorem of Algebra is equivalent to the statement that every non-constant polynomial function $\mathbb{C} \to \mathbb{C}$ has a zero:

If $p(z)$ is a polynomial in $z$ with complex coefficients, of degree at least 1, then there is a $\zeta \in \mathbb{C}$ with $p(\zeta) = 0$.

For given any $\zeta \in \mathbb{C}$

$$p(z) = (z - \zeta)q(z) + r$$

where $q(z)$ is a polynomial function of $z$ whose degree is precisely one less than that of $p(z)$ and $r$ is a complex number.

By direct substitution

$$p(\zeta) = r$$

whence $(z - \zeta)$ is a factor of $p(z)$ if and only if $p(\zeta) = 0$, and then

$$p(z) = (z - \zeta)q(z)$$

Repeat the argument with $q(z)$, and use downward induction until the quotient has degree 0.

The key technical result we need is contained in the next lemma, which explains the need for the complex numbers, rather than just real numbers.

Lemma 13.22. Let $f(z) := \sum_{j=0}^{n} c_j z^j$, with $n > 0$, each $c_j \in \mathbb{C}$ and $c_n \neq 0$.

Then either $f(0) = 0$ or there is a $\zeta \in \mathbb{C}$ with $|f(\zeta)| < |f(0)|$.

Proof. Suppose that $f(0) \neq 0$, that is, that $c_0 \neq 0$.

Then

$$f(z) = c_0 + c_k z^k + z^{k+1}g(z)$$

where $k := \min\{j > 0 \mid c_j \neq 0\}$ and $g(z)$ is a suitable polynomial in $z$. (Note that $k \leq n$).

Choose $\omega \in \mathbb{C}$ with \footnote{This is where the real numbers are not enough, and the complex numbers are needed.}

$$\omega^k = -\frac{c_0}{c_k} \quad (\ast)$$

As $h: \mathbb{R} \to \mathbb{C}$, $t \mapsto \omega^{k+1} g(t\omega)$ is a polynomial function of $t$,

$$t\omega^{k+1} g(t\omega) \to 0 \quad \text{as} \quad t \to 0$$
Hence there is a $T > 0$ such that for all $t \in \mathbb{R}$ with $|t| < T$,
\[ t \omega^{k+1} g(t\omega) < |c_0| \quad (**) \]

Putting $t := \min \left\{ \frac{T}{2}, \frac{1}{2} \right\}$, we see that $0 < t^k < 1$ and
\[
|f(t\omega)| \leq |c_0 + c_k \omega^k| + t^k |\omega^{k+1} g(t\omega)|
= |c_0(1 - t^k) + t^k |\omega^{k+1} g(t\omega)|
\]
by (*)
\[
< |c_0(1 - t^k) + t^k|c_0|
\]
by (**) and
\[
= |f(0)|
\]

\[ \square \]

**Theorem 13.23 (Fundamental Theorem of Algebra).** A polynomial with complex coefficients has a zero unless it is constant.

**Proof.** Let $p$ be a polynomial of degree at least 1. Then $|p(z)| \to \infty$ as $|z| \to \infty$.

Hence there is a $K > 0$ such that for all $z$ with $|z| > K$
\[ |p(z)| > |p(0)| \]

Put
\[ D := \{ z \in \mathbb{C} \mid |z| \leq K \}. \]

$D$ is compact and
\[ |p| : D \to \mathbb{R}, \quad z \mapsto |p(z)| \]
is continuous.

Hence, by Corollary 13.8, it has a minimum $|p(\zeta)|$.

Define $q : \mathbb{C} \to \mathbb{C}$ by $q(z) := p(z + \zeta)$.

Then $q(0) = p(\zeta)$ and $q(z)$ is a polynomial function of $z$ which takes precisely the same values as $p(z)$.

Hence $|q(0)|$ is the minimum value taken by $|q(z)|$.

By Lemma 13.22 $p(\zeta) = q(0) = 0$. \[ \square \]

2. **Exercises**

13.1. Prove that the intersection of any family of compact subsets of a topological space is compact.

13.2. Put $S^1 := \{(x, y) \mid x^2 + y^2 = 1\} \subseteq \mathbb{R}^2$.

Prove that $S^1$ is compact with respect to the subspace topology.

13.3 (The Cantor Set). This exercise is devoted to the construction of the Cantor set which, with its generalisations, provides important and instructive examples in measure and integration theory, in the theory of dynamical systems, in particular, in chaos theory, as well as elsewhere.

We establish some of its elementary properties.

We begin with the closed unit interval $I := [0, 1]$, and remove from it the open middle third, to obtain the union of two closed intervals, each of length $\frac{1}{3}$.

At the first stage, remove the open middle third from each of these closed intervals, and continue in this fashion, doubling the number of closed intervals at each stage, and reducing their lengths by a factor of $\frac{1}{3}$. 
After the $k^{\text{th}}$ stage, we obtain $2^k$ disjoint closed sub-intervals of $\mathcal{I}$, each of length $3^{-k}$.

We express this in terms of a decreasing sequence, $(I_n)_{n \in \mathbb{N}}$, of subsets of $\mathcal{I}$.

We put

$$ I_0 := \mathcal{I} := [0, 1] $$

Suppose that for some $n \in \mathbb{N}$

$$ I_n \subseteq \mathcal{I} $$

is the union of $2^n$ disjoint closed sub-intervals of $\mathcal{I}$ of the form

$$ \left[ \frac{j}{3^n}, \frac{j + 1}{3^n} \right], \quad j \in N_n $$

a suitable set of natural numbers not exceeding $3^n$.

Then

$$ I_{n+1} := \bigcup_{j \in N_n} \left( \left[ \frac{3j}{3^{n+1}}, \frac{3j + 1}{3^{n+1}} \right] \cup \left[ \frac{3j + 2}{3^{n+1}}, \frac{3j + 3}{3^{n+1}} \right] \right), $$

and so

$$ N_0 := \{0\} $$

$$ N_{n+1} := \{3j, 3j + 2 \mid j \in N_n\}. $$

Clearly $I_{n+1} \subseteq I_n$ for all $n \in \mathbb{N}$.

The Cantor set, $C$, is defined by

$$ C := \bigcap_{n \in \mathbb{N}} I_n. $$

(1) Prove that $C$ is in bijection with $\mathcal{I}$.

(2) Prove that $C$ is totally disconnected, that is to say, each connected component of $C$ is a singleton set.

(3) Prove that $C$ is compact.

13.4. Show that a union of compact sets need not be compact.

13.5. Find an example of a continuous function $f : X \rightarrow Y$ and a compact subset, $K$, of $Y$ such that $f^{-1}(K)$ is not a compact subset of $X$. 
Abstract mathematics combined with physics can discover and see more than any eye ever could.

Klaudio Marash

Chapter 14

Locally Compact Spaces

The set of real numbers \( \mathbb{R} \) has served as a source of inspiration for many notions in topology: For example, closed finite intervals are the archetypical infinite compact sets.

While \( \mathbb{R} \) is not a compact space, it is “almost” a compact space, in the sense that every point in \( \mathbb{R} \) is contained in a compact interval. We formulate this precisely to apply to arbitrary topological spaces.

Definition 14.1. The topological space \((X, T)\) is locally compact if and only if every element of \( X \) is contained in an open set whose closure is compact: For each \( x \in X \) there is an open subset, \( G \), of \( X \) with \( x \in G \) and \( \overline{G} \) compact.

Example 14.2. Every compact topological space is locally compact.

Example 14.3. Every discrete space is locally compact.

Example 14.4. \( \mathbb{R} \), the set of all real numbers with its Euclidean topology, is not compact. For example, there is no finite sub-covering of the open covering

\[
\left\{ \left[ n-1, n+1 \right] \mid n \in \mathbb{Z} \right\}
\]

But \( \mathbb{R} \) is locally compact because given any \( x \in \mathbb{R} \) and any \( \varepsilon > 0 \), \( G := \left[ x - \varepsilon, x + \varepsilon \right] \) is open and its closure, \( \overline{G} = [x - \varepsilon, x + \varepsilon] \), is compact.

We show that a locally compact Hausdorff space is almost compact, in that it can be regarded as a dense subspace of a compact space. We first formulate this precisely.

Definition 14.5. A compactification of the topological space \((X, T)\) comprises a compact topological space \((\hat{X}, \hat{T})\) and a continuous injective function

\[
\gamma: X \longrightarrow \hat{X}
\]

such that \( \text{im}(\gamma) \) is dense in \( \hat{X} \) and

\[
\pi_\gamma: X \longrightarrow \text{im}(\gamma), \quad x \mapsto \gamma(x)
\]

is a homeomorphism.

Example 14.6. Take \( A = \left[ -1, 1 \right] \), \( B = \left] -1, 1 \right] \), \( C = \left[ -1, 1 \right] \) and \( D = \left[ -1, 1 \right] \), each with its Euclidean topology.

Then none of \( A, B \) or \( C \) is compact, since none is a closed subset of the Hausdorff space \( \mathbb{R} \).

On the other hand, by the Heine-Borel Theorem (Theorem 13.5), \( D \) is compact.
The inclusions

\[ i^D_A : A \to D, \quad x \mapsto x \]
\[ i^D_B : B \to D, \quad x \mapsto x \]
\[ i^D_C : C \to D, \quad x \mapsto x \]

are continuous injections with image dense in \( D \).

Each map

\[ \pi^D_A : A \to \text{im}(i^D_A), \quad x \mapsto i^D_A(x) = x \]
\[ \pi^D_B : A \to \text{im}(i^D_B), \quad x \mapsto i^D_B(x) = x \]
\[ \pi^D_C : A \to \text{im}(i^D_C), \quad x \mapsto i^D_C(x) = x \]

is a homeomorphism.

Thus, \( D \) is a compactification of each of \( A, B \) and \( C \).

**Observation 14.7.** Notice that both \( A \) and \( B \) were compactified in Example 14.6 by the addition of one single new point. This is true of any locally compact Hausdorff space, as we next show.

**Theorem 14.8 (Alexandroff (1-Point) Compactification).** Let \((X, T)\) be a locally compact Hausdorff space, which is not compact.

Then there is a compact Hausdorff space \((\hat{X}, \hat{T})\), obtained by adjoining a single point to \(X\)\(^1\), and a continuous injective function,

\[ \alpha : X \to \hat{X}, \quad x \mapsto x \]

such that \( \text{im}(\alpha) = \hat{X} \) and

\[ \pi_\alpha : X \to \text{im}(\alpha) \]

is a homeomorphism.

**Proof.** Take \( \infty \notin X \), put

\[ \hat{X} := X \cup \{\infty\}, \]

and define

\[ \alpha : X \to \hat{X}, \quad x \mapsto x. \]

Plainly, \( \alpha \) is injective and for any \( H \subseteq \hat{X} \),

\[ \alpha^{-1}(H) = H \setminus \{\infty\} \]

We define the topology, \( \hat{T} \), on \( \hat{X} \) by extending the topology, \( T \), on \( X \): we add open sets to contain the additional element \( \infty \).

Formally, let \( \hat{T} \) be the topology generated by the sets of the form

(i) \( G \in T \),\(^2\) and

(ii) \( (X \setminus K) \cup \{\infty\} \), where \( K \) is a compact subset of \( X \).

In other words, the topology, \( \hat{T} \), has a sub-base

\[ B := \{G, (X \setminus K) \cup \{\infty\} | G \in T, \ K \text{ is a compact subset of } X\} \]

We show \( B \) is, in fact, a base for \( \hat{T} \), by proving that \( B \) is closed under finite intersections.

Take \( U, V \in B \).

\(^1\)The additional point is often referred to as the point at infinity

\(^2\)Strictly speaking, we should write \( \alpha(G) \) for \( G \in T \), but we identify \( x \in X \) with \( \alpha(x) \in \hat{X} \) because \( \hat{X} = X \cup \{\infty\} \).
(a) If \( U \) and \( V \) are both of type (i), then they are open subsets of \( X \), in which case \( U \cap V \) is also an open subset of \( X \), whence \( U \cap V \) is also of type (i).

(b) If \( U \) and \( V \) are both of type (ii), there are compact subsets of \( X \), \( K \) and \( L \), with
\[
\begin{align*}
U &= (X \setminus K) \cup \{\infty\} \text{ and } V = (X \setminus L) \cup \{\infty\}. \\
U \cap V &= ((X \setminus K) \cup \{\infty\}) \cap ((X \setminus K) \cup \{\infty\}) \\
&= \{(X \setminus K) \cap (X \setminus K)\} \cup \{\infty\} \\
&= (X \setminus (K \cup L)) \cup \{\infty\}
\end{align*}
\]
which is again of type (ii), since the union of two compact sets is again compact.

(c) If \( U \) is of type (ii) and \( V \) is of type (i), then there is a compact subset, \( K \), of \( X \) with
\[
U = (X \setminus K) \cup \{\infty\}.
\]
Thus, \( U \cap V = (X \setminus K) \cap V \).
Since \( K \) is a compact subset of the Hausdorff space \( X \), it follows by Theorem 13.13, that \( K \) is closed.
Thus \((X \setminus K)\) is open in \( X \), whence \( U \cap V \) is of type (i).

Induction now shows that \( \mathcal{B} \) is closed under finite intersections.

It is immediate from the definition of \( \hat{T} \) that \( \alpha \) is an open mapping.

To show that \( \alpha \) is continuous, it is enough, by Theorem 7.15, to show that \( \alpha^{-1}(H) \) is an open subset of \( X \) whenever \( H \) is either of type (i) or of type (ii).

By (\(\diamond\)),
\[
\alpha^{-1}(H) = \begin{cases} 
  H & \text{when } H \text{ is of type (i)} \\
  X \setminus K & \text{when } H \text{ is of type (ii)}
\end{cases}
\]

In the former case, \( H = \alpha^{-1}(H) \) is an open subset of \( X \), by the definition of open subsets of \( \hat{X} \) of type (i).

In the latter case, \( K \), being a compact subset of the Hausdorff space \( X \), is a closed subset of \( X \), rendering \( X \setminus K \) and open subset.

Hence, in both case, \( \alpha^{-1}(H) \) is an open subset of \( X \), showing that \( \alpha \) is continuous.

Since \( \alpha \) is a continuous, open injective mapping, it defines the homeomorphism
\[
\pi_{\alpha} : X \xrightarrow{\alpha} \alpha(X), \quad x \mapsto \alpha(x)
\]
which we use to identify \( X \) with \( \alpha(X) \).

To show that \( \text{im}(\alpha) = \alpha(X) \) is dense in \( \hat{X} \), it is sufficient to show that if \( H \) is a basic open subset of \( \hat{X} \) and contains \( \infty \), then \( H \cap \alpha(X) \neq \emptyset \).

Such a basic open set is of type (ii), so that there is compact subset, \( K \), of \( X \), with
\[
H = (X \setminus K) \cup \{\infty\}
\]
Then
\[
H \cap \alpha(X) = \left((X \setminus K) \cup \{\infty\}\right) \cap X \\
= X \setminus K \\
\neq \emptyset 
\]
since \( X \) is not compact.

To see that \( \hat{X} \) is Hausdorff, it is enough to show that for each \( x \in \alpha(X) \) there are disjoint open subsets, \( G, H \), of \( \hat{X} \) with \( x \in G \) and \( \infty \in H \).

Since \( X \) is locally compact, there is a \( G \in \mathcal{T} \) containing \( x \), with \( \overline{G} \) compact. Put
\[
H := (X \setminus \overline{G}) \cup \{\infty\}. 
\]
Then $H$ is a basic open subset of $\hat{X}$ containing $\infty$ and

$$G \cap H = G \cap (X \setminus \overline{G}) = \emptyset$$

as $G \subseteq \overline{G}$

Finally, to see that $\hat{X}$ is compact, let $\{G_\lambda \mid \lambda \in \Lambda\}$ be an open covering of $\hat{X}$ by basic open sets.

There is a $\mu \in \Lambda$ with $\infty \in G_\mu$.

By definition, $K := X \setminus G_\mu$ is a compact subset of $X$.

Then $\{G_\lambda \mid \lambda \in \Lambda \setminus \{\mu\}\}$ is an open covering of $K$.

Since $K$ is compact, there is a finite subset, $\{\lambda_1, \ldots, \lambda_n\}$, of $\Lambda$ with

$$K \subseteq \bigcup_{j=1}^n G_{\lambda_j}.$$  

Thus $\{G_{\lambda_1}, \ldots, G_{\lambda_n}, G_\mu\}$ is a finite open sub-covering of $\hat{X}$.  

\[\square\]

**Corollary 14.9.** The Alexandroff compactification of a non-compact, locally compact Hausdorff space is unique up to homeomorphism.

**Proof.** Exercise.  

**Example 14.10.** Since $\mathbb{R}$ (with the Euclidean topology) is locally compact but not compact, we may form its one-point compactification. This is homeomorphic with $S^1$.

The verification of this is left as an exercise, with a hint: Consider stereographic projection from the “north pole”, $N$ of $S^1$, so that

$$\alpha : \mathbb{R} \to S^1, \quad P \mapsto Q$$

is given as indicated in the following diagram

\[\text{This example can be easily generalised to show that the one-point compactification of } \mathbb{R}^n \text{ is (homeomorphic with)}\]

$$S^n = \{(x_1, \ldots, x_{n+1}) \in \mathbb{R}^{n+1} \mid \sum_{j=1}^{n+1} x_j^2 = 1\}$$

The details are left as an exercise.

In particular, $\mathbb{C}$ with its Euclidean topology is homeomorphic with $\mathbb{R}^2$ with its Euclidean topology. The one-point compactification of $\mathbb{C}$ is therefore $S^2$. When $S^2$ is regarded as the Alexandroff compactification of $\mathbb{C}$, it is referred to as the Riemann sphere.

\[\section{1. Compactification and Calculus}\]

Compactification provides an insight into the significance of limits in calculus, for, as the next examples illustrate, the existence of limits in elementary calculus is equivalent to continuous functions having continuous extensions to a compactification of the domain.
Example 14.11 (Convergent Sequences of Real Numbers). Recall that identifying the sequence\(^3\) of real numbers \((x_n)_{n \in \mathbb{N}}\) with the function
\[
s : \mathbb{N} \to \mathbb{R}, \quad n \mapsto x_n
\]
allows us to identify the set of all sequences of real numbers with the set of all functions \(s : \mathbb{N} \to \mathbb{R}\), where we take \(\mathbb{N}\) with its Euclidean topology.

Since \(\mathbb{N}\) is a discrete space, every function defined on \(\mathbb{N}\) is continuous.

Moreover, since \(\mathbb{N}\) is homeomorphic with the subspace
\[
X := \left\{ \frac{1}{n+1} \mid n \in \mathbb{N} \right\}
\]
of \(\mathbb{R}\), under the function
\[
i : \mathbb{N} \to X, \quad n \mapsto \frac{1}{n+1}
\]
we may identify the set of all sequences in \(\mathbb{R}\) with the set of all continuous functions \(X \to \mathbb{R}\).

It is left as an exercise (Exercise 14.4) for the reader to show that the Alexandroff compactification of \(X\) is (homeomorphic with)
\[
\hat{X} = \left\{ \frac{1}{n+1} \mid n \in \mathbb{N} \right\} \cup \{0\}
\]
with its Euclidean topology.

Identifying the sequence of real numbers, \((x_n)_{n \in \mathbb{N}}\), with the function
\[
f : X \to \mathbb{R}, \quad \frac{1}{n+1} \mapsto x_n
\]
we can reformulate the definition of convergent sequences in terms of the Alexandroff compactification of \(\mathbb{N}\) (up to homeomorphism).

**Theorem 14.12.** The sequence of real numbers \((x_n)_{n \in \mathbb{N}}\) converges if and only if the corresponding (continuous) function \(f : X \to \mathbb{R}\) can be extended to a continuous functions \(\hat{f} : \hat{X} \to \mathbb{R}\).

Moreover, in such a case,
\[
\lim_{n \to \infty} x_n = \hat{f}(0).
\]

**Observation 14.13.** Note that Example 14.11 did not use property of \(\mathbb{R}\) other than its being a Hausdorff topological space. This means that Theorem 14.12 can be generalised without further ado to sequences in any topological space.

**Theorem.** The sequence \((x_n)_{n \in \mathbb{N}}\) of elements of the Hausdorff topological space \((Y, U)\) converges if and only if the corresponding (continuous) function \(f : X \to Y\) can be extended to a continuous functions \(\hat{f} : \hat{X} \to Y\).

Moreover, in such a case,
\[
\lim_{n \to \infty} x_n = \hat{f}(0).
\]

**Example 14.14 (“Improper Limits” of Real Functions).** We turn to considering \(\lim_{x \to \infty} f(x)\) and \(\lim_{x \to -\infty} f(x)\) for the continuous function \(f : \mathbb{R} \to \mathbb{R}\).

Since both the bijection
\[
\tan : ] -1, 1[ \to \mathbb{R}, \quad t \mapsto x := \tan \left( \frac{\pi t}{2} \right)
\]
\(^3\)We investigate sequences more thoroughly in Chapter 15
and its inverse are analytic functions, we may identify the set of all functions $\mathbb{R} \rightarrow \mathbb{R}$ with the set of all functions $]-1,1[ \rightarrow \mathbb{R}$, preserving all continuity and differentiality properties.

Let the function $f: \mathbb{R} \rightarrow \mathbb{R}$ be identified with $g: ]-1,1[ \rightarrow \mathbb{R}$, where

$$g(t) := f \left( \tan \left( \frac{\pi t}{2} \right) \right)$$

From elementary calculus,

$$\lim_{x \to -\infty} f(x) = a \quad \text{if and only if} \quad \lim_{t \to -1} g(t) = a$$

$$\lim_{x \to \infty} f(x) = b \quad \text{if and only if} \quad \lim_{t \to 1} g(t) = b$$

We summarise these considerations, using the notation above and from Example 14.6,

**Theorem.** Let $f: \mathbb{R} \rightarrow \mathbb{R}$ be continuous. Then the limits $\lim_{x \to \pm \infty} f(x)$ exist if and only if the associated continuous function

$$g: ]-1,1[ \rightarrow \mathbb{R}$$

has a continuous extension

$$\hat{g}: [-1,1] \rightarrow \mathbb{R}$$

to the compactification, $[-1,1]$ of $]-1,1[$. 

We note in passing that this generalises immediately to continuous functions from $\mathbb{R}$ to any topological space.

**Observation 14.15.** Our discussion illustrates how problems in mathematics can reduce to finding extensions of functions:

**Given** $A \subseteq D$ and $f: A \rightarrow Y$ with some specified properties, **find** $\hat{f}: D \rightarrow Y$ with some specified properties such that $\hat{f}$ is the restriction of $f$ to $A$.

Extension problems, and their dual, **lifting problems** are central to modern mathematics and its applications.

The specific examples we have just considered illustrate, in addition, the importance of compactness even in elementary calculus.

### 2. Exercises

14.1. Prove that the Alexandroff compactification of a non-compact, locally compact Hausdorff space is unique up to homeomorphism.

14.2. Show that the one-point compactification of Euclidean $n$-space $\mathbb{R}^n$ is homeomorphic to the $n$-sphere $S^n := \{ (x_0, \ldots, x_n) \in \mathbb{R}^{n+1} \mid \sum_{j=0}^{n} x_j^2 = 1 \}$ with its Euclidean topology.

14.3. Use stereographic projection to identify $\mathbb{R}^n$ with

$$S^n \setminus \{(0, \ldots, 0, 1)\} = \{ (x_1, \ldots, x_{n+1}) \in \mathbb{R}^{n+1} \mid \sum_{j=1}^{n+1} x_j^2 = 1, x_{n+1} \neq 1 \}$$

Take a function $f: \mathbb{R}^n \rightarrow \mathbb{R}^m$, with $\|f(x)\| \rightarrow \infty$ as $\|x\| \rightarrow \infty$.

Prove that there is a unique function $\hat{f}: S^n \rightarrow S^m$ with $\hat{f}|_{\mathbb{R}^n} = f$.

14.4. Prove that the Alexandroff compactification of $\mathbb{N}$ is homeomorphic with

$$\left\{ \frac{1}{n+1} \mid n \in \mathbb{N} \right\} \cup \{0\}$$

with its Euclidean topology.
14.5. Take \( A = ]0, 1[ \) and \( B = ]0, 1] \), each with its Euclidean topology. Show that

\[ \mathbb{S}^1 = \{(x, y) \in \mathbb{R}^2 \mid x^2 + y^2 = 1\} \]

with its Euclidean topology, is a compactification of \( A \), but not of \( B \).
Chapter 15

Sequences, Convergence and Completeness

The starting point for our study of topological spaces was calculus, where sequences comprise one of the central tools. We now turn to the study of sequences in a more general setting.

**Definition 15.1.** A sequence in the set $X$, or a sequence of elements of $X$, is a function $f: \mathbb{N} \rightarrow X$. It is common to write $x_n$ for $f(n)$ and to denote the sequence by $(x_n)_{n \in \mathbb{N}}$ or $(x_n)_{n=0}^\infty$.

The sequence $(x_n)_{n \in \mathbb{N}}$ is an infinite sequence if and only if $\{x_n \mid n \in \mathbb{N}\}$ is an infinite subset of $X$.

Recall that the sequence $(x_n)_{n \in \mathbb{N}}$ of real numbers converges to $a$ if and only if for every $\varepsilon > 0$ there is an $N \in \mathbb{N}$ with $|x_n - a| < \varepsilon$ whenever $n > N$.

We easily generalise this to arbitrary metric spaces, using the stenographic procedure of replacing $|a - b|$ by $\varrho(a, b)$. Formally,

**Definition 15.2.** The sequence $(x_n)_{n \in \mathbb{N}}$ of elements in the metric space $(X, \varrho)$ converges to $a \in X$ if and only if for given $\varepsilon > 0$ all but finitely many terms of the sequence $(x_n)_{n \in \mathbb{N}}$ lie in $B(a; \varepsilon)$ whenever $n > N$.

In such a case, we write $x_n \rightarrow a$ as $n \rightarrow \infty$ or

$$\lim_{n \rightarrow \infty} x_n = a.$$

To see how this definition could be extended to sequences in a topological space, which is not necessarily metrisable, we examine the definition above more closely.

Given the sequence $(x_n)_{n \in \mathbb{N}}$ in the metric space $(X, \varrho)$, Definition 15.2 states that $\lim x_n = a$ if and only if given any $\varepsilon > 0$ all but finitely many terms of the sequence $(x_n)_{n \in \mathbb{N}}$ lie in $B(a; \varepsilon)$.

Let $G$ be an open subset of $X$ with $a \in G$. Then there is an $\varepsilon > 0$ such that $B(a; \varepsilon) \subseteq G$.

---

1The classical definition of a sequence in the set $X$ is that it consists of an ordered set $(x_0, x_1, \ldots, x_n, \ldots)$ of elements of $X$, one element for each $n \in \mathbb{N}$, allowing the possibility that $x_m = x_n$ when $m \neq n$.

This means that each sequence $(x_0, x_1, \ldots, x_n, \ldots)$ of elements of $X$ determines the function $s: \mathbb{N} \rightarrow X, \ n \mapsto x_n$.

Moreover, the function $s: \mathbb{N} \rightarrow X$ determines the sequence $(s(0), s(1), \ldots, s(n), \ldots)$ of elements of $X$.

This relation is easily seen to be a bijection, which allows us to think of sequences as functions.

This has several advantages, not the least of which is that we can use the theory of functions to study sequences.

This is why we have chosen to define a sequence as a function defined on the set of all natural numbers.
Thus, all but finitely many terms of the sequence \((x_n)_{n \in \mathbb{N}}\) must lie in \(G\).

This generalises readily to the arbitrary topological space \((X, T)\).

**Definition 15.3.** The sequence \((x_n)_{n \in \mathbb{N}}\) converges to \(a \in X\) if and only if given \(G\), an open subset of \(X\) with \(a \in G\), then there is an \(N \in \mathbb{N}\) such that \(x_n \in G\) whenever \(n > N\).

Since we take \(\mathbb{N}\) with its Euclidean topology, which is the discrete topology on \(\mathbb{N}\), every function \(s: \mathbb{N} \rightarrow X\) is continuous, for any topological space \((X, T)\):

**A sequence in the topological space, \((X, T)\), is a continuous function** \(s: \mathbb{N} \rightarrow X\).

Since \(\mathbb{N}\) is an infinite discrete space, it is a locally compact, but not compact, Hausdorff space. By Theorem 14.8, we can form its Alexandroff compactification

\[ \widehat{\mathbb{N}} := \mathbb{N} \cup \{\infty\} \]

which, together with the embedding

\[ \alpha: \mathbb{N} \rightarrow \widehat{\mathbb{N}}, \]

allows the formulation of the convergence of a sequence as an extension problem.

**Theorem 15.4.** The sequence, \((x_n)_{n \in \mathbb{N}}\), in the topological space \((X, T)\) converges in \(X\) if and only if the (continuous) function

\[ s: \mathbb{N} \rightarrow X, \quad n \mapsto x_n \]

can be extended to a continuous function

\[ \widehat{s}: \widehat{\mathbb{N}} \rightarrow X. \]

More precisely, \((x_n)_{n \in \mathbb{N}}\), in the topological space \((X, T)\) converges in \(X\) if and only if there is a continuous function \(\widehat{s}: \widehat{\mathbb{N}} \rightarrow X\) with \(s = \widehat{s} \circ \alpha\).

This is expressed by the commutative diagram

\[
\begin{array}{ccc}
\mathbb{N} & \xrightarrow{s} & X \\
\alpha \downarrow & & \downarrow \exists \widehat{s} \\
\widehat{\mathbb{N}} & & \\
\end{array}
\]

**Proof.** Let \((x_n)_{n \in \mathbb{N}}\) be a sequence in \((X, T)\) and consider the function

\[ s: \widehat{\mathbb{N}} \rightarrow X, \quad n \mapsto x_n \]

Suppose that there is a continuous function

\[ \widehat{s}: \widehat{\mathbb{N}} \rightarrow X \]

with \(s = \widehat{s} \circ \alpha\).

We show that

\[ \lim_{n \to \infty} x_n = \widehat{s}(\infty) \]

Put \(a := \widehat{s}(\infty)\).

Let \(G\) be an open subset of \(X\) with \(a \in X\).

Since \(\widehat{s}\) is continuous at \(\infty\), there is a basic open subset, \(H\), of \(\widehat{\mathbb{N}}\) with

\[ \infty \in H \subseteq (\widehat{s})^{-1}(G) \]
By the definition of the topology on \( \hat{N} \),
\[
H = (\mathbb{N} \setminus K) \cup \{\infty\}
\]
with \( K \) a compact subset of \( \mathbb{N} \).
Since \( \mathbb{N} \) is a discrete space, \( K \subset \mathbb{N} \) is compact if and only if it is finite.
Thus, \( \{n \in \mathbb{N} \mid x_n = s(n) \notin G\} \) is finite, whence
\[
\lim_{n \to \infty} x_n = a = \hat{s}(\infty)
\]
For the converse, suppose that
\[
\lim_{n \to \infty} x_n = a.
\]
Define
\[
\hat{s}: \hat{N} \to X
\]
by
\[
\hat{s}(u) = \begin{cases} 
  s(n) & \text{if } u = \alpha(n) \\
  a & \text{if } u = \infty
\end{cases}
\]
It is immediate that
(i) \( \hat{s} \) is a function,
(ii) \( s = \hat{s} \circ \alpha \),
(iii) \( \hat{s} \) is continuous on \( \hat{N} \setminus \{\infty\} \).
It remains only to establish the continuity of \( \hat{s} \) at \( \infty \in \hat{N} \).
Let \( G \) be an open subset of \( X \) with \( a \in G \).
Since \( (x_n)_{n \in \mathbb{N}} \) converges to \( a \), all but finitely many terms of the sequence \( (x_n)_{n \in \mathbb{N}} \) lie in \( G \).
Hence, by the definition of \( \hat{s} \), \( \hat{s}(u) \in G \) for all but finitely many \( u \in \hat{N} \). In other words
\[
(\hat{s})^{-1}(G) = (\mathbb{N} \setminus K) \cup \{\infty\}
\]
with \( K \) a finite — and therefore compact — subset of \( \mathbb{N} \).
Thus \( (\hat{s})^{-1}(G) \) is a basic open subset of \( \hat{N} \), showing that \( \hat{s} \) is continuous at \( \infty \) as well. \( \square \)

**Corollary 15.5.** A continuous function maps convergent sequences to convergent sequences.

**Proof.** Let \( f: (X, \mathcal{T}) \to (Y, \mathcal{U}) \) be continuous.
Let the sequence \( (x_n)_{n \in \mathbb{N}} \) converge in \( X \) to \( a \).
By Theorem 15.4,
\[
s: \mathbb{N} \to X, \quad n \mapsto x_n
\]
has a continuous extension
\[
\hat{s}: \hat{N} \to X
\]
Since the composition of continuous functions is continuous, we obtain the continuous function
\[
f \circ \hat{s}: \hat{N} \to Y
\]
which is plainly an extension of
\[
f \circ s: \mathbb{N} \to Y, \quad n \mapsto (f \circ s)(n) = f(x_n)
\]
Hence, by Theorem 15.4,
\[
\lim_{n \to \infty} f(x_n) = f(\hat{s}(\infty)) = f\left(\lim_{n \to \infty} x_n\right)
\]

\[\square\]

While our generalisation of the limit of a sequence in metric spaces to the limit of a sequence in topological spaces is natural enough, a new phenomenon arises: A sequence can have more than one limit.

Example 15.6. Let \(X\) be an infinite set, endowed with the finite-complement topology, \(\mathcal{T}_{FC}\),
\[
\mathcal{T}_{FC} = \{G \subseteq X \mid G = \emptyset \text{ or } X \setminus G \text{ is finite}\}
\]

In this case the intersection of any two non-empty open subsets of \(X\) must contain all but finitely many elements of \(X\).

Hence, if \((x_n)_{n \in \mathbb{N}}\) is a sequence in \(X\) with infinitely many different elements of \(X\) appearing as terms in the sequence, then, by Definition 15.3, the sequence \((x_n)_{n \in \mathbb{N}}\) converges to every \(a \in X\).

A concrete instance of this is the set of all real numbers, \(\mathbb{R}\), with its Zariski topology, which coincides with the finite-complement topology on \(\mathbb{R}\). Here the sequence \((x_n)_{n \in \mathbb{N}}\) converges to every real number.

Observation 15.7. The space \((X, \mathcal{T}_{FC})\) in Example 15.6 is a \(T_1\) space, but not a Hausdorff \((T_2)\) space. This is not coincidental.

Lemma 15.8. No sequence in a Hausdorff space can converge to more than one point.

Proof. Take \(a \neq b\) in the Hausdorff space \((X, \mathcal{T})\).

Then there are \(G, H \in \mathcal{T}\) with \(a \in G, b \in H\) and \(G \cap H = \emptyset\).

Let \((x_n)_{n \in \mathbb{N}}\) be a sequence in \(X\) which converges to \(a\).

Take \(N \in \mathbb{N}\) with \(x_n \in G\) whenever \(n > N\).

Then \(x_n \notin H\) if \(n > N\).

Since only finitely many terms of the sequence \((x_n)_{n \in \mathbb{N}}\) are elements of \(H\), the sequence \((x_n)_{n \in \mathbb{N}}\) cannot converge to \(b\).

Remark 15.9. The crucial feature of metric spaces which enables us to speak of the limit of a convergent sequence is that fact that metric spaces are Hausdorff spaces. It is not the metric itself, but the topology generated by the metric, which is essential.

Convergent sequences can be used to characterise closed sets in metric spaces.

Theorem 15.10. Let \(A\) be a subset of the metric space \((X, \varrho)\).

Then \(x \in X\) is a point of accumulation of \(A\) if and only if there is an infinite sequence \((a_n)_{n \in \mathbb{N}}\) in \(A\) which converges to \(x\).

Proof. Let \(x \in X\) be the limit of the infinite sequence \((a_n)_{n \in \mathbb{N}}\) of elements of \(A\).

Let \(G\) be an open subset of \(X\) containing \(x\).

Since \(x\) is the limit of the sequence \((a_n)_{n \in \mathbb{N}}\), there is a \(K \in \mathbb{N}\) with \(a_n \in G\) whenever \(n > K\).

Since \((a_n)_{n \in \mathbb{N}}\) is an infinite sequence of elements of \(A\), \(A \cap G\) is an infinite set.

Thus, \((A \cap G) \setminus \{x\} \neq \emptyset\), showing that \(x\) is a point of accumulation of \(A\).

For the converse, let \(x\) be a point of accumulation of \(A\).

Since \(B(x; 1)\) is an open subset of \(X\) containing \(x\), there is an \(a_0 \in (A \cap B(x; 1)) \setminus \{x\}\).
Theorem 15.10 showed that $A'$, the derived set of $A$, comprises precisely the limits of convergent infinite sequences of elements of $A$.

If the convergent sequence $(a_n)_{n \in \mathbb{N}}$ has only finitely many distinct elements, then its limit must be one of these and so is itself an element of $A$, and therefore of $\overline{A}$.

Conversely, $a \in A$ is the limit of the sequence $(a_n)_{n \in \mathbb{N}}$ defined by $a_n := a$ for every $n \in \mathbb{N}$.

Hence $\overline{A} = A \cup A'$ comprises the limits of all sequences in $A$ converging to some element of $X$. □

Many important applications of topology deal with metric spaces, and it is often convenient to work with the metric itself, rather than the topology it induces, even when the properties investigated depend only on the topology, and not the metric itself.

The presence of a metric enables us to use techniques familiar from calculus, and explains the use of these techniques in calculus.

Convergent sequences also characterise continuity of functions between metric spaces, as we next prove.

**Theorem 15.12.** Let $(X, \varrho)$ and $(Y, \sigma)$ be metric spaces.

The function $f : X \to Y$ is continuous at $a \in X$ if and only if $\lim_{n \to \infty} f(x_n) = f(a)$ whenever $\lim_{n \to \infty} x_n = a$.

**Proof.** Let $f$ be continuous at $a \in X$ and $(x_n)_{n \in \mathbb{N}}$ a sequence in $X$ converging to $a$.

Take $\varepsilon > 0$.

Since $f$ is continuous at $a$, there is a $\delta > 0$ with $\sigma(f(x), f(a)) < \varepsilon$ whenever $\varrho(x, a) < \delta$.

Since $\lim_{n \to \infty} x_n = a$, there is an $N \in \mathbb{N}$ with $\varrho(x_n, a) < \delta$ whenever $n > N$.

Hence $\sigma(f(x_n), f(a)) < \varepsilon$ whenever $n > N$.

Thus $\lim_{n \to \infty} f(x_n) = f(a)$ whenever $\lim_{n \to \infty} x_n = a$.

For the converse, suppose that $f$ is not continuous at $a \in X$.

Take $\varepsilon > 0$ such that for every $\delta > 0$ there is an $x \in X$ with $\varrho(x, a) < \delta$ and $\sigma(f(x), f(a)) \geq \varepsilon$.

Given $n \in \mathbb{N}$, $\frac{1}{n+1} > 0$.

Hence, there is an $a_n \in X$ with $\varrho(a_n, a) < \frac{1}{n+1}$ and $\sigma(f(a_n), f(a)) \geq \varepsilon$.

Since $\varrho(a, a_n) < \frac{1}{n+1}$, the sequence $(a_n)_{n \in \mathbb{N}}$ converges to $a$. 

\textbf{Corollary 15.11.} The subset $A$ of the metric space $(X, \varrho)$ is closed if and only if the limit of every convergent sequence of elements of $A$ is also in $A$.

Equivalently, $x \in \overline{A}$ if and only if $x$ is the limit of a sequence of elements of $A$. 

\textbf{Proof.} Theorem 15.10 showed that $A'$, the derived set of $A$, comprises precisely the limits of convergent infinite sequences of elements of $A$.

If the convergent sequence $(a_n)_{n \in \mathbb{N}}$ has only finitely many distinct elements, then its limit must be one of these and so is itself an element of $A$, and therefore of $\overline{A}$.

Conversely, $a \in A$ is the limit of the sequence $(a_n)_{n \in \mathbb{N}}$ defined by $a_n := a$ for every $n \in \mathbb{N}$.

Hence $\overline{A} = A \cup A'$ comprises the limits of all sequences in $A$ converging to some element of $X$. □
Since $\sigma(f(a), f(a_n)) > \varepsilon$ for every $n \in \mathbb{N}$, the sequence $(f(a_n))_{n \in \mathbb{N}}$ cannot converge to $f(a)$. □

Remark 15.13. We summarise the above:

A function between metric spaces is continuous if and only if it maps convergent sequences to convergent sequences.

Given the importance of convergent sequences to the theory of metric spaces and continuous functions, we investigate them further, beginning with a necessary condition for a sequence in a metric space to converge, namely Cauchy’s condition.

Definition 15.14. The sequence $(x_n)_{n \in \mathbb{N}}$ in the metric space $(X, \rho)$ satisfies Cauchy’s criterion or is a Cauchy sequence if and only if given $\varepsilon > 0$ there is a $K \in \mathbb{N}$ with $\rho(x_n, x_m) < \varepsilon$ for all $m, n > K$.

Theorem 15.15. Every convergent sequence in a metric space is a Cauchy sequence.

Proof. Let $(x_n)_{n \in \mathbb{N}}$ be a sequence in the metric space $(X, \rho)$ converging to $a \in X$. Given $\varepsilon > 0$ take $K \in \mathbb{N}$ such that $\rho(x_n, a) < \frac{\varepsilon}{2}$ whenever $n > K$.

For $m, n > K$

$$\rho(x_m, x_n) \leq \rho(x_m, a) + \rho(a, x_n)$$

$$\leq \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon.$$ □

The converse of this theorem is not true.

Example 15.16. We take as our metric space the interval $[0, 1]$ together with its Euclidean metric. Then $(\frac{1}{n+1})_{n \in \mathbb{N}}$ is a Cauchy sequence in $[0, 1]$ which does not converge.

Definition 15.17. The metric space $(X, \rho)$ is (Cauchy) complete if and only if every Cauchy sequence in $X$ converges in $X$.

Example 15.18. Every discrete metric space is complete, for a sequence in a discrete metric space is a Cauchy sequence if and only if it is eventually constant.

To see this, let $(x_n)_{n \in \mathbb{N}}$ be a Cauchy sequence in the discrete metric space $(X, d)$ and take $\varepsilon := \frac{1}{2}$. Since $(x_n)_{n \in \mathbb{N}}$ is a Cauchy sequence, there is a $K \in \mathbb{N}$ with $d(x_n, x_m) < \frac{1}{2}$ whenever $m, n > K$.

Since $d$ is the discrete metric, $d(x_n, x_m) < \frac{1}{2}$ if and only if $d(x_n, x_m) = 0$, that is if $x = x_m$. Thus $x = x_n$, whenever $m, n > K$.

Example 15.19. The set of real numbers, $\mathbb{R}$, is complete with respect to its Euclidean metric. This follows from the fact that every bounded set of real numbers has both a supremum (least upper bound) and an infimum (greatest lower bound).²

Let $(x_n)_{n \in \mathbb{N}}$ be a Cauchy sequence of real numbers.

Choose $K \in \mathbb{N}$ with $|x_m - x_n| < 1$ whenever $m, n \geq N$.

Put $M := \max\{|x_0|, \ldots, |x_N|\}$.

Then $|x_n| < M + 1$ for every $n \in \mathbb{N}$, which shows that $(x_n)_{n \in \mathbb{N}}$ is bounded.

²This property of the real numbers is established in the appendix, where the real numbers are constructed from the rational numbers by means of Dedekind cuts.
Given $n \in \mathbb{N}$, define
\[ u_n := \sup\{x_k \mid k \geq n\} \]
\[ \ell_n := \inf\{x_k \mid k \geq n\} \]
Since every non-empty subset of a bounded set is bounded, $u_n$ and $\ell_n$ are well-defined, and
\[ \ell_n \leq \ell_{n+1} \leq u_{n+1} \leq u_n \]
Thus, \((\ell_n)_{n \in \mathbb{N}}\) is a monotonic non-decreasing sequence of real numbers bounded above (by \(u_1\)) and \((u_n)_{n \in \mathbb{N}}\) is a monotonic non-increasing sequence of real numbers bounded below (by \(\ell_1\)).
Hence \((\ell_n)_{n \in \mathbb{N}}\) has a supremum, \(\ell\), and \((u_n)_{n \in \mathbb{N}}\) has an infimum, \(u\).
Plainly, \(\ell \leq u\).
Take \(\varepsilon > 0\).
There is a \(K \in \mathbb{N}\) with \(|\ell - \ell_n| < \frac{\varepsilon}{5}\), \(|u - u_n| < \frac{\varepsilon}{5}\) and \(|x_j - x_k| < \frac{\varepsilon}{5}\) for \(j, k, n \geq K\).
Take \(n > K\).
By the definition of \(\ell_n\) and \(u_n\), there are \(j, k > n\) with \(|\ell_n - x_k| < \frac{\varepsilon}{5}\) and \(|u_n - x_j| < \frac{\varepsilon}{5}\). Thus
\[
|u - \ell| = |u - u_n + u_n - x_j + x_j - x_k + x_k - \ell_n + \ell_n - \ell| \\
\leq |u - u_n| + |u_n - x_j| + |x_j - x_k| + |x_k - \ell_n| + |\ell_n - \ell| \\
< \frac{\varepsilon}{5} + \varepsilon + \frac{\varepsilon}{5} + \frac{\varepsilon}{5} + \frac{\varepsilon}{5} \\
= \varepsilon.
\]
Since \(|u - \ell| < \varepsilon\) for every \(\varepsilon > 0\), \(u = \ell\).
Finally, we show that \(u(= \ell)\) is the limit of the Cauchy sequence \((x_n)_{n \in \mathbb{N}}\).
Take \(\varepsilon > 0\).
There is a \(K \in \mathbb{N}\) with \(|\ell - \ell_K| < \varepsilon\) and \(|u - u_K| < \varepsilon\).
By the properties of the sequences \((u_n)_{n \in \mathbb{N}}\) and \((\ell_n)_{n \in \mathbb{N}}\),
\[
\ell - \varepsilon < \ell_K \leq u_K < u + \varepsilon
\]
and \(\ell_K \leq x_n \leq u_K\) whenever \(n \geq K\).
Given \(n > K\), \(\ell - \varepsilon < x_n < u + \varepsilon\), whence \(|u - x_n| < \varepsilon\) as \(\ell = u\).
Completeness is a metric, but not a topological property, for a metric space homeomorphic with a complete metric space need not be complete.

**Example 15.20.** Take the interval \(X := ]-1, 1[\) with the Euclidean metric.
\[
f: X \longrightarrow \mathbb{R}, \quad x \longmapsto \tan\frac{x\pi}{2}
\]
is a homeomorphism from \(X\) to the complete metric space \(\mathbb{R}\).
\(X\) is not complete because the Cauchy sequence \((\frac{n}{n+1})_{n \in \mathbb{N}}\) does not converge in \(X\).

We use Corollary 15.11 to characterise complete subspaces of a complete metric space.

**Theorem 15.21.** A subspace of a complete metric space is complete if and only if it is closed.

**Proof.** Let \(A\) be a subset of the complete metric space \((X, g)\).
Since a Cauchy sequence in \(A\) is a Cauchy sequence in \(X\), every Cauchy sequence in \(A\) converges.
By Corollary 15.11, the limit of a sequence in \(A\) lies in the closure of \(A\).
We have seen that every convergent sequence is a Cauchy sequence, and that continuous functions map convergent sequences to convergent sequences.

But continuous functions need not map Cauchy sequences to Cauchy sequences.

**Example 15.22.** Take \( \mathbb{R} = \{ x \in \mathbb{R} \mid x > 0 \} \) with its Euclidean metric.

\[
  f: \mathbb{R}^+ \to \mathbb{R}^+, \quad x \mapsto \frac{1}{x}
\]

is continuous and

\[
  \left( \frac{1}{n+1} \right)_{n \in \mathbb{N}}
\]

is a Cauchy sequence, but

\[
  (f(x_n))_{n \in \mathbb{N}} = (n + 1)_{n \in \mathbb{N}}
\]

is not a Cauchy sequence.

However, there is a stronger form of continuity for functions between metric space, such that if a map has this stronger continuity property, it does map Cauchy sequences to Cauchy sequences.

**Definition 15.23.** Given metric spaces \((X, \rho)\) and \((Y, \sigma)\) the function \( f: X \to Y \) is **uniformly continuous** if and only if given \( \varepsilon > 0 \), there is a \( \delta > 0 \) with \( \sigma(f(u), f(v)) < \varepsilon \) whenever \( \rho(u, v) < \delta \).

**Lemma 15.24.**

(a) Every uniformly continuous function is continuous.

(b) Every metric preserving map is uniformly continuous.

(c) The composite of two uniformly continuous functions is uniformly continuous.

**Proof.**

(a) is immediate from the definition of continuity.

To prove (b), recall that \( f: (X, \rho) \to (Y, \sigma) \) preserves the metric if and only if for all \( u, v \in X \)

\[
  \sigma(f(u), f(v)) = \rho(u, v)
\]

Hence, if for \( \varepsilon > 0 \), we put \( \delta := \varepsilon \), then \( \sigma(f(u), f(v)) < \varepsilon \) if and only if \( \rho(u, v) < \delta \).

The proof of (c) is left to the reader as an exercise. \( \square \)

**Theorem 15.25.** A uniformly continuous function maps Cauchy sequences to Cauchy sequences.

**Proof.** Exercise. \( \square \)

The converse of Theorem 15.25 is not true, as the next example shows.

**Example 15.26.** Consider the function

\[
  f: \mathbb{R} \to \mathbb{R}, \quad x \mapsto x^2
\]

It is elementary that \( f \) is continuous.

Take \( \delta > 0 \).

Given \( a \in \mathbb{R} \), put \( b = a + \frac{\delta}{2} \).

Plainly, \( |b - a| = \frac{\delta}{2} < \delta \).

\[
  |f(b) - f(a)| = |(a + \frac{\delta}{2})^2 - a^2| = |a\delta + \frac{\delta^2}{4}| \to \infty \text{ as } a \to \infty
\]

Hence, for \( \varepsilon > 0 \), given any \( \delta > 0 \), there is a \( K \in \mathbb{R} \) such that for all \( a > K \),

\[
  |f(a + \frac{\delta}{2}) - f(a)| > \varepsilon
\]

Thus, \( f \) is not uniformly continuous.
Let \((x_n)_{n \in \mathbb{N}}\) be a Cauchy sequence in \(\mathbb{R}\).

Since \(\mathbb{R}\) is a complete metric space, \((x_n)_{n \in \mathbb{N}}\) converges to some \(a \in \mathbb{R}\).

Since \(f\) is continuous at \(a\), it follows from Theorem ?? that
\[
\lim_{n \to \infty} f(x_n) = f(a)
\]
Thus, \((f(x_n))_{n \in \mathbb{N}}\) converges.

By Theorem 15.15, \((f(x_n))_{n \in \mathbb{N}}\) is a Cauchy sequence.

Hence, \(f\) maps Cauchy sequences to Cauchy sequences.

Observation 15.27. While a function which maps Cauchy sequences to Cauchy sequences need not be uniformly continuous, it must, nevertheless, be continuous.

The proof is left to the reader as an exercise.

1. A Fixed-Point Theorem

Our next theorem, due to Banach, illustrates the significance of completeness. It states that a **contracting self-map**, \(f\), of a complete metric space has a unique fixed point, that is to say the equation \(f(x) = x\) has a unique solution. This theorem has many important applications, including to the theory of (partial) differential equations.

But first we define the notion of a **contracting self-map**. As this is a special case of an important, more general concept, we introduce the general concept first.

**Definition 15.28.** Given metric spaces \((X, \rho)\) and \((Y, \sigma)\), the function \(f : X \to Y\)

is a **Lipschitz function** if and only if there is a \(K > 0\) such that for all \(u, v \in X\)
\[
\sigma(f(u), f(v)) \leq K \rho(u, v)
\]
Such a \(K\) is a **Lipschitz constant** for \(f\).

The self-map \(f : X \to X\), of the metric space \((X, \rho)\) is a **contracting map** if and only if is a Lipschitz function for a Lipschitz constant less than 1, or equivalently, there is a real number, \(K\), with \(0 < K < 1\) such that for all \(u, v \in X\)
\[
\rho(f(u), f(v)) \leq K \rho(u, v).
\]

**Lemma 15.29.** Every Lipschitz function is uniformly continuous.

**Proof.** Let \(K > 0\) be a Lipschitz constant for the function
\[f : (X, \rho) \to (Y, \sigma)\]
Given \(\varepsilon > 0\), put \(\delta := \frac{\varepsilon}{K}\).

Take \(u, v \in X\) with \(\rho(u, v) < \delta\). Then
\[
\rho(f(u), f(v)) \leq K \rho(x, u) < K \delta = \varepsilon.
\]

\(\Box\)

**Theorem 15.30 (Banach Fixed-Point Theorem).** Every contracting map of a complete metric space has a unique fixed point.

In other words, given a contracting map, \(f : X \to X\), of the complete metric space \((X, \rho)\), the equation \(f(x) = x\) has a unique solution.
Proof. We first show that $f$ cannot have more than one fixed point.

If $f(a) = a$ and $f(b) = b$, then
\[
\varrho(a, b) = \varrho(f(a), f(b)) \leq K\varrho(a, b)
\]
by assumption

Thus $(1 - K)\varrho(a, b) \leq 0$.

Since $1 - K > 0$, $\varrho(a, b) \leq 0$.

Hence $\varrho(a, b) = 0$, and so $a = b$.

It remains to show that $f$ has at least one fixed point.

Take any $x \in X$ and define the sequence $(x_n)_{n \in \mathbb{N}}$ inductively by putting
\[
x_0 := x, \quad x_{n+1} := f(x_n) \quad \text{for } n \in \mathbb{N}
\]

We show, by induction, that for every $r \in \mathbb{N}$,
\[
\varrho(x_{r+1}, x_r) \leq K^r \varrho(x_1, x_0)
\]
(*)

The statement is trivially true for $r = 0$, as $K^0 = 1$.

Suppose that for some $r \in \mathbb{N}$, $\varrho(x_{r+1}, x_r) \leq K^r \varrho(x_1, x_0)$. Then
\[
\varrho(x_{r+2}, x_{r+1}) = \varrho(f(x_{r+1}), f(x_r)) \leq K \varrho(x_{r+1}, x_r)
\]
by the definition of $(x_n)_{n \in \mathbb{N}}$
\[
\leq K.K^r \varrho(x_1, x_0)
\]
by the inductive hypothesis
\[
= K^{r+1} \varrho(x_1, x_0)
\]
Further, for $n, r \in \mathbb{N}$, with $r \neq 0$,
\[
\varrho(x_{n+r}, x_n) \leq \sum_{j=0}^{r-1} \varrho(x_{n+j+1}, x_{n+j})
\]
\[
\leq \sum_{j=0}^{r-1} K^{n+j} \varrho(x_1, x_0) \quad \text{by (*)}
\]
\[
= \frac{1 - K^r}{1 - K} K^n \varrho(x_1, x_0)
\]
\[
< K^n \frac{\varrho(x_1, x_0)}{1 - K}
\]
\[
\rightarrow 0 \quad \text{as } n \rightarrow \infty \quad \text{as } 0 < K^r < 1
\]

Hence, given $\varepsilon > 0$, there is an $N \in \mathbb{N}$ such that for all $n \geq N$,
\[
\frac{K^n \varrho(x_1, x_0)}{1 - K} < \varepsilon
\]

In particular, given $m > n \geq N$, put $r = m - n$. Then
\[
\varrho(x_m, x_n) = \varrho(x_{n+r}, x_n)
\]
\[
< K^n \frac{\varrho(x_1, x_0)}{1 - K}
\]
\[
< \varepsilon
\]

Thus, $(x_n)_{n \in \mathbb{N}}$ is a Cauchy sequence in $X$.

Since $(X, \varrho)$ is complete, $(x_n)_{n \in \mathbb{N}}$ converges to some $a \in X$. 
But then
\[
\begin{align*}
  f(a) &= f(\lim_{n \to \infty} x_n) \\
  &= \lim_{n \to \infty} f(x_n) \\
  &= \lim_{n \to \infty} x_{n+1} \\
  &= a
\end{align*}
\]
by the choice of \(a\), since \(f\) is continuous by the definition of \((x_n)_{n \in \mathbb{N}}\), by the choice of \(a\).

Hence, \(a\) is a fixed point of \(f\). \(\square\)

### 1.1. Newton’s Method

An application of the Banach Fixed Theorem is Newton’s method for finding zeroes of functions, \(\varphi: I \rightarrow \mathbb{R}\), where \(I \subseteq \mathbb{R}\) is an interval.

The idea is to approximate the zero, with, say \(x_0\), and to improve the approximation by replacing \(x_0\) with the point of intersection of the tangent at \((x_0, \varphi(x_0))\) to the graph of \(\varphi\) and the \(x\)-axis, and then to iterate this. We depict the \(n^{th}\) stage diagrammatically.

Since the equation of the tangent at \((x_n, \varphi(x_n))\) is
\[
y - \varphi(x_n) = \varphi'(x_n)(x - x_n),
\]
its intersection with the \(x\)-axis has co-ordinates
\[
\left( x_n - \frac{\varphi(x_n)}{\varphi'(x_n)}, 0 \right)
\]
(as long as \(\varphi'(x_n) \neq 0\)), we see that
\[
x_{n+1} = x_n - \frac{\varphi(x_n)}{\varphi'(x_n)}
\]

With this preparation, we show how to apply the Banach Fixed-Point Theorem.

Let \(I \subseteq \mathbb{R}\) be an interval.

Let the function \(\varphi: I \rightarrow \mathbb{R}\) have continuous second derivative, \(\varphi''\).

Take \(z \in I\) such that \(\varphi(z) = 0\) and \(\varphi'(z) \neq 0\).

Since \(\varphi\) is twice differentiable, \(\varphi'\) is continuous.

Hence, there is an open interval, \(J\), with \(x \in J \subseteq I\) such that for all \(x \in J\), \(\varphi'(x) \neq 0\).

Define
\[
\psi: J \rightarrow \mathbb{R}, \quad x \mapsto x - \frac{\varphi(x)}{\varphi'(x)}
\]
Plainly,
\[ \psi(z) = z \]
\[ \psi'(x) = 1 - \frac{(\phi'(x))^2 - \phi''(x)\phi(x)}{(\phi'(x))^2} = \frac{\phi''(x)\phi(x)}{(\phi'(x))^2} \]
whence \( \psi' \) is continuous and
\[ \psi'(z) = 0 \]
By the continuity of \( \psi' \), there is an \( r > 0 \) such that \([z - r, z + r] \subseteq J \) and
\[ |\psi'(x)| \leq \frac{1}{2} \]
whenever \( |x - z| < r \).
By the Mean Value Theorem of Differential Calculus, there is a \( y \) between \( x \) and \( z \) with
\[ |\psi(x) - z| = |\psi(x) - \psi(z)| = |\psi'(y)||x - z| \leq \frac{1}{2}r < r \]
so that
\[ \psi([z - r, z + r]) \subseteq [z - r, z + r] \]
Hence, putting \( X = [z - r, z + r] \), we have the function
\[ f: X \rightarrow X, \quad x \mapsto \psi(x) \]
Take \( x, y \in X \).
By the Mean Value Theorem of Differential Calculus, there is a \( w \) between \( x \) and \( y \) with
\[ |f(x) - f(y)| = |f'(w)||x - y| \leq \frac{1}{2}|x - y| \]
showing that \( f \) is a contraction map.
Since \( X \) is a closed subset of the complete metric space \( \mathbb{R} \), it is complete by Theorem 15.21.
Hence, by the Banach Fixed-Point Theorem, \( f \) has a unique fixed point (which must be \( z \)) and, for any \( x_0 \in X \), the sequence \( (x_n)_{n \in \mathbb{N}} \) with
\[ x_{n+1} = x_n - \frac{\phi(x_n)}{\phi'(x_n)} \]
converges to \( z \).

2. Exercises

15.1. Show that the composite of uniformly continuous functions is uniformly continuous.

15.2. Let \( (X, \varrho) \) and \( (Y, \sigma) \) be metric spaces.
(a) Show that the image of a Cauchy sequence in \( X \) under a uniformly continuous function \( f: X \rightarrow Y \) is a Cauchy sequence in \( Y \).
(b) Show that if \( f \) maps Cauchy sequences in \( X \) to Cauchy sequences in \( Y \), then \( f \) is continuous.
15.3. This exercise should be done by all those who intend to study functional analysis. It consists of showing that every metric space is either complete or can be identified with a dense subset of a complete metric space which is then called its completion.

Given the metric space \((X, \varrho)\), let \(\mathcal{B}(X) := \{f : X \to \mathbb{R} \mid f \text{ is bounded}\}\) be the set of bounded real valued functions defined on \(X\). Define

\[
\varrho_{\infty} : \mathcal{B}(X) \times \mathcal{B}(X) \to \mathbb{R}, \quad (f, g) \mapsto \sup_{x \in X} |f(x) - g(x)|
\]

(a) Show that \(\varrho_{\infty}\) is a metric on \(\mathcal{B}(X)\).

(b) Show that \((\mathcal{B}(X), \varrho_{\infty})\) is a complete metric space.

Fix \(\xi \in X\). Define

\[
i : X \to \mathcal{B}(X), \quad x \mapsto (f_x : X \to \mathbb{R})
\]

where

\[
f_x : X \to \mathbb{R}, \quad y \mapsto \varrho(x, y) - \varrho(y, \xi).
\]

(c) Show that \(i\) preserves the metric.

Let \(\hat{X}\) be the closure of \(i(X)\) in \(\mathcal{B}(X)\) and \(\hat{\varrho}\) the restriction of \(\varrho_{\infty}\) to \(\hat{X}\).

The metric space \((\hat{X}, \hat{\varrho})\) is the completion of \((X, \varrho)\).

(d) Show that given any complete metric space \((Y, \sigma)\) and any uniformly continuous function \(g : X \to Y\), there is a unique uniformly continuous function \(\hat{g} : \hat{X} \to Y\) with \(g = \hat{g} \circ i\).

(e) Show that if \(X\) occurs as a subspace of a complete metric space, then there is a natural isometry between \(X\) and \(\hat{X}\) and that hence there is a natural isometry between any complete metric space which contains \(X\) as a dense subset and \(\hat{X}\).

(f) Show that the completion of \((X, \varrho)\) is uniquely determined up to unique isometry.

(g) Show that \(\mathbb{R}\) is the completion of \(\mathbb{Q}\).
Chapter 16

Compact Metric Spaces

The topology used in functional analysis and differential geometry, and hence in theoretical physics, often appeals to properties specific to metric spaces: Banach spaces, Hilbert spaces and Riemannian manifolds are all metric spaces.

Moreover, the presence of a metric allows for more specialised results and methods, and more computation.

Compact metric spaces play a distinguished rôle, and there are several different characterisations of compactness for metric spaces that have proved to be convenient and useful. We investigate these here.

The main theorem we prove is that a metric space is compact if and only if it is complete and totally bounded. The proof we present makes use of several auxiliary notions in addition to the definition of total boundedness. We introduce these concepts as needed.

Until now we have used the notion of boundedness only for subsets of \( \mathbb{R}^n \): the subset \( A \) of \( \mathbb{R}^n \) is bounded if and only if it is contained in some open ball of finite diameter.

This suggests notion of boundedness to arbitrary metric spaces.

**Definition 16.1.** The subset \( A \) of the metric space \( (X, \varrho) \) is **bounded** if and only if there are an \( a \in X \) and an \( r > 0 \) such that \( A \subseteq B(a; r) \), or equivalently,

\[
\{ \varrho(x, y) \mid x, y \in A \}
\]

is a bounded subset of \( \mathbb{R} \).

If \( A \) is not empty, the **diameter of \( A \)**, \( \text{diam}(A) \), is defined by

\[
\text{diam}(A) := \sup \{ \varrho(x, y) \mid x, y \in A \}.
\]

If \( A \) is not bounded, we sometimes say that “\( A \) has infinite diameter” and write \( \text{diam}(A) = \infty \).

Our investigation of compactness for metric spaces commences with a related notion for sequences, and an associated property.

**Definition 16.2.** The metric space \( (X, \varrho) \) is **sequentially compact** if and only if each sequence, \( (x_n)_{n \in \mathbb{N}} \), in \( X \) has a convergent subsequence \( (x_{n_k})_{k \in \mathbb{N}} \).

The metric space \( (X, \varrho) \) has the **Bolzano-Weierstrass property** if and only if every infinite subset of \( X \) has a point of accumulation.

**Theorem 16.3.** The metric space \( (X, \varrho) \) is sequentially compact if and only if it has the Bolzano-Weierstrass property.
16. COMPACT METRIC SPACES

PROOF. Let $A$ be an infinite subset of the sequentially compact metric space $(X, \rho)$.
Then $A$ contains a sequence $(x_n)_{n \in \mathbb{N}}$ of distinct elements.
By hypothesis, $(x_n)_{n \in \mathbb{N}}$ has a subsequence, $(x_{n_k})_{k \in \mathbb{N}}$, converging to some $a \in X$.
Then $a$ is a point of accumulation of $A$, since given any $\varepsilon > 0$, $B(a; \varepsilon)$ contains all but finitely many terms of $(x_{n_k})_{k \in K}$.
For the converse, let the metric space $(X, \rho)$ have the Bolzano-Weierstrass property.
Take a sequence $(x_n)_{n \in \mathbb{N}}$ in $X$.
Put $A = \{x_n \mid n \in \mathbb{N}\}$
If $A$ is a finite set, there are infinitely many $n \in \mathbb{N}$ with $x_n = x$ for some $x \in X$.
Let $\{n_k \mid k \in \mathbb{N}\}$ be the set of all such $n$’s ordered so that $n_k > n_\ell$ if and only if $k > \ell$.
Then $(x_{n_k})_{k \in \mathbb{N}}$ is a convergent subsequence of the sequence $(x_n)_{n \in \mathbb{N}}$.
If $A$ is an infinite set, then, by hypothesis, it has a point of accumulation, $a$.
As $B(a; 1)$ is an open subset of $X$ containing $a$, there is an $x_{n_0} \in A$ with $\rho(x_{n_0}, a) < 1$.
Suppose that for $k \in \mathbb{N}$, we have chosen $x_{n_0}, \ldots, x_{n_k} \in A$ with $\rho(x_{n_k}, a) < \frac{1}{k+1}$.
Put $r_k := \min\{\frac{1}{k+1}, \rho(x_{n_k}, a) \mid n \leq n_k\}$.
Since $B(a; r_k)$ is open and contains $a$, there is an $x_{n_{k+1}} \in A$ with $\rho(x_{n_{k+1}}, a) < r_k \leq \frac{1}{k+1}$.
By the choice of $r_k$, $n_{k+1} > n_k$ and, since $\rho(x_{n_k}, a) < \frac{1}{k+1}$ for all $k \in \mathbb{N}$, $\lim_{k \to \infty} x_{n_k} = a$. \hfill $\square$

We next show that the Bolzano-Weierstrass property is closely related to compactness.

THEOREM 16.4. Every compact metric space has the Bolzano-Weierstrass property.

PROOF. Let $(X, \rho)$ be a compact metric space.
Let $A \subseteq X$ have no point of accumulation.
Then, given $x \in X$ there is an $\varepsilon_x > 0$ with
$A \cap B(x; \varepsilon_x) \subseteq \{x\}$
Plainly, $\{B(x; \varepsilon_x) \mid x \in X\}$ is an open covering of $X$.
Since $(X, \rho)$ is compact, there is a finite set $\{x_1, \ldots, x_n\}$ with $X = \bigcup B(x_j; \varepsilon_{x_j})$.
By construction, $A \subseteq \{x_1, \ldots, x_n\}$.
Hence, $A$ must be finite. \hfill $\square$

COROLLARY 16.5. Every compact metric space is sequentially compact.

Our aim is to prove the converse of this lemma.
To do so, we first investigate further properties of open coverings in metric spaces.

Let $\{G_\lambda \mid \lambda \in \Lambda\}$ be an open covering of the metric space $(X, \rho)$.
For each $x \in X$ there is a $\lambda_x \in \Lambda$ with $x \in G_{\lambda_x}$.
This being an open subset of $X$, there is an $\varepsilon_x$ with $B(x; \varepsilon_x) \subseteq G_{\lambda_x}$.
Then $\{B(x; \varepsilon_x) \mid x \in X\}$ is also an open covering of $X$.
Put $\varepsilon := \inf \{\varepsilon_x \mid x \in X\}$.
Clearly, $\varepsilon \geq 0$.
Suppose that $\varepsilon > 0$, and take $r \in ]0, \varepsilon[$.
Let $A$ be a subset of $X$ with $\text{diam}(A) < r$, and take $a \in A$. 

Then there is a $\lambda_A \in \Lambda$ with $a \in B(a; \varepsilon_a)$.

For $b \in A$, $g(b, a) < \varepsilon_a$, whence

$$b \in B(a; r) \subseteq B(a; \varepsilon_a)$$

as $0 < r < \inf\{\varepsilon_x \mid x \in X\}$

showing that $A \subseteq G_{\lambda_a}$.

In other words, any subset of $X$ which is “small enough” is contained within one of the sets of the given open covering. This leads to our next definition.

**Definition 16.6.** Let $G$ be a covering of the metric space $(X, \varrho)$.

The positive real number, $\ell$, is a *Lebesgue number* for $G$ if and only if given any subset $A$ of $X$ with $\text{diam}(A) < \ell$, there is a $G \in G$ with $A \subseteq G$.

**Theorem 16.7 (Lebesgue Covering Lemma).** Every open covering of a sequentially compact metric space has a Lebesgue number.

**Proof.** Take an open covering, $\{G_{\lambda} \mid \lambda \in \Lambda\}$, of the sequentially compact space $(X, \varrho)$.

Let $\mathcal{A}$ be the set of all those subsets of $X$ which are not subsets of any of the $G_{\lambda}$'s, so that

$$\mathcal{A} := \{A \subseteq X \mid A \setminus G_{\lambda} \neq \emptyset \text{ for every } \lambda \in \Lambda\}.$$ 

If no $A \in \mathcal{A}$ has finite diameter, then every positive number is a Lebesgue number for $\{G_{\lambda} \mid \lambda \in \Lambda\}$.

Suppose that at least one $A \in \mathcal{A}$ has finite diameter.

Since the diameter of a set cannot be negative, $\{\text{diam}(A) \mid A \in \mathcal{A}\}$ has finite diameter is a non-empty set of real numbers bounded below by 0.

Hence, $\alpha := \inf\{\text{diam}(A) \mid A \in \mathcal{A}\}$ is a well-defined non-negative real number.

If $\alpha > 0$, then, by the discussion preceding Definition 16.6, any $\ell \in \]0, \alpha[$ is a Lebesgue number for $\{G_{\lambda} \mid \lambda \in \Lambda\}$.

If $\alpha = 0$, then for each $n \in \mathbb{N}$ there is an $A_n \in \mathcal{A}$ with $0 < \text{diam}(A_n) < \frac{1}{n+1}$.

For each $n \in \mathbb{N}$, choose $x_n \in A_n$ and consider the sequence $(x_n)_{n \in \mathbb{N}}$.

Since $X$ is sequentially compact, $(x_n)_{n \in \mathbb{N}}$ has a convergent subsequence $(x_{n_k})_{k \in \mathbb{N}}$.

Put $x := \lim_{k \to \infty} x_{n_k}$.

Since $\{G_{\lambda} \mid \lambda \in \Lambda\}$ is a covering of $X$, there is a $\mu \in \Lambda$ with $x \in G_{\mu}$.

Since $G_{\mu}$ is open, there is an $\varepsilon > 0$ with $B(x; \varepsilon) \subseteq G_{\mu}$.

Moreover there is a $K \in \mathbb{N}$ such that $g(x, x_{n_k}) < \frac{\varepsilon}{2}$ and $\text{diam}(A_{n_k}) < \frac{\varepsilon}{2}$ whenever $k > K$.

Take $k > K$ and $y \in A_{n_k}$. Then

$$g(x, y) \leq g(x, x_{n_k}) + g(x_{n_k}, y) < \varepsilon$$

so that

$$A_{n_k} \subseteq B(x; \varepsilon) \subseteq G_{\mu},$$

contradicting the definition of $\mathcal{A}$.

Thus $\alpha > 0$. \hfill \Box

We next show that every compact metric space is bounded.

**Lemma 16.8.** Every compact metric space is bounded.

**Proof.** Let $(X, \varrho)$ be a compact metric space.

Clearly, $\{B(x; 1) \mid x \in X\}$ is an open covering of $(X, \varrho)$.
Since $X$ is compact, there are $x_1, \ldots, x_n \in X$ such that $X = \bigcup B(x_j; 1)$.

Put $K := \max\{d(x_i, x_j) : i, j = 1, \ldots, n\}$.

Take $x, y \in X$. There are $i, j \in \{1, \ldots, n\}$ with $x \in B(x_i; 1)$ and $y \in B(x_j; 1)$, whence

$$d(x, y) \leq d(x, x_i) + d(x_i, x_j) + d(x_j, y) \leq K + 2.$$  

Hence $\text{diam}(X) \leq K + 2$. □

The converse of Lemma 16.8 is not true in general.

**Example 16.9.** While $[0, 1]$ (with the Euclidean topology) is clearly bounded, it cannot be compact, since it is not a closed subset of the Hausdorff space $\mathbb{R}$.

A simpler example is provided by any non-empty set $X$ taken with its discrete metric $d_X$, for any $a \in X$

$$X \subseteq N_{d_X}(a; 2)$$

There is a stronger form of boundedness closely related to compactness.

**Definition 16.10.** Let $(X, d)$ be a metric space and take $\varepsilon > 0$.

An $\varepsilon$-net in $X$ is a finite subset, $A$, of $X$ such that $X = \bigcup_{a \in A} B(a; \varepsilon)$.

The metric space $(X, d)$ is **totally bounded** if and only if it has an $\varepsilon$-net for every $\varepsilon > 0$.

**Lemma 16.11.** Every totally bounded metric space is bounded.

**Proof.** A totally bounded metric space has a 1-net.

The argument in the proof of Lemma 16.8 now proves this lemma. □

The converse to Lemma 16.11 is not true in general.

**Example 16.12.** Take $X := \mathbb{R}$ with the metric

$$d : \mathbb{R} \times \mathbb{R} \to \mathbb{R}_0^+, \quad (x, y) \mapsto \frac{|x - y|}{1 + |x - y|}.$$  

Clearly $\text{diam}(X) = 1$, but $X$ admits no $\frac{1}{2}$-net.

For $d(x, y) < \frac{1}{2}$ if and only if $|x - y| < 1$, and so, every $\frac{1}{2}$-net in $(\mathbb{R}, d)$ is a 1-net in $\mathbb{R}$.

But, plainly, $\mathbb{R}$ has no 1-net.

Total boundedness is closely related to compactness, as the following theorem indicates.

**Theorem 16.13.** Every sequentially compact metric space is totally bounded.

**Proof.** Suppose that $(X, d)$ is not totally bounded.

Take $\varepsilon > 0$ for which $X$ has no $\varepsilon$-net.

Then we can choose a sequence $\{a_n\}_{n \in \mathbb{N}}$ in $X$ such that for each $n \in \mathbb{N}$, $a_{n+1} \notin \bigcup_{j=0}^{n} B(a_j; \varepsilon)$.

The sequence $\{a_n\}_{n \in \mathbb{N}}$ cannot have a convergent subsequence, since $d(a_n, a_m) \geq \varepsilon$ whenever $m \neq n$.

Hence $(X, d)$ is not sequentially compact. □

**Theorem 16.14.** Every sequentially compact metric space is compact.
Proof. Let $\ell$ be a Lebesgue number for the open covering, $\{G_\lambda \mid \lambda \in \Lambda\}$, of the sequentially compact metric space $(X, \varrho)$.

Put $\varepsilon := \frac{\ell}{3}$.

Choose an $\varepsilon$-net $A = \{a_1, \ldots, a_n\}$ for $X$.

Then, for each $j \in \{1, \ldots, n\}$,

$$\text{diam}(B(a_j; \varepsilon)) \leq 2\varepsilon \leq \frac{2\ell}{3} < \ell$$

As $\ell$ is a Lebesgue number for the given covering, for each $j \in \{1, \ldots, n\}$ there is a $\lambda_j \in \Lambda$ with $B(a_j; \varepsilon) \subseteq G_{\lambda_j}$. Then

$$X \subseteq \bigcup_{j=1}^{n} B(a_j; \varepsilon) \subseteq \bigcup_{j=1}^{n} G_{\lambda_j} \subseteq X$$

Thus, $\{G_{\lambda_j} \mid j = 1, \ldots, n\}$ is a finite subcover of $X$. 

We have seen that the following statements are equivalent for the metric space $(X, \varrho)$.

1. $(X, \varrho)$ is compact.
2. $(X, \varrho)$ is sequentially compact.
3. $(X, \varrho)$ has the Bolzano-Weierstrass property.

We have also shown that these (equivalent) properties imply that $X$ is totally bounded. We now prove our main theorem, namely, that totally boundedness together with completeness is equivalent to compactness.

**Theorem 16.15.** A metric space is compact if and only if it is complete and totally bounded.

Proof. We have shown that if $(X, \varrho)$ is compact, then it is sequentially compact, so that every sequence has a convergent subsequence.

In particular, any Cauchy sequence has a convergent subsequence.

But a Cauchy sequence converges if and only if it has a convergent subsequence.

Thus $X$ must be complete.

We have also shown that every sequentially compact metric space is totally bounded.

Hence $X$ is complete and totally bounded whenever it is compact.

For the converse, let $X$ be a complete and totally bounded metric space.

Let $(x_n)_{n \in \mathbb{N}}$ be a sequence in $X$, such that $\{x_n \mid n \in \mathbb{N}\}$ an infinite subset of $X$.

Define the sequence $S_1 := (y_{1n})_{n \in \mathbb{N}}$ by

$$y_{1n} := x_n$$

(We have simply changed notation.)

By the total boundedness of $X$, there is a finite set of open balls in $X$, each of radius $\frac{1}{2}$, whose union is $X$.

Since $S_1$ is infinite, at least one of these open balls of radius $\frac{1}{2}$ must contain infinitely many terms of $S_1$. 


In other words, \( S_1 \) has a subsequence \( S_2 = (y_{2n})_{n \in \mathbb{N}} \) all of whose terms lie in some one open sphere of radius \( \frac{1}{2} \).

Appealing again to the total boundedness of \( X \), there is a \( \frac{1}{3} \)-net in \( X \) and so, arguing as above, \( S_2 \) has a subsequence \( S_3 = (y_{3n})_{n \in \mathbb{N}} \) all of whose terms lie within some one open ball of radius \( \frac{1}{3} \).

Iteration now leads to a family of sequences 
\[(y_{mn})_{n \in \mathbb{N}} \] such that for each \( m \in \mathbb{N} \), the sequence \( (y_{(m+1)n})_{n \in \mathbb{N}} \) a subsequence of \( (y_{mn})_{n \in \mathbb{N}} \).

Then \( S := (y_{nn})_{n \in \mathbb{N}} \) is a Cauchy subsequence of \( S_1 \).

Since \( X \) is complete, \( S \) is a convergent subsequence of \( S_1 \).

Thus \( X \) is sequentially compact and hence compact. \( \square \)

Corollary 16.16 (Heine-Borel Theorem). The subset \( A \) of \( \mathbb{R} \) is compact with respect to the Euclidean metric if and only if it is closed and bounded.

Proof. Since \( \mathbb{R} \) is complete, the subset \( A \) is complete if and only if it is closed.

So, we only need to show that a closed subset, \( A \), is totally bounded if and only if it is bounded.

By Lemma 16.11, every totally bounded set is bounded.

Thus, it is enough to that if \( A \neq \emptyset \) is bounded, then it is totally bounded.

Suppose that the non-empty subset \( A \) of \( \mathbb{R} \) is bounded.

Then \( A \subseteq [K, L] \) where \( K = \inf A \) and \( L = \sup A \).

Take \( \varepsilon > 0 \).

Then, by the choice of \( K \), there is an \( a_1 \in A \) with \( K \leq a_1 < K + \varepsilon \).

If \( A \subseteq B(a_1; \varepsilon) \), then we are finished.

Otherwise, suppose that \( a_1, \ldots, a_j \) have been chosen and that
\[ A \setminus (B(a_1; \varepsilon) \cup \ldots \cup B(a_j; \varepsilon)) \neq \emptyset. \]

Put
\[ b_{j+1} := \inf \{ a \in A \mid a \notin \bigcup_{k=1}^{j} B(a_k; \varepsilon) \} \]

There is an \( a_{j+1} \in A \) with \( b_{j+1} \leq a_{j+1} < b_{j+1} + \varepsilon \).

If \( A \subseteq (B(a_1; \varepsilon) \cup \ldots \cup B(a_{j+1}; \varepsilon)) \), we are finished.

But this must be the case after at most \( n \) steps, where \( n \) is the least integer not less than \( \frac{L - K}{\varepsilon} \). \( \square \)

The compactness of a metric space is a topological condition which significantly affects the continuity of functions defined on that space, as shown in the next theorem.

Theorem 16.17. Let \((X, \rho)\) be a compact metric space and \((Y, \sigma)\) any metric space.

The function \( f : X \to Y \) is continuous if and only if it is uniformly continuous.

Proof. It is enough to show if \( f \) is continuous then it is uniformly continuous.

Choose \( \varepsilon > 0 \).

For each \( x \in X \) there is a \( \delta_x \) with \( \sigma(f(x), f(x')) < \frac{\varepsilon}{2} \) whenever \( \rho(x, x') < \delta_x \).

As \( \{B(x, \delta_x) \mid x \in X \} \) is an open covering of the compact metric space \( X \), it has a Lebesgue number, \( \delta > 0 \).
Take \( y, z \in X \) with \( \varrho(y, z) < \delta \).

Then \( \text{diam}\{y, z\} < \delta \) whence there is an \( x \in X \) with \( y, z \in B(x, \delta_x) \). Thus

\[
\sigma(f(y), f(z)) \leq \sigma(f(y), f(x)) + \sigma(f(x), f(z)) \\
< \frac{\varepsilon}{2} + \frac{\varepsilon}{2} \\
= \varepsilon
\]

showing that \( f \) is uniformly continuous. \( \Box \)

1. Exercises

16.1. Take \( \mathbb{R}^n \) is taken with its Euclidean topology.

Prove that \( A \subseteq \mathbb{R}^n \) is totally bounded if and only if it is bounded.

(This shows that a subset of \( \mathbb{R}^n \) is compact if and only if it is closed and bounded.)

16.2. Find an example of a metric space in which not every bounded subset is totally bounded.

16.3. Recall from Exercise 2.5 that given the metric space \((X, \varrho)\) the distance of the point \( x \in X \) from the subset \( A \subseteq X \) is

\[
\varrho(x, A) := \inf\{a \in A \mid \varrho(x, a)\}
\]

Show that if \( A \) is compact, then \( \varrho(x, A) = \min\{a \in A \mid \varrho(x, a)\} \).

Show that if \( B \) is also compact, then \( \{\varrho(x, A) \mid x \in B\} \) is bounded and has a maximum.

For compact subsets \( A, B \subseteq X \), define

\[
d(B, A) := \max\{\varrho(x, A) \mid x \in B\}.
\]

Show that for compact subsets \( A, B \subseteq X \), in general \( d(B, A) \neq d(A, B) \).

Now let \( \mathcal{C}(X) \) denote the set of all compact subsets of \((X, \varrho)\).

Show that

\[
\varrho_h : \mathcal{C}(X) \times \mathcal{C}(X) \rightarrow \mathbb{R}^+_0, \quad (A, B) \mapsto \max\{d(A, B), d(B, A)\}
\]

defines a metric on \( \mathcal{C}(X) \).

(This is the Hausdorff metric, which is important rôle to the theory of dynamical systems and fractals.)
Index

A, 54
0-sphere, 82
\(\mathbb{A}\), 54
0
\(\mathbb{S}\), 82
\(\mathbb{Y}^X\), 5
\(\emptyset\), 3
ext(\(A\)), 56
\(\mathbb{C}\), 3
1, 26
\(\mathbb{N}\), 3
\(\mathbb{Q}\), 3
\(\mathbb{R}\), 3
\(\mathbb{Z}\), 3
2, 4, 82
\(\mathcal{P}(X,Y)\), 5
\(\varepsilon\)-net, 130
id\(X\), 6
p-adic metric, 29
\(\mathcal{P}(A)\), 2
1–1, 9
1-point compactification, 106
Alexandroff compactification, 106
Axiom of Choice, 3, 22, 101
Banach Fixed-Point Theorem, 121
base for a topology, 59
basic open set, 59
bijective, 9
Bolzano-Weierstrass property, 127
boundedness, 127
total, 127
boundedness, total, 130
Cantor set, 102
Cartesian product, 3
Cauchy completeness, 118
Cauchy sequence, 118
Cauchy’s criterion, 118
Cauchy-Schwarz Inequality, 25
characteristic function, 7
closed mapping, 77
closed set, 73
closure, 74
co-domain, 4
coarser topology, 52
commutative diagram, 12
compactification, 105
compactness, 96
local, 105
sequential, 127
completely regular, 94
completeness, 118
order, 137
completion, 125
component, 84
composition (of functions), 7
connected, path, 85
connected component, 84
connectedness, 81
continuity, 23, 27, 49
at a point, 23, 27, 49
uniform, 120
continuity, uniform, 41
contracting map, 121
convergence, 113
covering, 96
finite, 96
open, 96
Dedekind cut, 118, 136
dense subset, 76
derived set, 76
dense subset, 76
disconnected, 81
discrete metric, 26
discrete topology, 51
distance, 23
domain, 4
epi, 9
equivalence class, 13
equivalence relation, 13
Euclidean metric, 26
Euclidean topology, 50
exterior, 56
Extreme Value Theorem, 97
fibre, 5iner topology, 52
Fréchet spaces, 89
function, 4
invertible, 8
Fundamental Theorem of Algebra, 102
generated topology, 57, 60
graph, 5
Hamming metric, 28
Hausdorff metric, 133
Hausdorff space, 89
Heine-Borel property, 96
homeomorphism, 69

identity function, 6
image, 4
inclusion map, 6
indexed family, 3
indexing set, 3
indiscrete topology, 51
induced metric, 33
induced topology, 61
infinite sequence, 113
injective, 9
interior, 54, 75
Intermediate Theorem of Calculus, 83
intersection, 2
interval, 82
inverse, 8
inverse image, 5
iso, 9
isometry, 31

Lebesgue Covering Lemma, 129
Lebesgue number, 129
left inverse, 8
Lipschitz constant, 121
Lipschitz function, 121
local compactness, 105

map, 4
mapping, 4
metric, 25
discrete, 26
product, 40
metric preserving function, 31
metric space, 25
metric topology, 49
metrizability, 52
mono, 9
mono-epi factorisation, 11

net, 130
Newton’s method, 123
norm, 29
normal space, 90
normed vector space, 29

onto, 9
open covering, 96
open ball, 43
open mapping, 54
open set in a metric space, 44
open subset, 44
order completeness, 137

partition, 13
path, 85
path-connected, 85
point of accumulation, 76
power set, 2
pre-image, 5
product metric, 37, 40
product topology, 64, 65
proper subset, 2

quotient function, 13
quotient map, 13
quotient set, 13
quotient topology, 66, 67

reflexive, 13
regular space, 90
restriction, 7
restriction of a metric, 32
Riemann sphere, 108
right inverse, 8

separation, 89
sequence, 113
infinite, 113
sequential compactness, 127
set, 1
sub-base for a topology, 59
sub-basic open set, 59
sub-covering, 96
sub-space of a topological space, 63
sub-space topology, 63
subset, 2
open, 44
proper, 2
subspace of a metric space, 33
subspace topology, 61
surjective, 9
symmetric, 13
systematic ambiguity, 14
topological space, 49
topologists’ sine curve, 86
topology, 49
product, 65
topology
subspace, 61
total boundedness, 127, 130
total disconnectedness, 84
transitive, 13
triangle inequality, 24
Tychonov’s Theorem, 100
uniform continuity, 41, 120
union, 2
universal property, 15
Urysohn’s Lemma, 92
Zariski topology, 52, 79
It has become almost a cliche to remark that nobody boasts of ignorance of literature, but it is socially acceptable to boast ignorance of science and proudly claim incompetence in mathematics.

Richard Dawkins

Appendix A

A Construction of the Real Numbers

We outline the construction of the real numbers, \( \mathbb{R} \), from the rational numbers, \( \mathbb{Q} \), using Dedekind cuts.

1. Dedekind’s Comments

Before presenting the construction, we reproduce Richard Dedekind’s own introduction.

My attention was first directed toward the considerations which form the subject of this pam-
phef in the autumn of 1858. As professor in the Polytechnic School in Zürich I found myself
for the first time obliged to lecture upon the elements of the differential calculus and felt
more keenly than ever before the lack of a really scientific foundation for arithmetic. In dis-
cussing the notion of the approach of a variable magnitude to a fixed limiting value, and
especially in proving the theorem that every magnitude which grows continually, but not
beyond all limits, must certainly approach a limiting value, I had recourse to geometric evi-
dences. Even now such resort to geometric intuition in a first presentation of the differential
calculus, I regard as exceedingly useful, from the didactic standpoint, and indeed indis-
pensable, if one does not wish to lose too much time. But that this form of introduction into
the differential calculus can make no claim to being scientific, no one will deny. For my-
selves this feeling of dissatisfaction was so overpowering that I made the fixed resolve to keep
meditating on the question till I should find a purely arithmetic and perfectly rigorous foun-
dation for the principles of infinitesimal analysis. The statement is so frequently made that the
differential calculus deals with continuous magnitude, and yet an explanation of this conti-
nuity is nowhere given; even the most rigorous expositions of the differential calculus do not
base their proofs upon continuity but, with more or less consciousness of the fact, they either
appeal to geometric notions or those suggested by geometry, or depend upon theorems
which are never established in a purely arithmetic manner. Among these, for example, be-
longs the above-mentioned theorem, and a more careful investigation convinced me that
this theorem, or any one equivalent to it, can be regarded in some way as a sufficient basis
for infinitesimal analysis. It then only remained to discover its true origin in the elements of
arithmetic and thus at the same time to secure a real definition of the essence of continu-
ity. I succeeded Nov. 24, 1858, and a few days afterward I communicated the results of my
meditations to my dear friend Durège with whom I had a long and lively discussion. Later
I explained these views of a scientific basis of arithmetic to a few of my pupils, and here in
Braunschweig read a paper upon the subject before the scientific club of professors, but I
could not make up my mind to its publication, because, in the first place, the presentation
did not seem altogether simple, and further, the theory itself had little promise. Nevertheless
I had already half determined to select this theme as subject for this occasion, when a few
days ago, March 14, by the kindness of the author, the paper *Die Elemente der Funktionenlehre* by E. Heine (*Crelle’s Journal*, Vol. 74) came into my hands and confirmed me in my decision. In the main I fully agree with the substance of this memoir, and indeed I could hardly do otherwise, but I will frankly acknowledge that my own presentation seems to me to be simpler in form and to bring out the vital point more clearly. While writing this preface (March 20, 1872), I am just in receipt of the interesting paper *Ueber die Ausdehnung eines Satzes aus der Theorie der trigonometrischen Reihen*, by G. Cantor (*Math. Annalen*, Vol. 5), for which I owe the ingenious author my hearty thanks. As I find on a hasty perusal, the axiom given in Section II. of that paper, aside from the form of presentation, agrees with what I designate in Section III. as the essence of continuity. But what advantage will be gained by even a purely abstract definition of real numbers of a higher type, I am as yet unable to see, conceiving as I do of the domain of real numbers as complete in itself.

2. The Construction

While we assume familiarity with the rational numbers, their arithmetic and order properties, we take this opportunity to summarise these in the form of axioms.

**Properties of** \( \mathbb{Q} \)  
*Let* \( a, b \) and \( c \) *be rational numbers. Then*

(i) \((a + b) + c = a + (b + c)\);
(ii) \(a + b = b + a\);
(iii) \(a + 0 = a\);
(iv) \(a + (-a) = 0\);
(v) \((a.b).c = a.(b.c)\);
(vi) \(a.b = b.a\);
(vii) \(a.1 = a\);
(viii) \(a.(b + c) = a.c + b.c\);
(ix) if \( a \neq 0 \) then there is a \( \frac{1}{a} \in \mathbb{Q} \) with \( a.\frac{1}{a} = 1\);
(x) if \( a \leq b \) and \( c \geq 0 \), then \( a.c \leq b.c\).

While the geometric picture of the real numbers as the points of a straight line is visually appealing, and the ability to perform the four “basic arithmetic operations” geometrically is reassuring, it is also unsatisfying if we wish to construct and discuss mathematics and mathematical entities intrinsically. It was this consideration which prompted Richard Dedekind to construct the real numbers from the rational numbers without appealing to geometric realisations. He completed the task in 1857\(^1\), but did not publish his work until the 1870s.

We present his construction, which uses what are now called Dedekind’s cuts.

**Definition A.1.** The subset \( \alpha \) of \( \mathbb{Q} \) is a (Dedekind) cut if and only if it satisfies the following conditions.

\((D0)\) \( \alpha \neq \emptyset \) and \( \alpha \neq \mathbb{Q}\).
\((D1)\) If \( a, b \in \mathbb{Q} \) with \( a \in \alpha \) and \( a \leq b \), then \( b \in \alpha \).
\((D2)\) \( \alpha \) contains no least element.

**Observation A.2.** If \( \kappa \subseteq \mathbb{Q} \) satisfies \((D0)\) and \((D1)\) but not \((D2)\), then it must have a least element, \( m \).

In this case, \( \kappa \setminus \{m\} \) is a Dedekind cut.

**Definition A.3.** Let \( \kappa \subseteq \mathbb{Q} \) satisfy \((D0)\) and \((D1)\). The Dedekind cut determined by \( \kappa \) is

\[
\kappa^* = \begin{cases} 
\kappa \setminus \{m\} & \text{if } m \text{ is the least element of } \kappa \\
\kappa & \text{if } \kappa \text{ is already a Dedekind cut}
\end{cases}
\]

\(^1\)Dedekind explained that he felt compelled to do this when he was teaching calculus to first year engineering students, in order to give them a proper account of the real numbers.
We define $\mathbb{R}$ to be the set of all cuts and define an order on $\mathbb{R}$.

**Definition A.4.** $\mathbb{R} := \{ \alpha \subset \mathbb{Q} \mid \alpha \text{ is a Dedekind cut} \}$

Given $\alpha, \beta \in \mathbb{R}$, $\alpha \leq \beta$ if and only if $\beta \subseteq \alpha$.

The next lemma establishes a technically useful property of Dedekind cuts.

**Lemma A.5.** Let $\alpha$ be a Dedekind cut.

Given a positive rational number $d > 0$, there is an $a \in \alpha$ with $a - d \notin \alpha$.

**Proof.** By (D0), there is an $a_1 \in \alpha$ and an $a_0 \notin \alpha$.

By (D1), $a_0 < a_1$, so that $a_1 = a_0 + \frac{m}{n}d$ for suitable $m, n \geq 1$.

The set $M := \{ k \in \mathbb{N} \mid a_0 + \frac{k}{n}d \in \alpha \}$ is then a non-empty set of natural numbers since it contains $m$.

Hence, it has a least element, say $k_0$. Clearly $k_0 > 0$.

Put $a := a_0 + \frac{k_0}{n}d$.

Then $a \in \alpha$ and

$$a - \frac{d}{n} = a_0 + \frac{k_0}{n}d - \frac{d}{n} = a_0 + \frac{(k_0 - 1)}{n}d \notin \alpha$$

by the definition of $k_0$.

But $a - d \leq a - \frac{d}{n}$ as $n \geq 1$ and so $a - d \notin \alpha$. \qed

The real numbers are order complete, as shown in the next theorem.

**Theorem A.6.** Every non-empty subset of $\mathbb{R}$, which is bounded below, has an infimum (greatest lower bound).

**Proof.** Let $\mathcal{A}$ be a non-empty subset of $\mathbb{R}$, which is bounded below.

Then there is a $\beta \in \mathbb{R}$ with $\beta \leq \kappa$ for all $\kappa \in \mathcal{A}$, that is, $\kappa \subseteq \beta$ whenever $\kappa \in \mathcal{A}$.

Define $\alpha := \bigcup_{\kappa \in \mathcal{A}} \kappa$.

(D0): Since $\mathcal{A}$ is not empty, it contains at least one cut, say $\kappa_0$.

Then $\kappa_0 \neq \emptyset$ and $\kappa_0 \subseteq \alpha$.

Hence $\alpha \neq \emptyset$.

Moreover, since $\kappa \subseteq \beta$ whenever $\kappa \in \mathcal{A}$,

$$\alpha = \bigcup \{ \kappa \mid \kappa \in \mathcal{A} \} \subseteq \beta$$

But $\beta \neq \mathbb{Q}$ as $\beta$ is a Dedekind cut.

Hence $\alpha \neq \mathbb{Q}$.

(D1): Take $a \in \alpha$ and $b \in \mathbb{Q}$ with $a \leq b$.

Since $a \in \alpha$, $a \in \kappa$ for some $\kappa \in \mathcal{A}$, and since $\kappa$ is a Dedekind cut, $b \in \kappa$.

But then $b \in \alpha$ as $\kappa \subseteq \alpha$.

(D2): Take $a \in \alpha$. 

Then $a \in \kappa$ for some $\kappa \in A$.

Hence, since $\kappa$ is a Dedekind cut and $a \in \kappa$, there is a $y \in \kappa$ with $y < a$.

But then $y \in \alpha$ and $y < a$.

Thus $\alpha$ has no least element.

Thus, $\alpha$ is indeed a cut.

Take $\kappa \in A$. Then $\kappa \subseteq \alpha$ and so $\alpha \leq \kappa$.

Thus $\alpha$ is a lower bound for $A$.

If $\gamma$ is any lower bound for $A$, then $\gamma \leq \kappa$ whenever $\kappa \in A$. Equivalently, $\kappa \subseteq \gamma$ whenever $\kappa \in A$.

Hence, $\alpha$ is the infimum (greatest lower bound) of $A$. $\square$

We next define the arithmetic operations on cuts.

**Definition A.7.** Let $\alpha$ and $\beta$ be Dedekind cuts.

Their sum, $\alpha + \beta$, is

$\alpha + \beta := \{a + b \mid a \in \alpha, b \in \beta\}$.

We verify that $\alpha + \beta$ is in fact a Dedekind cut.

(D0): Since $\alpha, \beta \neq \emptyset$, we can choose $a \in \alpha$ and $b \in \beta$.

Then $a + b \in \alpha + \beta$, so that $\alpha + \beta \neq \emptyset$.

Since $\alpha, \beta \neq \mathbb{Q}$, there are rational numbers $c \notin \alpha$ and $d \notin \beta$.

Hence, given any $a \in \alpha$ and $b \in \beta$, $c < a$ and $d < b$.

Hence $c + d < a + b$ for all $a \in \alpha$ and $b \in \beta$. Hence $c + d \in \mathbb{Q} \setminus (\alpha + \beta)$.

(D1): Take $c \in \alpha + \beta$ and $d \in \mathbb{Q}$ with $c \leq d$.

Then $c = a + b$ for suitable $a \in \alpha$ and $b \in \beta$, so that $a + b < d$.

Thus, $a < d - b$, and hence $d - b \in \alpha$.

But then $d = (d - b) + b \in \alpha + \beta$.

(D2): Take $c \in \alpha + \beta$.

Then $c = a + b$ for some $a \in \alpha$ and $b \in \beta$.

But then there is an $a' \in A$ with $a' < a$.

Putting $c' := a' + b$, we see that $c' \in \alpha + \beta$ and $c' < c$.

The properties of rational numbers listed above imply that for $\alpha, \beta, \gamma \in \mathbb{R}$

$(\alpha + \beta) + \gamma = \alpha + (\beta + \gamma)$

$\beta + \alpha = \alpha + \beta$

**Definition A.8.** Given $q \in \mathbb{Q}$, put $\sigma_q := \{a \in \mathbb{Q} \mid a > q,\}$.

It is easy to see that

$i : \mathbb{Q} \to \mathbb{R}, \quad q \mapsto \sigma_q$

imbeds $\mathbb{Q}$ into $\mathbb{R}$, and that $\sigma_a \leq \sigma_b$ if and only if $a \leq b$.

We identify $\mathbb{Q}$ with $i(\mathbb{Q})$ — that is $q$ with $\sigma_q$ — whenever convenient.
Definition A.9. Given $\alpha \in \mathbb{R}$, define $\kappa_{\alpha} := \{ x \in \mathbb{Q} \mid x + a > 0 \text{ for every } a \in \alpha \}$ and $-\alpha := \kappa_{-\alpha}$.

Example A.10. Clearly $0 \in \kappa_{0}$, so that $\kappa_{\sigma_{0}} = \sigma_{0} \cup \{0\}$. This has a least element. Thus $\kappa_{\sigma_{0}}$ cannot be a Dedekind cut. This illustrates why it is necessary to use $\kappa_{\alpha}^{*}$ in Definition A.9.

Lemma A.11. Take $\alpha \in \mathbb{R}$. Then

(i) $\alpha + \sigma_{0} = \alpha$,

(ii) $-\alpha \in \mathbb{R}$ and

(iii) $\alpha + (-\alpha) = \sigma_{0}$.

Proof. (i) Take $a \in \alpha$ and $x \in \sigma_{0}$.

Then $x > 0$ and so $a + x > a$, showing that $a + x \in \alpha$, that is $\alpha + \sigma_{0} \subseteq \alpha$.

Conversely, take $a \in \alpha$.

Then there is a $b \in \alpha$ with $b < a$.

Thus $a - b > 0$ and so $a - b \in \sigma_{0}$.

But $a = b + (a - b) \in \alpha + \sigma_{0}$, so that $\alpha + \sigma_{0} \subseteq \alpha$.

Thus $\alpha + \sigma_{0} = \alpha$.

(ii) Take $a \in \alpha$.

Then $a + (1 + |a|) \geq 1 > 0$ so that $1 + |a| \in \kappa_{\alpha}$.

Thus, $\kappa_{\alpha} \neq \emptyset$.

Moreover, since $a - a = 0$, $-a \notin \kappa_{\alpha}$, so that $\kappa_{\alpha} \neq \mathbb{Q}$, which verifies (D0).

Take $x \in \kappa_{\alpha}$ and $y \in \mathbb{Q}$ with $x < y$.

For every $a \in \alpha$, $y + a > x + a > 0$, so that $y \in \kappa_{\alpha}$, verifying (D1).

Thus $-\alpha := \kappa_{\alpha}^{*} \in \mathbb{R}$.

(iii) Take $x \in -\alpha$ and $a \in \alpha$.

By definition, $x + a > 0$ so that $x + a \in \sigma_{0}$.

Hence $\alpha + (-\alpha) \subseteq \sigma_{0}$.

Conversely, take $e \in \sigma_{0}$, so that $e > 0$.

Then $d := \frac{e}{2} > 0$ and so, by Lemma A.5, there is a $a \in \alpha$ with $a - d \notin \alpha$.

But then $a - d < x$ for every $x \in \alpha$ and so $x + (-a + d) > 0$ for every $x \in \alpha$.

Hence $-a + d \in \kappa_{\alpha}$ and so $-a + 2d \in \kappa_{\alpha}^{*} =: -\alpha$.

Since $a \in \alpha$ we see that $e = 2d = a + (-a + 2d) \in \alpha + (-\alpha)$ and so $\sigma_{0} \subseteq \alpha + (-\alpha)$.

Thus $\alpha + (-\alpha) = \sigma_{0}$.

For multiplication, we begin by restricting attention to $\mathbb{R}_{0}^{+} := \{ \alpha \in \mathbb{R} \mid \alpha \geq \sigma_{0} \}$, the set of non-negative Dedekind cuts.

We also define $\mathbb{R}^{+} := \{ \alpha \in \mathbb{R} \mid \alpha > \sigma_{0} \}$, the set of positive Dedekind cuts.

Definition A.12. Given $\alpha, \beta \in \mathbb{R}_{0}^{+}$, their product $\alpha \cdot \beta$ is defined by

$\alpha \cdot \beta := \{ a \cdot b \mid a \in \alpha, b \in \beta \}$.

The arguments used to show that the sum of two Dedekind cuts is itself a Dedekind cut can be used mutatis mutandis to show that the product of two non-negative Dedekind cuts is itself a
Dedekind cut and the properties of multiplication of rational numbers imply that
\[(\alpha\beta)\gamma = \alpha(\beta\gamma);\]
\[\alpha\beta = \beta\alpha;\]
\[\alpha(\beta + \gamma) = \alpha\beta + \alpha\gamma.\]

**Definition A.13.** Given \(\alpha \in \mathbb{R}^+_0\)
\[\lambda_\alpha := \{x \in \mathbb{Q} | x.a > 1 \text{ for every } a \in \alpha\}\]
\[\frac{1}{\alpha} := \lambda^*_\alpha\]

**Lemma A.14.** Take \(\alpha \in \mathbb{R}^+\). Then
(i) \(\alpha \sigma_1 = \alpha\),
(ii) \(\frac{1}{\alpha} \in \mathbb{R}^+\) and
(iii) \(\alpha \cdot \frac{1}{\alpha} = \sigma_1\).

**Proof.** (i) Take \(a.x \in \alpha \sigma_1\), so that \(a \in \alpha\) and \(x \in \sigma_1\).
But \(x \in \sigma_1\) if and only if \(x \in \mathbb{Q}\) and \(x > 1\).
Hence \(a.x > a\) and so, by (D1) \(a.x \in \alpha\), showing that \(\alpha \sigma_1 \subseteq \alpha\).
Now take \(a \in \alpha\).
By (D2) there is an \(a_1 \in \alpha\) with \(a_1 < a\).
Then \(x := \frac{a}{a_1}1\) and so \(x \in \sigma_1\).
Thus, \(a = a_1.x \in \alpha \sigma_1\), so that \(\alpha \subseteq \alpha \sigma_1\).
Hence, \(\alpha \sigma_1 = \alpha\).

(ii) Since \(\alpha \in \mathbb{R}^+\) there is a \(b \in \mathbb{Q}\) such that \(b > 0\) and \(b \notin \alpha\).
We keep this \(b\) fixed for the rest of this proof.
Note that for each \(a \in \alpha\), \(b < a\) and so \(\frac{1}{b}.a > 1\).
Hence \(\frac{1}{b} \in \lambda_\alpha\) so that \(\lambda_\alpha \neq \emptyset\).
Moreover, since \(0 \notin \lambda_\alpha, \lambda_\alpha \neq \mathbb{Q}\), verifying (D0).
Take \(x \in \lambda_\alpha\) and \(y \in \mathbb{Q}\) with \(x < y\). Take any \(a \in \alpha\).
Then \(y.a > x.a > 1\) and so \(y \in \lambda_\alpha\), verifying (D1).
Hence \(\frac{1}{a} \in \mathbb{R}\).
Moreover, since \(0 \notin \frac{1}{a}, 0 < x\) whenever \(x \in \frac{1}{a}\).
Thus, \(\frac{1}{a} \in \mathbb{R}^+\).

(iii) Since \(a.x > 1\) for every \(a \in \alpha\) and \(x \in \frac{1}{a}\),
\[\alpha \cdot \frac{1}{a} \subseteq \sigma_1\]
Take \(z \in \sigma_1\).
Then \(z = 1 + 2d\) for some suitable \(d \in \mathbb{Q}, d > 0\).
Since \( \frac{bd}{1+d} > 0 \) Lemma A.5 implies that there is a \( c \in \alpha \) such that \( c - \frac{bd}{1+d} \not\in \alpha \). Now
\[
\frac{c}{1+d} = c - \frac{cd}{1+d} < c - \frac{bd}{1+d} \not\in \alpha
\]
since \( b < c \).
Thus \( \frac{c}{1+d} < a \) for every \( a \in \alpha \).
In other words \( \frac{1+d}{c} \cdot a > 1 \) for every \( a \in \alpha \), whence \( \frac{1+d}{c} \in \lambda_\alpha \).

Hence
\[
\frac{z}{c} = \frac{1+2d}{c} \in \frac{1}{\alpha}
\]
Thus,
\[
z = \frac{z}{c} \cdot c \in \frac{1}{\alpha} \cdot \alpha
\]
so that
\[
\sigma_1 \subseteq \frac{1}{\alpha} \cdot \alpha
\]
Thus \( \sigma_1 = \frac{1}{\alpha} \cdot \alpha \).

It remains to extend multiplication to all of \( \mathbb{R} \) and to verify the rest of the axioms for the real numbers.

We outline procedures for doing so, leaving the details to the reader.

It follows from the properties of addition that for any \( \alpha \in \mathbb{R} \),
\[
(-\alpha) = \alpha
\]
\[
-(\alpha + \beta) = (-\alpha) + (-\beta)
\]

We write \((-1)^2\alpha\) for \((-\alpha)\) and define
\[
(-1)^{n+1}\alpha := -(-1)^n\alpha
\]
for \( n \in \mathbb{N} \), so that given \( k \in \mathbb{N} \), \((-1)^{2k}\alpha = \alpha\) and \((-1)^{2k+1}\alpha = -\alpha\).

It follows from the properties of \( \mathbb{Q} \) that given any \( \alpha \in \mathbb{R} \), there is a \( \mu \in \mathbb{R}_0^+ \) and a \( k \in \mathbb{N} \) with \( \alpha = (-1)^k\mu \).
Moreover, this \( \mu \) is uniquely determined by \( \alpha \). We therefore denote it by \( |\alpha| \).

Note that except when \( \alpha = \sigma_0 \) the \textit{parity} of \( k \) (that is, whether \( k \) is odd or even) is also uniquely determined by \( \alpha \).

This allows us to define the product of any two elements of \( \mathbb{R} \).

**Definition A.15.** Take \( \alpha, \beta \in \mathbb{R} \). Then
\[
\alpha \cdot \beta := (-1)^{k+l}|\alpha||\beta|,
\]
where \( \alpha = (-1)^k|\alpha| \) and \( \beta = (-1)^l|\beta| \).

(Note that the parity of \( k \) and the parity of \( l \) — and hence that of \( k+l \) — are well defined unless at least one of \( \alpha \) and \( \beta \) is \( \sigma_0 \), in which case the product is \( \sigma_0 \), rendering the parity of \( k+l \) irrelevant.)
The verification that the axioms for the real numbers, which are precisely those for the rational numbers listed above together with the property of order completeness proved above, are satisfied, is now straightforward.

Moreover it is easy to see that given rational numbers \( a \) and \( b \)

\[
\sigma_{a+b} = \sigma_a + \sigma_b \\
\sigma_{a\cdot b} = \sigma_a \cdot \sigma_b
\]

so that the imbedding

\[
i : \mathbb{Q} \longrightarrow \mathbb{R}, \quad a \mapsto \sigma_a
\]

is compatible with the arithmetic operations, that is,

\[
\sigma_{a+b} = \sigma_a + \sigma_b \\
\sigma_{a\cdot b} = \sigma_a \cdot \sigma_b
\]
The Greek Alphabet

The Greek alphabet is frequently used in the mathematical sciences. We list its characters and their names for the benefit of those who are not yet familiar with them.

<table>
<thead>
<tr>
<th>Character</th>
<th>Greek</th>
<th>Latin</th>
</tr>
</thead>
<tbody>
<tr>
<td>alpha</td>
<td>α</td>
<td>A</td>
</tr>
<tr>
<td>beta</td>
<td>β</td>
<td>B</td>
</tr>
<tr>
<td>gamma</td>
<td>γ</td>
<td>Γ</td>
</tr>
<tr>
<td>delta</td>
<td>δ</td>
<td>Δ</td>
</tr>
<tr>
<td>epsilon</td>
<td>ε, ε</td>
<td>E</td>
</tr>
<tr>
<td>zeta</td>
<td>ζ</td>
<td>Z</td>
</tr>
<tr>
<td>eta</td>
<td>η</td>
<td>H</td>
</tr>
<tr>
<td>theta</td>
<td>θ, θ</td>
<td>Θ</td>
</tr>
<tr>
<td>iota</td>
<td>ι</td>
<td>I</td>
</tr>
<tr>
<td>kappa</td>
<td>κ</td>
<td>K</td>
</tr>
<tr>
<td>lambda</td>
<td>λ</td>
<td>Λ</td>
</tr>
<tr>
<td>mu</td>
<td>μ</td>
<td>M</td>
</tr>
<tr>
<td>nu</td>
<td>ν</td>
<td>N</td>
</tr>
<tr>
<td>xi</td>
<td>ξ</td>
<td>Ξ</td>
</tr>
<tr>
<td>omicron</td>
<td>ο</td>
<td>O</td>
</tr>
<tr>
<td>pi</td>
<td>π, ϖ</td>
<td>Π</td>
</tr>
<tr>
<td>rho</td>
<td>ρ, ϑ</td>
<td>P</td>
</tr>
<tr>
<td>sigma</td>
<td>σ, ς</td>
<td>Σ</td>
</tr>
<tr>
<td>tau</td>
<td>τ</td>
<td>T</td>
</tr>
<tr>
<td>upsilon</td>
<td>υ</td>
<td>Υ</td>
</tr>
<tr>
<td>phi</td>
<td>φ, ϕ</td>
<td>Φ</td>
</tr>
<tr>
<td>chi</td>
<td>χ</td>
<td>X</td>
</tr>
<tr>
<td>psi</td>
<td>ψ</td>
<td>Ψ</td>
</tr>
<tr>
<td>omega</td>
<td>ω</td>
<td>Ω</td>
</tr>
</tbody>
</table>