An introduction to Real Analysis

Laurent W. Marcoux

February 9, 2023

For be a man's intellectual superiority what it will, it can never assume the practical, available supremacy over other men, without the aid of some sort of external arts and entrenchments, always, in themselves, more or less paltry and base. This it is, that for ever keeps God's true princes of the Empire from the world's hustings; and leaves the highest honors that this air can give, to those men who become famous more through their infinite inferiority to the choice hidden handful of the Divine inert, than through their undoubted superiority over the dead level of the mass. Such large virtue lurks in these small things when extreme political superstitions invest them, that in some royal instances even to idiot imbecility they have imparted potency.

Herman Melville: *Moby Dick*

Preface to the Fourth Edition - February 07, 2023

We continue to correct errors, and add minor exercises and notes. We *strongly* recommend that the reader not limit themselves to one single text when learning any mathematical subject, including Real Analysis. A excellent resource is the book of Rudin [**4**], and for general topology, one can't go wrong with the Willard's *General Topology* [**3**].

André's *Axioms and set theory* [**1**] is an excellent introduction to set theory, as is Halmos' *Naive set theory* [**2**].

Preface to the Third Edition - May 22, 2019

We continue to correct errors, and add minor exercises and notes. A special "Thank You!" to Prof. Y. Zhang, for interesting and helpful discussions about the notes, as well as pointing out the use of the Axiom of Choice in the proof of Theorem 1.4.8.

Preface to the Second Edition - December 6, 2016

A number of typos from the first edition have now been corrected. Presumably, many others remain, and there may even be new ones! Please read the preface below, and bring any remaining typos/errors to my attention.

I'd like to thank Mr. Nihal Pednekar for pointing out a number of typos.

Preface to the First Edition - May 6, 2014

The following is a set of class notes for a Real Analysis course I taught in 2014. As mentioned on the front page, they are a work in progress, and - this being the "first edition" - they are replete with typos. A student should approach these notes with the same caution he or she would approach buzz saws; *they can be very useful, but you should be thinking the whole time you have them in your hands.* Enjoy.

The reviews are in!

From the moment I picked your book up until I laid it down I was convulsed with laughter. Someday I intend reading it.

Groucho Marx

This is not a novel to be tossed aside lightly. It should be thrown with great force.

Dorothy Parker

The covers of this book are too far apart.

Ambrose Bierce

I read part of it all the way through.

Samuel Goldwyn

Reading this book is like waiting for the first shoe to drop.

Ralph Novak

Thank you for sending me a copy of your book. I'll waste no time reading it.

Moses Hadas

Sometimes you just have to stop writing. Even before you begin.

Stanislaw J. Lec

Contents

CHAPTER 1

Set Theory and Cardinality

1. Introduction

Somewhere on this globe, every ten seconds, there is a woman giving birth to a child. She must be found and stopped.

Sam Levenson

1.1. The purpose of this introduction is not to develop a formal, axiomatic theory of sets, but rather to discuss in an informal way those concepts which we shall need for this course. For those of you with an interest in set theory, we *strongly* recommend the wonderful text *Axioms and Set Theory*, by **Robert André** [**1**].

We shall adopt as "primitive" the notions of *classes*, *sets* and *belongs to*, which we denote by $\in$. To say that these notions are "primitive" means that we do not define them explicitly, but rather, that we assume that we know what is meant by these terms. All of the objects in our theory will be **classes**. **Sets** will refer to special classes. If A is a class and $A \in B$ for another class B, we shall say that A is an **element** of B. Following [**1**], we shall use lower case letters to denote classes that are also elements. The axioms of set theory will declare operations that we may perform on sets and classes in order to obtain new sets and classes, and to describe the existence of particular sets and classes.

1.2. A *mathematical statement* is a declaration that is either true or false. In other words, if it is not true, it is not false. We shall use these statements to describe our axioms.

For example, we can consider two basic types of sentences:

- *belonging*: "$x \in X$";
- *equality*: $Y = Z$.

Other sentences are obtained from given sentences by repeated applications of logical operators

- $\in$ is an element of;
- $\vee$ or;
- $\wedge$ and;
- $\neg$ not;

- $\forall$ for all; and
- $\exists$ there exists.

If S_1 and S_2 are sentences, we might consider:

(i) S_1 and S_2;
(ii) S_1 or S_2 (either S_1 or S_2 or both);
(iii) not S_1;
(iv) if S_1, then S_2;
(v) S_1 if and only if S_2;

If P is a sentence describing a property that an element x of a class X may or may not have, we may also consider statements of the form:

(vi) for all elements x of a class X, S_1 holds;
(vii) there exists an element x of the class X for which S_1 holds.

Note: in items (vi) and (vii), x is a variable name. For example, the sentence:

$$\text{For some } y, x \in A$$

means the same as $x \in A$, since y does not appear in the sentence $x \in A$. We leave it to the reader to express items (iv) and (v) in terms of "and", "or" and "not".

Set theory has as its core the following Axioms, named after **Ernst Zermelo** and **Abraham Fraenkel**, who helped to develop the axiomatic theory still in use today. (André [**1**] credits T. Skolem and J. von Neumann as having made slight modifications to the axioms.)

1.3. The Zermelo-Fraenkel Axioms.

(**A1**) (**Axiom of extension**): If x, A and B are classes and x is an element, then $A = B$ if and only if

$$([x \in A] \text{ if and only if } [x \in B]) .$$

(**A2**) (**Axiom of class construction**): Let $P(x)$ designate a statement about x which can be expressed entirely in terms of the symbols $\in, \vee, \wedge, \neg, \implies, \forall$, brackets and variables $x, y, z, ..., A, B, ...$ Then there exists a class C which consists of all the elements x which satisfy $P(x)$.

(**A3**) (**Axiom of pair**): If A and B are sets, then the doubleton $\{A, B\}$ is a set.

(**A4**) (**Axiom of specification/Axiom of Subsets/Aussonderungsaxiom**): If S is a set and P is a formula describing a particular property, then the class of all sets in S which satisfy this property P is a set. More succinctly, every subclass of a set (of sets) is a set. This is sometimes also referred to as the **Axiom of Subsets**, or by its German name, the **Aussonderungsaxiom**.

(**A5**) (**Axiom of power set**): If A is a set, then the power set $\mathcal{P}(A)$ is a set.

(**A6**) (**Axiom of union**): If $\mathcal{C}$ is a set of sets, then $\cup_{C \in \mathcal{C}} C$ is a set.

(**A7**) (**Axiom of replacement**): Let A be a set, and let $\varphi(x,y)$ be a formula which associates to each x of A an element y in such a way that whenever both ($\varphi(x,y)$ and $\varphi(x,z)$ hold true, then $y = z$. Then there exists a set B which contains all elements y such that $\varphi(x,y)$ holds true for some $x \in A$.

(**A8**) (**Axiom of infinity**): There exists a non-empty class A called a *set* that satisfies the condition:

$$\text{“}X \in A\text{”} \implies \text{“}X \cup \{X\} \in A\text{”}.$$

(**A9**) (**Axiom of regularity**): Every non-empty set A contains an element x whose intersection with A is empty.

1.4. Given two classes A and B, the statement $A \subseteq B$ is understood to mean $[x \in A]$ implies $[x \in B]$. Thus we may rewrite (A1) as follows:

The Axiom of Extension: Let A and B be classes. Then

$$A = B \text{ if and only if } A \subseteq B \text{ and } B \subseteq A.$$

This relates equality of sets to the concept of elements belonging to a class.

It follows from the Axiom of subsets that there exists a *set* $\varnothing$ which contains no elements. Namely, if S is a set, let $P(x)$ be the statement $x \neq x$. Then, setting $T := \{x \in S : x \neq x\}$ shows that $T = \varnothing$ is a set.

1.5. The Axiom of Subsets refers to a *set* S (of sets), as opposed to a *class* S. The property P referenced therein in turn refers to *sets in* S, which therefore define *elements* of S. That is, P describes a property which an element of a set may or may not have, and the Axiom states that given a set S, the subclass of S consisting of those elements for which the property P holds forms a *subset* of S (as opposed to just a *subclass* of S).

1.6. Example. Does there exist a "universal *set*" that contains every set? That is, can we define a set B which would be the "*set of all sets*"?

Suppose so, and let $P(x)$ be the sentence: "$\neg(x \in x)$", or equivalently, "$x \notin x$".

By the Axiom of Subsets, we may then define the *set*

$$A = \{x \in B : P(x)\} = \{x \in B : x \notin x\}.$$

Thus $y \in A$ if and only if $y \in B$ and $y \notin y$.

Since A is a set and B is the "set of all sets", it follows that $A \in B$. Of course, there are only two possibilities: either $A \in A$ or $A \notin A$.

- If $A \in A$, then $A \in B$ and $A \in A$ so that $A \notin A$, a contradiction.
- If $A \notin A$, then $A \in B$ and $A \notin A$ which implies $A \in A$, a contradiction.

This situation is unacceptable. We therefore conclude that there is no "universal *set*" - that is, there is no *set* that contains as its elements every set. At one time, it was assumed that such a universe did exist. The above example was known as *Russel's paradox*.

Observe, however, that we may still define the *class* of all sets: if $P(x)$ is the statement that $x = x$, then by the Axiom of class construction, the class $U \coloneqq \{x : x = x\}$ defines the class of all sets. Note that in order to do this, one would need to know that every set is not just a class, but also an element, since the Axiom of class construction requires C to be constructed from *elements*. Fortunately, the Axiom of power sets ensures that if A is a set, then its power set $\mathcal{P}(A)$ is a set, and thus A is not just a set, but an element (of $\mathcal{P}(A)$)! We're all good.

Exercise: If X is a finite set with n elements, then $\mathcal{P}(X)$ has 2^n elements.

1.7. Let $\Lambda \neq \varnothing$ and suppose that $\{X_\lambda\}_{\lambda\in\Lambda}$ is a set of sets. Then we would like to define the **union** of the X_λ's to be

$$\cup_{\lambda\in\Lambda} X_\lambda = \{x : x \in X_\lambda \text{ for some } \lambda \in \Lambda\},$$

and the **intersection** of the X_λ's to be

$$\cap_{\lambda\in\Lambda} X_\lambda = \{x : x \in X_\lambda \text{ for all } \lambda \in \Lambda\}.$$

Let's see what happens if $\Lambda = \varnothing$. Now $x \notin \cap_{\lambda\in\Lambda} X_\lambda$ implies that there exists $\lambda_0 \in \varnothing$ so that $x \notin X_{\lambda_0}$, which is false. It follows that $x \in \cap_{\lambda\in\Lambda} X_\lambda$ for all x. But we have just seen that there is no universal set that contains *every set*, so what can this mean?

To get beyond such problems, we shall always assume that in dealing with set constructions, we are beginning with a "universe" X which *is* a set, and that $X_\lambda \subseteq X$ for all λ.

1.8. Definition. *Let $\varnothing \neq \Lambda, X$ be sets and suppose that $\{X_\lambda\}_{\lambda\in\Lambda}$ is a set of subsets of X. Then we define the **union** of the X_λ's to be*

$$\cup_{\lambda\in\Lambda} X_\lambda = \{x \in X : x \in X_\lambda \textit{ for some } \lambda \in \Lambda\},$$

*and the **intersection** of the X_λ's to be*

$$\cap_{\lambda\in\Lambda} X_\lambda = \{x \in X : x \in X_\lambda \textit{ for all } \lambda \in \Lambda\}.$$

Then it follows from the above argument that $\cap_{\lambda\in\varnothing} X_\lambda = X$.

Exercise: What should $\cup_{\lambda\in\varnothing} X_\lambda$ mean?

1.9. Definition. *Let $\Lambda \neq \varnothing$ and let $\{X_\lambda\}_{\lambda\in\Lambda}$ be a set of subsets of a universe X. We define the **product** of the sets X_λ to be:*

$$\prod_{\lambda\in\Lambda} X_\lambda = \{f : \Lambda \to \cup_{\lambda\in\Lambda} X_\lambda : f(\lambda) \in X_\lambda \textit{ for all } \lambda \in \Lambda\}.$$

*If such a function f exists, it is called a **choice function**.*

Note: If $X_{\lambda_0} = \varnothing$ for some $\lambda_0 \in \Lambda$, then $f(\lambda_0) \in X_{\lambda_0}$ is false, and so $\prod_{\lambda\in\Lambda} X_\lambda = \varnothing$.

Given non-empty sets X and Y, we define

$$X^Y = \{f : Y \to X : f \text{ is a function}\} = \prod_{y\in Y} X_y,$$

where $X_y = X$ for all $y \in Y$.

1.10. Do choice functions always exist?

(a) Suppose that Λ is a finite, non-empty set and that for all $\lambda \in \Lambda$, $\varnothing \neq X_\lambda$ is a set. Then the answer is "yes". This follows from the basic axioms of Zermelo-Fraenkel set theory. For example, suppose that $\Lambda = \{1, 2\}$ and that X_1, X_2 are non-empty sets. In ZF-theory, the elements of X_1 and X_2 are themselves sets. Since $X_1 \neq \varnothing$, the statement "there exists a set x_1 such that $x_1 \in X_1$" is true, and similarly the statement "there exists a set x_2 such that $x_2 \in X_2$" is true. The Axiom of Pairing says implies that $\{x_1, x_2\}$ is a set, and hence that $\{x_1, \{x_1, x_2\}\}$ is a set. But this may be identified as an ordered pair (x_1, x_2), since the first instance of x_1 tells us which "comes first". This ordered pair may be thought of as an element of $X_1 \times X_2$, showing that the latter is non-empty, which is the Axiom of Choice for two sets. It fails for infinitely many sets because one would have to apply the Axiom of Pairing infinitely many times, and that would take all day and all night and a lot, lot, lot more. Now really, this is not a course in set theory, so let us never speak of this again. The reader is referred to [**1**] or a course in set theory/logic for more detail.)

(b) Let Λ be an arbitrary non-empty set. For each $\lambda \in \Lambda$, suppose that $\varnothing \neq X_\lambda \subseteq \mathbb{N}$. Given $\lambda \in \Lambda$, define $f(\lambda)$ to be the least element of X_λ. Then $f \in \prod_{\lambda \in \Lambda} X_\lambda$ is a choice function.

(c) Let Λ be an arbitrary non-empty set. Suppose that for each $\lambda \in \Lambda$, P_λ consists of a pair $\{L_\lambda, R_\lambda\}$ of shoes (where L_λ is the left shoe, and R_λ is the right shoe). Given $\lambda \in \Lambda$, set $g(\lambda) = L_\lambda$. Then $g \in \prod_{\lambda \in \Lambda} P_\lambda$ is a choice function.

(d) For each $n \geq 1$, let B_n denote a pair of identical socks. How do we specify a choice function $f \in \prod_{n \in \mathbb{N}} B_n$?

2. The Axiom of Choice

2.1. The above question prompted the following quote from the mathematician (and philosopher) Bertrand Russel(1872-1970):

> To choose one sock from each of infinitely many pairs of socks requires the Axiom of Choice, but for shoes the Axiom is not needed.

As the reader will undoubtedly come to appreciate over their undergraduate career, it is twentieth century's obsession with socks which drove most of the mathematics discovered over the last 116 years.

2.2. One way to circumvent the question of how we can choose one sock from amongst each pair in an infinite collection of pairs of socks is to *assume* we can.

The Axiom of Choice [AC]. If $\Lambda \neq \varnothing$ is a set and for each $\lambda \in \Lambda$, X_λ is a non-empty subset of a universe X, then $\prod_{\lambda \in \Lambda} X_\lambda \neq \varnothing$.

Exercise: Prove that the Axiom of Choice is equivalent to the following:

The Axiom of Choice - disjoint set version [ACD]. Suppose that $\Lambda \neq \varnothing$ is a set and that

(i) for all $\lambda \in \Lambda$, X_λ is a non-empty subset of a universe X, and
(ii) $X_\lambda \cap X_\beta = \varnothing$ if $\lambda \neq \beta \in \Lambda$.

Then $\prod_{\lambda\in\Lambda} X_\lambda \neq \varnothing$.

Exercise: Prove that the Axiom of Choice is equivalent to the following statement:

given a non-empty set X there exists a function $f : \mathcal{P}(X) \setminus \{\varnothing\} \to X$ so that $f(A) \in A$ for all $A \in \mathcal{P}(X) \setminus \{\varnothing\}$.

2.3. At first glance, it would seem madness to even try to imagine that the Axiom of Choice is not true. As it turns out, we can appeal to the Principle of *"you're damned if you do and you're damned if you don't"* to begin to appreciate the can of worms we have just opened.

It can be (in fact it *has been*) shown that the Axiom of Choice implies the following: it is possible to "carve up" the unit ball in $\mathbb{R}^3$ into finitely many pieces and, using only rotations and translations, to reassemble those pieces into two balls each having the same volume as the original unit ball. This is known as the *Banach-Tarski Paradox*. As one might imagine, this result is non-constructive. It does not tell you *how* to cut the unit ball. It would be unwise yet strangely thirst-quenching to test this out on a bag of oranges using a typical kitchen knife.

On the other hand, the negation of (AC) implies the existence of two sets A and B so that neither of these can be mapped injectively into the other. This too is not good.

Our next goal is to obtain a couple of equivalent formulations of the Axiom of Choice which will prove useful both in analysis and in algebra. Before describing these equivalent formulations, we shall pause to develop some notation and definitions.

2.4. Definition. *A* ***relation*** *R on a set X is a subset of the Cartesian product $X \times X = \{(x,y) : x, y \in X\}$. We write xRy if $(x,y) \in R$.*

A relation $\leq$ is called a ***partial order*** *on X if it satisfies*

(i) *$x \leq x$ for all $x \in X$ (reflexivity);*
(ii) *$x \leq y$ and $y \leq z$ implies that $x \leq z$ (transitivity);*
(iii) *$x \leq y$ and $y \leq x$ implies that $x = y$ (anti-symmetry).*

The ordered pair $(X, \leq)$ is called a ***partially ordered set****, or simply a* ***poset****. Informally, it is also customary to refer to X as the poset with partial order $\leq$.*

A ***chain*** *$\mathcal{C}$ in X is a subset of X such that for any $x, y \in \mathcal{C}$, either $x \leq y$ or $y \leq x$. Alternatively, these are called* ***totally ordered sets*** *or* ***linearly ordered sets****.*

2.5. Example.

(a) $(\mathbb{R}, \leq)$ is a totally ordered (and hence a partially ordered) set using the usual order on $\mathbb{R}$. Similarly, $(\mathbb{Q}, \leq)$ is a totally ordered set using the same partial order.

(b) The list of words in the dictionary forms a totally ordered set with the usual lexicographic ordering.

2.6. Example. Let $X \neq \varnothing$ be a set. Consider the power set $\mathcal{P}(X)$. For $A, B \in \mathcal{P}(X)$, define $A \leq B$ to mean $A \subseteq B$. We say that $\mathcal{P}(X)$ is **(partially) ordered by inclusion**. Then $(\mathcal{P}(X), \leq)$ is a poset. If X has more than one element, then $(\mathcal{P}(X), \leq)$ is not a chain.

Suppose $X = \{1,2,3,4,5\}$ and that $\mathcal{P}(X)$ is ordered by inclusion. Then

$$\mathcal{C} = \{\{2\}, \{2,5\}, \{2,3,5\}\}$$

is a chain in $\mathcal{P}(X)$. The set $\mathcal{D} = \{\{2\}, \{2,5\}, \{1,3,5\}\}$ is not a chain.

2.7. Example. Let $X \neq \varnothing$ be a set. Consider the power set $\mathcal{P}(X)$. For $A, B \in \mathcal{P}(X)$, define $A \leq B$ to mean $A \supseteq B$. We say that $\mathcal{P}(X)$ is **ordered by containment**. Then $(\mathcal{P}(X), \leq)$ is a poset. If X has more than one element, then $(\mathcal{P}(X), \leq)$ is not a chain.

2.8. Example. Let

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

For $f, g \in X$, define $f \leq g$ if $f(x) \leq g(x)$ for all $x \in [0,1]$. Then $(X, \leq)$ is a partially ordered set.

2.9. Example. Let $n \geq 1$, and suppose that V is an n-dimensional vector space over $\mathbb{R}$. Let $X = \{W : W \text{ is a subspace of } V\}$, partially ordered by inclusion. Then $(X, \leq)$ is a poset. If $B = \{v_1, v_2, ..., v_n\}$ is a basis for V and $W_k = \operatorname{span}\{v_1, ..., v_k\}$, $1 \leq k \leq n$, then $\mathcal{C} = \{\{0\}, W_1, W_2, ..., W_n\}$ is a chain in X. We leave it as an **exercise** to the reader to show that $\mathcal{C}$ is not contained in any bigger chain in X.

2.10. Definition. *Let $(X, \leq)$ be a poset. We say that $x \in X$ is **maximal** in X if $y \in X$ and $x \leq y$ implies $x = y$. We say that $m \in X$ is a **maximum element** in X if $m \geq y$ for all $y \in X$.*

*We say that $z \in X$ is **minimal** in X if $y \in X$ and $y \leq z$ implies $y = z$. The element $n \in X$ is a **minimum element** in X if $y \in X$ implies that $n \leq y$.*

The distinction between a maximal element and a maximum element is that a maximum element must be comparable to (and at least as big as) *every* element of the poset $(X, \leq)$. A maximal element need only be as big as those elements in X to which it is actually comparable.

2.11. Example.

(a) Let $X = \{1,2,3,4,5,6\}$, and denote by $\mathcal{P}_0(X)$ the collection of *proper* subsets of X, partially ordered by inclusion. (Recall that a subset $A \subseteq X$ is *proper* if $A \neq X$.) Then $N_1 = \{1,2,3,4,5\}$ and $N_2 = \{1,3,4,5,6\}$ are two distinct maximal elements of $\mathcal{P}_0(X)$. Neither of these is a maximum element; for example, $Y = \{6\} \in \mathcal{P}_0(X)$, but $Y \nleq N_1$. In fact, $\mathcal{P}_0(X)$ does not have a maximum element at all.

(b) Let $X = (0,1)$, equipped with the usual order inherited from $(\mathbb{R}, \leq)$. Again, X does not have a maximum element. In this case, it also does not have a maximal element. Moreover, $(\mathbb{R}, \leq)$ itself does not have any maximal elements.

(c) Let $n \geq 1$ be an integer and consider the algebra $\mathcal{T}_n$ of upper triangular $n \times n$ matrices over $\mathbb{C}$. Let $(X, \leq)$ denote the set of *proper* ideals of $\mathcal{T}_n$, partially ordered with respect to inclusion.

Exercise: prove that $M \in X$ is maximal if and only if there exists $1 \leq j \leq n$ so that $M = \{T = [t_{ij}] \in \mathcal{T} : t_{jj} = 0\}$.

(d) Consider $\mathcal{C}([0,1],\mathbb{C}) = \{f : [0,1] \to \mathbb{C} : f \text{ is continuous}\}$. Let $(X, \leq)$ denote the set of proper ideals of $\mathcal{C}([0,1],\mathbb{C})$, partially ordered with respect to inclusion. Observe that $\{0\} \in X$, so that $X \neq \varnothing$.

For each $y \in [0,1]$, define the set $\mathcal{K}_y := \{f \in \mathcal{C}([0,1],\mathbb{C}) : f(y) = 0\}$. It is not difficult to check that $\mathcal{K}_y$ is an ideal of $\mathcal{C}([0,1],\mathbb{C})$. If

$$\begin{array}{rccc} \varphi_y : & \mathcal{C}([0,1],\mathbb{C}) & \to & \mathbb{C} \\ & f & \mapsto & f(y) \end{array},$$

then φ_y is a (multiplicative) linear map, and clearly φ_y is surjective. By linear algebra,

$$\mathbb{C} = \operatorname{ran} \varphi_y \simeq \mathcal{C}([0,1],\mathbb{C}) / \ker \varphi_y.$$

But $\ker \varphi_y = \mathcal{K}_y$. Since $\mathcal{K}_y$ has co-dimension 1 in $\mathcal{C}([0,1],\mathbb{C})$, it must be maximal.

Exercise*: show that these are the only maximal ideals of $\mathcal{C}([0,1],\mathbb{C})$.

(e) **Exercise:** Every finite poset has a maximal element. (It is also a worthwhile exercise to describe all 3 element posets to get a feeling for what is going on.)

2.12. Definition. *Let $(X, \leq)$ be a poset and $A \subseteq X$. We say that $y \in X$ is an* ***upper bound*** *for A if $a \leq y$ for all $a \in A$. We say that $x \in X$ is a* ***lower bound*** *for A if $x \leq a$ for all $a \in A$.*

We say that $\beta \in X$ is the ***least upper bound (lub)****, or* ***supremum (sup)*** *for A if*

- *β is an upper bound for A, and*
- *if y is any upper bound for A, then $\beta \leq y$.*

*Similarly, we say that $\alpha \in X$ is the **greatest lower bound (glb)** or **infimum (inf)** for A if*

- *α is a lower bound for A, and*
- *if x is any lower bound for A, then $x \le \alpha$.*

2.13. Example.

(a) $(\mathbb{R}, \le)$ has the least upper bound property, where $\le$ is the usual ordering on $\mathbb{R}$. If $\varnothing \ne A \subseteq \mathbb{R}$ is bounded above, then A has a least upper bound β.

If $A = \varnothing$, then for any $b \in \mathbb{R}$, b is an upper bound for A. Indeed, if b were not an upper bound for $A = \varnothing$, then there would exist an element $a \in A$ such that $b \nleq a$, which is false. Since b is an upper bound for $\varnothing$ for all $b \in \mathbb{R}$, we say that the *least upper bound* of $\varnothing$ is $-\infty$.

Here, $-\infty$ is *not* a number! The statement $\sup \varnothing = -\infty$ is to be interpreted as saying that any $b \in \mathbb{R}$ is an upper bound for $\varnothing$.

Note that using the same logic, we write $\infty = \inf \varnothing$, as every $b \in \mathbb{R}$ is also a lower bound for $\varnothing$.

(b) Let X be a non-empty set and let $\mathcal{P}(X)$ denote its power set, partially ordered by inclusion. If $\{X_\lambda\}_{\lambda\in\Lambda} \subseteq \mathcal{P}(X)$, then $\cup_{\lambda\in\Lambda} X_\lambda$ is the l.u.b. of $\{X_\lambda\}_{\lambda\in\Lambda}$, and $\cap_{\lambda\in\Lambda} X_\lambda$ is the g.l.b. of $\{X_\lambda\}_{\lambda\in\Lambda}$.

(c) Consider $(\mathbb{Q}, \le)$ where $\le$ denotes the usual total order inherited from $\mathbb{R}$. The set $A = \{x \in \mathbb{Q} : x^2 < 2\}$ is bounded above, but there is no least upper bound for A *in* $\mathbb{Q}$. Indeed, if $b \in \mathbb{Q}$ and $b > \sqrt{2}$, then there exists another rational number $d \in (\sqrt{2}, b)$, and thus d is an upper bound for A and $d < b$, so b is not the supremum of A. If $b \in \mathbb{Q}$ and $b < \sqrt{2}$, then clearly b is not even an upper bound for A, so it is not a supremum for A.

The Axiom of Choice was introduced by Zermelo in order to prove his Well-ordering Principle. To explain this, we first need a couple of definitions.

2.14. Definition. *A non-empty poset $(X, \le)$ is said to be **well-ordered** if every non-empty subset $A \subseteq X$ has a mimimum element.*

It immediately follows that every well-ordered set is totally ordered.

2.15. Example.

(a) The set $\mathbb{N}$ is well-ordered with the usual ordering, whereas $\mathbb{R}$ is not.

(b) Let $\omega + 7 = \{1, 2, 3, ..., \omega, \omega + 1, \omega + 2, ..., \omega + 6\}$. Define a partial order on $\omega + 7$ by setting $n \le \omega + k$ for all $n \ge 1$, $0 \le k \le 6$ and $\omega + i \le \omega + j$ if $0 \le i \le j \le 6$. The ordering on $\mathbb{N} \subseteq \omega + 7$ is the usual ordering on $\mathbb{N}$. Then $\omega + 7$ is well-ordered.

2.16. Theorem. *The following are equivalent:*

(i) *The Axiom of Choice (AC): given a non-empty collection $\{X_\lambda\}_{\lambda\in\Lambda}$ of non-empty sets, $\prod_{\lambda\in\Lambda} X_\lambda \neq \varnothing$.*
(ii) *Zorn's Lemma (ZL): Let $(Y,\leq)$ be a poset. Suppose that every chain $\mathcal{C} \subseteq Y$ has an upper bound. Then Y has a maximal element.*
(iii) *The Well-Ordering Principle (WO): Every non-empty set Z admits a well-ordering.*

Proof. This result has been moved to PM433. You may consult the appendix to this Chapter for a proof.

□

2.17. Theorem. *Every vector space V has a basis.*
Proof. See **Assignment One**.

□

2.18. Remark. Many results are known to be equivalent to the Axiom of Choice (in Zermelo-Fraenkel theory). In fact, the above result from Theorem 2.17 is amongst these.

Others include:

- Let V be a vector space over a field $\mathbb{F}$ and suppose that $J \subseteq S \subseteq V$, where J is a linearly independent subset of V, and $\operatorname{span} S = V$. Then there exists a basis B for V with $J \subseteq B \subseteq S$.
- If X and $Y \neq \varnothing$ are disjoint sets and X is infinite, then there exists a bijection between $X \times Y$ and $X \cup Y$.
- If a set A is infinite, then there is a bijection between the sets A and $A \times A$.
- Given sets A and B, either there exists an injection from A into B, or there exists an injection from B into A.
- Every unital ring R contains a maximal ideal.
- Let X be a set and $\mathcal{F}$ be a collection of subsets of X. Suppose that $\mathcal{F}$ has **finite character**, that is: $Y \in \mathcal{F}$ if and only if each finite subset F_Y of Y lies in $\mathcal{F}$. Then any member of $\mathcal{F}$ is a subset of some maximal (with respect to containment) member of $\mathcal{F}$.

The following results are known to be weaker than the Axiom of Choice. They are mentioned for information purposes at this point - some definitions will eventually follow.

- Any countable union of countable sets is countable.
- Every Hilbert space has an orthonormal basis.
- The Hahn-Banach Theorem.

Throughout this course we shall follow the standard usage of the Axiom of Choice, namely: we shall assume that it holds, but we shall mention it explicitly whenever it is used.

2.19. Well-ordered sets allow us to extend the usual Principle of Induction.
The Principle of Transfinite Induction:

Suppose that $(X, \leq)$ is a well-ordered set, and that $S \subseteq X$. Suppose that $x \in X$ and that $P(X,x) := \{y \in X : y < x\} \subseteq S$ implies that $x \in S$. Then $S = X$.

Proof. If $S \neq X$, let x_0 denote the mimimum element of $X \setminus S$. Then $P(X, x_0) \subseteq S$, so $x_0 \in S$, a contradiction. Hence $S = X$.

□

2.20. Example. Suppose that $X = \mathbb{N} \cup \{\omega\}$, where we define $n < \omega$ for each $n \in \mathbb{N}$, and the ordering on $\mathbb{N} \subseteq X$ is the usual ordering on $\mathbb{N}$. Since ω does not have an immediate predecessor, the condition that statement $S(1)$ is true and that $S(k)$ implies $S(k+1)$ does not mean that $S(\alpha)$ is true for all $\alpha \in X$. On the other hand, one can use Transfinite Induction to handle such cases. We remark that one usually uses the notation $\omega + 1$ for X.

2.21. Culture: We end this section with a bird's eye view of **ordinal numbers**. We emphasize that the bird in question is flying at a very high altitude and has lousy vision and it is two hours before dawn on a moonless night.

One can define an ordinal number to be an equivalence class of well-ordered sets as follows.

A bijection $f : (A, \leq_A) \to (B, \leq_B)$ between two well-ordered sets is said to be **order-preserving** provided that $a_1 \leq_A a_2$ implies that $f(a_1) \leq_B f(a_2)$.

We shall say that two well-ordered sets $(A, \leq_A)$ and $(B, \leq_B)$ are **equivalent** if there exists an order preserving bijection $f : A \to B$. We write $A \equiv B$ in this case. It is not hard to verify that $\equiv$ is indeed an equivalence relation in the sense that $A \equiv A$, $A \equiv B$ implies $B \equiv A$, and if $A \equiv B$ and $B \equiv C$, then $A \equiv C$.

Given a set $\mathcal{A}$ of well-ordered sets, we define an ordinal number to be an equivalence class of an element of $\mathcal{A}$ under the relation $\equiv$.

Usually, one defines 0 to be $\varnothing$, 1 to be $\{0\} = \{\varnothing\}$, $2 = \{0,1\} = \{\varnothing, \{\varnothing\}\}$, $3 = \{0,1,2\} = \big\{\varnothing, \{\varnothing\}, \{\varnothing, \{\varnothing\}\}\big\}$, etc. We denote the first infinite ordinal by $\omega = \cup_{n \geq 0} n$. We can then define $\omega+1, \omega+2, ...\omega+\omega, \omega+\omega+1, \omega+\omega+2, ...$ and so on.

For ordinals μ, ν, we define $\mu + \nu$ to be (the equivalence class of) $\mu \cup \nu'$, where $\nu' \equiv \nu$ is disjoint from μ and the partial order is given by: $x \leq_{\mu+\nu'} y$ if $x, y \in \mu$ and $x \leq_\mu y$, $x, y \in \nu'$ and $x \leq_{\nu'} y$, or $x \in \mu$ and $y \in \nu'$.

Caution: Let $\omega = \{1, 2, 3, ...\}$ as above. Observe that $1 + \omega = \omega$ under this definition. That is, sticking a "1" before the natural numbers results in an ordered set that still looks like the natural numbers. On the other hand, $\omega + 1 = \{1, 2, 3, ..., \omega\} \neq \omega$, since $\omega + 1$ has a maximum element, which ω does not. Thus addition of ordinals is not commutative. This will not happen when we look at addition of cardinals.

3. Cardinality

3.1. In this section we introduce a notion of *size* of a set A. As we shall see, our notion is based upon the comparison of the size of two sets through embeddings of one into the other. Assuming the Axiom of Choice, we have seen in Section 2.18 that it is always possible to compare two sets in this manner.

3.2. Definition. *Recall that a relation R on a set A is a subset $R \subseteq A \times A$. The relation R is called an **equivalence relation** if the following conditions are met:*

(i) *xRx for all $x \in A$;*
(ii) *if xRy, then yRx (symmetry); and*
(iii) *if xRy and yRz, then xRz.*

If A is a set and R is an equivalence relation on A, then given $x \in A$, the set $\{y \in A : xRy\}$ is called the **equivalence class** of x. Since two equivalence classes are either equal or disjoint, this allows us to partition A as the disjoint union of equivalence classes under this relation.

3.3. Example. Let $A = \mathbb{Z}$ and define xRy if $y - x$ is even. Then R is an equivalence relation. There are exactly two classes for $\mathbb{Z}$ under this relation.

3.4. Definition. *Two sets A and B are said to be **equipotent**, and we write $A \sim B$, if either $A = \varnothing = B$, or there exists a bijection $f : A \to B$.*

3.5. Note that if $\mathcal{F}$ is a collection of sets, then equipotence is an equivalence relation on $\mathcal{F}$. With each equivalence class we shall associate an object which we call its **cardinal number**, and which we shall think of as the *size* of an element of that equivalence class. Technically, a cardinal number $|A|$ might be defined as the equivalence class of $A \in \mathcal{F}$ itself under equipotence. This raises the question: what should $\mathcal{F}$ be? This is a non-trivial question, and beyond the scope of this course. For our purposes, we shall content ourselves with saying that two sets A and B have the same **cardinality** if they are equipotent.

3.6. Definition. *A set A is said to be:*

(i) ***finite*** *if either $A = \varnothing$, and so A has cardinality 0, written $|A| = 0$, or there exists $n \in \mathbb{N}$ and a bijection $f : \{1, 2, ..., n\} \to A$, in which case we write $|A| = n$;*
(ii) ***denumerable*** *if there exists a bijection $g : \mathbb{N} \to A$. We write $|A| = \aleph_0$;*
(iii) ***countable*** *if A is either finite or denumerable;*
(iv) ***uncountable*** *if A is not countable.*

3.7. Example. The set $\mathbb{Z}$ is denumerable, hence countable.

3.8. Proposition.

(i) *If* $A \subseteq \mathbb{N}$, *then* A *is countable.*

(ii) *Any subset of a countable set is countable.*

Proof.

(i) If A is a finite set, then we are done. Otherwise, let k_1 be the smallest element of A, and more generally, for $n > 1$, let k_n be the smallest element of $A \setminus \{k_1, k_2, ..., k_{n-1}\}$. (Here we are using the fact that $\mathbb{N}$ is well-ordered.)

Define $f : \mathbb{N} \to A$ by $f(n) = k_n$. We claim that f is the desired bijection.

To see that f is injective, suppose $m < n$. Then

$$f(n) = k_n \in A \setminus \{k_1, k_2, ..., k_m, ..., k_{n-1}\},$$

so $f(n) = k_n \neq k_m = f(m)$.

To see that f is surjective, suppose that $p \in A$ but $p \notin \{k_n\}_{n=1}^{\infty}$. Since $k_1 < k_2 < k_3 < \cdots$ by construction, there must exist some $j \in \mathbb{N}$ so that $k_j < p < k_{j+1}$. But k_{j+1} is the smallest element of $A \setminus \{k_1, k_2, ..., k_j\}$, so $k_{j+1} \leq p$, a contradiction. Thus f is onto.

We conclude that A is denumerable.

(ii) Suppose that $B \subseteq C$, where C is countable. If C is finite, then so is B and we are done.

Otherwise, C is denumerable, so there exists a bijection $g : \mathbb{N} \to C$. Let $A = \{n \in \mathbb{N} : g(n) \in B\}$. Then A is countable by part (i) above. Let $h = g|_A$ be the restriction of h to A. Then h is a bijection of A onto B (*why?*), and so B is countable.

□

3.9. Proposition. *Let* S *be a non-empty set. The following are equivalent.*

(a) S *is countable.*

(b) *There exists an injection* $f : S \to \mathbb{N}$.

(c) *There exists a surjection* $g : \mathbb{N} \to S$.

Proof.

(a) implies (b): Since S is countable there exists a bijection $f : S \to J$ where $J = I_m := \{1, 2, ..., m\}$ for some $m \in \mathbb{N}$, or $J = \mathbb{N}$. Either way, $J \subseteq \mathbb{N}$, and f is the desired injection.

(b) implies (c): Let $f : S \to \mathbb{N}$ be an injection and $s \in S$. Define

$$\begin{array}{rccl} g: & \mathbb{N} & \to & S \\ & n & \mapsto & \begin{cases} f^{-1}(n) & \text{if } n \in f(S), \\ s & \text{otherwise.} \end{cases} \end{array}$$

Then g is the desired surjection.

(c) implies (a): Define $h : S \to \mathbb{N}$ via $h(s) = \min\{n \in \mathbb{N} : g(n) = s\}$. Then h is an injection from S into $\mathbb{N}$ and a bijection between S and $h(S)$. Since $h(S)$ is countable, so is S.

□

3.10. Example.

(a) Recall that every natural number has a unique factorization as a product of primes (up to the order of the factors). Suppose that A and B are non-empty countable sets. We claim that the Cartesian product $A \times B$ is countable.

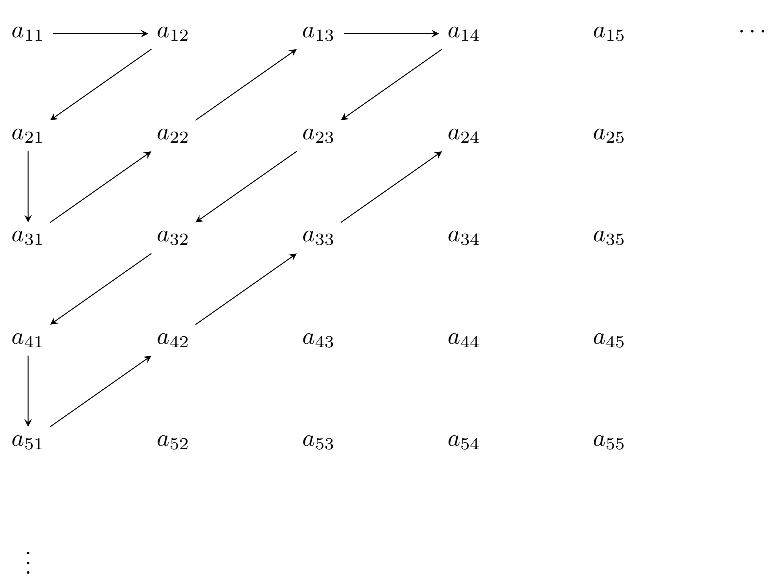
Let $f : A \to \mathbb{N}$ and $g : B \to \mathbb{N}$ be injections. Define

$$\begin{array}{rccc} h: & A \times B & \to & \mathbb{N} \\ & (a, b) & \mapsto & 2^{f(a)} 3^{g(b)}. \end{array}$$

It is routine to verify that h is injective, and hence that $A \times B$ is countable.

(b) Suppose that $\{A_n\}_{n=1}^{\infty}$ is a collection of denumerable sets. Then $A = \cup_{n \geq 1} A_n$ is denumerable.

We shall prove this in the case where $A_n \cap A_m = \varnothing$ if $n \neq m$. The general case follows easily from this (**exercise**). Since A_n is denumerable, we can write $A_n = \{a_{n1}, a_{n2}, a_{n3}, ...\}$ for each $n \geq 1$. Construct a new sequence by following the arrows below. (We have only included the first 13 arrows; there are infinitely many!)



Then $A = \{a_{11}, a_{12}, a_{21}, a_{31}, a_{22}, a_{13}, a_{14}, a_{23}, a_{32}, \ldots\}$. Thus A is countable.

Alternatively, observe that $\mathbb{N} \times \mathbb{N}$ is denumerable by part (a). Let $f : \mathbb{N} \to \mathbb{N} \times \mathbb{N}$ be a bijection. Next, define a surjection $g : \mathbb{N} \times \mathbb{N} \to \cup_{n \geq 1} A_n$ via $g(n, k) := a_{nk}$. Then $g \circ f : \mathbb{N} \to \cup_{n \geq 1} A_n$ is a surjection, and thus $\cup_{n \geq 1} A_n$ is countable, by Proposition 3.9 above.

(c) The set $\mathbb{Q}$ of rational numbers is countable. Indeed, for each $n \geq 1$, let $A_n = \{0/n, 1/n, -1/n, 2/n, -2/n, 3/n, -3/n, ...\}$. Then $\mathbb{Q} = \cup_{n \geq 1} A_n$. By (b) above, $\mathbb{Q}$ is countable.

Recall that between any two distinct rational numbers there are infinitely many rational numbers, and that between any two distinct irrational numbers one can find infinitely many rational numbers. One might be tempted to believe that the set $\mathbb{I}$ of irrational numbers and the set $\mathbb{Q}$ of rational numbers are equipotent. That would be a mistake.

3.11. Theorem. *The set* $\mathbb{R}$ *of real numbers is uncountable.*

Proof. Using Cantor's diagonal process, we shall prove that $(0,1)$ is uncountable. By Proposition 3.8, this implies that $\mathbb{R}$ is uncountable.

If $(0,1)$ were countable, then we could write $(0,1) = \{x_n\}_{n=1}^{\infty}$, where

$$\begin{aligned} x_1 &= 0.x_{11}\ x_{12}\ x_{13}\ x_{14}... \\ x_2 &= 0.x_{21}\ x_{22}\ x_{23}\ x_{24}... \\ x_3 &= 0.x_{31}\ x_{32}\ x_{33}\ x_{34}... \\ x_4 &= 0.x_{41}\ x_{42}\ x_{43}\ x_{44}... \\ &\vdots \end{aligned}$$

in decimal form, so that x_{ij} is an integer between 0 and 9.

Let $y = 0.y_1\ y_2\ y_3\ y_4...$ where $y_n = 7$ if $x_{nn} \in \{0,1,2,3,4,5\}$ and $y_n = 3$ if $x_{nn} \in \{6,7,8,9\}$. Then $y \neq x_n$, since $y_n \neq x_{nn}$ for all $n \geq 1$.

In other words, $y \in (0,1)$, but $y \notin \{x_n\}_{n=1}^{\infty}$, a contradiction.

□

3.12. Although $\mathbb{N} \subset \mathbb{Z} \subset \mathbb{Q}$, all three sets are equipotent, and we write $|\mathbb{N}| = |\mathbb{Z}| = |\mathbb{Q}| = \aleph_0$. Thus it is not enough for one set to properly contain another in order to have greater cardinality (although this does work for finite sets). Nevertheless, our intuition tells us that if $f : A \to B$ is an injection, then we would expect A to be no larger than B.

Letting $|A|$ denote the cardinality of A, we shall *define* $|A| \leq |B|$ to mean that there exists an injection $f : A \to B$. Then $|A| < |B|$ means that $|A| \leq |B|$, but that $|A| \neq |B|$. That is, although an injection $f : A \to B$ exists, there is no bijection $g : A \to B$. This agrees with the usual notion of size when the sets are finite.

3.13. Theorem. *Let* A, B, *and* C *be sets.*

(i) *If* $A \subseteq B$, *then* $|A| \leq |B|$.
(ii) $|A| \leq |A|$.
(iii) *If* $|A| \leq |B|$ *and* $|B| \leq |C|$, *then* $|A| \leq |C|$.
(iv) *If* $m, n \in \mathbb{N}$ *and* $m \leq n$, *then* $|\{1, 2, ..., m\}| \leq |\{1, 2, ..., n\}|$.
(v) *If* E *is finite, then* $|E| < \aleph_0$.

Proof. Exercise.

□

3.14. Let us denote $|\mathbb{R}|$ by c (for *continuum*). Since $\mathbb{N} \subseteq \mathbb{R}$, $\aleph_0 \le c$. In fact, since $\mathbb{R}$ is uncountable, $\aleph_0 < c$. That is, there are at least two infinite cardinals. In fact, there are many, many more.

3.15. Theorem. *For any set X, $|X| < |\mathcal{P}(X)|$, where $\mathcal{P}(X)$ is the power set of X.*

Proof. First note that the map

$$\begin{array}{rccc} f: & X & \to & \mathcal{P}(X) \\ & x & \mapsto & \{x\} \end{array}$$

is injective, and so $|X| \le |\mathcal{P}(X)|$.

Next, suppose that $g : X \to \mathcal{P}(X)$ is any surjective map. Given $x \in X$, either $x \in g(x)$, or $x \notin g(x)$. Let $T = \{x \in X : x \notin g(x)\}$. Since g is surjective, there exists $z \in X$ so that $g(z) = T$.

If $z \in T$, then $z \in g(z)$, so $z \notin T$, a contradiction.

If $z \notin T$, then $z \notin g(z)$, so $z \in T$, again a contradiction.

Thus g can not be surjective, and *a fortiori* it can not be bijective, so $|X| < |\mathcal{P}(X)|$.

□

3.16. It follows that $\aleph_0 = |\mathbb{N}| < |\mathcal{P}(\mathbb{N})| < |\mathcal{P}(\mathcal{P}(\mathbb{N}))| < |\mathcal{P}(\mathcal{P}(\mathcal{P}(\mathbb{N})))| < \cdots$, and that there exist infinitely many infinite cardinal numbers. Where does $c = |\mathbb{R}|$ fit? We have seen that $\aleph_0 < c$, but does there exist an infinite cardinal λ such that $\aleph_0 < \lambda < c$? Writing $\aleph_1$ for the first cardinal bigger than $\aleph_0$, the question becomes: is $c = \aleph_1$?

The conjecture that there does *not* exist such a λ is known as the *Continuum Hypothesis* and is due to Georg Cantor. In 1938, Kurt Gödel proved that the Continuum Hypothesis does not contradict the usual axioms of set theory. In 1963, Paul Cohen prove that the negation of the Continuum Hypothesis is also consistent with the usual axioms of set theory.

Whereas the Axiom of Choice is freely used (but cited) by the majority of mathematicians, it is not standard to assume the Continuum Hypothesis nor its negation. In the few instances where it is used, it *must* be explicitly stated that one is using it. If it is possible to prove something without assuming the Continuum Hypothesis, then it is generally considered best to prove it without using it.

4. Cardinal Arithmetic

4.1. In this section we shall briefly examine sums, products and powers of cardinal numbers. Finite numbers do not provide the best intuition, since we don't expect numbers other than 0 and 1 to satisfy $\lambda^2 = \lambda$, for example. This equality will be satisfied by infinite cardinals, as we shall soon see. We begin with an extremely useful result which is the usual tool for proving that two sets have the same cardinality. Although the result looks obvious, its proof is surprisingly non-obvious.

4.2. Theorem. ***The Schröder-Bernstein Theorem.*** *Let A and B be sets. If $|A| \le |B|$ and $|B| \le |A|$, then $|A| = |B|$.*

Proof. • **Step One:** If Z is any set and $\varphi : \mathcal{P}(Z) \to \mathcal{P}(Z)$ is increasing in the sense that $X \subseteq Y \subseteq Z$ implies that $\varphi(X) \subseteq \varphi(Y)$, then φ has a **fixed point**; that is, there exists $T \subseteq X$ such that $\varphi(T) = T$.

Indeed, let $T = \cup\{X \subseteq Z : X \subseteq \varphi(X)\}$. If $X \subseteq Z$ and $X \subseteq \varphi(X)$, then $X \subseteq T$ and so $\varphi(X) \subseteq \varphi(T)$. That is, $X \subseteq Z$ and $X \subseteq \varphi(X)$ implies $X \subseteq \varphi(T)$, and thus

$$T = \cup\{X \subseteq Z : X \subseteq \varphi(X)\} \subseteq \varphi(T).$$

But then $\varphi(T) \subseteq Z$ and $\varphi(T) \subseteq \varphi(\varphi(T))$, so that $\varphi(T)$ is one of the sets appearing in the definition of T - i.e. $\varphi(T) \subseteq T$.

Together, these imply that $\varphi(T) = T$. (We remark that it is entirely possible that $T = \varnothing$.)

• **Step Two:** Given sets A, B as above and injections $\kappa : A \to B$ and $\lambda : B \to A$, define

$$\begin{array}{rccc} \varphi : & \mathcal{P}(A) & \to & \mathcal{P}(A) \\ & X & \mapsto & A \setminus \lambda[B \setminus \kappa(X)]. \end{array}$$

Suppose $X \subseteq Y \subseteq A$. Then $\kappa(X) \subseteq \kappa(Y)$.

Hence

$$\begin{aligned} B \setminus \kappa(X) &\supseteq B \setminus \kappa(Y), \text{ so} \\ \lambda(B \setminus \kappa(X)) &\supseteq \lambda(B \setminus \kappa(Y)), \text{ which implies} \\ A \setminus \lambda(B \setminus \kappa(X)) &\subseteq A \setminus \lambda(B \setminus \kappa(Y)), \text{ which in turn implies} \\ \varphi(X) &\subseteq \varphi(Y). \end{aligned}$$

• **Step Three:** By Steps One and Two, there exists $T \subseteq A$ such that $T = \varphi(T) = A \setminus \lambda[B \setminus \kappa(T)]$.

Define

$$\begin{array}{rccc} f : & A & \to & B \\ & a & \mapsto & \begin{cases} \kappa(a) & \text{if } a \in T, \\ \lambda^{-1}(a) & \text{if } a \in A \setminus T. \end{cases} \end{array}$$

Observe that λ is a bijection between $B \setminus \kappa(T)$ and $A \setminus T$, and that κ is a bijection between T and $\kappa(T)$, so that f is a bijection between A and B.

□

Using the Schröder-Bernstein Theorem, we can prove the following:

4.3. Theorem. $c = |\mathbb{R}| = |\mathcal{P}(\mathbb{N})|$.

Proof. By the Schröder-Bernstein Theorem, it suffices to prove that $c \le |\mathcal{P}(\mathbb{N})|$ and $|\mathcal{P}(\mathbb{N})| \le c$.

To see that $|\mathcal{P}(\mathbb{N})| \le c$, define

$$\begin{array}{rccc} f: & \mathcal{P}(\mathbb{N}) & \to & \mathbb{R} \\ & A & \mapsto & 0.a_1\ a_2\ a_3\ ... \end{array}$$

where $a_n = 0$ if $n \notin A$ and $a_n = 1$ if $n \in A$. It is not hard to verify that f is injective.

To see that $c \le |\mathcal{P}(\mathbb{N})|$, first note that $|\mathbb{N}| = |\mathbb{Q}|$ and hence (**exercise**) $|\mathcal{P}(\mathbb{N})| = |\mathcal{P}(\mathbb{Q})|$. Next, let

$$\begin{array}{rccc} g: & \mathbb{R} & \to & \mathcal{P}(\mathbb{Q}) \\ & x & \mapsto & \{y \in \mathbb{Q} : y < x\}. \end{array}$$

If $x_1 < x_2$ in $\mathbb{R}$, then there exists $q \in \mathbb{Q}$ so that $x_1 < q < x_2$, and so $q \notin g(x_1)$ but $q \in g(x_2)$, showing that $g(x_1) \ne g(x_2)$ and so g is injective.

□

4.4. Definition. *Let α, β be cardinal numbers.*

*The **sum** $\alpha + \beta$ of α and β is defined to be the cardinal $|A \cup B|$, where A and B are* disjoint *sets such that $|A| = \alpha$ and $|B| = \beta$.*

*The **product** $\alpha\ \beta$ of α and β is the cardinal number $|A \times B|$, where A and B are sets with $|A| = \alpha$ and $|B| = \beta$.*

*The **power** β^α is defined as $|B^A|$, where A and B are sets with $|A| = \alpha$ and $|B| = \beta$.*

4.5. In "**ordinal arithmetic**", one defines $0 = \varnothing$, $1 = \{0\} = \{\varnothing\}$, $2 = \{0, 1\} = \{\varnothing, \{\varnothing\}\}$, $3 = \{0, 1, 2\}$, etc. In a mild abuse of notation, the same notation is used to denote the corresponding cardinal.

The proof of Theorem 4.3 shows that if A is any set and B is a subset of A, then B corresponds to a unique function $f_B : A \to \{0, 1\}$ given by $f_B(a) = 0$ if $a \notin B$, and $f_B(a) = 1$ if $a \in B$. (This is often called the **characteristic function** or the **indicator function** of B in A.)

The map $B \mapsto f_B$ is a bijection between $\mathcal{P}(A)$ and $\{0, 1\}^A = 2^A$. Thus $|\mathcal{P}(A)| = |2^A|$, and in particular, $|2^{\mathbb{N}}| = |\mathcal{P}(\mathbb{N})| = c$. But, from the definition of cardinal powers, this says that $c = |2^{\mathbb{N}}| = |2|^{|\mathbb{N}|} = 2^{\aleph_0}$.

4.6. Lemma.

(i) *If A is an infinite set, then A contains a denumerable subset.*
(ii) *If A is an infinite set and B is a finite set, then $|A| + |B| = |A|$.*

Proof.

(i) Since $A \ne \varnothing$, there exists $x_1 \in A$. Then $A \setminus \{x_1\} \ne \varnothing$, otherwise A would be finite.

In general, for $n \ge 1$, having chosen $\{x_1, x_2, ..., x_n\} \subseteq A$, we know that $A \setminus \{x_1, x_2, ..., x_n\} \ne \varnothing$, so we can find $x_{n+1} \in A \setminus \{x_1, x_2, ..., x_n\}$. (Doing this for all $n \geq 1$ requires the Axiom of Choice - or at least a weak version of it.)

The function $f : \mathbb{N} \to A$ defined via $f(n) = x_n$ is an injection, and it is a bijection between $\mathbb{N}$ and $B = \operatorname{ran} f = \{x_n\}_{n=1}^{\infty}$. Thus $|B| = \aleph_0$ and B is a denumerable subset of A.

(ii) Let A be an infinite set, and let $D \subseteq A$ be a denumerable subset of A. Suppose that $B = \{b_1, b_2, ..., b_n\}$. (We may suppose that $B \cap A = \varnothing$ (*why?*)). Define a map

$$\begin{array}{rccc} f: & A \cup B & \to & A \\ & z & \mapsto & z \text{ if } z \in A \setminus D \\ & b_i & \mapsto & d_i \text{ if } 1 \leq i \leq n, \\ & d_k & \mapsto & d_{k+n}, \text{ for all } k \geq 1. \end{array}$$

Then f is an bijection of $A \cup B$ onto A, and so $|A| + |B| = |A|$.

□

4.7. Theorem. *Let α, β, and γ be cardinal numbers. Then*

(i) *$\alpha + \beta$ is well-defined. That is, if $|A| = |C|$, $|B| = |D|$ and $A \cap B = \varnothing = C \cap D$, then $|A \cup B| = |C \cup D|$.*

(ii) *$\alpha + \beta = \beta + \alpha$ and $(\alpha + \beta) + \gamma = \alpha + (\beta + \gamma)$.*

(iii) *If β is infinite and $\alpha \leq \beta$, then $\alpha + \beta = \beta$.*

Proof.

(i) **Exercise.**

(ii) **Exercise.**

(iii) The case where $\alpha < \aleph_0$ is Lemma 4.6.

First let us show that $\beta + \beta = \beta$. Choose a set B with $|B| = \beta$. Then $B \times 2 = (B \times \{0\}) \cup (B \times \{1\})$ is the union of two disjoint sets equipotent with B, so it suffices to show that $|B \times 2| = |B| = \beta$.

Let $\mathcal{F} = \{(X, f) : X \subseteq B$ and $f : X \to X \times 2$ is a bijection$\}$, partially ordered by $(X_1, f_1) \leq (X_2, f_2)$ if $X_1 \subseteq X_2$ and $f_2|_{X_1} = f_1$.

If $X \subseteq B$ is denumerable, then $|X \times 2| = |X| = \aleph_0$ by Example 3.10, and hence $\mathcal{F} \neq \varnothing$.

Suppose that $\mathcal{C} = \{(X_\alpha, f_\alpha)\}_{\alpha \in \Lambda}$ is a chain in $\mathcal{F}$.

Let $X = \cup_{\alpha \in \Lambda} X_\alpha$. For $x \in X$, choose $\alpha \in \Lambda$ such that $x \in X_\alpha$. Define $f(x) = f_\alpha(x)$. Then f is well-defined (*why?*). Moreover, (X, f) is an upper bound for $\mathcal{C}$; i.e. $f : X \to X \times 2$ is a bijection (**exercise**).

By Zorn's Lemma, $\mathcal{F}$ has a maximal element (Y, g). We claim that $B \setminus Y$ is finite. Otherwise, choose a denumerable set $Z \subseteq B \setminus Y$. Since $|Z| = |Z \times 2| = \aleph_0$, there exists a bijection $h : Z \to Z \times 2$. Define a bijection

$$\begin{array}{rccc} h: & Y \cup Z & \to & (Y \cup Z) \times 2 \\ & w & \mapsto & \begin{cases} g(w) & \text{if } w \in Y, \\ h(w) & \text{if } w \in Z. \end{cases} \end{array}$$

Then $(Y \cup Z, h) > (Y, g)$, contradicting the maximality of (Y, g).

This shows that $B \setminus Y$ is finite. Hence $|Y| = |B| = \beta$, and so $\beta = |Y| = |Y \times 2| = |Y| + |Y| = \beta + \beta$.

Finally, in general we have $\beta \le \alpha + \beta \le \beta + \beta = \beta$, so that by the Schröder-Bernstein Theorem, $\alpha + \beta = \beta$.

□

4.8. Theorem. *Let $\alpha, \beta, \gamma, \delta$ be cardinal numbers. Then*

(i) *$\alpha \cdot \beta$ is well-defined. That is, if $|A| = |C|$, $|B| = |D|$, then $|A \times B| = |C \times D|$.*
(ii) *$\alpha \cdot \beta = \beta \cdot \alpha$; $\alpha(\beta \cdot \gamma) = (\alpha \cdot \beta)\gamma$; and $\alpha(\beta + \gamma) = \alpha \cdot \beta + \alpha \cdot \gamma$.*
(iii) *$0 \cdot \alpha = 0$.*
(iv) *If $\alpha \le \beta$ and $\gamma \le \delta$, then $\alpha \cdot \gamma \le \beta \cdot \delta$.*
(v) *$\alpha \cdot \alpha = \alpha$ if α is infinite.*

Proof.

(i) **Exercise.**
(ii) **Exercise.**
(iii) **Exercise.**
(iv) **Exercise.**
(v) Suppose that $|A| = \alpha$. Let $\mathcal{F} = \{(X, f) : X \subseteq A, f : X \to X \times X \text{ is a bijection}\}$, partially ordered by $(X_1, f_1) \le (X_2, f_2)$ if $X_1 \subseteq X_2$ and $f_2|_{X_1} = f_1$.

Since A is infinite, it contains a denumerable set X. Now by Example 3.10, if $X \subseteq A$ is denumerable, then there exists a function f so that $(X, f) \in \mathcal{F}$, and thus $\mathcal{F} \ne \varnothing$. By Zorn's Lemma (as before), there exists a maximal element (Y, g) in $\mathcal{F}$.

Then $|Y| \cdot |Y| = |Y|$, so it suffices to show that $|Y| = \alpha$.

Assume that $|Y| < \alpha$. Since $\alpha = |Y| + |A \setminus Y|$, it follows that $|A \setminus Y| = \alpha$, and so $|Y| < |A \setminus Y|$. Thus $A \setminus Y$ has a subset Z with $|Z| = |Y|$. Then $Y \times Z, Z \times Y$ and $Z \times Z$ are disjoint, infinite sets with cardinality $|Y|$, and so

$$\begin{aligned} |(Y \times Z) \cup (Z \times Y) \cup (Z \times Z)| &= |Y \times Z| + |Z \times Y| + |Z \times Z| \\ &= (|Y| \cdot |Y|) + (|Y| \cdot |Y|) + (|Y| \cdot |Y|) \\ &= |Y| + |Y| + |Y| \\ &= |Y| \\ &= |Z|. \end{aligned}$$

Thus there exists a bijection $h : Z \to (Y \times Z) \cup (Z \times Y) \cup (Z \times Z)$.

Define the map

$$\begin{aligned} m : \quad Y \cup Z &\to (Y \cup Z) \times (Y \cup Z) \\ x &\mapsto \begin{cases} g(x) & \text{if } x \in Y \\ h(x) & \text{if } x \in Z \end{cases} . \end{aligned}$$

Then m is a bijection and so $(Y \cup Z, m) \in \mathcal{F}$ with $(Y, g) < (Y \cup Z, m)$, contradicting the maximality of (Y, g) in $\mathcal{F}$.

This contradiction shows that $|Y| = \alpha$, and we are done, as $g : Y \to Y \times Y$ is the bijection which implies that $\alpha = \alpha \cdot \alpha$.

□

4.9. Remark. The perspicacious reader (and just about no one else) will have picked up a very subtle use of the Axiom of Choice in the above proof. You will recall that we said that with $Y \subseteq A$, $\alpha = |A| = |Y| + |A \setminus Y|$ and $|Y| < \alpha$ implies that $|A \setminus Y| = \alpha$.

In fact, how do we know this? Theorem 4.7 assures us that if $|Y| \le |A \setminus Y|$ or if $|A \setminus Y| \le |Y|$, then

$$\alpha = |A| = |Y| + |A \setminus Y| = \max(|Y|, |A \setminus Y|),$$

and thus $|A \setminus Y| = \alpha$. But how do we know that $|Y|$ and $|A \setminus Y|$ are comparable? In fact, for that we must go back to Remark 2.18, and the statement that given any two sets V and W, either there exists an injection from V into W (i.e. $|V| \le |W|$), or there exists an injection from W into V (i.e. $|W| \le |V|$). In other words, any two cardinal numbers are comparable! As pointed out in that Remark, however, this assumption is equivalent to the Axiom of Choice.

We thank Y. Zhang for pointing this out.

4.10. Theorem. *Let α, β and γ be cardinal numbers. Then*

(i) α^β *is well-defined. That is, if A_1, A_2, B_1, B_2 are sets with $|A_1| = \alpha = |A_2|$ and $|B_1| = \beta = |B_2|$, then $|A_1^{B_1}| = |A_2^{B_2}|$.*

(ii) $(\alpha^\beta)^\gamma = \alpha^{(\beta\gamma)}$.

(iii) $\alpha^{(\beta+\gamma)} = \alpha^\beta \; \alpha^\gamma$.

Proof.

(i) **Exercise.**

(ii) Let A, B and C be sets with $|A| = \alpha$, $|B| = \beta$ and $|C| = \gamma$. We must show that $|(A^B)^C| = |A^{B\times C}|$.

Now if $f \in A^{B\times C}$, then for each $c \in C$, the function f_c given by $f_c(b) := f(b,c)$ defines an element of A^B. Define $\varphi_f : C \to A^B$ by $\varphi_f(c) = f_c$. Then the correspondence

$$\begin{array}{cccc} \Phi : & A^{B\times C} & \to & (A^B)^C \\ & f & \mapsto & \varphi_f \end{array}$$

is a bijection. Indeed, if $\Phi(f) = \Phi(g)$ for $f, g \in A^{B\times C}$, then $\varphi_f = \varphi_g$, and so for every $c \in C$, $f_c = \varphi_f(c) = \varphi_g(c) = g_c$. But $f_c = g_c$ for all $c \in C$ implies that $f(b,c) = f_c(b) = g_c(b) = g(b,c)$ for all $b \in B$ and for all $c \in C$, so that $f = g$. This shows that Φ is injective.

Given $\tau \in (A^B)^C$, we see that $\tau(c) \in A^B$ for all $c \in C$, and so we define $f : B \times C \to A$ via $f(b,c) = [\tau(c)](b)$. It is clear that $f \in A^{B\times C}$ and that $\Phi(f) = \tau$, so that Φ is onto. Finally, since Φ is a bijection, $|A^{B\times C}| = |(A^B)^C|$, completing the proof.

(iii) Now suppose that $B \cap C = \varnothing$. We must show that $|A^{B\cup C}| = |A^B \times A^C|$. But every $f : B \cup C \to A$ is defined by its restrictions to B and C, so we are done.

□

4.11. Example.

(a) $\aleph_0 + \aleph_0 = \aleph_0$.
(b) $\aleph_0 + c = c$.
(c) $c \cdot c = c$.
(d) $c^{\aleph_0} = (2^{\aleph_0})^{\aleph_0} = 2^{(\aleph_0 \cdot \aleph_0)} = 2^{\aleph_0} = c$.

5. Appendix

In this Appendix we shall provide a proof of the equivalence of the Axiom of Choice, Zorn's Lemma and the Well-Ordering Principle. We begin with the definition of an **initial segment**, which will be required in the proof.

5.1. Definition. *Let $(X, \leq)$ be a poset, $C \subseteq X$ be a chain in X and $d \in C$. We define*

$$P(C, d) = \{c \in C : c < d\}.$$

*An **initial segment** of C is a subset of the form $P(C, d)$ for some $d \in C$.*

5.2. Example.

(a) For each $r \in \mathbb{R}$, $(-\infty, r)$ is an initial segment of $(\mathbb{R}, \leq)$.
(b) For each $n \in \mathbb{N}$, $\{1, 2, ..., n\}$ is an initial segment of $\mathbb{N}$.

5.3. Theorem. *The following are equivalent:*

(i) *The Axiom of Choice (AC): given a non-empty collection $\{X_\lambda\}_{\lambda \in \Lambda}$ of non-empty sets, $\prod_{\lambda \in \Lambda} X_\lambda \neq \varnothing$.*
(ii) *Zorn's Lemma (ZL): Let $(Y, \leq)$ be a poset. Suppose that every chain $\mathcal{C} \subseteq Y$ has an upper bound. Then Y has a maximal element.*
(iii) *The Well-Ordering Principle (WO): Every non-empty set Z admits a well-ordering.*

Proof.

(i) implies (ii): This is the most delicate of the three implications. We shall argue by contradiction.

Suppose that $(X, \leq)$ is a poset such that every chain in X is bounded above, but that X no maximal elements. Given a chain $C \subseteq X$, we can find an upper bound u_C for C. Since u_C is not a maximal element, we can find $v_C \in X$ with $u_C < v_C$. We shall refer to such an element v_C as a **strict upper bound** for C.

By the Axiom of Choice, for each chain C in X, we can choose a strict upper bound $f(C)$. If $C = \varnothing$, we arbitrarily select $x_0 \in X$ and set $f(\varnothing) = x_0$.

We shall say that a subset $A \subseteq X$ satisfies **property L** if

(I) The partial order $\leq$ on X when restricted to A is a well-ordering of A, and
(II) for all $x \in A$, $x = f(P(A, x))$.

• **Claim 1:** *if $A, B \subseteq X$ satisfy property L and $A \neq B$, then either A is an initial segment of B, or B is an initial segment of A.*

Without loss of generality, we may assume that $A \setminus B \neq \varnothing$. Let

$$x = \min\{a \in A : a \notin B\}.$$

Note that x exists because A is well-ordered. Then $P(A,x) \subseteq B$. We shall argue that $B = P(A,x)$. If not, then $B \setminus P(A,x) \neq \varnothing$, and using the well-orderedness of B,

$$y = \min \{b \in B : b \notin P(A,x)\}$$

exists. Thus $P(B,y) \subseteq P(A,x)$.

Let $z = \min (A \setminus P(B,y))$. Then $z \leq x = \min (A \setminus B)$.

- **Subclaim 1:** $P(A,z) = P(B,y)$.
  By definition, $P(A,z) \subseteq P(B,y)$.
  To obtain the reverse inclusion, we first argue that if $t \in P(B,y) = A \cap P(B,y)$, then $P(A,t) \cup \{t\} \subseteq P(B,y)$. By hypothesis, $t \in P(B,y)$, so suppose that $u \in P(A,t)$. Now $t \in P(B,y) \subseteq P(A,x)$, so $u < t < x$ implies that $u \in P(A,x)$. In other words, $P(A,t) \subseteq P(A,x) \subseteq B$. But then $u \in B$ and $u < t < y$ implies that $u \in P(B,y)$.
  We now have that if $s \in P(B,y)$, then $P(A,s) \cup \{s\} \subseteq P(B,y) \subseteq P(A,x) \subseteq A$. This forces $s < z := \min (A \setminus P(B,y))$, so that $s \in P(A,z)$.
  Together, we find that $P(B,y) \subseteq P(A,z) \subseteq P(B,y)$, which proves the subclaim.

Returning to the proof of the claim, we now have that $z = f(P(A,z)) = f(P(B,y)) = y$. But $y \in B$, so $y \neq x$. Hence $z < x$. Thus $y = z \in P(A,x)$, contradicting the definition of y. We deduce that $P(A,x) = B$, and hence that B is an initial segment of A, thereby proving our claim.

Suppose that $A \subseteq X$ has property L, and let $x \in A$. It follows from the above argument that given $y < x$, either $y \in A$ or y does not belong to any set B with property L.

Let $V = \cup\{A \subseteq X : A \text{ has property } L\}$.

- **Claim 2:** *We claim that if $w = f(V)$, then $V \cup \{w\}$ has property L.*

Suppose that we can show this. Then $V \cup \{w\} \subseteq V$, so $w \in V$, a contradiction. This will complete the proof.

- **Subclaim 2a:** *First we show that V itself has property L.* We must show that V is well-ordered, and that for all $x \in V$, $x = f(P(V,x))$.

(a) V is well-ordered.
Let $\varnothing \neq B \subseteq V$. Then there exists $A_0 \subseteq X$ so that A_0 has property L and $B \cap A_0 \neq \varnothing$. Since A_0 is well-ordered and $\varnothing \neq B \cap A_0 \subseteq A_0$, $m := \min(B \cap A_0)$ exists. We claim that $m = \min(B)$.
Suppose that $y \in B$. Then there exists $A_1 \subseteq X$ so that A_1 has property L and $y \in A_1$. Now, both A_0 and A_1 have property L:
  - ⋄ if $A_0 = A_1$, then $m = \min(B \cap A_1)$, so $m \leq y$.
  - ⋄ if $A_0 \neq A_1$, then either

- A_0 is an initial segment of A_1, so $A_0 = P(A_1, d)$ for some $d \in A_1$. Then

  $$m = \min(B \cap A_0) = \min(B \cap A_1),$$

  since $r \in A_1 \setminus A_0$ implies that $m < d \le r$. Hence $m \le y$;, or
- A_1 is an initial segment of A_0, say $A_1 = P(A_0, d) \subseteq A_0$ for some $d \in A_0$. Then

  $$m = \min(B \cap A_0) \le \min(B \cap A_1).$$

  Hence $m \le y$.

In both cases we see that $m \le y$. Since $y \in B$ was arbitrary, $m = \min(B)$.
Thus, any non-empty subset B of V has a minimum element, and so V is well-ordered.

(b) Let $x \in V$. Then there exists $A_2 \subseteq X$ with property L so that $x \in A_2$. Then $x = P(A_2, x)$. Suppose that $y \in V$ and $y < x$. Then there exists $A_3 \subseteq X$ with property L so that $y \in A_3$. Since A_2 and A_3 both have property L, either
- $A_2 = A_3$, and so $y \in A_2$; or
- $A_2 = P(A_3, d)$ for some $d \in A_3$. Since $x \in A_2$, $P(A_2, x) = P(A_3, x)$ and therefore $y \in A_2$; or
- $A_3 = P(A_2, d)$ for some $d \in A_2$. Then $y \in A_3$ implies that $y \in A_2$.

In any of these three cases, $y \in A_2$. Hence $P(V, x) \subseteq P(A_2, x)$. Since $A_2 \subseteq V$, we have that $P(A_2, x) \subseteq P(V, x)$, whence $P(A_2, x) = P(V, x)$. But then

$$x = f(P(A_2, x)) = f(P(V, x)).$$

By (a) and (b), V has property L.

We now return to the proof of Claim 2. That is, we prove that if $w = f(V)$, then $V \cup \{w\}$ has property L.

(I) $V \cup \{w\}$ is well-ordered.
We know that V is well-ordered by part (a) above. Suppose that $\varnothing \ne B \subseteq V \cup \{w\}$. If $B \cap V \ne \varnothing$, then by (a) above, $m := \min(B \cap V)$ exists. Clearly $m \in V$ implies $m \le f(V) = w$, so $m = \min(B \cap (V \cup \{w\}))$.
If $\varnothing \ne B \subseteq V \cup \{w\}$ and $B \cap V = \varnothing$, then $B = \{w\}$, and so $w = \min(B)$ exists.
Hence $V \cup \{w\}$ is well-ordered.

(II) Let $x = V \cup \{w\}$. If $x \in V$, then $x = f(P(V, x))$ by part (a). If $x = w$, then

$$P(V \cup \{w\}, x) = P(V \cup \{w\}, w) = V,$$

so $x = w = f(V) = f(P(V \cup \{w\}, x))$.

By (I) and (II), $V \cup \{w\}$ has property L. As we saw in the statement following Claim 2, this completes the proof that the Axiom of Choice implies Zorn's Lemma. Now let us never speak of this again.

(ii) implies (iii): Let $X \neq \varnothing$ be a set. It is clear that every finite subset $F \subseteq X$ can be well-ordered. Let $\mathcal{A}$ denote the collection of pairs $(Y, \leq_Y)$, where $Y \subseteq X$ and $\leq_Y$ is a well-ordering of Y. For $(A, \leq_A)$, $(B, \leq_B) \in \mathcal{A}$, observe that A is an **initial segment** of B if the following two conditions are met:

- $A \subseteq B$ and $a_1 \leq_A a_2$ implies that $a_1 \leq_B a_2$;
- if $b \in B \setminus A$, then $a \leq_B b$ for all $a \in A$.

Let us partially order $\mathcal{A}$ by setting $(A, \leq_A) \leq (B, \leq_B)$ if A is an initial segment of B. Let $\mathcal{C} = \{C_\lambda\}_{\lambda \in \Lambda}$ be a chain in $\mathcal{A}$.

Then (**exercise**): $\cup_{\lambda \in \Lambda} C_\lambda$ is an upper bound for $\mathcal{C}$.

By Zorn's Lemma, $\mathcal{A}$ admits a maximal element, say $(M, \leq_M)$. We claim that $M = X$. Suppose otherwise. Then we can choose $x_0 \in X \setminus M$ and set $M_0 = M \cup \{x_0\}$. define a partial order on M_0 via: $x \leq_{M_0} y$ if either (a) $x, y \in M$ and $x \leq_M y$, or (b) x is arbitrary and $y = x_0$. Then $(M_0, \leq_{M_0})$ is a well-ordered set and $(M, \leq_M) < (M_0, \leq_{M_0})$, a contradiction of the maximality of $(M, \leq_M)$. Thus $M = X$ and $\leq_M$ is a well- ordering of X.

(iii) implies (i): Suppose that $\{X_\lambda\}_{\lambda \in \Lambda}$ is a non-empty collection of non-empty sets. Let $X = \cup_{\lambda \in \Lambda} X_\lambda$. By hypothesis, X admits a well-ordering $\leq_X$. Since each $\varnothing \neq X_\lambda \subseteq X$, it has a minimum element relative to the ordering on X. Define a choice function f by setting $f(\lambda)$ to be this minimum element of X_λ for each $\lambda \in \Lambda$.

□

We include a proof of an exercise mentioned earlier in the notes:

5.4. Proposition. *The following are equivalent:*

(a) *The Axiom of choice: if $\Lambda \neq \varnothing$ and for each $\lambda \in \Lambda$ there exists a non-empty set X_λ, then*

$$\Pi_{\lambda \in \Lambda} X_\lambda \neq \varnothing.$$

(b) *If $\varnothing \neq \varnothing$, then there exists a function*

$$g : \mathcal{P}(X) \setminus \{\varnothing\} \to X$$

such that $g(Y) \in Y$ for all $Y \subseteq X$.

Proof.

(a) implies (b).

Suppose (a) holds. Let $\varnothing \neq X$ be a set and set $\Lambda = \mathcal{P}(X) \setminus \{\varnothing\}$. For each $Y \in \Lambda$, set $Z_Y = Y \neq \varnothing$.

By the Axiom of Choice, there exists a choice function

$$f \in \Pi_{Y \in \Lambda} Z_Y = \Pi_{Y \in \Lambda} Y.$$

But then $f(Y) \in Z_Y = Y$ for each $Y \in \Lambda = \mathcal{P}(X) \setminus \{\varnothing\}$.

That is, (b) holds.

(b) implies (a).

Suppose that (b) holds.

Let $\varnothing \neq \Lambda$ be a set and suppose that X_λ is a non-empty set for each $\lambda \in \Lambda$. Let $Y = \cup_{\lambda \in \Lambda} X_\lambda$.

By hypothesis, there exists a function $g : \mathcal{P}(Y) \setminus \{\varnothing\} \to Y$ so that $g(W) \in W$ for all $W \in \mathcal{P}(Y) \setminus \{\varnothing\}$. In particular, each $X_\lambda \in \mathcal{P}(Y) \setminus \{\varnothing\}$, and so $g(X_\lambda) \in X_\lambda$ for all $\lambda \in \Lambda$.

Define $f(\lambda) = g(X_\lambda)$, $\lambda \in \Lambda$. Then f is a choice function, so (a) holds.

□

Culture.

(a) The basic axioms of set theory are referred to as the **Zermelo-Fraenkel Axioms**, or (ZF).

Gödel proved that the Axiom of Choice is consistent with (ZF), but that (ZF) does not by itself imply the Axiom of Choice. Cohen then developed the theory of "forcing" to prove that (ZF) plus the *negation* of the Axiom of Choice is also consistent.

(b) It is known that the Riemann hypothesis is true in (ZF) if and only if it is true in (ZFC), namely (ZF) plus the Axiom of Choice.

(c) The generalized Continuum hypothesis (GCH) is known to be independent of (ZFC), however (ZF) plus (GCH) together imply the Axiom of Choice (AC).

(d) Tarski tried to publish the result which says that the Axiom of Choice (AC) is equivalent to the assertion that $|A| = |A \times A|$ whenever A is infinite in the Comptes Rendus. It was not accepted. Fréchet said that the equivalence of two true statements is not something new, while Lebesgue said that any implication between two false propositions is of no interest.

Exercises for Chapter 1.

Exercise 1.1.

Let X be a set with $0 \le n < \infty$ elements. Prove that the power set $\mathcal{P}(X)$ has 2^n elements.

Exercise 1.2.

Given a universe $\varnothing \ne X$ and a collection of subsets X_α, $\alpha \in \Lambda$ of X, what does $\cup_{\lambda \in \varnothing} X_\lambda$ mean?

Exercise 1.3.

Prove that the Axiom of Choice is equivalent to each of the following:

(a) **The Axiom of Choice - disjoint set version [ACD].** Suppose that $\Lambda \ne \varnothing$ and that
  (i) for all $\lambda \in \Lambda$, X_λ is a non-empty set, and
  (ii) $X_\lambda \cap X_\beta = \varnothing$ if $\lambda \ne \beta \in \Lambda$.
  Then $\prod_{\lambda \in \Lambda} X_\lambda \ne \varnothing$.

(b) Given a non-empty set X there exists a function $f : \mathcal{P}(X) \setminus \varnothing \to X$ such that $f(A) \in A$ for all $A \in \mathcal{P}(X) \setminus \varnothing$.

Exercise 1.4.

(a) Prove that the only maximal ideals of $\mathcal{C}([0,1], \mathbb{C})$ are of the form $\mathcal{K}_y := \{f \in \mathcal{C}([0,1], \mathbb{C}) : f(y) = 0\}$.

(b) Can you identify what a general (closed) ideal of $\mathcal{C}([0,1], \mathbb{C})$ might look like?

Exercise 1.5.

Prove that every finite poset has a maximal element.

Exercise 1.6.

Recall that a subset $Y \subseteq \mathbb{R}$ is said to be **open** if for every $y \in Y$ there exists $\delta > 0$ such that $(y - \delta, y + \delta) \subseteq Y$.

Let $G \subseteq \mathbb{R}$ be an open set, and define a relation R on G by setting xRy if the line segment $[\min(x,y), \max(x,y)] \subseteq G$. Prove that this defines an equivalence relation on G, and that each equivalence class is an open interval. Furthermore, prove that the collection of equivalence classes is countable.

In other words, every open set G in $\mathbb{R}$ is a countable union of disjoint open intervals.

Is there a generalization of this result to higher dimensions? If so, what is it?

Exercise 1.7.

Let A, B, and C be sets. Prove the following statements.

(i) If $A \subseteq B$, then $|A| \le |B|$.
(ii) $|A| \le |A|$.
(iii) If $|A| \le |B|$ and $|B| \le |C|$, then $|A| \le |C|$.

(iv) If $m, n \in \mathbb{N}$ and $m \leq n$, then $|\{1,2,...,m\}| \leq |\{1,2,...,n\}|$.
(v) If E is finite, then $|E| < \aleph_0$.

Exercise 1.8.

Let $\alpha, \beta, \gamma, \delta$ be cardinal numbers. Prove that

(i) $\alpha \cdot \beta$ is well-defined. That is, if $|A| = |C|$, $|B| = |D|$, then $|A \times B| = |C \times D|$.
(ii) $\alpha \cdot \beta = \beta \cdot \alpha$; $\alpha(\beta \cdot \gamma) = (\alpha \cdot \beta)\gamma$; and $\alpha(\beta + \gamma) = \alpha \cdot \beta + \alpha \cdot \gamma$.
(iii) $0 \cdot \alpha = 0$.
(iv) If $\alpha \leq \beta$ and $\gamma \leq \delta$, then $\alpha \cdot \gamma \leq \beta \cdot \delta$.
(v) $\alpha \cdot \alpha = \alpha$ if α is infinite.

Exercise 1.9.

Let α and β be cardinal numbers. Prove that α^β is well defined. That is, if A_1, A_2, B_1, B_2 are sets with $|A_1| = \alpha = |A_2|$ and $|B_1| = \beta = |B_2|$, then $|A_1^{B_1}| = |A_2^{B_2}|$.

CHAPTER 2

Metric spaces and normed linear spaces

1. An introduction

If you haven't got anything nice to say about anybody, come sit next to me.

Alice Roosevelt Longworth

1.1. In this Chapter we shall turn our attention to metric spaces. These are sets endowed with a well-behaved notion of distance. Notions such as limits and continuity extend naturally to this context. Our reasons for studying metric spaces are two-fold:

- we prove general results once, rather than proving equivalent versions many times; and
- by removing the non-essential properties of a specific metric space, we arrive at a better understanding of the underlying concept.

We pause briefly to introduce some notation. When we do not wish to specify if we are using the real numbers $\mathbb{R}$ or the complex numbers $\mathbb{C}$, we shall use $\mathbb{K}$ to denote the field.

1.2. Definition. *A* ***metric space*** *is a pair* (X, d) *where* $\varnothing \neq X$ *is a non-empty set and* $d : X \times X \to \mathbb{R}$ *is a function (called a* ***metric****) which satisfies:*

(i) $d(x, y) \geq 0$ *for all* $x, y \in X$*;*
(ii) $d(x, y) = 0$ *if and only if* $x = y$*;*
(iii) $d(x, y) = d(y, x)$ *for all* $x, y \in X$*; and*
(iv) $d(x, z) \leq d(x, y) + d(y, z)$ *for all* $x, y, z \in X$*.*

The last property is known as the **triangle inequality**.

1.3. Example.

(a) The motivating example of a metric space is $(\mathbb{K}, d)$ and $d : \mathbb{K} \times \mathbb{K} \to \mathbb{R}$ is the map $d(x, y) = |x - y|$. We shall refer to this as the **standard metric** on $\mathbb{K}$.

(b) Let $\varnothing \neq X$ be a set. The **discrete metric** on X is the function $\mu : X \times X \to \mathbb{R}$ defined by

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

We leave it to the reader to verify that this is indeed a metric.

(c) Let $\varnothing \neq X$ be a set. Suppose that $f : X \to [0, \infty)$ is a function and suppose that $|f^{-1}(0)| \le 1$; that is, there exists at most one point x for which $f(x) = 0$. Define $d : X \times X \to \mathbb{R}$ via:

$$d(x,y) = \begin{cases} 0 & \text{if } x = y \\ f(x) + f(y) & \text{if } x \neq y. \end{cases}$$

Then (X, d) is a metric space. (The proof is left to the assignments.) This metric is sometimes referred to as the **SNCF metric**. The connection is that in order to travel from one point in France to another by rail, one would always have to return to Paris.

(d) Let $X = \mathbb{M}_n(\mathbb{K})$. We may define a metric $d : X \times X \to \mathbb{R}$ via $d(A, B) = \operatorname{rank}(A - B)$.

(e) Let $\varnothing \neq G$ be an undirected connected graph, and let V denote the set of vertices of G. For $x, y \in V$, define $d(x, y)$ to be the length of the shortest path connecting x to y (where the trivial path from x to x has length zero). Then (V, d) is a metric space.

(f) Let $\varnothing \neq X$ be a set, and denote by $\mathcal{F}$ the collection of all finite subsets of X. For $A, B \in \mathcal{F}$, define $d(A, B) = |A \Delta B|$, where

$$A \Delta B = (A \setminus B) \cup (B \setminus A)$$

denotes the **symmetric difference** of A and B. Then $(\mathcal{F}, d)$ is a metric space. The proof is left to the assignment.

(g) Let (X, d) be a metric space and $\varnothing \neq Y \subseteq X$. For $y_1, y_2 \in Y$, define $d_Y(y_1, y_2) = d(y_1, y_2)$. Then (Y, d_Y) is a metric space. The proof is routine.

One of the most important sources of metric spaces comes from the following construction:

1.4. Definition. *Let $\mathfrak{X}$ be a vector space over $\mathbb{K}$. A* ***seminorm*** *on $\mathfrak{X}$ is a function*

$$\nu : \mathfrak{X} \to \mathbb{R}$$

satisfying:

(a) $\nu(x) \ge 0$ *for all* $x \in \mathfrak{X}$;

(b) $\nu(kx) = |k| \nu(x)$ *for all* $k \in \mathbb{K}$ *and* $x \in \mathfrak{X}$; *and*

(c) $\nu(x + y) \le \nu(x) + \nu(y)$ *for all* $x, y \in \mathfrak{X}$.

If ν satisfies the extra condition:

(d) $\nu(x) = 0$ *if and only if* $x = 0$, *then we say that ν is a **norm**, and we usually denote $\nu(\cdot)$ by $\|\cdot\|$. In this case, we say that $(\mathfrak{X}, \|\cdot\|)$ (or, with a mild abuse of nomenclature, $\mathfrak{X}$) is a **normed linear space**.*

1.5. A norm on $\mathfrak{X}$ is a generalisation of the absolute value function on $\mathbb{K}$. Of course, as pointed out in Example 1.3 above, one may define a metric $d : \mathbb{K} \times \mathbb{K} \to \mathbb{R}$ by setting $d(x, y) = |x - y|$.

In exactly the same way, the norm $\|\cdot\|$ on a normed linear space $\mathfrak{X}$ induces a metric

$$d: \begin{array}{ccc} \mathfrak{X} \times \mathfrak{X} & \to & \mathbb{R} \\ (x, y) & \mapsto & \|x - y\|. \end{array}$$

Unless we explicitly make a statement to the contrary, this will always be the metric we consider when dealing with a normed linear space $(\mathfrak{X}, \|\cdot\|)$.

1.6. Example. Let $n \ge 1$ and $\mathfrak{X} = \mathbb{K}^n$. We define three norms on $\mathfrak{X}$ as follows: for $x = (x_1, x_2, \ldots, x_n) \in \mathfrak{X}$, we set

(a) $\|x\|_1 = \sum_{k=1}^n |x_k|$;
(b) $\|x\|_\infty = \max(|x_1|, |x_2|, \ldots, |x_n|)$; and
(c) $\|x\|_2 = \left(\sum_{k=1}^n |x_k|^2\right)^{\frac{1}{2}}$.

These are referred to as the **1-norm**, the **infinity-norm**, and the **2-norm** respectively. That the 1-norm and the infinity-norm are norms is a routine exercise. That the 2-norm is a norm is left as an assignment exercise.

We denote the metrics induced by these norms by $d_1(\cdot,\cdot)$, $d_\infty(\cdot,\cdot)$ and $d_2(\cdot,\cdot)$ respectively.

1.7. Example.

Let $\ell_1 := \{x = (x_n)_n \in \mathbb{K}^{\mathbb{N}} : \sum_n |x_n| < \infty\}$. Then ℓ_1 is a vector space over $\mathbb{K}$ and

$$\|x\|_1 := \sum_{n=1}^{\infty} |x_n|$$

defines a norm on ℓ_1, once again called the **1-norm** on ℓ_1. We denote by $d_1(x, y) := \|x - y\|_1$ the corresponding metric on ℓ_1.

Let $\ell_\infty := \{y = (y_n)_n \in \mathbb{K}^{\mathbb{N}} : \sup_n |y_n| < \infty\}$. Then ℓ_∞ is a vector space over $\mathbb{K}$ and

$$\|y\|_\infty := \sup_{n \ge 1} |y_n|$$

defines a norm on ℓ_∞, also called the **infinity-norm** or **supremum norm** on ℓ_∞. We denote by $d_\infty(x, y) := \|x - y\|_\infty$ the corresponding metric on ℓ_∞.

1.8. Example.

Let $\mathfrak{X} = \mathcal{C}([0,1],\mathbb{K}) = \{f : [0,1] \to \mathbb{K} : f \text{ is continuous}\}$. We may once again define a norm which we call the **supremum norm** on $\mathfrak{X}$ via

$$\|f\|_\infty := \sup\{|f(x)| : x \in [0,1]\} = \max\{|f(x)| : x \in [0,1]\}.$$

Moreover, the function

$$\|f\|_1 := \int_0^1 |f(x)| dx$$

defines a norm on $\mathfrak{X}$, which we call the **1-norm** on $\mathcal{C}([0,1],\mathbb{K})$.

1.9. Example. Let $\mathfrak{X} = \mathcal{C}_b(\mathbb{R},\mathbb{K}) = \{f : \mathbb{R} \to \mathbb{K} : f \text{ is bounded}\}$. Then $\mathcal{C}_b(\mathbb{R},\mathbb{K})$ is a vector space and

$$\|f\|_\infty := \sup\{|f(x)| : x \in \mathbb{R}\}$$

defines a norm on $\mathfrak{X}$.

1.10. Example. Recall that if V and W are vector spaces over a field $\mathbb{F}$, then

$$\mathcal{L}(V,W) = \{T : V \to W : T \text{ is linear}\}$$

is again a vector space over $\mathbb{F}$.

Let $(\mathfrak{X}, \|\cdot\|_{\mathfrak{X}})$ and $(\mathfrak{Y}, \|\cdot\|_{\mathfrak{Y}})$ be normed linear spaces. Suppose that $T \in \mathcal{L}(\mathfrak{X},\mathfrak{Y})$, so that T is linear. We define

$$\|T\| := \sup\{\|Tx\|_{\mathfrak{Y}} : \|x\|_{\mathfrak{X}} \le 1\}.$$

If $\|T\| < \infty$, we say that T is **bounded**. The set $\mathcal{B}(\mathfrak{X},\mathfrak{Y}) := \{T \in \mathcal{L}(\mathfrak{X},\mathfrak{Y}) : \|T\| < \infty\}$ of all bounded linear operators from $\mathfrak{X}$ to $\mathfrak{Y}$ is a vector space over $\mathbb{K}$ and $(\mathcal{B}(\mathfrak{X},\mathfrak{Y}), \|\cdot\|)$ is a normed linear space. This is discussed in the assignment, where it is also seen that an operator $T : \mathfrak{X} \to \mathfrak{Y}$ is bounded if and only if T is continuous on $\mathfrak{X}$.

2. Topological structure of metric spaces

2.1. Definition. *Let (X,d) be a metric space, $x \in X$ and $\delta > 0$. The* ***ball of radius*** *δ,* ***centred at*** *x is the set*

$$B(x,\delta) := \{y \in X : d(x,y) < \delta\}.$$

We shall also refer to this as the ***δ-neighbourhood*** *of x in X.*

More generally, a set H is said to be a ***neighbourhood*** *of x if there exists $\delta > 0$ so that $B(x,\delta) \subseteq H$.*

A subset $G \subseteq X$ is said to be ***open*** *if for every $x \in G$ there exists $\delta > 0$ so that $B(x,\delta) \subseteq G$. The important thing to remember here is that δ depends upon x. Thus G is open if it is a neighbourhood of each of its points. We denote by $\tau (= \tau_X)$ the collection of all open sets in X.*

Finally, a set $F \subseteq X$ is said to be ***closed*** *if $X \setminus F$ is open.*

Warning! Many authors require neighbourhoods of a point to be *open*. We do not.

2.2. Proposition. *Let (X,d) be a metric space, $x \in X$ and $\delta > 0$. Then $B(x,\delta)$ is open.*

Proof. Let $y \in B(x,\delta)$. Then $\rho := d(x,y) < \delta$. Consider $\varepsilon = \delta - \rho > 0$. If $z \in B(y,\varepsilon)$, then

$$\begin{aligned} d(x,z) &\le d(x,y) + d(y,z) \\ &< \rho + \varepsilon \\ &= \delta, \end{aligned}$$

and thus $z \in B(x,\delta)$. That is, $B(y,\varepsilon) \subseteq B(x,\delta)$. Since $y \in B(x,\delta)$ was arbitrary, the latter is open.

□

For this reason, we shall refer to $B(x,\delta)$ as the **open** ball of radius δ.

2.3. Proposition. *Let (X,d) be a metric space, $x \in X$ and $\delta > 0$. Let*

$$B[x,\delta] := \{y \in X : d(x,y) \le \delta\}.$$

Then $B[x,\delta]$ is closed.

Proof. Let $G = X \setminus B[x,\delta]$ and let $y \in G$. Then $\rho := d(x,y) > \delta$. Set $\varepsilon = \rho - \delta > 0$. If $z \in B(y,\varepsilon)$, then

$$\begin{aligned} d(x,z) &\ge d(x,y) - d(y,z) \\ &> \rho - \varepsilon \\ &= \delta, \end{aligned}$$

and so $z \in G$. Hence G is open, and so $B[x,\delta]$ is closed.

□

In light of this result, we refer to $B[x, \delta]$ as the **closed ball of radius** δ, **centred at** x.

2.4. Remark. Most subsets of a metric space are *neither* open nor closed! For example, in $(\mathbb{R}, d)$, where d is the standard metric, $(0, 1)$ is open, $[0, 1]$ is closed, but $(0, 1]$ is neither open nor closed.

2.5. It is interesting to investigate the geometry of the open and closed balls of radius 1 in various metric spaces.

(a) Let (X, μ) be a discrete metric space, and $x \in X$. Then $B(x, 1) = \{x\}$, while $B[x, \delta] = X$. In particular, each point is open in a discrete metric space. (Is this true of all metric spaces?)

(b) Consider $(X, d) = (\mathbb{R}^2, \|\cdot\|_1)$.

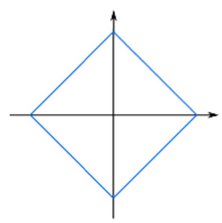


FIGURE 2.5(B). THE UNIT SPHERE OF $(\mathbb{R}^2, \|\cdot\|_1)$.

(c) Consider $(X, d) = (\mathbb{R}^2, \|\cdot\|_2)$.

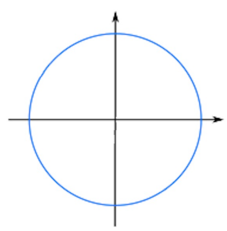


FIGURE 2.5(C). THE UNIT SPHERE OF $(\mathbb{R}^2, \|\cdot\|_2)$.

(d) Consider $(X, d) = (\mathbb{R}^2, \|\cdot\|_\infty)$.

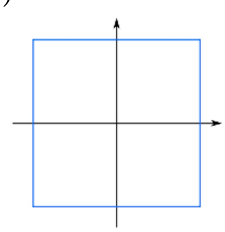


FIGURE 2.5(D). THE UNIT SPHERE OF $(\mathbb{R}^2, \|\cdot\|_\infty)$.

2.6. Definition. *Let (X, d) be a metric space, $x \in X$ and $(x_n)_n \in X^{\mathbb{N}}$ be a sequence in X. We say that $(x_n)_n$* ***converges to*** *x, and we write*

$$\lim_{n \to \infty} x_n = x,$$

if for each $\varepsilon > 0$ *there exists* $N > 0$ *so that* $n \geq N$ *implies that* $d(x_n, x) < \varepsilon$*; that is,* $n \geq N$ *implies that* $x_n \in B(x, \varepsilon)$*.*

We say that the sequence $(x_n)_n$ *is* ***Cauchy*** *if for all* $\varepsilon > 0$ *there exists* $N > 0$ *so that* $m, n \geq N$ *implies*

$$d(x_n, x_m) < \varepsilon.$$

As is the case for sequences of real numbers, every convergent sequence in a metric space is Cauchy. The proof is identical to the real case, and is left as an exercise.

The following familiar result from Calculus holds in the more general setting of metric spaces.

2.7. Proposition. *Let* (X, d) *be a metric space, and suppose that* $(x_n)_n$ *is a Cauchy sequence in* X*. Suppose that there exists a subsequence* $(x_{n_k})_{k=1}^{\infty}$ *of* $(x_n)_n$ *and an element* $x \in X$ *so that*

$$\lim_{k \to \infty} x_{n_k} = x.$$

Then $\lim_n x_n = x$*.*

Proof. See the homework problems.

□

2.8. Example. Let $(x_n)_n$ be a sequence in a discrete metric space (X, μ). Then $\lim_n x_n = x$ if and only if there exists $N > 0$ so that $n \geq N$ implies that $x_n = x$.

Indeed,

- Suppose that there exists $N > 0$ so that $n \geq N$ implies that $x_n = x$. Given any $\varepsilon > 0$, $n \geq N$ implies that $d(x_n, x) = 0 < \varepsilon$, and so $\lim_n x_n = x$.
- Now suppose that $\lim_n x_n = x$. Let $0 < \varepsilon \leq 1$, and choose $N > 0$ so that $n \geq N$ implies that $d(x_n, x) < \varepsilon$. Then $n \geq N$ implies that $x_n = x$.

2.9. Example. Let $\mathfrak{X} = \mathcal{C}([0,1], \mathbb{R})$. Consider the sequence $(f_n)_n$, where

$$f_n(x) = \begin{cases} 2nx & x \in [0, \frac{1}{2n}] \\ 2 - 2nx & x \in [\frac{1}{2n}, \frac{1}{n}] \\ 0 & x \in [\frac{1}{n}, 1]. \end{cases}$$

Then

- $(f_n)_n$ converges *pointwise* to 0. That is, for each $x \in [0,1]$, $\lim_n f_n(x) = 0$.
  Indeed, $f_n(0) = 0$ for all $n \geq 1$, and so clearly $\lim_n f_n(0) = 0$. If $0 < x \leq 1$, then there exists $N > 0$ so that $\frac{1}{N} < x$, and then $n \geq N$ implies that $f_n(x) = 0$, so that $\lim_n f_n(x) = 0$ as well.
- $(f_n)_n$ converges to 0 in $(\mathcal{C}([0,1],\mathbb{R}), \|\cdot\|_1)$.
  Consider
  $$d_1(f_n, 0) := \|f_n - 0\|_1 = \int_0^1 |f_n(x)| dx = \frac{1}{2n}.$$
  Since $\lim_n \frac{1}{2n} = 0$, it follows that $(f_n)_n$ converges to 0 in $(\mathcal{C}([0,1],\mathbb{R}), \|\cdot\|_1)$.
- $(f_n)_n$ does not converge in $(\mathcal{C}([0,1],\mathbb{R}), \|\cdot\|_\infty)$.
  It suffices to prove that $(f_n)_n$ is not Cauchy. Let $0 < \varepsilon \leq 1$, and suppose that $N \geq 1$. Choose $m, n \geq N$ with $m > 2n$. Then $f_m(\frac{1}{2n}) = 0$ and $f_n(\frac{1}{2n}) = 1$, so
  $$d_\infty(f_n, f_m) = \|f_n - f_m\|_\infty \geq |f_n(\frac{1}{2n}) - f_m(\frac{1}{2n})| = 1,$$
  proving that $(f_n)_n$ is not Cauchy in $(\mathcal{C}([0,1],\mathbb{R}), \|\cdot\|_\infty)$.

2.10. Remark. Convergence in $(\mathcal{C}([0,1],\mathbb{K}), \|\cdot\|_\infty)$ is also referred to as **uniform convergence**.

We leave it as an exercise for the reader to prove that if a sequence $(f_n)_n$ in $\mathcal{C}([0,1],\mathbb{K})$ converges uniformly to f, then $(f_n)_n$ converges pointwise to f, and also $(f_n)_n$ converges in $(\mathcal{C}([0,1],\mathbb{K}), \|\cdot\|_1)$.

2.11. Definition. *Let (X,d) be a metric space and $E \subseteq X$.*

*A point $q \in X$ is said to be a **limit point** of E if there exists a sequence $(x_n)_n \in E^{\mathbb{N}}$ such that $\lim_n x_n = q$.*

*We say that the point $p \in X$ is an **accumulation point** of E if for every neighbourhood H of p, the **punctured neighbourhood** $H \setminus \{p\}$ of p intersects E non-trivially. That is,*
$$(H \setminus \{p\}) \cap E \neq \varnothing.$$

*We write E' for the set of all accumulation points of E. (This is sometimes referred to as the **derived set** of E.)*

*If $q \in E$ and q is not an accumulation point of E, then q is said to be an **isolated point** of E.*

Remarks: in checking that p is an accumulation point of E, it suffices to consider punctured neighbourhoods of the form $B(p,\delta) \setminus \{p\}$. (Why?)

Secondly, it is a routine exercise to prove that if $E_1 \subseteq E_2 \subseteq X$, then $E_1' \subseteq E_2'$.

Exercise. Every accumulation point of E in (X,d) is a limit point of E.

2.12. Example. Let (X, μ) be a discrete metric space and $E \subseteq X$. If $p \in X$ and $\varepsilon = \frac{1}{2}$, then either

(a) $p \in E$ and $B(p, \varepsilon) \cap E = \{p\}$, or
(b) $p \notin E$, and $B(p, \varepsilon) \cap E = \varnothing$.

Either way, we see that p is not an accumulation point of E. If $p \in E$, then p must be an isolated point of E.

2.13. Example. Consider $\mathbb{R}$ with the standard metric d. Let $E = \{\frac{1}{n}\}_{n=1}^{\infty}$. If $p = 0$, then for each $\varepsilon > 0$, we can find $n > \frac{1}{\varepsilon}$, so that

$$(B(0, \varepsilon) \setminus \{0\}) \cap E \supseteq \{\frac{1}{n}\} \neq \varnothing.$$

Hence $p = 0$ is an accumulation point of E. Note that $0 \notin E$.

We leave it as an exercise for the reader to prove that every point of E is an isolated point of E.

2.14. Theorem. *Let (X, d) be a metric space and $F \subseteq X$. The following are equivalent:*

(a) *F is closed.*
(b) *$F' \subseteq F$; that is, every accumulation point of F lies in F.*
(c) *Every limit point of F lies in F.*

Proof.

(a) implies (b):

Suppose that F is closed and that $y \in F'$. If $y \notin F$, then $y \in G := X \setminus F$, and so there exists $\delta > 0$ so that $B(y, \delta) \subseteq G$. Then

$$(B(y, \delta) \setminus \{y\}) \cap F \subseteq G \cap F = \varnothing,$$

contradicting the fact that y is an accumulation point of F. Thus $y \in F$, as required.

(b) implies (c): Suppose that $F' \subseteq F$, and let q be a limit point of F. Let $(x_n)_n$ be a sequence in F such that $\lim_n x_n = q$. If there exists $n \geq 1$ so that $x_n = q$, then $q \in F$ and we are done. Otherwise, let $\delta > 0$, and choose $N > 0$ so that $n \geq N$ implies that $d(x_n, q) < \delta$, i.e. $x_n \in B(q, \delta)$. Since $x_N \neq q$, x_N lies in the punctured neighbourhood $(B(q, \delta) \setminus \{q\})$ of q. Since $\delta > 0$ was arbitrary, this shows that q is an accumulation point of F, and hence lies in F by our hypothesis in (b).

Hence F contains all of its limit points.

(c) implies (a): Suppose that F contains all of its limit points, and let $G = X \setminus F$. We shall prove that G is open, which is equivalent to showing that F is closed.

Let $y \in G$. Since y is not a limit point of F, there exists $\delta > 0$ so that $B(y, \delta) \cap F = \varnothing$. But then G is open, so F is closed.

$\square$

2.15. Example. Let $X = \mathbb{R}$, equipped with the standard metric d. Let $H = \{\frac{1}{n}\}_{n=1}^{\infty}$. Then H is not closed, since $0 \in H'$ is an accumulation point of H, but $0 \notin H$.

We note that H is not open, either, since there does not exist $\delta > 0$ so that $B(1, \delta) \subseteq H$.

2.16. Example. Let $X = \mathcal{C}_b(\mathbb{R}, \mathbb{R})$, equipped with the supremum norm $\|\cdot\|_\infty$. For each $n \geq 1$, define

$$f_n(x) = \begin{cases} 0 & \text{if } x \leq n; \\ x - n & \text{if } n < x \leq n+1 \\ 1 & \text{if } n+1 \leq x. \end{cases}$$

It is easy to verify that $f_n \in \mathcal{C}_b(\mathbb{R}, \mathbb{R})$ for all $n \geq 1$. Let $H = \{f_n\}_{n=1}^{\infty}$.

Then $\|f_n - f_m\|_\infty = 1$ for all $1 \leq n \neq m$, and so $H' = \varnothing$ (why?). In particular, H is closed, since $H' = \varnothing \subseteq H$.

2.17. Theorem. *Let (X, d) be a metric space.*

(a) *If $G_\lambda \subseteq X$ is open for all $\lambda \in \Lambda$, then*

$$G = \cup_{\lambda \in \Lambda} G_\lambda$$

is open.

(b) *If $F_\lambda \subseteq X$ is closed for all $\lambda \in \Lambda$, then*

$$F = \cap_{\lambda \in \Lambda} F_\lambda$$

is closed.

(c) *If $n \geq 1$ is an integer and $G_1, G_2, ..., G_n$ are open in X, then*

$$H = \cap_{k=1}^{n} G_k$$

is open.

(d) *If $n \geq 1$ is an integer and $F_1, F_2, ..., F_n$ are closed in X, then*

$$K = \cup_{k=1}^{n} F_k$$

is closed.

Proof.

(a) Let $x \in G$. Then there exists $\alpha \in \Lambda$ so that $x \in G_\alpha$. Since G_α is open, there exists $\delta > 0$ so that $B(x, \delta) \subseteq G_\alpha \subseteq G$. Hence G is open.

(b) Set $G_\lambda = X \setminus F_\lambda$ for all $\lambda \in \Lambda$, so that each G_λ is open. By part (a), $G = \cup_{\lambda \in \Lambda} G_\lambda$ is open. Hence $F = X \setminus G$ is closed.

(c) Let $x \in H$. Then $x \in G_k$, $1 \leq k \leq n$. Since each G_k is open, there exist $\delta_k > 0$, $1 \leq k \leq n$, so that $B(x, \delta_k) \subseteq G_k$. Let $\delta = \min(\delta_1, \delta_2, ..., \delta_n) > 0$. Then

$$B(x, \delta) \subseteq B(x, \delta_k) \subseteq G_k, 1 \leq k \leq n,$$

and so $B(x, \delta) \subseteq H$. Hence H is open.

(d) The proof is analogous to that of (b) and is left as an exercise.

□

2.18. Remark. In parts (c) and (d) of Theorem 2.17, the fact that n is finite is crucial. For example, if we set $X = \mathbb{R}$, equipped with the standard metric d, and if $G_n = (-1, \frac{1}{n})$, $n \geq 1$, then each G_n is open in $(\mathbb{R}, d)$, but $H = \cap_{n\geq 1} G_n = (-1, 0]$ is neither open nor closed.

2.19. Definition. *Let (X, d) be a metric space and $H \subseteq X$. We define the* ***closure*** *of H to be*

$$\overline{H} := \cap\{F \subseteq X : H \subseteq F \text{ and } F \text{ is closed}\}.$$

It should be clear from the definition of $\overline{H}$ that

- $H \subseteq \overline{H}$.
- $\overline{H}$ is closed, being the intersection of closed sets.
- $\overline{H}$ is the smallest closed set containing H, in the sense that if $K \subseteq X$ is closed and $H \subseteq K$, then $\overline{H} \subseteq K$.

2.20. Proposition. *Let (X, d) be a metric space and $H \subseteq X$. Then*

(a) *H is closed if and only if $H = \overline{H}$, and*
(b) *$\overline{H} = H \cup H'$.*

In particular, every point of $\overline{H}$ is a limit point of H.

Proof.

(a) Suppose that $H = \overline{H}$. By the comments following Definition 2.19, $\overline{H}$ is closed and thus H is closed.

Conversely, suppose that H is closed. Now $H \subseteq \overline{H}$ always holds. On the other hand, H is a closed set which contains H, and thus - by the third comment following Definition 2.19, $\overline{H} \subseteq H$. Hence $H = \overline{H}$.

(b) Let $K = H \cup H'$.

- Suppose that $x \notin K$. Then $x \notin H$ and $x \notin H'$. Since $x \notin H'$, there exists $\delta > 0$ such that $(B(x, \delta) \setminus \{x\}) \cap H = \varnothing$. Since $x \notin H$, we in fact find that $B(x, \delta) \cap H = \varnothing$. Now, by Proposition 2.2, $B(x, \delta)$ is open, and thus $F := X \setminus B(x, \delta)$ is closed. Furthermore, the fact that $H \subseteq F$ implies that $\overline{H} \subseteq F$.
  Since $x \notin F$, $x \notin \overline{H}$. Thus $x \notin K$ implies that $x \notin \overline{H}$, or equivalently, $\overline{H} \subseteq K$.
- Suppose that $x \notin \overline{H}$. Since $\overline{H}$ is closed by part (a), $G := X \setminus \overline{H}$ is open and $x \in G$. Hence there exists $\delta > 0$ such that $B(x, \delta) \subseteq G$; i.e. $B(x, \delta) \cap H = \varnothing$. But then $x \notin H$ and $x \notin H'$, so $x \notin K$. In other words, $K \subseteq \overline{H}$.

Finally, if $x \in \overline{H}$, then either $x \in H$, in which case, $x = \lim_n x_n$, where $x_n = x \in H$ for all $n \geq 1$, or $x \notin H$, in which case x is an accumulation point of H by (b).

□

2.21. Example. Let $X = \mathbb{R}^2$ with the Euclidean metric $d_2(x, y) = \|x - y\|_2$. Then

$$\begin{aligned}\overline{B((0,0),1)} &= B[(0,0),1] \\ &= \{(y_1, y_2) \in \mathbb{R}^2 : \|(y_1, y_2) - (0,0)\|_2 \le 1\} \\ &= \{(y_1, y_2) \in \mathbb{R}^2 : \sqrt{|y_1 - 0|^2 + |y_2 - 0|^2 \le 1}\}.\end{aligned}$$

2.22. Example. Let $X = \mathbb{R}$ with the usual metric. If $H = \mathbb{Q}$, then $\overline{H} = \mathbb{R} = X$, since every point of $\mathbb{R}$ is an accumulation point of $\mathbb{Q}$.

2.23. Definition. *A subset Y of a metric space (X, d) is said to be* ***dense*** *if $\overline{Y} = X$. If X admits a countable dense subset, then X is said to be* ***separable****.*

2.24. Example. With $X = \mathbb{R}$ and d the usual metric, the set $\mathbb{I} = \mathbb{R} \setminus \mathbb{Q}$ of irrational numbers is dense, and uncountable. However, we saw in Example 2.22 that the set $\mathbb{Q}$ of rational numbers is also dense in $\mathbb{R}$, and $\mathbb{Q}$ is countable, so $\mathbb{R}$ is separable.

Exercise. Suppose that (X, μ) is a discrete metric space. Which are the dense subsets of X? When is (X, μ) separable?

CHAPTER 3

Topology

1. Topological spaces

What I am looking for is a blessing not in disguise.

Jerome K. Jerome

1.1. Definition. *Let $\varnothing \neq X$ be a non-empty set. A subset $\tau \subseteq \mathcal{P}(X)$ is called a* ***topology*** *on X, and (X, τ) is said to be a* ***topological space****, if*

(a) $\varnothing, X \in \tau$;
(b) $\cup_{\lambda \in \Lambda} T_\lambda \in \tau$ *whenever* $\{T_\lambda\}_{\lambda \in \Lambda} \subseteq \tau$;
(c) $\cap_{k=1}^n T_k \in \tau$ *whenever* $\{T_k\}_{k=1}^n \subseteq \tau$.

The elements of τ are said to be **τ-open** - or, if τ is fixed, just **open**. A subset $F \subseteq X$ is said to be **closed** if $G = X \setminus F \in \tau$.

1.2. Example. Let (X, d) be a metric space. Consider

$$\tau := \{G \subseteq X : \text{ for all } x \in G \text{ there exists } \delta > 0 \text{ so that } B(x, \delta) \subseteq G\}.$$

Clearly $\varnothing$ and X belong to τ. Also, by Theorem 2.2.17, conditions (b) and (c) from Definition 1.1 also hold. Thus τ is a topology on X, called the **metric topology** on X.

Observe that we have defined the metric topology on X so that a set G is open in the sense of Definition 2.2.1 for metric spaces if and only if it is open in the topological sense of Definition 1.1 above.

If $X = \mathbb{K}$ and d is the standard metric on X, then we refer the metric topology induced by d as the **standard topology** on X. Unless otherwise specified, this will be the topology we shall always consider for $\mathbb{K}$.

1.3. Examples.

(a) Let X be any non-empty set. Then $\tau = \{\varnothing, X\}$ is a topology on X, called the **trivial topology** on X.
(b) At the other extreme: Let X be any non-empty set. Then $\tau = \mathcal{P}(X)$ is a topology on X, called the **discrete topology** on X. Note that this is the metric topology induced on X when X is equipped with the discrete metric μ.
(c) Let $X = \{a, b\}$, and set $\tau = \{\varnothing, \{a\}, \{a, b\}\}$. Then τ is a topology on X.

Question: does there exist a metric d on X so that τ is the metric topology induced by d?
(d) Let $X = \mathbb{Z}$ and set $\tau = \{\varnothing\} \cup \{T \subseteq \mathbb{Z} : 5 \in T\}$. Then τ is a topology on $\mathbb{Z}$.

1.4. Example. Let X be any non-empty set. Define

$$\tau_{cf}(X) = \{\varnothing\} \cup \{T \subseteq X : |X \setminus T| < \infty\}.$$

Then $\tau_{cf}(X)$ defines a topology on X, called the **co-finite topology**.

Borrowing shamelessly from our metric space notions, we obtain the following:

1.5. Definition. *Let (X, τ) be a topological space and $H \subseteq X$. The* ***closure*** *of H is*

$$\overline{H} = \cap\{F \subseteq X : H \subseteq F \text{ and } F \text{ is closed}\}.$$

We also say that a subset $D \subseteq X$ is ***dense*** *in X if $\overline{D} = X$, and that X is* ***separable*** *if X admits a countable, dense subset.*

As was the case for metric spaces, we see that $\overline{H}$ is therefore the smallest closed subset of X which contains H, in the sense that it is contained in every closed subset of X which contains H.

1.6. Examples.

(a) Suppose that X is a non-empty set equipped with the trivial topology τ. If $H = \varnothing \subseteq X$, then $\overline{H} = \varnothing = H$, so H is closed. If $H \neq \varnothing$, then $\overline{H} = X$. In particular, by taking H to be a singleton set, we see that (X, τ) is separable.
(b) Let $X = \{a, b\}$, and $\tau = \{\varnothing, \{a\}, \{a, b\}\}$. Then
- $\overline{\varnothing} = \varnothing$;
- $\overline{\{a\}} = \{a, b\}$;
- $\overline{\{b\}} = \{b\}$;
- $\overline{\{a, b\}} = \{a, b\}$.

(c) Let $X = \mathbb{Z}$ and let $\tau = \{\varnothing\} \cup \{H \subseteq \mathbb{Z} : 5 \in H\}$.
- If $A = \{5, 7\}$, then $\overline{A} = \mathbb{Z}$;
- If $A = 2\mathbb{N}$, then $\overline{A} = A = 2\mathbb{N}$.

1.7. Definition. *Let (X, τ) be a topological space and $x \in X$. A **neighbourhood** of x is a set $U \subseteq X$ for which there exists $G \in \tau$ such that $x \in G \subseteq U$. The collection*

$$\mathcal{U}_x = \{U : U \text{ is a neighbourhood of } x\}$$

*is called the **neighbourhood system** at x.*

Remark: Some authors require neighbourhoods to be open. We do not.

1.8. Example. This notion of a neighbourhood is consistent with our previous notion of a neighbourhood in a metric space. Thus if (X, d) is a metric space and $x \in X$, then U is a neighbourhood of x if and only if there exists $\delta > 0$ so that $B(x, \delta) \subseteq U$.

1.9. Theorem. *Let (X, τ) be a topological space and $x \in X$.*

(a) *If $U \in \mathcal{U}_x$, then $x \in U$.*
(b) *If $U, V \in \mathcal{U}_x$, then $U \cap V \in \mathcal{U}_x$.*
(c) *If $U \in \mathcal{U}_x$, then there exists $V \in \mathcal{U}_x$ such that $U \in \mathcal{U}_y$ for each $y \in V$.*
(d) *If $U \in \mathcal{U}_x$ and $U \subseteq V$, then $V \in \mathcal{U}_x$.*
(e) *The set $G \in \tau$ if and only if G contains a neighbourhood of each of its points.*

*Conversely: suppose that Y is a non-empty set and for each $x \in Y$ we are given a non-empty collection $\mathcal{U}_x \subseteq \mathcal{P}(Y)$ satisfying conditions (a) through (d). Suppose furthermore that we **declare** a set $G \in Y$ to be **open** if for each $x \in G$ there exists $U \in \mathcal{U}_x$ so that $x \in U \subseteq G$. If we then set $\rho = \{G \subseteq Y : G \text{ is open}\}$, then ρ is a topology on Y in which the neighbourhood system at x is exactly $\mathcal{U}_x$.*

Proof.

(a) This is clear from the definition.
(b) Let $G_1, G_2 \in \tau$ such that $x \in G_1 \subseteq U$ and $x \in G_2 \subseteq V$. Then $G = G_1 \cap G_2 \in \tau$, and $x \in G \subseteq (U \cap V)$, so $U \cap V \in \mathcal{U}_x$.
(c) Choose $G \in \tau$ so that $x \in G \subseteq U$, and set $V = G$. If $y \in V$, then $y \in G \subseteq U$, so $U \in \mathcal{U}_y$.
(d) Choose $G \in \tau$ so that $x \in G \subseteq U$. Then $x \in G \subseteq V$, so $V \in \mathcal{U}_x$.
(e)
- First suppose that $G \in \tau$, and let $x \in G$. Then $x \in G \subseteq G$, so $G \in \mathcal{U}_x$. Hence G is itself a neighbourhood of each of its points.
- Conversely, suppose that G contains a neighbourhood of each of its points. Let $x \in G$. Then there exists $T_x \in \tau$ so that $x \in T_x \subseteq G$. But then
$$G = \cup\{x : x \in G\} \subseteq \cup\{T_x : x \in G\} \subseteq G,$$
and thus $G = \cup\{T_x : x \in G\}$. Since the union of open sets is open, $G \in \tau$.

Next, suppose that $\varnothing \neq Y$ is a set satisfying (a)-(d), and that we have defined ρ as in the statement of the Theorem.

- It is clear that $\varnothing$ belongs to ρ trivially.

  As for Y, note that for each $y \in Y$, we have that $\mathcal{U}_y \neq \varnothing$ (by hypothesis). Hence there exists $U_y \in \mathcal{U}_y$ and by hypothesis (a), $y \in U_y$. But then for all $y \in Y$, $y \in U_y \subseteq Y$, and thus Y is declared open.
- Suppose that $T_\lambda \in \rho$, $\lambda \in \Lambda$. Let $T = \cup_{\lambda \in \Lambda} T_\lambda$, and choose $x \in T$. Then $x \in T_{\lambda_0}$ for some $\lambda_0 \in \Lambda$, and so there exists $U \in \mathcal{U}_x$ so that $x \in U_x \subseteq T_{\lambda_0} \subseteq T$. Thus T is open, i.e. $T \in \rho$.
- Suppose that $T_1, T_2, ..., T_n \in \rho$ and let $x \in T = \cap_{k=1}^n T_k$. Fix $1 \leq k \leq n$. Since $x \in T_k \in \rho$, we may find $U_k \in \mathcal{U}_x$ so that $x \in U_k \subseteq T_k$. By (finite) induction using condition (b), $U_x = \cap_{k=1}^n U_k \in \mathcal{U}_x$, and clearly

$$x \in U_x \subseteq T.$$

  But $x \in T$ was arbitrary, so $T \in \rho$ by definition of ρ.

Thus ρ is a topology. We leave the last statement as an exercise for the reader. The proof can be found in the appendix to this Chapter.

□

1.10. Definition. *Let (X, τ) be a topological space. A* ***directed set*** *is a set Λ with a relation $\leq$ that satisfies:*

(i) *$\lambda \leq \lambda$ for all $\lambda \in \Lambda$;*
(ii) *if $\lambda_1 \leq \lambda_2$ and $\lambda_2 \leq \lambda_3$, then $\lambda_1 \leq \lambda_3$; and*
(iii) *if $\lambda_1, \lambda_2 \in \Lambda$, then there exists λ_3 so that $\lambda_1 \leq \lambda_3$ and $\lambda_2 \leq \lambda_3$.*

The relation $\leq$ is called a ***direction*** *on Λ.*

A ***net*** *in X is a function $P : \Lambda \to X$, where Λ is a directed set. The point $P(\lambda)$ is usually denoted by x_λ, and we often write $(x_\lambda)_{\lambda \in \Lambda}$ to denote the net.*

A ***subnet*** *of a net $P : \Lambda \to X$ is the composition $P \circ \varphi$, where $\varphi : M \to \Lambda$ is an increasing* ***cofinal*** *function from a directed set M to Λ; that is,*

(a) *$\varphi(\mu_1) \leq \varphi(\mu_2)$ if $\mu_1, \mu_2 \in M$ and $\mu_1 \leq \mu_2$ (**increasing**), and*
(b) *for each $\lambda \in \Lambda$, there exists $\mu \in M$ so that $\lambda \leq \varphi(\mu)$ (**cofinal**).*

For $\mu \in M$, we often write x_{λ_μ} for $P \circ \varphi(\mu)$, and speak of the subnet $(x_{\lambda_\mu})_\mu$.

Remark: It might be worth recalling that a sequence in a set X defined to be a function $f : \mathbb{N} \to X$, and that we normally write x_n instead of $f(n)$, and $(x_n)_{n=1}^\infty$ instead of f. Thus, our notation $(x_\lambda)_{\lambda \in \Lambda}$ for a net has been chosen to mimic the notation we use for sequences. The same is true of our definition of convergence for nets, which mimics the definition of convergence in a metric space.

1.11. Definition. *Let (X, τ) be a topological space. The net $(x_\lambda)_\lambda$ is said to* ***converge to*** *$x \in X$ if for every $U \in \mathcal{U}_x$ there exists $\lambda_0 \in \Lambda$ so that $\lambda \geq \lambda_0$ implies $x_\lambda \in U$.*

We write $\lim_\lambda x_\lambda = x$, or $\lim_{\lambda \in \Lambda} x_\lambda = x$.

1.12. Examples.

(a) Since $\mathbb{N}$ is a directed set under the usual order $\le$, every sequence is a net. Any subsequence of a sequence is also a subnet. The converse to this is false, however. A subnet of a sequence need not be a subsequence, since its domain need not be $\mathbb{N}$ (or any countable set, for that matter).

(b) Let A be a non-empty set and Λ denote the power set of all subsets of A, partially ordered with respect to inclusion. Then Λ is a directed set, and any function from Λ to $\mathbb{R}$ is a net in $\mathbb{R}$.

(c) Although the definitions of a directed set and of a net might seem new and wondrous, you will have been introduced to the concept of convergence of nets when you studied Riemann integration (though you may not have used this terminology).

Let $a < b \in \mathbb{R}$. A **partition** of $[a, b]$ is a *finite* set $P = \{a = p_0 < p_1 < \cdots < p_N = b\}$ of $[a, b]$ which includes both a and b. We denote by $\mathcal{P}[a, b]$ the set of all partitions of $[a, b]$. Given a second partition $Q = \{a = q_0 < q_1 < \cdots < q_M = b\}$ of $[a, b]$, we say that Q is a **refinement** of P if $P \subseteq Q$.

Let $P = \{a = p_0 < p_1 < \cdots < p_N = b\} \in \mathcal{P}[a, b]$ be a partition of $[a, b]$. Recall that a set $P^* := \{p_1^*, p_2^*, \ldots, p_n^*\}$ is said to be a set of **test values** for P if $p_{n-1} \le p_n^* \le p_n$, $1 \le n \le N$.

We denote by $\mathcal{P}^\bullet[a, b]$ the set of all order pairs of partitions of $[a, b]$ and sets of test values for those partitions: that is,

$$\mathcal{P}^\bullet[a, b] := \{(P, P^*) : P \in \mathcal{P}[a, b] \text{ and } P^* \text{ a set of test values for } P\}.$$

We may partially order $\mathcal{P}^\bullet[a, b]$ as follows: that is, for $(P, P^*), (Q, Q^*) \in \mathcal{P}^\bullet[a, b]$, we set

$$(P, P^*) \le (Q, Q^*) \text{ if } P \subseteq Q;$$

i.e. if Q is a refinement of P.

Let $f : [a, b] \to \mathbb{R}$ be a bounded function. Corresponding to each pair $(P, P^*) \in \mathcal{P}^\bullet[a, b]$ is the real number

$$S(f, P, P^*) := \sum_{n=1}^{N} f(p_n^*)(p_n - p_{n-1}),$$

called the **Riemann sum for** f **corresponding to the pair** (P, P^*).

Recall that in our integral Calculus course, we defined f to be **Riemann integrable over** $[a, b]$ with **Riemann integral** $\int_a^b f$ if for every $\varepsilon > 0$, there exists a partition $P \in \mathcal{P}[a, b]$ such that for *any* refinement Q of P, and for *any* set Q^* of test values for Q, we have that

$$|S(f, Q, Q^*) - \int_a^b f| < \varepsilon.$$

So, what does all of this have to do with nets?

First note that $(\mathcal{P}^\bullet[a, b], \le)$ clearly satisfies the first two conditions of the definition of a direction above. Moreover, given $(P, P^*), (Q, Q^*) \in$

$\mathcal{P}^\bullet[a,b]$, the set $R = P \cup Q \in \mathcal{P}[a,b]$, and $P \subseteq R$, $Q \subseteq R$. Thus if R^* is any set of test values for R, then $(P,P^*),(Q,Q^*) \le (R,R^*)$, which completes the proof of the fact that $\mathcal{P}^\bullet[a,b]$ is a directed set.

It follows that for a bounded function $f : [a,b] \to \mathbb{R}$, the map

$$\begin{array}{rccc}\Phi_f : & \mathcal{P}^\bullet[a,b] & \to & \mathbb{R} \\ & (P,P^*) & \mapsto & S(f,P,P^*)\end{array}$$

is a net. Furthermore, the function f is Riemann integral with integral $\int_a^b f$ if and only if: given $\varepsilon > 0$, there exists $(P,P^*) \in \mathcal{P}^\bullet[a,b]$ such that if $(Q,Q^*) \in \mathcal{P}^\bullet[a,b]$ and $(P,P^*) \le (Q,Q^*)$, then

$$|S(f,Q,Q^*) - \int_a^b f| < \varepsilon.$$

This is precisely the statement that f is Riemann integral with integral $\int_a^b f$ if and only if

$$\lim_{(P,P^*)\in\mathcal{P}^\bullet[a,b]} \Phi_f((P,P^*)) = \lim_{(P,P^*)\in\mathcal{P}^\bullet[a,b]} S(f,P,P^*) = \int_a^b f.$$

Whether or not you are willing to admit it in front of your friends and family, this is very, very cool.

(d) Let $\mathcal{F} = \{F \subseteq \mathbb{N} : |F| < \infty\}$. We partially order $\mathcal{F}$ by inclusion: $F_1 \le F_2$ if $F_1 \subseteq F_2$.

Given $F_1, F_2 \in \Lambda$, $F_3 = F_1 \cup F_2 \in \mathcal{F}$ and $F_1 \le F_3$, $F_2 \le F_3$. Thus $(\mathcal{F}, \le)$ is a directed set.

Consider $X = \mathbb{R}$, equipped with the standard topology. For each $F \in \mathcal{F}$, define $x_F = \frac{1}{|F|+1} \in \mathbb{R}$. Then $(x_F)_{F\in\mathcal{F}}$ is a net in $\mathbb{R}$.

Suppose that $U \in \mathcal{U}_0$ is a neighbourhood of 0 in $\mathbb{R}$. Then there exists $\delta > 0$ so that $(-\delta,\delta) \subseteq U$. Choose $N > 0$ so that $\frac{1}{N} < \delta$. Fix $F_0 \in \mathcal{F}$ with $|F_0| \ge N$. If $F \in \mathcal{F}$ and $F \ge F_0$, then $|F| \ge |F_0| \ge N$, so $|x_F - 0| = \frac{1}{|F|+1} \le \frac{1}{N+1} < \delta$. Thus

$$\lim_{F\in\mathcal{F}} x_F = 0.$$

1.13. Example. Let (X, τ_X) be a topological space and $x \in X$. Let $\mathcal{U}_x$ denote the nbhd system at x. If, for $U_1, U_2 \in \mathcal{U}_x$ we define the relation $U_1 \le U_2$ if $U_2 \subseteq U_1$, then $(\mathcal{U}_x, \le)$ forms a directed set.

For each $U \in \mathcal{U}_x$, choose $x_U \in U$. Then $(x_U)_{U\in\mathcal{U}_x}$ forms a net in X. It is not hard to see that $\lim_{U\in\mathcal{U}_x} x_U = x$. Indeed, given $V \in \mathcal{U}_x$, we have that $x_U \in V$ for all $U \ge V$.

1.14. Definition. *Let (X,τ_X) and (Y,τ_Y) be topological spaces, and let $x_0 \in X$. We say that a function $f : X \to Y$ is* ***continuous at*** *x_0 if for all $V \in \mathcal{U}^Y_{f(x_0)}$, there exists $U \in \mathcal{U}^X_{x_0}$ such that $x \in U$ implies that $f(x) \in V$.*

We say that f is ***continuous on*** *X if f is continuous at x_0 for each $x_0 \in X$.*

The second condition in the next proposition gives a wonderfully concise way of defining (global) continuity of a function f between topological spaces.

1.15. Proposition. *Let (X, τ_X) and (Y, τ_Y) be topological spaces, and let $f : X \to Y$ be a function. The following conditions are equivalent:*

(a) *f is continuous on X; that is, f is continuous at each $x_0 \in X$ in the sense above.*
(b) *If $G \subseteq Y$ is open, then $f^{-1}(G) := \{x \in X : f(x) \in G\}$ is open in X.*

Proof. See the Assignments.

□

That this extends our usual notion of continuity for functions between metric space is made clear by the following result:

1.16. Proposition. *If (X, d_X) and (Y, d_Y) are metric spaces with metric space topologies τ_X and τ_Y respectively, then the following are equivalent for a function $f : X \to Y$:*

(a) *f is continuous on X, i.e. $f^{-1}(G) \in \tau_X$ for all $G \in \tau_Y$.*
(b) *$\lim_n f(x_n) = f(x)$ whenever $(x_n)_{n=1}^\infty$ is a sequence in X converging to $x \in X$.*

Proof. See the Assignments.

□

As we shall see in the Assignments, sequences are not enough to describe convergence, nor are they enough to characterize continuity of functions between general topological spaces. On the other hand, nets are sufficient for this task, and serve as the natural replacement for sequences. (The following result also admits a *local* version, which we shall also see in the Assignments.)

1.17. Theorem. *Let (X, τ_X) and (Y, τ_Y) be topological spaces. Let $f : X \to Y$ be a function. The following are equivalent:*

(a) *f is continuous on X.*
(b) *Whenever $(x_\lambda)_{\lambda \in \Lambda}$ is a net in X which converges to $x \in X$, it follows that $(f(x_\lambda))_{\lambda \in \Lambda}$ is a net in Y which converges to $f(x)$.*

Proof. See the Assignments.

□

1.18. Definition. *Let (X, τ) be a topological space, (Y, d) be a metric space, and suppose that $(f_n)_{n=1}^\infty$, f are functions from X into Y. We say that $(f_n)_{n=1}^\infty$* ***converges uniformly*** *to f if for all $\varepsilon > 0$ there exists $N > 0$ so that $n \geq N$ implies that*

$$\sup_{x \in X} d(f(x), f_n(x)) < \varepsilon.$$

1.19. Example. Let $X = \mathbb{R}$, $f : \mathbb{R} \to \mathbb{R}$ be the function $f(x) = x$, $x \in \mathbb{Q}$, $f(x) = 0$, $x \notin \mathbb{Q}$. Set $f_n(x) = f(x) + (1/n) \sin x$, $x \in \mathbb{R}$. Then $(f_n)_{n=1}^\infty$ converges uniformly to f. Observe that none of the functions f or f_n, $n \geq 1$ are bounded.

1.20. Proposition. *Let (X, τ) be a topological space, (Y, d) be a metric space, and suppose that $(f_n)_{n=1}^\infty$ is a sequence of continuous functions from X into Y which converges uniformly to a function $f : X \to Y$. Then f is continuous.*

Proof. Let $x_0 \in X$ be arbitrary and $\varepsilon > 0$. Find $N > 0$ so that $n \geq N$ implies that $d(f(x), f_n(x)) < \varepsilon/3$ for all $x \in X$. Since f_N is continuous, we can find a neigbourhood $U \in \mathcal{U}_{x_0}$ so that $x \in U$ implies $d(f_N(x), f_N(x_0)) < \varepsilon/3$. For $x \in U$,

$$\begin{aligned} d(f(x), f(x_0)) &\leq d(f(x), f_N(x)) + d(f_N(x), f_N(x_0)) + d(f_N(x_0), f(x_0)) \\ &< \varepsilon/3 + \varepsilon/3 + \varepsilon/3. \end{aligned}$$

It follows that f is continuous at x_0. Since $x_0 \in X$ was arbitrary, f is continuous on X.

□

2. A characterization of compactness for metric spaces

2.1. The notion of compactness should be viewed as a generalization of "finiteness".

2.2. Definition. *Let (X, τ) be a topological space and $E \subseteq X$. A collection $\{H_\gamma\}_{\gamma \in \Gamma}$ is said to be a* ***cover*** *of E if $E \subseteq \cup_{\gamma \in \Gamma} H_\gamma$. The cover is said to be* ***open*** *if each $H_\gamma \in \tau$, $\gamma \in \Gamma$.*

Given a cover of E as above, a ***finite subcover*** *of E is a finite subcollection $\{H_{\gamma_k}\}_{k=1}^n$ of $\{H_\gamma\}_{\gamma \in \Gamma}$ so that $E \subseteq \cup_{k=1}^n H_{\gamma_k}$.*

Finally, a subset $K \subseteq X$ is said to be ***compact*** *if every open cover of K admits a finite subcover.*

2.3. Example. Suppose that (X, τ) is a topological space and that $F \subseteq X$ is finite. Then F is compact.

2.4. Theorem. *Consider $\mathbb{R}$, equipped with the standard topology. If $a < b \in \mathbb{R}$, then $[a, b]$ is compact.*

Proof. We argue by contradiction. Assume otherwise, and let $\{G_\lambda\}_{\lambda \in \Lambda}$ be an open cover of $[a, b]$ which admits no finite subcover. Let $I_1 = [a, b]$.

We split I_1 into two subintervals $J_{1,1}$ and $J_{1,2}$ of equal length (i.e. $J_{1,1} = [a, c]$ and $J_{1,2} = [c, b]$, where $c = (a+b)/2$. Either $J_{1,1}$ or $J_{1,2}$ can not be covered by finitely many of the G_λ's (otherwise the union of the finitely many G_λ's would cover I_1).

If $J_{1,1}$ can not be so covered, set $I_2 = J_{1,1}$. Otherwise, let $I_2 = J_{1,2}$. We now proceed inductively. Given $I_k = [a_k, b_k]$ which can not be covered by finitely many of the G_λ's, set $J_{k,1} = [a_k, c_k]$ and $J_{k,2} = [c_k, b_k]$, where $c_k = (a_k + b_k)/2$. One of $J_{k,1}$, $J_{k,2}$ can not be covered by finitely many of the G_λ's. If $J_{k,1}$ cannot be so covered, set $I_{k+1} = J_{k,1}$, otherwise set $I_{k+1} = J_{k,2}$. Write $I_{k+1} = [a_{k+1}, b_{k+1}]$.

Observe that

$$I_1 \supseteq I_2 \supseteq I_3 \supseteq \cdots,$$

and that $\operatorname{diam} I_k := (b_k - a_k) = \frac{b-a}{2^{k-1}}$ converges to 0 as k tends to infinity. Thus $\cap_{k=1}^\infty I_k = \{z\}$, a single point (by the Nested Intervals Theorem). (Recall that $z =$

$\lim_k a_k = \sup_k a_k = \inf_k b_k = \lim_k b_k$.) Now, $z \in [a, b] \in \cup_{\lambda \in \Lambda} G_\lambda$, and so $z \in G_{\lambda_0}$ for some $\lambda_0 \in \Lambda$.

Since G_{λ_0} is open, there exists $\delta > 0$ so that $(z - \delta, z + \delta) \subseteq G_{\lambda_0}$. But then if m is sufficiently large, $I_m \subseteq (z - \delta, z + \delta) \subseteq G_{\lambda_0}$, contradicting the fact that I_m could not be covered by finitely many of the G_λ's.

Thus $[a, b]$ must be compact.

□

2.5. Example. Consider $\mathbb{R}$, equipped with the standard topology. Then $E = (0, 1]$ is not compact. For example, if we let $G_x = (x, 2)$, $x \in (0, 1)$, then $\{G_x\}_{x \in (0,1)}$ is an open cover of E which does not admit a finite subcover.

2.6. Examples.

(a) Let (X, τ) be a space endowed with the trivial topology $\tau = \{\varnothing, X\}$. Then every subset $K \subseteq X$ is compact.

(b) Let (X, μ) be a space endowed with the discrete topology. Then $K \subseteq X$ is compact if and only if K is finite.

2.7. Definition. *Let (X, τ) be a topological space, and suppose that $Y \subseteq X$. The **relative topology** on Y inherited from X is*

$$\tau_Y := \{G \cap Y : G \in \tau\}.$$

*We say that $H \in \tau_Y$ is **relatively open** in Y. Similarly, $K \subseteq Y$ is **relatively closed** in Y if $Y \setminus K$ is relatively open in Y.*

We leave it as an exercise for the reader to show that this is indeed a topology on Y.

2.8. Example. Suppose that (X, τ) is a topological space and that $Y \subseteq X$ is open. Then

$$\tau_Y = \{G \cap Y : G \in \tau\} = \{H \subseteq Y : H \in \tau\}.$$

That is, $H \in \tau_Y$ if and only if $H \subseteq Y$ and H is open in X.

2.9. Example.

(a) Consider $\mathbb{R}$ equipped with the standard topology τ. Let $\mathbb{Z} \subseteq \mathbb{R}$. For any $n \in \mathbb{Z}$,

$$\{n\} = (n - 1, n + 1) \cap \mathbb{Z} \in \tau_{\mathbb{Z}}.$$

That is, each point is open in $\mathbb{Z}$ in the relative topology. It follows that if $A \subseteq \mathbb{Z}$, then $A = \cup_{n \in A} \{n\}$ is a union of open sets, so every subset of $\mathbb{Z}$ is open in the relative topology. That is, the relative topology on $\mathbb{Z}$ is the discrete topology on $\mathbb{Z}$, which coincides with the metric topology on $\mathbb{Z}$ inherited from the discrete metric.

(b) Consider $\mathbb{R}$ equipped with the standard topology τ, and let $Y = (0,1]$. Then $(0,1/2)$ is relatively open in Y, since $(0,1/2)$ is open in $\mathbb{R}$ and $(0,1/2) = (0,1/2) \cap Y$.

More interestingly, $(1/2,1]$ is relatively open in Y, since $(1/2,2)$ is open in $\mathbb{R}$ and $(1/2,1] = (1/2,2) \cap Y$.

It is clear from the above example that if (X,τ) is a topological space and $H \subseteq Y \subseteq X$, then H may be relatively open in Y without being open in X. Compactness, however, does not depend upon the ambient space.

2.10. Theorem. *Let (X,τ) be topological space and $Y \subseteq X$. Suppose that $K \subseteq Y \subseteq X$. The following are equivalent:*

(a) *K is compact as a subset of (X,τ).*
(b) *K is compact as a subset of (Y,τ_Y).*

Proof.

(a) Suppose that K is compact as a subset of (X,τ), and let $\{H_\lambda\}_{\lambda\in\Lambda}$ be a τ_Y-open cover of K. That is, $H_\lambda \in \tau_Y$ for all $\lambda \in \Lambda$, and $K \subseteq \cup_\lambda H_\lambda$.

By definition of the relative topology on Y, for each $\lambda \in \Lambda$, there exists $G_\lambda \in \tau$ so that $H_\lambda = G_\lambda \cap Y$. Thus

$$K \subseteq \cup_\lambda H_\lambda \subseteq \cup_\lambda G_\lambda.$$

Since K is compact in (X,τ), we can find $\lambda_1, \lambda_2, ..., \lambda_n \in \Lambda$ so that

$$K \subseteq \cup_{k=1}^n G_{\lambda_k}.$$

But then

$$K = K \cap Y \subseteq (\cup_{k=1}^n G_{\lambda_k}) \cap Y = \cup_{k=1}^n (G_{\lambda_k} \cap Y) = \cup_{k=1}^n H_{\lambda_k}.$$

It follows that every τ_Y-open cover of K admits a finite subcover, and so K is compact in (Y,τ_Y).

(b) Conversely, suppose that K is compact in (Y,τ_Y) and let $\{G_\lambda\}_{\lambda\in\Lambda}$ be a τ-open cover of K in X. For all λ, let $H_\lambda = G_\lambda \cap Y$, so that $H_\lambda \in \tau_Y$.

Since $K = K \cap Y$, it follows that $K \subseteq \cup_\lambda H_\lambda$. Since K is compact in (Y,τ_Y), there exist $\lambda_1, \lambda_2, ..., \lambda_n \in \Lambda$ so that

$$K \subseteq \cup_{k=1}^n H_{\lambda_n} \subseteq \cup_{k=1}^n G_{\lambda_n}.$$

It follows that K is compact in (X,τ).

□

2.11. Definition. *A topological space (X,τ) is said to be* ***Hausdorff*** *or T_2 if for each pair $x, y \in X$ with $x \neq y$, there exist neighbourhoods $U \in \mathcal{U}_x$ and $V \in \mathcal{U}_y$ so that $U \cap V = \varnothing$. We also say that x and y can be* ***separated.***

2.12. Examples.

(a) Let (X, d) be a metric space and τ denote the metric topology on X induced by d. If $x, y \in X$ and $x \neq y$, then $\beta := \frac{1}{2}\, d(x, y) > 0$. Also, $B(x, \beta),\ B(y, \beta) \in \tau$ and $B(x, \beta) \cap B(y, \beta) = \varnothing$. Hence (X, τ) is Hausdorff.

(b) Let $X = \{a, b\}$ and $\tau = \{\varnothing, \{a\}, X\}$. Then a and b can not be separated, since the only open set which contains b is X itself, which also contains a. Thus X is not Hausdorff.

As we shall see in the Assignments, the Hausdorff property of a topological space is what guarantees that limits of nets are unique.

2.13. Theorem. *Suppose that (X, τ) is a Hausdorff space and that $K \subseteq X$ is compact. Then K is closed. In particular, every singleton set in a Hausdorff space is closed.*

Proof. Our goal is to show that $M := X \setminus K$ is open. To that end, let $x \in M$. Since X is Hausdorff, for each $y \in K$, we can find disjoint neighbourhoods U_y of y and V_y of x. Without loss of generality, we may assume that U_y and V_y are open (*why?*).

Thus $\{U_y\}_{y \in K}$ is an open cover of K, and since K is compact, we can find $y_1, y_2, ..., y_n \in K$ so that

$$K \subseteq \cup_{k=1}^n U_{y_k}.$$

Let $V = \cap_{k=1}^n V_{y_k}$. Then V is open, being the finite intersection of open sets, and $V \cap U_{y_k} = \varnothing$, $1 \leq k \leq n$, which implies that $V \cap K = \varnothing$. That is, $x \in V \subseteq M$. Hence M is open, or equivalently, K is closed.

The last statement follows from above, keeping in mind that every finite set in a topological space is compact.

□

2.14. Corollary. *Every compact subset of a metric space is closed in the metric topology.*

2.15. Proposition. *Let (K, τ) be a compact topological space, and $F \subseteq K$ be closed. Then F is compact.*

Proof. Let $\{G_\lambda\}_{\lambda \in \Lambda}$ be an open cover of F in K. Since F is closed, observe that $G = K \setminus F$ is also open, and that

$$K = F \cup G \subseteq (\cup_\lambda G_\lambda) \cup G.$$

That is, $\{G_\lambda\}_\lambda \cup \{G\}$ is an open cover of K. Since K is compact, there exist $\lambda_1, \lambda_2, ..., \lambda_n \in \Lambda$ so that

$$K \subseteq (\cup_{k=1}^n G_{\lambda_k}) \cup G.$$

But then

$$F = K \cap F \subseteq ((\cup_{k=1}^n G_{\lambda_k}) \cup G) \cap F \subseteq \cup_{k=1}^n G_{\lambda_k}.$$

That is, $\{G_\lambda\}_{\lambda \in \Lambda}$ admits a finite subcover of F.

This shows that F is compact.

□

The duality between open and closed sets in a topological space means that every property which may be described in terms of one admits an equivalent formulation in terms of the other. Let us now derive such a formulation of compactness in terms of closed sets.

2.16. Definition. *A collection $\{F_\lambda\}_{\lambda\in\Lambda}$ is said to have the* ***finite intersection property*** FIP *if every finite subcollection $\{F_{\lambda_1}, F_{\lambda_2}, ..., F_{\lambda_n}\}$ has non-empty intersection; i.e.*

$$\cap_{k=1}^n F_{\lambda_k} \neq \varnothing.$$

2.17. Example. For each $t \in (0,1)$, consider the set $H_t = (t,2) \subseteq \mathbb{R}$. Let $\mathcal{H} = \{H_t : t \in (0,1)\}$. Then $\mathcal{H}$ has the FIP.

2.18. Theorem. *Let (X,τ) be a topological space. The following are equivalent.*

(a) *X is compact.*
(b) *If $\mathcal{F} = \{F_\lambda\}_{\lambda\in\Lambda}$ is a collection of closed subsets of X and $\mathcal{F}$ has the* FIP*, then*

$$\cap_{\lambda\in\Lambda} F_\lambda \neq \varnothing.$$

Proof.

(a) implies (b).

Suppose that X is compact. Let $\mathcal{F} = \{F_\lambda\}_{\lambda\in\Lambda}$ be a collection of closed subsets of X which has the FIP. For all $\lambda \in \Lambda$, set $G_\lambda = X \setminus F_\lambda$, so that each G_λ is open. If, to the contrary, $\cap_\lambda F_\lambda = \varnothing$, then $\cup_\lambda G_\lambda = X$. Since X is compact, we can find $\lambda_1, \lambda_2, ..., \lambda_n \in \Lambda$ so that $X = \cup_{k=1}^n G_{\lambda_k}$. But then $\varnothing = \cap_{k=1}^n F_{\lambda_k}$, contradicting the fact that $\mathcal{F}$ has the FIP. Thus $\cap_{\lambda\in\Lambda} F_\lambda \neq \varnothing$.

(b) implies (a).

Now suppose that (b) holds but that X is not compact. Then there exists an open cover $\{G_\lambda\}_{\lambda\in\Lambda}$ which does not admit a finite subcover. Letting $F_\lambda = X \setminus G_\lambda$, $\lambda \in \Lambda$, we see that $\mathcal{F} = \{F_\lambda\}_{\lambda\in\Lambda}$ consists of closed sets, has the FIP, and yet $\cap_{\lambda\in\Lambda} F_\lambda = \varnothing$, contradicting the hypothesis of (b). Thus (b) implies that X is compact.

□

Using the above result, we get the following extension of the Nested Intervals Theorem.

2.19. Corollary. *Let (X,τ) be a compact topological space, and suppose that*

$$F_1 \supseteq F_2 \supseteq F_3 \supseteq \cdots$$

is a sequence of non-empty closed subsets of X. Then $\cap_{n\geq 1} F_n \neq \varnothing$.

Recall the definition of an *accumulation point* in a metric space from Definition 2.2.11. Closer inspection shows that the definition did not depend upon the metric in any significant way, and so it extends naturally to topological spaces.

2.20. Definition. *Let (X,τ) be a topological space and $E \subseteq X$. A point $x \in X$ is said to be an* ***accumulation point*** *of E if for all $U \in \mathcal{U}_x$, the* ***punctured neighbourhood*** *$U \setminus \{x\}$ of x intersects E non-trivially. That is, $(U \setminus \{x\}) \cap E \ne \varnothing$.*

As we did in the metric space setting, we denote by E' the set of accumulation points of E.

2.21. Example. Consider $\mathbb{N}$, equipped with the co-finite topology τ_{cf} from Example 1.4.

Let $H \subseteq \mathbb{N}$ be any infinite set. We claim that $H' = \mathbb{N}$; that is, every point $n \in \mathbb{N}$ is an accumulation point of H.

Indeed, let $U \in \mathcal{U}_n$. Then $\mathbb{N} \setminus U$ is finite, and so there exists a positive integer $N > 0$ so that $\{N+1, N+2, N+3, ...\} \subseteq U$. In particular, if $M = \max(N, n)$, then $\{M+1, M+2, M+3, ...\} \subseteq U \setminus \{n\}$

But $U \setminus \{n\}$ is a punctured nbhd of n, and $(U \setminus \{n\}) \cap H \ne \varnothing$, as H is an infinite set, and as such it is unbounded.

Since every punctured nbhd of n intersects H non-trivially, $n \in H'$. Since $n \in \mathbb{N}$ was arbitrarily chosen, $H' = \mathbb{N}$.

2.22. Definition. *A subset D of a topological space (X,τ) is said to be* ***countably compact*** *if every infinite subset of D has an accumulation point in D.*

A subset S of (X,τ) is said to be ***sequentially compact*** *if every sequence in S contains a subsequence which converges to a point in S.*

The most general relationship which exists between these different notions of compact is the following:

2.23. Proposition. *Let (X,τ) be a topological space and $H \subseteq X$. If H is either compact or sequentially compact, then H is countably compact.*

Proof.

- Suppose that H is compact. If H is not countably compact, then there must exist an infinite subset B of H which does not have an accumulation point in H. As such, for each $x \in H$, there exists an (open) neighbourhood $U_x \in \mathcal{U}_x$ so that $(U_x \setminus \{x\}) \cap B = \varnothing$, or equivalently, $U_x \cap B \subseteq \{x\}$. The collection $\{U_x\}_{x \in H}$ is clearly an open cover of H. Since H is compact, there exists a finite subcover $\{U_{x_k}\}_{k=1}^n$. But then

$$B = H \cap B \subseteq \cup_{k=1}^n (U_{x_k} \cap B) \subseteq \cup_{k=1}^n \{x_k\},$$

  contradicting the fact that B is infinite. Thus H must be countably compact.
- Suppose that H is sequentially compact. Let $B \subseteq H$ be an infinite set. Then B admits a denumerable subset $\{b_k\}_{k=1}^\infty$. Since H is sequentially compact, we can find a subsequence $(b_{k_m})_{m=1}^\infty$ of the *sequence* $(b_k)_{k=1}^\infty$ and $b \in H$ so that $\lim_m b_{k_m} = b$.

  Let $U \in \mathcal{U}_b$. By definition of convergence, there exists $M > 0$ so that $m \ge M$ implies that $b_{k_m} \in U$. Since $j \ne k$ implies that $b_j \ne b_k$, at most one

b_{k_m} can be equal to b, and so $U \setminus \{b\}$ intersects B non-trivially. That is, b is an accumulation point of B in H, proving that H is countably compact.

□

2.24. Example. Let τ be the topology on $\mathbb{N}$ **generated by the sets** $T_n = \{2n-1, 2n\}$, $n \ge 1$. (See the Assignments for the notion of a topology *generated by a subbase.*)

Let $\varnothing \ne H \subseteq \mathbb{N}$, and suppose $m \in H$.

- If m is odd, then $m+1$ is an accumulation point of H. Indeed, let $G \in \tau$ and suppose that $m+1 \in G$. Then $m \in G$. Let U be any neighbourhood of $m+1$. Then U contains an open neighbourhood of $m+1$, and so from above, $m \in U$. That is, $(U \setminus \{m+1\}) \cap H \supseteq \{m\} \ne \varnothing$.
- If m is even, then a similar argument shows that $m-1$ is an accumulation point of H.

It follows that $\mathbb{N}$ is countably compact.

Observe that $\mathbb{N}$ is *not* compact. Indeed, $\mathbb{N} = \cup_{n=1}^{\infty} T_n$ and each T_n is open (by our definition of τ!), but the cover $\{T_n\}_{n=1}^{\infty}$ does not admit a finite subcover.

Moreover, $(\mathbb{N}, \tau)$ is not sequentially compact. For example, consider the sequence $(x_k)_{k=1}^{\infty}$, where $x_k = k$, $k \ge 1$. Suppose that there exists $m \in \mathbb{N}$ and a subsequence $(x_{k_j})_{j=1}^{\infty}$ so that $\lim_j x_{k_j} = m$. Then there exists $m' \in \{m-1, m+1\}$ so that $U = \{m, m'\}$ is an (open) neighbourhood of m. But $\{x_{k_j}\}_{j=1}^{\infty}$ is not even bounded, so it cannot be that the tail of the sequence $(x_{k_j})_j$ would lie in U, thereby establishing our claim.

2.25. Remark. We mention in passing that the notions of compactness and of sequential compactness are distinct. It can be shown that there exist compact topological spaces which are not sequentially compact, and sequentially compact topological spaces which are not compact. Unfortunately, the proofs of each of these statements is beyond the scope of the course.

The standard example of a compact space which is not sequentially compact is obtained by considering an uncountable direct product of the compact interval $[0, 1]$, equipped with the so-called **product topology**.

The standard example of a sequentially compact space which is not compact is obtained by consider the first **uncountable ordinal** Ω, and imbuing Ω with the **order topology**. This is the topology on Ω generated by sets of the form

$$\{x \in \Omega : x < a\} \text{ and } \{x \in \Omega : x > a\}$$

where $a \in \Omega$. It follows that the open sets in the order topology are unions of *open intervals.*

The reader is directed to the excellent book by Stephen Willard for a further discussion of these topics.

Thus, while *compact*, *sequentially compact* and *countably compact* are, in general, distinct concepts, our goal is to prove that these notions coincide in the setting of metric spaces. We first require two more definitions.

2.26. Definition. *Let (X, d) be a metric space and let $E \subseteq X$. A finite set $\{x_1, x_2, ..., x_n\} \subseteq X$ is called an* ε***-net*** *for E if*

$$E \subseteq \cup_{k=1}^{n} B(x_k, \varepsilon).$$

We say that E is ***totally bounded*** *if for each $\varepsilon > 0$, E admits an ε-net.*

It is not hard to verify that if $E \subseteq X$ is totally bounded and $H \subseteq E$, then H is also totally bounded.

2.27. Example. Let $\mathbb{I} = [0,1] \subseteq \mathbb{R}$, with the standard metric $d(x,y) = |x - y|$, $x, y \in [0,1]$. First $\varepsilon > 0$ and choose $N > \frac{1}{\varepsilon}$. Then

$$\{0, \frac{1}{N}, \frac{2}{N}, \cdots, \frac{N-1}{N}, 1\}$$

is an ε-net for $\mathbb{I}$, since $\mathbb{I} \subseteq \cup_{k=0}^{N} (\frac{k}{N} - \varepsilon, \frac{k}{N} + \varepsilon)$.

Since $\varepsilon > 0$ was arbitrary, $\mathbb{I}$ is totally bounded.

2.28. Definition. *Let (X, d) be a metric space. Given $E \subseteq X$, the* ***diameter*** *of E is*

$$\operatorname{diam} E = \sup\{d(x,y) : x, y \in E\}.$$

We say that E is ***bounded*** *if* $\operatorname{diam} E < \infty$.

We leave it to the reader to check that a subset of $E \subseteq \mathbb{R}^n$ is bounded in the metric sense if and only if E is bounded in the usual sense.

2.29. Examples.

(a) Consider $\mathbb{R}^2$ equipped with the standard (Euclidean) metric

$$d_2((x_1, y_1), (x_2, y_2)) = \sqrt{|x_1 - x_2|^2 + |y_1 - y_2|^2}.$$

Let $E = [0,1] \times (0,1) = \{(x,y) : 0 \le x \le 1, 0 < y < 1\}$. Then $\operatorname{diam} E = \sqrt{2} < \infty$, so E is bounded.

(b) Let (X, d) be a metric space, $x \in X$ and $\varepsilon > 0$. Then $\operatorname{diam} B(x, \varepsilon) \le 2\varepsilon$. To see this, note that if $y, z \in B(x, \varepsilon)$, then $d(y,z) \le d(y,x) + d(x,z) < 2\varepsilon$.

2.30. Proposition. *Let (X,d) be a metric space and $E \subseteq X$. The following are equivalent.*

(a) *E is totally bounded.*

(b) *For every $\varepsilon > 0$, there exists a decomposition $E = \dot\bigcup_{k=1}^n E_k$ with $\operatorname{diam} E_k < \varepsilon$ for all $1 \le k \le n$.*

Proof.

(a) implies (b). Suppose that E is totally bounded. Let $\varepsilon > 0$ and let $\{x_k\}_{k=1}^n$ be an $\varepsilon/3$-net for E. Let $E_1 = B(x_1, \varepsilon/3)$, and for $2 \le k \le n$, let

$$E_k = \left(B(x_k, \varepsilon/3) \setminus (\dot\bigcup_{i=1}^{k-1} E_i) \right) \cap E.$$

By construction, the E_k's are disjoint. Also, $E_k \subseteq B(x_k, \varepsilon/3)$ implies that $\operatorname{diam} E_k \le \operatorname{diam} B(x_k, \varepsilon/3) \le \dfrac{2\varepsilon}{3} < \varepsilon$, and clearly

$$\cup_{k=1}^n E_k = (\cup_{k=1}^n B(x_k, \varepsilon/3)) \cap E = E.$$

(b) implies (a). Suppose that E satisfies condition (b), and let $\varepsilon > 0$. Choose $x_k \in E_k$, $1 \le k \le n$. Then $E_k \subseteq B(x_k, \varepsilon)$, as $\operatorname{diam} E_k < \varepsilon$, $1 \le k \le n$. Hence

$$E = \dot\bigcup_{k=1}^n E_k \subseteq \cup_{k=1}^n B(x_k, \varepsilon),$$

proving that E is totally bounded.

□

2.31. Proposition. *Let (X,d) be a metric space. If X is totally bounded, then X is bounded.*

Proof. Let $\varepsilon = 1$ and choose an ε-net $\{x_k\}_{k=1}^n$ for X, which exists because X is totally bounded. Thus $X = \bigcup_{k=1}^n B(x_k, \varepsilon)$.

Let $\rho = \max_{1 \le j,k \le n} d(x_j, x_k)$. Given $x, y \in X$, there exist $1 \le r, s \le n$ so that $x \in B(x_r, \varepsilon)$ and $y \in B(x_s, \varepsilon)$. Thus

$$\begin{aligned} d(x,y) &\le d(x, x_r) + d(x_r, x_s) + d(x_s, y) \\ &\le \varepsilon + \rho + \varepsilon \\ &= \rho + 2. \end{aligned}$$

Since $\rho + 2$ is a fixed constant, X is bounded.

□

2.32. Example. The converse to Proposition 2.31 is false. To see this, we begin with a simple observation. Suppose that a metric space (X,d) is totally bounded. Let $\{x_k\}_{k=1}^n$ be a $\frac{1}{2}$-net for X, so that

$$X = \bigcup_{k=1}^n B(x_k, \frac{1}{2}).$$

Let $m > n$, and suppose that $\{y_1, y_2, ..., y_m\}$ are m distinct points in X. Then there exist $1 \le i \ne j \le m$ so that $d(y_i, y_j) < 1$. Indeed, the Pigeonhole Principle implies that there exist $1 \le i \ne j \le m$ and $1 \le k \le n$ so that $y_i, y_j \in B(x_k, \frac{1}{2})$. An easy application of the triangle inequality show that $d(y_i, y_j) < 1$.

Consider ℓ_∞ (from Example 2.1.7), equipped with the supremum norm

$$\|(z_n)_n\|_\infty = \sup_{n \ge 1} |z_n|.$$

Let $X = \{z = (z_n)_n \in \ell_\infty : \|z\|_\infty \le 1\}$ be the unit ball of ℓ_∞, and denote by d the metric $d(w, z) = \|w - z\|_\infty$. Given $w = (w_n)_n$ and $z = (z_n)_n \in X$,

$$d(w, z) = \|w - z\|_\infty \le \|w\|_\infty + \|z\|_\infty \le 2.$$

Thus (X, d) is bounded.

For each $k \ge 1$, let $\mathbf{e}_k = (e_n^{(k)})_{n=1}^\infty$, where $e_n^{(k)} = 1$ if $n = k$ and $e_n^{(k)} = 0$ otherwise. It is easily seen that $\mathbf{e}_k \in X$ for all $k \ge 1$, and if $1 \le i \ne j$, then

$$d(\mathbf{e}_i, \mathbf{e}_j) = \|\mathbf{e}_i - \mathbf{e}_j\|_\infty = 1.$$

From the first paragraph, we see that (X, d) is not totally bounded. Indeed, suppose that $\{x_1, x_2, \ldots, x_N\}$ is a $\frac{1}{3}$-net for X. Consider $\{\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_{N+1}\} \in X$. By the first paragraph of this section, there must exist $1 \le i \ne j \le N + 1$ such that $d(\mathbf{e}_i, \mathbf{e}_j) \le 2\varepsilon < 1$, a contradiction. Hence X does not admit a $\frac{1}{3}$-net, and so X is not totally bounded.

As a second example, suppose that Y is an infinite set, and that μ represents the discrete metric on Y. Then clearly Y is bounded, since $\operatorname{diam} Y = \sup\{\mu(y, z) : y, z \in Y\} = 1$.

However, if $0 < \varepsilon < 1$, then any open ball $B(y_k, \varepsilon)$ of radius ε can contain at most one point of Y (check!), and thus Y does not admit an ε-net. Hence Y is not totally bounded.

2.33. Proposition. *Suppose that (X, d) is a metric space and that $H \subseteq X$ is sequentially compact. Then H is totally bounded.*

Proof. We argue by contradiction. Suppose that H is not totally bounded. Then there exists $\varepsilon > 0$ for which X does not admit an ε-net. Let $x_1 \in X$ be chosen arbitrarily. Since $\{x_1\}$ is not an ε-net for X, $X \ne B(x_1, \varepsilon)$. Choose $x_2 \in X \setminus B(x_1, \varepsilon)$. Since $\{x_1, x_2\}$ is not an ε-net for X, we can find $x_3 \in X \setminus (B(x_1, \varepsilon) \cup B(x_2, \varepsilon))$.

More generally, for $n \ge 2$, having chosen $\{x_1, x_2, ..., x_n\}$ as above, the fact that $\{x_k\}_{k=1}^n$ is not an ε-net implies that there exists $x_{n+1} \in X \setminus \bigcup_{k=1}^n B(x_k, \varepsilon)$.

Consider the sequence $(x_n)_n$ in $X^{\mathbb{N}}$. For any $1 \le n < m$, $x_m \notin \bigcup_{k=1}^{m-1} B(x_k, \varepsilon)$ implies that $x_m \notin B(x_n, \varepsilon)$, and so $d(x_n, x_m) \ge \varepsilon$. From this it follows that no subsequence $(x_{n_k})_k$ of $(x_n)_n$ is Cauchy, and thus no subsequence of $(x_n)_n$ can converge. This contradicts the hypothesis that H is sequentially compact.

The proof now follows.

□

2.34. Definition. *Let (X,d) be a metric space and $E \subseteq X$. Suppose that $\{G_\lambda\}_{\lambda\in\Lambda}$ is a cover of E. We say that $\delta > 0$ is a* ***Lebesgue number*** *for $\{G_\lambda\}_{\lambda\in\Lambda}$ if whenever $B \subseteq E$ and* $\operatorname{diam} B < \delta$ *it follows that there exists $\beta \in \Lambda$ so that $B \subseteq G_\beta$.*

2.35. Example. Let $\mathbb{I} = [0,1] \subseteq \mathbb{R}$, equipped with the standard metric topology as in Example 2.27. Let $G_1 = (-\frac{1}{2}, \frac{1}{2})$, $G_2 = (0,1)$ and $G_3 = (\frac{1}{2}, \frac{3}{2})$. Clearly $\{G_1, G_2, G_3\}$ is a cover of $\mathbb{I}$.

Let $\delta = \frac{1}{4}$, $B \subseteq \mathbb{I}$, and suppose that $\operatorname{diam} B < \delta$.

If $b \in B$, then either

- $b \in [0, \frac{1}{4}]$, so $B \subseteq (-\frac{1}{4}, \frac{1}{2}) \subseteq G_1$, or
- $b \in (\frac{1}{4}, \frac{3}{4})$ so $B \subseteq (0,1) = G_2$, or
- $b \in [\frac{3}{4}, 1]$, so $B \subseteq (\frac{1}{2}, \frac{5}{4}) \subseteq G_3$.

This proves that $\delta = \frac{1}{4}$ is a Lebesgue number for $\{G_1, G_2, G_3\}$.

2.36. Lemma. (Lebesgue) *Let (X,d) be a metric space and $H \subseteq X$ be sequentially compact. Let $\mathcal{G} = \{G_\lambda\}_{\lambda\in\Lambda}$ be an open cover of H. Then $\mathcal{G}$ has a (positive) Lebesgue number.*

Proof. Suppose otherwise.

Then, for each $n \ge 1$, we can find a set $B_n \subseteq H$ so that $\operatorname{diam} B_n < \frac{1}{n}$, and there does not exist $\lambda \in \Lambda$ for which $B_n \in G_\lambda$. Choose $b_n \in B_n$, $n \ge 1$. (This uses the countable version of the Axiom of Choice!) Since H is sequentially compact, there exists a subsequence $(b_{n_k})_k$ of $(b_n)_n$ which converges to some element $h \in H$.

But $\mathcal{G}$ is an open cover of H, and so $h \in G_\beta$ for some $\beta \in \Lambda$. Since G_β is open, there exists some $\delta > 0$ so that $B(h,\delta) \subseteq G_\beta$. Since $\lim_k b_{n_k} = h$, we may find $N > \frac{2}{\delta}$ so that $k \ge N$ implies that

$$d(b_{n_k}, h) < \frac{\delta}{2}.$$

If $k \ge N$, then $\frac{1}{n_k} \le \frac{1}{k} \le \frac{1}{N} < \frac{\delta}{2}$, and for all $b \in B_{n_k}$,

$$d(b,h) \le d(b, b_{n_k}) + d(b_{n_k}, h) < \operatorname{diam} B_{n_k} + \frac{\delta}{2} \le \frac{1}{n_k} + \frac{\delta}{2} < \delta.$$

Hence $b \in B(h,\delta) \subseteq G_\beta$, and so $B_{n_k} \subseteq G_\beta$, a contadiction.

Thus $\mathcal{G}$ has a positive Lebesgue number.

□

2.37. Theorem. *Let (X,d) be a metric space and $K \subseteq X$. The following are equivalent.*

(a) *K is compact.*
(b) *K is countably compact.*
(c) *K is sequentially compact.*

Proof.

(a) implies (b).

This is Proposition 2.23.

(b) implies (c).

Let $(x_n)_n$ be a sequence in K. If $(x_n)_n$ admits a constant subsequence $(x_{n_k})_k$, then that subsequence converges to its constant value, which lies in K.

If $(x_n)_n$ does not admit a constant subsequence, then $H = \{x_n\}_{n=1}^{\infty}$ is a denumerable subset of K which admits an accumulation point $h \in K$, by countable compactness of the latter. As we shall see in the Assignments, this implies that there exists a subsequence $(x_{n_k})_k$ of $(x_n)_n$ which converges to $h \in K$.

Thus K is sequentially compact.

(c) implies (a).

Suppose that K is sequentially compact, and let $\mathcal{G} = \{G_\lambda\}_{\lambda \in \Lambda}$ be an open cover of K. By Lebesgue's Lemma, $\mathcal{G}$ admits a positive Lebesgue number, say $\rho > 0$. Also, by Proposition 2.33, K is totally bounded.

Next, by Proposition 2.30, we can partition K as $K = \dot\bigcup_{k=1}^n E_k$, where $\operatorname{diam} E_k < \rho$ for each $1 \le k \le n$. But then by definition of the Lebesgue number, for each such $1 \le k \le n$, there exists $\beta_k \in \Lambda$ so that $E_k \subseteq G_{\beta_k}$, and so

$$K = \dot{\bigcup_{k=1}^{n}} E_k \subseteq \bigcup_{k=1}^{n} G_{\beta_k}.$$

In other words, $\mathcal{G}$ admits a finite subcover of K.

Thus K is compact.

□

2.38. Corollary. *Let $n \ge 1$ and $K \subseteq \mathbb{R}^n$. Let $\mathbb{R}^n$ be equipped with the standard (Euclidean) topology. The following are equivalent.*

(a) *K is compact.*
(b) *K is sequentially compact.*
(c) *K is closed and bounded.*

Proof. That (a) and (b) are equivalent is Theorem 2.37.

That (b) and (c) are equivalent is the Heine-Borel Theorem from Math 247.

□

3. Appendix

In this appendix we provide a proof of the last statement of Theorem 3.1.9, which we restate for the convenience of the reader.

Theorem. *Let (X,τ) be a topological space and $x \in X$.*

(a) *If $U \in \mathcal{U}_x$, then $x \in U$.*
(b) *If $U, V \in \mathcal{U}_x$, then $U \cap V \in \mathcal{U}_x$.*
(c) *If $U \in \mathcal{U}_x$, then there exists $V \in \mathcal{U}_x$ such that $U \in \mathcal{U}_y$ for each $y \in V$.*
(d) *If $U \in \mathcal{U}_x$ and $U \subseteq V$, then $V \in \mathcal{U}_x$.*
(e) *The set $G \in \tau$ if and only if G contains a neighbourhood of each of its points.*

Conversely: suppose that Y is a non-empty set and for each $x \in Y$ we are given a non-empty collection $\mathcal{U}_x \subseteq \mathcal{P}(Y)$ satisfying conditions (a) through (d). Suppose furthermore that we ***declare*** *a set $G \in Y$ to be* ***open*** *if for each $x \in G$ there exists $U \in \mathcal{U}_x$ so that $x \in U \subseteq G$. If we then set $\rho = \{G \subseteq Y : G \text{ is open}\}$, then ρ is a topology on Y in which the neighbourhood system at x is exactly $\mathcal{U}_x$.*

Our goal is therefore to show that ρ is a topology on Y in which the neighbourhood system at x is exactly $\mathcal{U}_x$.

Let

$$\rho = \{G \subseteq Y : \text{ for all } g \in G \text{ there exists } U \in \mathcal{U}_g \text{ such that } g \in U \subseteq G\}.$$

First we show that ρ is a topology on Y.

- That $\varnothing \in \rho$ is vacuously true. Also, given that for each $x \in Y$, the family $\mathcal{U}_x$ is non-empty, and $x \in U \subseteq Y$ for all $U \in \mathcal{U}_x$, we see that $Y \in \rho$ as well.
- Suppose that $\{G_\lambda\}_{\lambda\in\Lambda} \subseteq \rho$, and let $G = \cup_{\lambda\in\Lambda} G_\lambda$. Let $x \in G$. Then there exists $\alpha \in \Lambda$ so that $x \in G_\alpha$. But $G_\alpha \in \rho$, so there exists $U \in \mathcal{U}_x$ so that $x \in U \subseteq G_\alpha \subseteq G$.

  That is, for all $x \in G$, there exists $U \in \mathcal{U}_x$ so that $x \in U \subseteq G$, so $G \in \rho$ by definition of ρ.
- Suppose that $G_1, G_2 \in \rho$ and that $G = G_1 \cap G_2$. If $G = \varnothing$, then $G \in \rho$ from above. Otherwise, let $x \in G$. Then $x \in G_1$, so there exists $U_1 \in \mathcal{U}_x$ so that $x \in U_1 \subseteq G_1$. Similarly, $x \in G_2$, so there exists $U_2 \in \mathcal{U}_x$ so that $x \in U_2 \subseteq G_2$. But then $U := U_1 \cap U_2 \in \mathcal{U}_x$ by our hypotheses, and so $x \in U \subseteq G_1 \cap G_2 = G$.

  By definition of ρ, $G \in \rho$.

This shows that ρ is a topology on Y. For $x \in Y$, denote by $\mathcal{V}_x$ the nhood system of x in (Y,ρ).

- Suppose that $V \in \mathcal{V}_x$. Then $x \in V$ and so there exists $G \in \rho$ so that

$$x \in G \subseteq V.$$

  But then $x \in G \in \rho$ implies that there exists $U \in \mathcal{U}_x$ so that $x \in U \subseteq G$. Hence $G \in \mathcal{U}_x$, since $U \in \mathcal{U}_x$ and $U \subseteq G$. But then $G \in \mathcal{U}_x$ and $G \subseteq V$ implies that $V \in \mathcal{U}_x$.

  That is, $\mathcal{V}_x \subseteq \mathcal{U}_x$.
- Conversely, suppose that $U \in \mathcal{U}_x$. We wish to show that $U \in \mathcal{V}_x$.

  Let $W = \{w \in U : U \in \mathcal{U}_w\}$. Note that $x \in W$, so in particular, $W \neq \varnothing$. We shall prove that $W \in \rho$. If we can do this, then $W \in \rho$ and $x \in W$ implies that $W \in \mathcal{V}_x$. But then $W \subseteq U$ implies that $U \in \mathcal{V}_x$, completing the proof.

  Indeed, suppose that $w \in W$. Then $U \in \mathcal{U}_w$ and so by condition (c) there exists $Z \in \mathcal{U}_w$ so that $y \in Z$ implies that $U \in \mathcal{U}_y$. We claim that $Z \subseteq W$.

  To see this, let $y \in Z$. Then $U \in \mathcal{U}_y$, so $y \in U$ and $U \in \mathcal{U}_y$. By definition of W, $y \in W$, which proves the claim.

  Hence, $W \subseteq U \subseteq Y$, and for all $w \in W$, there exists $Z \in \mathcal{U}_w$ so that $w \in Z \subseteq W$. By definition of ρ, this means that $W \in \rho$. As we have seen, this is sufficient to prove the result.

CHAPTER 4

Completeness

1. Completeness and normed linear spaces

Politicians and diapers have one thing in common. They should both be changed regularly, and for the same reason.

ascribed to multiple people, including **Mark Twain** and **José Maria de Eça de Quieroz**

1.1. From the point of view of an analyst, the real numbers are far better behaved than the rational numbers[1]. For example, the Intermediate Value Theorem asserts that if $f : [a, b] \to \mathbb{R}$ is continuous and $f(a) \le \kappa \le f(b)$, then there exists $d \in [a, b]$ so that $f(d) = \kappa$. Such a result fails if we consider a continuous function $f : (\mathbb{Q} \cap [a, b]) \to \mathbb{R}$. For example, let $f : \mathbb{Q} \to \mathbb{R}$ be the function defined by $f(q) = q^2$. Observe that $f(0) = 0$ and $f(2) = 4$, and f is continuous on $\mathbb{Q}$ (with respect to the standard topology on $\mathbb{Q}$ which it inherits as a subspace of $\mathbb{R}$). Neverthless, there does not exist $r \in \mathbb{Q}$ so that $f(r) = 2$. Given a non-empty bounded subset of $\mathbb{Q}$, there is no least upper bound for that set which lies in $\mathbb{Q}$.

These are the kinds of properties which makes analysis possible, and it is to this property that we now turn our attention.

1.2. Definition. *A subset H of a metric space (X, d) is said to be* ***complete*** *if every Cauchy sequence in H converges to some element of H.*

1.3. Examples.

(a) Let $n \ge 1$ be an integer, and consider $(\mathbb{K}^n, d)$ where d denotes the standard (Euclidean metric)

$$d((x_1, x_2, ..., x_n), (y_1, y_2, ..., y_n)) = \sqrt{\sum_{k=1}^{n} |x_k - y_k|^2}.$$

Then $(\mathbb{K}^n, d)$ is complete, as seen in Math 147/148.

[1]This might be a good time to concede that from the point of view of algebraists, the rational numbers are far better behaved than analysts.

(b) Let $\varnothing \neq X$ be a non-empty set, equipped with the discrete metric μ. Then (X, μ) is complete. Indeed, suppose that $(x_n)_n$ is a Cauchy sequence and let $\varepsilon = 1$. Choose $N > 0$ so that $m, n \geq N$ implies that $\mu(x_n, x_m) < \varepsilon = 1$. Then $n, m \geq N$ implies that $x_n = x_m$, or equivalently, $n \geq N$ implies that $x_n = x_N$. In particular, if $\delta > 0$, then $n \geq N$ implies that $\mu(x_n, x_N) = 0 < \delta$, proving that $\lim_n x_n = X_N$.

(c) The open interval $(0, 1) \subseteq \mathbb{R}$ when equipped with the standard metric d is not complete. It is easily seen that $(\frac{1}{n})_n$ is a Cauchy sequence which does not converge to any element of $(0, 1)$.

1.4. Definition. *Two topological spaces* (X, τ_X) *and* (Y, τ_Y) *are said to be* ***homeomorphic*** *if there exists a bijection* $f : X \to Y$ *such that both* f *and* f^{-1} *are continuous. When this is the case, we write* $X \simeq Y$.

If X and Y are topological spaces as above and $X \simeq Y$, then $G \in \tau_X$ if and only if $f(G) \in \tau_Y$. Any topological property of X is shared by Y and vice-versa. For example, X is Hausdorff if and only if Y is. Homeomorphism is the notion of isomorphism in the category of topological spaces.

1.5. Remark. Consider the map

$$\begin{array}{rccc} f: & \mathbb{R} & \to & (-1,1) \\ & x & \mapsto & \frac{x}{1+|x|}. \end{array}$$

It is readily verified that f is a continuous bijection whose inverse is also continuous. Thus $\mathbb{R}$ is homeomorphic to $(-1, 1)$. It is worth pointing out, however, that $\mathbb{R}$ is complete, while $(-1, 1)$ is not.

This shows that completeness is not a topological property of a space.

1.6. Proposition. *Let* (X, d) *be a metric space and* $H \subseteq X$ *be complete. Then* H *is closed.*

Proof. Suppose that $x \in \overline{H}$. Then x is a limit point of H, so there exists a sequence $(x_n)_n$ in H which converges to x. But then that sequence is Cauchy, and since H is complete, $x \in H$.

□

One of the more interesting contexts in which to study completeness is that of normed linear spaces.

1.7. Definition. *A series $\sum_n x_n$ in a normed linear space $(\mathfrak{X}, \|\cdot\|)$ is said to be* ***summable*** *if there exists $x \in \mathfrak{X}$ such that*

$$\lim_N \|x - \sum_{n=1}^{N} x_n\| = 0.$$

We then write $x = \sum_n x_n$.

We say that $\sum_n x_n$ is ***absolutely summable*** *if*

$$\sum_n \|x_n\| < \infty.$$

The following result provides a very practical tool when trying to decide whether or not a given normed linear space is complete. We remark that the second half of the proof uses the standard fact that if $(y_n)_n$ is a Cauchy sequence in a metric space (Y, d), and if $(y_n)_n$ admits a convergent subsequence with limit y_0, then the original sequence $(y_n)_n$ converges to y_0 as well. The proof of this is identical to the proof of the corresponding result in $\mathbb{R}$, and is left as an (important!) exercise for the reader.

1.8. Theorem. *Let $(\mathfrak{X}, \|\cdot\|)$ be a normed linear space. The following statements are equivalent:*

(a) *$\mathfrak{X}$ is complete, and hence $\mathfrak{X}$ is a Banach space.*
(b) *Every absolutely summable series in $\mathfrak{X}$ is summable.*

Proof.

(a) implies (b): Suppose that $\mathfrak{X}$ is complete, and that $\sum x_n$ is absolutely summable. For each $k \ge 1$, let $y_k = \sum_{n=1}^{k} x_n$. Given $\varepsilon > 0$, we can find $N > 0$ so that $m \ge N$ implies $\sum_{n=m}^{\infty} \|x_n\| < \varepsilon$. If $k \ge m \ge N$, then

$$\begin{aligned}\|y_k - y_m\| &= \|\sum_{n=m+1}^{k} x_n\| \\ &\le \sum_{n=m+1}^{k} \|x_n\| \\ &\le \sum_{n=m+1}^{\infty} \|x_n\| \\ &< \varepsilon,\end{aligned}$$

so that $(y_k)_k$ is Cauchy in $\mathfrak{X}$. Since $\mathfrak{X}$ is complete, $y = \lim_{k\to\infty} y_k = \lim_{k\to\infty} \sum_{n=1}^{k} x_n = \sum_{n=1}^{\infty} x_n$ exists, i.e. $\sum_{n=1}^{\infty} x_n$ is summable.

(b) implies (a): Next suppose that every absolutely summable series in $\mathfrak{X}$ is summable, and let $(y_j)_j$ be a Cauchy sequence in $\mathfrak{X}$. For each $n \ge 1$ there exists $N_n > 0$ so that $k, m \ge N_n$ implies $\|y_k - y_m\| < 1/2^{n+1}$. Let $x_1 = y_{N_1}$

and for $n \ge 2$, let $x_n = y_{N_n} - y_{N_{n-1}}$. Then $\|x_n\| < 1/2^n$ for all $n \ge 2$, so that

$$\begin{aligned}\sum_{n=1}^{\infty} \|x_n\| &\le \|x_1\| + \sum_{n=2}^{\infty} \frac{1}{2^n} \\ &\le \|x_1\| + \frac{1}{2} < \infty.\end{aligned}$$

By hypothesis, $y = \sum_{n=1}^{\infty} x_n = \lim_{k\to\infty} \sum_{n=1}^{k} x_n$ exists. But $\sum_{n=1}^{k} x_n = y_{N_k}$, so that $\lim_{k\to\infty} y_{N_k} = y \in \mathfrak{X}$. Recalling that $(y_j)_j$ was Cauchy, we conclude from the remark preceding the Theorem that $(y_j)_j$ also converges to y. Since every Cauchy sequence in $\mathfrak{X}$ converges, $\mathfrak{X}$ is complete.

□

1.9. Definition. *Let $(\mathfrak{X}, \|\cdot\|)$ be a normed linear space and d be the metric on $\mathfrak{X}$ induced by the norm. If $(\mathfrak{X}, d)$ is complete, we say that $(\mathfrak{X}, \|\cdot\|)$ is a* ***Banach space****.*

Thus a Banach space is a complete normed linear space.

1.10. Theorem. *The normed linear space $(\ell_1, \|\cdot\|_1)$ is a Banach space.*

Proof.

By Theorem 1.8, it suffices to prove that every absolutely summable series in ℓ_1 is summable.

Suppose that $\sum_n x_n$ is absolutely summable with $M := \sum_n \|x_n\|_1 < \infty$. Writing $x_n = (x_{n,k})_{k=1}^{\infty}$ for each $n \ge 1$, we easily see that each $|x_{n,k}| \le \|x_n\|_1$, and thus for each $k \ge 1$,

$$\sum_{n=1}^{\infty} |x_{n,k}| \le M < \infty.$$

Since $(\mathbb{K}, |\cdot|)$ is complete (from first-year Calculus), the series $\sum_{n=1}^{\infty} x_{n,k}$ is summable for each $k \ge 1$. Define

$$z_k = \sum_{n=1}^{\infty} x_{n,k},$$

and set $z = (z_k)_{k=1}^{\infty}$. We must show that $z \in \ell_1$, and that $z = \sum_n x_n$.

- $z \in \ell_1$:
  Consider

$$\begin{aligned}\sum_{k=1}^{\infty} |z_k| &\le \sum_{k=1}^{\infty} \left(\sum_{n=1}^{\infty} |x_{n,k}| \right) \\ &= \sum_n \sum_k |x_{n,k}| \\ &= \sum_n \|x_n\|_1 \\ &= M < \infty.\end{aligned}$$

(Note that interchanging the order of summation in the second equation is justified by the fact that all of the terms are non-negative.)

- $z = \sum_n x_n$:

  Let $\varepsilon > 0$ and choose $N > 0$ so that $m \ge N$ implies that $\sum_{m=N}^{\infty} \|x_n\|_1 < \varepsilon$. This is possible since the series is absolutely summable. If $m \ge N$, then

$$\begin{aligned}\|z - \sum_{n=1}^{m} x_n\|_1 &= \sum_{k=1}^{\infty} |z_k - \sum_{n=1}^{m} x_{n,k}| \\ &= \sum_{k=1}^{\infty} |\sum_{n=m+1}^{\infty} x_{n,k}| \\ &\le \sum_{k=1}^{\infty} \sum_{n=m+1}^{\infty} |x_{n,k}| \\ &= \sum_{n=m+1}^{\infty} \sum_{k=1}^{\infty} |x_{n,k}| \\ &= \sum_{n=m+1}^{\infty} \|x_n\|_1 \\ &< \varepsilon.\end{aligned}$$

  Thus $z = \lim_{m\to\infty} \sum_{n=1}^{m} x_n = \sum_n x_n$.

It follows that $(\ell_1, \|\cdot\|_1)$ is complete, i.e. a Banach space.

□

1.11. Theorem. *The normed linear space $(\ell_\infty, \|\cdot\|_\infty)$ is a Banach space.*

Proof.

This is an Assignment problem.

□

1.12. Definition. *Let $(\mathcal{H}, \langle\cdot,\cdot\rangle)$ be an inner product space, and let $\|\cdot\|$ denote the norm induced by the inner product, i.e. $\|x\| = \sqrt{\langle x, x\rangle}$ for all $x \in \mathcal{H}$.*

If $(\mathcal{H}, \|\cdot\|)$ is complete, we say that $\mathcal{H}$ is a ***Hilbert space****. Thus a Hilbert space is a complete inner product space.*

It follows that every Hilbert space is a Banach space. The converse is false, though it is not trivial to prove. Given a Banach space $(\mathfrak{X}, \|\cdot\|)$, one must show that there does not exists *any* inner product on $\mathfrak{X}$ which induces the given norm $\|\cdot\|$.

1.13. Theorem. *The inner product space $(\ell_2, \langle\cdot,\cdot\rangle)$ is a Hilbert space.*

Proof.

As we saw in the Assignments, every inner product space is a normed linear space.

Again, we suppose that $\sum_n x_n$ is an absolutely summable series in ℓ_2, and set

$$M := \sum_n \|x_n\|_2 < \infty.$$

As was the case for ℓ_1, writing $x_n = (x_{n,k})_{k=1}^\infty$, it is easy to see that for each $k \geq 1$, $|x_{n,k}| \leq \|x_n\|_2$, and thus

$$\sum_n |x_{n,k}| \leq \sum_n \|x_n\|_2 = M < \infty.$$

As before, the completeness of $(\mathbb{K}, |\cdot|)$ implies that $z_k = \sum_{n=1}^\infty x_{n,k}$ exists for each $k \geq 1$. let $z = (z_k)_{k=1}^\infty$.

We must show that $z \in \ell_2$ and that $z = \sum_{n=1}^\infty x_n$.

The idea behind the following proof is to estimate the quantity which defines $\|z - \sum_{n=1}^p x_n\|_2$, *despite the fact that we do not yet know that* $z \in \ell_2$! In fact, by setting $p = 0$, we shall obtain *a proof* of the fact that $z \in \ell_2$.

Now for each $p \geq 1$, $z_k - \sum_{n=1}^p x_{n,k} = \sum_{n=p+1}^\infty x_{n,k}$, and thus

$$\begin{aligned}
\sum_{k=1}^\infty \left| \sum_{n=p+1}^\infty x_{n,k} \right|^2 &\leq \sum_{k=1}^\infty \left(\sum_{n=p+1}^\infty |x_{n,k}| \right)^2 \\
&= \sum_{k=1}^\infty \quad \sum_{n=p+1}^\infty \sum_{m=p+1}^\infty |x_{n,k}|\,|x_{m,k}| \\
&= \sum_{n=p+1}^\infty \sum_{m=p+1}^\infty \left(\sum_{k=1}^\infty |x_{n,k}|\,|x_{m,k}| \right) \\
&= \sum_{n=p+1}^\infty \sum_{m=p+1}^\infty \langle y_n, y_m \rangle,
\end{aligned}$$

where for each $n \geq 1$, $y_n = (|x_{n,1}|, |x_{n,2}|, |x_{n,3}|, \ldots) \in \ell_2$ with $\|y_n\|_2 = \|x_n\|_2$. (Again, note that the fact that we can interchange the order of summation as of the third line of this equation is due to the fact that all of the terms are non-negative.)

Thus, by the Cauchy-Schwarz Inequality,

$$\begin{aligned}
\sum_{k=1}^\infty \left(\left| \sum_{n=p+1}^\infty x_{n,k} \right|^2 \right) &\leq \sum_{n=p+1}^\infty \sum_{m=p+1}^\infty \langle y_n, y_m \rangle \\
&\leq \sum_{n=p+1}^\infty \sum_{m=p+1}^\infty \|y_n\|_2 \, \|y_m\|_2 \\
&= \left(\sum_{n=p+1}^\infty \|y_n\|_2 \right)^2 \\
&= \left(\sum_{n=p+1}^\infty \|x_n\|_2 \right)^2.
\end{aligned}$$

If $p = 0$, then this shows that $\sum_{k=1}^\infty |z_k|^2 < \infty$, and thus $z \in \ell_2$.

More generally, however, if we choose $\varepsilon > 0$ and $N > 0$ so that $p \ge N$ implies that $\sum_{n=p+1}^{\infty} \|x_n\|_2 < \varepsilon$, then $p \ge N$ implies that

$$\|z - \sum_{n=1}^{p} x_n\|_2 \le \sum_{n=p+1}^{\infty} \|x_n\|_2 < \varepsilon.$$

Thus $z = \sum_n x_n$.

By Theorem 1.8 above, $(\ell_2, \|\cdot\|_2)$ is a Banach space, i.e. $(\ell_2, \langle\cdot,\cdot\rangle)$ is a Hilbert space.

□

1.14. Theorem. *The normed linear space* $(\mathcal{C}([0,1],\mathbb{K}), \|\cdot\|_\infty)$ *is a Banach space.*

Proof. We have already seen that $\mathcal{C}([0,1],\mathbb{K})$ is a normed linear space when equipped with the norm

$$\|f\|_\infty = \sup\{|f(x)| : x \in [0,1]\}.$$

There remains to show that it is complete.

This follows immediately, however, from the Weierstraß M-test from Math 148, combined with Theorem 1.8. That is, if we let $\sum_n f_n$ be an absolutely summable series of continuous functions on $[0,1]$, then with $M_n := \|f_n\|_\infty$, we see that for each $n \ge 1$ we have $\sup_{x\in[0,1]} |f_n(x)| \le M_n$ (by definition of M_n) and that $\sum_n M_n < \infty$ (by definition of absolute summability).

By the Weierstraß M-test, $\sum_n f_n$ converges uniformly and absolutely to a continuous function $f \in \mathcal{C}([0,1],\mathbb{K})$. But uniform convergence is precisely convergence in the $\|\cdot\|_\infty$ norm. By Theorem 1.8, $(\mathcal{C}([0,1],\mathbb{K}), \|\cdot\|_\infty)$ is complete.

□

For those of you who may not have taken Math 148, an alternative proof of this is given in the Appendix to this Chapter.

The following is an analogue of the Nested Intervals Theorem for metric spaces.

1.15. Theorem. *Let* (X, d) *be a metric space. The following are equivalent.*

(a) (X,d) *is complete.*

(b) *If*

$$F_1 \supseteq F_2 \supseteq F_3 \supseteq \cdots$$

is a nested sequence of closed, non-empty subsets of X *and suppose that* $\lim_n \operatorname{diam} F_n = 0$. *Then*

$$\cap_{n=1}^{\infty} F_n \ne \varnothing.$$

Proof.

(a) implies (b). Suppose that (X,d) is complete.

Let $(F_n)_n$ be a nest sequence of closed, non-empty subsets of X as above, with

$$\lim_n \operatorname{diam} F_n = 0.$$

Fix $\varepsilon > 0$ and choose $N > 0$ so that $n \ge N$ implies that $\operatorname{diam} F_n < \varepsilon$.

Choose $x_n \in F_n$ arbitrarily. If $m > n \geq N$, then $x_m \in F_m \subseteq F_n$ and $x_n \in F_n$, so

$$d(x_m, x_n) \leq \operatorname{diam} F_n < \varepsilon.$$

It follows that $(x_n)_n$ is a Cauchy sequence in X. Since X is assumed to be complete, there exists $x \in X$ so that $\lim_n x_n = x$.

Note that if $p \geq 1$, then $n \geq p$ implies that $x_n \in F_n \subseteq F_p$. From this it follows that $x = \lim_n x_n \in F_p$. But $p \geq 1$ was arbitrary, so $x \in \cap_p F_p \neq \varnothing$.

(b) implies (a). Next, suppose that condition (b) holds, and let $(x_n)_n$ be a Cauchy sequence in X. For each $m \geq 1$, set

$$F_m = \overline{\{x_n\}_{n=m}^{\infty}}.$$

Then each F_m is closed, non-empty, and

$$F_1 \supseteq F_2 \supseteq F_3 \supseteq \cdots.$$

Furthermore, since $(x_n)_n$ is Cauchy, given $\varepsilon > 0$, we can find $N \geq 1$ so that $m, n \geq N$ implies that $d(x_n, x_m) < \varepsilon$. It follows that $\operatorname{diam} F_N \leq \varepsilon$. But $n \geq N$ implies that $F_n \subseteq F_N$, so that $\operatorname{diam} F_n \leq \varepsilon$ for all $n \geq N$. This shows that $\lim_n \operatorname{diam} F_n = 0$.

Our hypothesis in (b) implies that $\cap_n F_n \neq \varnothing$. Choose $x \in \cap_n F_n$. We claim that $\lim_n x_n = x$. Indeed, if $\varepsilon > 0$ and N are as above, then noting that for $n \geq N$ we have $x_n, x \in F_n$, we conclude that

$$d(x_n, x) \leq \operatorname{diam} F_n \leq \varepsilon.$$

Hence $\lim_n x_n = x$.

□

Note that both conditions of the previous Theorem are required.

1.16. Examples.

(a) Let $X = \mathbb{R}$, equipped with the standard metric and $F_n = [n, \infty)$, $n \geq 1$. Each F_n is closed, and $F_1 \supseteq F_2 \supseteq F_3 \supseteq \cdots$. Nevertheless, $\cap_n F_n = \varnothing$.

The issue is that $\lim_n \operatorname{diam} F_n = \infty \neq 0$.

(b) Let $X = \mathbb{R}$ again and this time set $F_n = (0, \frac{1}{n}]$. Now $\mathbb{R}$ is complete, $F_n \supseteq F_{n+1}$ for all $n \geq 1$ and $\lim_n \operatorname{diam} F_n = 0$, but $\cap_n F_n = \varnothing$.

In this case, the issue is that the F_n's are not closed.

2. Completions of metric spaces

2.1. Although $\mathbb{Q}$ is not complete (with respect to the standard metric), nevertheless, $\mathbb{Q}$ "sits inside" the complete metric space $\mathbb{R}$. In this section, we make precise what we mean by this, and show that every metric space embeds isometrically in a complete metric space. This is extremely useful when studying normed linear spaces.

2.2. Definition. *A metric space (X^*, d^*) is called a **completion** of a metric space (X, d) if*

- *(X^*, d^*) is complete, and*
- *(X, d) is isometrically isomorphic to a **dense** subset of (X^*, d^*); that is, there exists a map $\rho : X \to X^*$ satisfying*
  - (i) *$d^*(\rho(x), \rho(y)) = d(x, y)$ for all $x, y \in X$, and*
  - (ii) *$X^* = \overline{\rho(X)}$.*

2.3. Example. The motivating example is the one we have already mentioned: $(\mathbb{R}, d)$ is completion of $(\mathbb{Q}, d|_{\mathbb{Q}})$, where $d(x, y) = |x - y|$ is the standard metric on $\mathbb{R}$ and $d|_{\mathbb{Q}}$ is the restriction of d to $\mathbb{Q} \times \mathbb{Q}$. The map $\rho : \mathbb{Q} \to \mathbb{R}$ defined by $\rho(q) = q$ for all $q \in \mathbb{Q}$ satisfies the conditions of Definition 2.2 above.

2.4. Our next goal is to show that every metric space has a completion, and that this completion is in some sense unique. To that end:

let (X, d) be a metric space, and denote by $\Gamma[X]$ the collection of all Cauchy sequences in X. We define a relation $\sim$ on $\Gamma[X]$ by setting

$$(x_n)_n \sim (y_n)_n \qquad \text{if and only if} \qquad \lim_n d(x_n, y_n) = 0.$$

Intuitively, under this relation $\sim$, we identify Cauchy sequences which would have the same limit if that limit existed in X.

2.5. Lemma. *With the notation of Section 2.4, the relation $\sim$ is an equivalence relation on $\Gamma[X]$.*

Proof.

(a) $\sim$ is reflexive.

Clearly $(x_n)_n \sim (x_n)_n$ since $\lim_n d(x_n, x_n) = \lim_n 0 = 0$.

(b) $\sim$ is symmetric.

Since $d(x, y) = d(y, x)$ for all $x, y \in X$, it follows that given sequences $(x_n)_n$ and $(y_n)_n \in \Gamma[X]$,

$$\lim_n d(x_n, y_n) = \lim_n d(y_n, x_n),$$

and so $(x_n)_n \sim (y_n)_n$ if and only if $(y_n)_n \sim (x_n)_n$.

(c) $\sim$ is transitive.

Suppose that $(x_n)_n \sim (y_n)_n$ and that $(y_n)_n \sim (z_n)_n$. For each $n \geq 1$,

$$d(x_n, z_n) \leq d(x_n, y_n) + d(y_n, z_n),$$

and so

$$0 \leq \lim_n d(x_n, z_n) \leq \lim_n d(x_n, y_n) + \lim_n d(y_n, z_n) = 0 + 0 = 0.$$

Thus $(x_n)_n \sim (z_n)_n$, and so $\sim$ is transitive.

This three properties indicate that $\sim$ is an equivalence relation.

□

2.6. Recall that an equivalence relation $\sim$ on a set A partitions A into equivalence classes.

Let (X,d) be a metric space. We define the set $X^* = \Gamma[X]/\sim$, so that X^* consist of the equivalence classes of Cauchy sequences in $\Gamma[X]$ as determined by the equivalence relation $\sim$ of Section 2.4.

That is,

$$X^* = \{[(x_n)_n] : (x_n)_n \in \Gamma[X]\},$$

where $[(x_n)_n]$ denotes the equivalence class of $(x_n)_n$ under $\sim$.

We define the map:

$$\begin{array}{rccc} d^* : & X^* \times X^* & \to & \mathbb{R} \\ & ([(x_n)_n],[(y_n)_n]) & \mapsto & \lim_n d(x_n,y_n). \end{array}$$

Our first result shows that d^* is well-defined.

2.7. Proposition. *Let (X,d) be a metric space and define X^*, d^* as in Section 2.6. Then d^* is well-defined. That is, if $(x_n)_n \sim (w_n)_n$ and $(y_n)_n \sim (z_n)_n$, then $\lim_n d(x_n,y_n)$ exists and*

$$\lim_n d(x_n,y_n) = \lim_n d(w_n,z_n).$$

Proof. First we check that $\lim_n d(x_n,y_n)$ exists.

Let $\varepsilon > 0$, and choose $N > 0$ so that $m,n \geq N$ implies that $d(x_n,x_m) < \frac{\varepsilon}{2}$ and $d(y_n,y_m) < \frac{\varepsilon}{2}$. If $m,n \geq N$, then

$$\begin{aligned} |d(x_n,y_n) - d(x_m,y_m)| &\leq |d(x_n,y_n) - d(x_n,y_m)| + |d(x_n,y_m) - d(x_m,y_m)| \\ &\leq d(y_n,y_m) + d(x_n,x_m) \\ &< \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon. \end{aligned}$$

Thus $(d(x_n,y_n))_n$ is a Cauchy sequence in $\mathbb{R}$. But $\mathbb{R}$ is complete (with the standard metric), so $(d(x_n,y_n))_n$ converges to some point in $\mathbb{R}$.

Next,

$$\begin{aligned} \lim_n d(x_n,y_n) &\leq \lim_n d(x_n,w_n) + d(w_n,z_n) + d(z_n,y_n) \\ &= 0 + \lim_n d(w_n,z_n) + 0 \\ &= \lim_n d(w_n,z_n), \end{aligned}$$

and by symmetry, $\lim_n d(w_n,z_n) \leq \lim_n d(x_n,y_n)$. Thus we conclude that

$$\lim_n d(x_n,y_n) = \lim_n d(w_n,z_n).$$

□

2.8. Proposition. *The function d^* defined in Section 2.6 is a metric on X^*.*
Proof. Suppose that $x = [(x_n)_n]$, $y = [(y_n)_n]$ and $z = [(z_n)_n] \in X^*$. Then

(a) $d^*(x,y) = \lim_n d(x_n, y_n) \geq 0$, since each $d(x_n, y_n) \geq 0$, by virtue of the fact that d is a metric on X. Furthermore, $d^*(x,y) = 0$ if and only if $\lim_n d(x_n, y_n) = 0$, which by definition happens if and only if $(x_n)_n \sim (y_n)_n$, ie. if and only if $x = y$.

(b)
$$d^*(x,y) = \lim_n d(x_n, y_n) = \lim_n d(y_n, x_n) = d^*(y,x),$$
where the second equality again holds because d is a metric on X.

(c)
$$\begin{aligned} d^*(x,z) &= \lim_n d(x_n, z_n) \\ &\leq \lim_n d(x_n, y_n) + d(y_n, z_n) \\ &= \lim_n d(x_n, y_n) + \lim_n d(y_n, z_n) \\ &= d^*(x,y) + d^*(y,z), \end{aligned}$$
where the first inequality holds because d is a metric on X, and the second equality holds because each of the limits exists, as was shown in the previous Proposition.

From these three verifications it follows that d^* is a metric on X^*.

□

Our next goal is to prove that (X^*, d^*) is a completion of (X, d).

2.9. Proposition. *Let (X,d) be a metric space and let (X^*, d^*) be the metric space of equivalence classes of Cauchy sequences in X modulo the relation $\sim$ defined in Section 2.6. The map*
$$\begin{array}{rccc} \rho: & X & \to & X^* \\ & x & \mapsto & [(x,x,x,...)] \end{array}$$
is an isometric embedding of X into X^*, and $\rho(X)$ is dense in X^*.
Proof. Suppose that $x, y \in X$. Then
$$\begin{aligned} d^*(\rho(x), \rho(y)) &= d^*([(x,x,x,...)], [(y,y,y,...)]) \\ &= \lim_n d(x,y) \\ &= d(x,y). \end{aligned}$$
From this it follows that ρ is isometric, and hence ρ is injective. Indeed, if $x \neq y$, then $d^*(\rho(x), \rho(y)) = d(x,y) > 0$, so $\rho(x) \neq \rho(y)$.

Next, let $z = [(z_n)_n] \in X^*$. Let $\varepsilon > 0$. Since $(z_n)_n \in \Gamma[X]$, it is a Cauchy sequence and so we can find $N \geq 1$ so that $m, n \geq N$ implies that $d(z_n, z_m) < \varepsilon$. In particular, $d(z_n, z_N) < \varepsilon$ for all $n \geq N$.

Set $y = [(x_N, x_N, x_N, ...)]$, i.e. set $y_n = x_N$ for all $n \geq 1$. Then

$$\begin{aligned} d^*(z,y) &= \lim_n d(z_n, y_n) \\ &= \lim_n d(z_n, x_N) \\ &\leq \varepsilon. \end{aligned}$$

Since $y = \rho(x_N) \in \rho(X)$ and since $\varepsilon > 0$ is arbitrary, it follows that $\rho(X)$ is dense in (X^*, d^*).

□

2.10. Theorem. *The metric space (X^*, d^*) constructed above is complete, and thus it is a completion of (X, d). In particular, every metric space (X, d) admits a completion.*

Proof. Let $(\beta_n)_n$ be a Cauchy sequence in X^*, that is; $\beta_n = [(x_{n,k})_{k\geq 1}]$ for each $n \geq 1$. Let $\rho : X \to X^*$ denote the isometric embedding described above, so that $\rho(X)$ is dense in X^*. Then we can find a sequence $(p_n)_n$ in X so that $d^*(\rho(p_n), \beta_n) < \frac{1}{n}$, $n \geq 1$.

Note that the fact that $(\beta_n)_n$ is Cauchy in X^* implies that $(p_n)_n$ is Cauchy in X. Indeed, if $\varepsilon > 0$ there exists $N > \frac{3}{\varepsilon}$ such that $m, n \geq N$ implies that $d^*(\beta_m, \beta_n) < \frac{\varepsilon}{3}$. But then $m, n \geq N$ implies that

$$\begin{aligned} d(p_m, p_n) &= d^*(\rho(p_m), \rho(p_n)) \\ &\leq d^*(\rho(p_m), \beta_m) + d^*(\beta_m, \beta_n) + d^*(\beta_n, \rho(p_n)) \\ &< \frac{1}{N} + \frac{\varepsilon}{3} + \frac{1}{N} \\ &< \varepsilon. \end{aligned}$$

Let $\beta := [(p_k)_k] \in X^*$.

If $n \geq N$, then

$$\begin{aligned} d^*(\beta_n, \beta) &\leq d^*(\beta_n, \rho(p_n)) + d^*(\rho(p_n), \beta) \\ &\leq \frac{1}{n} + \lim_{k\to\infty} d(p_n, p_k) \\ &< \frac{1}{N} + \varepsilon \\ &< 2\varepsilon. \end{aligned}$$

Thus $\lim_n \beta_n = \beta$, and so $(\beta_n)_n$ converges to β in (X^*, d^*), showing that (X^*, d^*) is complete.

□

2.11. Theorem. *Let (X, d) be a metric space, and denote by (X^*, d^*) the metric space constructed above.*

If (Y, d_Y) is any completion of X, then (Y, d_Y) is isometrically isomorphic to (X^*, d^*); that is, there exists a bijective map $\gamma : Y \to X^*$ so that $d^*(\gamma(y), \gamma(z)) = d_Y(y, z)$ for all $y, z \in Y$.

It follows that every metric space admits a completion, and that this completion is unique up to isometric isomorphism.

Proof.

That every metric space (X, d) admits a completion is Theorem 2.10.

Suppose that (Y, d_Y) is a completion of X. Let $\beta : X \to Y$ be the isometric map for which $\beta(X)$ is dense in Y. By identifying (X, d) with $(\beta(X), d_Y|_{\beta(X)})$, we may assume *a priori* that $X \subseteq Y$, and that $d = d_Y|_X$.

Let $y \in Y$. Since X is dense in Y, we can find a sequence $(x_n)_n \in X^{\mathbb{N}}$ so that

$$\lim_n d_Y(x_n, y) = 0.$$

Let $\varepsilon > 0$, and choose $N \geq 1$ so that $n \geq N$ implies that $d_Y(x_n, y) < \frac{\varepsilon}{2}$. Then for $m, n \geq N$,

$$d_Y(x_n, x_m) \leq d_Y(x_n, y) + d_Y(y, x_m) < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon.$$

This shows that $(x_n)_n$ is a Cauchy sequence in X, i.e. $(x_n)_n \in \Gamma[X]$.

Consider the map

$$\begin{array}{rccc} \gamma : & Y & \to & X^* \\ & y & \mapsto & [(x_n)_n]. \end{array}$$

First we show that γ is well-defined. That is, if $(w_n)_n \in X^{\mathbb{N}}$ and $\lim_n w_n = y$, then $[(x_n)_n] = [(w_n)_n]$. Indeed, letting $\varepsilon > 0$ as before, we choose $M > 0$ so that $m \geq M$ implies that $d_Y(x_m, y) < \frac{\varepsilon}{2}$ and $d_Y(w_m, y) < \frac{\varepsilon}{2}$.

Then $m \geq M$ implies that

$$d(x_m, w_m) = d_Y(x_m, w_m) \leq d_Y(x_m, y) + d_Y(y, w_m) < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon.$$

Hence $(x_n)_n \sim (w_n)_n$, i.e. $[(x_n)_n] = [(w_n)_n]$, and the map is well-defined.

Next, note that if $\gamma(y) = [(y_n)_n]$ and $\gamma(z) = [(z_n)_n]$, then

$$\begin{aligned} d^*(\gamma(y), \gamma(z)) &= \lim_n d_Y(y_n, z_n) \\ &\leq \lim_n d_Y(y_n, y) + \lim_n d_Y(y, z) + \lim_n d_Y(z, z_n) \\ &= 0 + d_Y(y, z) + 0 \\ &= d_Y(y, z). \end{aligned}$$

Conversely,

$$\begin{aligned} d^*(\gamma(y), \gamma(z)) &= \lim_n d_Y(y_n, z_n) \\ &\geq \lim_n (d_Y(y_n, z) - d_Y(z, z_n)) \\ &= \lim_n d_Y(y_n, z) - 0 \\ &\geq \lim_n (d_Y(y, z) - d_Y(y, y_n)) \\ &= d_Y(y, z) - 0 \\ &= d_Y(y, z). \end{aligned}$$

Thus γ is isometric, and in particular, γ is injective.

Suppose that $[(x_n)_n] \in X^*$. By definition, $(x_n)_n \in \Gamma[X]$, so that $(x_n)_n$ is a Cauchy sequence in X, and hence in Y. Since Y is complete, $x = \lim_n x_n \in Y$. But then $\gamma(x) = [(x_n)_n]$, so that γ is onto.

That is, (Y, d_Y) is isometrically isomorphic to (X^*, d^*), and so the completion of (X, d) is unique up to isometric isomorphism.

□

2.12. Remark. We know that $(\mathbb{R}, d)$ is a completion of $(\mathbb{Q}, d)$. It follows that $\mathbb{R}$ is isometrically isomorphic to

$$\mathbb{Q}^* = \{[(q_n)_n] : (q_n)_n \in \Gamma[\mathbb{Q}]\}.$$

Some of you will have seen Dedekind's construction of the real numbers via *Dedekind cuts*. Since this construction of the real numbers also results in a completion of $\mathbb{Q}$, it follows that this presentation of the reals is also isometrically isomorphic to $(\mathbb{R}, d)$.

2.13. Example. Let $(\mathfrak{X}, \|\cdot\|_{\mathfrak{X}})$ be a normed linear space. Let d denote the metric on $\mathfrak{X}$ induced by the norm.

Let

$$\mathfrak{X}^* := \{[(x_n)_n] : (x_n)_n \in \Gamma[\mathfrak{X}]\}$$

denote the equivalence classes of Cauchy sequences in $\Gamma[X]$. (The notation is a bit unfortunate, and should not be confused with an identical notation for dual spaces of normed linear spaces.)

We may define two operations $\cdot$ and $+$ on $\mathfrak{X}^*$ under which $(\mathfrak{X}^*, \cdot, +)$ becomes a vector space, namely:

(a) For $k \in \mathbb{K}$ and $[(x_n)_n] \in \mathfrak{X}^*$, we define $k \cdot [(x_n)_n] = [(kx_n)_n]$, and
(b) for $[(x_n)_n]$ and $[(y_n)_n] \in \mathfrak{X}^*$ we define $[(x_n)_n] + [(y_n)_n] = [(x_n + y_n)_n]$.

We leave it to the reader to verify that these operations are well-defined, and that $(\mathfrak{X}^*, \cdot, +)$ is indeed a vector space.

If we define

$$\|[(x_n)_n]\| := \lim_n \|x_n\|_{\mathfrak{X}},$$

then $(\mathfrak{X}^*, \|\cdot\|)$ is a norm on $\mathfrak{X}^*$, and $(\mathfrak{X}^*, \|\cdot\|)$ is the completion of $(\mathfrak{X}, \|\cdot\|_{\mathfrak{X}})$. The proof of this is left to the Assignments.

3. The relation between completeness and compactness in metric spaces

3.1. Let (X, d) be a metric space and $H \subseteq X$. Recall that H is compact if and only if H is sequentially compact, and that this happens if and only if every sequence $(x_n)_n$ in H admits a subsequence $(x_{n_k})_k$ which converges to an element of H. But any convergent subsequence in H is Cauchy.

This suggests that there should be some connection between compactness and completeness in metric spaces. This is indeed the case, as we shall now see.

3.2. Proposition. *Every compact metric space is complete.*
Proof. Let (K,d) be a compact metric space, and let $(x_n)_n$ be a Cauchy sequence in K. Recall from Theorem 3.2.37 that this is equivalent to saying that K is sequentially compact.

Thus, there exists a subsequence $(x_{n_k})_{k=1}^\infty$ of $(x_n)_n$ and an element $x \in K$ so that $\lim_{k\to\infty} x_{n_k} = x$. By Proposition 2.2.7, $\lim_n x_n = x \in K$. Thus Cauchy sequences in K converge, and so K is complete.

□

3.3. Proposition. *Let (X,d) be a metric space and suppose that $H \subseteq X$ is totally bounded. Then every sequence in H contains a Cauchy subsequence.*
Proof. Let $(x_n)_n$ be a sequence in H.

Since H is totally bounded, we may apply Proposition 3.2.30 to write $H = \dot\bigcup_{k=1}^n E_k$, where $\operatorname{diam} E_k < 1$ for all $1 \le k \le n$. At least one of the sets E_k contains infinitely many terms of the sequence - call it H_1. Note that $H_1 \subseteq H$ implies that H_1 is also totally bounded.

Suppose that we have constructed

$$H_1 \supset H_2 \supseteq \cdots \supseteq H_m$$

such that for each $1 \le k \le m$,

- H_k is totally bounded,
- $\operatorname{diam} H_k < \frac{1}{k}$, and
- H_k contains infinitely many terms of the sequence.

Since H_m is totally bounded, as above, we may write it as a disjoint union of finitely many sets, each with diameter less than $\frac{1}{m+1}$. At least one of these finitely many sets must contain infinitely many terms of the sequence - call it H_{m+1}. Observe that $H_{m+1} \subseteq H_m$ implies that H_{m+1} is again totally bounded, and by construction, $\operatorname{diam} H_{m+1} < \frac{1}{m+1}$.

Thus

$$H \supseteq H_1 \supset H_2 \supseteq \cdots,$$

each H_n contains infinitely many terms of the sequence, and $\operatorname{diam} H_n < \frac{1}{n}$ for all $n \ge 1$.

Choose $n_1 \ge 1$ so that $x_{n_1} \in H_1$. In general, if $m \ge 2$ and we are given $n_1 < n_2 < \dots < n_{m-1}$, the fact that H_m contains infinitely many terms of the sequence implies that we may choose $n_m > n_{m-1}$ so that $x_{n_m} \in H_m$.

Let $\varepsilon > 0$, and choose $N > \frac{1}{\varepsilon}$. If $k, l \ge N$, then $x_{n_k}, x_{n_l} \in H_N$ and so

$$d(x_{n_k}, x_{n_l}) \le \operatorname{diam} H_N < \frac{1}{N} < \varepsilon.$$

Thus $(x_{n_k})_{k=1}^\infty$ is Cauchy.

□

3.4. Theorem. *Let (X,d) be a metric space and $\varnothing \neq H \subseteq X$. The following are equivalent:*

(a) *H is compact.*
(b) *H is complete and totally bounded.*

Proof.

(a) implies (b). Suppose that H is compact. By Proposition 3.2, H is complete. By Theorem 3.2.37, H is sequentially compact. Finally, by Proposition 3.2.33, H is totally bounded.
(b) implies (a). Suppose that H is complete and totally bounded. Let $(x_n)_n$ be a sequence in H. By Proposition 3.3, $(x_n)_n$ admits a Cauchy subsequence $(x_{n_k})_{k=1}^{\infty}$. Since H is complete, $x = \lim_k x_{n_k} \in H$. Thus H is sequentially compact, and therefore compact, by Theorem 2.37.

□

3.5. Theorem. *Let (X,d) be a complete metric space and $\varnothing \neq H \subseteq X$. The following are equivalent:*

(a) *H is compact.*
(b) *H is closed and totally bounded.*

Proof.

(a) implies (b). Since X is Hausdorff in the metric topology, any compact subset of X is closed, by Theorem 3.2.13. Thus H is closed. But H is compact and therefore sequentially compact by Theorem 3.2.37, and so Proposition 3.2.33 implies that H is totally bounded.
(b) implies (a). Let $(x_n)_n$ be a Cauchy sequence in H. Then $(x_n)_n$ is Cauchy in X, and since X is complete, $(x_n)_n$ converges to some element $x \in X$. But then $x \in H$, since H is closed. Thus H is complete. Applying Theorem 3.4, we conclude that H is compact.

□

4. Appendix

Let us now provide alternative proofs of Theorems 4.1.10 and 4.1.14.

4.1. Theorem. *The normed linear space* $(\ell_1, \|\cdot\|_1)$ *is a Banach space.*

Proof. We have seen earlier that $(\ell_1, \|\cdot\|_1)$ is a normed linear space, and so it suffices to prove that it is complete. By Theorem 4.1.8, it suffices to prove that every absolutely summable series in ℓ_1 is summable.

Let $(x_n)_n$ be an absolutely summable series in ℓ_1, where $x_n = (x_{n,k})_{k=1}^\infty$. For each $k \ge 1$,

$$|x_{n,k}| \le \|x_n\|_1,$$

and therefore $\sum_{n=1}^\infty |x_{n,k}| \le \sum_{n=1}^\infty \|x_n\|_1 < \infty$. Since $\sum_{n=1}^\infty x_{n,k}$ is absolutely summable in the complete metric space $(\mathbb{K}, |\cdot|)$, it is summable. Set

$$z_k := \sum_{n=1}^\infty x_{n,k}, \qquad k \ge 1,$$

and $z = (z_k)_k$. We now model the proof after the proof of Theorem 4.1.13, namely: we compute the quantity that defines $\|z - \sum_{n=1}^p x_n\|_1$ and use this to show both that $z \in \ell_1$ and that $z = \lim_{p\to\infty} \sum_{n=1}^p x_n$.

Note that for each $k \ge 1$, $z_k - \sum_{n=1}^p x_{n,k} = \sum_{n=p+1}^\infty x_{n,k}$, and therefore

$$\begin{aligned}\sum_{k=1}^\infty |z_k - \sum_{n=1}^p x_{n,k}| &= \sum_{k=1}^\infty |\sum_{n=p+1}^\infty x_{n,k}| \\ &\le \sum_{k=1}^\infty \sum_{n=p+1}^\infty |x_{n,k}| \\ &= \sum_{n=p+1}^\infty \sum_{k=1}^\infty |x_{n,k}| \\ &= \sum_{n=p+1}^\infty \|x_n\|_1,\end{aligned}$$

where the fact that each $|x_{n,k}|$ is non-negative is what allows us to change the order of summation without affecting the sum.

Setting $p = 0$ then shows that

$$\sum_{k=1}^\infty |z_k| \le \sum_{n=1}^\infty \|x_n\|_1 < \infty,$$

(as $\sum_n x_n$ is absolutely summable), and therefore $z \in \ell_1$.

Also, given $\varepsilon > 0$, the fact that $\sum_n \|x_n\|_1 < \infty$ implies that there exists $P \ge 1$ so that $p \ge P$ implies that $\sum_{n=p+1}^\infty \|x_n\|_1 < \varepsilon$.

But then the above estimate shows that for $p \ge P$,

$$\|z - \sum_{n=1}^p x_n\|_1 = \sum_{k=1}^\infty |z_k - \sum_{n=1}^p x_{n,k}| \le \sum_{n=p+1}^\infty \|x_n\|_1 < \varepsilon.$$

This says that $z = \sum_n x_n$ in $(\ell_1, \|\cdot\|_1)$, and completes the proof.

□

4.2. Theorem. *The normed linear space* $(\mathcal{C}([0,1],\mathbb{K}), \|\cdot\|_\infty)$ *is a Banach space.*

Proof. We have already seen that $\mathcal{C}([0,1],\mathbb{K})$ is a normed linear space when equipped with the norm

$$\|f\|_\infty = \sup\{|f(x)| : x \in [0,1]\}.$$

There remains to show that it is complete.

Suppose that $\sum_n f_n$ is an absolutely convergent series in $\mathcal{C}([0,1],\mathbb{K})$ with $\sum_n \|f_n\|_\infty = M < \infty$. For each $x \in [0,1]$, clear $|f_n(x)| \le \|f_n\|_\infty$, and so

$$\sum_n |f_n(x)| \le \sum_n \|f_n\|_\infty \le M.$$

Since $\mathbb{K}$ is complete, $f(x) := \sum_n f_n(x)$ exists, by Theorem 4.1.8. We must show that

(a) $f \in \mathcal{C}([0,1],\mathbb{K})$, and
(b) $f = \sum_{n=1}^\infty f_n$.

Let $\varepsilon > 0$ and choose $N > 0$ so that $\sum_{n=N+1}^\infty \|f_n\|_\infty < \varepsilon/3$. Fix $x_0 \in [0,1]$. Now $g_N = \sum_{n=1}^N f_n$ is continuous, being a finite sum of continuous functions, and thus g_N is uniformly continuous, since $[0,1]$ is compact. Thus there exists $\delta > 0$ so that $|y - x| < \delta$ implies that

$$|g_N(x) - g_N(y)| < \frac{\varepsilon}{3}.$$

If $|x_0 - y| < \delta$, then

$$\begin{aligned} |f(x_0) - f(y)| &\le |f(x_0) - g_N(x_0)| + |g_N(x_0) - g_N(y)| + |g_N(y) - f(y)| \\ &\le \Big|\sum_{n=N+1}^\infty f_n(x_0)\Big| + \frac{\varepsilon}{3} + \Big|\sum_{n=N+1}^\infty f_n(y)\Big| \\ &\le \sum_{n=N+1}^\infty |f_n(x_0)| + \frac{\varepsilon}{3} + \sum_{n=N+1}^\infty |f_n(y)| \\ &\le \sum_{n=N+1}^\infty \|f_n\|_\infty + \frac{\varepsilon}{3} + \sum_{n=N+1}^\infty \|f_n\|_\infty \\ &< \frac{\varepsilon}{3} + \frac{\varepsilon}{3} + \frac{\varepsilon}{3} = \varepsilon. \end{aligned}$$

This shows that f is continuous.

Moreover, for $x \in [0,1]$, $m \geq N$,

$$\begin{aligned}|f(x) - g_m(x)| &= |\sum_{n=m+1}^{\infty} f_n(x)| \\ &\leq \sum_{n=m+1}^{\infty} |f_n(x)| \\ &\leq \sum_{n=m+1}^{\infty} \|f_n\|_\infty \\ &< \frac{\varepsilon}{3}.\end{aligned}$$

Hence $\|f - g_m\|_\infty \leq \frac{\varepsilon}{3} < \varepsilon$. Since $\varepsilon > 0$ was arbitrary,

$$f = \lim_{m\to\infty} g_m = \sum_n f_n.$$

By Theorem 4.1.8, $(\mathcal{C}([0,1],\mathbb{K}), \|\cdot\|_\infty)$ is complete, i.e. it is a Banach space

□

CHAPTER 5

The Baire Category Theorem

1. The Uniform Boundedness Principle

Once, during prohibition, I was forced to live for days on nothing but food and water.

W.C. Fields

1.1. The Baire-Category Theorem is a useful device in Functional Analysis, where it is used to prove the Open Mapping Theorem, the Closed Graph Theorem and the Uniform Boundedness Principle. We shall prove a version of the Uniform Boundedness Principle here. We shall then see the Banach-space version of the Uniform Boundedness Principle in the Assignments.

1.2. Definition. *Let (X, τ) be a topological space, and let $H \subseteq X$. A point $x \in H$ is said to be an* ***interior point*** *of H if there exists $G \in \tau$ so that $x \in G \subseteq H$; that is, if H is a neighbourhood of x.*

We denote by $\operatorname{int} H$ the set of interior points of H.

The ***exterior*** *of H is the set*

$$\operatorname{ext} H := \operatorname{int}(X \setminus H).$$

Finally, the ***boundary*** *of H is*

$$\partial H = X \setminus (\operatorname{int} H \cup \operatorname{ext} H).$$

1.3. Examples.

(a) Let $X = \{a, b, c, d, e\}$, and suppose that $\tau = \{\varnothing, \{a\}, \{c, d\}, \{a, c, d\}, \{b, c, d, e\}, X\}$. It is routine to check that τ is a topology on X.

Consider $H = \{b, c, d\}$. Then

- $\operatorname{int} H = \{c, d\}$;
- $\operatorname{ext} H = \{a\}$, and
- $\partial H = \{b, e\}$.

(b) Let $\mathbb{Q} \subseteq \mathbb{R}$, equipped with the standard topology. Given $x \in \mathbb{R}$ and $\delta > 0$, we see that

$$(x-\delta, x+\delta) \cap \mathbb{Q} \neq \varnothing \neq (x-\delta, x+\delta) \cap (\mathbb{R} \setminus \mathbb{Q}),$$

from which it follows that $\operatorname{int} \mathbb{Q} = \operatorname{ext} \mathbb{Q} = \varnothing$. Thus $\partial \mathbb{Q} = \mathbb{R}$.

1.4. Proposition. *Let (X, τ) be a topological space and $H \subseteq X$. Then*

(a) $\operatorname{int} H = \cup\{G \in \tau : G \subseteq H\}$.
(b) $\operatorname{int} H$ *is open.*
(c) $\operatorname{int} H$ *is the largest open set contained in H; that is, if $G \subseteq H$ is open, then $G \subseteq \operatorname{int} H$.*
(d) $H \in \tau$ *if and only if* $H = \operatorname{int} H$.

Proof.

(a) This is elementary, and is left as an exercise.
(b) By (a), $\operatorname{int} H$ is a union of open sets, and as such it is open.
(c) If $G \subseteq H$ is open, then $G \subseteq \operatorname{int} H$ by (a).
(d) By (b), $\operatorname{int} H$ is open, so if $H = \operatorname{int} H$, then H is open.
Conversely, if $H \in \tau$, then by (c), $H \subseteq \operatorname{int} H$ while by (a), $\operatorname{int} H \subseteq H$, so that $H = \operatorname{int} H$.

□

1.5. Proposition. *Let (X, τ) be a topological space, and let $H \subseteq X$. Then*

$$\overline{H} = \operatorname{int} H \cup \partial H.$$

Proof.

- First we observe that $\operatorname{int} H \subseteq H \subseteq \overline{H}$. Also, if $x \in \partial H$, then for any $G \in \tau$ for which $x \in G$, we have $G \cap H \neq \varnothing$. Thus $x \in \overline{H}$.
  Hence $\operatorname{int} H \cup \partial H \subseteq \overline{H}$.
- Suppose that $x \in \operatorname{ext} H$. Then there exists $G \in \tau$ so that $x \in G \subseteq (X \setminus H)$. But then $F := X \setminus G$ is closed, and $H \subseteq F$, so $\overline{H} \subseteq F$. Since $x \notin F$, it follows that $x \notin \overline{H}$. That is, $\overline{H} \subseteq (X \setminus \operatorname{ext} H) = \operatorname{int} H \cup \partial H$.

Taken together, these imply that

$$\overline{H} = \operatorname{int} H \cup \partial H.$$

□

1.6. Definition. *Let (X, τ) be a topological space. A subset $H \subseteq X$ is said to be* ***nowhere dense*** *in X if* $\operatorname{int}(\overline{H}) = \varnothing$.

1.7. Proposition. *Let (X, τ) be a topological space. A subset H of X is nowhere dense if and only if $X \setminus \overline{H}$ is dense in X.*

Proof.

- Suppose first that X is not nowhere dense. Then $\text{int}\,(\overline{H}) \neq \varnothing$. Choose $p \in G := \text{int}\,(\overline{H})$. Then $F := X \setminus G$ is closed, and $F \supseteq (X \setminus \overline{H})$. Thus $F \supseteq \overline{(X \setminus \overline{H})}$. Since $p \in G$, we have $p \notin F$, so $\overline{(X \setminus \overline{H})} \neq X$, i.e. $X \setminus \overline{H}$ is not dense in X.

  Thus $X \setminus \overline{H}$ dense in X implies that X is nowhere dense.
- Conversely, suppose that $X \setminus \overline{H}$ is not dense in X. Choose $q \in G := X \setminus \overline{(X \setminus \overline{H})}$. Then G is open and $q \in G \subseteq X \setminus (X \setminus \overline{H}) = \overline{H}$, so $\text{int}\,\overline{H} \neq \varnothing$, i.e. H is not nowhere dense in X.

  Thus X nowhere dense in X implies that $X \setminus \overline{H}$ is dense in X.

□

1.8. Examples.

(a) Let $H = \mathbb{Z} \subseteq \mathbb{R}$. Then H is closed, and $\overline{\mathbb{R} \setminus \mathbb{Z}} = \mathbb{R}$, so $\mathbb{Z}$ is nowhere dense in $\mathbb{R}$.

(b) Let $H = \mathbb{Q} \cap (0,1) \subseteq \mathbb{R}$. Then $\overline{H} = [0,1]$, and so $\text{int}\,\overline{H} = (0,1) \neq \varnothing$. Thus H is *not* nowhere dense in $\mathbb{R}$.

1.9. Definition. *Let (X, τ) be a topological space. A subset $H \subseteq X$ is said to be of the* ***first category*** *(or H is* ***meagre****) if H is contained in the countable union of closed, nowhere dense subsets of X.*

Otherwise, H is said to be of the ***second category*** *(or* ***non-meagre****).*

1.10. Theorem. (*The Baire Category Theorem - 1*)

Let (X, d) be a complete metric space and $\{G_n\}_{n=1}^{\infty}$ be a countable collection of dense, open sets in X. Then

$$\bigcap_{n=1}^{\infty} G_n \neq \varnothing.$$

Proof. Let $x_1 \in G_1$. Since G_1 is open, we can find $\delta_1 > 0$ so that $B(x_1, \delta_1) \subseteq G_1$.

Next, G_2 is dense in X, and so we can find $x_2 \in G_2 \cap B(x_1, \delta_1)$. Since $G_2 \cap B(x_1, \delta_1)$ is open, there exists $0 < \delta_2 \leq \frac{\delta_1}{2}$ so that

$$\overline{B}(x_2, \delta_2) \subseteq G_2 \cap B(x_1, \delta_1).$$

Similarly, since G_3 is dense in X, we may find $x_3 \in G_3 \cap B(x_2, \delta_2)$. Since $G_3 \cap B(x_2, \delta_2)$ is open, there exists $0 < \delta_3 < \frac{\delta_2}{2}$ so that

$$\overline{B}(x_3, \delta_3) \subseteq G_3 \cap B(x_2, \delta_2),$$

and proceeding recursively, for each $n \geq 1$, we may find $\delta_n < \min \frac{\delta_{n-1}}{2}$ so that

$$\overline{B}(x_n, \delta_n) \subseteq G_{n-1} \cap B(x_{n-1}, \delta_{n-1}).$$

Thus

$$\overline{B}(x_n, \delta_n) \subseteq B(x_{n-1}, \delta_{n-1}) \subseteq \overline{B}(x_{n-1}, \delta_{n-1})$$

for all $n \ge 2$, and $\operatorname{diam} \overline{B}(x_n, \delta_n) = 2\delta_n$ converges to 0 as n tends to infinity.

By Theorem 4.1.15, $\cap_{n=1}^{\infty} \overline{B}(x_n, \delta_n) \ne \varnothing$. Choose $p \in \cap_{n=1}^{\infty} \overline{B}(x_n, \delta_n)$.

Then for each $n \ge 2$, $p \in \overline{B}(x_n, \delta_n) \subseteq G_{n-1}$, and thus

$$p \in \bigcap_{n=1}^{\infty} G_n \ne \varnothing.$$

□

1.11. Corollary. (***The Baire Category Theorem - 2***)

Let (X, d) be a complete metric space. Then X is of the second category.

Proof. Let $\{F_n\}_n$ be a countable collection of closed, nowhere dense subsets of X. For each $n \ge 1$, set $G_n = X \setminus F_n$, so that G_n is open and dense. By Theorem 1.10,

$$\bigcap_{n=1}^{\infty} G_n \ne \varnothing,$$

or equivalently,

$$\bigcup_{n=1}^{\infty} F_n \ne X.$$

Thus X is of the second category.

□

One of the best known and most important applications of the Baire Category Theorem is the following:

1.12. Theorem. (***The Uniform Boundedness Principle***)

Let (X, d) be a complete metric space and suppose that $\varnothing \ne \mathcal{F} \subseteq \mathcal{C}(X, \mathbb{K})$, where $\mathcal{C}(X, \mathbb{K}) = \{f : X \to \mathbb{K} : f \text{ is continuous}\}$.

Suppose that for each $x \in X$ there exists $\kappa_x > 0$ so that

$$|f(x)| \le \kappa_x \text{ for all } f \in \mathcal{F}.$$

Then there exists a non-empty open set $G \subseteq X$ and $\kappa > 0$ so that

$$|f(x)| \le \kappa \text{ for all } x \in G \text{ and for all } f \in \mathcal{F}.$$

In other words, if $\mathcal{F}$ is a family of continuous, $\mathbb{K}$-valued functions on a complete metric space X which are *pointwise bounded*, then there exists an open set $G \subseteq X$ where the collection $\mathcal{F}$ is *uniformly bounded*.

Proof. For each $m \ge 1$, let

$$H_{m,f} := \{x \in X : |f(x)| \le m\},$$

and let

$$H_m := \bigcap_{f \in \mathcal{F}} H_{m,f}.$$

Since each $f \in \mathcal{F}$ is continuous, each $H_{m,f}$ is easily seen to be closed, and so H_m is closed, being the intersection of closed sets. Moreover, for each $x \in X$, there exists $m \ge 1$ so that $|f(x)| \le m$ for all $f \in \mathcal{F}$, and so there exists $m \ge 1$ so that $x \in H_m$. That is, $X = \cup_{m=1}^{\infty} H_m$.

But X is complete, so by the Baire Category Theorem, at least one of the sets H_m fails to be nowhere dense.

That is, there exists $N \geq 1$ so that $G := \operatorname{int} H_N \neq \varnothing$. Let $\kappa = N$ to get that for $x \in G$,

$$|f(x)| \leq \kappa \text{ for all } f \in \mathcal{F}.$$

$\square$

CHAPTER 6

Spaces of continuous functions

1. Urysohn's Lemma and Tietze's Extension Theorem

Do you know what it means to come home at night to a woman who'll give you a little love, a little affection, a little tenderness? It means you're in the wrong house, that's what it means.

Henny Youngman

1.1. A great deal can be learnt about a topological space (X, τ) by studying the algebra $\mathcal{C}(X, \mathbb{K})$ of continuous, $\mathbb{K}$-valued functions that act upon it. This is the central theme of this Chapter. The motivating example will be $\mathcal{C}([0,1], \mathbb{R})$.

1.2. Definition. *A topological space (X, τ) is said to be* ***normal*** *if, given disjoint, closed subsets F_1, F_2 of X, we can find disjoint open sets G_1, G_2 of X such that $F_1 \subseteq G_1$ and $F_2 \subseteq G_2$.*

We say that (X, τ) is $\mathbf{T_1}$ *if, given points $x \neq y$ in X, we can find an open set $G \in \tau$ such that $y \in G$ but $x \notin G$.*

Finally, we say that (X, τ) is $\mathbf{T_4}$ *if X is both normal and T_1.*

1.3. Example. As we saw in the Assignments, every metric space (X, d) is T_4.

1.4. Example. If τ_{cf} denotes the co-finite topology on $\mathbb{N}$, then as we have seen, $(\mathbb{N}, \tau_{cf})$ is not T_2 (i.e. not Hausdorff), since any two non-empty open sets must intersect. It follows that $(\mathbb{N}, \tau_{cf})$ is not normal. On the other hand, by definition, every singleton set is closed, and thus $(\mathbb{N}, \tau_{cf})$ is T_1.

1.5. Lemma. *Let (X, τ) be a Hausdorff topological space. The following are equivalent.*

(a) *(X, τ) is T_4;*
(b) *Given $F \subseteq X$ closed, and $G \in \tau$ satisfying $F \subseteq G$, there exists a set $U \in \tau$ so that*

$$F \subseteq U \subseteq \overline{U} \subseteq G.$$

Proof.

(a) implies (b). Suppose that X is T_4. Then, with F, G as in (b), observe that F and $X \setminus G$ are disjoint, closed subsets of X. Since X is normal, we can find disjoint open sets U and V such that $F \subseteq U$ and $X \setminus G \subseteq V$. But then $U \subseteq X \setminus V$, and so $\overline{U} \subseteq \overline{X \setminus V} = X \setminus V \subseteq G$. That is,

$$F \subseteq U \subseteq \overline{U} \subseteq G.$$

(b) implies (a). Conversely, suppose that (b) holds, and let F_1, F_2 be disjoint closed subsets of X.

Then $X \setminus F_2$ is open, and $F_1 \subseteq X \setminus F_2$. By hypothesis, we can find $U \in \tau$ so that

$$F_1 \subseteq U \subseteq \overline{U} \subseteq X \setminus F_2.$$

Setting $V = X \setminus \overline{U}$, we see that V is open and $F_2 = X \setminus (X \setminus F_2) \subseteq V$. Moreover, $U \cap V = U \cap (X \setminus \overline{U}) = \varnothing$. Hence X is normal. Since X is Hausdorff, it is also T_1, and therefore T_4 as well.

□

1.6. Lemma. *Let (X, τ) be T_4-space, $F \subseteq X$ be closed, and suppose that $F \subseteq G$ for some set $G \in \tau$.*

Then there exists a continuous function $f : X \to \mathbb{R}$ such that

(a) $0 \le f(x) \le 1$ *for all* $x \in X$*;*
(b) $f(x) = 0$ *for all* $x \in F$*, and*
(c) $f(x) = 1$ *for all* $x \in X \setminus G$*.*

Proof. By Lemma 1.5 above, we can find $U_{\frac{1}{2}} \in \tau$ so that

$$F \subseteq U_{\frac{1}{2}} \subseteq \overline{U_{\frac{1}{2}}} \subseteq G.$$

Since $U_{\frac{1}{2}}$ is open and $\overline{U_{\frac{1}{2}}}$ is closed, a second application of Lemma 1.5 yields open sets $U_{\frac{1}{4}}$ and $U_{\frac{3}{4}}$ so that

$$F \subseteq U_{\frac{1}{4}} \subseteq \overline{U_{\frac{1}{4}}} \subseteq U_{\frac{1}{2}} \subseteq \overline{U_{\frac{1}{2}}} \subseteq U_{\frac{3}{4}} \subseteq \overline{U_{\frac{3}{4}}} \subseteq G.$$

We may continue in this manner and so by obtain for each dyadic rational in $(0, 1)$ (i.e. for each $q \in \mathcal{D} := \{\frac{p}{2^n} : 0 < p < 2^n, n \ge 1\}$) an open set U_q with the property that if $q_1, q_2 \in \mathcal{D}$ and $q_1 < q_2$, then

$$\overline{U_{q_1}} \subseteq U_{q_2}.$$

Let $U_1 = X$ and define

$$\begin{array}{cccc} f : & X & \to & \mathbb{R} \\ & x & \mapsto & \inf\{t \in \mathcal{D} : x \in U_t\}. \end{array}$$

If $x \notin G$, then – since $U_{\frac{p}{2^n}} \subseteq G$ for all n, p as above – we have

$$\inf\{t \in \mathcal{D} : x \in U_t\} = 1, \text{ i.e., } f(x) = 1.$$

If $x \in F$, then $x \in U_{\frac{1}{2^n}}$ for all $n \ge 1$, and thus $f(x) = 0$.

There remains to prove that f is continuous.

- First, we claim that

$$f^{-1}([0,a)) = \cup\{U_t : t \in \mathcal{D}, t < a\}.$$

Indeed, if $x \in f^{-1}([0,a))$, then $0 < f(x) < a$, and thus we can find $q \in \mathcal{D}$ with $f(x) < q < a$. Hence $x \in U_q$, and so $x \in \cup\{U_q : q \in \mathcal{D}, q < a\}$.

Conversely, suppose that $x \in \cup\{U_t : t \in \mathcal{D}, t < a\}$, say $x \in U_q$ for some fixed $q \in \mathcal{D}$ with $q < a$. Then $f(x) \le q < a$, so $x \in f^1([0,a))$.

Together, these prove the claim.

- Next, we show that $f^{-1}((b,1]) = \cup\{X \setminus \overline{U_t} : t \in \mathcal{D}, b < t\}$.

Suppose that $x \in f^{-1}((b,1])$, so that $f(x) > b$. Then we can find $q_1 < q_2 \in \mathcal{D}$ with $b < q_1 < q_2 < f(x)$. In particular, $\inf\{t \in \mathcal{D} : x \in U_t\} > q_2$, and so $x \notin u_{q_2}$, i.e. $x \in X \setminus U_{q_2}$. Since $\overline{U_{q_1}} \subseteq U_{q_2}$ by construction, we have that $x \in X \setminus \overline{U_{q_1}}$ with $q_1 > b$, and hence

$$x \in \cup\{X \setminus \overline{U_t} : t \in \mathcal{D}, b < t\}.$$

Conversely, suppose that $x \in \cup\{X \setminus \overline{U_t} : t \in \mathcal{D}, b < t\}$. Choose $q \in \mathcal{D}$, $b < q$ so that $x \in X \setminus \overline{U_q}$. Since $U_q \subseteq \overline{U_q}$, we therefore have that $x \notin U_q$ and so $b < q \le f(x)$.

That is, $x \in f^{-1}((b,1])$, so that

$$f^{-1}((b,1]) = \cup\{X \setminus \overline{U_t} : b < t\}.$$

Finally, since $f^{-1}([0,a))$, $f^{-1}((b,1])$ are unions of open sets, they are themselves open. Hence

$$f^{-1}(b,a) = f^{-1}((b,1]) \cap f^{-1}(([0,a))$$

is open for each $b < a$. Since every open set L in $[0,1]$ is a union of sets of the form (b,a), $(b,1]$, or $[0,a)$, $f^{-1}(L)$ is again a union of open sets, and hence it is open.

Thus f is continuous, as required.

□

1.7. Lemma. (***Urysohn's Lemma - special case.***) *Let $(X,\mathcal{T})$ be a T_4-topological space, and let A, B be disjoint, closed sets in X. Then there exists a continuous function $f : X \to [0,1]$ so that $f|_A \equiv 0$ and $f|_B \equiv 1$.*

Proof. Since X is T_4, we can find disjoint open sets U and V in X with $A \subseteq U$ and $B \subseteq V$. By Lemma 1.6, we can find a continuous function $f : X \to [0,1]$ so that $f|_A \equiv 0$ and $f \equiv 1$ on $X \setminus U$. But $B \subseteq V \subseteq (X \setminus U)$, and so in particular, $f|_B \equiv 1$.

□

1.8. Theorem. (***Urysohn's Lemma - general case.***)
Let $(X, \mathcal{T})$ be a T_4-topological space, and let A, B be disjoint, closed sets in X. Let $a < b \in \mathbb{R}$. Then there exists a continuous function $g : X \to [a, b]$ so that $g(y) = a$ for all $y \in A$ and $g(z) = b$ for all $z \in B$.

Proof. By the special case of Urysohn's Lemma 1.7, we can find a continuous function $f : X \to [0, 1]$ so that $f(y) = 0$ for all $y \in A$ and $f(z) = 1$ for all $z \in B$. Let $g = (b - a)f + a\mathbf{1}$, where $\mathbf{1}$ is the constant function $\mathbf{1} : X \to \mathbb{R}$, $\mathbf{1}(x) = 1$ for all $x \in X$.

It is routine to verify that g satisfies the stated conditions.

□

1.9. Lemma. *The function*

$$\begin{array}{rccc} \tau : & \mathbb{R} & \to & (-1,1) \\ & z & \mapsto & \frac{z}{1+|z|} \end{array}$$

is a homeomorphism. The inverse of this function is

$$\begin{array}{rccc} \tau^{-1} : & (-1,1) & \to & \mathbb{R} \\ & w & \mapsto & \frac{w}{1-|w|} \end{array}.$$

Proof. Exercise.

□

Before proving Tietze's Extension Theorem, we pause to recall a result from the Assignments which we shall need below.

1.10. Theorem. *Let (X, τ) be a topological space and let*

$$\mathcal{C}_b(X, \mathbb{K}) = \{f : X \to \mathbb{K} : f \text{ is bounded -- i.e. } \sup_{x \in X} |f(x)| < \infty\},$$

equipped with the norm $\|f\|_\infty = \sup_{x \in X} |f(x)|$. Then $(\mathcal{C}_b(X, \mathbb{K}), \|\cdot\|_\infty)$ is a Banach space.

1.11. Theorem. (***Tietze's Extension Theorem.***)
Let $(X, \mathcal{T})$ be a T_4-topological space, and let $E \subseteq X$ be a closed set. Suppose that $f : E \to \mathbb{R}$ is a continuous function. Then there exists a continuous function $g : X \to \mathbb{R}$ so that $g|_E = f$.

Proof.

- **Case One: $|f(x)| \le 1$ for all $x \in E$.**

  Let

  $$A_1 := \{x \in E : -1 \le f(x) \le -1/3\} = f^{-1}([-1, -1/3]),$$

  and let

  $$B_1 := \{x \in E : 1/3 \le f(x) \le 1\} = f^{-1}([1/3, 1]).$$

  Since f is continuous and $[-1, -1/3]$, $[1/3, 1]$ are closed, A_1 and B_1 are closed.

  By the general case of Urysohn's Lemma, there exists a function $h_1 : X \to [-\frac{1}{3}, \frac{1}{3}]$ such that $h_1|_{A_1} \equiv -\frac{1}{3}$, and $h_1|_{B_1} \equiv \frac{1}{3}$.

We claim that $|f(x) - h_1(x)| < \frac{2}{3}$ for all $x \in E$. Indeed,

- If $x \in A_1$, then $h_1(x) = -\frac{1}{3}$ and $-1 \le f(x) \le -\frac{1}{3}$, so $|f(x) - h_1(x)| \le \frac{2}{3}$;
- if $x \in B_1$, then $h_1(x) = \frac{1}{3}$ and $\frac{1}{3} \le f(x) \le 1$, so $|f(x) - h_1(x)| \le \frac{2}{3}$;
- if $x \in E\backslash(A_1 \cup B_1)$, then $-\frac{1}{3} \le h_1(x) \le \frac{1}{3}$, $-\frac{1}{3} \le f(x) \le \frac{1}{3}$, and so $|f(x) - h_1(x)| \le |f(x)| + |h_1(x)| \le \frac{2}{3}$.

Let $t = \frac{2}{3}$. We now argue by induction that there exists, for each $n \ge 1$, a continuous function $h_n : X \to \mathbb{R}$ so that

(i) $\|h_n\|_\infty := \sup_{x\in X} |h_n(x)| \le \frac{1}{3}t^{n-1}$, and
(ii) $|f(x) - \sum_{j=1}^n h_j(x)| \le t^n, \quad x \in E$.

The argument above shows that the statement holds in the case where $n = 1$. Suppose that $m \ge 1$ and that the statement holds for $n \le m$.

Let

$$A_{m+1} := \left\{ x \in E : -t^m \le f(x) - \sum_{j=1}^m h_j(x) \le -\frac{1}{3}t^m \right\},$$

and let

$$B_{m+1} := \left\{ x \in E : \frac{1}{3}t^m \le f(x) - \sum_{j=1}^m h_j(x) \le t^m \right\}.$$

By the general case of Urysohn's Lemma, we can find a continuous function $h_{m+1} : X \to [-\frac{1}{3}t^m, \frac{1}{3}t^m]$ so that

(i) $h_{m+1}|_{A_{m+1}} \equiv -\frac{1}{3}t^m$, and
(ii) $h_{m+1}|_{B_{m+1}} \equiv \frac{1}{3}t^m$.

Now $|h_{m+1}(x)| \le \frac{1}{3}t^m$ for all $x \in X$, so $\|h_{m+1}\|_\infty \le \frac{1}{3}t^m$. We leave it to the reader to verify that $|f(x) - \sum_{j=1}^{m+1} h_j(x)| \le \frac{2}{3}t^m = t^{m+1}$ for all $x \in E$. This completes the induction step.

For each $n \ge 1$, h_n is continuous and bounded, and so $h_n \in \mathcal{C}_b(X, \mathbb{R})$. Consider the series

$$\sum_{n=1}^\infty \|h_n\|_\infty \le \sum_{n=1}^\infty \frac{1}{3}t^{n-1} = \frac{1}{3}\frac{1}{1-t} = \frac{1}{3}(3) = 1.$$

Since $\sum_{n=1}^\infty h_n$ is absolutely summable in the complete normed linear space $(\mathcal{C}_b(X, \mathbb{R}), \|\cdot\|_\infty)$, it is summable.

Let $g = \sum_{n=1}^\infty h_n \in \mathcal{C}_b(X, \mathbb{R})$. Then $\|g\|_\infty \le \sum_{n=1}^\infty \|h_n\|_\infty \le 1$ - i.e. $g : X \to [-1, 1]$.

For each $x \in E$,

$$\begin{aligned} 0 \le |f(x) - g(x)| &= \lim_{n\to\infty} |f(x) - \sum_{j=1}^n h_j(x)| \\ &\le \lim_{n\to\infty} t^n \\ &= 0, \end{aligned}$$

as $t = \frac{2}{3} < 1$. Thus $g(x) = f(x)$ for all $x \in F$, and so $g|_E \equiv f$.

This completes the proof of the case where $|f(x)| \le 1$ for all $x \in E$.

- **Case Two: $|f(x)| > 1$ for some $x \in E$.**

  Let $\tau : \mathbb{R} \to (-1,1)$ be the homeomorphism from the previous Lemma. Consider the function $f_0 := \tau \circ f$, so that $f_0 : E \to (-1,1) \subseteq [-1,1]$. Since f is continuous on E and τ is continuous on $\mathbb{R}$, we know that f_0 is continuous on E. We can therefore apply **Case One** above to the function f_0 to obtain a continuous function $g_0 : X \to [-1,1]$ so that $g_0|_E = f_0$.

  Let $D = g_0^{-1}(\{-1,1\})$. Since $\{-1,1\}$ is closed in $[-1,1]$, and since g_0 is continuous, D is closed in X. Recall that $E \subseteq X$ is also closed. If $x \in E$, then $g_0(x) = f_0(x) = \tau \circ f(x) \subseteq \operatorname{ran}\tau \subseteq (-1,1)$, and so $x \notin D$. That is, $D \cap E = \varnothing$.

  Thus D and E are disjoint, closed subsets of the T_4 space X. Once again, we can apply (the special case of) Urysohn's Lemma to obtain a function $p : X \to [0,1]$ so that $p|_D \equiv \mathbf{0}$, while $p|_E \equiv \mathbf{1}$.

  Let $q(x) = p(x)g_0(x)$, $x \in X$. Since p and g_0 are both continuous on X, so is q. Moreover, for all $x \in X$, $|q(x)| < 1$. Indeed,
  - if $x \notin D$, then $|g_0(x)| < 1$. Since $|q(x)| \le |g_0(x)|$ for all $x \in X$, we have $|q(x)| < 1$ for $x \notin D$.
  - If $x \in D$, then $p(x) = 0$, so $q(x) = p(x)g_0(x) = 0$.

  We have shown that $\operatorname{ran} q \subseteq (-1,1) \subseteq \operatorname{Domain}\tau^{-1}$. Let $g = \tau^{-1} \circ q : X \to \mathbb{R}$. Since q and τ^{-1} are continuous, so is g.

  If $x \in E$, then

$$\begin{aligned} g(x) &= \tau^{-1} \circ q(x) \\ &= \tau^{-1} \circ (pg_0)(x) \\ &= \tau^{-1} \circ g_0(x) \qquad \text{since } p(x) = 1 \text{ for all } x \in E \\ &= \tau^{-1} \circ (\tau \circ f)(x) \\ &= (\tau^{-1} \circ \tau) \circ f(x) \\ &= f(x). \end{aligned}$$

  Thus $g|_E = f$, and g is the desired extension of f.

□

2. The Stone-Weierstraß Theorem

2.1. One of the most beautiful results in this course is Weierstraß's Approximation Theorem, which says that every continuous $\mathbb{K}$-valued function on a closed interval $[a,b]$ can be uniformly approximated by polynomials. This and Stone's generalization of Weierstraß's Theorem find ubiquitous applications in Analysis.

2.2. Definition. *An **algebra** $\mathcal{A}$ is a vector space over $\mathbb{K}$ equipped with a multiplication map $\mu : \mathcal{A} \times \mathcal{A} \to \mathcal{A}$ which is $\mathbb{K}$-bilinear; that is,*

$$k\mu(a,b) = \mu(ka,b) = \mu(a,kb)$$

or equivalently

$$k(ab) = (ka)b = a(kb)$$

for all $a, b \in \mathcal{A}$ and $k \in \mathbb{K}$, and

$$(a_1 + a_2)b = \mu(a_1 + a_2, b) = \mu(a_1, b) + \mu(a_2, b) = a_1 b + a_2 b,$$

while

$$a(b_1 + b_2) = \mu(a, b_1 + b_2) = \mu(a, b_1) + \mu(a, b_2) = ab_1 + ab_2$$

for all $a, a_1, a_2, b, b_1, b_2 \in \mathcal{A}$.

It is perhaps worth mentioning that every algebra over $\mathbb{C}$ is automatically an algebra over $\mathbb{R}$.

2.3. Examples.

(a) The space $\mathbb{M}_n(\mathbb{K})$ is an algebra over $\mathbb{K}$.

(b) Let (X, τ) be a topological space. Then

$$\mathbb{C}(X, \mathbb{K}) = \{f : X \to \mathbb{K} : f \text{ is continuous}\}$$

is an algebra over $\mathbb{K}$, where $(fg)(x) = f(x)g(x)$ for all $x \in X$.

(b) Recall that

$$\ell_\infty = \ell_\infty(\mathbb{N}) = \{(x_n)_n \in \mathbb{K}^{\mathbb{N}} : \sup_n |x_n| < \infty\}$$

is a Banach space with the norm $\|(x_n)_n\|_\infty = \sup_n |x_n|$.

If we set

$$(x_n)_n \cdot (y_n)_n = (x_n y_n)_n,$$

then ℓ_∞ becomes an algebra.

2.4. Our goal in this section is to study the density of subalgebras $\mathcal{A}$ of $\mathcal{C}(X, \mathbb{K})$. We shall find conditions both on X and on $\mathcal{A}$ to allow us to conclude that $\mathcal{A}$ is dense in $\mathcal{C}(X, \mathbb{K})$. Interestingly, the key point will be that $\mathcal{C}(X, \mathbb{R})$ admits an interesting *lattice* structure, which we now define.

2.5. Definition. *A **lattice** is a poset $(\mathcal{L}, \leq)$ in which each pair $\{x, y\}$ of elements has both a greatest lower bound $x \wedge y \in \mathcal{L}$ and a least upper bound $x \vee y \in \mathcal{L}$.*

The notations are analogous to, and intended to suggest, the notations for intersections and unions of sets.

2.6. Example. Let $\varnothing \neq X$ be a set and consider the power set $(\mathcal{P}(X), \leq)$, partially ordered by inclusion. If $A, B \in \mathcal{P}(X)$, set $A \vee B = A \cup B$, and $A \wedge B = A \cap B$. Then $(\mathcal{P}(X), \leq)$ is a lattice.

2.7. Example. Let (X, τ) be a topological space. As a generalization of Example 1.2.8 we may partially order $\mathcal{C}(X, \mathbb{R})$ by the relation

$$f \le g \text{ if } f(x) \le g(x) \text{ for all } x \in X.$$

We claim that $\mathcal{C}(X, \mathbb{R})$ admits a lattice structure where we set

$$[f \wedge g](x) = \min(f(x), g(x)) \quad \text{and}$$
$$[f \vee g](x) = \max(f(x), g(x)), x \in X.$$

It is clear that *if* $f \wedge g$ and $f \vee g$ as defined above are continuous, then they will be the meet and join of f and g. To see that they are continuous, observe that for each $x \in X$,

$$\min(f(x), g(x)) = \frac{f(x) + g(x)}{2} - \frac{|f(x) - g(x)|}{2},$$

so

$$\min(f, g) = \frac{f + g}{2} - \frac{|f - g|}{2} \qquad \text{and}$$
$$\max(f, g) = \frac{f + g}{2} + \frac{|f - g|}{2}.$$

Since sums, quotients (where defined) and absolute values preserve continuity, $f \wedge g$ and $f \vee g$ are continuous.

2.8. Proposition. *Let (X, τ) be a compact topological space and $\mathcal{L} \subseteq \mathcal{C}(X, \mathbb{R})$ be a lattice. Suppose that the function*

$$h(x) := \inf_{f \in \mathcal{L}} f(x)$$

is continuous on X.

Then, given $\varepsilon > 0$ there exists $g \in \mathcal{L}$ such that

$$0 \le g(x) - h(x) < \varepsilon \quad \textit{for all } x \in X.$$

Proof. Let $\varepsilon > 0$. For each $x \in X$ we may choose $f_x \in \mathcal{L}$ so that

$$f_x(x) < h(x) + \frac{\varepsilon}{3}.$$

Note that $f_x \in \mathcal{L} \subseteq \mathcal{C}(X, \mathbb{R})$ and h continuous implies that there exists an open set $G_x \in \tau$ so that

- $|f_x(y) - f_x(x)| < \frac{\varepsilon}{3}$ and
- $|h(y) - h(x)| < \frac{\varepsilon}{3}$

for all $y \in G_x$. In particular,

$$f_x(y) - h(y) < \varepsilon, \quad y \in G_x.$$

Of course, $X \subseteq \bigcup_{x \in X} G_x$, so $\{G_x\}_{x \in X}$ is an open cover of the compact set X, and as such, it admits a finite subcover. Choose $x_1, x_2, ..., x_N \in X$ so that

$$X \subseteq \cup_{n=1}^{N} G_{x_n}.$$

Let $g = f_{x_1} \wedge f_{x_2} \wedge \cdots \wedge f_{x_N}$. Then $g \in \mathcal{L}$, since $\mathcal{L}$ is a lattice, and if $y \in X$, then $y \in G_{x_n}$ for some $1 \leq n \leq N$, so

$$g(y) - h(y) \leq f_{x_n}(y) - h(y) < \varepsilon.$$

□

2.9. Definition. *Let $\varnothing \neq E$ be a set, and suppose that $\mathcal{A} \subseteq \mathbb{K}^E$. We say that $\mathcal{A}$* ***separates points*** *of E if for each pair $x \neq y \in E$ we can find $f \in \mathcal{A}$ such that $f(x) \neq f(y)$.*

We say that $\mathcal{A}$ is ***non-vanishing*** *on E if for each $x \in E$ there exists $f \in \mathcal{A}$ so that $f(x) \neq 0$.*

2.10. Example. Let $\mathbb{K}[x]$ denote the algebra of all polynomials on $\mathbb{K}$. That is,

$$\mathbb{K}[x] = \{\mathbf{p} = p_0 + p_1 x + p_2 x^2 + \cdots + p_n x^n : n \geq 1,\ p_k \in \mathbb{K}, 0 \leq k \leq n\}.$$

Then $\mathbb{K}[x]$ separates the points of $\mathbb{K}$. In fact, the single polynomial $\mathbf{p}(x) = x$ already separates the points of $\mathbb{K}$, and $\{\mathbf{p}\} \subseteq \mathbb{K}$.

Let

$$\mathcal{E} = \{\mathbf{q} = q_0 + q_1 x^2 + q_2 x^4 + \cdots + q_m x^{2m} : m \geq 1,\ q_k \in \mathbb{K}, 0 \leq k \leq m\}.$$

Then $\mathcal{E}$ consists of the algebra of **even polynomials** on $\mathbb{K}$. Note that $\mathcal{E}$ does not separate points of $\mathbb{K}$, since $\mathbf{q}(x) = \mathbf{q}(-x)$ for all $x \in \mathbb{K}$. In particular, $\mathcal{E}$ does not separate 1 from -1.

Since each of $\mathbb{K}[x]$ and $\mathcal{E}$ contains the constant functions (and in particular, the non-zero constant functions), they are both non-vanishing on $\mathbb{K}$.

2.11. Example. Let $\mathbb{D} = \{z \in \mathbb{C} : |z| < 1\}$. Let

$$\mathcal{P}_0 = \{\mathbf{p} = p_1 x + p_2 x^2 + \cdots p_n x^n : n \geq 1,\ p_k \in \mathbb{C}, 1 \leq k \leq n\}.$$

Then $\mathcal{P}_0$ is separating on $\mathbb{D}$, since the identity map $\mathbf{p}(x) = x$ lies in $\mathcal{P}_0$, but $\mathcal{P}_0$ vanishes at 0, since $0 \in \mathbb{D}$ and $\mathbf{q} \in \mathcal{P}_0$ implies that $\mathbf{q}(0) = 0$.

2.12. Lemma. *Let (X, τ) be a compact, topological space and $\mathcal{M} \subseteq \mathcal{C}(X, \mathbb{K})$ be a $\mathbb{K}$-vector space of continuous functions. Suppose that $\mathcal{M}$ separates the points of X and contains the constant function* $\mathbf{1}$.

Given any two elements $a, b \in \mathbb{K}$ and $x \neq y \in X$, there exists $f \in \mathcal{M}$ so that $f(x) = a$ and $f(y) = b$.

Proof. Since $\mathcal{M}$ is separating for X, we can find $g \in \mathcal{M}$ so that $g(x) \neq g(y)$. Then

$$f = \frac{a - b}{g(x) - g(y)} g + \frac{b g(x) - a g(y)}{g(x) - g(y)} \mathbf{1}$$

is the desired function.

□

2.13. Lemma. *Let (X, τ) be a compact, topological space. Suppose that $\mathcal{L} \subseteq \mathcal{C}(X, \mathbb{R})$ is both a vector space over $\mathbb{R}$ and a lattice (under the partial ordering of Example 2.7). Suppose furthermore that $\mathcal{L}$ separates the points of X and that $\mathcal{L}$ contains the constant functions.*

If $a, b \in \mathbb{R}$, $F \subseteq X$ is closed and $p \in X \setminus F$, then there exists $f \in \mathcal{L}$ such that

- *$f(x) \ge a$ for all $x \in X$;*
- *$f(p) = a$; and*
- *$f(x) > b$ for all $x \in F$.*

Proof. Since $F \subseteq X$ is closed and X is compact, F is also compact, by Proposition 3.2.15. By Lemma 2.12, for each $x \in F \subseteq X \setminus \{p\}$, we can find an element $f_x \in \mathcal{L}$ so that $f_x(p) = a$, and $f_x(x) = b + 1$. Let $G_x = \{y \in X : f_x(y) > b\}$, so that G_x is open. Then $x \in G_x$ for all $x \in F$, and $\{G_x\}_{x \in F}$ is an open cover of the compact set F. Choose $x_1, x_2, ..., x_N \in F$ so that $F \subseteq \bigcup_{n=1}^N G_{x_n}$.

Set

$$g = f_{x_1} \vee f_{x_2} \vee \cdots \vee f_{x_N}.$$

Clearly $g \in \mathcal{L}$ since $\mathcal{L}$ is a lattice. Moreover $g(p) = [f_{x_1} \vee f_{x_2} \vee \cdots \vee f_{x_N}](p) = a$, and if $x \in F$, then $x \in G_{x_n}$ for some $1 \le n \le N$, so that $g(x) \ge f_{x_n}(x) > b$.

Letting $f = g \vee a\mathbf{1}$ completes the proof.

□

2.14. Proposition. *Let (X, τ) be a compact, topological space. Suppose that $\mathcal{L} \subseteq \mathcal{C}(X, \mathbb{R})$ is both a vector space over $\mathbb{R}$ and a lattice (under the partial ordering of Example 2.7). Suppose furthermore that $\mathcal{L}$ separates the points of X and that $\mathcal{L}$ contains the constant functions.*

If $h \in \mathcal{C}(X, \mathbb{R})$ and $\varepsilon > 0$, then there exists $g \in \mathcal{L}$ such that for all $x \in X$ we have

$$0 \le g(x) - h(x) < \varepsilon.$$

In particular, $\mathcal{L}$ is dense in $(\mathcal{C}(X, \mathbb{R}), \|\cdot\|_\infty)$.

Proof. Let $h \in \mathcal{C}(X, \mathbb{R})$ and $\mathcal{L}_h = \{f \in \mathcal{L} : h \le f\}$. It is not hard to verify that $\mathcal{L}_h$ is again a lattice (of continuous, real-valued functions). Suppose that we can prove that for each $x \in X$,

$$h(x) = \inf\{f(x) : f \in \mathcal{L}_h\}.$$

Then, by applying Proposition 2.8, we see that for any $\varepsilon > 0$ there exists $g \in \mathcal{L}$ such that for all $x \in X$ we have

$$0 \le g(x) - h(x) < \varepsilon.$$

So there remains only to show that $h(x) = \inf\{f(x) : f \in \mathcal{L}_h\}$ for all $x \in X$.

Let $\delta > 0$ and $p \in X$. Since h is continuous and the set X is compact, it follows that $h(X) \subseteq \mathbb{R}$ is compact. But then $h(X)$ is closed and bounded, and thus it follows that h is bounded on X, say $|h(x)| \le \kappa$ for all $x \in X$.

Moreover, h is continuous at p, and thus there exists an open nbhd G_p of p such that $|h(y) - h(p)| < \delta$ for all $y \in G_p$. In particular, $h(y) < h(p) + \delta$ for all $y \in G_p$. Set $F_p = X \setminus G_p$, so that F_p is closed in X, hence compact.

By Lemma 2.13, with $a = h(p) + \delta$ and $b = \kappa$, we can find a function $f \in \mathcal{L}$ (the original lattice – not $\mathcal{L}_h$!) so that

- $f(x) \geq h(p) + \delta$ for all $x \in X$;
- $f(p) = h(p) + \delta$; and
- $f(x) > \kappa$ for all $x \in F_p$.

Note that since $|h(x)| \leq \kappa$ for all $x \in X$, it is clear that $h(x) \leq f(x)$ for $x \in F_p$. Since $x \in X \setminus F_p = G_p$ implies that $h(x) < h(p) + \delta$, it follows that $h \leq f$ on $X \setminus F_p$. Together, these imply that $h \leq f$ on X. Hence $f \in \mathcal{L}_h$ after all!

Thus $f \in \mathcal{L}_h$ and $f(p) = h(p) + \delta$. Since $\delta > 0$ was arbitrary,

$$h(p) = \inf\{f(p) : f \in \mathcal{L}_h\},$$

and we are done.

$\square$

Before starting in on the proof of the next Theorem, we remind the reader that if p and q are polynomials in a single variable x, then so is their composition $p \circ q(x) := p(q(x))$. This will be used implicitly below.

2.15. Theorem. (*Weierstraß's Approximation Theorem*)

Let $a < b \in \mathbb{R}$, and suppose that $f : [a,b] \to \mathbb{R}$ is a continuous function. Then there exists a sequence $(q_n)_n$ of polynomials in $\mathcal{C}([a,b],\mathbb{R})$ which converge uniformly to f on $[a,b]$, i.e.

$$\lim_n \|q_n - f\|_\infty = 0.$$

Proof.

STEP ONE. Observe that $f : [a,b] \to \mathbb{R}$ is continuous if and only if the function

$$\begin{array}{rccc} g: & [0,1] & \to & \mathbb{R} \\ & x & \mapsto & (f(a+(b-a)x) - f(a)) - x(f(b) - f(a)) \end{array}$$

is.

Indeed, the function $\alpha : [0,1] \to [a,b]$, $\alpha(x) = a + (b-a)x$ is continuous, being linear, as is $\beta : [0,1] \to \mathbb{R}$ defined by $\beta(x) = f(a) + x(f(b) - f(a))$. But then $g = (f \circ \alpha) - \beta$ is continuous as well.

Conversely, suppose that g is continuous. Then

$$f(x) = (g \circ \alpha^{-1}) + \beta \circ \alpha^{-1},$$

which is clearly continuous as g, β and α^{-1} are.

Suppose that we can approximate g uniformly by polynomials $(r_n)_n \in \mathcal{C}([0,1],\mathbb{R})$. An elementary calculation shows that

$$q_n = (r_n \circ \alpha^{-1}) + \beta \circ \alpha^{-1}$$

is a polynomial on $[a,b]$, and that $(q_n)_n$ converges to f uniformly.

STEP TWO. By Step One above, we have reduced the problem to the case where $[a,b] = [0,1]$, and $f(0) = 0 = f(1)$. We may then extend the domain of f to all of

$\mathbb{R}$ by setting $f(x) = 0$ if $f \notin [0,1]$. We denote this new, extended, function by f as well.

The basic idea is to define

$$Q_n(x) = c_n(1 - x^2)^n, \quad n \geq 1,$$

where the coefficient $c_n > 0$ is chosen so that

$$\int_{-1}^{1} Q_n(x)dx = 1 \qquad \text{for all } n \geq 1.$$

We wish to obtain an upper estimate on the size of c_n.

Note that if $\gamma_n(x) = (1 - x^2)^n - (1 - nx^2)$, $x \in [0,1]$, $n \geq 1$, then

- $\gamma_n(0) = 0$ and
- $\gamma_n'(x) = n(1 - x^2)^{(n-1)}2x + 2nx > 0$ for all $x \in (0,1)$.

Thus $(1 - x^2)^n \geq (1 - nx^2)$ for all $x \in [0,1]$, and so

$$\begin{aligned}\int_{-1}^{1}(1 - x^2)^n dx &= 2\int_0^1 (1 - x^2)^n dx \\ &\geq 2\int_0^{1/\sqrt{n}} (1 - x^2)^n dx \\ &\geq 2\int_0^{1/\sqrt{n}} (1 - nx^2) dx \\ &= \frac{4}{3\sqrt{n}} > \frac{1}{\sqrt{n}}.\end{aligned}$$

From this it follows that $c_n < \sqrt{n}$ for all $n \geq 1$. In particular, for any $\delta > 0$, we have

$$0 \leq Q_n(x) \leq \sqrt{n}(1 - \delta^2)^n, \quad \delta \leq |x| \leq 1.$$

Next, set

$$q_n(x) = \int_{-1}^{1} f(x + t)Q_n(t)dt, \qquad 0 \leq x \leq 1.$$

By a simple change of variable,

$$q_n(x) = \int_{-x}^{1-x} f(x + t)Q_n(t)dt = \int_0^1 f(t)Q_n(t - x)dt,$$

and this last integral is clearly a polynomial in x. Thus $(q_n)_n$ is a sequence of polynomials, whose coefficients are clearly real-valued if f is.

Given $\varepsilon > 0$, we choose $\delta > 0$ so that $|y - x| < \delta$ implies that

$$|f(y) - f(x)| < \frac{\varepsilon}{2}.$$

Let $\kappa = \sup\{|f(x)| : x \in [0,1]\}$. Recalling that $0 \le Q_n(x)$ for all $x \in [0,1]$, $n \ge 1$, we find that

$$\begin{aligned} |q_n(x) - f(x)| &= \left| \int_{-1}^{1} [f(x+t) - f(x)] Q_n(t) dt \right| \\ &\le \int_{-1}^{1} |f(x+t) - f(x)| Q_n(t) dt \\ &\le 2\kappa \int_{-1}^{-\delta} Q_n(t) dt + \frac{\varepsilon}{2} \int_{-\delta,\delta} Q_n(t) dt + 2\kappa \int_{\delta}^{1} Q_n(t) dt \\ &\le 4\kappa(\sqrt{n}(1-\delta^2)^n) + \frac{\varepsilon}{2} \\ &< \varepsilon \end{aligned}$$

when n is sufficiently large.

This completes the proof of the Theorem.

□

2.16. Remark. The construction of the sequence $(Q_n)_n$ above is not as strange as it might at first appear. It is actually an example of an **approximate identity** (or a **positive summability kernel**) for $\mathcal{C}([0,1],\mathbb{K})$ consisting of polynomials. In PMath 450, we shall examine in much greater detail such constructions.

2.17. Corollary. *For each $0 < a \in \mathbb{R}$, there exists a sequence $q_n \in \mathbb{R}[x]$ of real-valued polynomials with $q_n(0) = 0$ such that $(q_n)_n$ converges uniformly to the function $f(x) = |x|$ on the interval $[-a,a]$.*

Proof. Fix $a > 0$. Since f is clearly continuous on $[-a,a]$, using Weierstraß's Approximation Theorem above, we may find a sequence $(p_n)_n \in \mathbb{R}[x]$ so that $(p_n)_n$ converges uniformly to f on $[-a,a]$.

In particular, observe that $p_n(0)$ converges to $f(0) = 0$. If we set $q_n = p_n - (p_n(0))\mathbf{1}$ for each $n \ge 1$, then it is routine to check that $q_n(0) = 0$ for all $n \ge 1$ and that $(q_n)_n$ still converges uniformly to f.

□

The following result is Stone's sweeping generalization of Weierstraß's Approximation Theorem. It is an impressive and incredibly useful result, whose importance can not be over-stated, even if you try your hardest and even if you are known amongst your friends and family as having a special knack for over-stating things.

2.18. Theorem. (*The Stone-Weierstraß Theorem - real version*)
Let (X,τ) be a compact topological space and suppose that $\mathcal{A} \subseteq \mathcal{C}(X,\mathbb{R})$ is an algebra of continuous, real-valued functions on X which separates the points of X and contains the constant functions. Then $\mathcal{A}$ is dense in $(\mathcal{C}(X,\mathbb{R}), \|\cdot\|_\infty)$.

That is, given $f \in \mathcal{C}(X,\mathbb{R})$ and $\varepsilon > 0$, there exists $g \in \mathcal{A}$ so that

$$|g(x) - f(x)| < \varepsilon, \quad x \in X.$$

Proof. Let $\overline{\mathcal{A}}$ denote the norm-closure of $\mathcal{A}$ in $(\mathcal{C}(X,\mathbb{R}), \|\cdot\|_\infty)$. It is routine to verify that $\overline{\mathcal{A}}$ is an algebra. We claim that it is also a lattice.

To that end, let $f \in \overline{\mathcal{A}}$. Since $\overline{\mathcal{A}}$ is closed under scalar multiplication, we may suppose without loss of generality that $\|f\|_\infty \le 1$. Given $\varepsilon > 0$, by Corollary 2.17, there exists a polynomials $p \in \mathbb{R}[x]$ so that

$$\| |f| - p(f) \|_\infty < \varepsilon.$$

Since $\overline{\mathcal{A}}$ is an algebra, $p(f) \in \overline{\mathcal{A}}$, and since $\overline{\mathcal{A}}$ is closed, $|f| \in \overline{\mathcal{A}}$.

Then

$$f \vee g = \frac{f+g}{2} + \frac{|f-g|}{2} \in \overline{\mathcal{A}},$$

and similarly

$$f \wedge g = \frac{f+g}{2} - \frac{|f-g|}{2} \in \overline{\mathcal{A}}.$$

Hence $\overline{\mathcal{A}}$ is a lattice.

By Proposition 2.14, $\overline{\mathcal{A}} = \mathcal{C}(X,\mathbb{R})$.

□

2.19. Remarks.

- Some references will impose the condition that the space X occurring in the Stone-Weierstraß Theorem be Hausdorff. This, however, is already a consequence of the hypotheses. Indeed, using the notation of Theorem 2.18, note that if $x \ne y \in X$ then by hypothesis, there exists a continuous function $f \in \mathcal{A}$ such that $f(x) \ne f(y)$. But if $\delta := |f(x) - f(y)|/2 > 0$, then $B(f(x),\delta)$ and $B(f(y),\delta)$ are disjoint open sets in $\mathbb{R}$, and therefore $U = f^{-1}(B(f(x),\delta))$ and $V = f^{-1}(B(f(y),\delta))$ are disjoint open sets in X. Since $x \in U$ and $y \in V$, this shows that X is Hausdorff.
- The above theorem does not apply to complex algebras. A counterexample will be explored in the Assignments. However, if we add one extra condition, we do retrieve the conclusion.

2.20. Definition. *Let (X,τ) be a topological space and $\varnothing \ne \mathcal{S} \subseteq \mathcal{C}(X,\mathbb{K})$. We say that $\mathcal{S}$ is* ***self-adjoint*** *if $f \in \mathcal{S}$ implies that $f^* \in \mathcal{S}$, where*

$$f^*(x) = \overline{f(x)}, \quad x \in X$$

is the complex conjugate function of f.

2.21. Theorem. (*The Stone-Weierstraß Theorem - complex version*)

Let (X,τ) be a compact topological space and suppose that $\mathcal{A} \subseteq \mathcal{C}(X,\mathbb{C})$ is a ***self-adjoint*** *algebra of continuous, complex-valued functions on X which separates the points of X and contains the constant functions. Then $\mathcal{A}$ is dense in $(\mathcal{C}(X,\mathbb{C}), \|\cdot\|_\infty)$.*

Proof. Let $\mathcal{A}_\mathbb{R} = \{f \in \mathcal{A} : f(x) \in \mathbb{R} \text{ for all } x \in X\}$ denote the space of all real-valued functions in $\mathcal{A}$, viewed as a vector space over $\mathbb{R}$. In fact, $\mathcal{A}_\mathbb{R}$ is an algebra (over $\mathbb{R}$).

Indeed, $f, g \in \mathcal{A}_{\mathbb{R}}$ implies that $f, g \in \mathcal{A}$, and thus $fg \in \mathcal{A}$ since $\mathcal{A}$ is an algebra. But fg is real-valued, since each of f and g is, and therefore $fg \in \mathcal{A}_{\mathbb{R}}$.

If $f \in \mathcal{A}$, then we may write $f = u+iv$, where $u = \mathbf{Re}\, f := \frac{f+f^*}{2}$ and $v = \mathrm{Im}\, f = \frac{f-f^*}{2i}$ each lie in $\mathcal{A}_R$. Suppose that $x_1 \neq x_2 \in X$. By Lemma 2.12, there exists $f \in \mathcal{A}$ such that $f(x_1) = 1$ and $f(x_2) = 0$. Writing $f = u + iv$ as above, $0 = u(x_2) \neq u(x_1) = 1$, which shows that $\mathcal{A}_{\mathbb{R}}$ separates points of X. Clearly $\mathcal{A}_{\mathbb{R}}$ contains all (real-valued) constant functions, because $\mathcal{A}$ contains all complex-valued constant functions.

By the real version of the Stone-Weierstraß Theorem 2.18, $\overline{\mathcal{A}_{\mathbb{R}}} = \mathcal{C}(X, \mathbb{R})$. But if $g \in \mathcal{C}(X, \mathbb{C})$, then we may also write $g = \mathrm{Re}\, g + i \,\mathrm{Im}\, g$. Since each of these lies in $\overline{\mathcal{A}_{\mathbb{R}}} \subseteq \overline{\mathcal{A}}$, and since the latter is a complex algebra, $g \in \overline{\mathcal{A}}$, i.e.

$$\overline{\mathcal{A}} = \mathcal{C}(X, \mathbb{C}).$$

□

3. The Arzela-Ascoli Theorem

3.1. According to the Bolzano-Weierstrass Theorem, every bounded sequence of real (or complex) numbers admits a convergent subsequence. It is reasonable to ask to what extent such a result can be extended to bounded sequences of continuous functions from a topological space (X, τ_X) to a topological space (Y, τ_Y). To answer this, one must first decide specify the notion of convergence to which one is referring, and whether we are interested in "local" or "uniform" boundedness of our sequence.

3.2. Definition. *Let $\varnothing \neq E$ be a set, and suppose that $\mathcal{F}$ is a non-empty collection of $\mathbb{K}$-valued functions defined on E.*

*We say that $\mathcal{F}$ is **pointwise bounded** on E if we can find a function $\kappa : E \to [0, \infty)$ so that*

$$\sup_{f \in \mathcal{F}} |f(x)| \le \kappa(x) \text{ for all } x \in E.$$

*We say that $\mathcal{F}$ is **uniformly bounded** on E if there exists a constant $M > 0$ independent of x so that*

$$\sup_{f \in \mathcal{F}} |f(x)| \le M \text{ for all } x \in E.$$

3.3. Example. Fix an integer $N \ge 1$, and consider the set

$$\mathcal{E}_N := \{p \in \mathcal{C}(\mathbb{R}, \mathbb{K}) : p(x) = p_0 + p_1 x + p_2 x^2 + \cdots p_N x^N, p_n \in \mathbb{K}, |p_n| \le n, 0 \le n \le N\}.$$

For any $x \in \mathbb{R}$, $p \in \mathcal{E}_N$, observe that

$$\begin{aligned} |p(x)| &= |p_0 + p_1 x + p_2 x^2 + \cdots p_N x^N| \\ &\le |p_0| + |p_1||x| + |p_2||x^2| + \cdots |p_N||x^N| \\ &\le 0 + |x| + 2|x^2| + \cdots + N|x^N|, \end{aligned}$$

so that $\mathcal{E}_N$ is pointwise bounded on $\mathbb{R}$ (simply take $\kappa(x) = 0 + |x| + 2|x^2| + \cdots + N|x^N|$ in the definition above).

Since $q(x) = x$ lies in $\mathcal{E}_1 \subseteq \mathcal{E}_N$ for all $N \geq 1$, and since q itself is not a bounded function, we see that no $\mathcal{E}_N$ is uniformly bounded on $\mathbb{R}$.

Note, however, that if we consider $\mathcal{F}_N := \{p|_{[0,10]} : p \in \mathcal{E}_N\}$, then the above calculation shows that $\mathcal{F}_N$ is uniformly bounded by $M = 0 + 10 + 2(10^2) + \cdots N(10^N)$.

3.4. Remark. As we shall see below, if $(f_n)_n$ is a pointwise bounded sequence of functions on E, and if $E_1 \subseteq E$ is countable, then we can find a subsequence $(f_{n_k})_k$ of $(f_n)_n$ such that $(f_{n_k}(x))_k$ converges for each $x \in E_1$. However, even if $(f_n)_n$ is a *uniformly* bounded sequence of continuous functions on a compact set E, there need not be a subsequence which converges pointwise on E.

The proof is technical to write down, but the idea is not terribly complicated. By basic idea is that if a continuous function $h : [a, b] \to [-1, 1]$ satisfies $h(a) = 1 = h(b)$ and $h(c) = -1$ for some $c \in (a, b)$, then by choosing a regular partition of $[a, b]$ into sufficiently small subintervals, we can find a subinterval P where $h \geq \frac{1}{2}$ and an interval Q where $h \leq -\frac{1}{2}$. The issue for us is that we wish to produce an iterative procedure, and so we must be careful to clearly indicate how to choose these subintervals.

3.5. Lemma. *Given integers $0 \leq n$ and $0 \leq k \leq 10^n - 1$, define*

$$\alpha_{k,n} := \frac{k}{10^n},$$

and set

$$I_{k,n} := \left[\alpha_{k,n}, \alpha_{k+1,n}\right],$$

so that $I_{k,n} \subseteq [0, 1]$ is a closed interval of length $\frac{1}{10^n}$. Define $\gamma_{k,n} = \frac{\alpha_{k,n}+\alpha_{k+1,n}}{2}$ to be the midpoint of the interval $I_{k,n}$.

Next, suppose that $h : I_{k,n} \to [-1, 1]$ is a function which is linear on $\left[\alpha_{k,n}, \gamma_{k,n}\right]$ and on $\left[\gamma_{k,n}, \alpha_{k+1,n}\right]$, with $h(\alpha_{k,n}) = 1 = h(\alpha_{k+1,n})$, and $h(\gamma_{k,n}) = -1$.

Let $m > n$ be an integer. Then there exist integers $0 \leq p, q \leq 10^m - 1$ and intervals

$$P(I_{k,n}) := \left[\frac{p}{10^m}, \frac{p+1}{10^m}\right]$$

and

$$Q(I_{k,n}) := \left[\frac{q}{10^m}, \frac{q+1}{10^m}\right]$$

satisfying

(i) $P(I_{k,n}) \subseteq I_{k,n}$ *and* $\operatorname{diam} P(I_{k,n}) = \frac{1}{10^m}$.
(ii) $h(x) \geq \frac{1}{2}$ *for all* $x \in P(I_{k,n})$.
(iii) $Q(I_{k,n}) \subseteq I_{k,n}$ *and* $\operatorname{diam} Q(I_{k,n}) = \frac{1}{10^m}$.
(iv) $h(x) \leq -\frac{1}{2}$ *for all* $x \in Q(I_{k,n})$.

Proof. As we shall see, the proof of the result is not much longer than its statement.

Set $p = 10^{m-n}k$, let $P(I_{k,n}) = \left[\frac{p}{10^m}, \frac{p+1}{10^m}\right]$, and note that $\operatorname{diam} P(I_{k,n}) = \frac{1}{10^m}$. Observe that

$$\frac{p}{10^m} = \frac{k}{10^n} = \alpha_{k,n}.$$

Thus $h(\frac{p}{10^m}) = 1$. For $x \in [\alpha_{k,n}, \gamma_{k,n}]$, we have that $h(x) = 1 - 4(10^n(x - \alpha_{k,n}))$. But

$$\frac{p+1}{10^m} - \frac{p}{10^m} = \frac{1}{10^m} \le \frac{1}{10}\frac{1}{10^n} < \frac{1}{2}\frac{1}{10^n} = \gamma_{k,n} - \alpha_{k,n},$$

and so for all $x \in P(I_{k,n}) \subseteq [\alpha_{k,n}, \gamma_{k,n}] \subseteq I_{k,n}$ we have that

$$h(x) = 1 - 4(10^n(x - \alpha_{k,n})) = 1 - 4(10^n(x - \frac{p}{10^m})) \ge 1 - 4(10^n(\frac{1}{10^m})) \ge 1 - \frac{4}{10} \ge \frac{1}{2}.$$

This proves (i) and (ii).

As for $Q(I_{k,n})$, we choose q so that

$$\frac{q}{10^m} = \gamma_{k,n} = \frac{2k+1}{2 \cdot 10^n},$$

and set $Q(I_{k,n}) = \left[\frac{q}{10^m}, \frac{q+1}{10^m}\right]$. Again, it is clear that $\operatorname{diam} Q(I_{k,n}) = \frac{1}{10^m}$.

For $x \in [\gamma_{k,n}, \alpha_{k+1,n}]$, we note that $h(x) = -1 + 4(10^n(x - \gamma_{k,n}))$, and that

$$\frac{q+1}{10^m} - \frac{q}{10^m} = \frac{1}{10^m} \le \frac{1}{10}\frac{1}{10^n} < \frac{1}{2}\frac{1}{10^n} = \alpha_{k+1,n} - \gamma_{k,n}.$$

Thus, in a similar fashion to that above, $x \in Q(I_{k,n}) \subseteq [\gamma_{k,n}, \alpha_{k+1,n}] \subseteq I_{k,n}$ implies that

$$h(x) = -1 + 4(10^n(x - \gamma_{k,n})) = -1 + 4(10^n(x - \frac{q}{10^m})) \le -1 + 4\frac{10^n}{10^m} \le -1 + \frac{4}{10} < -\frac{1}{2}.$$

This completes the proof.

□

3.6. Example. There exists a uniformly bounded sequence $(f_n)_{n=1}^\infty$ of continuous functions on $[0,1]$ with the property that if $(f_{n_r})_{r=1}^\infty$ is *any* subsequence of $(f_n)_{n=1}^\infty$, then there exists a point $y \in [0,1]$ such that $(f_{n_r}(y))_{r=1}^\infty$ does not converge. In other words, no subsequence of $(f_n)_{n=1}^\infty$ converges pointwise on $[0,1]$.

Let $g_0 : [0,1] \to [-1,1]$ be the function defined by

$$g_0(x) := \begin{cases} 1 - 4x & \text{if } x \in [0, \frac{1}{2}] \\ -1 + 4x & \text{if } x \in [\frac{1}{2}, 1], \end{cases}$$

and let $g : \mathbb{R} \to [-1,1]$ be the 1-periodic extension of g_0 defined by

$$g(x + k) = g_0(x) \qquad \text{for all } x \in [0,1], k \in \mathbb{Z}.$$

For $n \ge 0$, define $f_n(x) = g(10^n x)$, $x \in [0,1]$, and observe that on any interval $I_{k,n} = \left[\frac{k}{10^n}, \frac{k+1}{10^n}\right]$, $1 \le n$, $0 \le k \le 10^n - 1$, the graph of f_n coincides with the graph of the function h defined in the above Lemma.

Now let $(f_{n_r})_{r=1}^\infty$ be a subsequence of $(f_n)_{n=1}^\infty$, and set

$$J_0 := I_{0,n_1} = \left[0, \frac{1}{10^{n_1}}\right].$$

Note that $n_2 > n_1$, and so by Lemma 3.5, there exists an interval $J_1 := P(J_0) = P(I_{0,n_1}) \subseteq J_0$ such that $f_{n_1}(x) \ge \frac{1}{2}$ for all $x \in J_1$. Moreover, $J_1 = I_{k_1,n_2}$ for some $0 \le k_1 \le 10^{n_2} - 1$ and $\operatorname{diam} J_1 = \dfrac{1}{10^{n_2}}$.

Noting that $n_3 > n_2$ and applying Lemma 3.5 once again, we can find an interval $J_2 := Q(J_1) \subseteq J_1$ such that $f_{n_2}(x) \le -\frac{1}{2}$ for all $x \in J_2$.

Having chosen $J_1 \supseteq J_2 \supseteq J_3 \supseteq \cdots \supseteq J_r$ as above, by applying Lemma 3.5 with $n_{r+1} > n_r$, we can choose an interval

$$J_{r+1} = \begin{cases} P(J_r) \subseteq J_r & \text{if } r \text{ is even,} \\ Q(J_r) \subseteq J_r & \text{if } r \text{ is odd,} \end{cases}$$

such that $f_{n_{r+1}}(x) \ge \frac{1}{2}$ if $x \in J_{r+1}$ and r is even, while $f_{n_{r+1}}(x) \le -\frac{1}{2}$ if $x \in J_{r+1}$ and r is odd. Furthermore, $\operatorname{diam} J_{r+1} = \dfrac{1}{10^{n_{r+2}}}$.

Since

$$[0,1] \supseteq J_0 \supseteq J_1 \supseteq J_2 \supseteq J_3 \supseteq \cdots$$

is a nested sequence of closed intervals whose diameters are converging to zero, we may apply the Nested Interval Theorem from Calculus to conclude that

$$\cap_{r=1}^\infty J_r = \{y\},$$

a singleton point. Of course, $y \in [0,1]$.

Now, $y \in J_r$ for all $r \ge 1$, so $f_{n_r}(y) \ge \frac{1}{2}$ if r is odd, while $f_{n_r}(y) \le -\frac{1}{2}$ if r is odd. Obviously, $(f_{n_r}(y))_{r=1}^\infty$ will not converge.

3.7. A second question which arises is whether every pointwise convergent sequence of functions on an interval $E = [a,b]$ necessarily contains a uniformly convergent subsequence. Our next example shows that this is not true, even if the original sequence is uniformly bounded.

3.8. Example. Let $E = [0,1]$, and for $n \ge 1$, define $f_n : E \to \mathbb{R}$ via

$$f_n(x) = \frac{x^2}{x^2 + (1-nx)^2}.$$

Then $\|f_n\|_\infty \le 1$ for all $n \ge 1$, and thus $(f_n)_n$ is uniformly bounded on E. Also, if $x \in E$, then $\lim_n f_n(x) = 0$.

Nevertheless,

$$f_n(\frac{1}{n}) = \frac{\frac{1}{n^2}}{\frac{1}{n^2} + 0} = 1, \quad n \ge 1,$$

so that no subsequence can converge uniformly on E.

The concept we shall need is the following:

3.9. Definition. *Let (X,τ) be a topological space and (Y,d) be a metric space. A family $\mathcal{F}$ of functions from X to Y is said to be **equicontinuous** at the point $x \in X$ if, given $\varepsilon > 0$ there is an open neighbourhood $G \in \mathcal{U}_x$ such that $y \in G$ implies*

$$d(f(x), f(y)) < \varepsilon \text{ for all } f \in \mathcal{F}.$$

*We say that $\mathcal{F}$ is **equicontinuous on** X if it is equicontinuous at x for all $x \in X$.*

3.10. Remark. If (X, d_X) and (Y, d_Y) are metric spaces and $x_0 \in X$, then a family $\mathcal{F}$ of functions from X to Y is equicontinuous at x_0 if for each $\varepsilon > 0$ there exists $\delta > 0$ so that $d_X(x, x_0) < \delta$ implies that

$$d_Y(f(x), f(x_0)) < \varepsilon \quad \text{for all } f \in \mathcal{F}.$$

In other words, not only is each function $f \in \mathcal{F}$ continuous at x_0, but given $\varepsilon > 0$, *the same* $\delta > 0$ works *for all* $f \in \mathcal{F}$ simultaneously.

Suppose now that (X, d_X) is compact, and that $\mathcal{F}$ is equicontinuous on X. In this case, the compactness of X allows us to conclude something much stronger. Indeed, let $\varepsilon > 0$ as above, and for each $x \in X$, use the equicontinuity of $\mathcal{F}$ at x to find $\delta_x > 0$ so that $d_X(x, z) < \delta_x$ implies that $d_Y(f(x), f(z)) < \frac{\varepsilon}{2}$ for all $f \in \mathcal{F}$. Note that $\{B(x, \frac{\delta_x}{2})\}_{x \in X}$ is an open cover of X, and so by compactness of X, there exist $x_1, x_2, \ldots, x_N \in X$ so that

$$X = \bigcup_{n=1}^{N} B(x_n, \frac{\delta_{x_n}}{2}).$$

Let $\delta := \frac{1}{2}\min(\delta_{x_1}, \delta_{x_2}, \ldots, \delta_{x_N}) > 0$. Suppose that $x, z \in X$ and that $d_X(x, z) < \delta$.

Since $X = \bigcup_{n=1}^{N} B(x_n, \frac{\delta_{x_n}}{2})$, there exists $1 \le m \le N$ so that $x \in B(x_m, \frac{\delta_{x_m}}{2})$. Thus $d(x, x_m) < \delta_{x_m}$ and so $d_Y(f(x), f(x_m)) < \frac{\varepsilon}{2}$ for all $f \in \mathcal{F}$.

Since $d_X(x, z) < \delta \le \frac{\delta_{x_m}}{2}$, we see that

$$d_X(x_m, z) < d_X(x_m, x) + d_X(x, z) < \frac{\delta_{x_m}}{2} + \delta \le \delta_{x_m},$$

and therefore $d_Y(f(x_m), f(z)) < \frac{\varepsilon}{2}$ for all $f \in \mathcal{F}$, again by the equicontinuity of $\mathcal{F}$ at x_m.

Together, these imply that

$$d_Y(f(x), f(z)) \le d_Y(f(x), f(x_m)) + d_Y(f(x_m), f(z)) < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon.$$

That is, not only is each f *uniformly continuous* on X, but indeed, *the same* δ works for all $f \in \mathcal{F}$ *simultaneously*!

Our next goal is to show that if $(f_n)_n$ is sequence in an equicontinuous family $\mathcal{F}$ of functions from a topological space (X,τ) to a metric space (Y,d), and if at each $x \in X$ there exists a convergent subsequence $(f_{n_k}(x))_k$ (where $(n_k)_k$ *a priori* depends upon x), then there exists a subsequence of $(f_n)_n$ which converges uniformly on each compact subset of X.

3.11. Lemma. *Let D be a countable set, and let $(f_n)_n$ be a sequence of functions from D into a topological space (Y, τ_Y). Suppose that for each $d \in D$, then closure of $\{f_n(d)\}_n$ is sequentially compact. Then there exists a subsequence $(f_{n_k})_k$ of $(f_n)_n$ which converges for each $d \in D$.*

Proof. We leave the case where D is finite as an exercise for the reader.

Let $D = \{d_n\}_{n=1}^\infty$. By the sequential compactness of $\overline{\{f_n(d_1)\}_n}$, we can pick a subsequence $(f_{1,n})_{n=1}^\infty$ of $(f_n)_n$ such that the sequence $(f_{1,n}(d_1))_{n=1}^\infty$ converges.

Next, choose a subsequence $(f_{2,n})_{n=1}^\infty$ of $(f_{1,n})_{n=1}^\infty$ so that $(f_{2,n}(d_2))_{n=1}^\infty$ converges. Observe that $(f_{2,n}(d_1))_{n=1}^\infty$ still converges, since $(f_{2,n})_{n=1}^\infty$ is a subsequence of $(f_{1,n})_{n=1}^\infty$.

Continuing in this manner, for each $k \ge 1$, we obtain a subsequence $(f_{k,n})_{n=1}^\infty$ of $(f_{k-1,n})_{n=1}^\infty$ so that $(f_{k,n}(d_j))_{n=1}^\infty$ converges for all $1 \le j \le k$.

Consider the *diagonal* sequence $(f_{n,n})_{n=1}^\infty$. Since $(f_{n,n})_n$ is a subsequence of $(f_{k,n})_n$ for each $1 \le k \le n$, it follows that $(f_{n,n}(d_j))_n$ converges for all $j \ge 1$.

□

3.12. Lemma. *Let $(f_n)_n$ be an equicontinuous sequence of functions from a topological space (X, τ) to a complete metric space (Y, d). If $D \subseteq X$ is dense and if $(f_n(d))_n$ converges in Y for each $d \in D$, then $(f_n(x))_n$ converges for each $x \in X$. Furthermore, the map*

$$\begin{array}{rccc} f: & X & \to & Y \\ & x & \mapsto & \lim_n f_n(x) \end{array}$$

is continuous on X.

Proof. Let $x \in X$ and $\varepsilon > 0$. By equicontinuity, we can find an open set $G \in \mathcal{U}_x$ so that $z \in G$ implies that

$$d(f_n(x), f_n(z)) < \frac{\varepsilon}{3}$$

for all $n \ge 1$. Since D is dense in X, we can choose $d_0 \in D \cap G$. Since $(f_n(d_0))_n$ then converges in Y by hypothesis, it must be a Cauchy sequence. Choose $N > 0$ so that $m, n \ge N$ implies that

$$d(f_n(d_0), f_m(d_0)) < \frac{\varepsilon}{3}.$$

Then, for $m, n \ge N$ we have

$$\begin{aligned} d(f_n(x), f_m(x)) &\le d(f_n(x), f_n(d_0)) + d(f_n(d_0), f_m(d_0)) + d(f_m(d_0), f_m(x)) \\ &< \frac{\varepsilon}{3} + \frac{\varepsilon}{3} + \frac{\varepsilon}{3} \\ &= \varepsilon. \end{aligned}$$

Thus $(f_n(x))_n$ is Cauchy for all $x \in X$.

But (Y, d) was assumed to be complete, and thus we may define

$$f(x) = \lim_n f_n(x), \quad x \in X.$$

With $x \in X$, $\varepsilon > 0$ and G as in the first line of the proof, observe that for all $z \in G$,

$$d(f(x), f(z)) = \lim_n d(f_n(x), f_n(z)) \le \frac{\varepsilon}{3} < \varepsilon,$$

and so f is continuous at x.

□

3.13. Lemma. *Let (K, τ) be a compact topological space and $(f_n)_n$ be an equicontinuous sequence of functions from K to a metric space (Y, d). Suppose that $(f_n)_n$ converges pointwise to some function $f : K \to Y$. Then $(f_n)_n$ converges uniformly to f.*

Proof. Let $\varepsilon > 0$. By equicontinuity, for each $x \in K$ we can find an open neighbourhood $G_x \in \mathcal{U}_x$ so that

$$d(f_n(x), f_n(y)) < \frac{\varepsilon}{3}, \quad y \in G_x, n \geq 1.$$

Thus

$$d(f(x), f(y)) \leq \frac{\varepsilon}{3}, \quad y \in G_x.$$

Since K is compact and $\bigcup_{x \in K} G_x$ is an open cover of K, we can find $x_1, x_2, ..., x_N \in K$ so that $K \subseteq \bigcup_{n=1}^N G_{x_n}$. Fix $M > 0$ so that $n \geq M$ implies that

$$d(f_n(x_k), f(x_k)) < \frac{\varepsilon}{3}, \quad 1 \leq k \leq N.$$

Given $y \in K$, we can find $1 \leq k < N$ so that $y \in G_{x_k}$. Thus $n \geq M$ implies that

$$\begin{aligned} d(f_n(y), f(y)) &\leq d(f_n(y), f_n(x_k)) + d(f_n(x_k), f(x_k)) + d(f(x_k), f(y)) \\ &< \frac{\varepsilon}{3} + \frac{\varepsilon}{3} + \frac{\varepsilon}{3} = \varepsilon. \end{aligned}$$

Thus $(f_n)_n$ converges uniformly to f on K.

□

Taken together, these last three Lemmas imply:

3.14. Theorem. (*Ascoli's Theorem*) *Let (X, τ) be a separable topological space and (Y, d) be a complete metric space. Let $\mathcal{F}$ be an equicontinuous family of functions from X into Y. Suppose that $(f_n)_n$ is a sequence in $\mathcal{F}$ such that for each $\overline{\{f_n(x)\}_n}$ is compact.*

Then there exists a subsequence $(f_{n_k})_k$ of $(f_n)_n$ which converges pointwise to a continuous function f. Moreover, the convergence is uniform on each compact subset of X.

Proof. Since (X, τ) is separable, there exists a countable dense subset D. By Lemma 3.11, there exists a subsequence $(f_{n_k})_k$ of $(f_n)_n$ which converges at each $x \in D$.

By Lemma 3.12, $(f_{n_k})_k$ converges pointwise at each $x \in X$ to a continuous function $f : X \to Y$.

By Lemma 3.13, if $K \subseteq X$ is compact, then $(f_{n_k})_k$ converges uniformly to f on K.

□

3.15. Corollary. *Let (X, τ) be a separable topological space, and let $\mathcal{F}$ be an equicontinuous family of $\mathbb{K}$-valued functions on X. Suppose that $(f_n)_n$ is a sequence on X which is pointwise bounded on a dense subset D of X. Then $(f_n)_n$ admits a subsequence $(f_{n_k})_k$ which converges pointwise to a continuous function f, and the convergence is uniform on each compact subset of X.*

The next result is perhaps the most important consequence of Ascoli's Theorem.

3.16. Theorem. *Let $\mathcal{A}$ be a norm-closed subset of $\mathcal{C}([0,1],\mathbb{R})$. The following statements are equivalent.*

(a) *$\mathcal{A}$ is compact.*
(b) *$\mathcal{A}$ is sequentially compact.*
(c) *$\mathcal{A}$ is uniformly bounded and equicontinuous.*

Proof.

(a) if and only if (b).
Since $(\mathcal{C}([0,1],\mathbb{R}), \|\cdot\|_\infty)$ is a metric space (using the metric induced by the norm), compactness and sequential compactness are equivalent, by Theorem 3.2.37.

(b) implies (c).
Since $\mathcal{A}$ is a sequentially compact subset of a metric space, $\mathcal{A}$ is totally bounded by Proposition 3.2.33, hence bounded by Proposition 3.2.31. That is, $\mathcal{A}$ is uniformly bounded as a set of functions on $[0,1]$. There remains to show that $\mathcal{A}$ is equicontinuous.

Let $\varepsilon > 0$. Since $\mathcal{A}$ is compact, it admits a finite $\frac{\varepsilon}{3}$-net, say $B = \{f_1, f_2, ..., f_N\}$. Then, for any $f \in \mathcal{A}$, there exists $f_m \in B$ so that

$$\|f - f_m\|_\infty \leq \frac{\varepsilon}{3}.$$

Thus, for any $x, y \in [0,1]$,

$$\begin{aligned}|f(x) - f(y)| &\leq |f(x) - f_m(x)| + |f_m(x) - f_m(y)| + |f_m(y) - f(y)| \\ &\leq |f_m(x) - f_m(y)| + \frac{2\varepsilon}{3}.\end{aligned}$$

Since each f_n is uniformly continuous on $[0,1]$, $1 \leq n \leq N$, we can find $\delta_n > 0$ so that $|x-y| < \delta_n$ implies that $|f_n(x)-f_n(y)| < \frac{\varepsilon}{3}$. Let $\delta = \min(\delta_1, \delta_2, ..., \delta_N)$. Thus for each $f \in \mathcal{A}$, $|x - y| < \delta$ implies that

$$\begin{aligned}|f(x) - f(y)| &\leq |f_m(x) - f_m(y)| + \frac{2\varepsilon}{3} \\ &< \frac{\varepsilon}{3} + \frac{2\varepsilon}{3} = \varepsilon.\end{aligned}$$

Thus $\mathcal{A}$ is equicontinuous.

(c) implies (a).

Suppose $\mathcal{A}$ is uniformly bounded and equicontinuous. Since $\mathcal{A}$ is closed and $\mathcal{C}([0,1],\mathbb{R})$ is complete, we need only show that $\mathcal{A}$ is totally bounded (see Theorem 4.3.4).

Step One. Let $f:[0,1]\to\mathbb{R}$ be continuous and $\varepsilon>0$. Then there exists $n_0\in\mathbb{N}$ and points

$$p_i=(\frac{i}{n_0},\frac{\varepsilon k_i}{5}),\quad 0\le i\le n_0,$$

where k_i are integers chosen such that, if g is the piecewise linear curve connecting p_{i-1} to p_i, $1\le i\le n_0$, then $\|f-g\|_\infty<\varepsilon$. In other words, the piecewise linear functions are dense in $(\mathcal{C}([0,1],\mathbb{R}),\|\cdot\|_\infty)$.

Indeed, since f is uniformly continuous on $[0,1]$, there exists $n_0\in\mathbb{N}$ so that $|x-y|\le\frac{1}{n_0}$ implies that $|f(x)-f(y)|<\frac{\varepsilon}{5}$.

Let

$$\mathbb{L}=\{(x,y):x=\frac{i}{n_0},y=\frac{k\varepsilon}{5},\ 0\le i\le n_0,k\in\mathbb{Z}\}.$$

Choose $p_i=(x_i,y_i)\in\mathbb{L}$ such that

$$y_i\le f(x_i)<y_i+\frac{\varepsilon}{5}.$$

If g is linear from p_i to p_{i+1}, then

$$|f(x_i)-g(x_i)|=|f(x_i)-y_i|<\frac{\varepsilon}{5},$$

and $|f(x_i)-f(x_{i+1})|<\frac{\varepsilon}{5}$.

Now

$$\begin{aligned}|g(x_i)-g(x_{i+1})|&\le|g(x_i)-f(x_i)|+|f(x_i)-f_{i+1})|+|f(x_{i+1})-g(x_{i+1})|\\&<\frac{3\varepsilon}{5}.\end{aligned}$$

Given $0\le z\le1$, choose j so that $x_j\le z<x_{j+1}$. Then

$$\begin{aligned}|f(z)-g(z)|&\le|f(z)-f(x_j)|+|f(x_j)-g(x_j)|+|g(x_j)-g(z)|\\&\le\frac{\varepsilon}{5}+\frac{\varepsilon}{5}+\frac{3\varepsilon}{5}\\&=\varepsilon.\end{aligned}$$

Thus $\|f-g\|_\infty<\varepsilon$.

Step Two. We now argue that the uniform boundedness of the family $\mathcal{A}$ implies the existence of a finite ε-net of piecewise linear curves.

Let $\varepsilon>0$. Since $\mathcal{A}$ is equicontinuous, there exists $n_0\in\mathbb{N}$ so that

$$|x-y|<\frac{1}{n_0}\qquad\text{implies that}\qquad|f(x)-f(y)|<\frac{\varepsilon}{5},\ f\in\mathcal{A}.$$

For each $f\in\mathcal{A}$, by Step One, we can find a piecewise linear function p_f so that $\|f-p_f\|_\infty<\varepsilon$ and p_f connects points in the set $\mathbb{L}$ defined above.

We claim that the set $\mathcal{B} = \{p_f : f \in \mathcal{A}\}$ is finite, and is this a finite ε-net for $\mathcal{A}$, proving that $\mathcal{A}$ is indeed totally bounded.

To see this, note that $\mathcal{A}$ is uniformly bounded, and hence $\mathcal{B}$ is uniformly bounded. Thus only a fixed, finite number of the points in $\mathbb{L}$ will appear as interpolating points of the polygonal curves in $\mathcal{B}$. Thus $\mathcal{B}$ is finite, proving that $\mathcal{A}$ is totally bounded, hence compact.

□

4. Appendix

4.1. In Example 3.6, we struggled mightily to bring you an example of a *uniformly* bounded sequence of continuous functions on the compact set $[0,1]$ for which no subsequence of the sequence converged pointwise on $[0,1]$. For those of you who have dabbled in the dark arts of measure theory, the *Lebesgue Dominated Convergence Theorem* provides a much shorter proof. Of course, it is only shorter because we will not give a proof of the Lebesgue Dominated Convergence Theorem.

We now state the version of this Theorem which we will employ. The general version is *much* stronger.

4.2. Theorem. (***The Lebesgue Dominated Convergence Theorem***)
Let $a < b \in \mathbb{R}$ and $E = [a,b]$. Suppose that $(f_n)_n$ is a uniformly bounded sequence of continuous functions on E which converges pointwise to 0. Then

$$\lim_n \int_a^b f_n(x)dx = 0.$$

4.3. Example. Let $E = [0,2\pi]$, and for $n \geq 1$, set $f_n(x) = \sin nx$. Clearly the sequence $(f_n)_n$ is uniformly bounded (by $M = 1$), and each f_n is continuous on E.

Suppose that we can find a subsequence $(f_{n_k})_k$ of $(f_n)_n$ which converges pointwise on E. Set $g_k = (f_{n_k} - f_{n_{k+1}})^2$, $k \geq 1$. Then $(g_k)_k$ is also uniformly bounded (by 4), each g_k is continuous, $k \geq 1$, and by hypothesis, $(g_k)_k$ converges pointwise to 0.

A routine calculation shows that for $n \neq m \geq 1$,

$$\int_0^1 (\sin nx - \sin mx)^2 dx = 2\pi,$$

which clearly implies that

$$\int_0^1 g_k(x)dx = 2\pi \quad \text{for all } k \geq 1.$$

This contradicts the Lebesgue Dominated Convergence Theorem, and thus no such subsequence can exist.

Bibliography

[1] R. André. *Axioms and set theory: a first course in set theory*. 2014.
[2] P.R. Halmos. *Naive set theory*. The University Series in Undergraduate Mathematics. D. van Nostrand Co., Princeton, N.J.-Toronto-London-New York, 1960.
[3] P.R. Halmos. *General Topology*. Addison–Wesley Publishing Co., Reading, Mass.–London–Don Mills, Ont., 1970.
[4] W. Rudin. *Principles of mathematical analysis*. International Series in Pure and Applied Mathematics. McGraw-Hill Book Co., New York-Auckland-Düsseldorf, Third edition, 1976.