Modern Applied Analysis. Part A.

(Draft version of December 22, 1997)

John Hunter and Bruno Nachtergaele

Department of Mathematics
University of California, Davis

Errata to *Modern Applied Analysis, Part A*

John Hunter and Bruno Nachtergaele

Erratum 1. The conclusion of the proof of Theorem 2.11 should read as follows:

We conclude the proof by deriving a contradiction. Indeed, starting from (2.6) we have

$$\epsilon < d(f(x_n), f(y_n)) \leq d(f(x_n), f(x)) + d(f(x), f(y_n))$$

As d and f are both continuous functions, and $x_n \to x$ and $y_n \to x$ the RHS tends to 0, which is a contradiction.

Erratum 2. The first paragraph of the proof of Theorem 2.16 should be:

We will first show that any open cover $\{G_\alpha\}$ of K contains a countable subcover. From Corollary 2.15, we know that K contains a countable dense subset A. Consider the family $\mathcal{B}$ of open balls with center in A and radius in $\mathbb{Q}$. Then, $\mathcal{B}$ is a countable family. Consider the subset $\mathcal{C}$ of $\mathcal{B}$ consisting of all $B \in \mathcal{B}$ that are contained in one of the G_α. As a subset of a countable set, $\mathcal{C}$ is still countable. For every $B \in C$, pick α_B such that $B \subset G_\alpha$. Then, $\{G_{\alpha_B}\}_{B\in\mathcal{C}}$ is a countable subset of $\{G_\alpha\}$, and that it is a cover of K can be seen as follows. For every $x \in K$, there is a member of the open cover G_α, such that $x \in G_\alpha$. As G_α is open, there is an open ball $B_\epsilon(x) \subset G_\alpha$, and we can assume that ϵ is rational. As A is dense, there exists $y \in A$ such that $d(x, y) < \epsilon/2$. Then $x \in B_{\epsilon/2}(y) \subset G_\alpha$, and, hence, $B_{\epsilon/2}(y) \in \mathcal{C}$, and it follows that $\mathcal{C}$ is a cover. By construction we have

$$\bigcup_{B\in\mathcal{C}} B \subset \bigcup_{B\in\mathcal{C}} G_{\alpha_B}$$

Therefore, $\{G_{\alpha_B}\}_{B\in\mathcal{C}}$ is a countable subcover of $\{G_\alpha\}$.

Prolegomena

These course notes very closely represent the material covered in class and the MAT203 exams will test your mastering of the math in these notes. Of course, this by no means implies that reading them over a couple of times will suffice to obtain a good grade or even pass the course.

You will find numerous exercises throughout these notes. They are usually based on concepts or results that have been introduced right there in the text preceding the exercise. Keep in mind, however, that their solution often requires some earlier material as well. We will discuss some of the exercises in class. Others will show up on a homework (or, who knows, on an exam!). You can also bring them up in the discussion session. In any case, it is important that you work through the exercises.

The statements of the exercises and the examples given in class, in the notes, and in homework assignments, form integral part of the course material.

At the beginning of each chapter you will find at least one reference for further reading.

We wish to thank all students, Sergio Lucero in particular, who reported typos and other goofs in previous versions of these notes. All remaining errors, however, are the sole responsibility of the authors. Of course, you are more than welcome to bring them to our attention.

Contents

1 The real numbers as a metric space **6**
1.1 Definition of Metric . . . 6
1.2 When distances tend to zero: convergence . . . 8
1.2.1 Sequences . . . 8
1.2.2 Upper and lower bound, inf and sup, lim inf and lim sup . . . 9
1.3 Series . . . 11
1.4 Open and closed sets . . . 12
1.5 Countably and uncountably infinite sets . . . 13
1.6 Compactness . . . 13

2 Continuous functions **16**
2.1 Continuity and semi-continuity . . . 16
2.2 Convergence of sequences of functions . . . 19
2.3 The spaces $C(K)$ . . . 21
2.4 Compact subsets of $C(K)$ . . . 26
2.5 Existence of local solutions of ODE's . . . 33

3 Topological Spaces **38**
3.1 General definition of a topological space . . . 38
3.2 Bases of open sets . . . 41
3.3 Comparing topologies . . . 43
3.4 Comments on non-metric topologies . . . 45

4 More on metric spaces: completeness; the contraction mapping theorem **47**
4.1 Completion of metric spaces . . . 48
4.2 The Contraction Mapping Theorem . . . 51

5 Applications of the Contraction Mapping Theorem **53**
5.1 An algorithm to compute $\sqrt{x}$ . . . 53
5.2 Discrete Dynamical Systems . . . 54
5.3 Integral Equations . . . 55
5.4 Differential Equations (Boundary Value Problems) . . . 57
5.5 Differential Equations (Existence and Uniqueness Theorem) . . . 59

6 Banach Spaces **63**
6.1 Bounded linear maps . . . 63
6.2 Topologies on $\mathcal{B}(X, Y)$ . . . 66
6.3 Remarks on Finite-Dimensional Banach Spaces . . . 72

7 Hilbert Spaces **76**
7.1 Inner Products 76
7.2 Hilbert Spaces 79

1 The real numbers as a metric space

In this section we recall some basic properties of the real numbers and real-valued functions. The familiar notions of limit, convergence, and continuity are expressed in terms of the metric or distance function. This allows us to define the above mentioned notions in the more general context of metric spaces, which we will need later on.

Good references for this and several of the following sections are
G.F. Simmons, *Introduction to Topology and Modern Analysis*, McGraw-Hill, New York, 1963.
J.E. Marsden, *Elementary Classical Analysis*, 2nd ed., W.H. Freeman and Company, New York, 1993.

The majority of the material of Sections 1 and 2 is also covered in
W. Rudin, *Principles of Mathematical Analysis*, 2nd ed. McGraw-Hill, New York, 1964

1.1 Definition of Metric

On a first reading, substitute $\mathbb{R}$ for the arbitrary set X in the following definition and subsequent discussion. This week we will discuss almost exclusively $\mathbb{R}$ and $\mathbb{R}^n$ (Euclidean space), but more general (and usually infinite-dimensional) metric spaces will follow soon.

The following is the definition of *metric*, also called *distance function*, on an arbitrary set X.

Definition 1.1 (metric) *A metric on X is a function $d : X \times X \to \mathbb{R}$, with the following properties:*
(M1) $d(x, y) \geq 0$ for all $x, y \in X$, and $d(x, x) = 0$.
(M2) If $d(x, y) = 0$, then $x = y$.
(M3) $d(x, y) = d(y, x)$, for all $x, y \in X$.
(M4) $d(x, y) \leq d(x, z) + d(z, y)$, for all $x, y, z \in X$.

In English the four defining properties read:
M1: Distances are non-negative, and the distance between a point and itself vanishes
M2: The only point at distance 0 from x is x itself.
M3: The distance is a *symmetric* function.
M4: This is the so-called *Triangle Inequality* . When applied to the Euclidean plane, i.e., $X = \mathbb{R}^2$ with the usual distance function, this inequality expresses the fact that the length of any side of a triangle is less than the sum of the lengths of the two other sides. M2 is also called *strict positivity* or *non-degeneracy.*

X, equipped with a metric d, is called a metric space, often denoted by (X, d), especially when it is important to specify which metric is being considered.

Examples:

1. $X = \mathbb{R}$, and $d(x, y) = |x - y|$.

2. $X = \mathbb{R}^n$, $n = 1, 2, 3, \ldots$, and the Euclidean metric given by
$$d(x, y) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \cdots (x_n - y_n)^2}$$
Note: the distance of $x \in \mathbb{R}^n$ to $0 \in \mathbb{R}^n$ (the origin), is called the length or *norm* of x, denoted by $\|x\|$. The obvious relation between the norm and the metric is then
$$d(x, y) = \|x - y\|$$

3. $X = \mathbb{R}^n$, $n = 1, 2, 3, \ldots$, and
$$d(x, y) = |x_1 - y_1| + |x_2 - y_2| + \cdots |x_n - y_n|$$
Note: the corresponding norm $\|x\| = d(0, x)$ is sometimes called the *sum norm.*

4. $X = \mathbb{R}^n$, $n = 1, 2, 3, \ldots$, and
$$d(x, y) = \max\{|x_1 - y_1|, |x_2 - y_2|, \ldots |x_n - y_n|\}$$
Note: the corresponding norm $\|x\| = d(0, x)$ is sometimes called the *maximum norm.*

5. Let X be a set people of one and the same generation and with a common ancestor, e.g. all greatgrandchildren of greatgrandmother $\mathcal{M}$. A metric d can then be defined on this set by putting $d(x, y) =$ the number of generations one has to go back to find the first common ancestor of any two individuals x and y. E.g., the distance between a brother and a sister is 1. Then
$$d(x, y) \leq \max\{d(x, z), d(z, y)\} \quad \text{for all } x, y, z \in X \tag{1.1}$$
and a fortiori M4 holds. (1.1) is stronger than the triangle inequality. When it holds the metric d is called an *ultrametric.* In taxonomy ultrametrics have been used to characterize genetic proximity of species.
Hint: to verify (1.1) it helps to draw a family tree on a piece of paper.

6. Let X be the set of all N-letter words in an alphabet A with k characters, i.e.,
$$X = \{(x_1, x_2, \ldots, x_N) \mid x_i \in A\}$$
and $A = \{a_1, a_2, \ldots, a_k\}$. A natural metric on X is then defined by
$$d(x, y) = \#\{i \mid x_i \neq y_i\}$$

7. The metric spaces most important for this course are infinite-dimensional ones, of which there is a large variety. Before we can start introducing those we need to review just a little more real analysis.

1.2 When distances tend to zero: convergence

1.2.1 Sequences

For concreteness let us consider again the real numbers with their usual metric. All definitions, however, do carry over to arbitrary metric spaces. Just replace $\mathbb{R}$ by X, $|x-y|$ by $d(x,y)$, etc.

Let (x_n) be a *sequence* of real numbers, i.e., for each $n \in \mathbb{N}, x_n \in \mathbb{R}$ (other common notations are $\{x_n\}, (x_n)_{n\geq 1}, \{x_n\}_{n=0}^{\infty}$ etc.

Definition 1.2 (convergence) *The sequence of real numbers (x_n) converges to $x \in \mathbb{R}$ iff for all $\epsilon > 0$ there is $N \in \mathbb{N}$ such that for all $n \geq N$, $|x_n - x| < \epsilon$.*

Note that in this definition N depends on ϵ, and you could write $N(\epsilon)$ to make this dependence explicit. The point x is called the *limit* of (x_n). Common notations expressing that (x_n) converges to x are:

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

Sequences that become eventually larger (smaller) than any number do not converge according to the definition given above. They are usually called *divergent.* But it is sometimes more useful to consider them as converging to $+\infty$, i.e., $x_n \to +\infty$ iff for all $M \in \mathbb{R}$, there is $N \in \mathbb{N}$, such that for all $n \geq N$, $x_n > M$. And similarly one defines $x_n \to -\infty$ iff for all $M \in \mathbb{R}$, there is $N \in \mathbb{N}$, such that for all $n \geq N$, $x_n < M$.

Next we define the concept of a *Cauchy* sequence.

Definition 1.3 (Cauchy sequence) *(x_n) is a Cauchy sequence iff for all $\epsilon > 0$ there is $N \in \mathbb{N}$ such that for all $n, m \geq N$, $|x_n - x_m| < \epsilon$.*

Cauchy sequences are sequences that *ought* to converge. The following propositon says that, in general, a necessary condition for any sequence to converge is that it is Cauchy.

Proposition 1.4 *If (x_n) converges to $x \in \mathbb{R}$, then (x_n) is Cauchy.*

This proposition holds for sequences in any metric space (X,d). A natural question now is whether the converse is also true: does every Cauchy sequence have a limit it converges to? For real sequences the answer is *yes*, but this is not so in general. The property that every Cauchy sequence in a metric space (X,d) converges to a limit *in* X, singles out a particularly interesting class of metric spaces. Such (X,d) are called *complete* . So, $\mathbb{R}$ with the usual metric is complete, but the rational numbers, $\mathbb{Q}$, considered as a subset of $\mathbb{R}$ and with the same distance function, is *not* complete. In fact, one way of constructing the set of real numbers is by considering all Cauchy sequences in $\mathbb{Q}$ and grouping those together in equivalence classes of sequences that ought to have the same limit. We will discuss completeness later this quarter.

Exercise 1.1 *Prove that $\mathbb{Q}$ with the usual distance function is not complete as a metric space. (Hint: find a sequence of rational numbers that is Cauchy, and hence converges in $\mathbb{R}$, but with an irrational limit.)*

Exercise 1.2 *Starting from the fact that $\mathbb{R}$ equipped with its usual distance function is complete, prove that*
a) $\mathbb{R}^n$ *with the Euclidean norm is a complete metric space,*
b) $\mathbb{R}^n$ *with the sum norm or the maximum norm is also complete.*

Another natural question is whether a sequence could converge to two different limits. In a metric space, the answer is *no*, in particular due to property M2 of the metric.

Proposition 1.5 *If $x_n \to x$ and $x_n \to y$, then $x = y$*

A *subsequence* of a sequence (x_n) is a sequence of the form $(x_{n_k})_{k=1}^{\infty}$, where for each $k \in \mathbb{N}$, $n_k \in \mathbb{N}$ and $n_k < n_{k+1}$, for all k. I.e., $k \mapsto n_k$ is a strictly increasing function from the set of natural numbers to itself. Clearly, if (x_n) is a convergent sequence with limit x, then all its subsequences also converge to x.

1.2.2 Upper and lower bound, inf and sup, lim inf and lim sup

There is a natural order relation on $\mathbb{R}$ that is nicely related to the metric structure. First, we define the notion of *upper* and *lower bound* for a subset A of $\mathbb{R}$.

Definition 1.6 (upper and lower bound) *Let $A \subset \mathbb{R}$. $M \in \mathbb{R}$ is an upper bound for A iff $x \leq M$, for all $x \in A$, and $m \in \mathbb{R}$ is a lower bound for A iff $m \leq x$, for all $x \in A$. The set A is called bounded iff both an upper and a lower bound for it exist.*

If A has an upper bound M, then A has many upper bounds. E.g, any $M' \geq M$ is also an upper bound. The *least upper bound* for A is called the *supremum* of A, denoted by $\sup A$. I.e., $\sup A$ is an upper bound with the property that if M is any upper bound for A, then $M \geq \sup A$.

Similarly one defines $\inf A$, the *infimum* of A, as the *greatest lower bound* for A. I.e., $\inf A$ is a lower bound with the property that if m is any lower bound for A, then $m \leq \inf A$.

By convention we set $\sup A = +\infty$ if A does not have an upper bound, i.e., if A contains arbitrarily large real numbers. In the same way, one sets $\inf A = -\infty$ if A does not have a lower bound.

Obviously, for each $A \neq \emptyset$, there can be only one least upper bound (greatest upper bound). This is due to the fact that $\mathbb{R}$ is *totally* ordered, i.e., for all $x, y \in \mathbb{R}$, $x \leq y$ or $y \leq x$. One also has $\inf A \leq \sup A$, for all $A \neq \emptyset$. For $A = \emptyset$ the convention $\inf A = +\infty$ and $\sup A = -\infty$ is sometimes useful.

That $\sup A$ ($\inf A$) exists for any set with an upper (lower) bound, is a consequence of (and actually equivalent to) the completeness of $\mathbb{R}$.

If A does not have an upper (lower) bound we put $\sup A = +(-)\infty$, by convention. If $\sup A \in A$, $\sup A$ is also the *maximum* of A, defined as the largest element of A, and often denoted by $\max A$. If $\inf A \in A$, $\inf A$ is also the *minimum* of A, defined as the smallest element of A, and often denoted by $\min A$. Note that every bounded set has a supremum and an infimum, but that it does not necessarily have a maximum or a minimum.

The following two quantities can again be defined in an arbritrary metric space (X, d).

For any $A \subset X$, the *diameter* of A is defined by

$$\mathrm{diam} A = \sup\{d(x, y) \mid x, y \in A\}$$

For a subset $A \subset X$ and a point $x \in X$ the distance from x to A is naturally defined by

$$\mathrm{dist}(x, A) \equiv d(x, A) = \inf\{d(x, y) \mid y \in A\}$$

Examples:

1. $A = \{1, 1/2, 1/3, 1/4, \ldots\}$. Then $\sup A = 1 = \max A$ and $\inf A = 0$. Note that, in this example, A does not have a minimum.

2. With A as given in 1), $\mathrm{diam} A = 1$. Note that $d(0, A) = 0$, although $0 \notin A$.

In order to define lim inf and lim sup we first observe that monotone sequences always converge (allowing for convergence to $\pm\infty$). Indeed, if (x_n) is *monotone increasing*, i.e., $x_n \leq x_{n+1}$, for all n, and (y_n) is *monotone decreasing*, i.e., $x_n \geq x_{n+1}$, for all n, then

$$\lim_{n\to\infty} x_n = \sup\{x_n\} \tag{1.2}$$

$$\lim_{n\to\infty} y_n = \inf\{y_n\} \tag{1.3}$$

lim inf (*limes inferior*) and lim sup (*limes superior*) can now be defined for an arbitray sequence (x_n), as the following two limits:

$$\liminf_{n\to\infty} x_n = \lim_{n\to\infty} \inf\{x_k \mid k \geq n\} \tag{1.4}$$

$$\limsup_{n\to\infty} x_n = \lim_{n\to\infty} \sup\{x_k \mid k \geq n\} \tag{1.5}$$

Another common notation is: $\liminf = \underline{\lim}, \limsup = \overline{\lim}$.

Exercise 1.3 *Let (x_n) be a sequence of real numbers. A point $c \in \mathbb{R} \cup \{\pm\infty\}$ is called a* cluster point *of (x_n) if there is a convergent subsequence of (x_n) with limit c. Let C denote the set of all cluster points of (x_n). Prove that*

$$\limsup x_n = \max C, \qquad \text{and } \liminf x_n = \min C \quad .$$

In contrast to the limit, the lim inf and lim sup always exist. Clearly, one always has $\liminf_n x_n \leq \limsup_n x_n$. Note that a sequence (x_n) converges to a limit x iff $\liminf_n x_n = \limsup_n x_n = x$.

Let A be an arbitrary set, and $x_\alpha \in \mathbb{R}$, for all $\alpha \in A$. Then $\sup\{x_\alpha \mid \alpha \in A\}$ is usually abbreviated to $\sup_{\alpha \in A} x_\alpha$, or, when this does not lead to confusion, by $\sup_\alpha x_\alpha$. Now consider a family of real numbers indexed by $n \in \mathbb{N}$ and $\alpha \in A$: $x_{n,\alpha}$. The following two relations are very useful and easy to prove:

$$\limsup_{n\to\infty} \inf_\alpha x_{n,\alpha} \leq \inf_\alpha \limsup_{n\to\infty} x_{n,\alpha} \tag{1.6}$$

$$\sup_\alpha \liminf_{n\to\infty} x_{n,\alpha} \leq \liminf_{n\to\infty} \sup_\alpha x_{n,\alpha} \tag{1.7}$$

Exercise 1.4 *Prove the inequalities (1.6) and (1.7)*

1.3 Series

Let (x_n) be a sequence in $\mathbb{R}$. The sequence of *partial sums* of (x_n) is the new sequence (s_n) defined by

$$s_n = \sum_{k \leq n} x_k$$

If (s_n) converges to a limit $s \in \mathbb{R}$, we say that the *series* $\sum_{n=1}^\infty x_n$ converges to s. If (s_n) does not converge the series is called *divergent.*

A series $\sum_{n=1}^\infty x_n$ is said to be *absolutely convergent* iff the series of absolute values $\sum_{n=1}^\infty |x_n|$ converges. Absolute convergence implies convergence, but not vice versa (see the assigned problems).

A useful feature of absolutely convergent series is that any series obtained from it by a permutation of its terms, is also convergent and has the same limit. For series of real numbers the property that any permutation converges to the same limit (this property is called *unconditional convergence*) is in fact equivalent to absolute convergence. This is no longer true for series in infinite-dimensional spaces (again, see later).

1.4 Open and closed sets

Consider a metric space (X, d), e.g., $\mathbb{R}^n$. The *open ball*, $B_r(x_0)$, with radius $r > 0$ and center $x_0 \in X$ is defined by

$$B_r(x_0) = \{x \in X \mid d(x_0, x) < r\}$$

The *closed ball*, $\overline{B}_r(x_0)$, is defined by

$$\overline{B}_r(x_0) = \{x \in X \mid d(x_0, x) \leq r\}$$

The basic concept of (set) topology is the notion of *open set.*

Definition 1.7 (open and closed sets) *A subset $G \subset X$ is open iff for all $x \in G$ there exist $r > 0$ such that $B_r(x) \subset G$.*
A subset $F \subset X$ is closed iff $F^c \equiv X \setminus F$ is open.

Proposition 1.8 *A subset F of a metric space is closed iff all convergent sequences of elements in F converge to a limit in F, i.e., if $x_n \to x$, and $x_n \in F$, for all n, then $x \in F$.*

A set $A \subset X$ is called bounded if it is contained in a ball $B_r(x_0)$, with $r < +\infty$. The following properties are easy to prove:

1. The empty set $\emptyset$ and X itself are open *and* closed.
2. The intersection $G_1 \cap G_2$ of two open sets $G_1, G_2 \subset X$ is open. Consequently, the intersection of any finite family of open sets is also open.
3. The union of an arbitrary family (finite or infinite) of open sets is open.
4. The union $G_1 \cup G_2$ of two closed sets $G_1, G_2 \subset X$ is closed, and hence any finite family of closed sets has a closed union.
5. The intersection of an arbitrary family (finite or infinite) of closed sets is closed.

Exercise 1.5 *Proof the properties 1-5.*

The closure $\overline{A}$ of a set $A \subset \mathbb{R}^n$, is, by definition, the smallest closed set containing A. The last property above implies that such a set indeed exists. In any metric space the closure of a subset A can be obtained by adding to A all limits of converging sequences of elements of A. I.e,

$$\overline{A} = \{x \in X \mid \text{there exist } x_n \in A, x_n \to x\}$$

Therefore, $\overline{A}$ is sometimes *defined* as the set of *points of adherence* of A. By definition, $x \in X$ is a *point of adherence* of A if every open set containing x also contains points of A.

For any $x \in X$ a *neighborhood* of x is a subset $V \subset X$ such that there is an open set $G \subset V$ containing x. Equivalently, V is a neighborhood of x if there is a ball $B_r(x) \subset V$, for some $r > 0$.

A set $A \subset X$ is called *dense* if $\overline{A} = X$.

1.5 Countably and uncountably infinite sets

This is a brief aside on infinite sets. Our only aim here is to clarify the distinction between countably infinite and uncountably infinite sets.

A set A is finite, i.e., has a finite number of elements, iff there is a bijection (a map that is one-to-one and onto), from A to one of the sets $\emptyset$ or $\{1, 2, \ldots, n\}$. Of course, n is the number of elements of A. If a set is not finite, it is called *infinite.*

A set A is *countable* if it is either finite or if there is a bijection from A to $\mathbb{N}$. In the latter case it is called *countably infinite*, i.e., in order to enumerate all the elements of A, you need an infinite number of natural numbers.

Sets that are not countable are *uncountable.* To enumerate (label) all the elements of an uncountable set, the natural numbers do not suffice.

Examples:

1. Any infinite subset of $\mathbb{N}$ (or any other countable set) is countably infinite. So, e.g., the even natural numbers form a countably infinite set.

2. $\mathbb{Z}$ is countably infinite.

3. $\mathbb{Q}$ is countably infinite. Do you know an enumeration of the rational numbers?

4. $\mathbb{R}$ is uncountably infinite. Do you know how to prove this?

1.6 Compactness

The full importance of the *compact sets* will become clear only in an infinite-dimensional setting, and we will have ample opportunity to see how useful they are later. It is nevertheless interesting to start with a finite-dimensional warmup.

Definition 1.9 *A subset K of a metric space X is called compact iff every sequence (x_n) of elements of K has a convergent subsequence with a limit $x \in K$.*

Sets satisfying the definition above are sometimes called *sequentially compact* in order to make a distinction with the more general definition of compactness in the framework of general topological spaces (of which the metric spaces are a special case). Using the general definition here would force us to also prove that for metric spaces compactness is equivalent to sequential compactness, which is the more useful notion for the applications we have in mind. Therefore, it is more economic not to go beyond the scope of metric spaces now. We will discuss the more general situation later.

Exercise 1.6 *Prove that every compact set is closed.*

The compact subsets of $\mathbb{R}^n$ have a simple and practical characterization given by the following theorem.

Theorem 1.10 (Heine-Borel) *A subset $K \subset \mathbb{R}^n$ is compact iff it is closed and bounded.*

The non-trivial part in the proof of the Heine-Borel Theorem is the following result attributed to Weierstrass. The property that, in a given metric space, every bounded sequence has a convergent subsequence is sometimes called the *Bolzano-Weierstrass property.*

Theorem 1.11 (Bolzano-Weierstrass) *Every bounded sequence in $\mathbb{R}^n$ has a convergent subsequence.*

Proof: This proof is constructive in the sense that we will give a procedure to construct a convergent subsequence starting from an arbitrary bounded sequence. In fact, this subsequence will be a Cauchy sequence by construction, and hence, as $\mathbb{R}^n$ is complete, it converges.

Let (x_i) denote the given sequence. As (x_i) is bounded, there is $M > 0$ such that $x_i \in [-M, M]^n$, for all i. The set $[-M, M]^n$ is a cube of side $2M$. The first element of the subsequence is chosen as follows: partition C_0 into 2^n cubes of side M, and pick C_1 to be one of these smaller cubes that contains an infinite number of elements of the sequence (x_i), i.e., $x_i \in C_1$, for infinitely many i. Such a cube exists because there are only a finite number of cubes and an infinite number of elements in the sequence. Let i_1 be the smallest index such that $x_{i_1} \in C_1$. This is the first element of the subsequence.

To find the second element, first form a new sequence (y_i) by delteing from (x_i), the element x_{i_1} and all $x_j \notin C_1$. Now, repeat the procedure described in the previous paragraph but with (x_i) replaced by (y_i), and $[-M, M]^n$ replaced by C_1. This will yield a subcube C_2 of C_1 of size $M/2$, and an element y_{i_1}. Put $x_{i_2} = y_{i_1}$. This is the second element of the subsequence.

Repeat this ad infinitum to obtain a subsequence (x_{i_j}) of (x_i). By construction we will have that for every j, and every $k \geq j$, $x_{i_k} \in C_j$, where C_j is a cube of size $2M/2^j$. Therefore, (x_{i_j}) is a Cauchy sequence and hence it converges. ∎

Exercise 1.7 *Complete the proof of Theorem 1.10.*

Note that for any metric space (X, d), and any $X_0 \subset X$ we can use the same distance function d on X_0, i.e., the restriction d_0 (often simply denoted by d) of d to X_0, to make X_0 into a metric space. If X_0 is a compact subset of X, (X_0, d) is a compact metric space.

The following is a useful property.

Exercise 1.8 *Let (K, d) be a compact metric space, and G a closed subset of K. Prove that G is compact.*

2 Continuous functions

For further reading and more exercises see
J.E. Marsden, *Elementary Classical Analysis*, 2nd ed., W.H. Freeman and Company, New York, 1993.
or
W. Rudin, *Principles of Mathematical Analysis*, 2nd ed. McGraw-Hill, New York, 1964

2.1 Continuity and semi-continuity

Let D be a subset of $\mathbb{R}$. A function $f : D \to \mathbb{R}$ is *continuous* at a point $x_0 \in D$ iff for all $\epsilon > 0$ there exists $\delta > 0$ such that for all $x \in D$, $|x - x_0| < \delta$ implies $|f(x) - f(x_0)| < \epsilon$. Continuity at x_0 is thus the property that the value of a function f in a point x close to x_0, is close to the value of f in x_0. If f is not continuous at x one says that it is *discontinuous* at x.

It is now obvious how to define continuity of a function from one metric space into another. Let (X, d_X) and (Y, d_Y) be two metric spaces.

Definition 2.1 (continuity) *A function $f : X \to Y$ is continuous at $x_0 \in X$ iff for all $\epsilon > 0$ there exists $\delta > 0$ such that for all $x \in X$, $d_X(x, x_0) < \delta$ implies $d_Y(f(x), f(x_0)) < \epsilon$.*

Remark: In the sequel we will usually denote all metrics by d when it is clear from the context (e.g., the arguments of d) which metric is meant.

We say that f is continuous on $D \subset X$ if f is continuous at all points $x \in D$. Note that this is not the same as saying that f restricted to D is continuous, e.g., if D contains only one point any function f restricted to D is continuous.

There are always continuous functions on X. E.g., any constant function is continuous. A less trivial example is the following. For any $x_0 \in X$, the function f defined by $f(x) = d(x_0, x)$ is continuous on X.

Definition 2.2 (uniform continuity) *A function $f : X \to Y$ is uniformly continuous on X iff for all $\epsilon > 0$ there exists $\delta > 0$ such that for all $x, y \in X$, $d_X(x, y) < \delta$ implies $d_Y(f(x), f(y)) < \epsilon$.*

Note the subtle difference between this and the previous definition.

Examples:

1. A function $f : \mathbb{R}^n \to \mathbb{R}^m$, satisfying

$$f(tx + (1-t)y) = tf(x) + (1-t)f(y), \qquad \text{for all } x, y \in \mathbb{R}^n, t \in [0,1]$$

is called *affine*. Affine functions are uniformly continuous.

2. The function $x \mapsto x^2, x \in \mathbb{R}$, is continuous but not uniformly continuous.

A function $f : D \to \mathbb{R}$, where $D \subset \mathbb{R}$ is an open set, is called *differentiable* at $x_0 \in D$, iff the following limit exists (and is finite):

$$f'(x_0) := \lim_{x \to x_0} \frac{f(x) - f(x_0)}{x - x_0}$$

Exercise 2.1 *Prove that if f is differentiable at x_0, then f is continuous at x_0.*

A useful characterization of continuous functions, which could also serve as an equivalent definition, is the following.

Proposition 2.3 (say it with sequences) *$f : X \to Y$ is continuous at $x \in X$ iff for every convergent sequence $x_n \to x$, the sequence $f(x_n)$ converges to $f(x)$.*

Proof:
The 'if' part: we show by construction that, if f is discontinuous at x, there is sequence (x_n), converging to x, but such that the sequence $(f(x_n))$ does not converge to $f(x)$. If f is discontinuous at x, there exists $\epsilon > 0$ such that for all $n \geq 1$, we can find $x_n \in X$ satisfying

$$d(x, x_n) < \frac{1}{n}, \text{ but } |f(x) - f(x_n)| > \epsilon$$

Clearly, the sequence x_n converges to x, but $f(x_n)$ does not converge to $f(x)$, as it stays at least a distance ϵ away from $f(x)$.
The 'only if' part: Suppose f is continuous at x, and $x_n \to x$. For any $\epsilon > 0$, there is $\delta > 0$ such that

$$\text{if } d(x, x_n) < \delta \text{ then } |f(x) - f(x_n)| < \epsilon$$

As $x_n \to x$, we can find N such that for all $n \geq N$, $d(x, x_n) < \delta$. Therefore, $f(x_n) \to f(x)$. ■

From the fact that continuous functions preserve convergence of sequences it immediately follows that they also preserve compactness. This is expressed in the following theorem.

Theorem 2.4 *Let $f : K \to Y$ be continuous on K, where K is a compact metric space and Y is any metric space. Then, $f(K) \subset Y$ is compact.*

In the rest of this section the target space Y will always be taken to be the real numbers. A function $f : D \to \mathbb{R}$ is called *bounded* on D iff there exist $M \in \mathbb{R}$ such that $|f(x)| \leq M$ for all $x \in D$, which is equivalent to saying that $f(D)$ is a bounded set.

Here is our first practical use of compactness:

Theorem 2.5 *Let K be a compact metric space (e.g., a compact subset of a metric space). If $f : K \to \mathbb{R}$ is continuous, then f is bounded and attains its supremum (i.e, maximum) and infimum (i.e., minimum). I.e.,*

$$\sup_{x\in K} f(x) = f(y) \quad \text{for some } y \in K \tag{2.1}$$

$$\inf_{x\in K} f(x) = f(z) \quad \text{for some } z \in K \tag{2.2}$$

Proof: First, we show that f is bounded. By Theorem 2.4, $f(K)$ is a compact subset of $\mathbb{R}$. Therefore, $f(K)$ is bounded by the Heine-Borel Theorem 1.10.

As (2.1) follows from (2.2) applied to the function $-f$, we only need to prove (2.2). We already know that f is bounded. In particular, f is bounded from below, i.e., there exists a constant $M \in \mathbb{R}$ such that $f(x) \geq M$, for all $x \in K$. Hence, $m := \inf_{x\in K} f(x) \geq M > -\infty$. By the definition of infimum, for all n there is $x_n \in K$ such that

$$|m - f(x_n)| < \frac{1}{n}, \text{ and hence } f(x_n) \to m \tag{2.3}$$

As K is compact it follows that (x_n) has a convergent subsequence. Let (y_n) be such a subsequence and denote its limit by y. Clearly, (2.3) implies $f(y_n) \to m$. As f is continuous $f(y_n) \to f(y)$. By the uniqueness of the limit, Proposition 1.5, we conclude $f(y) = m$, i.e., f attains its infimum in the point y. ■

Examples:

1. The function $f(x) = 1 - \sqrt{|x|}$ is continuous on $[-1, 1]$ and bounded. f attains its maximum in $x = 0$. (The theorem applies)

2. $f(x) = e^x$ is continuous on the whole of $\mathbb{R}$ but not bounded. ($\mathbb{R}$ is not compact)

3. $f(x) = 1/x$ on $(0, 1]$ is continuous but not bounded. ($(0, 1]$ is not compact)

4. $f(x) = \text{sign}x$ on $[-1, 1]$ is bounded, not continuous, but attains its maximum on any $x \in (0, 1]$ (by convention we put $\text{sign}0 = 0$). (continuity is not a necessary condition for attaining the maximum)

For functions that have only a certain kind of discontinuities, part of the conclusions of Theorem 2.5 still holds. There are two kinds of "half-continuous" functions, defined as follows.

Definition 2.6 (upper and lower semi-continuity) *$f : X \to \mathbb{R}$ is* upper *semi-continuous on X if for all $x \in X$ and any sequence $x_n \to x$ one has*

$$\limsup_{n\to\infty} f(x_n) \le f(x)$$

Similarly, f is lower *semi-continuous on X if for all $x \in X$ and any sequence $x_n \to x$ one has*

$$\liminf_{n\to\infty} f(x_n) \ge f(x)$$

Note that f is continuous iff f is both upper and lower semicontinuous. The analogue of Theorem 2.5 is then:

Theorem 2.7 *Let K be a compact metric space. If $f : K \to \mathbb{R}$ is upper semi-continuous, then f is bounded from above and attains its supremum. If $f : K \to \mathbb{R}$ is lower semicontinuous, then f is bounded from below and attains its infimum.*

Bounded (from) above here means that there is $M \in \mathbb{R}$ such that $f(x) \le M$ for all $x \in K$, and f is *bounded (from) below* iff $-f$ is bounded from above.

Exercise 2.2 *Let $(X, d_X), (Y, d_Y)$, and (Z, d_Z) be metric spaces and let $f : X \to Y$, and $g : Y \to Z$ be two continuous functions. Show that $h : X \to Z$, defined by $h(x) = g(f(x))$ is also continuous.*

2.2 Convergence of sequences of functions

In this section we consider sequences of real-valued functions on a common domain $D \subset X$ in a metric space (X, d), i.e., $f_n : D \to \mathbb{R}$, for $n \ge 1$. What would we mean by $f_n \to f$? Your "first thought" on this question is likely to be one of the following two:
1) The functions f_n are defined by the values they take on D. So, let us say that the functions converge if the values converge, i.e., if $f_n(x) \to f(x)$ for all $x \in D$. Thus, convergence of real-valued functions is reduced to convergence of real numbers, which we already know. This is called *pointwise* convergence of functions.
2) Find a set of functions with a reasonable metric on it and define $f_n \to f$ in the sense of that metric.

1) and 2) are both natural and useful ideas. But they are not compatible! It turns out that for most relevant domains D (e.g., whenever D is not a countable set), pointwise convergence cannot be expressed as convergence in a metric space

(see later). This could be good motivation to study more general topological spaces (beyond metric spaces). Let us postpone this study for another while, and first consider 1) and 2), separately, in some more detail.

Besides being non-metric, pointwise convergence has another limitation: it does not preserve continuity. Consider, e.g., the functions $f_n : [0,1] \to \mathbb{R}, f_n(x) = x^n$. Then, (f_n) converges pointwise to the function δ_1 defined by

$$\delta_1(x) = \begin{cases} 0 & \text{if } 0 \le x < 1 \\ 1 & \text{if } x = 1 \end{cases}$$

which is clearly discontinuous at $x = 1$.

A natural metric structure on the space of all continuous functions is given in terms of the *uniform* or *supremum* norm defined by:

$$\|f\| = \sup_{x \in D} |f(x)| \tag{2.4}$$

This norm is often denoted by $\|\cdot\|_{\mathrm{sup}}$ or $\|\cdot\|_\infty$. The origin of the latter notation will become clear when we introduce L^p spaces (MAT 203C). From the definition (2.4) it is clear that a function f is bounded iff $\|f\|_{\mathrm{sup}} < +\infty$.

Definition 2.8 (uniform convergence) *A sequence of bounded functions (f_n) converges uniformly to a function f iff*

$$\lim_{n\to\infty} \|f_n - f\|_{sup} = 0$$

Note that uniform convergence implies pointwise convergence. The example defined in (2.2) shows that the opposite does not hold. Here is another example.

Example:
$f_n(x) = \tanh nx \to \mathrm{sign} x$, pointwise, but not uniformly.

A crucial property of uniform convergence is that it preserves continuity. Below, we state the theorem, and its proof. Note the strategy of the proof.

Theorem 2.9 *Let (f_n) be a sequence of continuous real-valued functions on a metric space (X, d), and suppose $f_n \to f$ in the uniform norm. Then, f is continuous.*

Proof: In order to show that f is continuous at $x \in X$, we need to prove that for every $\epsilon > 0$, there is a $\delta > 0$ such that for all $y \in X$, with $d(x, y) < \delta$, $|f(x) - f(y)| < \epsilon$.

We start from

$$|f(x) - f(y)| \le |f(y) - f_n(y)| + |f_n(y) - f_n(x)| + |f_n(x) - f(x)|$$

First, use the *uniform* convergence to choose an n such that the first and the third term are less than $\epsilon/3$ *for all* $x, y \in X$. Then, use the continuity of f_n at x to find δ such that the middle term is less than $\epsilon/3$ for $d(x, y) < \delta$. The desired result follows. ■

2.3 The spaces $C(K)$

Let K be a compact metric space, and let $C(K)$ denote the set of all continuous functions $f : K \to \mathbb{R}$. Equipped with the metric derived from the uniform norm, $C(K)$ becomes a metric space. Explicitly:

$$d(f, g) = \|f - g\| = \sup_{x \in K} |f(x) - g(x)|$$

for all $f, g \in C(K)$. Note that due to Theorem 2.5, $\|f\| < +\infty$, for all $f \in C(K)$. It is straightforward to verify the properties M1-4 of Definition 1.1.

Theorem 2.10 *$C(K)$, with the uniform norm, is a complete metric space.*

Proof: Let (f_n) be a Cauchy sequence in $C(K)$ for the uniform norm. We have to show that (f_n) converges in the uniform norm. By Theorem 2.9 the limit function f will then automatically be continuous and hence $f \in C(K)$. We do this in two steps a) find a candidate function f for the limit of the sequence, and b) show that the sequence indeed converges uniformly to f.
a) Note that (f_n) converges pointwise due to the completeness of $\mathbb{R}$ and the observations that for all $x \in K$ the sequence of real numbers $(f_n(x))$ is Cauchy. Call the limit of $(f_n(x))$, $f(x)$. This defines f.
b) We show explicitly that $\lim_n \|f_n - f\|_{\text{sup}} = 0$.

$$\begin{aligned}
\|f_n - f\|_{\text{sup}} &= \sup_{x \in K} |f_n(x) - f(x)| \\
&= \sup_{x \in K} \lim_{m \to \infty} |f_n(x) - f_m(x)| \\
&= \sup_{x \in K} \liminf_{m \to \infty} |f_n(x) - f_m(x)| \\
&\leq \liminf_{m \to \infty} \sup_{x \in K} |f_n(x) - f_m(x)| \qquad (2.5)
\end{aligned}$$

The fact that (f_n) is Cauchy in the uniform norm implies that

$$\sup_{x \in K} |f_n(x) - f_m(x)| < \epsilon, \text{ for all } n, m \text{ large enough.}$$

■

$C(K)$ has additional structure: it is also a linear space for the natural operations of addition and multiplication by a scalar of functions, and its metric is derived from a norm. Normed linear spaces that are complete are called *Banach spaces*, and so, $C(K)$ is a Banach space. In fact, one can also multiply two functions (pointwise) and thus give $C(K)$ an *algebra* structure. One says that $C(K)$ is a *Banach algebra*.

Instead of real-valued functions on K, one can equally well consider complex-valued functions and the same result holds: the complex-valued continuous functions on K form a Banach space.

Exercise 2.3 *Let $f_n \in C([a,b])$ be a sequence of functions converging uniformly to a function f. Show that*

$$\lim_{n\to\infty} \int_a^b f_n(x)\,dx = \int_a^b f(x)\,dx$$

If X is not compact, continuous functions defined on it may be unbounded. But it is still true that the space of all *bounded continuous functions* on X, denoted by $C_b(X)$, is a Banach space for the supremum norm. In the case of $X = \mathbb{R}$, there is another commonly used Banach space of continuous functions: the space $C_0(\mathbb{R})$ consisting of all continuous functions vanishing at infinity, i.e.,

$$C_0(\mathbb{R}) = \{f : \mathbb{R} \to \mathbb{R} \mid f \text{ is continuous and } \lim_{x\to\pm\infty} f(x) = 0\}$$

Exercise 2.4 *Prove that $C_0(\mathbb{R})$ is a closed subspace of $C(\mathbb{R})$ with the sup norm.*

Exercise 2.5 *Show that $(C_0(\mathbb{R}), \|\cdot\|_{\mathrm{sup}})$ and $(C_b(\mathbb{R}), \|\cdot\|_{\mathrm{sup}})$, are Banach spaces. (Hint: adapt the proof of Theorem 2.10.)*

Theorem 2.11 *Let K be a compact metric space and let $f : K \to \mathbb{R}$ be a continuous real-valued function on K. Then, f is uniformly continuous.*

Proof: We prove the theorem by contraposition. Suppose f is not uniformly continuous. Then, there exists $\epsilon > 0$ such that for all $\delta > 0$, there are x, y, $d(x,y) < \delta$, but $d(f(x), f(y)) > \epsilon$. We use this for the sequence of δ's, $\delta_n = 1/n$ to obtain two sequences (x_n) and y_n in K with the properties:

$$d(x_n, y_n) < 1/n \quad \text{and} \quad d(f(x_n), f(y_n)) > \epsilon \tag{2.6}$$

Since K is compact there are convergent subsequences of (x_n) and (y_n), which for simplicity we again denote by (x_n) and (y_n), with limits x and y in K, and satisfying (2.6). Clearly, $x = y$, because

$$d(x,y) \le d(x,x_n) + d(x_n,y_n) + d(y_n,y)$$

and the three terms on the RHS of this inequality converge to 0. We conclude the proof by showing that assuming that all this holds for a continuous function f implies that f is not continuous at x (*reductio ad absurdum*)! Indeed, for the $\epsilon > 0$ as above, we now have that for all $\eta > 0$, there is x_n such that $d(x_n, x) < \eta$ but

$$d(f(x_n), f(x)) = \lim_m d(f(x_n), f(y_m)) \ge \epsilon$$

which implies that f is discontinuous at x. ∎

Let $[a, b] \subset \mathbb{R}$ be a closed interval. Consider the polynomial functions on $[a, b]$, i.e., functions p of the form

$$p(x) = \sum_{k=0}^{n} a_k x^k$$

where the coefficients a_k are arbitrary real numbers and $n \in \mathbb{N}$ is called the degree of p if $a_n \neq 0$.

Our next goal is the classical Weierstrass Approximation Theorem, which states that any continuous function $f : [a, b] \to \mathbb{R}$ can be approximated in the sense of the uniform norm by a polynomial (function), to arbitrary precision. Such an approximation property can be concisely formulated using the notion of dense set.

Theorem 2.12 (Weierstrass Approximation Theorem) *The set of all polynomial functions is dense in $C([a, b])$.*

Proof: We need to show that for any $f \in C([a, b])$ there is a sequence of polynomials p_n such that $\lim_n \|f - p_n\| = 0$.

We first show that it is sufficient to prove the theorem in the case $a = 0, b = 1$. This is so because rescaling and shifting the argument x are norm-preserving transformations from $C([a, b] \to C([0, 1])$ that map the polynomial functions into themselves. More explicitly, define $T : C([a, b] \to C([0, 1])$ by

$$Tf(x) = f(a + (b - a)x)$$

Then, $\|Tf\|_{\mathrm{sup}} = \|f\|_{\mathrm{sup}}$ and T is linear, in particular $T(f - g) = Tf - Tg$, and T is invertible with inverse

$$T^{-1}f = f(\frac{x - a}{b - a})$$

It is also easy to see that t and T^{-1} preserve continuity, and that for any polynomial p, Tp and $T^{-1}p$ are again polynomials. Then, for any $f \in C([a, b])$, $Tf \in C([0, 1])$, and assuming we have shown that the polynomials are dense in $C([0, 1])$, we then have a sequence of polynomials p_n such that $p_n \to Tf$ in $C([0, 1])$ with the uniform norm. From the properties discussed above it follows that the sequence $T^{-1}p_n$ is a sequence of polynomials converging to f.

So, we have to prove that the polynomial functions are dense in $C([0, 1])$. We follow the proof by Bernstein, who gave, for every f, an explicit formula for a sequence of polynomials converging to f. These are called the Bernstein polynomials and are defined as follows:

$$B_n(x; f) = \sum_{k=0}^{n} f(\frac{k}{n}) \binom{n}{k} x^k (1 - x)^{n-k} \tag{2.7}$$

$B_n(x;1) = 1$, by the binomial theorem. Therefore,

$$|f(x) - B_n(x;f)| \leq \sum_{k=0}^{n} |f(x) - f(\frac{k}{n})| \binom{n}{k} x^k (1-x)^{n-k} \tag{2.8}$$

Hence,

$$\begin{aligned} \|B_b(\cdot\,;f) - f\|_{\sup} &= \sup_{0\leq x\leq 1} \sum_{k=0}^{n} \left(f(\frac{k}{n}) - f(x) \right) \binom{n}{k} x^k (1-x)^{n-k} \\ &\leq \sup_{0\leq x\leq 1} \sum_{k=0}^{n} \left| f(\frac{k}{n}) - f(x) \right| \binom{n}{k} x^k (1-x)^{n-k} \end{aligned} \tag{2.9}$$

Next, we use the fact that f is *uniformly* continuous (Theorem 2.11), and pick a $\delta > 0$ such that $|f(k/n) - f(x)| < \epsilon$, for all k such that $|x - (k/n)| < \delta$. And define $I(\delta, x) = \{k||x - n/k| \geq \delta\}$. Then we can estimate the RHS of (2.9) as follows:

$$\begin{aligned} \|B_b(\cdot\,;f) - f\|_{\sup} &\leq \sup_{0\leq x\leq 1} \epsilon \sum_{0\leq k\leq n, k\notin I(\delta,x)} \binom{n}{k} x^k (1-x)^{n-k} \\ &\quad + \sup_{0\leq x\leq 1} \sum_{k\in I(\delta,x)} \left| f(\frac{k}{n}) - f(x) \right| \binom{n}{k} x^k (1-x)^{n-k} \\ &\leq \epsilon \sum_{k=0}^{n} \binom{n}{k} x^k (1-x)^{n-k} \\ &\quad + 2\|f\|_{\sup} \sup_{0\leq x\leq 1} \sum_{k\in I(\delta,x)} \binom{n}{k} x^k (1-x)^{n-k} \end{aligned} \tag{2.10}$$

The first term is equal to ϵ, and the second term can be estimated as follows:

$$\begin{aligned} & 2\|f\|_{\sup} \sup_{0\leq x\leq 1} \sum_{k\in I(\delta,x)} \binom{n}{k} x^k (1-x)^{n-k} \\ &\leq 2\|f\|_{\sup} \sup_{0\leq x\leq 1} \frac{1}{\delta^2} \sum_{k=0}^{n} (x - \frac{n}{k})^2 \binom{n}{k} x^k (1-x)^{n-k} \\ &\leq 2\|f\|_{\sup} \sup_{0\leq x\leq 1} \frac{1}{\delta^2} \sum_{k=0}^{n} (x^2 - 2\frac{k}{n}x + \frac{k^2}{n^2}) \binom{n}{k} x^k (1-x)^{n-k} \\ &\leq \frac{2\|f\|_{\sup}}{\delta^2} \sup_{0\leq x\leq 1} \left(x^2 B_n(x;1) - 2x B_n(x;x) + B_n(x;x^2) \right) \end{aligned} \tag{2.11}$$

Here, $1, x$, and x^2 denote the corresponding polynomial functions, when they serve as the second argument of B_n. We can now compute $B_n(x;1), B_n(x;x)$, and $B_n(x;x^2)$ explicitly to finish the estimate.

$$B_n(x;1) = \quad 1 \qquad \text{for all } n \geq 0$$

$$\begin{array}{rcll} B_n(x;x) = & x & \text{for all } n \geq 1 \\ B_n(x;x^2) = & \frac{n-1}{n}x^2 + \frac{1}{n}x & \text{for all } n \geq 1 \end{array} \tag{2.12}$$

All combined we obtain the estimate

$$\|B_n(\cdot\,;f) - f\|_{\sup} \leq \epsilon + \frac{\|f\|_{\sup}}{2n\delta^2}$$

which concludes the proof.

[Hint: prove (2.12) using the binomial theorem to expand $(x+y)^n$ and take derivatives with respect to x to obtain the identities:

$$\begin{aligned} x(x+y)^{n-1} &= \sum_{k=0}^{n} \binom{n}{k} \frac{k}{n} x^k y^{n-k} \\ \frac{1}{n}x(x+y)^{n-1} + \frac{n-1}{n}x^2(x+y)^{n-2} &= \sum_{k=0}^{n} \binom{n}{k} \frac{k^2}{n^2} x^k y^{n-k} \quad] \end{aligned} \tag{2.13}$$

■

Note that the proof works for both real-valued and complex-valued functions.

The space of all continuous functions (with values in $\mathbb{R}$ or $\mathbb{C}$) defined on the unit circle, T, is closely related to $C([0,1])$. A very usefull density result for $C(T)$ is the following.

If we parametrize the unit circle by an angle θ, i.e.,

$$T = \{e^{i\theta} \mid \theta \in [0, 2\pi)\}$$

any continuous function $f : T \to \mathbb{C}$ can be considered as a continuous periodic function of the real variable θ.E.g., the trigonometric polynomials are continuous. A *trigonometric polynomial* of degree n is a function of the form

$$P_n(\theta) = \sum_{k=-n}^{n} a_k e^{ik\theta}$$

the a_k are arbitrary complex coefficients.

The following density result will be proved later in the course (in MAT203B, when we discuss Fourier series). We will encounter several applications.

Theorem 2.13 *The set of all trigonometric polynomials is dense in $C(T)$ with the uniform norm.*

2.4 Compact subsets of $C(K)$

Recall that the proof of the Heine-Borel Theorem used the finite-dimensionality of $\mathbb{R}^n$ in a crucial way. In infinite-dimensional spaces compact sets are also necessarily bounded and closed, but these properties are no longer sufficient. The aim of this subsection is to prove a characterization of the compact subsets of $C(K)$, the space of all real-valued continuous functions on a compact space K with the uniform norm (the Arzelá-Ascoli Theorem). For this we will need the fact that every compact set has the *finite-open-subcover* property. First, we explain what the finite-open-subcover property is.

A family of open sets $G_\alpha \subset X$ is called an *open cover* of $A \subset X$ if

$$A \subset \bigcup_\alpha G_\alpha$$

The index α runs through an arbitrary index set that need not be countable. For any subset A of a metric space (X, d), it is obvious that for any $\epsilon > 0$, A can be covered with a family of open balls in X of radius ϵ, meaning that there exists a family of points $x_\alpha \in X$ such that

$$A \subset \bigcup_\alpha B_\epsilon(x_\alpha) \tag{2.14}$$

Indeed, it is always possible to take $\{x_\alpha\} = A$. Let us call such a cover an *ϵ-cover* of A. A subset $\{x_i\} \subset A$ with the property (2.14) is called an *ϵ-net* for A, and a finite ϵ-net if the set of points is finite. $A \subset X$ is called *totally bounded* if A has a finite ϵ-net for every $\epsilon > 0$.

We say that A has the finite-open-subcover property if for all open covers $\{G_\alpha\}$ of A there is a finite subcover. By this we mean that we can find a finite subset of the G_α, say $\{G_{\alpha_1}, \ldots, G_{\alpha_k}\}$ such that

$$A \subset \bigcup_{i=1}^{k} G_{\alpha_i}$$

The finite-open-subcover property is equivalent to compactness. We will prove one direction of this equivalence in Theorem 2.16. First, we prove some preliminary results.

Theorem 2.14 *Every compact metric space K is totally bounded, i.e., has a finite ϵ-net for every $\epsilon > 0$.*

Proof: Let $\epsilon > 0$, and suppose that K has no finite ϵ-net. Then, for every finite subset $\{x_1, \ldots, x_n\} \subset K$, there exists

$$x_{n+1} \in K \setminus \bigcup_{i=1}^{n} B_\epsilon(x_i)$$

So, we can find an infinite sequence (x_n) in K with the property that $d(x_n, x_m) > \epsilon$ for all $n \neq m$. Clearly, this sequence does not contain any Cauchy subsequence, and hence no convergent subsequence. We conclude that K must have a finite ϵ-net for every $\epsilon > 0$. $\blacksquare$

A metric space is called *separable* if it contains a countable dense subset.

Corollary 2.15 *Every compact metric space K is separable, i.e., contains a countable dense subset.*

Proof: For every $n \geq 1$, let A_n be a finite $\frac{1}{n}$-net of K. Such A_n exist due to Theorem 2.14. Then, $A := \bigcup_{n\in\mathbb{N}} A_n$ is dense in K, and, as a countable union of finite sets, A is countable. $\blacksquare$

The notation LHS := RHS is often used when the RHS *defines* the LHS. Other commonly used equivalent notations are $\equiv$ and $\stackrel{\text{def}}{=}$.

Theorem 2.16 *Let (X, d) be a metric space and suppose $K \subset X$ is compact. Then K has the finite-open-subcover property.*

Proof: We will first show that any open cover $\{G_\alpha\}$ of K contains a countable subcover. From Corollary 2.15, we know that K contains a countable dense subset A, say $A = \{y_1, y_2, \ldots\}$. For every i there is α_i such that $y_i \in G_{\alpha_i}$. $\{G_{\alpha_i} \mid i\}$ is then countable and that it is cover can be seen as follows: for every i there is $\epsilon_i > 0$ such that $B_{\epsilon_i}(y_i) \subset G_{\alpha_i}$, because G_{α_i} is open. Then

$$K = \overline{A} \subset \bigcup_i B_{\epsilon_i}(y_i) \subset \bigcup_i G_{\alpha_i}$$

(any points of adherence of A must be contained in one of the $B_{\epsilon_i}(y_i)$).

Next, we show that any countable open cover $\{G_i \mid i \in \mathbb{N}\}$ contains a finite subcover. To prove this we will perform a construction that either selects a finite subcover or yields a sequence (x_n) of elements $x_n \in K$ without a convergent subsequence with limit in K. If K is compact the last possibility must be excluded and hence there must be a finite subcover.

Define

$$R_n = K \setminus \bigcup_{i=1}^{n} G_i$$

If $R_n = \emptyset$ for some n, we have found a finite subcover and we are done. If $R_n \neq \emptyset$, for all n, we can construct a sequence $y_k \in K$ as follows. Pick $y_1 \in R_1$. Let n_1 be such that $y_1 \in G_{n_1}$. Then pick $y_2 \in R_{n_1}$, and so on. The result is a sequence y_k such that

$$y_{k+1} \notin \bigcup_{i=1}^{n_k} G_i \quad \text{and} \quad y_{k+1} \in \bigcup_{i=1}^{n_{k+1}} G_i \tag{2.15}$$

Suppose (y_{k_n}) is a convergent subsequence of (y_k), with limit y. Then, $y \notin G_i$, for any i, for if $y \in G_i$ for some i, (y_{k_n}) cannot converge to y. This follows from (2.15) and the fact that the G_i are open. We conclude that $y \notin K \subset \bigcup_i G_i$.

As K is assumed to be compact the possibility that all $R_n \neq \emptyset$ must be excluded and hence there is a finite subcover of K. ∎

Exercise 2.6 (Dini's Monotone Convergence Theorem)
Suppose $f_n \in C([0,1])$ converges pointwise and monotone decreasingly *to $f \in C([0,1])$. I.e., for all $x \in [0,1]$, the sequence $(f_n(x))$ is monotone decreasing and converges to $f(x)$. Prove that f_n converges uniformly to f.*

For the characterization of the compact subsets of $C(K)$ we will need the notion of *equicontinuity.*

Definition 2.17 (Equicontinuity) *Let F be a family of continous functions from (X,d) to (Y,d). F is equicontinuous iff for all $x \in X$, and for all $\epsilon > 0$, there is a $\delta > 0$ such that for all y with $d(x,y) < \delta$, one has $d(f(x), f(y)) < \epsilon$, for all $f \in F$.*

The crucial point in this definition is that δ can be chosen independent of f, while it may depend on x. If δ can be chosen independent of x as well, the family is *uniformly equicontinuous.* If X is compact equicontinuity implies uniform equicontinuity, as is shown in the following theorem.

Theorem 2.18 *Let F be a family of continuous functions from (K,d) to (Y,d), where K is compact. If F is equicontinuous, then F is uniformly equicontinuous.*

Proof: Suppose the equicontinuous family F is not uniformly equicontinuous. Then, there must be $\epsilon > 0$, such that for all $\delta > 0$ there are x_δ, and f_δ, such that there exist $y, d(x_\delta, y) < \delta$, but $d(f_\delta(y), f_\delta(x_\delta)) > 2\epsilon$.

In particular there are sequences x_n, y_n, and f_n, such that for all n, $d(x_n, y_n) < 1/n$, and

$$d(f_n(y_n), f_n(x_n)) > 2\epsilon \tag{2.16}$$

As K is compact, (x_n) has a convergent subsequence. Let such a subsequence again be denoted by (x_n). We then have $x_n \to x_0$, and hence $y_n \to 0$. Choose n large enough such that $d(x_n, x_0) < \delta$ and $d(y_n, x_0) < \delta$. Because of (2.16) we must have at least one of the following two: $d(f_n(y_n), f_n(x_0)) > \epsilon$ or $d(f_n(x_n), f_n(x_0)) > \epsilon$. As $\delta > 0$ is arbitrary, this is in contradiction with F being equicontinuous in x_0. ∎

Note that this theorem implies Theorem 2.11. It suffices to take $F = \{f\}$.

A set A is called *relatively compact*, or *precompact*, if its closure $\overline{A}$ is compact.

Theorem 2.19 (Arzelà-Ascoli) *Let K be compact. A set $A \subset C(K)$ with the uniform norm is compact iff A has the following two properties:*
i) A is closed and bounded, and
ii)A is equicontinuous.
Equivalently, A is precompact iff it satisfies:
i') A is bounded, and
ii') A is equicontinuous.

Proof: We divide the proof into four parts:
1) precompact $\Rightarrow$ bounded
2) compact $\Rightarrow$ closed
3) precompact $\Rightarrow$ equicontinuous
4) bounded + equicontinuous $\Rightarrow$ precompact.

1) If A is not bounded there is a sequence of functions $f_n \in A$, satisfying $\|f_{n+1}\| \geq \|f_n\| + 1$. Then, $\|f_n - f_m\| \geq 1$, for all $n \neq m$. It follows that (f_n) that does not have a Cauchy subsequence, and hence does not have a convergent subsequence.

2) If A is not closed, there is a sequence $f_n \in A$ converging to a limit $f \in \overline{A} \setminus A$. All subsequences of this sequence also converge to f, and hence do not have a limit *in* A.

3) If A is precompact, then $\overline{A}$ is compact by definition. Fix $\epsilon > 0$. A is dense in $\overline{A}$. Therefore

$$\overline{A} \subset \bigcup_{f \in A} B_\epsilon(f)$$

By Theorems 2.14 and 2.16 there is then a finite subset $\{f_1, \ldots, f_k\} \subset A$ such that

$$\overline{A} \subset \bigcup_{i=1}^{k} B_\epsilon(f_i)$$

As each f_i is uniformily continuous (Theorem 2.11), there exist $\delta_i > 0$ such that for all $x, y \in K$, we have

$$d(x,y) < \delta_i \Rightarrow |f_i(x) - f_i(y)| < \epsilon. \tag{2.17}$$

Then, for any $f \in A$, there is an i such that $\|f - f_i\| < \epsilon$ (because $\{f_i\}$ is an ϵ-net). Take $\delta = \min_i \delta_i$. We conclude that for $d(x,y) < \delta$

$$\begin{aligned} |f(x) - f(y)| &\leq |f(x) - f_i(x)| + |f_i(x) - f_i(y)| + |f_i(y) - f(y)| \\ &\leq 3\epsilon \end{aligned} \tag{2.18}$$

As ϵ is arbitrary, and δ is independent of f, this shows that A is equicontinuous.

4) We will show, using boundedness and equicontinuity, that every sequence in A has a convergent subsequence (then, necessarily, with a limit in $\overline{A}$). By Corollary

2.15 there is a countable dense set $\{x_1, x_2, x_3, \dots\} \subset K$. We now choose a first subsequence $(f_{1,n})$ of (f_n) such that the sequence $(f_{1,n}(x_1))$ converges in $\mathbb{R}$. Such a subsequence exists because A, and hence $(f_n(x_1))$, is bounded. Next, we choose a subsequence $(f_{2,n})$ of $(f_{1,n})$ such that $(f_{2,n}(x_2))$ converges, which exists for the same reason. And so on. $(f_{(k+1),n})$ is a subsequence of $(f_{k,n})$ such that $(f_{(k+1),n}(x_{k+1}))$ converges as $n \to \infty$. Finally, the subsequence we are after is the "diagonal" of the "table" of functions $f_{i,j}$, i.e., define the subsequence (g_k) of (f_n) by $g_k = f_{k,k}$.

By construction (g_k) has the property that $g_k(x_i)$ converges in $\mathbb{R}$ as $k \to \infty$ for all x_i in a dense subset of K. So far, we only used the boundedness of A. The equicontinuity is needed to ensure that (g_k) converges uniformly to a limit function g.

Let $\epsilon > 0$. Since A is equicontinuous, A is uniformily equicontinuous by Theorem 2.18. Therefore, and since all $g_k \in A$, there exists $\delta > 0$ such that for all x, y such that $d(x, y) < \delta$,

$$|g_k(x) - g_k(y)| < \epsilon$$

Since $\{x_i\}$ is dense in K we have

$$K \subset \bigcup_{i=1}^{\infty} B_\delta(x_i)$$

By Theorem 2.16 there is a finite subset $\{x_{i_1}, \dots, x_{i_l}\}$ such that

$$K \subset \bigcup_{j=1}^{l} B_\delta(x_{i_j})$$

Since for all $j = 1, \dots, l$, $(g_k(x_{i_j}))$ is convergent, and hence Cauchy, we can now find N such that

$$\text{for all } p, q \geq N, \quad |g_p(x_{i_j}) - g_q(x_{i_j})| < \epsilon \text{ for all } j = 1, \dots, l$$

Now, for any $x \in K$,

$$\begin{aligned} |g_p(x) - g_q(x)| &\leq |g_p(x) - g_p(x_{i_j})| + |g_p(x_{i_j}) - g_q(x_{i_j})| + |g_q(x_{i_j}) - g_q(x)| \\ &\leq 3\epsilon \end{aligned} \tag{2.19}$$

where we have made use of the δ-cover to choose x_{i_j} appropriately. It follows that (g_k) is a Cauchy sequence for the uniform norm. As $C(K)$ is complete (Theorem 2.10), it converges. ■

Remarks:

1. In the proof of 4) we used Cantor's "*diagonal argument*". Keep it in mind! It is used in other proofs. Also note the repeated use of the $\epsilon/3$-trick (in the disguise of a 3ϵ-trick).

2. The same result can be proved (in exactly the same way) for a more general situation: consider the space $C(K, Y)$ of all continuous functions $f : K \to Y$, with K a compact and Y an arbitrary metric space. Define a metric on $C(K, Y)$ by
$$d(f, g) = \sup_{x \in K} d(f(x), g(x))$$
Then $(C(K, Y), d)$ is a complete metric space and the compact subsets of $C(K, Y)$ are exactly those that have the properties i) and ii) stated in the Arzelà-Ascoli Theorem.

3. Parts 1) and 2) of the proof work as stated for arbitrary normed spaces.

Examples and counterexamples

We start with two examples of bounded sets in $C([0,1])$ that are not precompact (and hence their closures are bounded and closed but not compact). Both examples make use of the fact that, as a linear space, $C([0,1])$ is infinite-dimensional.

1. For each $n \geq 0$ define a function $f_n : [0,1] \to \mathbb{R}$ by

$$f_n(x) = \begin{cases} 0 & \text{if } \ 0 \leq x \leq 2^{-(n+1)} \\ 2^{n+2}(x - 2^{-(n+1)}) & \text{if } \ 2^{-(n+1)} \leq x \leq 3 \cdot 2^{-(n+2)} \\ 2^{n+2}(2^{-n} - x) & \text{if } \ 3 \cdot 2^{-(n+2)} \leq x \leq 2^{-n} \end{cases} \tag{2.20}$$

Then $\|f_n\|_\infty = 1$, for all $n \geq 1$, and so $\{f_n\}$ is bounded, but $\|f_n - f_m\|_\infty = 1$ for all $n \neq m$. Hence, the sequence (f_n) does not have any Cauchy subsequences, and therefore, as a set it can not be precompact. Make a graph of f_n and show explicitly that $\{f_n\}$ is not equicontinuous.

2. The same phenomenon occurs for the set of functions f_n defined on $[0,1]$ by $f_n(x) = \sin(n\pi x)$.

For a subset of an infinite-dimensional linear space to be (pre)compact "most of it" must be contained in a finite-dimensional subspace. Without making this more precise, we illustrate this with the following example.

3. Define a subset A of $C([0,1])$ as follows:

$$A = \{f \in C([0,1]) \mid f(x) = \textstyle\sum_{n=1}^\infty a_n \sin(n\pi x) \text{ with } a_n \in \mathbb{R}, \text{ and } \sum_{n=0}^\infty n^2 a_n^2 \leq 1\}$$

It can be shown that is precompact in $(C([0,1]), \|\cdot\|_\infty)$, by verifying the assumptions of the Arzelà-Ascoli Theorem.

Next, we give two examples of useful sufficient conditions for equicontinuity.

4. Let (X, d) be a metric space. A function $f : X \to \mathbb{R}$ is called (globally) *Lipschitz continuous* if there exists a constant $C \geq 0$ such that

$$|f(x) - f(y)| \leq Cd(x, y)$$

It is obvious that Lipschitz continuous $\Rightarrow$ continuous. For a Lipschitz continuous f we can define the *Lipschitz constant* $\mathrm{Lip}(f)$ of f by

$$\mathrm{Lip}(f) = \sup_{x \neq y \in X} \frac{|f(x) - f(y)|}{d(x, y)}$$

or, equivalently

$$\mathrm{Lip}(f) = \inf\{C \geq 0 \mid |f(x) - f(y)| \leq Cd(x, y), \text{ for all } x, y \in X\}$$

For any compact metric space K and any constant M define

$$A = \{f \in C(K) \mid \mathrm{Lip}(f) \leq M\} \tag{2.21}$$

Then, A is equicontinuous. Indeed, for all $\delta > 0$ we have

$$d(x, y) < \delta \Rightarrow |f(x) - f(y)| \leq M\delta \quad \text{for all } f \in A; x, y \in K$$

The set A defined by (2.21) is not bounded. E.g., it is easy to see that $f \in A \Rightarrow f + a \in A$, for any $a \in \mathbb{R}$, and hence, A is unbounded. Here is a bounded set $B \subset A$, in the case $K = [0, 1]$:

$$B = \{f \in A \mid f(0) = 0\}$$

Now, B is bounded because

$$|f(x)| = |f(x) - f(0)| \leq M|x - 0| \leq M$$

In fact, B is also closed (and hence compact). In fact, A is closed for arbitrary K and M. To check this we use once more our favorite relation involving lim inf. Let f_n be a sequence in B converging uniformly to f. Then

$$\begin{aligned}
\mathrm{Lip}(f) &= \sup_{x \neq y \in K} \frac{|f(x) - f(y)|}{d(x, y)} \\
&= \sup_{x \neq y \in K} \liminf_{n \to \infty} \frac{|f_n(x) - f_n(y)|}{d(x, y)} \\
&\leq \liminf_{n \to \infty} \sup_{x \neq y \in K} \frac{|f_n(x) - f_n(y)|}{d(x, y)} \\
&\leq \liminf_{n \to \infty} M = M
\end{aligned}$$

5. Let $C^1([0,1])$ denote the space of all continuously differentiable real-valued functions on $[0,1]$, and suppose $f \in C^1([0,1])$. Then, by the mean-value theorem, for every $x, y \in [0,1]$, $x < y$, there is z, $x < z < y$, such that

$$f(x) - f(y) = f'(z)(x-y)$$

It follows that $\mathrm{Lip}(f) \leq \|f'\|_\infty$, which is finite because $f' \in C([0,1])$ by assumption. For any constants M and N define a set $D \subset C^1([0,1])$ by

$$D = \{f \in C^1([0,1]) \mid \|f\|_\infty \leq M, \|f'\|_\infty \leq N\}$$

It follows that D is bounded and equicontinuous, and hence precompact in $C([0,1])$ with the uniform norm.

2.5 Existence of local solutions of ODE's

This section is devoted to a simple application of the Arzelà-Ascoli Theorem. Consider a scalar ordinary differential equation (ODE) of the type

$$y' = F(x,y)$$

on an open domain $D \subset \mathbb{R}$, and with "initial" condition $y(x_0) = y_0$. Solving this ODE on $D_0 \subset D$ means finding a function (or functions) $f : D_0 \to \mathbb{R}$ such that f and its derivative f' satisfy

$$\left.\begin{array}{rcll} f'(x) &=& F(x, f(x)) & \text{for all } x \in D_0 \\ f(x_0) &=& y_0 & \end{array}\right\} \tag{2.22}$$

We made a distinction between D and D_0, because it is a priori not clear one can find a solution f defined on the whole of D on which the problem is posed. But under mild conditions on F there is always an open interval D_0, containing x_0, on which at least one solution exists.

Theorem 2.20 (Local existence for scalar ODE) *If F is continuous on an open set $R \in \mathbb{R}^2$, containing (x_0, y_0), then there exists an open interval $D_0 \subset \mathbb{R}$, containing x_0, and a function $f : D_0 \to \mathbb{R}$ solving the system (2.22).*

Proof: The proof is mainly constructive, in the following sense. We will first construct a sequence of functions that "approximately" solve the ODE. This sequence will consist of elements in a precompact subset (Arzelà-Ascoli) of some $C([a,b])$, and hence it contains a uniformly convergent subsequence. We then show that any limit of such a subsequence must solve (2.22).

For any $\epsilon > 0$ we will define an approximate solution f_ϵ in the sense that

$$|f'_\epsilon(x) - F(x, f_\epsilon(x))| \leq \epsilon\,, \quad f_\epsilon(x_0) = y_0$$

for all x in some interval $[x_0 - d, x_0 + d]$, except for a finite number of points (where f'_ϵ does not exist). The f_ϵ are piecewise linear continuous functions obtained by the following, rather naive, construction:

$$f_\epsilon(x) = a_k + b_k(x - (x_0 + k\delta))\,, \qquad \text{for } x_0 + k\delta \leq x \leq x_0 + (k+1)\delta$$

for some appropriate $\delta > 0$, and a similar expression for $x \leq x_0$. The parameters a_k and b_k are recursively defined by

$$\begin{aligned} a_0 &= y_0 \\ b_0 &= F(x_0, y_0) \\ a_k &= a_{k-1} + b_{k-1}\delta \\ b_k &= F(x_0 + k\delta, a_k) \end{aligned}$$

for $k \geq 1$. Then $f'_\epsilon(x) = b_k$, for $x_0 + k\delta < x < x_0 + (k+1)\delta$, and therefore we have

$$\begin{aligned} |f'_\epsilon(x) - F(x, f_\epsilon(x))| &= |b_k - F(x, a_k + b_k(x - (x_0 + k\delta))| \\ &= |F(x_0 + k\delta, a_k) - F(x, a_k + b_k(x - (x_0 + k\delta))| \end{aligned} \tag{2.23}$$

for $x_0 + k\delta \leq x \leq x_0 + (k+1)\delta$, and

$$|x_0 + k\delta - x| \leq \delta \tag{2.24}$$

$$|a_k - a_k + b_k(x - (x_0 + k\delta))| \leq b_k\delta \tag{2.25}$$

Let $R \subset D \subset \mathbb{R}^2$ be a closed rectangle with center (x_0, y_0). We will choose R such that the graph of each f_ϵ, for $x \in [x_0 - d, x_0 + d]$ is contained in R (see figure[1]). Then

$$|b_k| \leq \sup_{(x,y)\in R} |F(x, y)| \leq M$$

for some $M \in \mathbb{R}$, because F is continuous. In fact, R is compact, and hence, F restricted to R is uniformly continuous. It then follows form (2.24) and (2.25) that for every ϵ, there exist $\delta > 0$ such that

$$|F(x_0 + k\delta, a_k) - F(x, a_k + b_k(x - (x_0 + k\delta))| \leq \epsilon$$

Combined with (2.23) this shows that f_ϵ is an approximate solution.

[1] Come to class to get the figure from the blackboard.

A suitable rectangle R is found as follows. First, consider a rectangle $R_1 \subset D$, with center (x_0, y_0). Define

$$M = \sup_{(x,y)\in R_1} |F(x,y)|$$

Then choose $R \subset R_1$ of the form $[x_0 - d, x_0 + d] \times [y_0 - dM, y_0 + dM]$. The figure makes it crystal clear why this choice of R works.

Let us summarize the properties of the approximate solution f_ϵ: for each ϵ, δ is found as described above. Then f_ϵ is piecewise linear on intervals of length δ. The slope of each linear piece is $\leq M$, and, therefore, the graph of f_ϵ lies in the cone bounded by $x_0 + Mx$ and $x_0 - Mx$ (see Figure). This also means that each f_ϵ is Lipschitz continuous with $\mathrm{Lip}(f_\epsilon) \leq M$. By construction we also have $f_\epsilon(x_0) = y_0$, for all ϵ. From example 4 in Section 2.4 we then know that the set $\{f_\epsilon \mid 0 < \epsilon \leq Md\}$ is bounded and equicontinuous, and hence, by the theorem of Arzelà-Ascoli (Theorem 2.19), precompact. It follows that there is a sequence $\epsilon_n \downarrow 0$ such that $f_{\epsilon_n} \to f$ uniformly on $[x_0 - d, x_0 + d]$.

It remains to be shown that this limiting function f solves (2.22). This goes as follows. As f_ϵ is piecewise linear we have

$$\begin{aligned} f_\epsilon(x) &= f_\epsilon(x_0) + \int_{x_0}^{x} f_\epsilon'(s)\, ds \\ &= y_0 + \int_{x_0}^{x} F(s, f_\epsilon(s))\, ds + \int_{x_0}^{x} [f_\epsilon' - F(s, f_\epsilon(s))]\, ds \end{aligned}$$

where f_ϵ' is not necessarily defined in points of the form $x_0 \pm k\delta$, but this does not affect the value of the integral (e.g., one could set the derivative equal to 0 in those points). Now, Exercises 2.2 and 2.3 come in handy to show that

$$\begin{aligned} f(x) &= \lim_{n\to\infty} \left\{ y_0 + \int_{x_0}^{x} F(s, f_{\epsilon_n}(s))\, ds + \int_{x_0}^{x} \left[f_{\epsilon_n}' - F(s, f_{\epsilon_n}(s))\right]\, ds \right\} \\ &= y_0 + \int_{x_0}^{x} F(s, f(s))\, ds \end{aligned} \tag{2.26}$$

Using the Fundamental Theorem of Calculus we conclude that the RHS of (2.26) is differentiable and therefore also the LHS is, and

$$f'(x) = F(x, f(x))$$

■

We observe that we also proved that there exists a differentiable solution.

Note that uniqueness of the solution is not part of Theorem 2.20. In fact, there are continuous F for which (2.22) has more than one solution on any open interval around x_0, as is shown in the following examples.

Examples of non-uniqueness:
1. Consider (2.22) with $F(x, y) = \sqrt{y}, y \geq 0$, and $x_0 = y_0 = 0$. I.e., we look for solutions of

$$\begin{aligned} f'(x) &= \sqrt{f(x)} && (2.27) \\ f(0) &= 0 && (2.28) \end{aligned}$$

Clearly, $f \equiv 0$ is a solution, but there is also the following one-parameter family of non-zero solutions:

$$f(x) = \begin{cases} 0 & \text{if } \ x \leq a \\ \frac{1}{4}(x-a)^2 & \text{if } \ x > a \end{cases}$$

for any $a \geq 0$. The construction used in the proof of Theorem 2.20 would only yield the zero solution. One could slightly modify it to produce other solutions as well. In any case this example illustrates clearly that under the conditions stated in Theorem 2.20, one does not necesarily specify a unique function by the system (2.22).

To get a clue as to what extra condition would be needed to guarantee uniqueness of the solution, consider the following example, which is a generalization of Example 1.
2. For $\alpha \geq 0$, we consider

$$\left.\begin{aligned} f'(x) &= |f(x)|^\alpha \\ f(0) &= 0 \end{aligned}\right\} \qquad (2.29)$$

Again there is always the solution $f \equiv 0$. When one tries to find other solutions of the form

$$f(x) = \begin{cases} 0 & \text{if } \ x \leq a \\ C(x-a)^\nu & \text{if } \ x > a \end{cases}$$

one arrives at the following condition on the parameters C and ν:

$$\alpha = 1 - \frac{1}{\nu} \quad , c^{1-\frac{1}{\nu}} = C\nu \qquad (2.30)$$

For $f \not\equiv 0$, we need $C \neq 0$, and the initial condition $f(0) = 0$ is only satisfied if $\nu > 0$. From (2.30) we conclude then conclude that $\alpha < 1$. If $\alpha \geq 1$ there are no non-zero solutions of the proposed from. In fact, as a direct application of the following proposition, one can show that (2.29) has indeed only one solution when $\alpha \geq 1$.

Proposition 2.21 *If F is continuous on an open set $R \in \mathbb{R}^2$, containing (x_0, y_0), and F is Lipschitz continuous in its second argument, with a Lipschitz constant independent of x (on the domain R). Then, on any interval containing x_0 there exists at most one solution of the system (2.22).*

Proof: The strategy of this uniqueness proof is a very common one: suppose there are two solutions, say f and g, show that the difference $\Delta := f - g$ satisfies an equation that implies $\Delta = 0$.

If f and g both satisfy (2.22) on a interval I containing x_0, it is easy to see that Δ must satisfy

$$\left.\begin{array}{rcl} \Delta'(x) & = & F(x, f(x)) - F(x, g(x)) \quad , \text{ for all } x \in I \\ \Delta(x_0) & = & 0 \end{array}\right\} \tag{2.31}$$

As F is Lipschitz continuous in its second argument, there is a constant C such that

$$|F(x, f(x)) - F(x, g(x))| \leq C|f(x) - g(x)| = C|\Delta(x)|$$

for all $x \in I$. This implies that the function G defined by

$$G(x) = \frac{F(x, f(x)) - F(x, g(x))}{\Delta(x)}$$

is bounded on I. On the other hand we have

$$\frac{\Delta'(x)}{\Delta(x)} = G(x)$$

for all $x \in I, x \neq x_0$. This equation can be integrated to obtain

$$\frac{\Delta(y)}{\Delta(x)} = \exp\{\int_x^y G(s)\, ds\}$$

for all $x \neq x_0$. Using the boundedness of G we conclude that Δ, which is by assumption a differentiable and hence continuous function, satisfies

$$e^{-M}|\Delta(x)| \leq |\Delta(y)| \leq e^{M}|\Delta(x)|$$

for all $x, y \in I$, and where M can be taken to be $C|I|$, with $|I|$ denoting the length of the interval I. Hence, Δ either never vanishes or is identically 0. As $\Delta(x_0) = 0$, we must have $\Delta = 0$. ■

3 Topological Spaces

In the previous weeks we have generalized concepts from the finite-dimensional spaces $\mathbb{R}^n$ with the usual Euclidean metric, to general metric spaces. We are fundamentally interested in convergence of sequences and continuity of functions. These properties have been expressed in terms of the metric. And as an important auxiliary tool we explored the notion of compactness.

As a particularly useful example of infinite-dimensional metric spaces we studied the function spaces $C(K)$ in detail. We also mentioned that pointwise convergence of functions (e.g., real-valued functions defined on an interval) can generally not be expressed in terms of a metric on some function space, suggesting that there is an even more general framework for expressing convergence, continuity, and compactness. Indeed there is: topological spaces.

Here, we limit ourselves to the basic definitions. For more information on point set topology and its use, see, e.g.,
W. Rudin, *Functional Analysis*, 2nd ed, McGraw-Hill, New York 1991
or
J.L. Kelley and I. Namioka, *Linear Topological Spaces*, D. Van Nostrand Company, Princeton, NJ 1963

3.1 General definition of a topological space

Let X be an arbitrary set.

Definition 3.1 *A set $\mathcal{T}$ of subsets of X is a topology on X if*
i) $\emptyset \in \mathcal{T}$ and $X \in \mathcal{T}$,
ii) for every family $\{G_\alpha\} \subset \mathcal{T}$, $\bigcup_\alpha G_\alpha \in \mathcal{T}$, and
iii) for every finite family $\{G_1, \dots, G_n\} \subset \mathcal{T}$, $\bigcap_{1\leq i\leq n} G_i \in \mathcal{T}$.

The elements of $\mathcal{T}$ are called *open sets.* A set $A \subset X$ is called *closed* iff its complement in X, $A^c := X \setminus A$, is open. The pair $(X, \mathcal{T})$ is called a *topological space.*

Examples:

1. For any X, $\mathcal{T} = \{\emptyset, X\}$ is a topology and it is obviously the smallest $\mathcal{T}$ that satisfies Definition 3.1. $\mathcal{T}$ is called the *trivial topology* or *indiscrete topology.*

2. For any X, $\mathcal{T} = \mathcal{P}(X)$, the set of *all* subsets of X, obviously also satisfies Definition 3.1. It is called the *discrete topology.*

3. Let $(X, \mathcal{T})$ be a topological space and $X_0 \subset X$. Then $\mathcal{T}_0 = \{G_0 \subset X_0 \mid G_0 = G \cap X_0, \text{ for some } G \in \mathcal{T}\}$ is a topology on X_0. So, the open subsets of X_0 are all sets obtained by intersecting the open sets of X with X_0. This topology is called the *induced topology* or *relative topology* on X_0 and X_0 is said to be a topological subspace of X.

4. Let (X, d) be a metric space. Then the set of all open sets, as defined in Section 1, is a topology on X.

We will encounter many concrete examples of topologies later on.

Some more terminology:

- A set $V \subset X$ is called a *neighborhood* of a point $x \in X$ if there exixts an open set contained in V and containing x, i.e., if there exists $G \in \mathcal{T}$ such that $x \in G \subset V$.
- A topology $\mathcal{T}$ is called *Hausdorff* (or $(X, \mathcal{T})$ is called a Hausdorff space) if every pair of distinct points $x, y \in X$, has a pair of non-intersecting neighborhoods, i.e., there exists a neighborhood V_x of x and a neighborhood V_y of y such that $V_x \cap V_y = \emptyset$. Almost all topological spaces used in analysis are Hausdorff. E.g., all metric topologies are Hausdorff (why ?). On the other hand, if X has at least two elements, the trivial topology on X is not Hausdorff.

Next, we express the basic notions of convergence, continuity, and compactness in terms of open sets. Let X and Y be a topological spaces.

Definition 3.2 (convergence) *A sequence (x_n) in X converges to a limit $x \in X$ iff for every neighborhood V of x, there exists N such that for all $n \geq N$, $x_n \in V$.*

This definition says that eventually the sequence lies entirely in any neighborhood of x.

Exercise 3.1 *What are the convergent sequences in a topological space X with the trivial topology? And with the discrete topology?*

Definition 3.3 (continuity) *A function $f : X \to Y$ is continuous at $x \in X$ if for each neighborhood W of $f(x)$ there exists a neighborhood V of x such that $f(V) \subset W$. f is continuous on X if it is continuous at every $x \in X$.*

Theorem 3.4 *Let $(X, \mathcal{T})$ and $(Y, \mathcal{S})$ be two topological spaces and $f : X \to Y$. Then f is continuous iff for every $G \subset \mathcal{S}$, $f^{-1}(G) \in \mathcal{T}$.*

Exercise 3.2 *Prove Theorem 3.4*

If a function $f : X \to Y$ is one-to-one and onto (a bijection), and both f and f^{-1} are continuous, f is called a *homeomorphism*, and any two topological spaces X and Y such that there exists a homeomorphism $f : X \to Y$, are called *homeomorphic.*

Homeomorphic spaces are indistinguishable as topological spaces. E.g., if X and Y are homeomorphic, one has that for any sequence (x_n) in X, (x_n) converges to x iff the sequence $(f(x_n))$ in Y converges to $f(x)$ (prove this).

Remark:
A bijection $f : X \to Y$ always has an inverse f^{-1}, but f^{-1} need not be continuous even if f is. E.g., consider $f : [0, 2\pi) \to S^1, f(\theta) = e^{i\theta}$, where $[0, 2\pi) \subset \mathbb{R}$ is considered with the topology induced by the standard topology on $\mathbb{R}$, and $S^1 \subset \mathbb{C}$ is the unit circle with the topology induced by the topology of the complex plane (or, equivalently, $\mathbb{R}^2$). f is continuous, but f^{-1} is not. For other examples see Section 3.3

Definition 3.5 (compactness) *$K \subset X$ is compact if every open cover of K contains a finite subcover.*

Note that $K \subset X$ is compact in the topology $\mathcal{T}$ on X iff K is compact as a subset of itself for the induced topology of X on K.

Exercise 3.3 *Let (X, d_1) and (Y, d_2) be metric spaces. Prove that the definitions of convergence and continuity are equivalent to the ones previously given (Definition 1.2 and 2.1).*

The following proposition states that, in the case of metric spaces, compactness as defined here, is equivalent to (sequential) compactness as defined in Definition 1.9. Hence, we can keep using the term compactness (instead of sequential compactness) when we are considering metric spaces.

Proposition 3.6 *Let (K, d) be a metric space. Then K is compact in the sense of Definition 3.5 iff every sequence (x_n) in K has a convergent subsequence.*

Proof: We already proved in Theorem 2.16 that sequential compactness implies the finite-open-subcover property, i.e., compactness. It remains to prove the other direction, which we do by contraposition.

Suppose there is a sequence (x_n) in K without any convergent subsequence. Then the sequence necessarily contains an infinite number of distinct x_n. Therefore, without loss of generality, we may as well assume that all elements of (x_n) are distinct. We will show that the existence of such a sequence implies that K does not have the finite-open-subcover property, i.e., that K is not compact.

We start by constructing an open cover of K as follows. For each $x \in K$, either $x \in \{x_n \mid n \geq 1\}$ or $x \notin \{x_n \mid n \geq 1\}$. In both cases, as (x_n) has no convergent subsequences, there is $\epsilon_x > 0$ such that $x_n \notin B_{\epsilon_x}(x)$ for all $x_n \neq x$. Clearly $\{B_{\epsilon_x}(x) \mid x \in K\}$ is an open cover of K. This cover has the property that every finite subcover $\{B_{\epsilon_x}(x) \mid x \in \{y_1, \ldots, y_k\} \subset K\}$ contains at most k elements of the sequence (x_n). As we assumed that all the elements of (x_n) are distinct, this implies that any finite subcover does not contain all elements of (x_n). Therefore, this open cover of K does not contain a finite subcover, which is in contradiction with the assumption that K is compact. ■

We repeat that in general, for non-metric topological spaces, sequential compactness does not imply compactness. In Proposition 1.8 we gave a characterization of closed sets, and the closure of a set, in terms of convergent sequences. Again, this breaks down for general topological spaces. Similarly, continuous functions $f : X \to Y$, in the case X and Y are metric spaces, were characterized in Proposition 2.3 as exactly those functions that preserve convergent sequences. Again, not so in arbitrary topological spaces. There are however objects that replace sequences, called nets, that can be used to express all the above-mentioned properties in an alternative way. We will not make use of nets in this course.

3.2 Bases of open sets

The set of all open sets can be huge and unwieldy. Recall that for metric spaces everything could be expressed in terms of open balls alone, a rather small subset of the open sets. In this section we introduce subsets of $\mathcal{T}$ (called *bases*) that can play the same role for an arbitrary topology as the open balls for a metric topology. We will also see that such special sets of open sets arise naturally when one defines new topologies.

Definition 3.7 *A family of open sets $\mathcal{B} \subset \mathcal{T}$ is a base for the topology $\mathcal{T}$, if for every $G \in \mathcal{T}$ there is a family of $B_\alpha \in \mathcal{B}$, such that $G = \bigcup_\alpha B_\alpha$.*

Examples:
1. The family of all open intervals $(a, b) \subset \mathbb{R}$, $a, b \in \mathbb{R}$, is a base for the standard topology on $\mathbb{R}$.
2. The family of all open intervals $(a, b) \subset \mathbb{R}$, with $a, b \in \mathbb{Q}$, is a base for the standard topology on $\mathbb{R}$.
3. Let (x, d) be a metric space. The set of all open balls $B_{1/n}(x)$, with $n \geq 1$ and $x \in A$, where A is dense in X, is a base for the metric topology on X.

Definition 3.8 *A family $\mathcal{N}$ of neighborhoods of a point $x \in X$ is called a neighborhood base (or a base of neighborhoods) of x if for any neighborhood V of x there exists a $W \in \mathcal{N}$ such that $W \subset V$.*

Example:
Let (X, d) be a metric space. The open (or closed) balls with center x and radius $1/n$, $n \geq 1$, form a neighborhood base for x.

Definition 3.9 (first countable) *A topological space X is first countable if every $x \in X$ has a countable neighborhood base.*

Examples:
1. As the examples above show, any metric space is first countable.
2. Any X with the discrete topology is first countable.
3. Any X with the trivial topology is first countable.
Often, it is useful to *define* a topology in terms of a base or a neighborhood base for every point.

Examples:
1. Let X be the space of all real-valued functions on $[a, b] \subset \mathbb{R}$. Let $f \in X$, $\epsilon > 0$, $n \geq 1$, and $x := \{x_1, \dots, x_n\} \subset [a, b]$. Define a subset $B_{f,x,\epsilon} \subset X$ by

$$B_{f,x,\epsilon} = \{g \in X \mid |g(x_i) - f(x_i)| < \epsilon, \text{ for all } i = 1, \dots, n\}$$

Then, define $\mathcal{T}$ as the smallest set of subsets of X satisfying the conditions of Definition 3.1 and containing $B_{f,x,\epsilon}$, for all $f \in X$, $\epsilon > 0$, and finite sets $x \subset X$. This defines the topology of pointwise convergence of functions in the sense that $f_n \to f$ in the topological space $(X, \mathcal{T})$ iff $f_n(x) \to f(x)$ for all $x \in [a, b]$. The sets $B_{f,x,\epsilon}$ are a neighborhood base for $f \in X$. A base for the topology can be obtained by considering arbitrary finite intersections of sets of the form $B_{f,x,\epsilon}$. In fact, it is easy to see that it is sufficient to consider sets of the form

$$\{g \in X \mid |g(x_i) - f(x_i)| < \epsilon_i, \text{ for all } i = 1, \dots, n\}$$

for all $f \in X$, $\epsilon_i > 0$, and finite sets $x \subset X$.
2. Let V be a linear space. A function $p : V \to \mathbb{R}$ is called a *seminorm* on V if it has the following properties:
i) $p(x) \geq 0$ for all $x \in V$.
ii) $p(ax) = |a|p(x)$, for all $x \in V$, $a \in \mathbb{R}$.
iii) $p(x + y) \leq p(x) + p(y)$, for all $x, y \in V$.
If, in addition, p satisfies $p(x) > 0$ for all $x \neq 0$, p is called a *norm*. We already know how to make a normed linear space into a metric space (and hence a topological space). More generally, suppose P is a set of seminorms on a linear space V. One can then define a neighborhood base for any $x \in V$ parametrized by $\epsilon > 0$ and finite subsets $\{p_1, \dots, p_n\} \subset P$:

$$B_{\epsilon,p_1,\dots,p_n}(x) = \{y \in V \mid p_1(x - y) < \epsilon, \dots, p_n(x - y) < \epsilon\}$$

We will define various interesting topologies on linear spaces exactly in this way, i.e., by defining appropriate sets of seminorms. Note that the topology of pointwise convergence on the space of all real-valued functions on some common domain is of this type. If the set of seminorms used is uncountable, chances are that the topology is not first countable and non-metric.

Theorem 3.10 *A family of open sets $\mathcal{B} \subset \mathcal{T}$ is a base for the topology $\mathcal{T}$ iff for every $x \in X$, $\mathcal{B}$ contains a base of neighborhoods for x.*

Proof: Suppose $\mathcal{B}$ is a base for $\mathcal{T}$. Then, if V is a neighborhood of x then there exists $G \in \mathcal{T}$, such that $x \in G \subset V$. As $\mathcal{B}$ is a base, there are B_α such that $\bigcup_\alpha B_\alpha = G$. Therefore, there exists $B \in \mathcal{B}$, such that $x \in B$ and $B \subset V$. We conclude that $\mathcal{B}$ contains a base of neighborhoods of x.

Conversely, if a family of open sets $\mathcal{B}$ contains a base of neighborhoods for every $x \in X$, then, for every open set $G \in \mathcal{T}$, and any $x \in G$, there exists $B_x \in \mathcal{B}$ such that $x \in B_x \subset G$. Therefore, $\bigcup_x B_x = G$, and we have proved that $\mathcal{B}$ is a base for the topology. ■

3.3 Comparing topologies

It is clear that different metrics on a space X can lead to one and the same notion of convergence of sequences and continuity of functions. E.g., if d_1 an d_2 are two metrics on X such that for all $x, y \in X$, $d_1(x, y) = 2d_2(x, y)$, this will certainly be the case. And in general, if two metrics lead to exactly the same set of open sets all topological properties should be the same. Ee will see that indeed the notion of topology (as the set of open sets) is more fundamental than, e.g., a norm or a metric.

As noted in the previous section, a topology if often given indirectly by a base of neighborhoods for every point. A natural question is then: how can one compare topologies by investigating a base of neighborhoods for them? When are the topologies the same?

Let $\mathcal{T}_1$ and $\mathcal{T}_2$ be two topologies on the same space X. Then, $\mathcal{T}_2$ is called *finer* than $\mathcal{T}_1$ if $\mathcal{T}_1 \subset \mathcal{T}_2$, i.e., if it has more open sets (one then also says that $\mathcal{T}_1$ is *coarser* than $\mathcal{T}_2$). Of course, not every pair of topologies are comparable, i.e., it is quite possible that $\mathcal{T}_1 \not\subset \mathcal{T}_2$ and $\mathcal{T}_2 \not\subset \mathcal{T}_1$.

Proposition 3.11 *Let X and Y be two spaces, each with two topologies, $\mathcal{T}_1, \mathcal{T}_2$ and $\mathcal{S}_1, \mathcal{S}_2$ respectively. If $f : (X, \mathcal{T}_1) \to (Y, \mathcal{S}_1)$ is continuous and*
i) if $\mathcal{T}_2$ is finer than $\mathcal{T}_1$, then $f : (X, \mathcal{T}_2) \to (Y, \mathcal{S}_1)$ is continuous
ii) if $\mathcal{S}_2$ is coarser than $\mathcal{S}_1$, then $f : (X, \mathcal{T}_1) \to (Y, \mathcal{S}_2)$ is continuous

Proof: These statements are a direct consequence of the general definition of continuity (Definition 3.3), or, more directly, of Theorem 3.4. ∎

The most regular bijection from a space X to itself one can imagine is the identity map $\mathrm{id} : X \to X, x \mapsto \mathrm{id}(x) = x$. As a map $\mathrm{id}^{-1} = \mathrm{id}$, and for every topology $\mathcal{T}$ on X, $\mathrm{id}^{-1} = \mathrm{id} : (X, \mathcal{T}) \to (X, \mathcal{T})$ are continuous, and hence *id* is a homeomorphism. The situation changes drastically when we consider two different topologies on X, say $\mathcal{T}_1$ and $\mathcal{T}_2$. E.g., consider

$$\mathrm{id} : (X, \{\emptyset, X\}) \to (X, \mathcal{P}(X))$$

Then, if X contains at least two points, id is not continuous at any point $x \in X$. More generally, as the following theorem shows, the continuity of the map id is a perfect tool for comparing topologies.

Theorem 3.12 *Let $\mathcal{T}_1$ and $\mathcal{T}_2$ be two topologies on X. Then the map* $\mathrm{id} : (X, \mathcal{T}_1) \to (X, \mathcal{T}_2)$ *is continuous iff $\mathcal{T}_1$ is finer than $\mathcal{T}_2$.*

Proof: This is a direct consequence of Theorem 3.4. ∎

Corollary 3.13 $\mathrm{id} : (X, \mathcal{T}_1) \to (X, \mathcal{T}_2)$ *is a homeomorphism iff $\mathcal{T}_1 = \mathcal{T}_2$.*

Theorem 3.14 *Let $\mathcal{T}_1$ and $\mathcal{T}_2$ be two topologies on X. Then $\mathcal{T}_1 = \mathcal{T}_2$ iff the following is true: for all topological spaces $(Y, \mathcal{S})$, a function $f : (X, \mathcal{T}_1) \to (Y, \mathcal{S})$ is continuous iff $f : (X, \mathcal{T}_2) \to (Y, \mathcal{S})$ is continuous.*

Proof: If $\mathcal{T}_1 = \mathcal{T}_2$, then the statement about continuous functions $f : X \to Y$ is trivial. For the other direction we use Corollary 3.13. So, we have to show that $\mathrm{id} : (X, \mathcal{T}_1) \to (X, \mathcal{T}_2)$ is a homeomorphism. This follows by once taking $(Y, \mathcal{S}) = (X, \mathcal{T}_1)$, and once $(Y, \mathcal{S}) = (X, \mathcal{T}_2)$. ∎

The next result is of practical importance, because often topologies are defined indirectly through bases of neighborhoods for every point. So, it is particularly useful to be able to distinguish (or identify) topologies on the basis of bases of neighborhoods alone.

Theorem 3.15 *Let $\mathcal{T}_1$ and $\mathcal{T}_2$ be two topologies on X. And suppose that for every $x \in X$ there exist bases of neighborhoods of x, $\mathcal{N}_1$ and $\mathcal{N}_2$, for $\mathcal{T}_1$ and $\mathcal{T}_2$ respectively, such that for every $V \in \mathcal{N}_1$ there is a $W \in \mathcal{N}_2$ such that $W \subset V$. Then $\mathcal{T}_2$ is finer than $\mathcal{T}_1$.*

Proof: We simply use Theorem 3.12. It is obvious that under the condition stated $\mathrm{id} : (X, \mathcal{T}_2) \to (X, \mathcal{T}_1)$ is continuous and the result follows. ∎

Corollary 3.16 *Let $\mathcal{T}_1$ and $\mathcal{T}_2$ be two topologies on X. Then, $\mathcal{T}_1 = \mathcal{T}_2$ iff for every $x \in X$ there are bases of neighborhoods of x, $\mathcal{N}_1$ and $\mathcal{N}_2$, for $\mathcal{T}_1$ and $\mathcal{T}_2$ respectively, such that for every $V \in \mathcal{N}_1$ there is a $W \in \mathcal{N}_2$, such that $W \subset V$, and, vice versa, there are bases of neighborhoods $\mathcal{M}_1$ and $\mathcal{M}_2$, for $\mathcal{T}_1$ and $\mathcal{T}_2$ respectively, such that for every $W \in \mathcal{M}_2$ there is a $V \in \mathcal{M}_2$, such that $V \subset W$.*

For metric spaces one can charatcerize the topology also by the set of all convergent sequences. Note that in general, the same topology can be obtained from many different metrics. Metrics leading to the same topology, i.e., the same set of open sets, are called *equivalent* .

Theorem 3.17 *Let $\mathcal{T}_1$ and $\mathcal{T}_2$ be two metric topologies on X, defined by metrics d_1 and d_2 respectively. Then, $\mathcal{T}_1 = \mathcal{T}_2$, i.e., d_1 and d_2 are equivalent metrics, iff $x_n \to x_0$ in $(X, d_1) \Leftrightarrow x_n \to x_0$ in (X, d_2).*

Proof: The proof is a direct application of Corollary 3.13. ■

For general topological spaces there is the analogous theorem with sequences replaced by nets.

3.4 Comments on non-metric topologies

1. The notion of Cauchy sequence is defined in terms of a metric. Therefore, also completeness is defined only for metric spaces.
2. Recall that the closure $\overline{A}$ of a set $A \in X$ is defined as the smallest closed set containing A. For metric spaces $\overline{A}$ can be obtained from A through addition of all the limits of converging sequences with elements in A (See Section 1.4). For a non-metric topological space this procedure might fail. In fact, the set thus obtained, sometimes called the *sequential closure* of A and denoted by $\overline{A}^S$, need not be closed. In general one has $A \subset \overline{A}^S \subset \overline{A}$. An example of $\overline{A}^S \neq \overline{A}$ is the following. Define

$$f_{m,n}(x) = [\cos(m!\pi x)]^{2n}$$

on $[0, 1]$, for all $m, n \geq 1$. The following pointwise limits are easy to compute:

$$\lim_{n\to\infty} f_{m,n}(x) := f_m(x) = \begin{cases} 1 & \text{if} \quad x = k/m!, k = 0, \ldots, m! \\ 0 & \text{else} \end{cases}$$

and

$$\lim_{m\to\infty} f_n(x) := f(x) = \begin{cases} 1 & \text{if} \quad x \in \mathbb{Q} \cap [0, 1] \\ 0 & \text{else} \end{cases} \tag{3.1}$$

For $A = \{f_{m,n} \mid m, n \geq 1\}$, we then have $f_n \in \overline{A}^S$, and certainly also

$$f \in \overline{\overline{A}^S}^S$$

The claim is that $f \notin \overline{A}^S$, and hence $\overline{A}^S \neq \overline{\overline{A}^S}^S$ and therefore, $\overline{A}^S \neq \overline{A}$. To show that $f \notin \overline{A}^S$ one can resort to the following proposition:

Proposition 3.18 *Any function $f : [0,1] \to \mathbb{R}$, that is the pointwise limit of continuous functions, is continuous on a dense subset of* $[0,1]$.

As f defined in (3.1) is nowhere continous in $[0,1]$, it is not the pointwise limit of some sequence of functions $f_{m,n}$, which are all continuous. This example also proves that the topology of pointwise convergence on the real-valued functions on $[0,1]$ is not metric.

4 More on metric spaces: completeness; the contraction mapping theorem

Recall that a metric space (X, d) is called *complete* iff every Cauchy sequence in X converges. Our next objective is the *Contraction Mapping Theorem* that we will first state and discuss and prove later in this section.

References for this section are
G.F. Simmons, *Introduction to Topology and Modern Analysis*, McGraw-Hill, New York, 1963,

and

J.E. Marsden, *Elementary Classical Analysis*, 2nd ed., W.H. Freeman and Company, New York, 1993.

Definition 4.1 (contraction) *Let (X, d) be a metric space. A map $T : X \to X$ is a contraction mapping (or contraction), if there exists a constant c, $0 \leq c < 1$, such that*

$$d(T(x), T(y)) \leq cd(x, y) \tag{4.1}$$

for all $x, y \in X$.

Exercise 4.1 *Prove that if T is a contraction on a metric space X, then T is uniformly continuous.*

Keep in mind that in some texts a map satisfying (4.1) with $c = 1$ is also called a contraction. Sometimes a map satisfying (4.1) for some $c < 1$ is then called *strictly contracting* or a a *strict contraction.*

Theorem 4.2 (Contraction Mapping Theorem) *If $T : X \to X$ is a contraction mapping and (X, d) is complete, then the equation*

$$T(x) = x \tag{4.2}$$

has exactly one solution $x \in X$.

A solution of (4.2) is called a *fixed point* of T. So, strict contractions have exactly one fixed point. The Contraction Mapping Theorem is only one example of what are more generally called fixed-point theorems. There are fixed-point theorems for maps satisfying (4.1) with $c = 1$, or even for arbitrary continuous transformations on certain metric spaces. In general, however, $c < 1$ is needed for uniqueness or even existence of a fixed point. E.g., consider the metric space $X = \{0, 1\}$, with the metric determined by $d(0, 1) = 1$. The map T defined by $T(0) = 1, T(1) = 0$, satisfies (4.1) with $c = 1$. But T does not have any fixed points. On the other extreme, the identity map on any metric space trivially satisfies (4.1) with $c = 1$, and every point is a fixed point.

It is worth noting that (4.2), and hence its solutions, do not refer to the metric d. If one finds a metric on X such that T is a contraction according to Definition 4.1, one can apply the theorem and obtain existence and uniqueness of the fixed point *if* X is complete for that same metric d. And there is the catch: the frustrating situation may occur that for none of the metrics for which one manages to prove that T is a contraction, X is complete. (X, d) might be complete for some metric, but for that metric T is then not a contraction. This can be an indication that the natural setting for the fixed point problem at hand is not X, but another, larger, space $(\tilde{X}, \tilde{d})$, called the *completion* of (X, d). Therefore, before we prove Theorem 4.2, we will discuss the notion of *completion* of a metric space.

4.1 Completion of metric spaces

A metric space $(\tilde{X}, \tilde{d})$ is called the *completion* of (X, d) (*the* completion, rather than *a* completion in anticipation of Theorem 4.3), if the following conditions are satisfied:

1. $(\tilde{X}, \tilde{d})$ is a complete metric space.
2. There is an embedding $i : X \to \tilde{X}$, satisfying
$$\tilde{d}(i(x), i(y)) = d(x, y) \tag{4.3}$$
for all $x, y \in X$.
3. $i(X)$ is dense in $\tilde{X}$, i.e., $\overline{i(X)} = \tilde{X}$.

A map i satisfying (4.3) is called *isometric*, or an *isometry.* Two metric spaces (X, d_X) and (Y, d_Y) are *isomorphic* if there exists an isometric bijection ($\equiv$ *isomorphism*) between them. They are then indistinguishable as metric spaces. Note that an isomorphism of metric spaces is also a homeomorphism of the corresponding topological spaces.

Note that isometries are automatically continuous and one-to-one. Some authors include the condition that the map be onto in the definition of isometry. With the latter convention isometries are isomorphisms of metric spaces. Do not let this confuse you. It is usually clear from the context whether an isometry is assumed to be onto or not.

We now have the terminology to state the theorem about the completion of metric spaces.

Theorem 4.3 *Every metric space (X, d) has a completion $(\tilde{X}, \tilde{d})$. The completion is unique up to isomorphism, i.e., any two completions $(\tilde{X}, \tilde{d})$ and $(\hat{X}, \hat{d})$ of (X, d) are isomorphic as metric spaces.*

Proof: We first construct a completion $(\tilde{X}, \tilde{d})$ of (X, d). Consider the set C of all Cauchy sequences in X, and define an equivalence relation on C by

$$x \sim y \quad \text{iff} \lim_{n\to\infty} d(x_n, y_n) = 0$$

for any two Cauchy sequences $x = (x_n), y = (y_n) \in C$. Note that if two Cauchy sequences x and y are convergent, then their limits are the same iff $x \sim y$. Verify that $\sim$ is indeed an equivalence relation. Let $\tilde{X}$ be the set of equivalence classes in C, i.e., $\tilde{X} = C/\sim$. Next, we define $\tilde{d} : \tilde{X} \times \tilde{X} \to \mathbb{R}$ by

$$\tilde{d}(\tilde{x}, \tilde{y}) = \lim_{n\to\infty} d(x_n, y_n) \tag{4.4}$$

where (x_n) and (y_n) are two representatives of $\tilde{x}$ and $\tilde{y}$ respectively. It is elementary to check that $(d(x_n, y_n))_{n\geq 1}$ is a Cauchy sequence of real numbers, and hence the limit exists. For this definition to make sense it is also crucial that the limit is independent of the chosen representatives. This can be checked as follows: suppose $(x_n), (x'_n) \in \tilde{x}$, and $(y_n), (y'_n) \in \tilde{y}$. Then, by the triangle inequality,

$$\begin{aligned} d(x_n, y_n) &\leq d(x_n, x'_n) + d(x'_n, y'_n) + d(y', y_n) \\ d(x_n, y_n) &\geq d(x'_n, y'_n) - d(x_n, x'_n) - d(y'_n, y_n) \end{aligned}$$

Taking the limit $n \to \infty$ of these inequalities, and using that $(x_n) \sim (x'_n)$, and $(y_n) \sim y'_n)$, we see that the limit in (4.4) is indeed independent of the representatives. So, $\tilde{d}$ is well-defined, and it is straightforward to check that it defines a metric on $\tilde{X}$.

Next, we check that the metric space $(\tilde{X}, \tilde{d})$ is indeed a completion of (X, d).
1) To check that $(\tilde{X}, \tilde{d})$ is complete, consider an arbitrary Cauchy sequence $(\tilde{x}_n)$ in $\tilde{X}$. We can chose a representative $(x_{n,k})_{k=1}^{\infty} \in \tilde{x}_n$ such that,

$$d(x_{n,k}, x_{n,l}) < \frac{1}{n} \quad \text{for all } k, l \geq n \tag{4.5}$$

This is possible because any representative is a Cauchy sequences and by selecting an appropriate subsequence, which is automatically also Cauchy and in the same equivalence class, we can make sure that (4.5) holds. We claim that the sequence $(x_{k,k})$ is a Cauchy sequence and that it is a representative of the limit of $(\tilde{x}_n)$. Indeed,

$$\begin{aligned} \lim_{n\to\infty} \tilde{d}((\tilde{x}_n), \widetilde{(x_{k,k})}) &= \lim_{n\to\infty} \lim_{k\to\infty} d(x_{n,k}, x_{k,k}) \\ &\leq \lim_{n\to\infty} \frac{1}{n} = 0 \end{aligned}$$

This shows that $(\tilde{X}, \tilde{d})$ is complete.
2) An isometric embedding $i : X \to \tilde{X}$ can be defined by mapping $x \in X$ to the

equivalence class of Cauchy sequences containing the constant sequence $x_n = x$. That this map is isometric follows from

$$\tilde{d}(i(x), i(y)) = \lim_{n\to\infty} d(x_n, y_n) \lim_{n\to\infty} d(x, y) = d(x, y)$$

3) $i(X)$ is dense in $\tilde{X}$ is seen as follows. Let $\tilde{x} \in \tilde{X}$ with representative (x_n). Because (x_n) is a Cauchy sequence the sequence of constant sequences $(y_{n,k} = x_n)_{k=1}^{\infty}$ satisfies

$$\lim_{n\to\infty} \tilde{d}(\tilde{(y_n)}, \tilde{(x_n)}) = 0$$

To conclude the proof of the theorem we still have to show that the completion is unique up to isomorphism, which we do in the following way. Let $(\tilde{X}, \tilde{d})$ be the completion of (X, d) constructed above and let $(\hat{X}, \hat{d})$ be any other completion, with its own isometric embedding $j : X \to \hat{X}$. We will extend j to a map $\tilde{j} : \tilde{X} \to \hat{X}$, in the sense that $\tilde{j} \circ i = j$. To define $\tilde{j}$ let us pick a representative (x_n) for any $\tilde{x} \in \tilde{X}$ and define

$$\tilde{j}(\tilde{x}) = \lim_{n\to\infty} j(x_n)$$

Because j is isometric $(j(x_n))$ is a Cauchy sequence in $\hat{X}$, and therefore converges. For the same reason the limit will be independent of the representative. It is also easy to see that j is isometric:

$$\hat{d}(\tilde{j}(\tilde{x}, \tilde{j}(\tilde{y})) = \lim_n \hat{d}(j(\tilde{x}_n), j(\tilde{y}_n)) = \lim_n d(x_n, y_n) = \tilde{d}(\tilde{x}, \tilde{y})$$

By chosing the constant sequences as representatives of $\tilde{x}, x \in X$ we see that, for $x \in X$, $\tilde{j}(i(x)) = j(x)$. To check that $\tilde{j}$ is onto, note that $\tilde{X}$ contains all Cauchy sequences in X. So, we have $\hat{X} = \overline{j(X)} \subset \tilde{j}(\tilde{X}) \subset \hat{X}$, which shows that $\tilde{j}$ is onto. ■

It is slightly annoying that $\tilde{X}$ is a space of sequences, rather than some more direct extension of X. E.g., there is no guarantee that the completion of a space of functions will still be a space of functions, just containing more functions than the original one. We will come back to this issue in concrete examples.

Examples:

1. $\mathbb{R}$ is the completion of $\mathbb{Q}$ with the usual metric. This is one of the standard ways of introducing (constructing, and mathematically defining) the real numbers. Note that when you write a real number in decimal notation, you in fact are giving a Cauchy sequence of rational numbers converging to it.

2. Consider the following metric on $C([a,b])$:

$$d(f,g) = \left[\int_a^b (f(x) - g(x))^2\,dx\right]^{1/2}$$

This is a very useful metric but $C([a,b])$, with this metric, is not complete. Its completion is denoted by $L^2([a,b])$, the space of all real-valued square-integrable functions on $[a,b]$. The dirty secret is that it is really not a space of functions, but a space of equivalence classes of (measurable) functions. Many of the usual operations with functions can still be defined, independent of the representative, but the evaluation of an element $f \in L^2([a,b])$ in a point, $f(x_0)$ *cannot* be unambiguously defined! We will meet L^2 again, and eventually come to think of it as a very nice space, consisting of objects as good as functions.

4.2 The Contraction Mapping Theorem

Recall the statement of Theorem 4.2: *If $T : X \to X$ is a contraction mapping and (X,d) is complete, then the equation $T(x) = x$ has exactly one solution $x \in X$.*
Proof: The proof is contructive in the sense that we will explicitly construct a sequence converging to the fixed point.

Let x_0 be any point in X and defined a sequence (x_n) by $x_1 = Tx_0$, $x_{n+1} = Tx_n$ (to alleviate the notations we often ommit the parentheses around the argument of a map). Denote by T^n the n^{th} power (n^{th} iterate) of T. So, $x_n = T^n x_0$.

First, we show that (x_n) is a Cauchy sequence. Let $n \geq m \geq 1$.

$$\begin{aligned}
d(x_n, x_m) &= d(T^n x_0, T^m x_0) \\
&\leq c^m d(T^{n-m} x_0, x_0) \\
&\leq c^m \left[d(T^{n-m}x_0, T^{n-m-1}x_0) + d(T^{n-m-1}x_0, T^{n-m-2}x_0) + \cdots + d(Tx_0, x_0)\right] \\
&\leq c^m \left[\sum_{k=0}^{n-m-1} c^k\right] d(x_1, x_0) \\
&\leq c^m \left[\sum_{k=0}^{\infty} c^k\right] d(x_1, x_0) \\
&\leq d(x_1, x_0)\frac{c^m}{1-c}
\end{aligned}$$

from which it follows that (x_n) is Cauchy. As X is complete x_n converges to a limit $x \in X$. That x is a fixed point of T follows using the continuity of T:

$$Tx = T\lim_n x_n = \lim_n Tx_n = \lim_n x_{n+1} = x$$

The uniqueness of the fixed point is also a consequence of the contraction property. Suppose x and y are two fixed points. Then

$$0 \leq d(x, y) = d(Tx, Ty) \leq c d(x, y)$$

As $c < 1$, this implies $d(x, y) = 0$, and hence $x = y$. So, the fixed point is unique. ■

5 Applications of the Contraction Mapping Theorem

5.1 An algorithm to compute $\sqrt{x}$

The Contraction Mapping Theorem (Theorem 4.2) not only proves the existence of a unique solution of certain equations of the form:

$$Tx = x$$

but also provides us with an algorithm to obtain arbitrarily good approximations of this solution (see the proof of Theorem 4.2). Here we apply the Contraction Mapping Theorem to a method of computing the square root of a real number.

Define a map $T : (0, \infty) \to (0, \infty)$ by

$$Ty = \frac{1}{2}(y + \frac{x}{y}) \tag{5.1}$$

where $x \in (0, \infty)$ is the number we want to compute the square root of. Clearly, $y = \sqrt{x}$ is a fixed point of T. And T has the following structure: let $0 < y_0$ be a first guess for $\sqrt{x}$, and suppose that $y_0^2 \leq x$, i.e., we guessed too small. Then,

$$\frac{x}{y_0} \geq \sqrt{x}$$

and, therefore, Ty, defined by (5.1), is the average of two approximations of $\sqrt{x}$, one that is too big and one that is too small. That should bring us closer to the correct value, and, hopefully, by iterating the process we obtain better and better approximations of $\sqrt{x}$. Before we start using this algorithm we would like to be sure that $T^n y_0 \to \sqrt{x}$. A way to to this is to set up T as a contraction on some metric space. Of course we are considering only real numbers here, and we are interested in approximations in the usual distance on the real numbers. The only remaining freedom is then the choice of a subset $I \subset \mathbb{R}$.

First, let us see whether T contracts at all: for $y_1, y_2 > 0$

$$\begin{aligned} |Ty_1 - Ty_2| &= \left| \frac{1}{2}\left(y_1 + \frac{x}{y_1}\right) - \frac{1}{2}\left(y_2 + \frac{x}{y_2}\right) \right| \\ &= \frac{1}{2}|y_1 - y_2| \left| 1 - \frac{x}{y_1 y_2} \right| \end{aligned}$$

So, T is a contraction if one necessarily has $x < 3y_1y_2$, which means we have to exclude arguments y that are too small. Therefore, we try a metric space of the form $([a,\infty),d)$, with $a>0$. This is a complete metric space because $[a,\infty)\subset\mathbb{R}$ is closed, and $(\mathbb{R},d)$ is complete (see Assignment 6, Problem 1).

In order to make a good choice for a we first observe that

$$Ty=\sqrt{x}+\frac{(y-\sqrt{x})^2}{2y}\geq\sqrt{x} \tag{5.2}$$

for all $y>0$. And for $y_1,y_2\geq\sqrt{x}$ we have $|1-x/(y_1,y_2)|<1$ and hence

$$|Ty_1-Ty_2|\leq\frac{1}{2}|y_1-y_2|$$

Therefore, the restriction of T to $[\sqrt{x},\infty)$ is well-defined (meaning that $([\sqrt{x},\infty))\subset[\sqrt{x},\infty))$, and is a nice contraction ($c=1/2$). We conclude that for any $y_0>0$, $y_i=Ty_0\in[\sqrt{x},\infty)$ and $y_n:=T^ny_0\to\sqrt{x}$. Moreover, the convergence is exponentially fast (see the proof of Theorem 4.2):

$$\begin{aligned}|T^ny_0-\sqrt{x}| &\leq |T^ny_0-T^my_0|\\ &\leq \frac{c^n}{1-c}|Ty_0-y_0|\\ &\leq \frac{1}{2^n}|\frac{x}{y_0}-y_0|\end{aligned}$$

Exercise 5.1 *a) Show that for any $y>0$, the convergence of T^ny to $\sqrt{x}$ is in fact faster than exponential. Start from (5.2) to obtain a good estimate for $|Ty-\sqrt{x}|$.*
b) Take $y=1$ and $x=2$. Find a good estimate of how large n needs to be for T^n1 and $\sqrt{2}$ to have identical first d digits. E.g., how many iterations of the map T are sufficient to compute the first 63 digits of $\sqrt{2}$?

5.2 Discrete Dynamical Systems

The term dynamical system refers to models for the evolution in time of the state of some "real life" system, as they arise in many disciplines: physics, chemistry, engineering. biology, physiology, population dynamics, economics and finance, … . In fact, they can also be an auxiliary tool for solving other problems in mathematics, then sometimes referred to as "symbolic" dynamics. A dynamical system is then defined by a state space X (often a metric space), whose elements describe the different states the system can be in, and a prescription relating the state $x_t\in X$ at time t to the state at previous (and/or subsequent times). E.g, this could be given as a differential equation, leading to a function $t\mapsto x_t$ for a suitable initial condition x_0.

The term discrete refers to the fact that time is taken to discrete, i.e., integer multiples of some time unit, i.e., we can take $t=n\in\mathbb{N}$. The dynamics is then usually defined by a map T relating the state x_{n+1} at time $t = n+1$ to the state x_n at time $t = n$:

$$x_{n+1} = Tx_n$$

If T is invertible, the dynamics can be defined for all $t = n \in Ir$.

One of the simplest and at the same time most famous dynamical systems is the *logistic map* of population dynamics. The state space is $[0, \infty) \subset \mathbb{R}$, representing the size of the population, and the dynamics is given by

$$Tx = 4\mu x(1-x)$$

where $\mu \geq 0$ is a parameter. The non-linear dependence of x_{n+1} on x_n expresses the fact that large populations lead to large off-spring, but too large populations have an adverse effect on the reproduction process.

The importance of this model is primarily that it convincingly illustrates the fact that very simple maps can have amazingly complex behavior. Analyzing the full variety of behavior of the logistic map is beyond the scope of this modest application of the Contraction Mapping Theorem. You can read more about it in : M. Schroeder, *Fractals, Chaos, Power Laws*, or R. Devaney, *An Introduction to Chaotic Dynamical Systems*.

Fixed points of the map T describe steady states (equilibrium states) of the system. If T is a contraction on a complete metric space the Contraction Mapping Theorem implies that there is a unique steady state, and that starting from any initial state, the system will evolve toward this steady state as time goes on. In the case of the logistic map we have the following.

Obviously, 0 is a fixed point of T. It is easy to show that when $\mu < 1/4$, T is a contraction on $[0, 1]$. The Contraction Mapping Theorem then tells us that for such values of μ, 0 is the unique fixed point, and that for an arbitrary initial population $x_0 \in [0, 1]$, $T^n x_0 \to 0$, i.e., the population dies out. For $\mu > 1/4$, but not too large, there is a second, non-zero, fixed point. As μ increases further, even more interesting things happen.

5.3 Integral Equations

Consider the (linear) Volterra equation for an unknown function $f : [a, b] \to \mathbb{R}$:

$$f(x) = g(x) + \int_a^b k(x, y) f(y)\, dy \tag{5.3}$$

where $g \in C([a, b])$ and $k : [0, 1] \times [0, 1] \to \mathbb{R}$, are given functions. The integral equation (5.3) is a fixed point equation, i.e., of the form $Tf = f$, for the map T defined by

$$Tf(x) = g(x) + \int_a^b k(x, y) f(y)\, dy \tag{5.4}$$

Theorem 5.1 *If there exists a constant c satisfying*

$$\sup_{a \le x \le b} \int_a^b |k(x,y)|\, dy \le c < 1 \tag{5.5}$$

then there is a unique continuous function f satisfying (5.3).

Proof: We prove this result by showing that, when (5.5) is satisfied, the map T is a contraction on the complete metric space $(C([a,b]), |\cdot\|_\infty)$ (see Theorem 2.10). For any $f_1, f_2 \in C([a,b])$ we have

$$\begin{aligned}
\|Tf_1 - Tf_2\|_\infty &= \sup_{a \le x \le b} \left| \int_a^b k(x,y)(f_1(y) - f_2(y))\, dy \right| \\
&\le \sup_{a \le x \le b} \int_a^b |k(x,y)|\, |f_1(y) - f_2(y)|\, dy \\
&\le \|f_1 - f_2\|_\infty \sup_{a \le x \le b} \int_a^b |k(x,y)|\, dy \\
&\le c\|f_1 - f_2\|_\infty
\end{aligned}$$

■

From the proof of the Contraction Mapping Theorem we also know that, when $c < 1$, we can obtain the unique fixed point f as a limit:

$$f = \lim_{n \to \infty} T^n f_0 \tag{5.6}$$

for any $f_0 \in C([a,b])$. It is interesting to reinterprete this limit as a series. Define a map $K : C([a,b]) \to C([a,b])$ by

$$Kf = \int_a^b k(x,y) f(y)\, dy$$

One has that $\|K^n f_0\|_\infty \le c^n |f_0\|_\infty \to 0$. Then, (5.3) becomes $f = g + Kf$, or

$$(\mathbb{1} - K)f = g \tag{5.7}$$

where $\mathbb{1}$ is the identity map, i.e., $\mathbb{1}f = f$. Solving (5.7) for f, given an arbitrary g, is equivalent to inverting the map $\mathbb{1} - K$, and (5.6) can be reinterpreted as the solution of this problem:

$$f = (\mathbb{1} - K)^{-1} g = \sum_{n=0}^{\infty} K^n g \tag{5.8}$$

This series is called the *Neumann series*, and using its partial sums to approximate f is often called the *Born approximation.* Note the similarity of (5.8) with the geometric series $(1-x)^{-1} = \sum_{n=0}^{\infty} x^n$, for $|x| < 1$. In fact, we will see later that it *is* a geometric series that is absolutey convergent for a natural norm on the linear operators on a Banach space. This also explains why we only needed a condition on k, and none on g: (5.5) is the condition needed for $\mathbb{1} - K$ to be invertible, which does not involve g.

5.4 Differential Equations (Boundary Value Problems)

Consider a copper rod or pipe, say of length 1 in some units, and partly insulated. The ends of the rod are kept at constant temperatures T_0 and T_1. Denote by x the coordinate along the length of the pipe ($x \in [0,1]$), and let t be time. We are interested in the temperature $u(x,t)$ as it varies along the rod, and with time. To a reasonable degree of accuracy $u(x,t)$ satisfies

$$u_t = u_{xx} - q(x)u \tag{5.9}$$

where $q(x)$ is a given function describing the cooling rate at x, and

$$\begin{aligned} u_t(x,t) &= \frac{\partial}{\partial t} u(x,t) && (5.10) \\ u_{xx}(x,t) &= \frac{\partial^2}{\partial x^2} u(x,t) && (5.11) \end{aligned}$$

is a standard notation for partial derivatives. Of course, this equation does not uniquely determine u by itself. It has to be supplemented with an inital condition $u_0(x)$, and boundary conditions, e.g., $u(0,t) = T_0, u(1,t) = T_1$, for all t. If one is interested in time independent solutions $u(x,t) = u(x)$, e.g., describing the situation as $t \to \infty$, one has to solve the following ODE (put $u_t = 0$ in the PDE (5.9)) + boundary conditions:

$$\begin{aligned} -u''(x) + q(x)u(x) &= 0 && (5.12) \\ u(0) = T_0, \quad u(1) &= T_1 && (5.13) \end{aligned}$$

If $q(x)$ is a constant function, (5.12) is easy to solve explicitly. But if q is not constant explicit integration of this equation is in general impossible. We will show that, when $\|q\|_\infty$ is not too large, the Contraction Mapping Theorem implies the existence of a unique solution, and provides a method for calculating approximations of that solution.

First, it is convenient to get rid of the inhomogeneous boundary conditions (5.13), by deriving an equation for a function v related to u by

$$u(x) = T_0 + (T_1 - T_0)x + v(x)$$

v then satisfies

$$v''(x) = q(x)v(x) + (T_0 + (T_1 - T_0)x)q(x) \tag{5.14}$$
$$v(0) = 0, \quad v(1) = 0 \tag{5.15}$$

Next, we will turn this problem into a fixed point problem of the Volterra type with the aid of the following proposition.

Proposition 5.2 *Let $f : [0,1] \to \mathbb{R}$ be a continuous function, then the unique solution v of*

$$-v'' = f \tag{5.16}$$
$$v(0) = v(1) = 0 \tag{5.17}$$

can be given as

$$v(x) = \int_0^1 g(x,y)f(y)\,dy \tag{5.18}$$

where

$$g(x,y) = \begin{cases} x(1-y) & \text{if } 0 \le x \le y \le 1 \\ y(1-x) & \text{if } 0 \le y \le x \le 1 \end{cases} \tag{5.19}$$

We postpone the proof of this proposition to the end of this section and first see where it takes us. Using this proposition, (5.14) can be written in the form

$$v(x) = -\int_0^1 g(x,y)q(y)v(y)\,dy - \int_0^1 g(x,y)(T_0 + (T_1 - T_0)y)q(y)\,dy$$

which is of the form

$$(\mathbb{1} - K)v = h$$

with

$$Kv(x) = -\int_0^1 \underbrace{g(x,y)q(y)}_{:=k(x,y)} v(y)\,dy \tag{5.20}$$
$$h(x) = -\int_0^1 g(x,y)(T_0 + (T_1 - T_0)y)q(y)\,dy \tag{5.21}$$

We can now apply Theorem 5.1 to obtain the following result.

Theorem 5.3 *If*

$$\sup_{0 \le x \le 1} \int_0^1 |g(x,y)q(y)|\,dy < 1 \tag{5.22}$$

the boundary value problem (5.12)-(5.13) has a unique continuous solution.

Note that it is not necessary that q be continuous, as long as h is and the condition (5.22) is satisfied.

We still have to prove Proposition 5.2.

Proof: (Proposition 5.2)
By integrating (5.16) twice we obtain:

$$v(x) = -\int_0^x \underbrace{\int_1^y f(s)\,ds}_{:=a(y)}\,dy + C_1 x + C_2$$

where C_1 and C_2 are two real constants to be determined later. Integration by parts of the integral over y in the RHS gives

$$\begin{aligned} v(x) &= -\left[ya(y)\right]_0^x + \int_0^x yf(y)\,dy + C_1 x + C_2 \\ &= x\int_x^1 f(y)\,dy + \int_0^x yf(y)\,dy + C_1 x + C_2 \end{aligned}$$

We can determine C_1 and C_2 by imposing (5.17). One finds

$$\begin{aligned} C_1 &= -\int_0^1 yf(y)\,dy \\ C_2 &= 0 \end{aligned}$$

It follows that

$$v(x) = \int_0^x y(1-x)f(y)\,dy + \int_x^1 x(1-y)f(y)\,dy$$

which is what we had to prove. ■

5.5 Differential Equations (Existence and Uniqueness Theorem)

Another important application of Contraction Mapping Theorem to differential equations is an Existence and Uniqueness Theorem for initial value problems of the type

$$\begin{cases} \dot{x}\,(t) = F(t, x(t)) \\ x(0) = x_0 \end{cases} \tag{5.23}$$

with F a Lipschitz continuous function on a suitable domain. Here, x takes values in $\mathbb{R}^n$, and so do its derivative and F. (5.23) is often presented as a set of coupled ODE's, e.g., describing a dynamical system. (2.22) can be reformulated as an integral equation:

$$x(t) = x_0 + \int_{t_0}^t F(t, x(t))\,dt \tag{5.24}$$

By the fundamental theorem of calculus the continuous solutions of (5.23) and (5.24) coincide. (5.24) is easily written as a fixed point equation for the map T defined by

$$Tx(t) = x(t) = x_0 + \int_{t_0}^{t} F(s, x(s))\, ds \tag{5.25}$$

We are then looking for solutions of

$$Tx = x \tag{5.26}$$

If we can find sufficient conditions that guarantee that T is a contraction on a suitable complete metric space of continuous functions, the Contraction Mapping Theorem immediately implies that (5.26) has a unique solution in that space.

Theorem 5.4 *Let $I \subset \mathbb{R}$ be a (possibly infinite) interval, and t_0 a point in the interior. Suppose $F : I \times \mathbb{R}^n \to \mathbb{R}^n$ is continuous in both arguments and uniformily Lipschitz continuous in its second argument, i.e., suppose there exists a constant $C > 0$ such that for all, $t \in I$, and $x, y \in \mathbb{R}^n$,*

$$\|F(t,x) - F(t,y)\| \leq C\|x - y\| \tag{5.27}$$

Then, there is a unique continuous function $x : I \to \mathbb{R}^n$ solving (5.23).

Proof: As mentioned above, our strategy is to show that T, defined in (5.25), is a contraction on a suitable metric space of continuous functions. Note that T depends on t_0 and t. It will be necessary to first consider solutions for sufficiently short times, i.e., $t \in [t_0, t_0 + \delta] \subset I$, with $\delta > 0$ to be determined later. Let $f, g : [t_0, t_0 + \delta] \to \mathbb{R}^n$, be two continuous functions. Then

$$\begin{aligned}
\|Tf - Tg\|_{\sup} &= \sup_{t_0 \leq t \leq t_0 + \delta} |Tf(t) - Tg(t)| \\
&= \sup_{t_0 \leq t \leq t_0 + \delta} \left| \int_{t_0}^{t} F(s, f(s)) - F(s, g(s))\, ds \right| \\
&\leq \sup_{t_0 \leq t \leq t_0 + \delta} \int_{t_0}^{t} |F(s, f(s)) - F(s, g(s))|\, ds \\
&\leq \sup_{t_0 \leq t \leq t_0 + \delta} \int_{t_0}^{t} C|f(s) - g(s)|\, ds \\
&\leq \delta C \|f - g\|_{\sup}
\end{aligned}$$

where we have used the definition of T and the Lipschitz condition (5.27). Now, it is clear that T is a contraction on $C([t_0, t_0 + \delta])$ with the supremum norm as long as $|t - t_0| \leq \delta < 1/C$. Therefore, there is a unique solution $x : [t_0, t_0 + \delta] \to \mathbb{R}^n$. The argument holds for *any* $t_0 \in I$. By covering I with overlapping intervals of length $< 1/C$, we see that (5.23) has a unique continuous solution defined on all of I. ∎

Note that, due to (5.24), the solution is in fact C^1 (continosly differentiable).

Examples:

1. Linear ODE with continuous coefficients have unique global solutions. It is sufficient to consider *first order* ODE's (Why?):

$$\begin{aligned} \dot{x}\,(t) &= a(t)x(t) + b(t) \\ x(0) &= x_0 \end{aligned}$$

   where $a : \mathbb{R} \to \mathbb{R}^{n \times n}$ is a continuous matrix-valued function (with respect to any norm on the $n \times n$ matrices with real entries), and $b : \mathbb{R} \to \mathbb{R}^n$ is continuous as well. Then, for any bounded interval $I \subset \mathbb{R}$, there exists a constant C such that

$$\|a(t)v\| \leq C\|v\|, \quad \text{for all } t \in I, v \in \mathbb{R}^n$$

   and it is then obvious that

$$\|a(t)f(t) - a(t)g(t)\|_{\text{sup}} \leq C\|f - g\|_{\text{sup}}$$

   Therefore, the hypotheses of Theorem 5.4 are satisfied and we have a unique continuous solution on I. As I was an arbitrary interval, we can conclude that there exists a unique continuous solution for all $t \in \mathbb{R}$.

2. The applications of Theorem 5.4 are by no means limited to linear ODE. E.g., consider the nonlinear scalar ODE given by

$$\begin{aligned} \dot{x}\,(t) &= \sqrt{a(t)^2 + x(t)^2} =: F(t, x(t)) \\ x(0) &= x_0 \end{aligned}$$

   where $a : \mathbb{R} \to \mathbb{R}$ is a continuous function. It is not hard to prove that

$$|F(t,x) - F(t,y)| \leq |x - y|, \quad \text{for all } t, x, y, \in \mathbb{R}$$

   Again we can conclude by Theorem 5.4 that there is a unique global solution of this ODE.

3. Nonlinear F can easily be continuous without being uniformly Lipschitz continuous. A simple example of this situation shows that global solutions may not exist in this case.

$$\begin{aligned} \dot{x} &= x^2 \\ x(0) &= 1 \end{aligned}$$

It is easy to check that

$$x(t) = \frac{1}{1-t}\,, \quad t \in [0,1)$$

is a solution. The singularity at $t = 1$ is possible because the RHS is not uniformly Lipschitz continuous:

$$|x^2 - y^2| = |x+y|\;|x-y|$$

and $|x+y|$ can become arbitrarily large. In this simple example it is easy to understand that the quadratic RHS makes x grow faster the larger it becomes (positive, nonlinear feedback). This leads to blowup of the solution in finite time, at which point the equation is not well-defined.

6 Banach Spaces

In this section we focus on a very important class of complete metric spaces called Banach spaces. A *Banach Space* is linear space V, equipped with a norm, usually denoted by $\|\cdot\|$, that is a complete metric space for the metric derived from this norm ($d(x,y) = \|x-y\|$).

If you do not remember much about linear spaces, see e.g.,
P.D. Lax, *Linear Algebra*, John Wiley & Sons, New York 1997
or
P.R. Halmos, *Finite-dimensional Vector Spaces*, 2nd ed D. Van Nostrand Company, Princeton 1958.

For more information on Banach spaces see
W. Rudin, *Functional Analysis*, 2nd ed, McGraw-Hill Inc., New York 1991
or
G.F. Simmons, *Introduction to Topology and Modern Analysis*, McGraw-Hill, New York, 1963.

Recall that a *norm* on a linear space V over the field of scalars $\mathbb{R}$ (or $\mathbb{C}$) is a function $\|\cdot\| : V \to \mathbb{R}$ with the following properties:
(N1) $\|x\| \geq 0$, for all $x \in V$. (positivity)
(N2) $\|\lambda x\| = |\lambda|\|x\|$, for all $x \in V$ and $\lambda \in \mathbb{R}$ (or $\mathbb{C}$). (homogeneity)
(N3) $\|x+y\| \leq \|x\| + \|y\|$, for all $x, y \in V$. (triangle inequality)
(N4) $\|x\| = 0 \Rightarrow x = 0$. (strict positivity)
If the (N1–3) hold but not necessarily (N4), $\|\cdot\|$ is called a *seminorm*.

All linear spaces that we will consider will be either over $\mathbb{R}$ or $\mathbb{C}$, respectively called *real* and *complex* linear spaces.

6.1 Bounded linear maps

Let X and Y be two normed linear spaces (not necessarily complete). As usual we will denote both norms by $\|\cdot\|$.

Definition 6.1 *A linear map $T : X \to Y$ is bounded if there exists a constant $M \geq 0$ such that*

$$\|Tx\| \leq M\|x\|, \quad \text{for all } x \in X \tag{6.1}$$

Theorem 6.2 *A linear map $T : X \to Y$ is bounded iff it is continuous.*

Proof:
1) Bounded $\Rightarrow$ Continuous. For all $x, y \in X$ we have

$$\|Tx - Ty\| = \|T(x-y)\| \leq M\|x-y\|$$

where M is a constant for which (6.1) holds. Therefore, one can take $\delta = \epsilon/M$ in the definition of continuity.
2) Continuous $\Rightarrow$ Bounded. If T is continuous, then T is continuous at $0 \in X$. So, for all $\epsilon > 0$, there exists $\delta > 0$ such that $\|x\| \leq \delta \Rightarrow \|Tx - T0\| = \|Tx\| \leq \epsilon$. Pick any $\delta > 0$. Then, for any $0 \neq x \in X$, $\tilde{x}$ defined by

$$\tilde{x} = \frac{\delta}{\|x\|} x$$

satisfies $\|\tilde{x}\| \leq \delta$. Therefore, $\|T\tilde{x}\| \leq \epsilon$. It follows that for all $x \in X$

$$\|Tx\| = \|\frac{\|x\|}{\delta} T\tilde{x}\| \leq \frac{\|x\|}{\delta}\epsilon$$

Hence, T satisfies Definition 6.1 with $M = \epsilon/\delta$. ■

Note that we have also proved that if T is continuous at $0 \in X$, then T is continuous on the whole of X. This can also be seen as a direct consequence of linearity.

The set of all bounded linear maps $T : X \to Y$ is denoted by $\mathcal{B}(X, Y)$.

Next, we define the *null space* or *kernel* of a linear map $T \in \mathcal{L}(X, Y)$, denoted by $\ker T$, and the *range* of T, denoted by $\mathrm{ran}T$.

$$\begin{aligned} \ker T &= \{x \in X \mid Tx = 0\} \\ \mathrm{ran}T &= \{y \in Y \mid \text{ there exists } x \in X, \text{ such that } Tx = y\} \end{aligned}$$

Theorem 6.3 *For any $T \in \mathcal{B}(X, Y)$, $\ker T$ is a closed linear subspace of X.*

Proof:
• $\ker T$ *is a linear subspace:* For all $\lambda \in \mathbb{R}$ or $\mathbb{C}$, and $x, y \in \ker T$, one has

$$T(x + \lambda y) = Tx + \lambda Ty = 0$$

and therefore, $x + \lambda y \in \ker T$ (here, we only used the fact that T is linear, not that it is bounded).
• $\ker T$ *is closed:* As X is a metric space, we only have to show that for any sequence of elements $x_n \in \ker T$, converging to x, we necesarily have $x \in \ker T$. This follows from the continuity of T:

$$Tx = T \lim_n x_n = \lim_n Tx_n = \lim_n 0 = 0$$

■

When $X = Y$ one usually simplifies the notation: $\mathcal{L}(X, X) = \mathcal{L}(X)$, and $\mathcal{B}(X, X) = \mathcal{B}(X)$. In that case the linear maps T are referred to as *linear transformations* of X, or *linear operators* on X.

Another important special case is when $Y = \mathbb{R}$ or $\mathbb{C}$ (depending on whether X is a real or complex linear space). The elements of $\mathcal{L}(X,\mathbb{C})$ and $\mathcal{B}(X,\mathbb{C})$ are then called linear *functionals* (or linear forms), and bounded ($\equiv$ continuous) linear functionals on X, respectively. $\mathcal{L}(X,\mathbb{C})$ is also called the *algebraic dual*, and $\mathcal{B}(X,\mathbb{C})$ the *topological dual* space of X. Somewhat confusingly, both are commonly denoted by X^*.

Examples:

1. Consider $X = C([a,b])$. For any $\rho \in L_1([a,b])$, the following formula defines a continuous linear functional φ on X:

$$\varphi(f) = \int_a^b f(x)\rho(x)\,dx \tag{6.2}$$

2. Not all continuous linear functionals are of the form (6.2). E.g., for any $x_0 \in [a,b]$, evaluation of f at x_0 is a continuous linear functional. I.e., define

$$\delta_{x_0}(f) = f(x_0),$$

then δ_{x_0} is a continuous linear functional on $C([a,b])$.

3. An example of a discontinuous linear functional on $C([0,1])$ is the following. It is easy to see that the set of monomials $M = \{x^n \mid n \geq 0\}$ is a linearly independent set of vectors in $C([0,1])$. By a general theorem (see, e.g., the reference by Simmons given at the beginning of this section), $C([0,1])$ has a Hamel basis , H, containing M. A *Hamel basis*, is a maximal linearly independent set of vectors. Any f, can be uniquely written as finite linear combination of vectors in H, i.e., for any $f \in C([0,1])$, there is a unique set of elements, $h_i \in H$, $i = 1,\ldots,n$, and unique non-zero coeffcients c_i such that

$$f = c_1 h_1 + \cdots c_n h_n \tag{6.3}$$

For each $n \geq 0$ define $a_n(f)$ by

$$a_n(f) = \begin{cases} c_i & \text{if } h_i = x^n \\ 0 & \text{else} \end{cases}$$

Due to the uniqueness of the decompositon this is well-defined. Define a linear functional φ on $C([0,1])$ by

$$\varphi(f) = \sum_{n=1}^{\infty} n a_n(f)$$

As, for each f, only a finite number of terms in this sum are non-zero, this is well-defined. However, it is quite obvious that φ is unbounded. E.g.,

$$\sup\{|\varphi(f)| \mid \|f\| = 1\} \geq \sup\{|\varphi(f)| \mid f \in M\} = \sup\{n \geq 1\} = +\infty$$

Hence, by Theorem 6.2, this implies that φ is discontinuous. In fact, with this example we have shown that every infinite-dimensional linear space has discontinuous linear functionals defined on it. On the other hand, as will be shown in Theorem 6.13, all linear functionals on a finite-dimensional linear space are continuous.

6.2 Topologies on $\mathcal{B}(X, Y)$

Let X and Y be normed linear spaces. Then, for any $T \in \mathcal{B}(X, Y)$ the following defines a norm on $\mathcal{B}(X, Y)$:

$$\|T\| = \inf\{M \mid \|Tx\| \leq M\|x\|, \text{for all } x \in X\} \tag{6.4}$$

Check that this indeed defines a norm. An alternative expression for $\|T\|$ is

$$\|T\| = \sup\{\|Tx\| \mid \|x\| \leq 1\} \tag{6.5}$$

When $X \neq \{0\}$ one also has

$$\|T\| = \sup\{\|Tx\| \mid \|x\| = 1\} \tag{6.6}$$

$$\|T\| = \sup\{\frac{\|Tx\|}{\|x\|} \mid x \neq 0\} \tag{6.7}$$

This norm on $\mathcal{B}(X, Y)$ is called the *uniform* or *operator* norm, or sometimes just the *usual* norm. And if for a sequence of $T_n \in \mathcal{B}(X, Y)$ one has $\lim_n \|T_n - T\| = 0$, for some $T \in \mathcal{B}(X, Y)$, one says that T_n converges uniformly to T, or that it converges in the uniform (operator) topology.

It is clear that $\mathcal{L}(X)$ is not only a linear space, but that it is also an algebra with the product given by the composition of linear maps. The following theorem implies that $\mathcal{B}(X)$ is a subalgebra of $\mathcal{L}(X)$, i.e., the product of two bounded linear maps is bounded.

Theorem 6.4 *Let X, Y, and Z be normed linear spaces. For any $T \in \mathcal{B}(X, Y)$ and $S \in \mathcal{B}(Y, Z)$, one has that $ST \in \mathcal{B}(X, Z)$, and*

$$\|ST\| \leq \|S\| \, \|T\| \tag{6.8}$$

Proof: For all $x \in X$ we have

$$\|STx\| \leq \|S\| \, \|Tx\| \leq \|S\| \, \|T\| \, \|x\|$$

■

Note that (6.8) also shows that the product of bounded linear maps is a continuous map $\mathcal{B}(Y, Z) \times \mathcal{B}(X, Y) \to \mathcal{B}(X, Z)$.

Theorem 6.5 *If Y is a Banach space, then $\mathcal{B}(X, Y)$ is a Banach space for the uniform norm.*

Proof: We have to prove that every Cauchy sequence in $\mathcal{B}(X, Y)$ converges (to a limit in $\mathcal{B}(X, Y)$). Let (T_n) be a Cauchy sequence. We have that for all $x \in X$

$$\|T_n x - T_m x\| \leq \|T_n - T_m\| \, \|x\|$$

which shows that the sequence $(T_n x)$ is a Cauchy sequence in Y. As Y is complete, there exists $y \in Y$ such that $T_n x \to y$. It is then straightforward to check that $Tx = y$ defines a linear map $X \to Y$. Next, we show that T is bounded. For any $\epsilon > 0$, let N_ϵ be such that $\|T_n - T_m\| < \epsilon/2$ for all $n, m \geq N_\epsilon$. Take $n \geq N_\epsilon$. Then, for all $x \in X$, $\|x\| = 1$, there exist $m(x) \geq N_\epsilon$ such that

$$\|T_n x - Tx\| \leq \|T_n x - T_{m(x)} x\| + \|T_{m(x)} x - Tx\| \leq \epsilon/2 \|x\| + \epsilon/2 \leq \epsilon \qquad (6.9)$$

Therefore, for all $x \in X$, $\|x\| = 1$, $\|Tx\| \leq \|T_n x\| + \epsilon \leq \|T_n\| + \epsilon$, and T is bounded. Finally, from (6.9) it also follows that $\lim_n \|T_n - T\| = 0$, Hence, $T_n \to T$ in the uniform norm. ■

Exercise 6.1 *Prove that if $T_n \to T$ uniformly, then $\|T_n\| \to \|T\|$.*

The following theorem says that a bounded linear map has a unique bounded extension to the closure (or completion) of its domain. Theorem is very useful when one defines linear maps (one only has to define them on a dense subspace) or when one has to show that two linear maps are equal (one has to show equality only on a dense subspace).

Theorem 6.6 *Let X be a normed linear space and let Y be a Banach space. For any dense linear subspace $X_0 \subset X$ and any $T_0 \in \mathcal{B}(X_0, Y)$, there exists a unique $T \in \mathcal{B}(X, Y)$ such that $Tx_0 = T_0 x_0$, for all $x_0 \in X_0$.*

Proof: We first define a map $T \in \mathcal{B}(X, Y)$ as follows. For all $x \in X$, there is a sequence (x_n) in X_0 converging to x. Define

$$Tx = \lim_n T_0 x_n$$

This limit exists because T_0 is bounded and Y is complete. Also, any two sequences (x_n) and (x'_n) converging to x, satisfy

$$\lim_n \|x_n - x'_n\| \leq \lim_n \|x_n - x\| + \lim_n \|x - x'_n\|$$

and therefore

$$\|\lim_n T_0 x_n - \lim_n T_0 x'_n\| = \lim_n \|T_0 x_n - T_0 x'_n\| \leq \lim_n \|T_0\| \, \|x_n - x'_n\| = 0$$

and Tx is independent of the choice of the sequence (x_n) converging to x.

T is an extension of T_0, i.e., $Tx_0 = T_0 x_0$, for all $x_0 \in X_0$, because we can take the constant sequence $x_n = x_0$, for all n, to define Tx_0. The linearity of T is a consequence of the linearity of the T_n, the fact that X_0 is a linear subspace, and the independence of the chosen sequence that we just proved.

Next, we prove that T is bounded. It follows from

$$\|Tx\| = \lim_n \|T_0 x_n\| \leq \lim_n \|T_0\| \, \|x_n\| = \|T_0\| \, \|x\|$$

that $\|T\| \leq \|T_0\|$. In fact, as $Tx_0 = T_0 x_0$, for $x_0 \in X_0$, we must have $\|T\| = \|T_0\|$.

Finally, we show that T is the unique bounded linear map $X \to Y$, coinciding with T_0 on X_0. Suppose S is another such map. Then, for all $x \in X$, and sequences $x_n \to x$, $x_n \in X_0$,

$$Sx = \lim_n Sx_n = \lim_n T_0 x_n = Tx$$

For the first inequality we used that S is bounded by assumption, for the second that S is an extension of T_0, and the third is the definition of T. ∎

The following exercise asks you to think about the importance of the linearity of T_0 in the previous theorem. Is there a property that can replace it for more general functions?

Exercise 6.2 *What can one say about continuous extensions of functions $f : X_0 \to Y$, where X_0 is a dense subset of a metric space X and Y is a complete metric space?*

Next, we look at an example of a convergent sequence of bounded linear maps on $C([0, 1])$, and also an example of a divergent sequence, which, however, convergence in a different topology on $\mathcal{B}(C([0, 1]))$.

Example:
Let $X = C([0,1])$, equipped with the supremum norm. For $n = 1, 2, \ldots$, let $k_n(x,y)$ be a sequence of real-valued continuous functions on $[0,1] \times [0,1]$ such that

$$\lim_n \left\{ \sup_{x \in [0,1]} \int_0^1 |k_n(x,y)|\, dy \right\} = 0 \tag{6.10}$$

For $f \in X$, define $K_n \in \mathcal{B}(X)$ by

$$K_n f(x) = \int_0^1 k_n(x,y) f(y)\, dy \tag{6.11}$$

Then $K_n \to 0$ in the uniform norm. Indeed, for all $f \in X$, we have

$$\begin{aligned} |K_n f(x)| &= \left| \int_0^1 k_n(x,y) f(y)\, dy \right| \\ &\leq \int_0^1 |k_n(x,y)|\, |f(y)|\, dy \\ &\leq \|f\| \int_0^1 |k_n(x,y)|\, dy \\ &\leq \|f\| \sup_x \int_0^1 |k_n(x,y)|\, dy \end{aligned}$$

which, by (6.10) implies that

$$\|K_n\| \leq \sup_{x \in [0,1]} \int_0^1 |k_n(x,y)|\, dy \to 0$$

An example of functions k_n satisfying (6.10) is

$$k_n(x,y) = xy^n$$

Next, Consider another sequence of linear transformations $K_n \in \mathcal{B}(X)$, of the form (6.11), but with the functions k_n given by

$$k_n(x,y) = \sin(n\pi y)$$

It is not hard to prove (see the Riemann-Lebesgue Lemma below), that, for any $f \in C([0,1])$

$$\lim_n \int_0^1 \sin(n\pi y)\, f(y)\, dy = 0$$

And, therefore,

$$\lim_n \|K_n f\| = 0 \tag{6.12}$$

for all $f \in X$. On the other hand, for $f_n = \sin(n\pi x)$,

$$\|f_n\| = 1, \quad \text{but} \quad \|K_n f_n\| = 1/2$$

which implies that $\|K_n\| \geq 1/2$, and so K_n does *not* converge to 0 in the uniform norm. This motivates us for the definition of a second notion of convergence of linear maps (and a second topology on $\mathcal{B}(X,Y)$, distinct from the uniform norm topology whenever X is infinite-dimensional).

Definition 6.7 (Strong Convergence) *A sequence (T_n) in $\mathcal{B}(X,Y)$ converges strongly (or in the strong operator topology) to $T \in \mathcal{B}(X,Y)$, if for all $x \in X$*

$$\lim_n T_n x = Tx$$

Note that, by (6.12), K_n of the previous example, converges strongly to zero.

Theorem 6.8 *If $T_n \to T$ uniformly, then $T_n \to T$ strongly.*

Exercise 6.3 *Prove Theorem 6.8.*

Appendix: The Riemann-Lebesgue Lemma

In this appendix we consider complex-valued functions on a finite interval $[a,b]$. Recall the definition of the L_1 norm: for all $f \in C([a,b])$

$$\|f\|_1 = \int_a^b |f(x)|\, dx$$

The completion of $C([a,b])$ with this norm is denoted by $L_1([a,b])$, and, loosely speaking, it is the space of all functions f for which $\|f\|_1$ is finite, and its elements are called L_1 functions. It is easy to see that the product of an L_1 function and a bounded continuous function is again L_1. The functions considered can be either real- or complex-valued, in which case $L_1([a,b])$ is a real or complex Banach space respectively.

The Riemann-Lebesgue Lemma can be considered as a statement about the Fourier coefficients of L_1 functions.

Theorem 6.9 (Riemann-Lebesgue Lemma) *For any $f \in L_1([0,2\pi])$ one has*

$$\lim_{n \to \pm\infty} \int_0^{2\pi} f(x) e^{inx}\, dx = 0$$

Proof: It is sufficient to prove this for f in a dense subspace of $L_1([0, 2\pi])$ (e.g., use Theorem 6.6).

A convenient choice for the dense subspace is the space of all trigonometric polynomials of the form

$$P_n(x) = \sum_{k=-n}^{n} a_k e^{ikx}$$

the a_k are arbitrary complex coefficients. By Theorem 2.13 the trigonometric polynomials are dense in $C(T)$ with the uniform norm (T is the unit circle). As the L^1-norm is weaker than the uniform norm, the trigonometric polynomials are also dense in $C(T)$ with the L^1 norm. As $C(T)$ is dense in $L_1([0, 2\pi])$ by construction, the trigonometric polynomials are also dense in we conclude $L_1([0, 2\pi])$. For any such P_n it is clear that

$$\int_0^{2\pi} P_n(x) e^{-ikx}\, dx = 2\pi a_k$$

And, therefore,

$$\int_0^{2\pi} P_n(x) e^{-ikx}\, dx = 0$$

for $|k| > n$, and so

$$\lim_{k \to \pm\infty} \int_0^{2\pi} P_n(x) e^{-ikx}\, dx = 0$$

■

6.3 Remarks on Finite-Dimensional Banach Spaces

Let X be a finite-dimensional linear space over $\mathbb{R}$ or $\mathbb{C}$. *Finite-dimensional* means that there is a finite set, $\{e_1, \ldots, e_n\} \subset X$, with the property that for any $x \in X$, there are constants $\alpha_1, \ldots, \alpha_n \in \mathbb{R}$ or $\mathbb{C}$, such that

$$x = \alpha_1 e_1 + \cdots + \alpha_n e_n \tag{6.13}$$

In fact, there is n, such that one can choose n *linearly independent* vectors $e_1, \ldots, e_n$ such that every $x \in X$ can be written as a linear combination of them. n is then the *dimension* of X. Linear independence of $e_1, \ldots, e_n$ is equivalent to the property that the coefficients α_i in (6.13) are uniquely determined by x.

Our main objective in these remarks on finite-dimensional Banach spaces is to show that every finite-dimensional normed linear space is a Banach space (i.e., is complete), and that, whenever X is finite-dimensional, every $T \in \mathcal{L}(X, Y)$ is bounded, i.e., $\mathcal{L}(X, Y) = \mathcal{B}(X, Y)$.

For the rest of this chapter, let X be a finite-dimensional normed linear space over $\mathbb{R}$ or $\mathbb{C}$. For simplicity of notation we will use $\mathbb{C}$ as the number field in the statements and proofs below, with the understanding that everywhere one can replace $\mathbb{C}$ by $\mathbb{R}$ and obtain the corresponding result about real linear spaces.

Lemma 6.10 *Let $\{e_1, \ldots, e_n\}$ be a basis for X. For any $i, 1 \leq 1 \leq n$, define a linear functional $\varphi_i : X \to \mathbb{C}$ by*

$$\varphi_i(x) = \alpha_i$$

where α_i is the coefficient of e_i appearing in the decomposition of x (6.13). Then, φ_i is bounded.

Proof: We will show that there exists a constant $m > 0$ such that

$$m \sum_{i=1}^{n} |\alpha_i| \leq \|x\| \tag{6.14}$$

It then follows that, for any i,

$$|\varphi_i(x)| = |\alpha_i| \leq \sum_{i=1}^{n} |\alpha_i| \leq \frac{1}{m}\|x\|$$

and, as $m > 0$, φ_i is bounded.

Clearly, it is sufficient to prove (6.14) for $x \in X$ such that $\sum_{i=1}^{n} |\alpha_i| = 1$ (use linearity and uniqueness of the coefficients α_i). The set

$$C = \{(\alpha_i) \in \mathbb{C}^n \mid \sum_{i=1}^{n} |\alpha_i| = 1\}$$

is a closed and bounded, and hence compact subset of $\mathbb{C}^n$ (Heine-Borel, Theorem 1.10). Define a function $f : C \to X$ by

$$f((\alpha_i)) = \sum_{i=1}^{n} \alpha_i e_i = x$$

As

$$\|f((\alpha_i) - f((\beta_i))\| = \left\| \sum_{i=1}^{n} (\alpha_i - \beta_i) e_i \right\| \leq \sum_{i=1}^{n} |\alpha_i - \beta_i| \|e_i\|$$

f is continuous, hence $(\alpha_i) \mapsto \|f((\alpha_i))\|$ is continuous, and, by Theorem 2.5, $\|f\|$ attains its infimum. Define

$$m = \min_{(\alpha_i) \in C} |f((\alpha_i))|$$

and suppose $|f((\alpha_i^*))| = m$. It follows from the linear independence of the basis vectors e_i that $m \neq 0$, and hence $m > 0$. For $(\alpha_i) \in C$ we have

$$m \sum_{i=1}^{n} |\alpha_i| = m \leq \|x\|$$

which implies (6.14). ■

Theorem 6.11 *Every finite-dimensional normed linear space is a Banach space.*

Proof: We need to show that every Cauchy sequence $(x_n)_{n \geq 1}$ in a finite-dimensional normed linear space X converges. Let $\{e_1, \ldots, e_k\}$ be a basis for X. Consider the sequences $(\alpha_{i,n})_{n \geq 1}$ in $\mathbb{C}$, where

$$x_n = \sum_{i=1}^{k} \alpha_{i,n} e_i$$

Then, for $i = 1, \ldots k$, $(\alpha_{i,n})_{n \geq 1}$ is Cauchy. Indeed

$$|\alpha_{i,n} - \alpha_{i,m}| = |\varphi_i(x_n - x_m)| \leq M \|x_n - x_m\|$$

where φ_i is the linear function considered in the previous lemma, and M is a common bound for the φ_i. It follows that there are $\beta_i \in \mathbb{C}, i = 1 \ldots, k$, such that

$$\lim_n \alpha_{i,n} = \beta_i$$

Define $y = \sum_{i=1}^{k} \beta_i e_i$. Then

$$\|x_n - y\| \leq \sum_{i=1}^{k} |\alpha_{i,n} - \beta_i| \|e_i\|$$

and therefore, $x_n \to y$. ■

Corollary 6.12 *Every finite-dimensional linear subspace of a Banach space is closed.*

Proof: Use the previous theorem and the following Exercise. ■

Exercise 6.4 *Let (X, d) be a complete metric space, and $X_0 \subset X$. Prove that (X_0, d) is complete iff X_0 is a closed subset of X.*

A linear map $T : \mathbb{C}^n \to \mathbb{C}^m$, can be defined by specifying a matrix $T = (t_{ij})$, such that

$$T(e_i) = \sum_{j=1}^{m} t_{ji} f_j, \quad \text{for all } i = 1, \ldots, n$$

where $\{e_i\}$ and $\{f_j\}$ are bases for $\mathbb{C}^n$ and $\mathbb{C}^m$ respectively. It is then easy to see that, with the usual norms on $\mathbb{C}^n$ and $\mathbb{C}^m$, T is bounded. E.g., $\|T\| \leq \max_{ij} |t_{ji}|$. The following theorem is a generalization of this observation.

Theorem 6.13 *Let X and Y be normed linear spaces. If X is finite-dimensional, $\mathcal{L}(X, Y) = \mathcal{B}(X, Y)$.*

Proof: Let $\{e_1, \ldots, e_n\}$ be a basis for X, and $x = \sum_{i=1}^{n} \alpha_i e_i \in X$. Then

$$\|Tx\| \leq \sum_{i=1}^{n} |\alpha_i| \, \|Te_i\| \leq M\|x\| \sum_{i=1}^{n} \|Te_i\|$$

where M is a common bound for the maps φ_i of Lemma 6.10. It follows that

$$\|T\| \leq M \sum_{i=1}^{n} \|Te_i\|$$

■

The next theorem shows that, although one can define many different norms on finite-dimensional linear spaces such as $\mathbb{R}^n$ and $\mathbb{C}^n$, there is only one topology derived from a norm on these spaces.

Theorem 6.14 *Let $\| \cdot \|_1$ and $\| \cdot \|_2$ be two norms on X. If X is finite-dimensional, then there exist constants $C_1 > 0$ and $C_2 > 0$ such that*

$$C_1 \|x\|_1 \leq \|x\|_2 \leq C_2 \|x\|_1, \quad \textit{for all } x \in X.$$

Proof: Consider the identity map $\mathbb{1} \in \mathcal{L}((X, \|\cdot\|_1), (X, \|\cdot\|_2))$. By Theorem 6.13 $\mathbb{1}$ is bounded. So, there exists a constant M such that

$$\|x\|_2 \leq M\|x\|_1, \quad \text{for all } x \in X.$$

Therefore, one can take $C_2 = M$. By the same argument with the roles of 1 and 2 interchanged one finds C_1. ■

Definition 6.15 *Let X be a linear space. Two norms $\|\cdot\|_1$ and $\|\cdot\|_2$ on X are called equivalent iff there exists constants $C_1, C_2 > 0$ such that*

$$C_1\|x\|_1 \leq \|x\|_2 \leq C_2\|x\|_1, \tag{6.15}$$

for all $x \in X$.

With this definition Theorem 6.14 can be rephrased as follows: any two norms on a finite-dimensional linear space are equivalent. In the following theorem we prove that equivalent norms, and only equivalent norms, generate the same topology on X. In combination with Theorem 6.14 this then proves that there is only one norm topology on a finite-dimensional linear space.

Theorem 6.16 *Let $\|\cdot\|_1$ and $\|\cdot\|_2$ be two norms on a (not necessarily finite-dimensional) linear space X. Then, $\|\cdot\|_1$ and $\|\cdot\|_2$ generate the same topology on X iff $\|\cdot\|_1$ and $\|\cdot\|_2$ are equivalent norms.*

Proof: Consider the identity map $\mathrm{id} : (X, \|\cdot\|_1) \to (X, \|\cdot\|_2))$. The topologies generated by the two norms are the same iff id and id^{-1} are continuous (Corollary 3.13). As id is linear, id is continuous iff it is bounded. By the same reasoning as in the proof of the previous theorem, the boundedness of the identity map is equivalent to the existence of constants C_1 and C_2 such that (6.15) holds. ■

Corollary 6.17 *A finite-dimensional linear space has exactly one norm topology.*

Exercise 6.5 *Prove that*
a) *in general, different norms on X lead to different but equivalent norms on $\mathcal{B}(X)$;*
b) *the set $\mathcal{B}(X)$ itself is independent of the chosen norm.*

7 Hilbert Spaces

In the course of the past couple of weeks we have considered different kinds of topological spaces. In increasing order of specialization we studied arbitrary toplogical spaces, metric spaces, normed linear spaces, and Banach spaces. Hilbert spaces are Banach spaces with a norm that is derived from an inner product. Thus, again, they have an extra feature as compared to arbitrary Banach spaces, which makes them more special. On the other hand, the extra structure will lead to new interesting properties.

The material of this sections elementary introduction to Hilbert space is covered in Chapter 4 of
W. Rudin, *Real and Complex Analysis*, 3rd ed, McGraw-Hill, New York 1986.

P.R. Halmos wrote a well-known study book on Hilbert space, which consists of 250 problems with hints and solutions:
P.R. Halmos, *A Hilbert Space Problem Book*, 2nd ed, Springer Verlag, New York 1982.

The new feature of Hilbert spaces is the *inner product*.

7.1 Inner Products

Definition 7.1 (Inner Product) *An inner product on a linear space X over $\mathbb{C}$ (or $\mathbb{R}$) is a map $(\,\cdot\,,\,\cdot\,) : X \times X \to \mathbb{C}$ (or $\mathbb{R}$), with the following properties: for all $x, y, z \in X$ and $\lambda \in \mathbb{C}$ (or $\mathbb{R}$) one has*
i) $(x, y + \lambda z) = (x, y) + \lambda(x, z)$ (linear in the second argument)
ii) $(y, x) = \overline{(x, y)}$, where $\bar{\cdot}$ denotes complex conjugation ("hermitian symmetric")
iii) $(x, x) \geq 0$ (non-negative, or positive semi-definite)
iv) $(x, x) = 0 \Rightarrow x = 0$ (positive definite).

A linear space with an inner product is often referred to as an *inner product space*.

From i) and ii) it follows that $(\,\cdot\,,\,\cdot\,)$ is *antilinear* in the first argument: $(y + \lambda z, x) = (y, x) + \overline{\lambda}(z, x)$, for all $x, y, z \in X$ and $\lambda \in \mathbb{C}$ (or $\mathbb{R}$). In the case of complex X, this means that $(\,\cdot\,,\,\cdot\,)$ is not bilinear. It is called *sesquilinear*, which literally means $1\frac{1}{2}$ linear.

Note that there are two conventions for the linearity of the inner product in the literature: in most of the mathematics oriented literature $(\,\cdot\,,\,\cdot\,)$ is linear in the first argument. We will adopt the convention that the inner product is linear in the second argument, which is the most common in applied mathematics and physics oriented works.

The following very useful inequality for inner products on a real or complex linear spaces is a consequence of the non-negative definiteness:

$$|(x, y)|^2 \leq (x, x)(y, y), \quad \text{for all } x, y \in X \tag{7.1}$$

It is usually called the *Cauchy-Schwarz inequality* .

Exercise 7.1 *Prove the Cauchy-Schwarz inequality.*

For simplicity of notation we will henceforth assume that X is a linear space over $\mathbb{C}$. Many statements carry over to the case of real linear spaces.

Let X be a linear space equipped with an inner product $(\cdot\,,\,\cdot)$. We can define a norm on X by

$$\|x\| = \sqrt{(x,x)}$$

Check that this indeed defines a norm.

Examples:

1. The standard inner product on $\mathbb{C}^n$ is given by

$$(x,y) = \sum_{i=1}^{n} \overline{x_i} y_i, \quad \text{for all } x, y \in \mathbb{C}^n$$

2. Let $C([a,b],\mathbb{C})$ denote the space of all complex-valued continuous functions defined on the interval $[a,b]$. An inner product on this space can be defined by

$$(f,g) = \int_a^b \overline{f(x)} g(x)\, dx, \quad \text{for all } f,g \in C([a,b],\mathbb{C})$$

By continuity this can be extended to $L_2([a,b],\mathbb{C})$, defined as the completion of $C([a,b],\mathbb{C})$ for the norm defined by $\|f\| = (f,f)^{1/2}$.

3. Define l_2 by

$$l_2 = \{(z_n)_{n=1}^{\infty} \mid \sum_{n=1}^{\infty} |z_n|^2 < \infty\}$$

l_2 is a complex linear space with the obvious operations of addition and multiplication by a scalar, and a natural inner product on it is given by

$$(x,y) = \sum_{n=1}^{\infty} \overline{x_n} y_n$$

4. Let M_n denote the space of all $n \times n$ matrices with complex entries. One can define a inner product on M_n by

$$(A,B) = \mathrm{Tr} A^* B$$

where Tr denotes the trace and * is hermitian conjugation of matrices. Check that this is the natural inner product one obtains by identifying M_n with $\mathbb{C}^{n^2}$.

Theorem 7.2 *A normed linear space X is an inner product space with the norm derived from the inner product iff*

$$\|x+y\|^2 + \|x-y\|^2 = 2\|x\|^2 + 2\|y\|^2, \quad \textit{for all } x,y \in X \tag{7.2}$$

and the inner product is necessarily given by the formula

$$(x,y) = \frac{1}{4}\left\{\|x+y\|^2 - \|x-y\|^2 + i\|x+iy\|^2 - i\|x-iy\|^2\right\}, \quad \textit{for all } x,y \in X \tag{7.3}$$

Proof: The proof of this theorem is homework. ■

The relation (7.2) has a geometric content and is called the *Parallelogram Law.* It says that, if one considers the parallelogram spanned by the vectors x and y in the two dimensional subspace ($\equiv$ plane) generated by x and y, the sum of the squares of two adjacent sides of the parallelogram is equal to half the sum of the squares of the diagonals.

Let $(X, (\cdot\,,\,\cdot)_X)$ and $(Y, (\cdot\,,\,\cdot)_Y)$ be two inner product spaces. Then $X \times Y = \{(x,y) \mid x \in X, y \in Y\}$ has a natural inner product defined by

$$((x_1,y_1),(x_2,y_2))_{X\times Y} = (x_1,x_2)_X + (y_1,y_2)_Y$$

which in turn leads to a natural norm on $X \times Y$:

$$\|(x,y)\| = \sqrt{\|x\|^2 + \|y\|^2}$$

Unless otherwise stated, this will be the norm considered on a Cartesian product of two inner product spaces.

Theorem 7.3 *Let X be an inner product space, considered as a normed space with the norm the norm derived from the inner product. Then, the inner product is continuous as a map from $X \times X \to \mathbb{C}$.*

Proof: For all $x, y \in X$ one has the bound

$$|(x,y)| \leq \|x\|\,\|y\|$$

which is nothing else than the Cauchy-Schwartz inequality. Just as for linear maps, continuity follows from this kind of boundedness:

$$|(x_1,y_1) - (x_2,y_2)| = |(x_1 - x_2, y_1) + (x_2, y_1 - y_2)| \leq \|x_1 - x_2\|\,\|y_1\| + \|x_2\|\,\|y_1 - y_2\|$$

■

7.2 Hilbert Spaces

Definition 7.4 *A Hilbert space is an inner product space X that is complete for the norm derived from the inner product.*

In particular, a Hilbert space is a Banach space.

Examples:
Except for $C([a,b],\mathbb{C})$, the examples of inner product spaces given in the previous section are all complete and, therefore, are Hilbert spaces.

The inner product structure of a Hilbert space allows us to introduce the concept of orthogonality of vectors. Let $\mathcal{H}$ be a Hilbert space. We say that x and $y \in \mathcal{H}$ are *orthogonal* iff $(x,y) = 0$. Shorthand for "x and y are orthogonal" is $x \perp y$. For two linear subspaces $\mathcal{H}_0$ and $\mathcal{H}_1$ of $\mathcal{H}$, we say that $\mathcal{H}_0$ and $\mathcal{H}_1$ are orthogonal if $x \perp y$ for all $x \in \mathcal{H}_0$ and $y \in \mathcal{H}_1$, in which case one writes $\mathcal{H}_0 \perp \mathcal{H}_1$. Similarly, $x \perp \mathcal{H}_0$ means $x \perp y$, for all $y \in \mathcal{H}_0$. The *orthogonal complement* $\mathcal{H}_0^\perp$ of a linear subspace $\mathcal{H}_0$ of $\mathcal{H}$ is defined by

$$\mathcal{H}_0^\perp = \{x \in \mathcal{H} \mid x \perp y, \text{for all } y \in \mathcal{H}_0\}$$

The notion of orthogonality makes it possible to think about vectors and linear subspaces of a Hilbert space in a geometric way. This is sometimes very helpful, much in the same way as our geometric intuition for the Euclidean spaces $\mathbb{R}^n$ helps us to understand linear transformations of $\mathbb{R}^n$. It should be mentioned that the geometry of Banach spaces can be surprisingly complicated and quite different from what naive extrapolations of the finite-dimensional situation would suggest. The following two results, and their applications (see MAT203B), rely very much on the notion of orthogonality.

Theorem 7.5 *Let $\mathcal{H}$ be a Hilbert space and $\mathcal{H}_0$ a linear subspace of $\mathcal{H}$. Then, $\mathcal{H}_0^\perp$ is a closed linear subspace of $\mathcal{H}$.*

Proof: Let us first prove that $\mathcal{H}_0^\perp$ is indeed a linear subspace. For this we must show that for all $x, y \in \mathcal{H}_0^\perp$, and $\lambda \in \mathbb{C}$, $x + \lambda y \in \mathcal{H}_0^\perp$. But this follows immediately form the linearity of the inner product:

$$(x_0, x + \lambda y) = (x_0, x) + \lambda (x_0, y) = 0 + 0, \quad \text{for all } x_0 \in \mathcal{H}_0$$

Next, we show that $\mathcal{H}_0^\perp$ is closed. A Hilbert space is a metric space, and therefore, it suffices to show that for any convergent sequence of elements $x_n \in \mathcal{H}_0^\perp$, the limit x also belongs to $\mathcal{H}_0^\perp$. Let $x_0 \in \mathcal{H}_0$. From Theorem 7.3 we have that the inner product is continuous and therefore, we can compute (x_0, x) as follows:

$$(x_0, x) = (x_0, \lim_n x_n) = \lim_n (x_0, x_n) = \lim_n 0 = 0$$

Hence, $x \in \mathcal{H}_0^\perp$. ■

The following theorem expresses a deep property of Hilbert spaces. Beware that there is no analogous theorem for arbitrary Banach spaces.

Theorem 7.6 *Let $\mathcal{H}_0$ a closed linear subspace of a Hilbert space $\mathcal{H}$.*
i) For any $x \in \mathcal{H}$ there is a unique $x_0 \in \mathcal{H}_0$ such that

$$||x - x_0|| = \min_{y \in \mathcal{H}_0} ||x - y|| \tag{7.4}$$

ii) The point $x_0 \in \mathcal{H}_0$, closest to $x \in \mathcal{H}$, is the unique element of $\mathcal{H}_0$ with the property that $x - x_0 \perp \mathcal{H}_0$.

Proof:
Step 1. First, we prove that there exists $x_0 \in \mathcal{H}_0$ such that

$$||x - x_0|| = d := \inf\{||x - y|| \mid y \in \mathcal{H}_0\} \tag{7.5}$$

From the definition of d it follows that there is a sequence of elements $y_n \in \mathcal{H}_0$, such that $\lim_n ||x - y_n|| = d$. So, for all $\epsilon > 0$, there exists N such that for all $n \geq N$, $||x - y_n|| \leq d + \epsilon$. Let us show that the sequence (y_n) is Cauchy. By the Parallelogram Law we have

$$||y_m - y_n||^2 + ||2x - y_m - y_n||^2 = 2||x - y_m||^2 + 2||x - y_n||^2$$

As $(y_m + y_n)/2 \in \mathcal{H}_0$, we have

$$||x - (y_m + y_n)/2|| \geq d$$

Bu combining these relations we obtain:

$$\begin{aligned} ||y_m - y_n||^2 &= 2||x - y_m||^2 + 2||x - y_n||^2 - ||2x - y_m - y_n||^2 \\ &\leq 4(d + \epsilon)^2 - 4d^2 \\ &\leq 8\epsilon(d + \epsilon) \end{aligned}$$

for all $m, n \geq N$. This proves that (y_n) is Cauchy. As a Hilbert space is complete by definition, there is y such tha $y_n \to y$, and as $\mathcal{H}_0$ is closed by assumption, we have that $y \in \mathcal{H}_0$, and we can take $x_0 = y$.
Step 2. Next, we prove the uniqueness of the vector $x_0 \in \mathcal{H}_0$ minimizing $||x - x_0||$. Suppose x_0 and x_0' both minimize the distance to x, i.e., $||x - x_0|| = ||x - x_0'|| = d$. Again, by the Parallelogram Law, we have

$$2||x - x_0||^2 + 2||x - x_0'||^2 = ||2x - x_0 - x_0'||^2 + ||x_0 - x_0'||^2$$

and hence, as $x_0 + x_0' \in \mathcal{H}_0$,

$$||x_0 - x_0'||^2 \leq 4d^2 - ||2x - x_0 - x_0'||^2 \leq 0$$

Therefore, $||x_0 - x_0'|| = 0$, and we must conclude that $x_0 = x_0'$. So there can be only one $x_0 \in \mathcal{H}_0$ such that $||x - x_0|| = d$.
Step 3. The unique x_0 found in Steps 1 and 2, must satisfy $(z, x - x_0) = 0$, for all $z \in \mathcal{H}_0$. This follows from the inequality

$$||x - x_0||^2 \leq ||x - (x_0 - \lambda z)||$$

for any $z \in H_0, \lambda \in \mathbb{C}$. Indeed, by expanding the RHS and using $||x - x_0||^2 = d^2$, we obtain

$$d^2 \leq d^2 + |\lambda|^2 ||z||^2 - 2\mathrm{Re}\,\lambda(x - x_0, z)$$

which simplifies to

$$|\lambda|^2 ||z||^2 \leq 2\mathrm{Re}\,\lambda(x - x_0, z)$$

Take λ of the form $\epsilon e^{i\varphi}$, with $\epsilon > 0$ and $\varphi \in \mathbb{R}$, divide by ϵ, and take the limit $\epsilon \downarrow 0$, and you obtain

$$2\mathrm{Re}\, e^{i\varphi}(x - x_0, z) \leq 0, \quad \text{for all } \varphi \in \mathbb{R}$$

It follows that $(x - x_0, z) = 0$ (consider this inequality for $\varphi = 0, \pi/2, \pi$, and $3\pi/2$).
Step 4. Finally, we show that x_0 is the only element in $\mathcal{H}_0$ such that $x - x_0 \perp \mathcal{H}_0$. Suppose x_0' is another such element in $\mathcal{H}_0$. Then $x_0 - x_0' \in \mathcal{H}_0$, and, for any $y \in \mathcal{H}_0$, we have

$$(y, x_0 - x_0') = (y, x - x_0') - (y, x - x_0) = 0 - 0 = 0$$

In particular, we can take $y = x_0 - x_0'$, and therefore we must have $x_0 = x_0'$. ■

Index

ϵ-cover, 26
ϵ-net, 26

homeomorphism, 39

adherence, 13
Arzelà-Ascoli, Theorem of, 29

Banach algebra, 21
Banach space, 21, 63
  of linear operators, 67
base of open sets, 41
Bernstein polynomials, 23
Bolzano-Weierstrass, 14
Born approximation, 57
boundary value problem, 57
bounded
  from above(below), 19
  function, 18
  linear operator, 63
  totally, 26

Cantor, 31
Cauchy sequence, 8
Cauchy-Schwarz inequality, 77
closed set, 12
closure, 12
  sequential, 45
cluster point, 11
compact
  metric space, 13
  relatively, 28
  sequentially, 13, 40
  topological space, 40
complete metric space, 8
  of continuous functions, 21
completion, 48
continuous, 16
  extension, 67
  topological space, 39
  uniformly, 16, 22
contraction, 47
Contraction Mapping Theorem, 47, 51
  applications, 53–62
convergence
  of a sequence, 8
  pointwise, 19
  topological space, 39
  uniform, 20
convergent
  absolutely, 11
  series, 11
  unconditional, 11
countable, 13
cover, 26

dense, 13
diagonal argument, Cantor's, 31
diameter, 10
differentiable, 17
differential equation
  existence and uniqueness, 59, 61
  existence of solutions, 33
  nonlinear, 61
dimension, 72
dist(ance), 10
distance function, 6
divergent
  series, 11
dual topological space, 65
dynamical system, 54, 59

equicontinuous, 28
Euclidean metric, 7
Euclidean space, 6

finite-open-subcover property, 26
first countable, 42
fixed point, 47, 60

functional, 65

Green's function, 58

Hamel basis, 65
Hausdorff, 39
heat equation, 57
Heine-Borel, Theorem of, 14
Hilbert space, 79

infimum, 9, 18, 19
inner product, 76
    standard, 77
inner product space, 76
integral equation, 55, 59
isometry, 48
isomorphic
    metric spaces, 48

kernel, 64

lim inf, 9
lim sup, 9
linear
    functional, form, 65
linear spaces
    finite-dimensional, 72
Lipschitz, 32, 36, 60
lower bound, 9

maximum, 10, 18
metric, 6
    equivalent, 45
metric space, 6
minimum, 10, 18
monotone, 10
monotone convergence theorem
    Dini's, 28

neighborhood, 13, 39
neighborhood base, 41
Neumann series, 57
nonlinear
    feedback, 62
    ODE, 33, 61
norm, 63
    equivalent, 75
    maximum, 7
    operator, 66
    sum, 7
    supremum, 20
    uniform, 20, 66
null space, 64

ODE
    blowup, 62
    existence and uniqueness, 59, 61
    existence of solutions, 33
    non-unique solutions, 36
    nonlinear, 61
open cover, 26
open set
    metric spaces, 12
    topological space, 38
operator
    bounded, 63
    topology, 66
orthogonal
    complement, 79
    vectors, 79

parallelogram law, 78
polynomial
    Berstein, 23
    trigonometric, 25, 71
population dynamics, 55
precompact, 28

range, 64
Riemann-Lebesgue Lemma, 69, 70

semi-continuous, 19
seminorm, 42, 63
separable
    metric space, 27

sequence
Cauchy, 8
convergent, 8
divergent, 8
sequentially compact, 13, 40
series, 11
sesquilinear, 76
strong operator topology, 70
subcover, 26
subsequence, 9
supremum, 9, 18, 19

topological space, 38
dual, 65
topology
discrete, 38
indiscrete, 38
induced, 39
non-metric, 45
operator, 66
relative, 39
strong operator, 70
trivial, 38
uniform, 66
totally bounded, 26
totally ordered, 10
triangle inequality, 6, 63

ultrametric, 7
uncountable, 13
upper bound, 9

Volterra equation, 55

Weierstrass, 14
approximation theorem, 23