TMA4145 – Linear Methods
Final version
November 28, 2017

Franz Luef

Abstract. These notes are for the course TMA4145 – Linear Methods at NTNU for the fall 2017.

Contents

CHAPTER 1

Sets and functions

Basic definitions and theorems about sets and functions are the content of this chapter and are presented in the setting of Naive Set Theory. These notions set the stage for turning our intuition about collections of objects and relations between these objects.

1.1. Sets

DEFINITION 1.1.1. A *set* is a collection of distinct objects, its *elements*. If an object x is an element of a set X, we denote it by $x \in X$. If x is not an element of A, then we write $x \notin X$.

A set is uniquely determined by its elements. Suppose X and Y are sets. Then they are identical, $X = Y$, if they have the same elements. More formalized, $X = Y$ if and only if for all $x \in X$ we have $x \in Y$, and for all $y \in Y$ we have $y \in X$.

DEFINITION 1.1.2. Suppose X and Y are sets. Then Y is a subset of X, denoted by $Y \subset X$, if for all $y \in Y$ we have $y \in X$.

If $Y \subseteq X$, one says that Y is contained in X. If $Y \subseteq X$ and $X \neq Y$, then Y is a proper subset of X and we use the notation $Y \subset X$. The most direct way to prove that two sets X and Y are equal is to show that

$$x \in X \iff x \in Y$$

for any element x. (Another way is to prove a double inclusion: if $x \in X$ then $x \in Y$, establishing that $X \subset Y$ and if $x \in Y$, then $x \in X$, establishing that $Y \subset X$.)

The *empty set* is a set with no elements, denoted by $\emptyset$.

PROPOSITION 1.1.3. *There is only one empty set.*

PROOF. Suppose E_1 and E_2 are two empty sets. Then for all elements x we have that $x \notin E_1$ and $x \notin E_2$. Hence $E_1 = E_2$. □

Some familiar sets are given by the various number systems:

(1) $\mathbb{N} = \{1, 2, 3, ...\}$ the set of natural numbers, $\mathbb{N}_0 = \{0, 1, 2, 3, ...\}$;
(2) $\mathbb{Z} = \{..., -2, -1, 0, 1, 2, ...\}$ the set of integerr;
(3) $\mathbb{Q} = \{p/q : p, q \in \mathbb{Z}\}$ the set of rational numbers;
(4) $\mathbb{R}$ denotes the set of real numbers;
(5) $\mathbb{C}$ denotes the set of complex numbers.

For real numbers a, b with $a < b < \infty$ we denote by $[a, b]$ the closed bounded interval, and by (a, b) the open bounded interval. The length of these bounded intervals is $b - a$.
Here are a few constructions related to sets.

DEFINITION 1.1.4. Let X and Y be sets.

- The *union* of X and Y, denoted by $X \cup Y$, is defined by
$$X \cup Y = \{z | \, z \in X \quad \text{or } z \in Y\}.$$
- The *intersection* of X and Y, denoted by $X \cap Y$, is defined by
$$X \cap Y = \{z | \, z \in X \quad \text{and } z \in Y\}.$$
- . The *difference set* of X from Y, denoted by $X \backslash Y$, is defined by
$$X \backslash Y = \{z \in X : \, z \in X \quad \textit{and } z \neq Y\}.$$
If all sets are contained in one set X, then the difference set $X \backslash Y$ is called the *complement* of Y and denoted by Y^c.
- The *Cartesian product* of X and Y, denoted by $X \times Y$, is the set
$$X \times Y = \{(x, y) | \, x \in X, y \in Y\},$$
i.e the set of all ordered pairs (x, y), with $x \in X$ and $y \in Y$. Recall an ordered pair has the property that $(x_1, y_1) = (x_2, y_2)$ if and only if $x_1 = x_2$ and $y_1 = y_2$.
- $\mathcal{P}(X)$ denotes the set of all subsets of X.

Here are some basic properties of sets.

LEMMA 1.1. *Let X, Y and Z be sets.*

(1) $X \cap (Y \cup Z) = (X \cap Y) \cup (X \cap Z)$ *and* $X \cup (Y \cap Z) = (X \cup Y) \cap (X \cup Z)$ *(distribution law)*
(2) $(X \cup Y)^c = X^c \cap Y^c$ *and* $(X \cap Y)^c = X^c \cup Y^c$ *(de Morgan's laws)*
(3) $X \backslash (Y \cup Z) = (X \backslash Y) \cap (X \backslash Z)$ *and* $X \backslash (Y \cap Z) = (X \backslash Y) \cup (X \backslash Z)$
(4) $(X^c)^c = X$.

PROOF. (1) Let us prove one of de Morgan's relations. Let us use the most direct approach. Keep in mind that $x \in E^c \iff x \notin E$. We then have:

$$x \in (X \cup Y)^c \iff x \notin X \cup Y \iff x \notin X \textbf{ and } x \notin Y$$
$$\iff x \in X^c \text{ and } x \in Y^c \iff x \in X^c \cap Y^c.$$

This proves the identity.

(2)
$$x \in (X^c)^c \iff x \notin X^c \iff x \in X.$$

□

Note that if you have a statement involving $\cup$ and $\cap$. Then you get another true statement if you interchange $\cup$ with $\cap$ and $\cap$ with $\cup$, as one can see in the lemma. This is part of the field Boolean algebra.

1.2. Functions

Let X and Y be sets. A function with *domain* X and *codomain* Y, denoted by $f : X \to Y$, is a relation between the elements of X and Y satisfying the properties: for all $x \in X$, there is a unique $y \in Y$ such that $(x, y) \in f$, we denote it by: $f(x) = y$.

By definition, for each $x \in X$ there is exactly one $y \in Y$ such that $f(x) = y$. We say that y the *image* of x under f. The *graph* $G(f)$ of a function f is the subset of $X \times Y$ defined by

$$G(f) = \{(x, f(x)) | \, x \in X\}.$$

The *range* of a function $f : X \to Y$, denoted by range(f), or $f(X)$, is the set of all $y \in Y$ that are the image of some $x \in X$:

$$\text{range}(f) = \{y \in Y | \, \text{there exists} \, x \in X \, \text{such that} \, f(x) = y\}.$$

The *pre-image* of $y \in Y$ is the subset of all $x \in X$ that have y as their image. This subset is often denoted by $f^{-1}(y)$:

$$f^{-1}(y) = \{x \in X | \, f(x) = y\}.$$

Note that $f^{-1}(y) = \emptyset$ if and only if $y \in Y \backslash \text{ran}(f)$.

Here are some simple examples of functions.

$$|x| = \begin{cases} x & \text{if } x > 0, \\ 0 & \text{if } x = 0, \\ -x & \text{if } x < 0. \end{cases}$$

Note that $|x| = \max\{x, -x\}$. We define the positive, x^+ and negative part, x^- of $x \in \mathbb{R}$:

$$x^+ = \max\{x, 0\}, \quad \text{and} \quad x^- = \max -x, 0,$$

so we have $x = x^+ - x^-$ and $|x| = x^+ + x^-$.

The following notions are central for the theory of functions.

Definition 1.2.1. Let $f : X \to Y$ be a function.

(1) We call f *injective* or one-to-one if $f(x_1) = f(x_2)$ implies $x_1 = x_2$, i.e. no two elements of the domain have the same image. Equivalently, if $x_1 \neq x_2$, then $f(x_1) \neq f(x_2)$.
(2) We call f *surjective* or *onto* if $\text{ran}(f) = Y$, i.e. each $y \in Y$ is the image of at least one $x \in X$.
(3) We call f *bijective* if f is both injective and surjective.

Note that a bijective function matches up the elements of X with those of Y so that in some sense these two sets have the same number of elements.

Let $f : X \to Y$ and $g : Y \to Z$ be two functions so that the range of f coincides with the domain of g. Then we define the *composition*, denoted by $g \circ f$, as the function $g \circ f : X \to Z$, defined by $x \mapsto g(f(x))$.

For every set X, we define the *identity map*, denoted by id_X or id where $\text{id}(x) = x$ for all $x \in X$.

Lemma 1.2. *Let $f : X \to Y$ and $g : Y \to Z$ be two bijections. Then $g \circ f$ is also a bijection and $(g \circ f)^{-1} = f^{-1} \circ g^{-1}$.*

Lemma 1.3. *Let $f \colon X \to Y$ be a function and let $C, D \subset Y$. Then*

$$f^{-1}(C \cup D) = f^{-1}(C) \cup f^{-1}(D).$$

PROOF.

$$x \in f^{-1}(C \cup D) \iff f(x) \in C \cup D \iff f(x) \in C \text{ or } f(x) \in D$$
$$\iff x \in f^{-1}(C) \text{ or } x \in f^{-1}(D) \iff x \in f^{-1}(C) \cup f^{-1}(D).$$

□

If one has a function f that maps elements in X to Y, then it is often desirable to reverse this assignment. Let us introduce some notions to address this basic problem.

DEFINITION 1.2.2. Let f be a function from X to Y.

- The mapping f is said to be *left invertible* if there exists a function $g : Y \to X$ such that $g \circ f = \mathrm{id}_X$. We call g a *left inverse* of f and denote it by f_l^{-1}.
- The mapping f is said to be *right invertible* if there exists a function $h : Y \to X$ such that $f \circ h = \mathrm{id}_Y$. We call h a *right inverse* of f and denote it by f_r^{-1}.
- The mapping f is said to be *invertible* if there exists a $g : Y \to X$ such that $g \circ f = f \circ g = \mathrm{id}$, the so-called *inverse* of f and denoted by f^{-1}.

One may think of a left and right inverse in layman terms: (i) If you map an element of the domain via a function to an element in the target space, then the left inverse tells you how to go back to where you started from;(ii) If one wants to get to a point in the target, then the right inverse tells you a possible place to start in the domain. The inverse of a function has some important properties.

LEMMA 1.4. *Given an invertible function $f : X \to Y$.*

(1) *The inverse function $f^{-1} : Y \to X$ is unique.*
(2) *The inverse function is also invertible and we have $(f^{-1})^{-1} = f$.*

PROOF. (1) Suppose there are two inverse functions $g_i : Y \to X$, $i = 1, 2$. By assumption we have that $f \circ g_1 = \mathrm{id}$ and $g_2 \circ f = \mathrm{id}$. Hence we have

$$g_2(y) = g_2(f g_1(y)) = g_2 f(g_1(y)) = g_1(y) \quad \text{for all } y \in Y,$$

i.e. $g_1 = g_2$.
(2) Exercise.

□

Let us give a description of left, right invertibility and invertibility in more concrete terms.

PROPOSITION 1.2.3. *Given a function $f : X \to Y$.*

(1) *f is left invertible if and only if it is injective.*
(2) *f is right invertible if and only if it is surjective.*
(3) *f is invertible if and only if it is injective and surjective, i.e. if f is bijective.*

PROOF. (1) Let us assume that f is injective. Then $f : x \to \mathrm{ran}(f)$ is invertible with $f^{-1} : \mathrm{ran}(f) \to X$. Let $g : Y \to X$ be any extension of this inverse. Then $g \circ f = \mathrm{id}_X$.

Suppose f is left invertible. Assume there are $x_1, x_2 \in X$ such that $f(x_1) = f(x_2) = y$. Then

$$x_1 = f_l^{-1}(f(x_1)) = f_l^{-1}(f(x_2)) = x_2,$$

i.e. f is injective.

(2) Let us assume that f is surjective. Pick an arbitrary element $z \in Y$, wich is by assumption an element of $\mathrm{ran}(f)$. Hence z has at least one pre-image in X and thus $f^{-1}(z) \neq \emptyset$.Take $y_1 \neq y_2$. Then the sets $f^{-1}(\{y_1\})$ and $f^{-1}(\{y_2\})$ in X are disjoint. Let us pick from each set $f^{-1}(\{y\})$ an element x and define $x := h(y)$. Then $h : Y \to X$ and $f \circ h = \mathrm{id}_Y$.
Suppose that f is right invertible. Then we have for $y \in Y$ that $f(f_r^{-1})(y) = f(x)$ where we set x to be $x = f_r^{-1}(y)$. In other words, y is in the range of f.

(3) Follows from the other assertions.

□

A consequence of the characterizations of left and right invertibility is the observation:

REMARK 1.2.4. If $f : X \to Y$ is left invertible such that $\mathrm{ran}(f) \neq Y$, then there are many left inverses. However the restriction of any left inverse of f to $\mathrm{ran}(f)$ is unique.
One the other hand if $f : X \to Y$ is right invertible such that f is surjective but not injective, then f will have many right inverses.

Our study of linear mappings will provide ample examples of the aforementioned notions. Here we just give one example.

EXAMPLE 1.2.5. Given the linear mapping $T : \mathbb{R}^3 \to \mathbb{R}^2$ given by $T = Ax$ with

$$A = \begin{pmatrix} -3 & -4 \\ 4 & 6 \\ 1 & 1 \end{pmatrix}.$$

Then the matrix

$$A_l^{-1} = \frac{1}{9}\begin{pmatrix} -11 & -10 & 16 \\ 7 & 8 & -11 \end{pmatrix}$$

induces a left inverse T_l^{-1} of T.
This left inverse is not unique, for example

$$\frac{1}{2}\begin{pmatrix} 0 & -1 & 6 \\ 0 & 1 & -4 \end{pmatrix}$$

also gives a left inverse. One can turn this example into one for right inverses as well, see problem set 1.

1.3. Cardinality of sets

Bijective functions provide us with a way to compare the size of two sets. We start with the case of finite sets.

DEFINITION 1.3.1. Two sets X and Y have equal cardinality, if there is a bijective map $f : X \to Y$. If there is an injective map from X to Y, then we say that the cardinality of X is less than or equal to the cardinality of Y.
A set X has n elements if there is a bijection between X and the set $\{0, 1, ..., n-1\}$.

We denote the set $\{0, 1, ..., n-1\}$ by n. A set X is countable if there is a bijection with $\mathbb{N}$. In other words, X is countable if we can arrange its elements in an infinite sequence $\{x_1, x_2, x_3, ...\}$ such that eqch element occurs exactly once in the sequence.

REMARK 1.3.2. There is some more terminology that we will not use in the course. A set X is called at most countable if there is an injection from X to $\mathbb{N}$.

EXAMPLES 1.3.3. We give some examples based on the set of natural numbers.

(1) The set of squares $X = \{1, 4, 9, ..., n^2, ...\}$ is countable, since $f : \mathbb{N} \to X$ defined by $f(n) = n^2$ is bijective.
(2) The set of odd numbers $X = \{1, 3, 5, ..., 2n-1, ...\}$ is countable, since $f : \mathbb{N} \to X$ defined by $f(n) = 2n - 1$ is a bijection.

Let us state a characterization of countable sets.

LEMMA 1.5. *A set X is countable. $\Leftrightarrow$ There exists a surjective map $f : \mathbb{N} \to X$.*

PROOF. ($\Rightarrow$) Suppose X is countable. Then there is a surjection $f : \mathbb{N} \to X$ which is in addition injective.

($\Leftarrow$) Given a surjective map $f : \mathbb{N} \to X$. We have to turn this map into an bijection g. The idea is to omit the repeated values of f. We proceed in a recursive manner. Define $g(1) := f(1)$. Suppose we have chosen n distinct values $g(1), g(2), ..., g(n)$. We collect the set of natural numbers where the values of f are not already included among the list $\{g(1), g(2), ..., g(n)\}$:

$$X_n := \{k \in \mathbb{N} : \ f(k) \neq g(j) \text{ for every } j = 1, 2, ..., n\}.$$

The set X_n can either be empty or not. Suppose $X_n = \emptyset$. Then $g : \{1, 2, ..., n\} \to X$ is a bijection and thus X is finite. Otherwise, if $X_n \neq \emptyset$, then we denote by k_n the least integer in X_n and set $g_{n+1} := f(k_n)$. Note that by construction $g(n+1)$ differs from $g(1), g(2), ..., g(n)$. We continue in this manner. If the process terminates, then X is finite, or we go through all the values of f and obtain a surjection $g : \mathbb{N} \to X$. □

The assignment of the number of elements of $\{0, 1, ..., n-1\}$ with the set n yields that for any set X, there is at most one natural number n such that X is bijective with the set n.

PROPOSITION 1.3.4. *If there is a bijection between the sets n and m, then they have the same number of elements.*

PROOF. We proceed by induction. For $n = 0$ the set $n = \{0, 1, ..., n-1\}$ is the empty set, and thus the only set bijective with it is the empty set. Suppose that $n > 0$ and that the result is true for $n - 1$. Hence there is a bijection $f : \{0, 1, ..., n-1\} \to \{0, 1, ..., m-1\}$. We assume that $f(n-1) = m - 1$. Then the restriction of f to the set $\{0, 1, ..., n-2\}$ gives a bijection to $\{0, 1, ..., m-2\}$. By the induction hypothesis we have $n - 1 = m - 1$. Let us now look at the case when $f(n-1) \neq m - 1$. We have that $f(n-1) = a$ for some a and $f(b) = m - 1$ and we define a function $\tilde{f}$ by $\tilde{f}(x) = f(x)$ if $x \neq b, n-1$; $\tilde{f}(k) = a$ and $\tilde{f}(n-1) = m - 1$. Then $\tilde{f}$ is a bijection and we conclude as before that $n = m$. □

We move on to sets that are bijective to the set of natural numbers $\mathbb{N} = \{1, ...\}$.

PROPOSITION 1.3.5. *A set is at most countable it is finite or countable.*

PROOF. Suppose $f : X \to \mathbb{N}$ is an injective function. We construct a function $g : X \to \mathbb{N}$ as follows: $g(x) = n$ if $f(x)$ is the nth element in the image of f. □

PROPOSITION 1.3.6. *$\mathbb{N} \times \mathbb{N}$ is countable.*

PROOF. The argument starts out with decomposing $N \times N$ into finite sets $F_0, F_1, ...$, where

$$F_k = \{(i, j) \in \mathbb{N} \times \mathbb{N} | \, i + j = k\}$$

and the cardinality of F_k is $k + 1$. Now we arrange these sets: first writing the one element of F_0, then the two elements of F_1 and so forth. Hence, we have established the assertion. In other words, we have arranged $N \times \mathbb{N}$ in a table:

$$\begin{array}{ccccc} (1,1) & (1,2) & (1,3) & (1,4) & \cdots \\ (2,1) & (2,2) & (2,3) & (2,4) & \cdots \\ (3,1) & (3,2) & (3,3) & (3,4) & \cdots \\ (4,1) & (4,2) & (4,3) & (4,4) & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{array}$$

and list the elements along sucessive (anti-)diagonals from bottom-left to top-right as

$$(1,1), (2,1)(1,2), (3,1), (2,2), (1,3), ...$$

We define $f : \mathbb{N} \to \mathbb{N} \times \mathbb{N}$ by $f(n) := nth$ pair in this order. Note that f is a bijection. □

Here are some facts about countable sets.

PROPOSITION 1.3.7. *We have the following assertions:*

(1) *The Cartesian product of two countable sets is countable.*
(2) *The union of countably many countable sets is countable.*

PROOF. (1) We show that the Cartesian product of two countable sets is countable which reduces to the statement that the set $N \times \mathbb{N}$ is countable which we have shown in 1.3.6.

(2) Let $X_0, X_1, ...$ be a countable family of countable sets. We denote the elements of X_i by $\{x_{0i}, x_{1i}, ...\}$ for $i = 0, 1, ...$ and define a map by $f(i, j) = x_{ij}$. Note that $f : N \times \mathbb{N} \to \cup_{i=0}^{\infty} X_i$ and thus the union $\cup_{i=0}^{\infty} X_i$. is countable. The map f is not injective in general, because the X_i's need not to be disjoint. The proposition preceding this statement yields the desired claim.

□

PROPOSITION 1.3.8. *The sets $\mathbb{Z}$ of integers and $\mathbb{Q}$ of rational numbers are countable.*

PROOF. One of the problems of problem set 1. □

Bernstein and Schröder observed an elementary characterization of two sets having the same cardinality, we state it without proof.

THEOREM 1.6. *Let X and Y be two sets. Suppose there are injective maps $f : X \to Y$ and $g : Y \to X$. Then there exists a bijection between X and Y.*

We give some examples of a non-countable sets.

THEOREM 1.7 (Cantor). *The set* $\mathbb{R}$ *of real numbers is* **not** *countable.*

If a set is not countable, then one often calls it uncountable.

PROOF. We argue by contradiction and assume that $\mathbb{R}$ is countable. Then a subset of $\mathbb{R}$ is also countable. Thus the open interval $(0,1)$ is a countable set, i.e.

$$(0,1) = \{x_0, x_1, ...\}.$$

Any $a_i \in (0,1)$ has an infinite decimal expansion (possibly terminating, in which case we let it continue forever with zeros):

$$a_i = 0.a_{i0}a_{i1}..., \qquad a_{ij} \in \{0,1,...,9\}.$$

We set b_i to be

$$b_i = \begin{cases} 3 & \text{if } a_{ii} \neq 3 \\ 1 & \text{if } a_{ii} = 7. \end{cases}$$

By construction we have $b_i \neq a_i i$ and thus the number

$$a = 0.b_1b_2...$$

differs from a_i. Note that $a \in (0,1)$ which is not included in the given enumeration of $(0,1)$. Hence we have deduced a contradiction to the countability of $(0,1)$. The number $b_i \in (0,1)$ and differs from a_i, since the ith place of a_i and b_i are by construction not the same digit. □

PROPOSITION 1.3.9. *Let* X *be the set of all binary sequences:* $X = \{(a_1, a_2, a_3, ...) : a_i \in \{0,1\}\}$. *Then* X *is not countable.*

PROOF. We apply the method from the preceding theorem, aka diagonal argument.
Suppose $X = \{(x_1, x_2, x_3, ...) : x_i \in \{0,1\}\}$ is countable. Then we have

$$x_1 = 010100....$$
$$x_2 = 101111....$$
$$\vdots$$

Then we define a sequence $x \notin X$ by moving down the diagonal and switching the values from 0 to 1 or from 1 to 0. Hence X is uncountable. □

PROPOSITION 1.3.10. *The power set* $\mathcal{P}(\mathbb{N})$ *of the natural numbers* $\mathbb{N}$ *is uncountable.*

PROOF. Let $C = \cup_{n\in\mathbb{N}}$ be a countable collection of subsets of $\mathbb{N}$. Define $X \subset \mathbb{N}$ by

$$X = \{n \in \mathbb{N} : n \in X_n\}$$

. Claim: $X \neq X_n$ for every $n \in \mathbb{N}$. Since either $n \in X$ and $n \notin X_n$ or $n \notin X$ and $n \in X_n$.
Thus $X \notin C$ and so no countable collection of subsets of $\mathbb{N}$ includes all of the subsets of $\mathbb{N}$. □

We introduce two crucial notions: the infimum and supremum of a set. First we provide some preliminaries.

DEFINITION 1.3.11. Let A be a non-empty subset of $\mathbb{R}$

- If there exists $M \in \mathbb{R}$ such that $a \leq M$ for all $a \in A$, then M is an *upper bound* of A. We call A *bounded above*.
- If there exists $m \in \mathbb{R}$ such that $m \leq a$ for all $a \in A$, then m is a *lower bound* of A.
- If there exist lower and upper bounds, then we say that A is *bounded*. We call A *bounded below*.

DEFINITION 1.3.12 (Infimum and Supremum). Let A be a subset of $\mathbb{R}$.

- If m is a lower bound of A such that $m \geq m'$ for every lower bound m', then m is called the *infimum* of A, denoted by $m = \inf A$. Furthermore, if $\inf A \in A$, then we call it the minimum of A, $\min A$.
- If M is an upper bound of A such that $M' \geq M$ for every upper bound M', then M is called the *supremum* of A, denoted by $M = \sup A$.Furthermore, if $\sup A \in A$, then we call it the maximum of A, $\max A$.

Note that the infimum of a set A, as well as the supremum, are unique. The elementary argument is left as an exercise.
If $A \subset \mathbb{R}$ is not bounded above, then we define $\sup A = \infty$. Suppose that a subset A of $\mathbb{R}$ is not bounded below, then we assign $-\infty$ as its infimum.
We state a different formulation of the notions $\inf A$ and $\sup A$ that is just a reformulation of the definition.

LEMMA 1.8. *Let A be a subset of $\mathbb{R}$.*

- *Suppose A is bounded above. Then $M \in \mathbb{R}$ is the supremum of A if and only if the following two conditions are satisfied:*
  (1) *For every $a \in A$ we have $a \leq M$.*
  (2) *Given $\varepsilon > 0$, there exists $a \in A$ such that $M - \varepsilon < a$.*
- *Suppose A is bounded below. Then $m \in \mathbb{R}$ is the infimum of A if and only if the following two conditions are satisfied:*
  (1) *For every $a \in A$ we have $m \leq a$.*
  (2) *Given $\varepsilon > 0$, there exists $a \in A$ such that $a < m + \varepsilon$.*

LEMMA 1.9. *Suppose A is a bounded subset of A. Then* $\inf A \leq \sup A$

For $c \in \mathbb{R}$ we define the *dilate* of a set A by $cA := \{b \in \mathbb{R} : b = ca \quad \text{for } a \in A\}$.

LEMMA 1.10 (Properties). *Suppose A is a subset of $\mathbb{R}$.*

(1) *For $c > 0$ we have* $\sup cA = c \sup A$ *and* $\inf cA = c \inf A$.
(2) *For $c < 0$ we have* $\sup cA = c \inf A$ *and* $\inf cA = c \sup A$.
(3) *Suppose A is contained in a subset B. If* $\sup A$ *and* $\sup B$ *exist, then* $\sup A \leq \sup B$. *In words, making a set larger, increases its supremum.*
(4) *Suppose A is contained in a subset B. If* $\inf A$ *and* $\inf B$ *exist, then* $\inf A \geq \inf B$. *In words, making a set smaller increases its infimum.*
(5) *Suppose $A \subset B$ are non-empty subsets of $\mathbb{R}$ such that $x \leq y$ for all $x \in A$ and $y \in B$. Then* $\sup A \leq \inf B$.
(6) *If A and B are non-empty subsets of $\mathbb{R}$, then* $\sup(A + B) = \sup A + \sup B$ *and* $\inf(A + B) = \inf A + \inf B$

PROOF. (1) We prove that $\sup cA = c \sup A$ for positive c. Suppose $c > 0$. Then $cx \leq M \Leftrightarrow x \leq M/c$. Hence M is an upper bound of cA if and only if M/c is an upper bound of A. Consequently, we have the desired result.

(2) Without loss of generality we set $c = -1$. Let $a \in A$ (we assume that the set A is non-empty, otherwise there is nothing interesting here). Then as a lower bound for A, $\inf A \leq a$. Moreover, as an upper bound for A, $a \leq \sup A$. Using transitivity, we conclude that $\inf A \leq \sup A$.

We now prove the second identity. Keep in mind that the supremum of a set is its **least upper bound**, while the infimum is its **greatest lower bound**.

For any $a \in A$, $\inf A \leq a$, so $-\inf A \geq -a$, showing that $-\inf A$ is an upper bound for $-A$. Therefore, $-\inf A \geq \sup(-A)$, which implies $\boxed{\inf A \leq -\sup(-A)}$.

For any $a \in A$ we have $-a \in -A$, so $-a \leq \sup(-A)$, which implies $a \geq -\sup(-A)$. Therefore, $-\sup(-A)$ is a lower bound for A, so $\boxed{-\sup(-A) \leq \inf A}$.

The two boxed inequalities prove the identity $\inf A = -\sup(-A)$.

(3) Since $\sup B$ is an upper bound of B, it is also an upper bound of A, i.e. $\sup A \leq \sup B$.

(4) Analogously to (iii).

(5) Since $x \leq y$ for all $x \in A$ and $y \in B$, y is an upper bound of A. Hence $\sup A$ is a lower bound of B and we have $\sup A \leq \inf B$.

(6) By definition $A + B = \{c : c = a + b \,\text{for some}\, a \in A, b \in B\}$ and thus $A + B$ is bounded above if and only if A and B are bounded above. Hence $\sup(A + B) < \infty$ if and only if $\sup A$ and $\sup B$ are finite. Take $a \in A$ and $b \in B$, then $a + b \leq \sup A + \sup B$. Thus $\sup A + \sup B$ is an upper bound of $A + B$:

$$\sup(A + B) \leq \sup A + \sup B.$$

The reverse direction is a little bit more involved. Let $\varepsilon > 0$. Then there exists $a \in A$ and $b \in B$ such that

$$a > \sup A - \varepsilon/2, \quad b > \sup B - \varepsilon/2.$$

Thus we have $a + b > \sup A + \sup B - \varepsilon$ for every $\varepsilon > 0$, i.e. $\sup(A + B) \geq \sup A + \sup B$.

The other statements are assigned as exercises. □

One reason for the relevance of the notions of supremum and infimum is in the formulation of properties of functions.

Definition 1.3.13. Let f be a function with domain X and range $Y \subseteq \mathbb{R}$. Then

$$\sup_X f = \sup\{f(x) : x \in X\}, \quad \inf_X f = \inf\{f(x) : x \in X\}.$$

If $\sup_X f$ is finite, then f is bounded from above on A, and if $\inf_X f$ is finite we call f bounded from below. A function is bounded if both the supremum and infimum are finite.

Lemma 1.11. *Suppose that $f, g : X \to \mathbb{R}$ and $f \leq g$, i.e. $f(x) \leq g(x)$ for all $x \in X$. If g is bounded from above, then $\sup_X f \leq \sup_A g$. Assume that f is bounded from below. Then $\inf_X f \leq \inf_X g$.*

Proof. Follows from the definitions. □

The supremum and infimum of functions do not preserve strict inequalities. Define $f, g : [0,1] \to \mathbb{R}$ by $f(x) = x$ and $g(x) = x + 1$. Then we have $f < g$ and

$$\sup_{[0,1]} f = 1, \quad \inf_{[0,1]} f = 0, \quad \sup_{[0,1]} g = 2, \quad \inf_{[0,1]} g = 1.$$

Hence we have $\sup_{[0,1]} f > \inf_{[0,1]} g$.

LEMMA 1.12. *Suppose f, g are bounded functions from X to $\mathbb{R}$ and c a positive constant. Then*

$$\sup_X(f + cg) \leq \sup_X f + c \sup_X g \quad \inf_X(f + cg) \geq \inf_X f + c \inf_X g.$$

The proof is left as an exercise. Try to convice yourself that the inequalities are in general strict, since the functions f and g may take values close to their suprema/infima at different points in X.

LEMMA 1.13. *Suppose f, g are bounded functions from X to $\mathbb{R}$. Then*

$$|\sup_X f - \sup_X g| \leq \sup_X |f - g|, \quad |\inf_X f - \inf_X g| \leq \sup_X |f - g|$$

LEMMA 1.14. *Suppose f, g are bounded functions from X to $\mathbb{R}$ such that*

$$|f(x) - f(y)| \leq |g(x) - g(y)| \quad \textit{for all } x, y \in X.$$

Then

$$\sup_X f - \inf_X f \leq \sup_X g - \inf_X g.$$

Recall that a sequence (x_n) of real numbers is an ordered list of numbers x_n, indexed by the natural numbers. In other words, (x_n) is a function f from $\mathbb{N}$ to $\mathbb{R}$ with $f(n) = x_n$. A sequence is a function from $\mathbb{N}$ to $\mathbb{R}$ or $\mathbb{C}$, so the properties of the inf and sup for functions apply to sequences as well.

CHAPTER 2

Normed spaces and innerproduct spaces

In order to measure the length of a vector and to define a distance between vectors we introduce the notion of a norm of a vector. Norms may be a tool to specify properties of a class of vectors in a convenient form. We review basic aspects of vector spaces before we define normed vector spaces.

2.1. Vector spaces

Vector spaces and linear mappings between them are a useful tool for engineers, scientists and mathematicians, aka Linear Algebra.

Vector spaces formalize the notion of linear combinations of objects that might be vectors in the plane, polynomials, smooth functions, sequences. Many problems in engineering, mathematics and science are naturally formulated and solved in this setting due to their linear nature. Vector spaces are ubiquitous for several reasons, e.g. as linear approximation of a non-linear object, or as building blocks for more complicated notions, such as vector bundles over topological spaces. We restrict our discussion to complex and real vector spaces.

A set V is a vector space if it is possible to build linear combinations out of the elements in V. More formally, on V we have the operations of addition of vectors and multiplication by scalars. The scalars will be taken from a field $\mathbb{F}$, which is either the real numbers $\mathbb{R}$ or $\mathbb{C}$. In various situations $\mathbb{F}$ might also be a finite field or a field different from $\mathbb{R}$ and $\mathbb{C}$. If it is necessary we will refer to these vector spaces as real or complex vector spaces.

Developing an understanding of these vector spaces is one of the main objectives of this course. The axioms for a vector space specify the properties that addition of vectors and scalar multiplication.

Definition 2.1.1. A *vector space* over a field $\mathbb{F}$ is a set V together with the operations of addition $V \times V \to V$ and scalar multiplication $\mathbb{F} \times V \to V$ satisfying the following properties:

(1) Commutativity: $u + v = v + u$ for all $u, v \in V$ and $(\lambda\mu v) = \lambda(\mu v)$ for all $\lambda, \mu \in \mathbb{F}$;
(2) Associativity: $(u + v) + w = u + (v + w)$ for all $u, v, w \in V$;
(3) Additive identity: There exists an element $0 \in V$ such that $0 + v = v$ for all $v \in V$;
(4) Additive inverse: For every $v \in V$, there exists an element $w \in V$ such that $v + w = 0$;
(5) Multiplicative identity: $1v = v$ for all $v \in V$;

(6) Distributivity: $\lambda(u+v) = \lambda u + \lambda v$ and $(\lambda+\mu)u = \lambda u + \mu u$ for all $u, v \in V$ and $\lambda, \mu \in \mathbb{F}$.

The elements of a vector space are called vectors. Given $v_1, ..., v_n$ be in V and $\lambda_1, ..., \lambda_n \in \mathbb{F}$ we call the vector

$$v = \lambda_1 v_1 + \cdots + \lambda_n v_n$$

a *linear combination.*

Our focus will be on three classes of examples.

Examples 2.1.2. We define some useful vector spaces.

- **Spaces of n-tuples:** The set of tuples $(x_1, ..., x_n)$ of real and complex numbers are vector spaces $\mathbb{R}^n$ and $\mathbb{C}^n$ with respect to component-wise addition and scalar multiplication: $(x_1, ..., x_n) + (y_1, ..., y_n) = (x_1 + y_1, ..., x_n + y_n)$ and $\lambda(x_1, ..., x_n) = (\lambda x_1, ..., \lambda x_n)$.
- The set of functions $\mathbb{F}(X, Y)$ of a set X to a set Y: $\lambda f + \mu g(x) := (\lambda f + \mu g)(x)$ for all $x \in X$.
- The space of polynomials of degree at most n, denoted by $\mathcal{P}_n$, where we define the operations of multiplication and addition coefficient-wise: For $p(x) = a_0 + a_1 x + \cdots a_n x^n$ and $q(x) = b_0 + b_1 x + \cdots b_n x^n$ we define

$(p+q)(x) = (a_0+b_0)+(a_1+b_1)x+\cdots(a_n+b_n)x^n$ and $(\lambda p)(x) = \lambda a_0+\lambda a_1 x+\cdots \lambda a_n x^n$ for $\lambda \in \mathbb{F}$.

  The space of all polynomials $\mathcal{P}$ is the vector space of polynomials of arbitrary degrees.
- **Sequence spaces**: s denotes the set of sequences, c the set of all convergent sequences, c_0 the set of all convergent sequences tending to 0, c_f the set of all sequences with finitely many non-zero elements.
- **Function spaces:** The set of continuous functions $C(I)$ on an interval of $\mathbb{R}$, popular choices for I are $[0, 1]$ and $\mathbb{R}$. We define addition and scalar multiplication as follows: For $f, g \in C(I)$ and $\lambda \in \mathbb{F}$

$$(f + g)(x) = f(x) + g(x) \quad \text{and } (\lambda f)(x) = \lambda f(x).$$

  We denote by $C^{(n)}(I)$ the space of n-times continuously differentiable functions on I and the space $C^\infty(I)$ of smooth functions on I is the space of functions with infinitely many continuous derivatives. More generally, the set $\mathcal{F}(X)$ of functions from a set X to $\mathbb{F}$ is a vector space for the operations defined above. Note that $\mathcal{F}(\{1, 2, ..., n\})$ is just $\mathbb{F}^n$ and hence the first class of examples.
- **Spaces of matrices**: Denote by $\mathcal{M}_{m\times n}(\mathbb{C})$ the space of complex $m \times n$ matrices where we define addition and scalar multiplication entry-wise: For $A = (a_{ij})_{i,j}$ and $B = (a_{ij})_{i,j}$ where $i = 1, ..., m$ and $j = 1, ...n$ we define

$$A + B := (a_{ij} + b_{ij})_{i,j} \qquad \text{and } \alpha(a_{ij})_{ij} = (\alpha a_{ij})_{ij},\ \alpha \in \mathbb{F}.$$

There are relations between the vector spaces in the aforementioned list. We start with clarifying their inclusion properties.

Definition 2.1.3. A subset W of a vector space V is called a *subspace* if W is a vector subspace with respect to addition and scalar multiplication of V.

One way to express this more concretely is stated in the next lemma:

LEMMA 2.1. *A subset W of a vector space V is a subspace if and only if W is closed under linear combinations: For any $\alpha, \beta \in \mathbb{F}$ and $w_1, w_2 \in W$ we have $\alpha_1 w_1 + \alpha_2 w2 \in W$. Equivalently, we have that the subset W of a vector space V is a subspace if and only if*

(1) $0 \in W$;
(2) $w_1 + w_2 \in W$ *for any* $w_1, w_2 \in W$;
(3) αw *for any* $\alpha \in \mathbb{F}$ *and any* $w \in W$.

Consequently, we have a way to decide when a subset of a vector space is not a subspace.

LEMMA 2.2. *A subset W of a vector space V is a not a subspace if one of the following conditions holds:*

(1) $0 \notin W$;
(2) *There are some* $w_1, w_2 \in W$ *such that* $w_1 + w_2 \in W$;
(3) *There is a vector* $w \in W$ *such that* $-w$ *is not in* $= W$.

This is the contrapositve of the preceding lemma.

Here are some examples of vector subspaces:

$$\mathcal{P}_n \subset \mathcal{P} \subset \mathcal{F}, \qquad C^\infty(I) \subset C^{(n)}(I) \subset C(I), \qquad c_f \subset c_0 \subset c \subset s$$

.

We define the linear span, span$S$, of a subset S of a vector space V to be the intersection of all subspaces of V containing S.

Linear transformations T between vector spaces V and W are mappings T that respect linear transformations:

$$T(\alpha_1 v_1 + \alpha_2 v_2) = \alpha_1 T(v_1) + \alpha_2 T(v_2) \qquad \text{for any } v_1, v_2 \in V, \alpha, \beta \in \mathbb{F}.$$

We denote by $\mathcal{L}(V, W)$ the set of all linear transformations between V and W and it is a subset of the vector space of all functions $f : V \to W$. Furthermore $\mathcal{L}(V, W)$ is a vector space:

$$\mathcal{L}(V, W) \subseteq \mathcal{F}(V, W).$$

EXAMPLE 2.1.4. Let D denote the differentiation operator $Df = f'$. Then $D : C^{(1)}(a, b) \to C(a, b)$ is a linear transformation.

Linear transformations have some useful properties.

LEMMA 2.3. *For any $T \in \mathcal{L}(V, W)$ we have $T(0) = 0$.*

PROOF. We have that $v + 0 = v$ for any $v \in V$; in particular for $v = 0$:

$$T(0) = T(0 + 0) = T(0) + T(0)$$

and after subtracting $T(0)$ we get $T(0) = 0$. □

The *kernel* of $T \in \mathcal{L}(V, W)$ is the set

$$\ker(T) := \{v \in V | \, Tv = 0\},$$

i.e. $\ker(T) = T^{-1}(0)$.

LEMMA 2.4. *For a linear transformation $T : V \to W$ the kernel of T is a subspace of V.*

PROOF. Suppose $v_1, v_2 \in \ker(T)$. Then for any scalars α_1, α_2 we have

$$T(\alpha v_1 + \alpha_2 v_2) = \alpha_1 T(v_1) + \alpha_2 T(v_2) = \alpha_1 \cdot 0 + \alpha_2 \cdot 0 = 0$$

and thus $\alpha v_1 + \alpha_2 v_2 \in \ker(T)$. □

The range of T is a subspace of W, too.

LEMMA 2.5. *The range of a linear transformation $T : V \to W$ is a subspace of W.*

PROOF. Exercise, see problem set 2. □

There is some natural operations for vector spaces.

DEFINITION 2.1.5. Let V and W be subspaces of Z.

(1) The **sum** of V and W is defined by $V + W := \{z \in Z | \, z = v + w \; v \in V, w \in W\}$.
(2) The **intersection** of V and W is defined by $V \cap W := \{z \in Z | \, z \in V \cap W\}$.

From the definitions we see that $V + W$ and $V \cap W$ are subspace of Z. We introduce some more notions: If the sum of the subspaces V and W equals Z, then we say that Z is the sum of V and W, i.e. $V + W = Z$. If in addition, the subspaces are disjoint subsets, $U \cap V = \{0\}$, then we refer to the sum of V and W as the **direct sum**.

LEMMA 2.6. *Let I be an index set. Given vector spaces V_i for any $i \in I$. Then $\cap_{i \in I} V_i$ is a vector space.*

PROOF. Exercise, see problem set 2. □

DEFINITION 2.1.6. Let S be a nonempty subset of a vector space V. Then we define the **span** of S, $\text{span}(S)$, as the intersection of all subspaces of V that contain S.

LEMMA 2.7. *Let $S \subset V$ be a nonempty subset. Then*

$$\text{span}(S) = \{\lambda_1 v_1 + \ldots + \lambda_n v_n \; : \; v_1, \ldots, v_n \in S \text{ and } \lambda_1, \ldots, \lambda_n \in \mathbb{F}\}.$$

By definition, $\text{span}(S)$ is the intersection of all subspaces W of V that contain the set S. From the preceding lemma, it follows that $\text{span}(S)$ is a subspace of V, hence it is the *smallest* subspace of V that contains S.

Let us denote

$$W := \{\lambda_1 v_1 + \ldots + \lambda_n v_n \; : \; v_1, \ldots, v_n \in S \text{ and } \lambda_1, \ldots, \lambda_n \in \mathbb{F}\},$$

so W is the set of all linear combinations with elements in S.

Being a subspace of V, $\text{span}(S)$ must contain all such linear combinations, so we must have that

$$W \subset \text{span}(S).$$

All we have left to show is that W is a subspace of V. This is not hard to see, since linear combinations of linear combinations are linear combinations as well.

Indeed, let $a, b \in \mathbb{F}$ and let $w_1, w_2 \in W$, so

$$\begin{aligned} w_1 &= \lambda_1 v_1 + \ldots + \lambda_n v_n \quad \text{with } v_1, \ldots, v_n \in S, \\ w_2 &= \mu_1 u_1 + \ldots + \mu_m u_m \quad \text{with } u_1, \ldots, u_m \in S. \end{aligned}$$

Then

$$a\,w_1 + b\,w_2 = a\lambda_1\,v_1 + \ldots + a\lambda_n\,v_n + b\mu_1\,u_1 + \ldots + b\mu_m\,u_m,$$

and since $v_1, \ldots, v_n, u_1, \ldots, u_m \in S$, it follows that $a\,w_1 + b\,w_2 \in W$.

Therefore, W is a subspace of V that contains S, so we must have

$$\operatorname{span}(S) \subset W.$$

Together with the previous inclusion, this proves the equality of the two sets.

2.2. Normed spaces

The norm on a general vector space generalizes the notion of the length of a vector in $\mathbb{R}^2$ and $\mathbb{R}^3$.

DEFINITION 2.2.1. A *normed space* is a vector space X together with a function $\|.\| : X \to \mathbb{R}$, the *norm* on X, such that for all $x, y \in X$ and $\lambda \in \mathbb{R}$:

(1) *Positivity:* $0 \leq \|x\| < \infty$ and $\|x\| = 0$ if and only if $x = 0$;
(2) *Homogeneity:* $\|\alpha x\| = |\alpha|\|x\|$ for $\alpha \in \mathbb{F}$;
(3) *Triangle inequality:* $\|x + y\| \leq \|x\| + \|y\|$.

We denote this normed space by $(X, \|.\|)$

A norm gives a way to measure the distance between two vectors by $d(x, y) := \|x - y\|$. We refer to d as the metric associated to the norm $\|.\|$.

PROPOSITION 2.2.2. *Let $(X, \|.\|)$ be a normed space. Then $d : X \times X \to \mathbb{R}$ defined by $d(x, y) = \|x - y\|$ satisfies for all $x, y, z \in X$*

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

PROOF. The properties (i)-(iii) are direct consequences of the axioms for a norm. In particular, (i) follows from property (1) of a norm, (ii) is derived from property (ii) of a norm for $\lambda = -1$ and (iii) is deduced from property (3) of a norm. □

The metric d on X is also compatible with the linear structure of a vector space:

- *Translation invariance:* $d(x + z, y + z) = d(x, y)$ for all $x, y, z \in X$;
- *Symmetry:* $d(\alpha x, \alpha y) = |\alpha| d(y, x)$ for all $x, y \in X$ and $\alpha \in \mathbb{F}$.

The function $d(x, y) = \|x - y\|$ on the vector space $\mathbb{R}$ is an example of a distance function on $\mathbb{R}$, aka as a metric.

The metric d on X gives us a way to generalize intervals in $\mathbb{R}$ to so-called balls.

DEFINITION 2.2.3. For $r > 0$ and $x \in X$ we define the *open ball* $B_r(x)$ of radius r and center x as the set

$$B_r(x) = \{y \in X : \|x - y\| < r\},$$

and the *closed ball* $\overline{B}_r(x)$ of radius r and center x as

$$\overline{B}_r(x) = \{y \in X : \|x - y\| \leq r\}.$$

The translation invariance and the homogeneity imply that the ball $B_r(x)$ is the image of the unit ball $B_1(0)$ centered at the origin under the (affine) mapping $f(y) = ry + x$.

The balls $B_r(x)$ have another peculiar feature. Namely, these are convex subsets of X.

DEFINITION 2.2.4. Let X be a vector space.

- For two points $x, y \in X$ the *interval* $[x, y]$ is the set of points $\{z | \, z = \lambda x + (1-\lambda)y \; 0 \leq \lambda \leq 1\}$.
- A subset E of X is called *convex* if for any two points $x, y \in E$ the interval $[x, y]$ is also in E.

The notion of convexity is central to the theory of vector spaces and enters in an intricate manner in functional analysis, numerical analysis, optimization, etc. .

LEMMA 2.8. *Let $(X, \|.\|)$ be a normed vector space. Then the unit ball $B_1(0) = \{x \in X | \, \|x\| \leq 1\}$ is a convex set.*

PROOF. For $x, y \in B_1(0)$ we have that $\|\lambda x + (1-\lambda)y\| \leq |\lambda|\|x\| + |1-\lambda|\|y\| = 1$, because $\|x\|, \|y\|$ are both less than or equal to 1. Thus $\lambda x + (1-\lambda)y \in B_1(0)$. □

The real numbers with the absolute value is a normed space $(\mathbb{R}, |.|)$ and the open ball $B_r(x)$ is the open interval $(x - r, x + r)$ and $\overline{B}_r(x)$ is the closed interval $[x - r, x + r]$.

LEMMA 2.9 (Reverse triangle inequality). *Let $(X, \|.\|)$ be a normed space. Then we have*

$$|\|x\| - \|y\|| \leq \|x - y\| \qquad \textit{for all } x, y \in X.$$

PROOF. See problem set 3. □

A fundamental class of normed spaces is $\mathbb{R}^n$ with the ℓ^p-norms.

DEFINITION 2.2.5. For $p \in [1, \infty)$ we define the **p-norm**, denoted by $\|.\|_p$, on $\mathbb{R}^n$ by assigning to $x = (x_1, ..., x_n) \in \mathbb{R}^n$ the number $\|x\|_p$:

$$\|x\|_p = (|x_1|^p + |x_2|^p + \cdots |x_n|^p)^{1/p}$$

. For $p = \infty$ we define the ℓ^∞-norm $\|.\|_\infty$ on $\mathbb{R}$ by

$$\|x\|_\infty = \max |x_1|, ..., |x_n|.$$

The notation for $\|.\|_\infty$ is justified by the fact that it is the limit of the $\|.\|_p$-norms.

LEMMA 2.10. *For $x \in \mathbb{R}^n$ we have that*

$$\|x\|_\infty = \lim_{p \to \infty} \|x\|_p.$$

PROOF. Without loss of generality we assume that the largest component of x, the $\|x\|_\infty$, to be x_n. For $1 \leq p < \infty$ we have

$$\|x\|_p = (|x_1|^p + |x_2|^p + \cdots |x_n|^p)^{1/p} = \|x\|_\infty ((\frac{|x_1|}{\|x\|})^p + \frac{|x_2|}{\|x\|})^p + \cdots + 1)^{1/p},$$

since $\frac{|x_i|}{\|x\|}) < 1$ for $i = 1, ..., n-1$ we have $\lim_{p\to\infty} \frac{|x_1|}{\|x\|})^p = 0$. Thus we have

$$\lim_{p\to\infty} \|x\|_p = \|x\|_\infty.$$

□

In the proof of the triangle inequality for the p-norms we have to rely on some inequalities: Hölder's inequality and Young's inequality.

For $p \in (1, \infty)$ we define its *conjugate* q as the number such that

$$\frac{1}{p} + \frac{1}{q} = 1.$$

If $p = 1$, then we define its conjugate q to be ∞ and vice versa for $p = \infty$ we set $q = 1$.

LEMMA 2.11 (Young's inequality). *For $p \in (1, \infty)$ and q its conjugate we have*

$$ab \leq \frac{a^p}{p} + \frac{b^q}{q},$$

for any non-negative real numbers a, b.

PROOF. Consider the function $f(x) = x^{p-1}$ and integrate this with respect to x from zero to a. Now take the inverse function of f given by $f^{-1}(y) = y^{q-1}$, where we used that $1/(p-1) = q-1$ for conjugate exponents p and q. Let us integrate f^{-1} from zero to b. Then the sum of these two integrals always exceeds the product ab, see figure. Note that the two integrals are given by a^p/p and b^q/q. Hence we have established Young's inequality. □

A consequence of Young's inequality is Hölder's inequality.

LEMMA 2.12. *Suppose $p \in (1, \infty)$ and $x = (x_1, ..., x_n)$ and $y = (y_1, ..., y_n)$ are vectors in $\mathbb{R}^n$. Then*

$$|\sum_{i=1}^n x_i y_i \leq \Big(\sum_{i=1}^n |x_i|^p\Big)^{1/p} \Big(\sum_{i=1}^n |y_i|^q\Big)^{1/q}.$$

PROOF. Set $a_i = |x_i|/(\sum_{i=1}^n |x_i|^p)^{1/p}$ and $b_i = |y_i|/(\sum_{i=1}^n |y_i|^q)^{1/q}$. Then we have $\sum_i a_i^p = 1$ and $\sum_i b_i^q = 1$. By Young's inequality

$$\sum_{i=1}^n |x_i||y_i| \leq (\sum_{i=1}^n |x_i|^p)^{1/p} (\sum_{i=1}^n |y_i|^q)^{1/q}.$$

□

The unit balls of $(\mathbb{R}^2, \|.\|_1)$, $(\mathbb{R}^2, \|.\|_2)$ and $(\mathbb{R}^2, \|.\|_\infty)$ indicate the different nature of these norms.

PROOF. Positivity and homogeneity are consequences of the corresponding properties of the absolute value of a real number. The triangle inequality is the non-trivial assertion that we split up in three cases $p = 1$, $p = \infty$ and $p \in (1, \infty)$. Let $x = (x_1, ..., x_n)$ and $y = (y_1, ..., y_n)$ be points in $\mathbb{R}^n$.

(1) For $p = 1$ we have

$$\|x + y\|_1 = |x_1 + y_1| + \cdots + |x_n + y_n| \leq |x_1| + |y_1| + \cdots + |x_n| + |y_n| \leq \|x\|_1 + \|y\|_1$$

.

(2) For $p = \infty$ the argument is similar:

$$\begin{aligned}\|x+y\|_\infty &= \max\{|x_1+y_1|, ..., |x_n+y_n|\} \\ &\leq \max\{|x_1|+|y_1|, ..., |x_n|+|y_n|\} \\ &\leq \max\{|x_1|, ..., |x_n|\} + \max\{|y_1|, ..., |y_n|\} = \|x\|_\infty + \|y\|_\infty.\end{aligned}$$

(3) The general case $p \in (1, \infty)$: The triangle inequality follows from Hölder's inequality.

$$\begin{aligned}\|x+y\|_p^p &= \sum_{i=1}^n |x_i+y_i|^p \\ &\leq \sum_{i=1}^n |x_i+y_i|^{p-1}(|x_i|+|y_i|) \\ &\leq \sum_{i=1}^n |x_i+y_i|)^{p-1}|x_i| + \sum_{i=1}^n |x_i+y_i|^{p-1}|y_i| \\ &\leq \Big(\sum_{i=1}^n |x_i+y_i|^p\Big)^{1/q}\Big(\Big(\sum_{i=1}^n |x_i|^p\Big)^{1/p} + \Big(\sum_{i=1}^n |y_i|^p\Big)^{1/p}\Big) \\ &= \|x+y\|_p^{1/q}(\|x\|_p + \|y\|_p)\end{aligned}$$

Dividing by $\|x+y\|_p^{1/q}$ and using $1 - 1/q = 1/p$ we obtain the triangle inequality:

$$\|x+y\|_p \leq \|x\|_p + \|y\|_p.$$

Thus the space $\mathbb{R}^n$ with the p-norm $\|.\|_p$ is a normed space for $p \in [1, \infty]$. □

The triangle inequality for p-norms on $\mathbb{R}^n$ is also known as **Minkowski's inequality**:

$$(\sum_{i=1}^n |x_i+y_i|^p)^{1/p} \leq (\sum_{i=1}^n |x_i|^p)^{1/p} + (\sum_{i=1}^n |y_i|^p)^{1/p}.$$

There are variations of the $(\mathbb{R}^n, \|.\|_p)$ with relevance in engineering, physics and mathematics. (i) Replace the real scalars by complex scalars $(\mathbb{C}^n, \|.\|_p)$; (ii) Replace $\mathbb{R}^n$ by the vector space of sequences s; (iii) Deal with complex-valued sequences, (iv) Consider continuous functions and define norms in terms of integrals instead of sums for sequences.

Before we present these classes of normed spaces, we show that the vector space of $m \times n$-matrices is a normed spaces, too.

Define a norm on $\mathcal{M}_{m\times n}(\mathbb{F})$ by picking a norm on $\mathbb{F}^{mn}$: For $1 \leq p < \infty$ we define $\|A\|_{(p)} = (\sum_{i=1}^m \sum_{j=1}^n |a_{ij}|^p)^{1/p}$ or $\|A\|_{(\infty)} = \max|a_{ij}|$ for $A \in \mathcal{M}_{m\times n}(\mathbb{F})$. The case $p = 2$ is of special interest and is known as the Frobenius norm.

PROPOSITION 2.2.6. *For $1 \leq p \leq \infty$ we have that $(\mathcal{M}_{m\times n}(\mathbb{F}), \|.\|_p)$ is a normed space.*

The identification of $\mathcal{M}_{m\times n}(\mathbb{F})$ with the vector space $\mathbb{F}^{mn}$ gives us this result.

PROPOSITION 2.2.7. *Let $\mathbb{C}^n$ be the vector space of complex n-tuples $z = (z_1, ..., z_n)^T$, $z_i \in \mathbb{C}$ for $i = 1, ..., n$. For $1 \leq p < \infty$ we define*

$$\|z\|_p = (\sum_{i=1}^{n})|z_i|^p)^{1/p},\ z \in \mathbb{C}^n$$

and for $p = \infty$ we have $\|z\|_\infty := \max|z_i| : i = 1, ..., n.$ where $z_i \in \mathbb{C}$ and $|z_i| = (z_i\overline{z_i})^{1/2}$ denotes the modulus of z_i. Then $(\mathbb{C}^n, \|.\|)_p$ is a normed space for $1 \leq p \leq \infty$. The proof of $\mathbb{R}^n$ goes through without any changes.

PROOF. Young's inequality is a statement about non-negative numbers which in this case are modulus of complex numbers. Hence Young's inequality is valid in this case as well and consequently Hölder's inequality. The later is the key to prove the triangle inequality. □

Recall that s denotes the vector space of all sequences with values in $\mathbb{R}$ or $\mathbb{C}$. We define for $1 \leq p < \infty$ the space ℓ^p as the set of all sequences $x = (x_1, x_2, ...)$ satisfying

$$\|x\|_p := (|x_1|^p + |x_2|^p + \cdots)^{1/p} < \infty,$$

and ℓ^∞ denotes the space of all bounded sequences $(s, \|.\|_\infty)$ with

$$\|x\|_\infty := \sup_{i \in \mathbb{N}} |x_i|,$$

where $|.|$ denotes the absolute value of a real number or the modulus of a complex number, respectively.

LEMMA 2.13 (Hölder's inequality). *For $1 \leq p \leq \infty$ and q its conjugate index we have for $x \in \ell^p$ and $y \in \ell^q$*

$$\sum_{i=1}^{\infty} |x_i||y_i| \leq (\sum_{i=1}^{\infty} |x_i|^p)^{1/p} (\sum_{i=1}^{\infty} |y_i|^q)^{1/q}.$$

Since Hölder's inequality is true for all $n \in \mathbb{N}$ we deduce that the limits of the partial sums in question also satisfy these inequalities. Hence we deduce the desired inequality for sequences instead of n-tuples.

PROPOSITION 2.2.8. *For $1 \leq p \leq \infty$ we have that ℓ^p is a normed vector space.*

PROOF. First we show that ℓ^p is a vector space for $p \in [1, \infty)$: For $\alpha \in \mathbb{F}$ and $x \in \ell^p$ we have $\alpha x \in \ell^p$. One has to work a little bit to see that for $x, y \in \ell^p$ also $x + y \in \ell^p$:

$$\begin{aligned}
\|x + y\|_p^p &= \sum_{i=1}^{\infty} |x_i + y_i|^p \\
&\leq 2^p \sum_{i=1}^{\infty} \max\{|x_i|, |y_i|\}^p \\
&= 2^p \sum_{i=1}^{\infty} |\max\{|x_i|, |y_i|\}|^p \\
&\leq 2^p (\sum_{i=1}^{\infty} |x_i|^p + \sum_{i=1}^{\infty} |y_i|^p) = 2^p (\|x\|_p^p + \|y\|_p^p) < \infty.
\end{aligned}$$

The norm properties may be deduced as in the case of $\mathbb{F}^n$ since we have Hölder's inequality at our disposal. □

For $1 \leq p < \infty$ the spaces $(\ell^p, \|.\|_p)$ are subspaces of the vector space of sequences converging to zero, c_0. In contrast $(\ell^\infty, \|.\|_\infty)$ is the space of bounded sequences and is much larger than the other ℓ^p-spaces. We have the following inclusions:

LEMMA 2.14. *For $p_1 < p_2$ the space ℓ^{p_1} is a proper subspace of ℓ^{p_2}, i.e.*

$$\ell^1 \subset \ell^2 \subset \ell^\infty.$$

PROOF. See problem set 4. □

For example $(1/n)_n$ is in ℓ^p for $p \geq 2$, but not in ℓ^1.

We finish this section with normed spaces based on continuous functions.

DEFINITION 2.2.9. For $f \in C[a,b]$ we define its p-norm for $1 \leq p < \infty$ by

$$\|f\|_p = (\int_a^b |f(x)|^p dx)^{1/p}$$

and $\|f\|_\infty = \sup_{x\in[a,b]} |f(x)|$. We denote by $(C[a,b], \|.\|_p)$ the set of all functions satisfying $\|f\|_p < \infty$.

LEMMA 2.15 (Hölder's inequality). *For $1 \leq p \leq \infty$ and its conjugate exponent q we have*

$$\int_a^b |f(x)||g(x)|dx \leq \|f\|_p \|g\|_q.$$

PROOF. We assume without loss of generality that $\|f\|_p = 1 = \|g\|_q$. By Young's inequality we have

$$|f(x)||g(x)| \leq |f(x)|^p/p + |g(x)|^q/q$$

and thus

$$\int_a^b |f(x)||g(x)| \leq \frac{1}{p}\int_a^b |f(x)|^p dx + \frac{1}{q}\int_a^n |g(x)|^q dx = \|f\|_p\|g\|_q.$$

As in the case of $\mathbb{F}^n$ we are able to turn this inequality in the desired one. □

PROPOSITION 2.2.10. *The space $(C[a,b], \|.\|_p)$ is a normed space for $p \in [1,\infty]$.*

PROOF. As for ℓ^p-spaces we deduce that the $\|.\|_p$ is a vector space. The norm part is based on the validity of Hölder's inequality as above. □

We close with a way to construct a normed space out of given normed spaces. Let $\{X_1, \|.\|_{X_1}), ..., (X_1, \|.\|_{X_1})\}$ be given normed spaces. Then the direct product $X_1 \times \cdots \times X_n$ is a normed space for

$$\|(x_1, ..., x_n)\| := \|x_1\|_{X_1} + \cdots + \|x_n\|_{X_n}.$$

2.3. Innerproduct spaces

In this section we consider innerproduct spaces and we start with the case of real vector spaces and afterwards treat complex vector spaces.

For vectors in $\mathbb{R}^3$ we have the 'dot product' aka 'scalar product' that assigns to a pair of vectors $x = (x_1, x_2, x_3)$ and $y = (y_1, y_2, y_3)$ the number

$$\langle x, y\rangle = x_1y_1 + x_2y_2 + x_3y_3.$$

Pythagoras' theorem gives the length of $x = (x_1, x_2, x_3)$ as $\sqrt{x_1^2 + x_2^2 + x_3^2}$. Note that $\langle x, x\rangle = \sqrt{x_1^2 + x_2^2 + x_3^2}$. Innerproduct spaces are a generalization of these basic facts from Euclidean geometry to general vector spaces.

DEFINITION 2.3.1. Let X be a real vector space. An *innerproduct* on X is a map $\langle ., .\rangle : X \times X \to \mathbb{R}$ satisfying:

(1) (Linearity) For vectors $x_1, x_2, y \in X$ and scalars $\alpha_1, \alpha_2 \in \mathbb{R}$ we have $\langle \alpha_1 x_1 + \alpha_2 x_2, y\rangle = \alpha_1 \langle x_1, y\rangle + \alpha_2 \langle x_2, y\rangle$.
(2) (Symmetry) For vectors $x, y \in X$ we have $\langle x, y\rangle = \langle y, x\rangle$.
(3) (Positive definiteness) For any $x \in X$ we have $\langle x, x\rangle \geq 0$ and $\langle x, x\rangle = 0$ if and only if $x = 0$.

We call $(X, \langle ., .\rangle)$ an *innerproduct space* and define by $\|x\| := \langle x, x\rangle^{1/2}$.

Here is a reformulation of the positive definiteness of innerproducts.

LEMMA 2.16. *Suppose X is an innerproduct space. If $\langle x, y\rangle = 0$ for all $y \in X$, then $x = 0$.*

PROOF. Since $\langle x, y\rangle = 0$ holds for all $y \in X$, in particular for $y = x$ and thus $\langle x, x\rangle = 0$. Hence $x = 0$. □

Note that the symmetry and linearity in the first entry gives that $\langle ., .\rangle$ is bilinear: For vectors $x, y_1, y_2 \in X$ and scalars $\alpha_1, \alpha_2 \in \mathbb{R}$ we have $\langle x, \alpha_1 y_1 + \alpha_2 y_2\rangle\, y = \alpha_1 \langle x, y_1\rangle + \alpha_2 \langle x, y_2\rangle$.

EXAMPLE 2.3.2. The family of p-norms on $\mathbb{R}^n$, the space of sequences s and on the space of continuous functions $C[a, b]$ include for $p = 2$ important examples of innerproduct spaces.

There is a link between innerproducts and the length of x. Namely $\langle x, x\rangle^{1/2}$ is the length $\|x\|$ of x. The proof of this fact is based on a well-known inequality.

PROPOSITION 2.3.3 (Cauchy-Schwarz). *Suppose X is a real innerproduct space. Then for all $x, y \in X$ we have*

$$|\langle x, y\rangle| \leq \|x\|\|y\|.$$

We have $|\langle x, y\rangle| = \|x\|\|y\|$ if and only if $x = \alpha y$ for some $\alpha \in \mathbb{R}$.

PROOF. For any $t \in \mathbb{R}$ and $x, y \in X$ we have $\|x - ty\| \geq 0$. More explicitly, we have

$$\begin{aligned}\|x - ty\| = \langle x - ty, x - ty\rangle &= \langle x, x\rangle - t(\langle y, x\rangle + \langle x, y\rangle) + t^2 \langle y, y\rangle \\ &= \langle x, x\rangle - 2t \langle x, y\rangle + t^2 \langle y, y\rangle\end{aligned}$$

Suppose $y \neq 0$, otherwise there is nothing to show.

Hence we have

$$\begin{aligned} t^2 \langle y,y\rangle - 2t\langle x,y\rangle + \langle x,x\rangle &= \langle y,y\rangle \left(t^2 - 2t\frac{\langle x,y\rangle}{\langle y,y\rangle} + \frac{\langle x,x\rangle}{\langle y,y\rangle}\right) \\ &= \langle y,y\rangle \left(\left(t - \frac{\langle x,y\rangle}{\langle y,y\rangle}\right)^2 - \frac{\langle x,y\rangle^2}{\langle y,y\rangle^2} + \frac{\langle x,x\rangle}{\langle y,y\rangle}\right) \\ &= \langle y,y\rangle \left(\left(t - \frac{\langle x,y\rangle}{\langle y,y\rangle}\right)^2 + \frac{\langle x,x\rangle\langle y,y\rangle - \langle x,y\rangle^2}{\langle y,y\rangle^2}\right) \end{aligned}$$

Hence we have $\langle x,x\rangle\langle y,y\rangle - \langle x,y\rangle^2 \geq 0$, i.e.

$$|\langle x,y\rangle| \leq \langle x,x\rangle^{1/2}\langle y,y\rangle^{1/2}.$$

The assertion about the equality follows from the proof of the Cauchy-Schwarz inequality, since $\|x - ty\| = 0$ if and only if $x = \alpha y$ for some $\alpha \in \mathbb{R}$. □

As a consequence we deduce that innerproduct spaces $(X, \langle .,.\rangle)$ are normed spaces for $\|x\| = \langle x,x\rangle^{1/2}$.

Proposition 2.3.4. *For $(X, \langle .,.\rangle)$ the expression $\|x\| = \langle x,x\rangle^{1/2}$ defines a norm on X.*

Proof. Homogeneity follows from the linearity of the innerproduct. The triangle inequality requires some work:

$$\|x+y\|^2 = \|x\|^2 + \|y\|^2 + 2\langle x,y\rangle \leq \|x\|^2 + \|y\|^2 + 2\|x\|\|y\|,$$

so the right side is $(\|x\| + \|y\|)^2$, where we applied Cauchy-Schwarz to bound the innerproduct in terms of the norms of its elements. Thus we have $\|x+y\| \leq \|x\| + \|y\|$. □

Example 2.3.5.

(1) The sequence space ℓ^2 is an innerproduct space for real-valued sequences $(x_i), (y_i)$

$$\langle x,y\rangle = \sum_{i=1}^{\infty} x_i y_i.$$

The sequence space ℓ^2 was the first example of an innerproduct space, studied by D. Hilbert in 1901 in his work on Fredholm operators.
Hölder's inequality for $p = 2$ gives $|\langle x,y\rangle| \leq \|x\|_2\|y\|_2$, which is the Cauchy-Schwarz inequality in this case.

(2) The 2-norm $\|.\|_2$ for the space of continuous functions on the interval $C[a,b]$ is inducded from the innerproduct

$$\langle f,g\rangle = \int_a^b f(x)g(x)dx.$$

The Cauchy-Schwarz inequality for $(C(\mathbb{R}), \langle .,.\rangle$ is due to Karl H. A. Schwarz in 1888.

The innerproduct $\langle .,.\rangle$ and its associated norm $\|.\| = \langle .,.\rangle^{1/2}$ are related by the *polarization identity.*

LEMMA 2.17 (Polarization identity). *Let $(X, \langle ., . \rangle)$ be an innerproduct space with norm $\|.\| = \langle ., . \rangle^{1/2}$. For a real innerproduct space we have $\langle x, y \rangle = \frac{1}{4}(\|x + y\|^2 - \|x - y\|^2)$ for all $x, y \in X$.*

PROOF. The arguments are based on the properties of innerproducts. $\|x + (-1)^k y\|^2 = \|x\|^2 + \|y\|^2 + (-1)^k \langle x, y \rangle$ for $k = 0, 1$. Adding these two identities yields the desired polarization identity. □

Jordan and von Neumann gave an elementary characterizations of norms that arise from innerproducts.

THEOREM 2.18 (Jordan-von Neumann). *Suppose $(X, \|.\|)$ is a complex normed space. If the norm satisfies the parallelogram identity*

$$\|x - y\|^2 + \|x + y\|^2 = 2\|x\|^2 + 2\|y\|^2 \quad forall\ x, y \in X,$$

then X is an innerproduct space for the innerproduct

$$\langle x, y \rangle = \frac{1}{4} \sum_{k=1}^{4} i^k \|x + i^k y\|^2.$$

PROOF. One direction is just a computation like the one done for the polarization identity. The reverse direction is based on defining an innerproduct in terms of the norms by turning the parallelogram identity into a definition and show that this is indeed an innerproduct. In the course of the argument one takes advantange of the paralellogram identity. □

Innerproduct spaces are the infinite-dimensional counterparts of $(\mathbb{R}^n, \|.\|_2)$ and share many properties with these finite-dimensional spaces, in contrast to general normed spaces such as $C(I)$ with the sup-norm or ℓ^p for $p \neq 2$.

EXAMPLE 2.3.6. The supremum norm of $C[0, 1]$ does not come from an innerproduct. Use the polarization identity to show this fact.

We consider the case of complex innerproduct spaces that are of relevance in quantum mechanics and signal analysis as well as mathematics.

For vectors in $\mathbb{C}^2$ we have the 'dot product' aka 'scalar product' that assigns to a pair of vectors $z = (z_1, z_2)$ and $z' = (z_1', z_2')$ the complex number

$$\langle z, z' \rangle = z_1 \overline{z_1}' + z_2 \overline{z_2}'.$$

The reason for adding the complex conjugates to the definition of the real case is to get the length of $z = (z_1, z_2) \in \mathbb{C}^2$:

$$\|z\|^2 = \langle z, z \rangle = z_1 \overline{z_1} + z_2 \overline{z_2} = |z_1|^2 + |z_2|^2.$$

DEFINITION 2.3.7. Let X be a complex vector space. An *innerproduct* on X is a map $\langle ., . \rangle : X \times X \to \mathbb{C}$ satisfying:

(1) (Linearity) For vectors $x_1, x_2, y \in X$ and scalars $\alpha_1, \alpha_2 \in \mathbb{F}$ we have $\langle \alpha_1 x_1 + \alpha_2 x_2, y \rangle = \alpha_1 \langle x_1, y \rangle + \alpha_2 \langle x_2, y \rangle$.
(2) (Conjugate Symmetry) For vectors $x, y \in X$ we have $\langle x, y \rangle = \overline{\langle y, x \rangle}$.
(3) (Positive definiteness) For any $x \in X$ we have $\langle x, x \rangle \geq 0$ and $\langle x, x \rangle = 0$ if and only if $x = 0$.

We call $(X, \langle ., . \rangle)$ an *innerproduct space* and define by $\|x\| := \langle x, x \rangle^{1/2}$.

Note that the conjugate symmetry and linearity in the first entry gives that $\langle .,. \rangle$ is conjugate linear in the second entry: For vectors $x, y_1, y_2 \in X$ and scalars $\alpha_1, \alpha_2 \in \mathbb{R}$ we have $\langle x, \alpha_1 y_1 + \alpha_2 y_2 \rangle\, y = \overline{\alpha_1}\, \langle x, y_1 \rangle + \overline{\alpha_2}\, \langle x, y_2 \rangle$.

PROPOSITION 2.3.8 (Cauchy-Schwarz). *Suppose X is a complex innerproduct space. Then for all $x, y \in X$ we have*

$$|\, \langle x, y \rangle\, | \leq \|x\| \|y\|.$$

We have $|\, \langle x, y \rangle\, | = \|x\| \|y\|$ if and only if $x = \alpha y$ for some $\alpha \in \mathbb{C}$.

PROOF. Suppose x and y are non-zero vectors of X.

$$\begin{aligned} 0 \leq \langle x - y, x - y \rangle &= \langle x, x \rangle + \langle y, y \rangle - \langle y, x \rangle - \langle x, y \rangle \\ &= \langle x, x \rangle + \langle y, y \rangle - 2\mathrm{Re}\, \langle x, y \rangle\, , \end{aligned}$$

and we obtain an additive inequality:

$$\mathrm{Re}\, \langle x, y \rangle \leq \frac{1}{2} \langle x, x \rangle + \frac{1}{2} \langle y, y \rangle\, .$$

The normalization method turns this one into a multiplicative one: We set $\tilde{x} = x/\langle x, x \rangle^{1/2}$ and $\tilde{y} = y/\langle y, y \rangle^{1/2}$ and plug $\tilde{x}$ and $\tilde{y}$ into the preceding inequality:

$$\mathrm{Re}\, \langle x, y \rangle \leq \langle x, x \rangle^{1/2} \langle y, y \rangle^{1/2}\, .$$

We want to have a bound on $|\, \langle x, y \rangle\, |$ based on the one on the real part of $\langle x, y \rangle$ via pre-multiplication. By the later one means that one pre-multiplies by a well-chosen complex number in order to guarantee that some quantity will be real. In our case we use the polar decomposition of $\langle x, y \rangle$: $\langle x, y \rangle = |\, \langle x, y \rangle\, | e^{i\varphi}$ for some $\varphi \in [0, 2\pi)$. We set $\tilde{x} := e^{-i\varphi} x$

$$|\, \langle x, y \rangle\, | = \mathrm{Re} \tilde{x} y \leq \langle \tilde{x}, \tilde{x} \rangle^{1/2} \langle y, y \rangle^{1/2} = \langle x, x \rangle^{1/2} \langle y, y \rangle^{1/2}\, ,$$

which yields the complex Cauchy-Schwarz inequality. The case of equality is a consequence of the argument. □

EXAMPLE 2.3.9.

(1) The space ℓ^2 of square-integrable complex-valued sequences $(z_i), (z_i')$ is an innerproduct space:

$$\langle z, z' \rangle = \sum_{i=1}^{\infty} z_i \overline{z_i'}.$$

Hölder's inequality for $p = 2$ gives $|\, \langle x, y \rangle\, | \leq \|x\|_2 \|y\|_2$, which is the Cauchy-Schwarz inequality in this case.

(2) The 2-norm $\|.\|_2$ for the space of continuous complex-valued functions on the interval $C[a, b]$ is induced from the innerproduct

$$\langle f, g \rangle = \int_a^b f(x) \overline{g(x)} dx.$$

This innerproduct is of utmost importance in Schrödinger's approach to quantum mechanics and in signal analysis. In physics one often denotes $\langle f, g \rangle$ by $\langle f | g \rangle$ and they tend to have it conjugate linear in the first entry and linear in the second.

By the same reasoning as for real innerproduct spaces X we deduce that $\|z\| := \langle z, z \rangle^{1/2}$ is a norm on X Innerproducts provide a generalization of the notion of *orthogonality* of elements.

DEFINITION 2.3.10. Two elements x, y in an innerproduct space $(X, \langle ., , \rangle)$ are *orthogonal* to each other if $\langle x, y\rangle = 0$

The theorem of Pythagoras is true for any innerproduct space $(X, \langle ., .\rangle)$.

PROPOSITION 2.3.11 (Pythagoras's Theorem). *Let $(X, \langle ., .\rangle)$ be an innerproduct space. For two orthogonal elements $x, y \in X$ we have*

$$\|x + y\|^2 = \|x\|^2 + \|y\|^2.$$

PROOF. The argument is based on the fact that $\langle x, x\rangle$ is a norm. By assumption we have $\langle x, y\rangle = 0$

$$\|x + y\|^2 = \|x\|^2 + 2\mathrm{Re}\,\langle x, y\rangle + \|y\|^2 = \|z\|^2 + \|y\|^2.$$

□

As an example we consider some orthogonal vectors in $(C([0, 1]), \langle ., .\rangle$. For $m \neq n$ we define the exponentials $e_m(x) = e^{2\pi imx}$ and $e_n(x) = e^{2\pi inx}$. Then

$$\langle e_m, e_n\rangle = \int_0^1 e^{2\pi i(m-n)x} dx = (2\pi i(m - n))^{-2}(e^{2\pi i(m-n)} - 1) = 0.$$

Note that $\langle e_n, e_n\rangle = 1$ for any $n \in \mathbb{Z}$. With the help of Kronecker's delta function we may express this as $\langle e_m, e_n\rangle = \delta_{m,n}$.

The theorem of Pythagoras is now at our disposal in any innerproduct spaces such as ℓ^2.

DEFINITION 2.3.12. A set of vectors $\{e_i\}_{i\in I}$ in an innerproduct space $(X, \langle ., , \rangle)$ is called an *orthogonal family* if $\langle e_i, e_j\rangle = 0$ for all $i \neq j$. In case that the orthogonal family $\{e_i\}_{i\in I}$ in X satisfies in addition $\|e_i\| = 1$ for any $i \in I$, then we refer to it as *orthonormal family*.

The exponentials $\{e^{2\pi nx}\}_{n\in\mathbb{Z}}$ is an orthonormal family in $C[0, 1]$ with respect to $\langle ., .\rangle$ and is a system of utmost importance, e.g. it lies at the heart of Fourier analysis or more generally harmonic analysis.

CHAPTER 3

Banach and Hilbert spaces

We extend the topological notions introduced for the real line to general normed spaces and we focus on completeness in this section. Complete normed spaces are nowadays called Banach spaces, after the numerous seminal contributions of the Polish mathematician Stefan Banach to these objects. The class of complete innerproduct spaces are named after David Hilbert, who introduced the sequence space ℓ^2. His students made numerous contributions to the theory of innerproduct spaces, e.g. Erhard Schmidt, Hermann Weyl, Otto Toeplitz,... .

3.1. Sequences in normed spaces

Norms on a vector space are the tool that provides us with a way to merge linear algebra and analysis, which is known as functional analysis. We will discuss some of the basic aspects of functional analysis in this course. We start with the notion of convergent sequences and will work our way up to completeness.

DEFINITION 3.1.1. Let $(X, \|.\|)$ be a normed space. A sequence $(x_n)_{n\in\mathbb{N}}$ in X is said to **converge to** $x \in X$ if for a given $\varepsilon > 0$ there exists a N such that $\|x - x_n\| < \varepsilon$ for $n \geq N$. The vector x is called the **limit** of the sequence $(x_n)_{n\in\mathbb{N}}$.

Suppose A is a subset of X. Given a convergent sequence $(a_n)_{n\in\mathbb{N}}$ in A, meaning all the a_n's are elements of A. Then the limit of the sequence $(a_n)_{n\in\mathbb{N}}$ is also known as a *limit point of* A. We denote the union of A and all its limit points by $\overline{A}$.

This notion of convergence for sequences in normed spaces is a natural generalization of the one for real and complex numbers. Note that the elements of the sequences are vectors in a normed space. For example, a sequence in ℓ^2 is a sequence where the elements themselves are also sequences. A more geometric view towards this notion of convergence is that for any $\varepsilon > 0$ there exists an N such that $(x_N, x_{N+1}, ...)$ lies in the ball, $B_\varepsilon(x)$, of radius ε around the limit x. Sometimes $(x_N, x_{N+1}, ...)$ is called the **tail** of the sequence $(x_n)_{n\in\mathbb{N}}$. Hence convergence of $x_n \to x$ means that for arbitrary small balls around the limit x the tail of $(x_n)_{n\in\mathbb{N}}$ lies in $B_\varepsilon(x)$.

Note that $x \in \overline{A}$ if there exists a sequence $(a_n)_{n\in\mathbb{N}}$ in A such that $a_n \to x$.

LEMMA 3.1. *Suppose the sequence* $(x_n)_{n\in\mathbb{N}}$ *in* $(X, \|.\|)$ *converges to a* $x \in X$. *Then*

$$\Big|\, \|x_n\| - \|x\| \,\Big| \to 0.$$

PROOF. By assumption we have that for any $\varepsilon > 0$ there exists an $N \in \mathbb{N}$ such that $\|x_n - x\| < \varepsilon$ for all $n \geq N$. By the reverse triangle inequality we have that

$$\Big| \|x_n\| - \|x\| \Big| \leq \|x_n - x\|$$

but the right hand side goes to zero by the convergence of (x_n) and thus we have that $\|x_n - x\| \to 0$. □

The notion of convergence depends on the norm the vector space is equipped with!

EXAMPLE 3.1.2. Consider the sequence $(f_n)_{n\in\mathbb{N}}$ in $C[0,1]$ defined by $f_n(t) = e^{-nt}$. Then we have that f_n converges to 0 in $(C[0,1], \|.\|_1)$:

$$\|f_n - 0\|_1 = \int_0^1 e^{-nt} dt = \frac{1}{n}(1 - e^{-n}) \to 0$$

as $n \to \infty$. Let us now discuss the convergence of $(f_n)_{n\in\mathbb{N}}$ in $(C[0,1], \|.\|_\infty)$. Since $\|f_m\|_\infty = \sup_{t\in[0,1]} |e^{-nt}| = 1$, so $(f_n)_{n\in\mathbb{N}}$ does not converge to the zero function with respect to $\|.\|_\infty$.

This example has a further feature.

EXAMPLE 3.1.3. Let A be the set of positive functions in $C[0,1]$, i.e. $A = \{f \in C[0,1] : f(t) > 0,\ t \in [0,1]\}$. Then the convergence of $(f_n)_{n\in\mathbb{N}}$ in $(C[0,1], \|.\|_1)$ of $(e^{-nt})_{n\in\mathbb{N}}$ to zero, gives us a sequence in A with a limit not contained in A; the zero function is the very example of a function attaining zero in $[0,1]$.

As for real sequences we have that limits of convergent sequences are unique.

LEMMA 3.2. *Let $(x_n)_{n\in\mathbb{N}}$ be a convergent sequence in the normed space $(X, \|.\|)$. Then its limit is unique.*

PROOF. Suppose there exist two limits x, y of $(x_n)_{n\in\mathbb{N}}$. Then for any $\varepsilon > 0$ there exist $N_1, N_2 \in \mathbb{N}$ such that for all $n \geq N_1$ $\|x_n - x\| \leq \varepsilon/2$ and for all $n \geq N - 2$ we have $\|x_n - y\| \leq \varepsilon/2$. Hence for all $n \geq \max N_1, N_2$ we have

$$\|x - y\| = \|x - x_n + x_n - y\| \leq \|x - x_n\| + \|x_n - y\| \leq \varepsilon/2 + \varepsilon/2 = \varepsilon.$$

□

A convergent sequence of real numbers is bounded, i.e. there exists a constant $M > 0$ such that $|a_n| \leq M$ for all $n \in \mathbb{N}$. Convergent sequences in normed spaces are also bounded if one defines the boundedness of a subset of this space in an analogous manner.

DEFINITION 3.1.4. A subset A of $(X, \|.\|)$ is called **bounded** if A is contained in some ball $B_{r_0}(x_0)$ for some radius r_0 and point $x_0 \in X$. In this case we define the **diameter** of A, $\operatorname{diam}(A)$, to be the real number $\sup\{\|x - y\| : x, y \in X\}$.

Let us state some reformulations of the notion of boundedness of a set.

LEMMA 3.3. *For a subset A of a normed space X the following statements are equivalent:*

(1) *A is bounded.*
(2) *There exists a constant $M > 0$ such that $\|x - y\| \leq M$ for all $x, y \in A$.*
(3) $\operatorname{diam}(A) < \infty$

(4) *For every $x \in X$ there exists a radius $r > 0$ such that $A \subseteq B_r(x)$.*
(5) *There exists a $m > 0$ such that $\|x\| \leq m$ for all $x \in A$.*

PROOF. We show $(i) \Rightarrow (ii) \Rightarrow (iii) \Rightarrow (iv) \Rightarrow (i)$, and finally $(v) \Rightarrow (i)$.

If (i) holds, then for some $x_0 \in X$ and $r_0 > 0$ we have $A \subset B_{r_0}(x_0)$:

$$\|x - y\| \leq \|x - x_0\| + \|x_0 - y\| \leq 2r_0 \text{ for all } x, y \in A,$$

i.e. $\|x - y\| \leq M = 2r_0$ for all $x, y \in A$.

If (ii) holds, then by the definition of supremum, as least upper bound, of the set $\{\|x - y\| : x, y \in A\}$ is less than or equal to the finite constant M, i.e. the diameter of A is finite.

If (iii) holds, then for all $x, y \in A$ we have $\|x - y\| \leq \operatorname{diam}(A) < \infty$. Choose an element $a_1 \in A$. Then given any $x \in X$ and $a \in A$ we have $\|x - a\| \leq \|x - x_1\| + \|x_1 - a\| \leq d(x, a_1) + \operatorname{diam}(A) =: r$ and $A \subseteq B_r(x)$. Hence we have shown that $(iii) \Rightarrow (iv)$.

The assertion $(iv) \Rightarrow (i)$ by definition of boundedness.

If (v) holds, then $A \subset B_m(0)$. Thus we have A is contained in a ball of radius m around the origin which is possible since in vector spaces we can translate its elements by a given vector such that the set gets centered at the origin. □

Further results about boundedness are posed as problems on the next problem set: (i) Any ball $B_r(x) \subset (X, \|.\|)$ is bounded and $\operatorname{diam}(B_r(x)) \leq 2r$. (ii) If A is a bounded subset, then for any $a \in A$ we have $A \subseteq B_{\operatorname{diam}(A)}(a)$.

LEMMA 3.4. *A convergent sequence in a normed space X is bounded.*

PROOF. See problem set. □

The definition of convergence of a sequence has one flaw: Namely one needs to have a candidate for the limit beforehand to actually set up the proof that the sequence converges to this particular object. Cauchy has noted that it is much more suitable to have a condition that only involves the sequence elements.

DEFINITION 3.1.5. Let $(x_n)_{n\in\mathbb{N}}$ be a sequence in $(X, \|.\|)$. Then we call $(x_n)_{n\in\mathbb{N}}$ a **Cauchy sequence** if for any $\varepsilon > 0$ there exists an $N \in \mathbb{N}$ such that for all $m, n \geq N$ we have

$$\|x_n - x_m\| < \varepsilon.$$

LEMMA 3.5. *Any Cauchy sequence in $(X, \|.\|)$ is bounded.*

PROOF. See problem set. □

LEMMA 3.6. *Every convergent sequence in $(X, \|.\|)$ is a Cauchy sequence.*

PROOF. Let $x_n \to x$ in $(X, \|.\|)$. Then for any $\varepsilon > 0$ there exists an $N \in \mathbb{N}$ such that $\|x_n - x\| < \varepsilon/2$ for all $n \geq N$. Hence for $m, n \geq N$ we have

$$\|x_n - x_m\| \leq \|x_n - x\| + \|x - x_m\| \leq \varepsilon.$$

□

EXAMPLE 3.1.6. We define a sequence in $(C[a,b], \|.\|_1)$ by a sequence of piecewise continuous functions f_n:

$$f_n(t) = \begin{cases} 0 & \text{for } a \le t \le \frac{a+b}{2}, \\ n(t - \frac{a+b}{2}) & \text{for } \frac{a+b}{2} < t \le \frac{a+b}{2} + \frac{1}{n}, \\ 1 & \text{for } \frac{a+b}{2} + \frac{1}{n} \le t \le b. \end{cases}$$

(f_n) is a Cauchy sequence in $(C[.a,b], \|.\|_1)$.
For $m > n$ the slope of f_m is greater than of f_n and thus the area of the function $f_m - f_n$ can be bounded by the triangle with sides 1 and $1/n$, i.e. $\|f_m - f_n\|_1 \le 1/2n$.

There are Cauchy sequences in $(C[a,b], \|.\|_1)$ that have no continuous limit function.

PROPOSITION 3.1.7. *$(C[a,b], \|.\|_1)$ is not complete.*

PROOF. The sequence (f_n) defined by

$$f_n(t) = \begin{cases} 0 & \text{for } a \le t \le \frac{a+b}{2}, \\ n(t - \frac{a+b}{2}) & \text{for } \frac{a+b}{2} < t \le \frac{a+b}{2} + \frac{1}{n}, \\ 1 & \text{for } \frac{a+b}{2} + \frac{1}{n} \le t \le b. \end{cases}$$

is Cauchy sequence in $(C[a,b], \|.\|_1)$ with discontinuous limit function:

$$\begin{cases} 0 & \text{for } a \le t \le \frac{a+b}{2}, \\ 1 & \text{for } \frac{a+b}{2} \le t \le b. \end{cases}$$

Suppose $f_n \to f$ in $\|.\|_1$ with $f \in C[a,b]$. Let us analyze the implications of $\|f_n - f\|_1 \to 0$ as $n \to \infty$.

$$\int_a^b |f_n(t) - f(t)|dt = \Big[\int_a^{\frac{a+b}{2}} + \int_{\frac{a+b}{2}}^{\frac{a+b}{2}+\frac{1}{n}} + \int_{\frac{a+b}{2}+\frac{1}{n}}^{b} \Big] |f_n(t) - f(t)|dt$$

breaks up into three integrals:

(1) $\int_a^{\frac{a+b}{2}} |f_n(t) - f(t)|dt \to 0$ only if $f = 0$ on $[a, \frac{a+b}{2}]$;

(2) $\int_{\frac{a+b}{2}}^{\frac{a+b}{2}+\frac{1}{n}} |f_n(t) - f(t)|dt \to 0$. Since f_n is continuous for all $n \in \mathbb{N}$ and f is continuous on $[a,b]$ we have

$$\int_{\frac{a+b}{2}}^{\frac{a+b}{2}+\frac{1}{n}} |f_n(t) - f(t)|dt \le (\max_{t\in[0,1]} |f(t)| + 1)\frac{1}{n} \to 0$$

as $n \to \infty$. Hence this imposes no condition on the limit function f.

(3) By the continuity of f we have that

$$\int_{\frac{a+b}{2}+\frac{1}{n}}^{b} |f_n(t) - f(t)|dt = \int_{\frac{a+b}{2}+\frac{1}{n}}^{b} |1 - f(t)|dt \to \int_{\frac{a+b}{2}}^{b} |1 - f(t)|dt,$$

as $n \to \infty$. Hence this limit is zero, we must have $1 - f(t) = 0$, i.e. $f(t) = 1$ for all $t \in [\frac{a+b}{2}, b]$.

In summary, the limit function f on $[a,b]$ has a jump discontinuity at $\frac{a+b}{2}$. □

3.2. Completeness

The difference between the the normed space $(\mathbb{Q}, |.|)$ and the real numbers $(\mathbb{R}, |.|)$ viewed as normed space is that not all Cauchy sequences in $\mathbb{Q}$ converge to a rational number but that is the case for $\mathbb{R}$. Cauchy established that any Cauchy sequence in $\mathbb{R}$ converges and its limit is again a real number. In order to show this we assume a property of the set of real numbers without proof, a so-called axiom. Namely, $\mathbb{R}$ is supposed to have the **least upper bound property: Any non-empty subset S that is bounded from above has a supremum** $\sup S$ **and** $\sup S$ **is a real number.**

For example the set $\{a \in \mathbb{Q} : a < \sqrt{3}\}$ is bounded above by $\sqrt{3}$, but $\sqrt{3}$ is not a rational number. We include the proof of this important fact.

PROPOSITION 3.2.1. *The equation*

$$x^2 - 3 = 0$$

has no solutions in $\mathbb{Q}$.

PROOF. We assume by contradiction that there is a rational number r such that $r^2 - 3 = 0$.

We represent r as a *reduced* fraction. That is, we write $r = \frac{p}{q}$ where p, q are integers, $q \neq 0$ and $\gcd(p, q) = 1$. We then have:

$$r^2 - 3 = 0 \implies r^2 = 3 \implies \frac{p^2}{q^2} = 3 \implies p^2 = 3\,q^2.$$

The last identity says that p^2 is a multiple of 3. Then p itself must be a multiple of 3 as well (why?), which means that $p = 3m$ for some integer m.

Substituting this into the identity $p^2 = 3q^2$ we get $9m^2 = 3q^2$, which implies $3m^2 = q^2$, and so q^2 must be a multiple of 3. But then q must also be a multiple of 3.

Let us step back and look at what we have: we started of with a completely reduced fraction $r = \frac{p}{q}$, assumed that $r^2 - 3 = 0$, which through a series of derivations led to the conclusion that both p and q must be multiples of 3. This contradicts the fraction $\frac{p}{q}$ being reduced.

Therefore, the equation $x^2 - 3 = 0$ cannot have any rational number as solution. □

THEOREM 3.7. *A sequence of real numbers* $(a_n)_{n\in\mathbb{N}}$ *converges if and only if for any* $\varepsilon > 0$ *there exists an index* N *such that for all* $m, n \in \mathbb{N}$ *we have* $|a_m - a_n| < \varepsilon$.

PROOF. The statement about convergent sequences satisfying the Cauchy property is one of the problems of problem set 5. The other implication is much more intricate. Suppose we have a Cauchy sequence $(a_n)_{n=1}^{\infty}$. Then we claim it converges to a real number. The argument is elementary but a little bit involved. Let A be the set of elements of our sequence (a_n), $A = \{a_1, a_2, ...\}$. Then A is a bounded subset of $\mathbb{R}$: there exists an $M > 0$ such that $a_n \in [-M, M]$ for $n = 1, 2, ...$. Take $\varepsilon = 1$ in the Cauchy condition: Then there exists an integer N_1 such that for all $m, n \geq N_1$ such that $|a_n - a_{N_1}| < 1$ and thus the set $\{a_1, a_2, ..., a_{N_1}, a_{N_1+1}\}$ is bounded by a constant M.

Now we consider the set

$$S := \{s \in [-M, M] : \text{there exist infinitely many } n \in \mathbb{N} \text{ for which } a_n \geq s\},$$

in other words we collect all the numbers s in $[-M, M]$ such that $a_n \geq s$ infinitely often. Definitely $-M \in S$ and S is bounded above by M. Thus by the least upper bound property of $\mathbb{R}$ there exists a real number a such that $a = \sup S$.

Claim: $a_n \to a$ as $n \to \infty$.
For any $\varepsilon > 0$ the Cauchy condition provides an N_2 s.t. for all $m, n \geq N_2$:

$$|a_m - a_n| < \varepsilon/2.$$

All elements of S are less than or equal to a, so the larger number $a + \varepsilon/2$ does not belong to S, and hence only finitely many often does a_n exceed $a + \varepsilon/2$. That is for some $N_3 \geq N_2$ we have for all $n \geq N_3$ that

$$a_n \geq a + \varepsilon/2.$$

Since a is a least upper bound for S, the smaller number $a - \varepsilon/2$ cannot also be an upper bound for S. Hence, there is some $s \in S$ such that $s \geq a - \varepsilon/2$. Consequently, we have infinitely many sequence elements such that

$$a - \varepsilon/2 < s \leq a_n.$$

In particular, there exists an $N \geq N_3$ such that

$$a_N > a - \varepsilon/2$$

. Since $N \geq N_3$ we have $a_N \leq a + \varepsilon/2$ and so $a_N \in (a - \varepsilon/2, a + \varepsilon/2)$. Now recall that $N \geq N_2$ which yields that

$$|a_n - a| \leq |a_n - a_N| + |a_N - a| < \varepsilon$$

for all $n \geq N$, i.e. $a_n \to a$ as $n \to \infty$.

□

The property of $\mathbb{R}$ that any Cauchy sequence converges in $\mathbb{R}$ is a favorable property that we would like to have for general normed spaces.

Definition 3.2.2. A normed space $(X, \|.\|)$ is called *complete* if every Cauchy sequence (x_k) in X has a limit x belonging to X. Moreover, a complete normed space is referred to as *Banach space* and a complete innerproduct space is known as *Hilbert space*.

Let us start with an elementary observation that is a straightforward consequence of the definitions.

Theorem 3.8. $(\mathbb{R}^n, \|.\|_\infty)$ *is a Banach space.*

The completeness of the normed space $(\mathbb{R}, |.|)$ has numerous ramifications.

Proof. The $\|.\|_\infty$-convergence of $(x_n)_{n\in\mathbb{N}}$ implies the coordinate wise convergence. Since any Cauchy sequence in $(\mathbb{R}^n, \|.\|_\infty)$ gives Cauchy sequences in each coordinate. Since $\mathbb{R}$ is complete we deduce that all these coordinate Cauchy sequences converge in $\mathbb{R}$. Thus we have that $(\mathbb{R}^n, \|.\|_\infty)$ is complete. □

Theorem 3.9. *The space of absolutely summable sequences is a Banach space with respect to* $\|.\|_1$*-norm; i.e.* $(\ell^1, \|.\|_1)$ *is a Banach space.*

PROOF. The argument is split into three steps.
Step 1: Find a candidate for the limit. Let $(x_n)_n$ be a Cauchy sequence in ℓ^1. We denote the n-th element of the sequence by $x_n = (x_1^{(n)}, x_2^{(n)}, ...)$.
Note that $|x_1^{(m)} - x_1^{(n)}| \leq \|x_m - x_n\|_1$, so the first coordinates $(x_1^{(n)})_n$ are a Cauchy sequence of real numbers and hence converge to some real number z_1. Similarly, the other coordinates converge: $z_j = \lim_{n\to\infty} x_j^{(n)}$. Hence our candidate for the limit of (x_n) is the sequence $z = (z_1, z_2, ...)$.
Step 2: Show that z is in ℓ^1. We have that

$$\sum_{j=1}^{N} |z_j| = \sum_{j=1}^{N} \lim_n |x_j^{(n)}| = \lim_n \sum_{j=1}^{N} |x_j^{(n)}|,$$

where the interchange of the limit with the sum of a finite number of real numbers is no problem. Since Cauchy sequences are bounded, there is a constant $C > 0$ such that $\|x_n\|_1 < C$ for all n. Thus for any N

$$\sum_{j=1}^{N} |x_j^{(n)}| \leq \sum_{j=1}^{\infty} |x_j^{(n)}| = \|x_n\|_1 < C.$$

Letting $n \to \infty$ we find that

$$\sum_{j=1}^{N} |z_j| \leq \|x_n\|_1 < C$$

for arbitrary N. Hence we have $z \in \ell^1$.
Step 3: Show the convergence. We want to prove that $\|x_n - z\|_1 \to 0$ for $n \to \infty$. Given $\varepsilon > 0$, pick N_1 so that if $m, n > N_1$ then $\|x_m - x_n\|_1 < \varepsilon$. Hence for any fixed N and $m, n > N_1$, we find

$$\sum_{j=1}^{N} |x_j^{(m)} - x_j^{(n)}| \leq \sum_{j=1}^{\infty} |x_j^{(m)} - x_j^{(n)}| = \|x_n - x_m\| < \varepsilon.$$

Fix $n > N_1$ and N, let $m \to \infty$ to obtain

$$\sum_{j=1}^{N} |x_j^{(n)} - z_j| = \lim_{n\to\infty} |x_j^{(n)} - x_j^{(m)}| \leq \varepsilon.$$

Since this is true for all N we have demonstrated that

$$\|x_n - z\|_1 < \varepsilon.$$

That is our desired conclusion. □

THEOREM 3.10. *The space of bounded sequences is a Banach space with respect to $\|.\|_\infty$-norm; i.e. $(\ell^\infty, \|.\|_\infty)$ is a Banach space.*

PROOF. The argument is once more split into three steps.
Step 1: Find a candidate for the limit. Let $(x_n)_n$ be a Cauchy sequence in ℓ^∞. We denote the n-th element of the sequence by $x_n = (x_1^{(n)}, x_2^{(n)}, ...)$.
Note that $|x_k^{(m)} - x_k^{(n)}| \leq \|x_m - x_n\|_\infty$ for all k and all $m, n > N$, so the k-th coordinates $(x_k^{(n)})_n$ are a Cauchy sequence of real numbers and hence converge to some real number z_k. Similarly, the other coordinates converge: $z_k = \lim_{m\to\infty} x_k^{(n)}$.

Hence our candidate for the limit of (x_n) is the sequence $z = (z_1, z_2, ...)$.
Step 2: Show that z is in ℓ^∞. We have that

$$\sup\{|z_j| : j = 1, ..., N\} = \sup\{\lim_n |x_j^{(n)}| j = 1, ..., N\} = \lim_n \{\sup |x_j^{(n)}| j = 1, ..., N\},$$

where the interchange of the limit with the sum of a finite number of real numbers is no problem. Since Cauchy sequences are bounded, there is a constant $C > 0$ such that $\|x_n\|_\infty < C$ for all n. Thus for any N

$$\lim_n \{\sup |x_j^{(n)}| j = 1, ..., N\}| \leq \|x_n\|_\infty < C.$$

Thus we find that $\|x_n\|_\infty < C$, i.e. we have $z \in \ell^\infty$.
Step 3: Show the convergence. We want to prove that $\|x_n - z\|_\infty \to 0$ for $n \to \infty$. Given $\varepsilon > 0$, pick N_1 so that if $m, n > N_1$ then

$$|x_m^{(k)} - x_n^{(k)}| \leq \|z_k - x_n^{(k)}\|_\infty < \varepsilon$$

for all k. Taking limits as $m \to \infty$ we have

$$|z_k - x_n^{(k)}| \leq \varepsilon$$

Taking supremum in k, we obtain

$$\sup_k |z_k - x_n^{(k)}| \leq \varepsilon$$

for all $n > N_1$, i.e. $\|x_n - z\|_\infty \leq \varepsilon$ for all $n > N$. Consequently we have that x_n converges to z in $(\ell^\infty, \|.\|_\infty)$.

□

Reasoning similar to the one for ℓ^1 gives us that all ℓ^p-spaces are Banach spaces for $\|.\|_p$ when $1 \leq p < \infty$.

THEOREM 3.11. *Let $[a, b]$ be a bounded interval of real numbers. Then the normed space $C[a, b]$ with respect to the sup-norm $\|.\|_\infty$ is a Banach space.*

The situation is different for the function spaces $(C[a, b], \|.\|_p)$, as we have seen before for $p = 1$ this is not a complete space and this is also true for $1 \leq p < \infty$. In contrast $(C[a, b], \|.\|_\infty)$ is a complete space. Before we are able to proof this statement we have to discuss different notions of convergence for sequences of functions and properties of continuous functions.

LEMMA 3.12. *For $f, g \in C[a, b]$ we have that $\sup\{|f(x) - g(x)| x \in [a, b]\}$ is finite, and there is a $y \in [a, b]$ such that $d_\infty(f, g) = |f(y) - g(y)| = \max\{|f(x) - g(x)| x \in [a, b]\}$.*

PROOF. We show that $d(x) = |f(x) - g(x)|$ is continuous on $[a, b]$ and thus by the Extreme Value Theorem the assertion follows. The continuity of d is deduced from

$$|d(x) - d(y)| \leq ||f(x) - g(x)| - |f(y) - g(y)|| \leq |f(x) - f(y)| + |g(y) - g(x)|.$$

Since f and g are continuous at x there is for any given $\varepsilon > 0$ a $\delta > 0$ such that $|f(x) - f(y)| < \varepsilon/2$ and $|g(x) - g(y)| < \varepsilon/2$ for $|x - y| < \delta$. Hence

$$|d(x) - d(y)| \leq |f(x) - f(y)| + |g(y) - g(x)| < \varepsilon/2 + \varepsilon/2 = \varepsilon$$

for all $y \in [a, b]$ with $|x - y| < \delta$. Consequently d is continuous. □

REMARK 3.2.3. Observe that the $\|f-g\|_\infty$-norm measures the distance between the functions f and g by looking at the point in $a[a,b]$ they are the furthest apart.

DEFINITION 3.2.4. Let (f_n) be a sequence of functions on a set X.

- We say that (f_n) *converges pointwise* to a limit function f if for a given $\varepsilon > 0$ and $x \in X$ there exists an N so that
$$|f_n(x) - f(x)| < \varepsilon \quad \text{for all } n \geq N.$$
- We say that (f_n) *converges uniformly* to a limit function f if for a given $\varepsilon > 0$ there exists an N so that
$$|f_n(x) - f(x)| < \varepsilon \quad \text{for all } n \geq N$$
holds for all $x \in X$.

There is a substantial difference between these two definitions. In pointwise convergence, one might have to choose a different N for each point $x \in X$. In the case of uniform convergence there is an N that holds for all $x \in X$. Note that uniform convergence implies pointwise convergence. If one draws the graphs of a uniformly convergent sequence, then one realizes that the definition amounts for a given $\varepsilon > 0$ to have a N so that the graphs of all the f_n for $n \geq N$, lie in an ε-band about the graph of f. In other words, the f_n's get uniformly close to f. Hence uniform convergence means that the maximal distance between f and f_n goes to zero. We prove this assertion in the next proposition.

PROPOSITION 3.2.5. *Let (f_n) be a sequence of continuous functions on $[a,b]$. Then the following are equivalent:*

(1) *(f_n) converges uniformly to f.*
(2) *$\|f_n - f\|_\infty = \sup\{|f_n(x) - f(x)| : x \in [a,b]\} \to 0$ as $n \to \infty$.*

PROOF. Assertion $(i) \Rightarrow (ii)$: Assume that (f_n) converges uniformly to f. Then for any $\varepsilon > 0$ there exists a N such that $|f_n(x) - f(x)| < \varepsilon$ for all $x \in [a,b]$ and all $n > N$. Hence $\sup\{|f_n(x) - f(x)| : x \in [a,b]\} \leq \varepsilon$ for all $n > N$. Since this holds for all $\varepsilon > 0$, we have demonstrated that $\sup\{|f_n(x) - f(x)| : x \in [a,b]\} \to 0$ for $n \to \infty$.
Assertion $(ii) \Rightarrow (i)$: Assume that $\sup\{|f_n(x) - f(x)| : x \in [a,b]\} \to 0$ for $n \to \infty$. Given an $\varepsilon > 0$, there is a N such that $\sup\{|f_n(x) - f(x)| : x \in [a,b]\} < \varepsilon$ for all $n > N$. Thus we have $|f_n(x) - f(x)| < \varepsilon$ for all $x \in [a,b]$ and all $n > N$, i.e. (f_n) converges uniformly to f. □

A reformulation of this result is that a sequence converges in $(C[a,b], \|.\|_\infty)$ to f is equivalent to the uniform convergence of (f_n) to f.

PROPOSITION 3.2.6. *A sequence (f_n) converges to f in in $(C[a,b], \|.\|_\infty)$ if and only if (f_n) converges uniformly to f.*

Uniform convergence has an important property.

THEOREM 3.13. *Let (f_n) be a uniformly convergent sequence in $C[a,b]$ with limit f. Then the limit function f is continuous on $[a,b]$.*

PROOF. Let $y \in I$ and $\varepsilon > 0$ be given. By the uniform convergence of $f_n \to f$, there exists an N such that $n \geq N$ implies that
$$|f_n(x) - f(x)| \leq \varepsilon/3 \quad \text{for all } x \in I.$$

The continuity of f_N implies that there exists a $\delta > 0$ such that

$$|f_N(x) - f(y)| \leq \varepsilon/3 \quad \text{for } |x - y| \leq \delta.$$

We want to show that f is continuous. For all x such that $|x - y| < \delta$ we have that

$$\begin{aligned} |f(x) - f(y)| &\leq |f(x) - f_N(x)| + |f_N(x) - f_N(y)| + |f_N(y) - f(y)| \\ &< \varepsilon/3 + \varepsilon/3 + \varepsilon/3 = \varepsilon. \end{aligned}$$

□

THEOREM 3.14. *$(C[a,b], \|.\|_\infty))$ is a Banach space.*

PROOF. Convergence of a sequence in $(C[a,b], \|.\|_\infty)$ to $f \in C[a,b]$ is equivalent to uniform convergence of the sequence to f.

Assume that (f_n) is a Cauchy sequence in $(C[a,b], \|.\|_\infty)$. Then we have to show that there exists a function $f \in C[a,b]$ that has (f_n) as its limit.
Fix $x \in [a,b]$ and note that $|f_n(x) - f_m(x)| \leq \|f_n - f_m\|_\infty$. Since (f_n) is a Cauchy sequence $(f_n(x))$ is a Cauchy sequence in $\mathbb{R}$. Since $\mathbb{R}$ is complete, $(f_n(x))$ converges to a point $f(x)$ in $\mathbb{R}$. In other words, $f_n \to f$ pointwise.
Next we show that $f \in C[a,b]$. Since (f_n) is a Cauchy sequence, we have for any $\varepsilon > 0$ a N such that $\|f_n - f_m\| < \varepsilon/2$ for all $m, n > N$. Hence we have $|f_n(x) - f_m(x)| < \varepsilon/2$ for all $x \in [a,b]$ and for all $m, n > N$. Letting $m \to \infty$ yields for all $x \in [a,b]$ and all $n > N$:

$$|f_n(x) - f(x)| = \lim_{m \to \infty} |f_n(x) - f_m(x)| \leq \varepsilon/2 < \varepsilon.$$

Consequently, $f_n \to f$ converges uniformly. Now by the preceding proposition f is a continuous function on $[a,b]$. In other words, we have established that $(C[a,b], \|.\|_\infty)$ is a Banach space. □

3.3. Banach's Fixed Point Theorem

In 1922 Banach established a theorem on the convergence of iterations of contractions that has become a powerful tool in applied and pure mathematics aka Contraction Mapping Theorem. Before we state Banach's fixed point theorem we define continuous functions between normed spaces. A natural and far-reaching generalization of the notion of continuous functions defined on $\mathbb{R}$.We will have much more to say about continuous functions in the next chapter.

DEFINITION 3.3.1. Let $(X, \|\cdot\|_X)$ and $(Y, \|\cdot\|_Y)$ be two normed spaces, let $A \subset X$ and let $f\colon A \to Y$ be a function.

(1) We say that f is **continuous** at a point $a \in A$ if for all $\varepsilon > 0$ there is $\delta > 0$ such that for all $x \in A$ with $\|x - a\|_X < \delta$ we have $\|f(x) - f(a)\|_Y < \varepsilon$.
(2) We say that f is **continuous** on A if it is continuous at each point of A.
(3) We say that f is **uniformly continuous** on A if, for all $\varepsilon > 0$, there exists a $\delta > 0$ such that $\|x - y\|_X < \delta$ implies $\|f(x) - f(a)\|_Y < \varepsilon$ for all $x, y \in A$.

A class of continuous functions on normed spaces is given by functions satisfying: There exists a finite constant L such that

$$\|f(x) - f(x')\| \leq L \|x - x'\| \quad \text{for all } x, x' \in A.$$

One calls such functions **Lipschitz continuous**, after the German mathematician R. Lipschitz, and often one refers to L as **Lipschitz constant.** On Problem set 6 you will show that any Lipschitz continuous function is continuous.

We have come across Lipschitz continuous functions in our discussion of normed spaces. Namely, the reverse triangle inequality shows that a norm $\|,\| : X \to \mathbb{R}$ on a vector space X is Lipschitz continuous with constant 1.

Here is a useful criterion for continuity of a function.

PROPOSITION 3.3.2. *Let $f\colon A \to Y$ be a function, where $A \subset X$ and X, Y are normed spaces. Let $a \in A$. Then the following two statements are equivalent.*

(i) f is continuous at a.

(ii) For every sequence $(x_n) \subset A$, if $x_n \to a$ then $f(x_n) \to f(a)$.

PROOF. i) $\Rightarrow$ (ii): We assume that f is continuous at a.

Let $(x_n) \subset A$ be a sequence such that $x_n \to a$. We prove that $f(x_n) \to f(a)$.

Let $\varepsilon > 0$. Since f is continuous at a, there is $\delta > 0$ such that if $\|x - a\| < \delta$ then $\|f(x) - f(a)\| < \varepsilon$.

Since $x_n \to a$, there is $N \in \mathbb{N}$ such that for all $n \geq N$ we have $\|x_n - a\| < \delta$. From the above, if $n \geq N$ we must then have $\|f(x_n) - f(a)\| < \varepsilon$.

As ε was arbitrary, this proves that $f(x_n) \to f(a)$.

(i) $\Leftarrow$ (ii): We assume by contradiction that f is *not* continuous at a. Let us write down carefully what that means.

Firstly, we recall the definition of continuity. f is continuous at the point $a \in A$ means:
for all $\varepsilon > 0$ there is $\delta > 0$ such that for all $x \in A$ with $\|x - a\| < \delta$ we have $\|f(x) - f(a)\| < \varepsilon$.

Next, we formulate the *negation* of this statement.

The function f is *not* continuous the point $a \in A$ means:
there is $\varepsilon_0 > 0$ such that for all $\delta > 0$ there is an element of A, which we denote by x_δ, such that $\|x_\delta - a\| < \delta$ but $\|f(x_\delta) - f(a)\| \geq \varepsilon_0$.

For every $n \geq 1$, we may choose $\delta = \frac{1}{n}$. Then for some element of A, which we denote by x_n, we have that $\|x_n - a\| < \frac{1}{n}$ but $\|f(x_n) - f(a)\| \geq \varepsilon_0$.

We have thus obtained a sequence $(x_n) \subset A$ such that $\|x_n - a\| < \frac{1}{n} \to 0$, so $x_n \to a$. However, since $\|f(x_n) - f(a)\| \geq \varepsilon_0$, the sequence $f(x_n) \not\to f(a)$, which is a contradiction.

Hence f must be continuous at a. □

Suppose we have a continuous function f on a normed space X. Take a point x_0 in X and build the sequence of iterates

$$x_0,\ x_1 = f(x_0),\ x_2 = f(x_1) = f^2(x_0),\ ...,\ x_{n+1} = f(x_n).$$

The existence of the limit of this sequence $x = \lim_n x_n = \lim_n f^n(x_0)$ is the basic question that underlies Banach's fixed point theorem. The limit x of the iterates (x_n) is a fixed point of the continuous map f:

$$f(x) = f(\lim_n x_n) = \lim_n f(x_n) = \lim_n x_{n+1} = \lim x_n = x.$$

A mapping f on a normed space X is called a **contraction** if there exists a $0 < K < 1$ such that

$$\|f(x) - f(y)\| \leq K\|x - y\| \quad x, y \in X,$$

a contraction is a Lipschitz continuous function with Lipschitz constant $L < 1$. Recall that $\|x - y\| = d(x, y)$ is the distance between x and y.

THEOREM 3.15 (Banach's Fixed Point). *Let X be a Banach space X. Any contraction $f : X \to X$ has a unique fixed point $\tilde{x}$ and the fixed point is the limit of every sequence generated from an arbitrary nonzero point $x_0 \in X$ by iteration $(x_n)_n$, where $x_{n+1} = f(x_n)$ for $n \geq 1$.*

PROOF. Let $x_0 \in X$ be arbitrary. Define $x_{n+1} = f(x_n)$ for $n = 1, 2, \ldots$. By the contractivity of T we have

$$\|x_n - x_{n-1}\| = \|f(x_{n-1}) - f(x_{n-2})\| \leq K\|x_{n-1} - x_{n-2}\|$$

and iterations yields

$$\|x_n - x_{n-1}\| \leq K^{n-1}\|x_1 - x_0\|.$$

The existence of a fixed point is based on the completeness of X. Hence we proceed to show that $(x_n)_n$ is a Cauchy sequence. Let m, n be greater than N and we choose $m \geq n$. Then by the preceding inequality and the triangle inequality we have

$$\begin{aligned}
\|x_m - x_n\| &\leq \|x_m - x_{m-1}\| + \|x_{m-1} - x_{m-2}\| + \cdots + \|x_{n+1} - x_n\| \\
&\leq (K^{m-1} + K^{m-2} + \cdots K^n)\|x_1 - x_0\| \\
&\leq (K^N + K^{N+1} + \cdots)\|x_1 - x_0\| \\
&= K^N(1 - K)^{-1}\|x_1 - x_0\|.
\end{aligned}$$

Since $0 \leq K < 1$, $\lim_N K^N = 0$ and thus (x_n) is a Cauchy sequence. Consequently, (x_n) converges to a point $\tilde{x}$ by the completeness of X. Furthermore $\tilde{x}$ is a fixed point by the contractivity of T.
Uniqueness: Suppose there is another fixed point $\tilde{y}$ of f. Then $\|\tilde{x} - \tilde{y}\| = \|f(\tilde{x}) - f(\tilde{y})\| \leq K\|\tilde{x} - \tilde{y}\|$ and $\|\tilde{x} - \tilde{y}\| > 0$. Thus we deduce that $K \geq 1$ which is a contradiction to f being a contraction. □

Lipschitz maps with constant 1 are not eligible in this fixed point theorem. Since the map $f(x) = x + 1$ on $[0, 1]$ has no fixed point, but the map $f(x) = x$ on $[0, 1]$ has infinitely many fixed points.

COROLLARY 3.3.3. *Under the assumption in Banach's fixed point theorem we have the following estimates about the rate of convergence of the iterates (x_n) towards the fixed point $\tilde{x}$:*

(1)

$$\|x_n - \tilde{x}\| \leq \frac{K^n}{1 - K}\|x_0 - f(x_0)\|,$$

tells us, in terms of the distance between x_0 and $f(x_0)$ how many times we need to iterate f starting from x_0 to be certain that we are within a specified distance from the fixed point.

(2)
$$\|x_n - \tilde{x}\| \leq K\|x_{n-1} - \tilde{x}\|,$$
is called an a priori estimate, meaning that it gives us an upper bound on how long we need to compute to reach the fixed point.

(3)
$$\|x_n - \tilde{x}\| \leq \frac{K}{1-K}\|x_{n-1} - x_n\|,$$
tells us, after each computation, how much closer we are to the fixed point in terms of the previous two iterations. This kind of estimate, called an a posteriori estimate, is very important because if two successive iterations are nearly equal, guarantees that we are very close to the fixed point.

PROOF. From the proof we have that for $m > n$

$$\|x_m - x_n\| \leq \frac{K^n}{1-K}\|x_0 - x_1\| = \frac{K^n}{1-K}\|x_0 - f(x_0)\|. \tag{3.1}$$

The right side is *independent* of m and so $m \to \infty$ gives

$$\|x_n - \tilde{x}\| \leq \frac{K^n}{1-K}\|x_0 - f(x_0)\|. \tag{3.2}$$

The second inequality comes along like that: Since $\tilde{x}$ is the unique fixed point of f:

$$\|x_n - \tilde{x}\| = \|f(x_n) - f(\tilde{x})\| \leq K\|x_{n-1} - \tilde{x}\|.$$

Applying the triangle inequality to $\|x_{n-1} - \tilde{x}\|$ gives the third inequality:

$$\|x_n - \tilde{x}\| \leq K(\|x_{n-1} - x_n\| + \|x_n - \tilde{x}\|),$$

which gives

$$\|x_n - \tilde{x}\| \leq \frac{K}{1-K}\|x_{n-1} - x_n\|. \tag{3.3}$$

□

Recall that we defined for a the closure $\overline{A}$ of A as the union of A and the set of limit points of A.

DEFINITION 3.3.4. A subset A of $(X, \|.\|)$ is called **closed** if $\overline{A} = A$.

For example $\{y \in X : \|x - y\| \leq r\}$ is a closed subset of X. We will discuss properties of closed sets in the next chapter.

Here is a variant of Banach's fixed point theorem:

THEOREM 3.16. *Let A be a closed subset of a Banach space X. If $f : A \to X$ is a contraction, then f has a unique fixed point and the fixed point is the limit of every sequence generated from an arbitrary nonzero point $x_0 \in A$ by iteration $(x_n)_n$, where $x_{n+1} = f(x_n)$ for $n \geq 1$.*
If the contraction $f : A \to X$ satisfies in addition, $f(A) \subseteq A$, then the fixed point lies also in A.

PROOF. See problem set. □

Two well-known applications are Newton's method for finding roots of general equations, solving systems of linear equations and the theorem of Picard-Lindelöf on the existence of solutions of ordinary differential equations. We discuss the first item and postpone the other items.

Newton's method:

How does one compute $\sqrt{3}$ up to a certain precision, i.e. we are interested in error estimates? Idea: Formulate it in the form $x^2 - 3 = 0$ and try to use a method that allows to compute zeros of general equations.

Newton came up with a method to solve $g(x) = 0$ for a differentiable function $g : I \to \mathbb{R}$.
Suppose x_0 is an approximate solution or starting point. Define recursively

$$x_{n+1} = x_n - \frac{g(x_n)}{g'(x_n)} \quad \text{for } n \geq 0.$$

Then (x_n) converges to a solution $\tilde{x}$, provided certain assumptions on g hold.
If $x_n \to \tilde{x}$, then by continuity of g we get $g(\tilde{x}) = 0$.

When does Netwon's method lead to a convergent sequence of iterates? Idea: Apply Banach's Fixed Point Theorem.
Set $f(x) := x - \frac{g(x)}{g'(x)}$. Then given $x_0 \in I$ and $x_{n+1} = x_n - \frac{g(x_n)}{g'(x_n)} = f(x_n)$. Moreover, $f(\tilde{x}) = \tilde{x}$ if and only if $g(\tilde{x}) = 0$.

Let us restrict our discussion to the computation of $\sqrt{3}$. The Banach space X is the space of real numbers $\mathbb{R}$ and $g(x) = x^2 - 3$, so

$$f(x) = x - \frac{x^2 - 3}{2x} = \frac{1}{2}(x + \frac{3}{x})$$

on $[\sqrt{3}, \infty) \to [\sqrt{3}, \infty)$. Note that $[\sqrt{3}, \infty)$ is a closed set of $\mathbb{R}$ containing $\sqrt{3}$. For $x \geq 0$ we have $\frac{1}{2}(x + 3/x) \geq \sqrt{3x/x} = \sqrt{3}$. Compute f' and note that a differentiable function $f : I \to \mathbb{R}$ with a bounded derivative, $|f'(x)| \leq L$ for $x \in I$ is Lipschitz continuous with constant L.

$$f'(x) = \frac{1}{2}(1 - \frac{3}{x^2})$$

and note that it's range is contained in $[0, 1/2]$ for $x \geq \sqrt{3}$. Hence we have $L = 1/2$ and by Banach's Fixed Point Theorem $\frac{1}{2}(x_n + \frac{3}{x_n}) \to \sqrt{3}$.
Let's pick $x_0 = 2$ and thus $x_1 = 7/4$ and so $|x_1 - x_0| = 1/4$. Furthermore, we have

$$|x_n - \sqrt{3}| \leq \frac{(1/2)^n}{1 - 1/2}|x_1 - x_0| = \frac{1}{2^n} \cdot 2 \cdot \frac{1}{4} = \frac{1}{2^{n+1}}.$$

Hence

$$|x_n - \sqrt{3}| \leq \frac{1}{2^{n+1}}.$$

For $n = 4$, we have $|x_n - \sqrt{3}| \leq 1/1024 < 0.001$.

Integral equations

Equations of the following type appear naturally in mathematics, physics and engineering: Given functions $f : [a,b] \to \mathbb{R}$ and $k : [a,b] \times [a,b] \to \mathbb{R}$, a parameter λ, where $[a,b]$ denotes a finite interval of $\mathbb{R}$. Solve the **integral equation**

$$f(x) = \lambda \int_a^b k(x,y) f(y) dy + g(x)$$

for g. We will restrict our discussion to continuous functions f and k. Note that the mapping

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

is a continuous analogue of matrix multiplication, where the function k on the rectangle $[a,b] \times [a,b]$ is the continuous variant of a matrix (a_{ij}) and one often calls T an **integral operator** and k its **kernel**. The fixed point theorem of Banach allows us to solve this integral equation for sufficiently small λ.

Note that $T : C[a,b] \to C[a,b]$ respects the vector space structure of $C[a,b]$: For any $\alpha, \beta \in \mathbb{R}$ and $f_1, f_2 \in C[a,b]$ we have

$$T(\alpha f_1 + \beta f_2) = \alpha T(f_1) + \beta T(f_2).$$

LEMMA 3.17. *Let $f \in C[a,b]$ and $k \in C\Big([a,b] \times [a,b]\Big)$. Then $Tf \in C[a,b]$.*

PROOF. For each fixed x the function $K(x,y)$ is a continuous function of y on $[a,b]$. Hence $K(x,y)f(y)$ is a continuous function of y and so the integral in the definition of T makes sense.
Claim: For $f \in C[a,b]$ we also have $Tf \in C[a,b]$.

As a preparation we look at $|T(f)(x_1) - T(f)(x_2)|$ for $x_1 \neq x_2$:

$$\begin{aligned} |T(f)(x_1) - T(f)(x_2)| &\leq \Big| \int_a^b \Big(k(x_1,y) - k(x_2,y) \Big) f(y) dy \Big| \\ &\leq \int_a^b |k(x_1,y) - k(x_2,y)| \, |f(y)| dy. \end{aligned}$$

Since k is continuous on $[a,b] \times [a,b]$, we have that k is bounded on $[a,b] \times [a,b]$: $\|k\|_\infty \leq \|k\|_\infty$. We also have more control over k as one would have for a continuous function. Namely, it is uniformly continuous on $[a,b] \times [a,b]$: For any $\delta > 0$ so that $|x_1 - x_2| < \delta$ we have

$$|k(x_1,y) - k(x_2,y)| \leq \varepsilon / \|f\|_\infty (b-a) \quad \text{for all } y \in [a,b].$$

Using this estimate we obtain that for $|x_1 - x_2| < \delta$

$$|T(f)(x_1) - T(f)(x_2)| \leq \varepsilon \quad \text{for all } y \in [a,b].$$

Hence Tf is continuous on $[a,b]$. □

Furthermore T is also compatible with the norm structure on $C[a,b]$, which follows from the estimates in the preceding proof:

$$\|T(f_1) - T(f_2)\|_\infty \leq \|k\|_\infty (b-a) \|f_1 - f_2\|_\infty.$$

Hence we are in the position to specify when $T_\lambda f(x) = g(x) + \lambda \int_a^b \mathrm{k(x,y)f(y)dy}$ is a contraction on $C[a,b]$: Namely, when $|\lambda| < \frac{1}{\|k\|_\infty (b-a)}$.

PROPOSITION 3.3.5. *Suppose $g \in C[a,b]$ and $k \in C\Big([a,b] \times [a,b]\Big)$. Then*

$$f(x) = \lambda \int_a^b k(x,y) f(y) dy + g(x)$$

has a unique continuous solution $\tilde{f}$ on $[a,b]$ for $|\lambda| < \frac{1}{\|k\|_\infty (b-a)}$. The solution can be found by iteration.

PROOF. Consider the mapping $f(x) \mapsto T_\lambda f(x) := g(x) + \lambda \int_a^b k(x,y) f(y) dy$. For $f_1, f_2 \in C[a,b]$ we have

$$|T_\lambda f_1(x) - T_\lambda f_2(x)| = |g(x) - g(x)| + |\lambda| \int_a^b |k(x,y)||f_1(y) - f_2(y)| dy$$
$$\leq |\lambda| \int_a^b |k(x,y)||f_1(y) - f_2(y)| dy.$$

Since k is bounded on $[a,b] \times [a,b]$ we have $|k(x,y)| \leq \|k\|_\infty$ for all $x, y \in [a,b]$:

$$|T_\lambda f_1(x) - T_\lambda f_2(x)| \leq |\lambda| \int_a^b |k(x,y)||f_1(y) - f_2(y)| dy \leq |\lambda| \|k\|_\infty \int_a^b |f_1(y) - f_2(y)| dy.$$

By the boundedness of $f_1 - f_2$ on $[a,b]$ we have that $|f_1(y) - f_2(y)| \leq \|f_1 - f_2\|_\infty$. Thus we have

$$|T_\lambda f_1(x) - T_\lambda f_2(x)| \leq |\lambda| \|k\|_\infty \|f\|_\infty \int_a^b 1 dy = |\lambda| (b-a) \|k\|_\infty \|f_1 - f_2\|_\infty$$

Hence T_λ is a contraction on the Banach space $(C[a,b], \|.\|_\infty)$ if $|\lambda|(b-a)\|k\|_\infty < 1$, i.e.

$$|\lambda| < ((b-a)\|k\|_\infty)^{-1}$$

and so Banach's fixed point theorem completes the argument.

□

Mappings of the form $T(f)(x) = \int_a^b k(x,y) f(y) dy$ are called **integral operators** and one may impose various conditions on $[a,b]$, the function f and the kernel k depending on your problem. We just point out that a specific choice of kernels gives integral operators with a one-dimensional range. Namely, if $k(x,y) = k_1(x) k_2(y)$, then

$$Tf(x) = \int_a^b k_1(x) k_2(y) f(y) dy = \langle k_2, f \rangle_2 \, k_1(x),$$

is a scalar multiple of k_1. We denote functions of the form $k_1(x)k_2(y)$ by $(k_1 \otimes k_2)(x,y)$. We call operators with one-dimensional range, rank-one operators. If the kernel is of the form $k(x,y) = k_1^{(1)}(x) k_2^{(2)}(y) + \cdots + k_1^{(n)}(x) k_2^{(n)}(y)$, then the range of the associated integral operator has a finite-dimensional range and we call these finite-rank operators.

Integral operators are ubiquitous in mathematics, science and engineering, e.g. as filters and channels in signal analysis and cybernetics, as pseudodifferential operators in the generalization of differential operators or in the description of quantization procedures in quantum mechanics. Kernels have often a special structure, e.g. it might be of the form $k(x-y)$. Then $Tf(x) = \int_a^b k(x-y) f(y) dy$ and Tf is a modified variant of f with respect to the function k, and is known as convolution $k * f$ of f with k. Hence $Tf = k * f$ might be viewed as weighted average of f, e.g. if k

equals the function $k(x-y) = (b-a)^{-1}$ on $[a,b]$, then $k*f(x) = (b-a)^{-1}\int_a^b f(y)dy$. Convolution operators are an integral part of engineering as time-variant filters or time-variant channels etc . We close with an application to an important topic: the existence and uniqueness of solutions of an ordinary differential equation (ODE), aka the **Picard-Lindelöf Theorem.**

We consider the following initial value problem:

$$x'(t) = \frac{dx}{dt} = f(t,x) \quad \text{and } x(t_0) = x_0 \tag{3.4}$$

for a function $f : A \subset \mathbb{R}^2 \to \mathbb{R}$ with $t_0 \in I$.

Definition 3.3.6. Let I be an interval and $t_0 \in I$. A differentiable function $x : I \to \mathbb{R}$ is a *solution* of the IVP (3.4) if for all $t \in \mathbb{R}$ we have $x'(t) = f(t, x(t))$ and $x(t_0) = x_0$.

We say that the IVP has a *local solution* if there exists a $\delta > 0$ such that (3.4) has a solution x on $(x_0 - \delta, x_0 + \delta)$.

Now we can state the theorem of Picard-Lindelöf and in the sketch of its proof we will also show how to construct approximately a solution to IVPs.

Theorem 3.18 (Picard-Lindelöf). *Consider the initial value problem:*

$$x'(t) = \frac{dx}{dt} = f(t,x) \quad \textit{and } x(t_0) = x_0), \tag{3.5}$$

where $f : U \times V \to \mathbb{R}$ is a function, U, V are intervals with t_0 in the interior of U and x_0 in the interior of V.
Assume that f is continuous and uniformly Lipschitz in x:

$$|f(t,x) - f(t,x')| \le L|x - x'| \quad \textit{for all } t \in U, x, x' \in V.$$

Then the IVP has a unique local solution.

We start with a more precise formulation of the assumptions on f.

We have that f is a continuous function defined $f : U \times V \to \mathbb{R}$ on the intervals $U = [t_0 - a, t_0 + a]$, $V = [x_0 - b, x_0 + b]$ for $a, b > 0$, such that

$$|f(t,x) - f(t,x')| \le L|x - x'| \quad \text{for all } t \in U, x, x' \in V.$$

The assumptions on f imply that it is bounded, i.e. there exists a $M > 0$ such that $|f(t,x)| \le M$ for all $(t,x) \in U \times V$. Hence, the theorem of Picard-Lindelöf asserts that for $\delta < \min a, 1/L, b/M$ the IVP has a solution on $[t_0 - \delta, t_0 + \delta]$.

A key step in the proof is the reformulation of the theorem in terms of an integral equation.

Lemma 3.19. *The IVP has a solution if and only if*

$$x(t) = x_0 + \int_{t_0}^{t} f(s, x(s))ds.$$

Proof. We define φ on U by $\varphi(t) = f(t, x(t))$. By the Fundamental Theorem of Analysis $x_0 + \int_{t_0}^{t} \varphi(s)ds$ is the anti-derivative of f whose value at t_0 is x_0. □

The next step is an iterative procedure to solve the integral equation, also known as *Picard iteration*.
We define an operator T on $(C(I), \|.\|_\infty)$ for an interval I by

$$T(x)(t) = x_0 + \int_{t_0}^{t} f(s, x(s))ds.$$

Then x solves the integral equation

$$x(t) = x_0 + \int_{t_0}^{t} f(s, x(s))ds.$$

if and only if $T(x) = x$. We leave the technical details out of the discussion, which just amount to modify the discussion of the solution of linear integral equations to the non-linear case.

CHAPTER 4

Continuous functions between normed spaces

Continuous functions may be viewed as the simplest functions. In this chapter we define continuous functions between normed spaces. Before are we shifting our focus to special classes of sets in normed spaces: open sets and closed sets. We also comment briefly on the concept of topologies on a set.

4.1. Closed and open sets

Here is a reminder about the definition of limit points of a set and closed subsets.

Suppose $(X, \|.\|)$ is a normed space and A a subset of X. Then $x \in X$ is a *limit point* of A if there exists a sequence of points in A that converges to x. We denote by $\overline{A}$ the set of all limit points of A. Other terminology for limit point is accumulation point. Note that a sequence might have several limit points.

LEMMA 4.1. *For a subset A of a normed space $(X, \|.\|)$ we have $A \subseteq \overline{A}$.*

PROOF. We have to show that any point a of A is limit point of A. Hence we have to find a sequence (a_n) in A that converges to a. There is an evident choice: Namely the constant sequence $(a, a, ...)$. □

As we have noted before, $\overline{A}$ might be strictly larger than A. This lead us to define a special class of sets, closed sets, where this cannot occur: A is called *closed* if $\overline{A} = A$.

The definition of the closure of a set involved sequences which we would like to formulate it more elementary.

LEMMA 4.2. *Suppose $(X, \|.\|)$ is a normed space and A a subset of X. Then $x \in X$ is a limit point of A if and only if we have that for every $\varepsilon > 0$ there exists $a \in A$ with $\|x - a\| < \varepsilon$.*

PROOF. ($\Rightarrow$) Suppose $x \in X$ is a limit point of A. Then there exists (a_n) in A such that $a_n \to x$. Hence for any $\varepsilon > 0$ there exists an $N \in \mathbb{N}$ such that $\|x - a_n\| < \varepsilon$ for any $n \geq N$. In particular, any ball $B_\varepsilon(x)$ contains a point different from a.

($\Leftarrow$) Since we have that $B_\varepsilon(x) \cap A \neq \emptyset$ for any $\varepsilon > 0$, we definitely have for $\varepsilon = 1/n$ for $n = 1, 2, 3, ..$ that there exists an element $a_n \in A$ with $\|x - a_n\| < 1/n$. The sequence (a_n) lies in A and converges to x. □

Let us turn this criterion into a useful characterization of closed subsets.

LEMMA 4.3. *Suppose $(X, \|.\|)$ is a normed space and A a subset of X. Then A is closed if and only if for each $x \in X \backslash A$ there exists $\varepsilon > 0$ such that $B_\varepsilon(x) \subseteq X \backslash A$.*

PROOF. By definition a closed subset contains all its limit points. Hence A is closed if and only if its complement $X\backslash A$ contains no limit points of A. Now, a point $x \in X$ is not a limit point of A if and only if there exists some $\varepsilon > 0$ such that for all $a \in A$ we have $\|x - a\| \geq \varepsilon$, i.e. $B_\varepsilon(x) \cap A = \emptyset$ or equivalently $B_\varepsilon(x) \subseteq X\backslash A$. □

LEMMA 4.4. *Suppose $(X, \|.\|)$ is a normed space and A a subset of X. Then $\overline{A}$ is a closed set.*

PROOF. Let x be a limit point of $\overline{A}$. Then for any $\varepsilon > 0$ there exists $a \in \overline{A}$ with $\|x - a\| < \varepsilon/2$. Since $a \in \overline{A}$, it is a limit point A, and there exists a point $a_1 \in A$ with $\|a - a_1\| < \varepsilon/2$. Thus we have

$$\|x - a_1\| \leq \|x - a\| + \|a - a_1\| < \varepsilon.$$

Hence $x \in \overline{A}$. □

An elementary and useful fact concerns closed subspaces of Banach spaces that is one of tasks of next weeks problem set.

LEMMA 4.5. *Suppose A is a closed subset of a Banach space $(X, \|.\|)$. Then $(A, \|.\|)$ is a complete subspace of X, i.e. $(A, \|.\|)$ is a Banach space.*

Let us close the discussion of closed sets with a useful characterization.

LEMMA 4.6. *Let $(X, \|.\|)$ be a normed spaces and $A \subseteq X$. Then $\overline{A}$ is the smallest closed subset containing A. Hence $\overline{A}$ is the intersection of all closed sets containing A.*

PROOF. Let B be a closed subset with $A \subseteq B$. If a is a limit point of A, then there exists (a_n) in $A \subseteq B$ such that $a_n \to a$. Hence $a \in \overline{B} = B$ and thus $\overline{A} \subseteq B$. Since the intersection of an arbitrary intersection of closed subsets is closed, we have that the intersection of all closed sets containing A is closed. By the minimality of $\overline{A}$ we deduce the required statement. □

COROLLARY 4.1.1. *Let $(X, \|.\|)$ be a normed spaces and $A \subseteq X$. Then*

$$\overline{A} = \cap_{n \in \mathbb{N}}(A + B_{1/n}(0)).$$

The proof of the corollary is part of the next problem set.

Our concept of open sets in normed spaces relies on the existence of a distance function since we define a notion of "closeness" of a point by the use of balls $B_r(x)$.

DEFINITION 4.1.2. Suppose X is a normed space and $U \subseteq X$.

(1) A point $x \in U$ is called an *interior point* of U if there is an $\varepsilon > 0$ such that $B_\varepsilon(x) \subseteq U$. We denote the set of all interior points of U by $\text{int}(U)$.
(2) The set U is called *open* if each point in U is an interior point, i.e. for each point $x \in U$ there is an $\varepsilon > 0$ such that $B_\varepsilon(x) \subseteq U$.

The characterization of closed sets via its complement becomes in terms of open sets:

LEMMA 4.7. *A subset A of a normed space X is closed if and only if its complement $X\backslash A$ is open.*

We collect some elementary observations about open sets.

LEMMA 4.8. *If U_1 and U_2 are open subsets of a normed space X, then $U_1 \cap U_2$ is also open.*

PROOF. If $x \in U_1 \cap U_2$ there is an $\varepsilon_1 > 0$ for which $B_{\varepsilon_1}(x) \subseteq U_1$ and an $\varepsilon_2 > 0$ for which $B_{\varepsilon_2}(x) \subseteq U_2$. Then $\varepsilon = \min\{\varepsilon_1, \varepsilon_2\}$ gives a ball $B_\varepsilon(x) \subseteq U_1 \cap U_2$. □

Furthermore, the empty set and X are open sets, the union of arbitrary collections of open sets is also an open set. As in the case of the closure of a set we have that the interior of a set is open and that it is the largest open set containing the set.

DEFINITION 4.1.3. For a set A in a normed space X we define its *boundary*, ∂A by $\partial A := \overline{A} \cap \overline{X \backslash A}$, i.e. the boundary of A is the set of all limit points which are arbitrarily close to both A and its complement $X \backslash A$.

The set ∂A is symmetric under changing A to $X \backslash A$ and thus we have $\partial A = \partial(X \backslash A)$.

DEFINITION 4.1.4. We call the set of all open sets of a normed space X, the *topology* of X.

One has that $\partial A = \overline{A} \backslash \text{int} A$ and we have $\text{int} A \cap \partial A = \emptyset$ and $\overline{A} = \text{int} A \cup \partial A$.

An important concept in this context is the one of a dense subset.

DEFINITION 4.1.5. A subset A of $(X, \|.\|)$ is said to be *dense* in $\mathbb{R}$ if its closure is equal to X, i.e. $\overline{A} = X$. If the dense subset A is countable, then X is called *separable*.
In other words, a subset A of a normed space X is dense in X if for each $x \in X$ and each $\varepsilon > 0$ there exists a vector $y \in A$ such that

$$\|x - y\| < \varepsilon.$$

The relevance of a dense subset of a normed space is that it provides a way to approximate elements of the normed space by ones from the dense subset up to any given precision.

LEMMA 4.9. *Suppose A is a dense subspace of a normed space X. For any $x \in X$ there exists a sequence of elements $x_k \in A$ such that $\|x_k - x\| \to 0$ as $k \to \infty$.*

The real numbers have the **Archimedean property:** For any $x, y \in \mathbb{R}$ there exists a natural number n such that $nx > y$.

As a consequence we deduce a that $\mathbb{Q}$ is a dense subspace of $\mathbb{R}$.

PROPOSITION 4.1.6. *For $x, y \in \mathbb{R}$ with $x < y$ there exists a $r \in \mathbb{Q}$ such that $x < r < y$.*

PROOF. Goal: Find $m, n \in \mathbb{Z}$ such that

$$x < \frac{m}{n} < y. \tag{4.1}$$

First step: Choose the denominator of n large such that there exists an $m \in \mathbb{Z}$ such that $x \in (\frac{m-1}{n}, \frac{m}{n})$ are separating x and y. The Archimedean property of $\mathbb{R}$ allows

us to find a $n \in \mathbb{N}$ with this property. More concretely, we pick $n \in \mathbb{N}$ large enough such that $1/n < y - x$ or equivalently

$$x < y - \frac{1}{n} \tag{4.2}$$

Second step: Inequality (4.1) is equivalent to $nx < m < ny$. From the first step we have n already chosen. Now we choose $m \in \mathbb{Z}$ to be the smallest integer greater than nx. In other words, we pick $m \in \mathbb{Z}$ such that $m - 1 \leq nx < m$. Thus we have $m - 1 \leq nx$, i.e. $m \leq nx + 1$. By inequality (4.2)

$$m \leq nx + 1 < n(y - \frac{1}{n}) + 1 = ny,$$

hence we have $m < ny$, i.e. $m/n < y$. Once more by (4.2) we have $x \leq m/n$. These two inequalities yield the desired assertion: $x < m/n < y$. □

In an similar manner one may deduce that the irrational numbers are dense in the real numbers.

LEMMA 4.10. *For $x, y \in \mathbb{R}$ with $x < y$ there exists a $r \in \mathbb{R}\backslash\mathbb{Q}$ such that $x < r < y$.*

PROOF. Pick your favorite irrational number, a popular choice is $\sqrt{2}$. Then by the density of the rational numbers there exists a rational number $r \in (x/\sqrt{2}, y/\sqrt{2})$. Hence $r\sqrt{2} \in (x, y)$. Note that $r\sqrt{2}$ is an irrational number in (x, y) that completes our argument. □

Let us state an example of a dense subset of the ℓ^p-spaces.

PROPOSITION 4.1.7. *For $1 \leq p \leq \infty$ the set c_f of all sequences with only finitely many non-zero entries is dense in ℓ^p.*

PROOF. See problem set. □

The sequence spaces ℓ^p behave differently whether $p \in [1, \infty)$ or $p = \infty$. One instance of this phenomenon have we deduced from the definition of the respective norms: in one case the spaces consist of convergent sequence while for $p = \infty$ we have just bounded sequences. A different incarnation of this prinicple is that ℓ^∞ is not separable while all the other ℓ^p-spaces are separable.

PROPOSITION 4.1.8.
(1) *For $1 \leq p < \infty$ the sequence spaces ℓ^p are separable.*
(2) *ℓ^∞ is not separable.*

PROOF. Let us split up the arguments:

(1) For $1 \leq p < \infty$ this is a consequence of $\overline{\mathbb{Q}} = \mathbb{R}$ and the definition of the $\|.\|_p$-norm. The details are part of a problem on the next problem set.
(2) We have to show that there exists no countable dense subset of ℓ^∞. Suppose that A is a countable dense subset of ℓ^∞. We denote by b the set of all binary sequences, i.e. all sequences where the elements are either 0 or 1. We have shown that b is uncountable in Chapter 1. Note that b is a subset of ℓ^∞.

    For a $x \in b$ there exists an $x_a \in A$ such that $\|x - x_a\|_\infty < 1/2$, since A is dense in ℓ^∞ and thus also in b. The mapping $a \mapsto x_a$ is injective,

since for any two distinct binary sequences a and a' we have $\|a-a'\|_\infty = 1$ and thus we have:

$$1 = \|a - a'\|_\infty \leq \|a - x_a\|_\infty + \|x_{a'} - a'\|_\infty + \|x_a - x_{a'}\|_\infty$$

and $\|a - x_a\|_\infty < 1/2$ and $\|x_{a'} - a'\|_\infty < 1/2$ implies $\|x_a - x_{a'}\|_\infty > 0$. In summary, we have constructed an injective map between the uncountable set b and the dense countable subset X, but b is uncountable and thus we have arrived at a contradiction.

□

Continuous functions on a bounded interval have as dense subset the space of polynomials, a result of utmost importance in analysis and its applications.

THEOREM 4.11 (Weierstraß). *Let $[a, b]$ be a bounded interval of $\mathbb{R}$. Then the space of polynomials $\mathcal{P}$ is dense in $C[a, b]$. In other words, for any $f \in C[a, b]$ we that for any $\varepsilon > 0$ there exists a polynomial p such that $\|f - p\|_\infty < \varepsilon$.*

There are a number of proofs of this deep theorem. We present the one given by Landau. First we note that we can after take instead of $[a, b]$ the interval $[0, 1]$ and assume that $f(0) = f(1) = 0$, and $\|f\|_\infty = 1$. Before we prove Weierstrass's theorem we introduce a sequence of polynomials (q_n) and state some of its properties.

We consider on $[-1, 1]$ the polynomials $q_n(x) = c_n(1 - x^2)^n$, where c_n is the normalization factor $c_n = (\int_{-1}^{1}(1-x^2)^n dx)^{-1}$. Then the inequality $(1-x^2)^n \geq 1-nx^2$ on $(0, 1)$ gives that $c_n < \sqrt{n}$. Consequently, we have $\int_{-1}^{1}(1-x^2)^n dx \geq 2\int_0^{1/\sqrt{n}}(1-nx^2)dx > 1/\sqrt{n}$. Note that $q_n(-x) = q_n(x)$ and

(1) $q_n(x) \geq 0$ for $x \in [-1, 1]$;
(2) $\int_{-1}^{1} q_n(x)dx = 1$;
(3) If $0 < \delta < |x|$, then $q_n(x) < \sqrt{n}(1 - x^2)^n \leq (1 - \delta^2)^n$. Hence $q_n(x) \to 0$ as $n \to \infty$ uniformly for $|x| \geq \delta$.

The sequence $(q_n)_{n\in\mathbb{N}}$ is a so-called *approximate identity* for convolution: $\int_{-1}^{1} f(t)p_n(x-t)dt \to f(x)$ uniformly as $n \to \infty$, i.e. $\|p_n * f - f\|_\infty \to 0$ as $n \to \infty$.

PROOF. Landau defines $p_n(x) = \int_{-1}^{1} f(x - t)q_n(t)dt$ which are polynomials of degree n. Since $x - t \in [-1, 1]$ when $t \in [0, 1]$ we may express p_n as $p_n(x) = \int_0^1 f(t)q_n(x - t)dt$.
Since f is uniformly continuous on $[-1, 1]$ we have $|f(x) - f(y)|\varepsilon$ for $|x - y| < \delta$. For $x \in [0, 1]$ we have

$$p_n(x) - f(x) = \int_{-1}^{1} f(x - t)q_n(t)dt - f(x)\int_{-1}^{1} q_n(t)dt.$$

By splitting $\int_{-1}^{1}$ in $\int_{-\delta}^{-1} + \int_{-\delta}^{\delta} + \int_{\delta}^{1}$ we obtain

$$\begin{aligned} |p_n(x) - f(x)| &\leq \int_{-1}^{1} |f(x-t) - f(t)| q_n(t) dt \\ &\leq \varepsilon \int_{-\delta}^{\delta} q_n(t) dt + \varepsilon \int_{\delta}^{1} q_n(t) dt \\ &\leq \varepsilon + 4\sqrt{n}(1-\delta^2)^n. \end{aligned}$$

Hence, there exists an N such that for $n \geq N : |p_n(x) - f(x)| \leq 2\varepsilon$ and so $\|p_n - f\|_\infty \leq 2\varepsilon$. □

Weierstraß established also a variant of this theorem for continuous periodic functions where trigonometric polynomials enter the picture instead of polynomials. We denote by $\mathcal{T}$ the space of all functions of the form

$$t_n(x) = c_{-n}e^{-inx} + \cdots + c_{-1}e^{-ix} + c_0 + + c_1 e^{ix} + \cdots + c_n e^{inx}$$

for $n \in \mathbb{N}$. A function of the form t_n is a *trigonometric polynomial* of degree n. Expressing e^{ix} in terms of $\cos x$ and $\sin x$ gives the following form for t_n:

$$t_n(x) = a_0 + \sum_{k=1}^{n} (a_k \cos kx + b_k \sin kx).$$

THEOREM 4.12 (Weierstraß). *Suppose f is a continuous function of period 2π. Then for every $\varepsilon > 0$ there exists a trigononmetric polynomial t such that $\|f - t\|_\infty < \varepsilon$. In other words, $\mathcal{T}$ is dense in the space of all 2π-periodic continuous functions with resepct to the $\|.\|_\infty$-norm.*

Our proof strategy resembles closely the one for the non-periodic case since we use once more an approximate identity to construct a trigonometric polynomial with the desired properties. In the periodic case the Poisson kernel $\{P_r\}_{r\in(0,1)}$ is a good choice. Recall that $P_r(\varphi) = \frac{1}{2\pi}\sum_{k\in\mathbb{Z}} r^{|k|}e^{ik\varphi}$. The Poisson kernel is used to solve the Dirichlet problem for the Laplacian on the open unit disc with prescribed boundary data given by a continuous function f. The solution of this problem is given by $u = P_r * f$ and so we have that u is a harmonic function, i.e. $\Delta u = 0$. The Poisson kernel has some useful properties.

(1) $P_r(\varphi) \geq 0$ on $[-\pi, \pi]$;
(2) $P_r(-\varphi) = P(\varphi)$;
(3) $\int_{-\pi}^{\pi} P(\varphi) d\varphi = 1$;
(4) $\sup_{0<\delta\leq|\varphi|\leq\pi} P_r(\varphi) \leq P_r(\delta)$ and thus $P_r(\delta) \to 0$ uniformly as $\delta \to 0$.

In particular, we have that for a continuous function f on $[-\pi, \pi]$ that $f_r(\varphi) := \frac{1}{2pi}\int_{-\pi}^{\pi} f(t)P_r(\varphi - t)dt$ satisfies

$$\|f_r - f\|_\infty \to 0 \qquad \text{as } r \to 1.$$

PROOF. We can restrict the discussion to $f : [-\pi, \pi] \to \mathbb{R}$ such that $\|f\|_\infty = 1$ and $f(-\pi) = f(\pi) = 0$. We know that for every $\varepsilon > 0$ there is some $r \in (0, 1)$ such that $\|f - f_r\|_\infty \leq \varepsilon$. Note that f_r may be expressed as

$$f_r(\varphi) = \sum_{k\in\mathbb{Z}} \widehat{f}(k) r^{|k|} e^{ik\varphi}$$

with $\widehat{f}(k) = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(t)e^{-ikt}dt$ the k-th Fourier coefficient of f. Then we have

$$\sum_{|k|\geq N} |\widehat{f}(k)r^{|k|}|e^{ik\varphi} \leq \sum_{|k|\geq N} r^{|k|} = 2\frac{r^N}{1-r}.$$

Now we choose N so that $2r^N/(1-r) < \varepsilon$. Consequently, the trigonometric polynomial $t_N(\varphi) = \sum_{|k|<N} \widehat{f}(k)r^{|k|}|e^{ik\varphi}$ is the looked-after approximation of f:

$$\|f - f_N\|_\infty \leq \|f - f_r\|_\infty + \|f_r - f\|_\infty \leq 2\varepsilon.$$

□

We close our interlude on dense subsets with the existence of a completion for any normed space X.

PROPOSITION 4.1.9. *For any normed space $(X, no._X)$ there exists a Banach space $(\tilde{X}, \|.\|_{\tilde{X}})$, the completion of X, and there is an injective linear mapping $\iota : X \to B$ such that $\|\iota(x)\|_{\tilde{X}} = \|x\|_X$ for all $x \in X$ and $\iota(X)$ is dense in $\tilde{X}$.*

There is several ways to prove that any normed space has a completion. Since we are just interested in the statement but not so much in its proof, we move on to just state some interesting consequences.

PROPOSITION 4.1.10. *For $1 \leq p < \infty$ the normed space $(C[a,b], \|.\|_p)$ has a completion, which we denote by $L^p[a,b]$.*

The spaces $L^p[a,b]$ are an important class of function spaces which have intimate ties to the theory of Lebesgues measure and measurable functions.

4.1.1. Continuous functions. Our definition of open sets is naturally aligned with the concept of continuous function. In fact, it enables us to define a continuous function without recourse to preliminary definition of continuity at a point.

PROPOSITION 4.1.11. *Let $(X, \|.\|_X)$ and $(Y, \|.\|_Y)$ be normed spaces. A function $f : X \to Y$ is continuous in the $\varepsilon - \delta$ sense if and only if the preimage $f^{-1}(U)$ is open for each open set $U \subseteq Y$.*

Recall that for a set $U \subseteq Y$ its preimage is a subset of X given by $f^{-1}(U) = \{x \in X : f(x) \in U\}$.

PROOF. ($\Rightarrow$) Let us start with the $\varepsilon - \delta$ defintion: For each $x_0 \in X$ and each $\varepsilon > 0$ there is a $\delta > 0$ such that

$$\|x - x_0\|_X < \delta \;\Rightarrow\; \|f(x) - f(x_0)\|_Y < \varepsilon.$$

In terms of open balls, this says

$$x \in B_\delta(x_0) \;\Rightarrow\; f(x) \in B_\varepsilon(f(x_0)).$$

Let us phrase this in terms of preimages: For each $x_0 \in X$ and $\varepsilon > 0$ there is a δ with

$$B_\delta(x_0) \subseteq f^{-1}(B_\varepsilon(f(x_0))).$$

Hence we have established our claim for any open ball in Y and by our definition of open sets it is possible to extend this argument to any open set: If $U \subseteq Y$ is open and if $x_0 \in f^{-1}(U)$, then $f(x) \in U$. Since U is open, it contains some ball $B_\varepsilon(f(x_0))$ and so $f^{-1}(B_\varepsilon(f(x_0))$ contains $B_\delta(x_0)$. Hence $f^{-1}(U)$ is open since any point x_0 in it also contains a ball around it.

($\Leftarrow$) If for any open set $U \subseteq Y$ its preimage $f^{-1}(U)$ is open, then $f^{-1}(B_\varepsilon(f(x_0)))$ is an open set of X that includes x_0 and hence some $B_\delta(x_0)$. In this way we get a δ for each ε and if you write out the conditions of being in a δ and ε ball around the respective points, then you get the familiar $\varepsilon - \delta$ definition. □

Linear mapping between normed spaces are an important class of continuous functions. In the treatment of integral equations we noted that integral operators T with continuous kernels satisfy $\|Tf\|_\infty \leq C\|f\|_\infty$ for some $C \geq 0$ which expresses that T is a bounded operator on the space of continuous functions with respect to the supremums-norm. We formalize this in the following definition.

Definition 4.1.12. Suppose $(X, \|.\|_X)$ and $(Y, \|.\|_Y)$ are normed spaces. A linear mapping $T : X \to Y$ is called *bounded* if there exists a constant $M \geq 0$ such that

$$\|Tx\|_Y \leq M\|x\|_X \qquad \text{for all } x \in X.$$

A more extensive account of bounded operators is given in the next section. The relevance of bounded operators is supported by the fact that a linear map between normed spaces is continuous if and only if the linear map is bounded.

Proposition 4.1.13. *Let X and Y be normed spaces. For a linear transformation $T : X \to Y$ the following conditions are equivalent:*

(1) *T is uniformly continuous.*
(2) *T is continuous on X.*
(3) *T is continuous at 0.*
(4) *T is a bounded operator.*

Proof. We will show the following implications to demonstrate the assertions. From the definitions we have (i) implies (ii) and (ii) implies (iii).

(iii) $\Rightarrow$ (iv) By the continuity of T at 0 there exists a $\delta > 0$ for $\varepsilon = 1$ such that $\|Tx\| < \varepsilon = 1$ for $\|x\| \leq \delta$. We want to show that there exists a constant $C > 0$ such that

$$\|Tx\| \leq C\|x\| \quad \text{for all x with } \|x\| \leq 1$$

Note that for $x \in \overline{B_1(0)}$ we have $\frac{\delta x}{2} \in B_\delta(0)$:

$$\|\tfrac{\delta x}{2}\| = \delta\|x\|/2 \leq \delta/2 < \delta.$$

Hence $\|T(\frac{\delta x}{2})\| < 1$ Since T is linear transformation this condition is equivalent to $\|T(\frac{\delta x}{2})\| = \delta\|T(x)\|/2 < 1$ and thus $\|Tx\| \leq 2/\delta$ for $x \in B_1(0)$. In other words, T is a bounded operator.

(iv) $\Rightarrow$ (i) Since T is linear we have

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

for all $x, y \in X$. Let $\varepsilon > 0$ and $\delta = \varepsilon/C$. Then for all $x, y \in X$ with $\|x - y\| < \delta$

$$\|Tx - Ty\| = \|T(x - y)\| \leq C\|x - y\| \leq C\varepsilon/C = \varepsilon.$$

Hence T is uniformly continuous.

□

We just state the equivalence between continuity and the boundedness of a linear mapping as a separate statement due to its relevance.

PROPOSITION 4.1.14 (Boundedness ⇔ Continuity). *A linear operator T between two normed spaces X and Y is continuous if and only if it is bounded.*

LEMMA 4.13. *Let X be an innerproduct space. Then the innerproduct is continuous in each component.*

PROOF. We have to show that $x \to \langle x, y\rangle$ is continuous for a fixed $y \in X$. By the symmetry of innerproducts this also yields the continuity with respect to the second component.
By Cauchy-Schwarz

$$| \langle x - x', y\rangle | \leq \|x - x'\|\|y\|$$

for a fixed y. Hence we have a bounded map and so we proved its continuity. □

EXAMPLE 4.1.15. For $a = (a_n) \in \ell^\infty$ we define $\varphi(x) = \sum_n a_n x_n$ for $(x_n)\ell^1$. Then φ is continuous, i.e. a bounded linear functional on ℓ^1.
First we show that φ is well-defined.

$$|\varphi(x)| \leq \sum_n |a_n||x_n| \leq \|a\|_\infty \sum_n |x_n| = \|a\|_\infty \|x\|_1.$$

Furthermore this yields that φ is a bounded linear mapping from ℓ^1 to $\mathbb{C}$ and hence continuous.

Recall that the kernel of a linear operator $T : X \to Y$ is the subset of X defined by

$$\ker T = \{x \in X : Tx = 0\}.$$

PROPOSITION 4.1.16. *Let T be a linear map between normed spaces X and Y. Then the kernel of T is a closed subspace of X.*

PROOF. Suppose (x_n) is a sequence in $\ker T$ with $x_n \to x$ in X. Then the continuity of T implies that

$$Tx = T(\lim_{n\to\infty} x_n) = \lim_{n\to\infty} Tx_n = 0,$$

and $x \in \ker T$. □

The range of a bounded linear map is in general not closed.

EXAMPLES 4.1.17. (1) We define the *Volterra integral operator* $V : (C[0,1], \|.\|_\infty) \to (C[0,1], \|.\|_\infty)$ by

$$Vf(x) := \int_0^x f(y)dy.$$

The operator V is continuous: $\|Vf\|_\infty \leq \sup_{x\in[0,1]} \int_0^x |f(y)|dy \leq \int_0^1 |f(y)|dy \leq \|f\|_\infty$. Since $V(1) = x$ and $\|x\|_\infty = 1$ we have that $\|V\| = 1$.

The range of V is the set of continuously differentiable functions on $[0,1]$ that vanish at $x = 0$. Thus the range of V is a subspace of $C[0,1]$, which is not closed (The proof is part of the next problem set).

(2) Let T be the multiplication operator $Tx = (\frac{x_n}{n})$ on ℓ^∞. Then T is a bounded linear operator. The range of T is not closed:

The sequence $x_0(n) = (1, \sqrt{2}, ..., \sqrt{n}), 0, 0, ...)$ is mapped to the sequence $y_0^(n) = (1, 1/\sqrt{2}, ..., 1/\sqrt{n}, 0, 0, 0, ...)$. Hence $y_0^(n)$ are in$T(\ell^\infty)$. The sequence $(x_0^(n))_{n\in\mathbb{N}}$ converges to $x_0 = (1, \sqrt{2}, ..., \sqrt{n}), ...)$ which is not in ℓ^∞. Hence the range of T is not closed.

We close the discussion of continuous mappings with a statement about the extension of a continuous map from a dense subspace to the whole space.

PROPOSITION 4.1.18. *Let X be a normed space and Y a Banach space. Suppose M is a dense subspace of X and $T : M \to Y$ is a continuous linear map. Then there is a unique continuous linear map $\overline{T} : X \to Y$ such that $\overline{T}x = Tx$ for all $x \in M$. The map $\overline{T}$ is bounded and we have $\|\overline{T}\| = \|T\|$.*

The map $\overline{T}$ is called the *extension* of T.

PROOF. Since M is dense in X, we have for every $x \in X$ that there is a sequence (x_n) in M that converges to x. A natural candidate for the extension of T is $\overline{T}x := \lim_{n\to\infty} Tx_n$. Consider $y^{(n)} = (1, ..., \sqrt{n}, 0, 0, 0, ...)$. Then $Ty^n = x^{(n)}$ and $y^{(n)} \in \ell^\infty$ for all $n \in \mathbb{N}$, but $(x_0^{(n)})$ is a sequence that converges to x_0 which is not in $T(X)$. Hence $T(X)$ is not closed in ℓ^∞.
There is two issues we have to address: (i) Does the limit in the definition of $\overline{T}x$ exist?, and (ii) Does the definition of $\overline{T}x$ dependent on the sequence (x_n) used to approximate x?

(1) Since T is bounded and (x_n) is a Cauchy sequence, we have that (Tx_n) is also a Cauchy sequence. By the completeness of Y, (Tx_n) has a limit and it is this limit that we denote by $\overline{T}x$.
(2) Suppose that (x_n) and (x'_n) are two sequences in M that converge to x, then

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

and thus we obtain as $n \to \infty$ that $\|x_n - x'_n\| \to 0$. By the continuity of T we have that

$$\|Tx_n - Tx'_n\| \leq \|T\|\|x_n - x'_n\|$$

and as a consequence we get $\|Tx_n - Tx'_n\| \to 0$ as $n \to \infty$. In other words, the definition of $\overline{T}x$ does not depend on the approximating sequence.

The map $\overline{T}$ is linear, since T is linear. For $x \in M$ we take the constant sequence $(x, x, ...)$ to see that $\overline{T}$ is an extension of T. The boundedness of $\overline{T}$ is once more a consequence of the continuity of T:

$$\|\overline{T}x\| = \lim_{n\to\infty} \|Tx_n\| = \|T\| \lim_{n\to\infty} \|x_n\| = \|T\|\|x\|,$$

which also gives $\|\overline{T}\| \leq \|T\|$ and since we have for $x \in M$ that $\overline{Tx} = Tx$ this gives $\|\overline{T}\| = \|T\|$.

The uniqueness of the extension $\overline{T}$ is deduced by contradiction. Suppose there

is another continuous linear map $\tilde{T}$ such that $\tilde{T}x = Tx$ for $x \in M$. For $x \in X$ we have a sequence (x_n) in M converging to x:

$$\tilde{T}x = \lim_{n\to\infty} \tilde{T}x_n = \lim_{n\to\infty} Tx_n = \overline{T}x$$

and thus $\tilde{T}x = \overline{T}x$ for all $x \in X$, i.e. $\tilde{T} = \overline{T}$. □

This statement is also true for general continuous functions between normed spaces.

4.1.2. Bounded linear operators between normed spaces. Mappings between vector spaces are of interest in a wide range of applications. We restrict our focus to mappings that respect the vector space structure: linear mappings aka linear operators.

Definition 4.1.19. Let X, Y be vector spaces over the same scalar field $\mathbb{F}$. Then a mapping $T : X \to Y$ is *linear* if

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

for all $x, y \in X$ and $\lambda \in \mathbb{F}$. We denote by $\mathcal{L}(X, Y)$ the set of all linear operators between X and Y.

Linear mappings are a special class of functions between two sets. Hence it has the structure of a vector space.Here are some examples of linear mappings for the classes of vector spaces of our interest.

(1) Linear mappings between $\mathbb{F}^n$ and $\mathbb{F}^m$ are given by $m \times n$ matrices A with entries in $\mathbb{F}$, $x \mapsto Ax$ for $x \in \mathbb{F}^n$.
(2) On the space of polynomials $\mathcal{P}_n$ of degree at most n we define the *differentiation operator* $Dp(x) = a_1 x + \cdots m a_n x^{n-1}$, the operator $p \mapsto \int p(x)dx$ and the evaluation operator $Tp(x) = p(0)$.
(3) Operators on sequence spaces: For an element of the vector space s, a sequence $x = (x_n)_n$, we define the *left shift* $Lx = (0, x_0, x_1, x_2, ...)$, the *right shift* $Rx = (x_1, x_2, ...)$ and the multiplication operator $T_a x = (a_0 x_0, a_1 x_1, ...)$ for a sequence $a = (a_0, a_1, ...) \in s$. On the vector space of convergent sequences c we define $Tx = \lim_n x_n$ for $x = (x_n) \in c$.
(4) Operators on function spaces: The set of continuous functions $C(I)$ on an interval of $\mathbb{R}$, popular choices for I are $[0, 1]$ and $\mathbb{R}$. For $f \in C(I)$ we define the *integral operator* $f \mapsto \int k(x, y)f(y)dx$ for a function k defined on $I \times I$, the kernel of the operator, and the evaluation operator $Tf(x) = f(a)$ for $a \in I$. For a differentiable continuous function f we are able to study the *differentiation operator* $Df(x) = f'(x)$.

Norms on these spaces provide a tool to understand the properties of these mappings via the notion of *operator norm* that measures the size of the measure of distortion of x induced by T: For normed spaces $(X, \|.\|_X)$, $(Y, \|.\|_Y)$ and a linear mapping $T : X \to Y$ we are interested in operators such that there exists a constant c such that

$$\|Tx\|_Y \leq c\|x\|_X \quad forall\ x \in X.$$

Often we will omit the subscripts to ease the notation. The operators with a finite c are of particular relevance and are called *bounded operators*. We denote by $\mathcal{B}(X, Y)$ the set of all bounded linear operators from X to Y.

Definition 4.1.20. Let T be a linear operator between the normed spaces $(X, \|.\|_X)$ and $(Y, \|.\|_Y)$. The *operator norm* of T is defined by

$$\|T\| = \sup\{\frac{\|Tx\|_Y}{\|x\|_X} : \|x\|_X \neq 0\}.$$

Sometimes we denote the operator norm of T by $\|T\|_{\text{op}}$.

Lemma 4.14. *For $T \in \mathcal{B}(X,Y)$ the following quantities are all equal to the operator norm $\|T\|$ of T:*

(1) $C_1 = \inf\{c \in \mathbb{R} : \|Tx\|_Y \leq c\|x\|_X\}$,
(2) $C_2 = \sup\{\|Tx\|_Y : \|x\|_X \leq 1\}$,
(3) $C_3 = \sup\{\|Tx\|_Y : \|x\|_X = 1\}$.

Proof. The argument is based on some inequalities:

(1) $C_2 \leq C_1$: By definition of C_1 we have $\|Tx\| \leq C_1\|x\|$. Hence for all $x \in \overline{B_1}(0)$ we have $\|Tx\| \leq C_1$ and thus we have $C_2 \leq C_1$.
(2) $C_3 \leq C_2$: For all $x \in \overline{B_1}(0)$ we have $\|Tx\| \leq C_2$. Pick an x with $\|x\| = 1$ and define the sequence of vectors $(x_n = (1 - 1/n)v)_n$ which all have $\|x_n\| \leq 1$ and hence $\|Tx_n\| \leq C_2$ for all $n \in \mathbb{N}$. Taking the limit gives $\|Tx\| \leq C_2$ and thus $C_3 \leq C_2$.
(3) $\|T\| \leq C_3$: By definition of C_3 we have $\|Tx\| \leq C_3$ for all x with $\|x\| = 1$. Take an arbitrary non-zero vector $x \in X$. Then $x/\|x\|$ has unit length and hence $\|T(\frac{x}{\|x\|})\| = \frac{\|Tx\|}{\|x\|} \leq C_3$, which establishes the desired inequality $\|T\| \leq C_3$.
(4) We have $\|Tx\|\|x\| \leq \|T\|$ for all $x \in X$. Hence $\|Tx\| \leq \|T\|\|x\|$ for all $x \in X$. Hence we have $C_1 \leq \|T\|$. Hence we have $C_1 \leq C_2 \leq C_3 \leq \|T\| \leq C_1$ and so the assertion is established.

□

These different expressions for the operator norm of a linear operator are elementary but nonetheless useful. Before we discuss some examples we note some properties of the operator norm.

Proposition 4.1.21. *For $S, T \in \mathcal{B}(X,Y)$ we have*

(1) $\|I\| = 1$ *for the identity operator $I : X \to X$.*
(2) $\|\lambda S + \mu T\| \leq |\lambda|\|S\| + |\mu|\|T\|$ *for $\lambda, \mu \in F$.*
(3) *Submultiplicativity:* $\|S \circ T\| \leq \|S\|\|T\|$.
(4) *If T has an inverse T^{-1}, then $\|T^{-1}\| \geq \|T\|^{-1}$.*

Proof. (1) By the definition of the operator norm we have $\|I\| = 1$.
(2) The triangle inequality for norms yields the assertion.
(3) By definition we have

$$\|S \circ T\| = \sup\{\|STx\| : \|x\| = 1\} \leq \sup\{\|S\|\|Tx\| : \|x\| = 1\} = \|S\|\|T\|.$$

(4) $T^{-1}T = I$ and hence $1 = \|I\| \leq \|T\|\|T^{-1}\|$, i.e. $\|T^{-1}\| \geq \|T\|^{-1}$.

□

Proposition 4.1.22. *The vector space $\mathcal{B}(X,Y)$ of bounded operators between two normed spaces is a normed spaces with respect to the operator norm.*

PROOF. The preceding proposition implies the homogeneity property and the triangle inequality. The operator norm is clearly positive definite, and we have $\|T\| = 0$ if and only if $T = 0$ because it is defined in terms of a norm on Y. fined in terms of a norm on Y. □

We treat some of the operators defined above.

(1) The right shift $Rx = (0, x_0, x_1, x_2, ...)$ has $\|R\| = 1$ and also the left shift $Lx = (x_2, x_3, ...)$ $\|L\| = 1$ on ℓ^∞. For the multiplication operator $T_a x = (a_0 x_0, a_1 x_1, ...)$ for a sequence $a = (a_0, a_1, ...) \in s$ we have $\|T_a\| = \|a\|_\infty$ on ℓ^∞. Let us look at the right shift operator. The operator norm is given by $\|R\| = \sup\{\|Rx\|_\infty : \|x\|_\infty = 1\}$:

$$\|Rx\|_\infty = 0 + |x_0|^2 + |x_1|^2 + \cdots = \|x\|_\infty = \|x\|_\infty,$$

for all $x \in \ell^\infty$, hence $\|R\| = 1$. In a similar way one gets the norms of the other operators.

(2) The operator norm of the integral operator $T_k f(x) = \int_a^b k(x,y) f(y) dy$ on $C[a,b]$ with $\|.\|_\infty$ for an interval of finite length with a kernel $k \in C([a,b] \times [a,b])$ is $(b-a)\,|k\|_\infty$. Note that

$$\begin{aligned} \|T_k f\|_\infty &= \sup\{|\int_a^b k(x,y) f(y) dy| : x \in [a,b]\} \\ &\leq \sup\{\int_a^b |k(x,y)||f(y)| dy : x \in [a,b]\} \\ &\leq \|k\|_\infty \|f\|_\infty (b-a), \end{aligned}$$

so we have $\|T_k f\|_\infty \leq \|k\|_\infty \|f\|_\infty (b-a)$ for all non-zero $f \in C[a,b]$, i.e. $\|T_k\| \leq \|k\|_\infty (b-a)$. For the constant function $f(x) = 1$ for all $x \in [a,b]$ we get $\|T_k\| = 1$.

Some classes of operators on a normed space X: (i) *isometries* on X are linear operators T with $\|Tx\| = \|x\|$ for all $x \in X$, (ii) projections are linear operators P on X satisfying $P^2 = P$. A different way is to specify norms $\|.\|_a$ and $\|.\|_b$ on $\mathbb{C}^n$ and $\mathbb{C}^m$, respectively. Then these norms induce a norm on $\mathcal{M}_{m\times n}(\mathbb{C})$, known as the *induced* norm. From a general perspective that is the operator norm of the induced linear transformation.

EXAMPLE 4.1.23. Let $A : \mathbb{C}^n \to \mathbb{C}^n$ be a linear operator given by a matrix $A = (a_i j)$ and we put on both spaces the 1-norm. Let $A = (a_1 | \cdots | a_n)$. Then $\|A\|_{\text{op}} = \max 1 \leq j \leq n \|a_j\|_1$, i.e. it is the maximum column sum.
We have $Ax = \sum_{j=1}^n a_{ij} x_j$ and thus

$$\|A\|_{\text{op}} = \|Ax\|_1 \leq \sum_{j=1}^n |a_{ij}||x_j| \leq \|x\|_1 \max_j \|a_j\|_1.$$

Hence $\max_{\|x\|_1=1} \|Ax\|_1 \leq \max_j \|a_j\|_1$.
Let e_j be the jth standard basis vector for $\mathbb{C}^n$. Then $\|A\|_{\text{op}} = \max_j \|a_j\|_1$.

We state a sufficient condition on the infinite matrix A that implies that the linear mapping $T(x) = Ax$ maps ℓ^p into ℓ^p for $p \in [1, \infty]$. This statement is often called Schur's test, after the eminent German mathematician I. Schur.

PROPOSITION 4.1.24. *Let $A = (a_{ij})_{i,j\in\mathbb{N}}$ be an infinite matrix such that*

$$M_c = \sup_{i\in\mathbb{N}} \sum_{j=1}^{\infty} |a_{ij}| < \infty, \qquad M_r = \sup_{j\in\mathbb{N}} \sum_{i=1}^{\infty} |a_{ij}| < \infty.$$

Then we have the following for each $p \in [1,\infty]$.

(1) *For $x \in \ell^p$ the series $(Ax)_i = \sum_{j=1}^{\infty} a_{ij}x_j$ converges for $i = 1, 2, ...,$ and $((Ax)_i)$ is in ℓ^p.*
(2) *The map $T(x) = Ax$ is bounded on ℓ^p and its operator norm satisfies:*

$$\|T\| \leq M_c^{1/q} M_r^{1/p},$$

where q is given by $q = p/(p-1)$.

PROOF. We leave the proof for the reader. Split it up in the case $p = 1, p = \infty$ and $p \in (1,\infty)$ and use Hölder's inequality. □

We move on to the properties of the class of all bounded operators between normed spaces.

PROPOSITION 4.1.25. *The normed space of bounded operators $(B(X,Y), \|.\|_{\text{op}})$ is complete if and only if Y is a Banach space.*

The Banach space $(B(X,\mathbb{C}), \|.\|_{\text{op}})$ is known as the *dual space* of X, denoted by X', and its elements are refered to as *functionals* on X.

PROOF. Let (T_n) be a Cauchy sequence in $B(X,Y)$, so for any $\varepsilon > 0$ there exists a $N \in \mathbb{N}$ such that for all $m, n \geq \mathbb{N}$ we have $\|T_m - T_n\|_{\text{op}} < \varepsilon$. Hence for any $x \in X$ we have

$$\|(T_m - T_n)x\|_Y \leq \|T_m - T_n\|_{\text{op}} \|x\|_X < \varepsilon \|x\|_X.$$

Hence for all $x \in X$ the sequence $(T_n x)$ is a Cauchy sequence in Y. Since Y is a Banach space, it has a limit denoted by Tx, and thus we define $Tx = \lim_{n\to\infty} T_n x$. The limit operator T is linear and bounded.

$$\|Tx\|_Y \leq \sup_n \|T_n x\|_Y \leq \|x\|_X \sup_n \|T_n\|_{\text{op}},$$

and thus we have $\|T\|_{\text{op}} \leq \sup_n \|T_n\|_{\text{op}}$, i.e. $T \in \mathrm{B}(X,Y)$.
We show that $\|T_n - T\|_{\text{op}} \to 0$. We assume otherwise that $\|T_n - T\|_{\text{op}}$ does not converge to 0. Then there exists an $\varepsilon > 0$ and a subsequence $(T_{n_k})_k$ of (T_n) such that

$$\|T_n - T\|_{\text{op}} \geq \varepsilon \quad \text{for all k.}$$

Consequently, for every k there exists a $x_k \in X$ with $\|x_k\| = 1$ and

$$\|T_{n_k}(x_k) - T_m(x_k)\| \geq \varepsilon.$$

By assumption (T_n) is a Cauchy sequence, so one can choose a N_0 such that for all $m, n_k \geq N_0$ we have

$$\|T_{n_k}(x_k) - T_m(x_k)\| \leq \varepsilon/2$$

and this gives

$$\varepsilon \leq \|T_{n_k}(x_k) - T(x_k)\|_Y \leq \|T_{n_k}(x_k) - T_m(x_k)\|_Y + \|T_m(x_k) - T(x_k)\|_Y.$$

Hence for all $m \geq N_0$ we have

$$\|T_m(x_k) - T(x_k)\|_Y \geq \varepsilon/2.$$

That is a contradiction to the definition of T, thus we have $T_m(x_k) - T(x_k) \to 0$ in $(\mathcal{B}(X,Y), \|.\|_{\text{op}})$. □

The operator norm on $\mathcal{B}(X,Y)$ induces a notion of convergence for operators.

DEFINITION 4.1.26. Suppose (T_n) is sequence of operators in $\mathcal{B}(X,Y)$. If there exists a T in $\mathcal{B}(X,Y)$ such that

$$\lim_{n\to\infty} \|T - T_n\| = 0,$$

then we say that T_n *converges uniformly* to T.

The definition of uniform convergence of operators is analogous to the uniform convergence of sequences of continuous functions. So one might wonder about the analog of pointwise convergent sequences of functions.

DEFINITION 4.1.27. Suppose (T_n) is sequence of operators in $\mathcal{B}(X,Y)$. If there exists a bounded operator T such that $T_n x \to Tx$ in Y for all $x \in X$, then we say that T_n *converges strongly* to T.

Strong convergence does not imply uniform convergence, i.e. it is at strictly weaker notion of convergence.

EXAMPLE 4.1.28. On ℓ^2 we define P_n by $P_n(x) = (x_1, ..., x_n, 0, 0, ...)$. Then $\|P_n - P_m\| = 1$ for $n \neq m$ and thus (P_n) does not converge uniformly but for any $x \in \ell^2$ we have $\|P_n x - x\|_2 \to 0$ as $n \to \infty$. In other words, P_n converges strongly to I on ℓ^2.

4.1.3. Applications of operator norm. The operator norm provides a way to measure the "size" of a linear mapping. The fixed point theorem of Banach relies on contractions on a Banach space and we indicate a way to use this theorem for solving systems of linear equations. We are interested in this system:

$$\begin{aligned} 8x_1 - x_3 &= 2 \\ -2x_1 + 5x_2 &= 3 \\ -4x_2 + 7x_3 &= 4 \end{aligned}$$

(i) Express the system in the form $x = Ax + b$ and view $A : \mathbb{R}^3 \to \mathbb{R}^3$ as a linear operator on $(\mathbb{R}^3, \|.\|_\infty)$, where $\|x\|_\infty = \max\{|x_1|, |x_2|, |x_3|\}$ is the supremum norm of $x = (x_1, x_2, x_3) \in \mathbb{R}^3$.

Our system is equivalent to

$$\begin{aligned} x_1 &= \tfrac{1}{4}x_3 + \tfrac{1}{4} \\ x_2 &= \tfrac{2}{5}x_1 + \tfrac{3}{5} \\ x_3 &= \tfrac{4}{7}x_2 + \tfrac{4}{7} \end{aligned}$$

and thus we are looking for solutions of

$$x = Ax + b,$$

where

$$A = \begin{pmatrix} 0 & 0 & 1/8 \\ 2/5 & 0 & 0 \\ 0 & 4/7 & 0 \end{pmatrix}, \qquad b = \begin{pmatrix} 1/4 \\ 3/5 \\ 4/7 \end{pmatrix}.$$

We have

$$\|Ax\|_\infty \leq \max 1/8, 2/3.4/7\|x\|_\infty$$

and thus A is a contraction: $\|Ax - Ay\|_\infty \leq \frac{4}{7}\|x\|_\infty$.

(ii) Since A is a contraction on $(\mathbb{R}^3, \|.\|_\infty)$ Banach's fixed point theorem yields that any $x_0 \in \mathbb{R}^3$ may be used as starting point for a solution of our linear system via iteration.

The space of n-tuples $\mathbb{F}^n$ is complete for any $\|.\|_p$-norm, so it is up to you, to pick appropriate norms in the domain and target space.

The other application is once more about solutions of linear systems. Suppose T is a bounded invertible linear operator on $(X, \|.\|)$ and $b \in X$ a given vector. We would like to solve $Tx = b$. In practice one has often to deal with faulty data and so one actually has to deal with the system

$$T\widehat{x} = \widehat{b}.$$

How does the solution $\widehat{x}$ of this system relate to the one of $Tx = b$?

We introduce the error vector $e = x - \widehat{x}$ and the residual vector $r = b - \widehat{b} = b - T\widehat{x}$. The residual vector is the vector which gives how much $T\widehat{x}$ fails to match b. Then we have by linearity of T

$$Te = Tx - T\widehat{x} = b - \widehat{b} = e$$

and since T is invertible: $e = T^{-1}r$:

$$\|e\| \leq \|T^{-1}\|\|r\|.$$

We also know that $\|b\| \leq \|T\|\|x\|$ and consequently,

$$\|e\|\|b\| \leq \|T\|\|T^{-1}\|\|x\|\|r\|$$

which may be written as

$$\frac{\|e\|}{\|x\|} \leq \|T\|\|T^{-1}\|\frac{\|r\|}{\|b\|}.$$

We denote $\|T\|\|T^{-1}\|$ by $\mathrm{cond}(T)$ and call it the *condition number* of T, $\frac{\|e\|}{\|x\|}$ is the *relative error* and $\frac{\|r\|}{\|b\|}$ the relative residual of the system. The condition number measures the quality of your linear system, which satisfies $\mathrm{cond}(T) \geq 1$. A system with $\mathrm{cond}(T)$ around 2 or 3 is considered good.

4.1.4. Equivalent norms. On a vector space X one may define different norms. We describe a way to compare these norms that respects basic properties, e.g. convergent sequences.

Definition 4.1.29. Given a vector space X. Two norms $\|.\|_a$ and $\|-\|_b$ are called *equivalent* if there exist (positive) constants C_1, C_2 such that

$$C_1\|x\|_a \leq \|x\|_b \leq C_2\|x\|_a \quad \text{for all } x \in X.$$

Two equivalent norms $\|.\|_a$ and $\|-\|_b$ on a vector space X give the same class of convergent sequences: Namely, a sequence (x_n) converges in $(X, \|.\|_a)$ if and only if (x_n) converges in $(X, \|.\|_b)$,i.e. there exists an $x \in X$ such that

$$\lim_{n\to\infty} \|x_n - x\|_a = 0 \qquad \Leftrightarrow \lim_{n\to\infty} \|x_n - x\|_b = 0.$$

LEMMA 4.15. *Suppose $\|.\|_a$ and $\|-\|_b$ are equivalent norms on X.*

(1) *Then a sequence (x_n) converges with respect to the $\|.\|_a$ if and only if it converges with respect to $\|.\|_b$.*
(2) *Then a sequence (x_n) is Cauchy with respect to the $\|.\|_a$ if and only if it is Cauchy with respect to $\|.\|_b$.*

PROOF. We just give the proof for (i) and leave the other case as an exercise.

($\Leftarrow$) Suppose $\lim \|x_n - x\|_a = 0$. Then our assumption implies the existence of a constant such that

$$\|x_n - x\|_b \leq C_2 \|x_n - x\|_a$$

and hence $\lim \|x_n - x\|_b = 0$. ($\Rightarrow$) The argument is as for the other direction but now we use that there exists a constant such that $C_1\|x\|_a \leq \|x\|_b$. □

An important consequence is the following fact:

PROPOSITION 4.1.30. *Suppose $\|.\|_a$ and $\|.\|_b$ are equivalent norms on X. Then $(X, \|.\|_a)$ is a Banach space if and only if $(X, \|.\|_b)$ is a Banach space.*

PROOF. See next problem set. □

EXAMPLES 4.1.31. In the infinite-dimensional setting one has norms on vector spaces that are not equivalent.

(1) Let s be the space of all real-valued sequences. Then the $\|.\|_1$-norm and the $\|.\|_\infty$ are not equivalent on s.

    Fix an $N \in \mathbb{N}$ and take the sequence $x = (1, 1, ..., 1, 0, 0, 0, ...)$ with N non-zero entries. Then $\|x\|_1 = N$ and $\|x\|_\infty = 1$. Hence we have $N\|x\|_\infty \leq \|x\|_1$ and thus there exists no finite constant M such that $\|x\|_1 \leq M\|x\|_\infty$ for all $x \in s$.
(2) Let us take the space of continuous functions $C[0, 1]$ and complete it with respect to $\|.\|_2$ and $\|.\|_\infty$. Then we have shown that $(C[0, 1], \|.\|_2)$ is not complete, but $(C[0, 1], \|.\|_\infty)$ is a Banach space.

Let us distill the general principle underlying this example.

PROPOSITION 4.1.32. *Suppose $\|.\|_a$ and $\|.\|_b$ are two norms on a vector space X. Then $\|.\|_a$ and $\|.\|_b$ are not equivalent if there exists a sequence (x_n) in X such that $\|x_n\|_b = 1$ for all $n \in \mathbb{N}$ but $\|x_n\|_a = n$ for all $n \in \mathbb{N}$.*

We continue with some properties of equivalent norms.

LEMMA 4.16. *Given three norms $\|.\|_a, \|.\|_b$ and $\|.\|_c$ on X. Suppose $\|.\|_a$ and $\|.\|_c$ are equivalent and $\|.\|_b$ and $\|.\|_c$. Then $\|.\|_a$ and $\|.\|_b$ are equivalent norms.*

PROOF. We have constants C_1, C_2, C_1', C_2' such that

$$C_1\|x\|_c \leq \|x\|_a \leq C_2\|x\|_c$$

and

$$C_1'\|x\|_c \leq \|x\|_b \leq C_2'\|x\|_c.$$

Hence $\|x\|_c \leq C_1^{-1}\|x\|_a$ and thus

$$\|x\|_b \leq C_2'C_1^{-1}\|x\|_a,$$

which by the second set of inequalities gives

$$\|x\|_b \leq C_2'C_1^{-1}\|x\|_a.$$

In a similar way, we obtain

$$C_1'C_2^{-1}\|x\|_a \leq \|x\|_b$$

and thus $\|x\|_b$ and $\|x\|_a$ are equivalent

$$C_1'C_2^{-1}\|x\|_a \leq \|x\|_b \leq C_2'C_1^{-1}\|x\|_a.$$

□

Open sets in normed spaces are defined in terms of open balls and thus there might be a relation between equivalent norms and open sets. Indeed, there is a close link. We denote by $B_r^a(x) = \{y \in X : \|x-y\|_a < r\}$ and $B_r^b(x) = \{y \in X : \|x-y\|_b < r\}$ the open balls of radius r and center $x \in X$ with respect to the norms $\|.\|_a$ and $\|.\|_b$.

PROPOSITION 4.1.33. *Let $\|.\|_a$ and $\|.\|_b$ be two norms on a vector space X. Then the following statements are equivalent:*

1. *$\|.\|_a$ and $\|.\|_b$ are equivalent norms.*
2. *There exists some $r > 0$ such that $B_{1/r}^a(0) \subseteq B_1^b(0) \subseteq B_r^a(0)$.*
3. *For a set $U \subseteq X$ we have that U is open in $(X, \|.\|_a)$ if and only if U is open in $(X, \|.\|_b)$.*
4. *For a set $F \subseteq X$ we have that F is closed in $(X, \|.\|_a)$ if and only if F is closed in $(X, \|.\|_b)$.*

PROOF. We just prove $(i) \Leftrightarrow (ii)$ and leave the other claims as an exercise.

$(i) \Leftarrow (ii)$ Suppose that $\|.\|_a$ and $\|.\|_b$ are equivalent norms. Then there exists an $r > 0$ such that

$$\frac{1}{r}\|x\|_b \leq \|x\|_a \leq r\|x\|_b \qquad \text{for all } x \in X.$$

Then for x with $\|x\|_b < 1$ we have $\|x\|_b \leq r\|x\|_a < r$ and thus we have $B_1^b(0) \subseteq B_r^a(0)$. Now we assume $x \in X$ and $\|x\|_a < 1/r$. Then we get that $\|rx\|_a < 1$. Since the norms are equivalent we have $\frac{1}{r}\|rx\|_b \leq \|rx\|_a < 1$ and thus we have $\|x\|_b < 1$, i.e. $B_{1/r}^a(0) \subseteq B_1^b(0)$.

$(ii) \Leftarrow (i)$ Suppose $B_{1/r}^a(0) \subseteq B_1^b(0) \subseteq B_r^a(0)$ holds for some $r > 0$. Then for any $x \in X$ we have that $\frac{x}{2\|x\|_b}$ is in $B_1^b(0)$ and consequently in $B_r^a(0)$, i.e. $\|\frac{x}{2\|x\|_a}\|_b < r$. Hence we have

$$\|x\|_b \leq 2r\|x\|_a.$$

The other inclusion follows by the same reasoning. □

On a finite-dimensional vector space X all norms are equivalent.

THEOREM 4.17. *All norms on $\mathbb{R}^n$ are equivalent.*

PROOF. By Lemma 4.16 it suffices to show a norm $\|.\|$ on $\mathbb{R}^n$ is equivalent to a fixed norm. We fix the $\|.\|_1$ on $\mathbb{R}^n$. Suppose $e_1, ..., e_n$ is a basis for $\mathbb{R}^n$. Then any $x \in \mathbb{R}^n$ has a unique expansion

$$x = \sum_{i=1} a_i e_i$$

and its 1-norm is defined by

$$\|x\|_1 = \sum_{i=1}^{n} \|a_i|.$$

The proof may be broken up into four steps. Step 1 is the reduction of the general case to the situation that we have to show that $\|.\|$ is equivalent to $\|.\|_1$. Step 2 is the elementary observation that it suffices to check the desired assertion

$$C_1\|x\|_1 \leq \|x\| \leq C_2\|x\|_1$$

not for all $x \in X$ but just for elements in the unit ball of $\|.\|_1$. Namely, the preceding inequalities are true for $x = 0$. Let us assume $x \neq 0$. Then we can divide the inequalities by $\|x\|_1$:

$$C_1 \leq \|x/\|x\|_1\| \leq C_2.$$

Since the elements we have to check our inequalities are now in $B_1(0)$ defined by the $\|.\|_1$.

The next step paves the way to make the problem accessible to methods from analysis. Step 4: $\|.\|$ is continuous under $\|.\|_1$. Explicitly, we have to show that for a given $\varepsilon > 0$ there exists a $\delta > 0$ such that $\|x - x'\|_1 < \delta$ implies that $|\,\|x\| - \|x'\|\,| < \varepsilon$. We know that

$$|\,\|x\| - \|x'\|\,| \leq \|x - x'\|.$$

Let us relate the $\|.\|$ with $\|.\|_1$. We represent x and x' with respect to the basis $\{e_1, ..., e_n\}$:

$$x = \sum_{i=1}^{n} a_i e_i \quad \text{and } x' = \sum_{i=1}^{n} a_i' e_i.$$

The triangle inequality implies

$$\|x - x'\| \leq \sum_{i=1}^{n} |a_i - a_i'|\|e_i\| \leq (\max_i \|e_i\|)\|x - x'\|_1.$$

Choose $\delta = \varepsilon/\max_i \|e_i\|$. Then we get the desired statement: If $\|x - x'\|_1 < \delta$, then

$$|\,\|x\| - \|x'\|\,| \leq \varepsilon.$$

The final step is to use the the Extreme Value Theorem for the continuous function $\|.\|$ on $\mathbb{R}^n$ and note that the set $\{x \in X : \|x\|_1 = 1\}$ is closed and bounded. Then $\|.\|$ has to achieve its minimum and maximum on the unit ball for the 1-norm:

$$C_1 := \max\{\|x\| : \|x\|_1 = 1\} \quad \text{and } C_2 := \min\{\|x\| : \|x\|_1 = 1\}.$$

By definition we have $C_2 \geq C_1$ and hence

$$C_1 \leq \|x\| \leq C_2$$

for $x \in X$ with $\|x\|_1 = 1$. □

A consequence of the equivalence of norms on $\mathbb{R}^n$ is that a sequence in $\mathbb{R}^n$ converges in norm if and only if converges coordinate-wise.

PROPOSITION 4.1.34. *Let $\|.\|$ be a norm on $\mathbb{R}^n$, and (x_j) a sequence in $\mathbb{R}^n$. Then $\|x_j - x\| \to 0$ if and only if $x_j^{(i)} \to x^{(i)}$ for $i = 1, ..., n$.*

PROOF. ($\Leftarrow$) Since all norms are equivalent we are allowed to pick a norm most appropriate for our problem. We pick the sup-norm.
Suppose $\lim_j \|x_j - x\| = 0$. Denote the components of x_j by $x_j = (x_j^{(1)}, ..., x_j^{(n)})$. Then $x_j^{(i)}$ converges to $x^{(i)}$ for $i = 1, ..., n$.

($\Rightarrow$) For this direction we use the 1-norm. Suppose $x_j^{(i)} \to x^{(i)}$ for $i = 1, ..., n$. Then $\|x_j - x\|_1 = \sum_{j=1}^n |x_j^{(i)} - x^{(i)}| \to 0$. □

CHAPTER 5

Best approximation and projection theorem

This chapter is based on a classical theorem of the theory of Hilbert spaces, the projection theorem, that was first proved by E. Schmidt in his Ph.D. thesis, where he also established the Gram-Schmidt orthogonalization method. The projection theorem indicates that Hilbert spaces are in some sense infinite dimensional Euclidean spaces. In contrast, the structure of Banach spaces is quite rich and full of strange phenomena.

Let M be a line through the origin of $\mathbb{R}^2$. Then any vector in the plane not in M may be projected onto this line and there is one way that has minimal distance. The latter arises via orthogonal projection. We demonstrate that this method also works in general Hilbert spaces as long as one takes closed subspaces.

We give some examples of subspaces in the Hilbert spaces $M_n(\mathbb{R})$, ℓ^2 and $L^2[0,1]$.

EXAMPLES 5.0.35. (1) The set of all symmetric matrices M_s, and the set of all anti-symmetric matrices M_{as} are closed subspaces of $M_n(\mathbb{R})$, the space of $n \times n$-matrices with real entries.

$$M_s = \{A \in M_n(\mathbb{R}) : A = A^T\} \qquad M_{as} = \{A \in M_n(\mathbb{R}) : A^T = -A\}.$$

(2) In ℓ^2 we will consider $M = \{x \in \ell^2 : (x_1, ..., x_n, 0, 0, 0, ...)\}$ and $M = \{x \in \ell^2 : (x_1, 0, x_3, ...)\}$. For the space of square-integrable sequences defined on the integers $\ell^2(\mathbb{Z})$ we have as subspace $M = \{x \in \ell^2(\mathbb{Z}) : (..., 0, 0, x_0, x_1, ...)\}$.

(3) Subspaces in $L^2[-1,1]$ we are interested in, are $M = \{f \in L^2[-1,1] : f(x) = 0 \text{ for } x \in [-1,0]\}$, the spaces of even and odd functions $M_e = \{f \in \mathrm{L}^2[-1,1] : f(-x) = f(x)\}$ and $M_o = \{f \in \mathrm{L}^2[-1,1] : f(-x) = -f(x)\}$.

All these subspaces are closed in the respective Hilbert spaces.

A consequence of our main result are decompositions of elements in Hilbert spaces with respect to subspaces and these generalize the well-known facts for matrices and functions. Any matrix may be decomposed into a symmetric and an anti-symmetric matrix:

$$A = \frac{A + A^T}{2} + \frac{A - A^T}{2}$$

and any $f \in L^2[-1,1]$ may be written as an even and odd function:

$$f(x) = \frac{f(x) + f(-x)}{2} + \frac{f(x) - f(-x)}{2}.$$

Let M be a subspace of X. Denote by $M^\perp$, its *orthogonal complement*, the set of all $x \in X$ that are orthogonal to all the elements of M. Formally we have

$$M^\perp = \{x \in X : \langle x, y\rangle = 0 \text{ for all } y \in M\}.$$

The linearity of an innerproduct implies that M is a vector space.

LEMMA 5.1. *Let M be a subspace of $(X, \langle ., .\rangle)$. Then $M^\perp$ is a closed subspace of X.*

PROOF. Let (x_n) be a sequence in $M^\perp$ converging to $x \in X$. We have to show that $x \in M^\perp$. Since $\langle x_n, y\rangle = 0$ for all $y \in M$ we note that

$$|\langle x_n - x, y\rangle| \leq \|x_n - x\|\|y\| \to 0.$$

Hence we have

$$\langle x_n, y\rangle \to \langle x, y\rangle,$$

but $\langle x_n, y\rangle = 0$ for all n. Consequently, $\langle x, y\rangle = 0$ and so $x \in M^\perp$. □

By definition of $M^\perp$ we have that M and $M^\perp$ are disjoint subspaces of X. For any proper closed subspace M of X its orthogonal complement $M^\perp$ is non-empty and there are sufficiently many elements in $M^\perp$ that allows one to decompose elements in X with respect to M and $M^\perp$. The precise formulations of these facts and their proofs are the main parts of our treatment of Hilbert spaces.

The best approximation property holds for proper closed subspaces of Hilbert spaces.

THEOREM 5.2 (Best Approximation Theorem). *Suppose M is a proper closed subspace of a Hilbert space X. Then for any $x \in X$ there exists a unique element $z \in M$ such that*

$$\|x - z\| = \inf_{m \in M} \|x - m\|.$$

The quantity $\inf_{m\in M} \|x - m\|$ measures the distance of x from M. In the chapter on metric spaces we show that it defines an honest metric on X.

REMARK 5.0.36. In general the theorem is not true in Banach spaces. Take ℓ^∞ and as closed subspace c_0, the space of sequences converging to zero. For $x = (1, 1, 1, ...)$ there exists no sequence in c_0 attaining the minimal distance 1.

PROOF. Denote by $d = \inf_{m\in M} \|x - m\|^2$. Note that d is finite, since the real numbers $\|x - m\|$ for $m \in M$ are all nonnegative and bounded below by 0. Since d is the greatest lower bound of this set, there exists a sequence $(m_k) \subset M$ such that for each $\varepsilon > 0$ there exists an N such that $\|x - m_k\|^2 \leq d + \varepsilon$ for all $k \geq N$.
Claim: The sequence (m_k) is a Cauchy sequence. Applying the parallelogram identity to $x - m_k$ and $x - m_l$ we get

$$\|2x - m_k - m_l\|^2 + \|m_k - m_l\|^2 = 2(\|x - m_k\|^2 + \|x - m_l\|^2),$$

which yields to

$$\|x - \frac{m_k + m_l}{2}\|^2 + \|m_k - m_l\|^2/2 = (\|x - m_k\|^2 + \|x - m_l\|^2)/2.$$

Since $\frac{m_k+m_l}{2} \in M$ we have $\|x - \frac{m_k+m_l}{2}\|^2 \geq d$ and so we have

$$\|m_k - m_l\|^2 \leq 2(\|x - m_k\|^2 + \|x - m_l\|^2) - 4d.$$

For any $\varepsilon > 0$ there exists a N such that $\|x - m_k\|^2 \leq d + \varepsilon/4$ for all $k \geq N$. Then we have for all $k, l \geq N$ that

$$\|m_k - m_l\|^2 \leq 2(\|x - m_k\|^2 + \|x - m_l\|^2) - 4d \leq \varepsilon.$$

Hence we have demonstrated that (m_k) is a Cauchy sequence. Since M is closed, (m_k) converges to some element $z \in M$ and we have that $\|x - z\|^2 = d$ and so z is the vector in M closest to x. We have established the existence of a closest vector. The uniqueness goes as follows: Suppose there is another element $y \in M$ such that $\|x - y\|^2 = d$. Consider the sequence $(y, z, y, z, ...)$, and note that it is a Cauchy sequence by the same argument as for (m_k). Hence $y = z$ and so z is the unique solution to our approximation problem. □

There is a characterization of best approximations in Hilbert spaces in terms of the orthogonal complement.

THEOREM 5.3 (Characterization of Best Approximation). *Suppose M is a proper closed subspace of a Hilbert space X. Then for any $x \in X$ there exists a best approximation $\tilde{x} \in M$ if and only if $x - \tilde{x} \in M^\perp$.*

PROOF. *First step:* Suppose $x - \tilde{x} \in M^\perp$. Then for any $y \in M$ with $y \neq \tilde{x}$ we have $\|y - x\|^2 = \|y - \tilde{x} + \tilde{x} - x\|^2$. Note that $y - \tilde{x} \in M$ and $\tilde{x} - x \in M^\perp$ so we have $\langle y - \tilde{x}, \tilde{x} - x\rangle = 0$. Hence Pythagoras yields $\|y - x\|^2 = \|y - \tilde{x}\|^2 + \|\tilde{x} - x\|^2$. By assumption $y - \tilde{x} \neq 0$ so we arrive at the desired assertion $\|y - x\|^2 > \|\tilde{x} - x\|^2$. *Second step:* Suppose $\tilde{x}$ minimizes $\|x - \tilde{x}\|$. We assume that there exists a $y \in M$ of unit length such that $\langle x - \tilde{x}, y\rangle = \delta \neq 0$.
Consider the element $z = \tilde{x} + \delta y$.

$$\begin{aligned}\|x - z\|^2 &= \|x - \tilde{x} - \delta y\|^2 \\ &= \langle x - \tilde{x}, x - \tilde{x}\rangle - \langle x - \tilde{x}, \delta y\rangle - \langle \delta y, x - \tilde{x}\rangle + \langle \delta y, \delta y\rangle \\ &= \|x - \tilde{x}\|^2 - |\delta|^2 - |\delta|^2 + |\delta|^2 \\ &= \|x - \tilde{x}\|^2 - |\delta|^2.\end{aligned}$$

Thus we have $\|x - z\|^2 \leq \|x - \tilde{x}\|^2$. Contradiction to the assumption that $\tilde{x}$ minimizes $\|x - \tilde{x}\|$. □

THEOREM 5.4 (Projection Theorem). *Let M be a closed proper subspace of a Hilbert space X. Then every $x \in X$ can be uniquely written as $x = y + z$ where $y \in M$ and $z \in M^\perp$.*

PROOF. For $x \in X$ there exists a best approximation $y \in M$. Note that $x = y + x - y$ with $y \in M$ and $x - y \in M^\perp$. Furthermore we have $M \cap M^\perp = \{0\}$ (if $x \in M \cap M^\perp$, then $\langle x, x\rangle = 0 = \|x\|^2$ and thus $x = 0$.) which completes the proof. □

COROLLARY 5.0.37. *Let M be proper closed subspace of a Hilbert space X. Then $M^\perp \neq \{0\}$.*

PROOF. If $x \neq M$, then the decomposition $x = y + z$ has a $z \neq 0$. Since $z \in M^\perp$ we have $M^\perp \neq \{0\}$. □

Recall that a *projection* on a normed space X is a linear mapping $P : X \to X$ satisfying $P^2 = P$.

Here is a reformulation of the preceding theorem in terms of projections, justifying the name.

PROPOSITION 5.0.38. *For any closed subspace M of a Hilbert space X, there is a unique projection P on X satisfying:*

(1) $\mathrm{ran}(P) = M$ *and* $\mathrm{ran}(I - P) = M^\perp$.
(2) $\|Px\| \leq \|x\|$ *for all* $x \in X$. *Moreover,* $\|P\| = 1$.

PROOF. (1) The decomposition of $x \in X$ into $x = y + z$ for $y \in M, z \in M^\perp$ allows one to define $Px := y$. By definition $\mathrm{ran}(P) \subseteq M$ and if $x \in M$, then $Px = x$. Thus $P^2 = P$ and $M \subseteq \mathrm{ran}(P)$.
Once more, by $x = y + z$ we have $(I - P)x = z \in M^\perp$ and as above we deduce that $\mathrm{ran}(I - P) = M^\perp$.

(2) By Pythagoras we have $\|x\|^2 = \|Px\|^2 + \|z\|^2$ and thus we have $\|Px\| \leq \|x\|$. Hence $\|P\| \leq 1$. On the other hand, there exists $x \in X$ with $Px \neq 0$ and $\|P(Px)\| = \|Px\|$, so that $\|P\| \geq 1$. Hence we conclude that $\|P\| = 1$. □

EXAMPLE 5.0.39. Let M be the line $\{t\xi : t \in \mathbb{R}\}$ given by a unit vector $\xi \in X$. Then

$$P_\xi x = \langle \xi, x \rangle \xi$$

projects a vector orthogonally onto its component in direction ξ

EXAMPLE 5.0.40. In $L^2[-1,1]$, consider the closed subspaces $M_e = \{f \in L^2[-1,1] : f(-x) = f(x)\}$ and $M_o = \{f \in L^2[-1,1] : f(-x) = -f(x)\}$ of even and odd functions. We will show that $M_e^\perp = M_o$. If $f \in M_e$ and $g \in M_o$, then

$$\langle f, g \rangle = \int_{-1}^{1} f(t)\overline{g(t)}\, dt = 0,$$

since the integrand $f\overline{g}$ is an odd function and we integrate from -1 to 1. This shows that $M_o \subset M_e^\perp$. To show that $M_e^\perp \subset M_o$, recall from an earlier example that any $f \in L^2[-1,1]$ may we written as a sum

$$f = f_e + f_o,$$

where $f_e \in M_e$ and $f_o \in M_o$. Now assume that $f \in M_e^\perp$. Since $f_e \in M_e$, this implies that

$$0 = \langle f, f_e \rangle = \langle f_e + f_o, f_e \rangle = \langle f_e, f_e \rangle = \|f_e\|^2,$$

hence $f_e = 0$. Thus $f = f_e + f_o = f_o$, and we see that $f \in M_o$. This proves that $M_e^\perp \subset M_o$, so $M_e^\perp = M_o$.
It is clear from the proof of the previous proposition that the associated projection $P : L^2[-1,1] \to M$ is given by

$$Pf = f_e,$$

where $f = f_e + f_o$ is the unique decomposition of f as the sum of an odd function f_o and an even function f_e.

We state some consequences of the projection theorem. In the mathematics literature the tensor product notation $\xi \otimes \xi$ is used to refer to P_ξ.

PROPOSITION 5.0.41. *Let X be a Hilbert space.*

(1) *For any closed subspace M of X we have $M^{\perp\perp} = M$.*
(2) *For any set A in X we have $A^{\perp\perp} = \overline{\mathrm{span}(A)}$.*

PROOF. (1) For any $x \in M$ we have $\langle x, y\rangle = 0$ for every $y \in M^\perp$. In other words, x is orthogonal to $M^\perp$, so $x \in (M^\perp)^\perp$.
Conversely, suppose that $x \in M^{\perp\perp}$. Since M is closed, we can decompose $x = y + z$ with $y \in M$ and $z \in M^\perp$. Since $x \in M^{\perp\perp}$ we have $\langle x, z\rangle = 0$. Furthermore, we have $y \in M \subseteq M^{\perp\perp}$, so we also have $\langle y, z\rangle = 0$. Consequently, $\|z\|^2 = \langle z, z\rangle = \langle x - y, z\rangle = \langle x, z\rangle - \langle y, z\rangle = 0$. Hence $z = 0$ and we have deduced that $x \in M$.

(2) For a general set A in X we note that $\overline{\operatorname{span}(A)}$ is the smallest closed subspace containing A. We set $M = \operatorname{span}(A)$. Then we have $M \subset \overline{M}$ and thus $\overline{M}^\perp \subseteq M^\perp$. Consequently, $M^{\perp\perp} \subseteq \overline{M}^{\perp\perp}$. But $\overline{M}$ is closed in X so $\overline{M}^{\perp\perp} = M^{\perp\perp}$. Since $\overline{M}^{\perp\perp} = M^{\perp\perp}$ we get that $M^{\perp\perp} \subseteq \overline{M}^{\perp\perp}$. Finally, $M \subseteq M^{\perp\perp}$ and $M^{\perp\perp}$ closed implies $\overline{M} \subseteq M^{\perp\perp}$, which completes the argument.

□

COROLLARY 5.0.42. *A subset A in a Hilbert space X is dense if and only if $A^\perp = \{0\}$. Moreover, $A^\perp = \{0\}$ is equivalent to x orthogonal to A and hence $x = 0$. In words, $\overline{\operatorname{span}(A)} = X$ if and only if the only element orthogonal to every element in A is the zero vector.*

PROOF. Suppose $\overline{\operatorname{span}(A)} = X$. Then A is a closed linear subspace and hence $A^\perp = A^{\perp\perp\perp} = X^\perp = 0$.
Conversely, $\overline{\operatorname{span}(A)} = A^{\perp\perp} = 0^\perp = X$. □

Many interesting theorems in analysis are about the identification of the dual spaces of normed spaces. A topic one is at the heart of functional analysis. Here we restrict our focus to the Hilbert space setting since its proof relies on the projection theorem.

Recall that the dual space X' of a normed space X is the space of bounded operators from X to $\mathbb{C}$.

LEMMA 5.5. *For $\varphi \in X'$ we have that $\ker(\varphi)$ is a closed subspace of X.*

PROOF. Let (x_n) be a sequence in $\ker(\varphi)$ converging to $x \in X$. Then $\varphi(x_n) = 0$ for all n and so $|\varphi(x_n) - \varphi(x)| \leq \|\varphi\|\|x - x_n\|$. Thus we have $\varphi(x) = 0$. □

THEOREM 5.6 (Riesz representation theorem). *Let X be a Hilbert space. For each $\xi \in X$ define $\varphi_\xi(x) = \langle x, \xi\rangle$. Then $\varphi_\xi \in X'$ is a bounded linear functional on X.*
Furthermore, every $\varphi \in X'$ is of the form φ_ξ for some unique $\xi \in X$.

The final assertion of the theorem is the subtle part and is due to F. Riesz.

PROOF. The Cauchy-Schwarz inequality gives $|\varphi_\xi(x)| \leq \|x\|\|\xi\|$ and thus $\varphi_\xi \in X'$.
Converse statement: For any $x, z \in X$ and a non-zero $\varphi \in X'$. Then $\varphi(x)z - \varphi(z)x \in \ker(\varphi)$.
Let us pick z in $\ker(\varphi)^\perp$, which we can do by the projection theorem, to get

$$0 = \langle z, \varphi(x)z - \varphi(z)x\rangle = \varphi(x)\|z\|^2 - \varphi(z)\langle x, z\rangle.$$

Hence,

$$\varphi(x) = \frac{\varphi(z)}{\|z\|^2}\langle x, z\rangle.$$

We set $\xi = \frac{\overline{\varphi(z)}}{\|z\|^2} z$. Then we have $\varphi(x) = \langle x, \xi\rangle$.

Since $\xi \to \varphi_\xi$ preserves sums and differences we have that $\|\varphi\|$ obeys the parallelogram law. Hence the theorem of Jordan-von Neumann implies that X' is a Hilbert space.
Uniqueness: Suppose $\tilde{\xi}$ is another representation of φ of the form $\varphi_{\tilde{\xi}}$. Then $\langle x, \xi - \tilde{\xi}\rangle = \langle x, \xi\rangle - \langle x, \tilde{\xi}\rangle = 0$ and $\xi = \tilde{\xi}$. □

The theorem yields that any bounded linear functional φ on ℓ^2 is of the form

$$\varphi(x) = \sum_{n=1}^{\infty} x_i \xi_i \quad \text{for a unique } \xi \in \ell^2.$$

A different description of operators is one consequence of Riesz' theorem, because it implies the existence of the adjoint of an operator.

LEMMA 5.7. *Suppose $T \in B(X)$, X a Hilbert space, and $x, x' \in X$.*

(1) *If $\langle x, y\rangle = \langle x', y\rangle$ for all $y \in X$, then we have $x = x'$.*
(2) $\|T\| = \sup\{\|Tx\| = \sup\{|\langle Tx, y\rangle| : x, y \in X \quad with \|x\|, \|y\| \leq 1\}$.

For motivation of the general result we indicate the main idea for linear operators T on $\mathbb{C}^2$. We represent T with respect to the standard basis of $\mathbb{C}^2$, so $T = Ax$ for a matrix $A = (a_{ij})$. We look for a matrix $B = (b_{ij})$ such that

$$\langle Ax, y\rangle = \langle x, By\rangle$$

for all $x, y \in \mathbb{C}^2$. Concretely, we have

$$\langle \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}, y\rangle = \langle x, \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}\rangle$$

and so

$$\langle \begin{pmatrix} a_{11}x_1 + a_{12}x_2 \\ a_{21}x_1 + a_{22}x_2 \end{pmatrix}, y\rangle = \langle x, \begin{pmatrix} b_{11}y_1 + b_{12}y_2 \\ b_{21}y_1 + b_{22}y_2 \end{pmatrix}\rangle$$

The equation is equivalent to

$$a_{11}x_1\overline{y_1} + a_{12}x_2\overline{y_1} + a_{21}x_1\overline{y_2} + a_{22}x_2\overline{y_2} = \\ = x_2\overline{b_{11}}\overline{y_1} + x_1\overline{b_{12}}\overline{y_2} + x_2\overline{b_{21}}\overline{y_1} + x_2\overline{b_{22}}\overline{y_2}$$

to hold for all $x_1, x_2, y_1, y_2 \in \mathbb{C}$. Hence we deduce that

$$a_{11} = \overline{b_{11}},\ a_{12} = \overline{b_{21}},\ a_{21} = \overline{b_{12}}, a_{22} = \overline{b_{22}}.$$

Thus

$$B = \begin{pmatrix} \overline{a_{11}} & \overline{a_{21}} \\ \overline{a_{21}} & \overline{a_{22}} \end{pmatrix}$$

is the *conjugate-transpose* of A. The adjoint of T, denoted by T^*, is in this way linked to the original transform.

THEOREM 5.8 (Adjoint). *Let T be a bounded operator on a Hilbert space X. Then there exists a unique operator $T^* \in \mathcal{B}(X)$ such that*

$$\langle Tx, y\rangle = \langle x, T^*y\rangle \quad \textit{for all } x, y \in X.$$

The operator T^* is called the adjoint of T.

PROOF. Fix $y \in X$ and let $\varphi : X \to \mathbb{C}$ be defined by $\varphi(x) = \langle Tx, y\rangle$. Then φ is linear and by Cauchy-Schwarz it is bounded:

$$|\varphi(x)| \leq |\langle Tx, y\rangle| \leq \|Tx\|\|y\| \leq \|T\|\|x\|\|y\|.$$

Hence φ is a bounded linear functional on X and so by the Riesz representation theorem there exists a unique $\xi \in X$ such that $\varphi(x) = \langle x, \xi\rangle$ for all $x \in X$.
The vector ξ depends on the vector $y \in X$. In order to keep track of this fact we set $T^*y := \xi$. Hence we have defined an operator T^* from X to X based on the structure of bounded linear functionals on X. In summary, we have demonstrated the existence of an operator T^* on X such that

$$\langle Tx, y\rangle = \langle x, T^*y\rangle \quad \text{for all } x, y \in X.$$

(1) T^* is linear.

$$\begin{aligned}\langle x, T^*(\lambda y_1 + \mu y_2)\rangle &= \langle Tx, \lambda y_1 + \mu y_2\rangle \\ &= \overline{\lambda}\langle Tx, y_1\rangle + \overline{\mu}\langle Tx, y_2\rangle \\ &= \overline{\lambda}\langle x, T^*y_1\rangle + \overline{\mu}\langle x, T^*y_2\rangle \\ &= \langle x, \lambda T^*y_1\rangle + \mu T^*y_2\rangle.\end{aligned}$$

(2) T^* is bounded. We use the Cauchy-Schwarz inequality:

$$\begin{aligned}\|T^*y\|^2 = \langle T^*y, T^*y\rangle &= \langle TT^*y, y\rangle \\ &\leq \|TT^*y\|\|y\| \\ &\leq \|T\|\|T^*y\|\|y\|.\end{aligned}$$

Hence we have shown

$$\|T^*y\|^2 \leq \|T\|\|T^*y\|\|y\|$$

If $\|T^*y\| > 0$, then we can through and obtain the desired result: $\|T^*y\| \leq \|T\|\|y\|$. Suppose $\|T^*y\| = 0$. Then the desired inequality holds, too. Consequently, we have proved that

$$\|T^*\| \leq \|T\|.$$

(3) T^* is unique. Suppose there exists another $S \in B(X)$ such that $\langle Tx, y\rangle = \langle x, Sy\rangle$ for all $x, y \in X$. Then we have

$$\langle x, Sy\rangle = \langle x, T^*y\rangle \; y \in X$$

and by a well-known fact about innerproducts we deduce that $T^*y = Sy$ for all $y \in Y$. Hence T^* is unique.

□

We collect a few properties of the adjoint.

LEMMA 5.9. *Let S, T be in $B(X)$ and $\lambda, \mu \in \mathbb{C}$.*

(1) $(\lambda S + \mu T)^* = \overline{\lambda}S^* + \overline{\mu}T^*$;
(2) $(ST^*) = T^*S^*$.

(3) *If T is invertible, then T^* is also invertible and $(T^*)^{-1} = (T^{-1})^*$.*

PROOF. The proofs of (i) and (iii) are left as an exercise. Here we show the second assertion:

$$\langle x, (ST)^* y\rangle = \langle STx, y\rangle = \langle Tx, S^* y\rangle = \langle x, T^* S^* y\rangle$$

holds for all $x \in X$ and so we have $(ST^*) = T^* S^*$. □

We continue with some useful facts about T^*.

LEMMA 5.10. *Let T be a bounded operator on a Hilbert space X.*

(1) $(T^*)^* = T$;
(2) $\|T^*\| = \|T\|$;
(3) $\|T^*T\| = \|T\|^2$ *(C^*-algebra identity)*

PROOF. (1) For $x, y \in X$ we have

$$\begin{aligned}\langle y, (T^*)^* x\rangle &= \langle T^* y, x\rangle \\ &= \overline{\langle x, T^* y\rangle} \\ &= \overline{\langle Tx, y\rangle} \\ &= \langle y, Tx\rangle,\end{aligned}$$

so $(T^*)^* x = Tx$ for all $x \in X$.

(2) In the proof of the existence of the adjoint we established that $\|T^*\| \leq \|T\|$. Applying this result to T^{**} and using (i) yields $\|T\| \leq \|T^*\|$. Hence we have $\|T^*\| = \|T\|$.

(3) By (ii) we have $\|T^*\| = \|T\|$ that implies

$$\|T^*T\| \leq \|T^*\|\|T\| = \|T\|^2.$$

For the reverse inequality we use

$$\begin{aligned}\|Tx\|^2 &= \langle Tx, Tx\rangle \\ &= \langle T^*Tx, x\rangle \\ &\leq \|T^*Tx\|\|x\| \\ &\leq \|T^*T\|\|x\|^2\end{aligned}$$

to deduce $\|T\|^2 \leq \|T^*T\|$.

□

Some examples should help to build up some intuition on adjoint operators.

EXAMPLE 5.0.43. We investigate some operators onℓ^2 and $L^2[0,1]$.

(1) The adjoint of $Lx = (0, x_1, x_2, ...)$ on ℓ^2 is the right shift operator $Rx = (x_2, x_3, ...)$.

By definition

$$\langle (0, x_1, x_2, ...), (y_1, y_2, ...)\rangle = \langle x, L^* y\rangle$$

for all $x, y \in \ell^2$. We denote L^*y by $z = (z_n)$ Therefore we have

$$x_1\overline{y_2} + x_2\overline{y_3} + \cdots = x_1\overline{z_1} + x_2\overline{z_2} + \cdots .$$

This equation is true for all x_i if $z_1 = y_2, z_2 = y_3, \ldots$. Hence by the uniqueness of the adjoint

$$L^* y = (y_2, y_3, \ldots),$$

i.e. $L^* = R$.

(2) The adjoint of the multiplication operator T_a for $a \in \ell^\infty$ is the multiplication operator for the sequence $\overline{a}$.

$$\langle T_a x, y\rangle = \langle x, T_a^* y\rangle$$

Hence

$$a_1 x_1 \overline{y_1} + a_2 x_2 \overline{y_2} + \cdots = x_1 \overline{\overline{a_1} y_1} + x_2 \overline{\overline{a_2} y_2} + \cdots,$$

which by the uniqueness of the adjoint gives that $T_{\overline{a}}$ is the adjoint of T_a.

(3) The multiplication operator T_a on $L^2[0,1]$ defined by $a \in C[0,1]$ has $T_{\overline{a}}$ as its adjoint.

$$\langle T_a f, g\rangle = \int_0^1 a(t) f(t) \overline{g(t)} dt = \overline{\int_0^1 f(t) \overline{a(t)} g(t) dt} = \langle f, T_{\overline{a}} g\rangle.$$

We introduce some classes of operators defined in terms of the adjoint.

Definition 5.0.44. Let T be a bounded operator on a Hilbert space X.

(1) T is called *normal* if $T^*T = TT^*$.
(2) T is called *unitary* if $T^*T = TT^* = I$.
(3) T is called *selfadjoint* if $T = T^*$.

Examples 5.0.45 (Operators on ℓ^2).

(1) The multiplication operator T_a for $a \in \ell^\infty$ is normal, since $T_a^* T_a = T_a^* T_a = T_{|a|^2}$. Hence it is unitary if $|a| = 1$ as in the example $(1, i, -1, -i, \ldots) = (-i^k)_{k=0}^\infty$. T_a is selfadjoint if and only if a is real-valued.

(2) The shift operator is not normal: $L^*L = I$ and $LL^* y = (0, y_2, y_3, \ldots) \neq I$. Hence L is not unitary.

We state a few properties of unitary operators. We denote the set of all unitary operators on X by $\mathcal{U}$

Lemma 5.11. *For S, T in $\mathcal{U}$ we have that ST and TS are also in $\mathcal{U}$. The identity operator is a unitary operator. Unitary operators are invertible and $T^{-1} = T^*$.*

Proof. Since $(ST)^*(ST) = T^* S^* S T^*$ we get from $S^*S = I$ and $T^*T = I$ that ST is also unitary. The invertibility follows from the definition of unitary operators. □

In some problems it is of interest to have control over linear operators that preserve the norm, known as isometries.

Definition 5.0.46. Let X be a normed space. Then $T \in B(X)$ is called an *isometry* if $\|Tx\| = \|x\|$ for all $x \in X$.

In one of the problem sets we have already settled that isometries are injective. We settle the structure of isometries for Hilbert spaces.

Proposition 5.0.47. *Let T be a bounded operator on a Hilbert space X.*

(1) *T is an isometry of X if and only if $T^*T = I$.*
(2) *If T is unitary then T is an isometry of X.*

(3) *T is a surjective isometry if and only if T is unitary.*

PROOF. (1) Suppose that $T^*T = I$. Then

$$\|Tx\|^2 = \langle Tx, Tx\rangle = \langle T^*Tx, x\rangle = \langle Ix, x\rangle = \|x\|^2,$$

so T is an isometry.
Conversely, suppose that T is an isometry. Then

$$\langle T^*Tx, x\rangle = \langle Tx, Tx\rangle = \|Tx\|^2 = \|x\|^2 = \langle Ix, x\rangle.$$

Hence $T^*T = I$.
(2) Suppose that T is unitary. By (i) T is an isometry.
(3)

□

EXAMPLE 5.0.48. The shift operator $Rx = (0, x_1, x_2, ...)$ is an isometry on ℓ^2, but it is not a unitary operator due to its lack of surjectivity.

EXAMPLE 5.0.49. Let U be a linear transformation on a finite-dimensional innerproduct space X. Consider U as a matrix relative to an orthonormal basis on X. Show that the following statements are equivalent.

(1) U is unitary, i.e. $U^*U = I = UU^*$.
(2) The columns of U are an orthonormal basis of X.
(3) The rows of U are an orthonormal basis of X.

We close our discussion of the adjoint, with some relations between the kernel and range of an operator and its adjoint. These statements are of utmost importance.

PROPOSITION 5.0.50. *Let T be a bounded operator on a Hilbert space X.*

(1) $\overline{\text{ran}(T)} = (\ker(T^*))^\perp$.
(2) $\ker(T) = (\text{ran}(T^*))^\perp$;

Equivalently,

$$\overline{\text{ran}(T)} = (\ker(T^*))^\perp, \quad \ker(T) = (\text{ran}(T^*))^\perp$$

and consequently:

$$X = \ker(T) \oplus \overline{\text{ran}(T^*)}.$$

PROOF. (1) $\overline{\text{ran}(T)} \subseteq (\ker(T^*))^\perp$: Let $z \in \ker(T^*)$ and let $y \in \text{ran}(T)$, i.e. there exists a $x \in X$ such that $y = Tx$. Hence

$$\langle y, z\rangle = \langle Tx, z\rangle = \langle x, T^*z\rangle = 0$$

and we have shown that $\text{ran}(T) \subseteq (\ker(T^*))^\perp$. Since $(\ker(T^*))^\perp$ is closed, we have that $\overline{\text{ran}(T)} \subseteq (\ker(T^*))^\perp$.
$(\text{ran}(T^*))^\perp \subseteq \ker(T)$: For $x \in \text{ran}(T))^\perp$ we have for all $y \in X$:

$$0 = \langle Ty, x\rangle = \langle y, T^*x\rangle,$$

that gives $T^*x = 0$ and thus $\text{ran}(T)^\perp \subseteq \ker(T^*)$. By taking the orthogonal complement of this relation, we get $\ker(T^*)^\perp \subseteq \text{ran}(T)^{\perp\perp} = \overline{\text{ran}(T)}$.
(2) Apply (i) to T^*.

For the equivalent formulation note, that we have as above $\text{ran}(T) = (\ker(T^*))^\perp$, but since $(\ker(T^*))^\perp$ is closed we also get $\overline{\text{ran}(T)} \subseteq (\ker(T^*))^\perp$. The rest of the argument follows similar lines as before. □

COROLLARY 5.0.51. *Let T be a bounded operator on a Hilbert space X. Then* $\ker(T^*) = \{0\}$ *if and only if* $\operatorname{ran}(T)$ *is dense in* X.

PROOF. Assume that $\ker(T^*) = \{0\}$. Then

$$\ker(T^*)^\perp = \{0\}^\perp = X$$

and the assertion (ii) of the proposition implies that

$$\ker(T^*)^\perp = (\operatorname{ran}(T))^{\perp\perp} = \operatorname{ran}(T).$$

Thus we have $\operatorname{ran}(T)$ is dense in X.

Suppose $\operatorname{ran}(T)$ is dense in X. Then by $(\operatorname{ran}(T))^{\perp\perp} = \overline{\operatorname{ran}(T)} = X$ and

$$\ker(T^*) = \operatorname{ran}(T)^\perp = ((\operatorname{ran}(T))^{\perp\perp})^\perp = X^\perp = \{0\}.$$

□

The corollary allows one to check if the range of an operator is dense in a Hilbert space by determining its adjoint and the computation of the kernel of the adjoint. In general, this is a good strategy, because it is very difficult to compute the range of an operator. Another important application of the preceding theorem is the Fredholm alternative.

THEOREM 5.12 (Fredholm alternative). *Suppose T is a bounded linear operator on a Hilbert space X with closed range. Then the equation*

$$Tx = b \quad , b \in X$$

has a solution x in X for every $b \in X$ if and only if

$$b \in (\ker(T^*))^\perp.$$

Hence operators with a closed range have a general criterion of existence. For example if $T \in \mathcal{B}(X)$ satisfies for all $x \in X$ and estimate of the form

$$\|Tx\| \geq c\|x\| \quad \text{for some } c > 0.$$

EXAMPLE 5.0.52. The range of the right shift operator R on ℓ^2 is closed since if consists of $\{(0, x_2, x_3, ...) : x_i \in \mathbb{C}\}$. The left shift is L not invertible since its kernel is one-dimensional and spanned by $(1, 0, 0, ...)$.
The equation

$$Rx = b \Leftrightarrow (0, x_1, x_2, ...) = (b_1, b_2, ...)$$

is solvable if and only if $b_1 = 0$, or $b \in (\ker(L))^\perp$.

On the other hand

$$Lx = b$$

is solvable for all $b \in \ell^2$ despite of L not being injective.

CHAPTER 6

Series and bases in normed spaces

In this chapter we investigate series and bases in normed spaces and in particular, we focus on Schauder bases for Banach spaces, the geometric series for bounded operators, and orthonormal bases for separable Hilbert spaces.

6.1. Schauder bases and series of operators

First some general facts about bases in vector spaces.

DEFINITION 6.1.1. A set $A = \{x_n : n \in \mathbb{N}\}$ is called *linearly independent* if we have

$$\sum_{i=1}^{n} \alpha_i x_i = 0 \Rightarrow \alpha_1 = \cdots \alpha_n = 0$$

for all $n \in \mathbb{N}, x_i \in A$ and $\alpha_i \in \mathbb{F}$ for $i = 1, ..., n$.

For example, we have that $\{1, x, ..., x^n\}$ is a linearly independent set of $\mathcal{P}_n$, the space of polynomials of degree at most n and that $\{x^n : n = 0, 1, 2, ...\}$ is a linearly independent set of the space of all polynomials $\mathcal{P}$. These two linearly independent sets have an additional property, namely they span the respective spaces.

DEFINITION 6.1.2. We call a linearly independent set $\mathcal{B}$ of a vector space X a *basis* if $\mathcal{B}$ spans X, i.e. if for any $x \in X$ there exist unique scalars $\alpha_1, ..., \alpha_n$ such that

$$x = \alpha_1 x_1 + \cdots + \alpha_n x_n.$$

If the basis consists of finitely many elements, then X is called *finite dimensional*. Otherwise, we call X *infinite dimensional*.

The aforementioned bases for a vector space are also known as Hamel bases, named after the German mathematician G. Hamel. The axiom of choice implies that any vector space has a (Hamel) basis. For most interesting cases are these bases uncountable and of little practical use. We are interested in substitutes of the notion of a basis for normed spaces. There is a variety of notions and we are just treating, so-called Schauder basis.

DEFINITION 6.1.3. A countable set $\mathcal{B}$ of a normed space X is called a *Schauder basis* if for any $x \in X$ there exists a unique sequence of scalars $(\alpha_n)_{n\in\mathbb{N}}$ such that for any $\varepsilon > 0$ there exists an $N \in \mathbb{N}$ such that

$$\|x - \sum_{n=1}^{N} \alpha_n x_n\| < \varepsilon \qquad \text{for } n \geq N,$$

and we write in this case

$$x = \sum_{n=1}^{\infty} \alpha_n x_n.$$

In other words, a Schauder basis allows to take infinite linear combinations to express elements in a normed space, since the norm on the vector spaces gives us a way to define a limiting process. Note that rational linear combinations of the basis elements of a Schauder basis span the normed space and thus normed spaces with a Schauder basis are separable.

DEFINITION 6.1.4. Let X be a normed space and (x_n) a sequence of vectors in X. Then we define the *series* $\sum_{n=1}^{\infty} x_n$ as the sequence of partial sums $(\sum_{k=1}^{n} x_k)_{n\in\mathbb{N}}$ converging with respect to the norm in X, i.e. $\sum_{n=1}^{\infty} x_n$ is the element $s \in X$ such that

$$\|\sum_{k=1}^{n} x_k - s\| \to 0 \qquad \text{as } n \to \infty.$$

Many of the Banach spaces that are of interest in applications, have a Schauder basis, such as the space of continuous functions on a bounded interval etc. We just describe a Schauder basis for the sequence spaces ℓ^p for $p \in [1, \infty)$.

PROPOSITION 6.1.5. *We denote by* $\{e_n : n \in \mathbb{N}\}$ *the standard basis, where* e_n *is the sequence that has a 1 at the nth component and is zero otherwise. Then* $\{e_n : n \in \mathbb{N}\}$ *is a Schauder basis for* ℓ^p *for* $p \in [1, \infty)$.

PROOF. The proof is part of one of the problems one the new problem set. □

Since ℓ^∞ is not separable, the standard basis $\{e_n : n \in \mathbb{N}\}$ for ℓ^∞. We are still interested in finding out under what conditions the series $\sum_{h\in\mathbb{N}} \alpha_n e_n$ converges in ℓ^∞. By the definition of a series one gets that this is precisely the case when the sequence (α_n) converges to zero. Filling in the details for this claim is also part of the new problem set.

Suppose we have a polynomial $p(x) = a_0 + a_1 x + \cdots + a_n x^n$. Then we associate to p the operator

$$p(T) = a_0\, I + a_1\, T + \cdots + a_n\, T^n$$

for any operator $T \in \mathcal{B}(X)$. If one thinks of taking powers of a number is the same as taking compositions of T, then this association between $p(x)$ and $p(T)$ is an instance of a so-called functional calculus and if we define $T^0 := I$, then $p(T) = \sum_{k=0}^{n} a_k T^k$.

Let us try to see how one could try to make sense of this procedure for power series. We will at the moment just focus on the one for the geometric series: We know that $\sum_{k=0}^{\infty} x^k$ converges for $x \in \mathbb{C}$ if and only if $|x| < 1$ and in the case of convergence the series sums to $1/(1-x)$. Based on our experience with convergent sequences in normed spaces, we expect that for bounded operators with norm less than one, the geometric series might make sense.

PROPOSITION 6.1.6 (Neumann series). *Let* X *be a normed space. For* $T \in \mathcal{B}(X)$ *with* $\|T\| < 1$ *the geometric series* $\sum_{k=0}^{\infty} T^k$ *exists and equals to*

$$1 + T + T^2 + \cdots = (1 - T)^{-1}.$$

This way of computing the inverse of $1 - T$ for operators that are sufficiently close to the identity operator, is named after the German mathematician F. Neumann, who used this relation for solving integral equations.

PROOF. Note that $\sum_{k=0}^{\infty} \|T^k\|$ converges for $\|T\| < 1$ since $\|T^k\| \leq \|T\|^k$. Consequently, the partial sums $S_n = \sum_{k=0}^{n} T^k$ form a Cauchy sequence (S_n) in $\mathcal{B}(X)$. Hence there exists a $S \in \mathcal{B}(X)$ such that $\|S - S_n\| \to 0$ as $n \to \infty$.

Note that $1 - S$ and S_n commute for each $n \in \mathbb{N}$ and that

$$(1 - T)S_n = S_n(1 - T) = 1 - T^{n+1}.$$

We have as n goes to infinity that $(1-T)S = I$ and $S(1-T) = I$, since $\|T\|^{n+1} \to 0$ as $n \to \infty$. Hence $S = \sum_{k=0}^{\infty} T^k$ is the inverse of $(1 - T)$, as claimed. □

Neumann series are an important tool in applications and many algorithms are based on this elementary fact. We close the discussion of series of operators with a consequence of Neumann series on the structure of the set (actually group) of invertible bounded operators on a normed space X, we denote this set by $G(X)$.

PROPOSITION 6.1.7. *Let X be a normed space. Then $G(X)$ is open in $\mathcal{B}(X)$, the space of bounded linear operators on X with respect to the operator norm.*

PROOF. For $S \in G(X)$ and $T \in \mathcal{B}(X)$ we have

$$T = S - (S - T) = S(I - S^{-1}(S - T))$$

and hence $T \in G(X)$ if $I - S^{-1}(S - T)$ is invertible. Our result on Neumann series yields that $I - S^{-1}(S - T)$ is invertible if $\|S^{-1}(S - T)\| < 1$, i.e. if $\|S - T\| < 1/\|S^{-1}\|$. Consequently, we have shown that the open ball

$$B := \{T \in \mathcal{B}(X) : \|S - T\| < 1/\|S^{-1}\|\}$$

around the invertible operator S consists of invertible operators. Hence $G(X)$ is open. □

6.2. Separable Hilbert spaces

We restrict our discussion to Hilbert spaces that contain a countable dense subset, i.e. separable Hilbert spaces, for two reasons: (i) The general case is way more technically involved, and (ii) most of the time one just has to deal with separable Hilbert spaces. There is one prominent example of a non-separable Hilbert space, the space of almost periodic functions. Since we are not going to discuss this function space in this course, we have decided to keep non-separable Hilbert spaces for another occasion.

DEFINITION 6.2.1. A set A of an innerproduct space is called *orthonormal* if for any two distinct elements $x, y \in A$ we have $\langle x, y\rangle = 0$ and $\|x\| = 1$ for any $x \in X$.

We have encountered orthonormal sets earlier in this course. Let $L^2[0, 2\pi]$ be the completion of $C[0, 2\pi]$ with respect to $\|f\| = (\int_0^{2\pi})|f(x)|^2 dx)^{1/2}$. Then the set of exponentials $\{e^{inx} : n \in \mathbb{Z}\}$ is an orthonormal sequence in $L^2[0, 2\pi]$.

LEMMA 6.1. *Any orthonormal set in an innerproduct space is linearly independent.*

The proof is left as an exercise.

LEMMA 6.2. *Let A be a finte orthonormal set in an innerproduct space X, then*

$$\|\sum_{j=1}^{n}\alpha_j e_j\|^2 = \sum_{j=1}^{n}|\alpha_j|^2$$

for any scalars $\alpha_1, ..., \alpha_n$.

PROOF. The argument is just an elementary computation:

$$\|\sum_{j=1}^{n}\alpha_j e_j\|^2 = \langle\sum_{j=1}^{n}\alpha_j e_j, \sum_{j=1}^{n}\alpha_j e_j\rangle = \sum_{j=1}^{n}\sum_{k=1}^{n}\alpha_j\overline{\alpha_k}\langle e_j, e_k\rangle = \sum_{j=1}^{n}|\alpha_j|^2.$$

□

Any infinite-dimensional innerproduct space contains a countable orthonormal set. Since the vector space is infinite-dimensional, it has to contain an infinite set of linearly independent vectors. Now use any procedure that turns this set into an orthonormal one. The most popular orthogonalization method is the one by Gram and Schmidt, but it is by any means the only method out there.

PROPOSITION 6.2.2. *Let X be an infinite-dimensional innerproduct space. Then X contains a countable orthonormal set.*

PROOF. By assumption there exists a linearly independent subset $\{x_1, x_2, ...\}$ in X. We show that there exists an orthonormal set $\{e_1, e_2, ...\}$ such that $\text{span}\{x_1, ..., x_n\} = \text{span}\{e_1, ..., e_n\}$ for all $n \in \mathbb{N}$.

The argument is based on an extension of the Gram-Schmidt algorithm to infinite-dimensional sets. Set $e_1 := x_1/\|x_1\|$. Then we have $\text{span}(x_1) = \text{span}(e_1)$. We continue by induction. Suppose that for some $n \geq 2$ we have constructed an orthonormal set $E_{n-1} = \{e_1, ..., e_{n-1}\}$ such that $\text{span}\{x_1, ..., x_{n-1}\} = \text{span}\{e_1, ..., e_{n-1}\}$. Then we project x_n onto E_{n-1} and subtract this from x_n:

$$\tilde{e}_n := x_N - \sum_{k=1}^{n-1}\langle x_n, e_k\rangle\, e_k.$$

Since $\sum_{k=1}^{n-1}\langle x_n, e_k\rangle\, e_k \in E_{n-1}$ it also lies in $\text{span}\{x_1, ..., x_{n-1}\}$ and thus $\{x_1, ..., x_{n-1}, x_n\}$ is linearly independent, which yields that $\tilde{e}_n$ is non-zero. By construction, $\langle\tilde{e}_n, e_k\rangle = 0$ for $j < k$. We normalize $\tilde{e}_n$ and add this vector to E_{n-1}, in order to get our E_n. Note that $E_n = \text{span}\{x_1, ..., x_n\}$. Hence in this way we have constructed a countable orthonormal set in X. □

Suppose $\{e_n : n \in \mathbb{N}\}$ is an orthonormal sequence in a Hilbert space. We study for which sequences the series $\sum_k \alpha_k e_k$ exists.

PROPOSITION 6.2.3. *Let $\{e_n : n \in \mathbb{N}\}$ be an orthonormal sequence in a Hilbert space X. Then the series $\sum_{k=1}^{\infty}\alpha_k e_k$ exists if and only if $(\alpha_k) \in \ell^2$, and $\|\sum_{k=1}^{\infty}\alpha_k e_k\| = \|(\alpha_k)\|_{\ell^2}$.*

PROOF. As shown above, we have that $\|\sum_{k=1}^{n}\alpha_k e_k\|^2 = \sum_{k=1}^{n}|\alpha_k|^2$. Hence the partial sums $(s_n = \sum_{k=1}^{n}\alpha_k e_k)$ satisfy for $n > m$

$$\|s_n - s_m\|^2 = \sum_{k=m+1}^{n}|\alpha_k|^2.$$

Hence (s_n) is a Cauchy sequence in X if and only if $(\|s_n\|^2)$ is a Cauchy sequence in $\mathbb{R}$. Since X and $\mathbb{R}$ are both complete, these two sequences converge or divergence simultaneously. In the case of convergence, we take the limit $n \to \infty$ and obtain the desired claim. □

Note that any x in the span of finitely many orthonormal vectors $\{e_1, ..., e_n\}$ may be uniquely written as

$$x = \sum_{k=1}^{n} \langle x, e_k \rangle \, e_k.$$

Since the vectors $e_1, ..., e_n$ are a basis for its span, we have that there exist some scalars $\alpha_1, ..., \alpha_n$ such that $x = \sum_{k=1}^{n} \alpha_k e_k$. Take the innerproduct with e_j yields $\langle x, e_j \rangle = \alpha_j$.

Our characterizations of series in Hilbert spaces gives us that the (generalized) *Fourier series*

$$\sum_{k=1}^{\infty} \langle x, e_k \rangle \, e_k$$

exists if and only if the sequence $(\langle x, e_k \rangle)$ is square-summable. These generalized Fourier series have some interesting properties. Consequently, we have that the closed span of an orthonormal sequence $\{e_n : n \in \mathbb{N}\}$ in a Hilbert space X is of the form

$$\overline{\text{span}}(\{e_n : n \in \mathbb{N}\}) = \{x \in X | \, x = \sum_{k=1}^{\infty} \alpha_k e_k \text{ for } (\alpha_k) \in \ell^2\}.$$

PROPOSITION 6.2.4. *Let $\{e_n : n \in \mathbb{N}\}$ be an orthonormal sequence in a Hilbert space. Then for any $x \in X$ the best approximation of x in $\overline{span}(\{e_n : n \in \mathbb{N}\})$ is given by*

$$Px = \sum_{k=1}^{\infty} \langle x, e_k \rangle \, e_k,$$

which also gives the projection of x onto $\overline{span}(\{e_n : n \in \mathbb{N}\})$. Furthermore, we have for any $n \in \mathbb{N}$ that

$$\|x - \sum_{k=1}^{n} \langle x, e_k \rangle \, e_k\|^2 = \|x\|^2 - \sum_{k=1}^{n} |\langle x, e_k \rangle|^2.$$

PROOF. Since for any x the sequence $(\langle x, e_n \rangle)$ is in ℓ^2 the series $\sum_{k=1}^{\infty} \langle x, e_k \rangle \, e_k$ exists and defines an element $\tilde{x} \in X$. We apply the characterization of best approximations in terms of orthogonal complements to deduce the claim. Note that $\langle \tilde{x} - x, e_n \rangle = 0$ for all $n \in \mathbb{N}$ and thus an $\tilde{x} - x \in \overline{\text{span}}(\{e_n : n \in \mathbb{N}\})^{\perp}$ and consequently $\tilde{x}$ is the best approximation of x in $\overline{\text{span}}(\{e_n : n \in \mathbb{N}\})$ and it also gives the orthogonal projection onto $\overline{\text{span}}(\{e_n : n \in \mathbb{N}\})$. □

The equation $\|x - \sum_{k=1}^{n} \langle x, e_k \rangle \, e_k\|^2 = \|x\|^2 - \sum_{k=1}^{n} |\langle x, e_k \rangle|^2$ yields that

$$\|x\|^2 \geq \sum_{k=1}^{n} |\langle x, e_k \rangle|^2.$$

Since this is true for all $n \in \mathbb{N}$ we may take the limit as n goes to ∞:

$$\sum_{k=1}^{\infty} |\langle x, e_k \rangle|^2 \leq \|x\|^2.$$

This inequality is known as *Bessel's inequality.*

PROPOSITION 6.2.5 (Bessel's inequality). *If $\{e_n : n \in \mathbb{N}\}$ is an orthonormal sequence in a Hilbert space C, then we have for any $x \in X$:*

$$\sum_{k=1}^{\infty} |\langle x, e_k \rangle|^2 \leq \|x\|^2.$$

Suppose $\{e_n : n \in \mathbb{N}\}$ is an orthonormal sequence in a Hilbert space X. We are interested when it is possible to extend it to an orthonormal set that spans all of X. This is exactly the case, when $\overline{\text{span}}(\{e_n : n \in \mathbb{N}\})^{\perp} = \{0\}$.

DEFINITION 6.2.6. An orthonormal sequence $\{e_n : n \in \mathbb{N}\}$ in a Hilbert space X is called *maximal* if

$$\overline{\text{span}}(\{e_n : n \in \mathbb{N}\})^{\perp} = \{0\}.$$

The classical example of a maximal orthonormal set in $L^2[0, 2\pi]$ is the exponentials $\{e^{inx} : n \in \mathbb{N}\}$. This follows from the Approximation Theorem of Weierstrass that the trigonometric polynomials are dense in $C[0, 2\pi]$ and by construction $C[0, 2\pi]$ is dense in $L^2[0, 2\pi]$ (which in the traditional approach to $L^2[0, 2\pi]$ via measure theory is known as Lusin's theorem).

DEFINITION 6.2.7. An orthonormal sequence $\{e_n : n \in \mathbb{N}\}$ in a Hilbert space X is called an *orthonormal basis* of X if

$$x = \sum_{k=0}^{\infty} \langle x, e_k \rangle e_k$$

holds for any $x \in X$.

By definition we have that an orthonormal sequence $\{e_n : n \in \mathbb{N}\}$ is maximal if and only if it is an orthonormal basis. We present a characterization of orthonormal bases in a Hilbert space X.

THEOREM 6.3 (Parseval's identity). *Let $\{e_n : n \in \mathbb{N}\}$ be an orthonormal sequence in a Hilbert space X. The following are equivalent:*

(1) *$\{e_n : n \in \mathbb{N}\}$ is an orthonormal basis of X.*
(2) *$\sum_{k=1}^{\infty} |\langle x, e_k \rangle|^2 = \|x\|^2$ for any $x \in X$.*

PROOF. Suppose (2) holds. Since $(\langle x, e_n \rangle)$ is in ℓ^2, the sequence of partial sums

$$(\sum_{k=1}^{n} \langle x, e_k \rangle e_k)$$

is a Cauchy sequence in X. Hence there exists a $\tilde{x} \in X$ such that $\tilde{x} = \sum_{k=1}^{\infty} \langle x, e_k \rangle e_k$. We want to show that $\tilde{x} = x$. Note that $x - \tilde{x}$ is orthogonal to $\tilde{x}$, because $\langle \tilde{x}, e_k \rangle = \tilde{x} e_k$ for any $k \in \mathbb{N}$. By the Pythagoras Theorem we get that $\|x\|^2 = \|x - \tilde{x}\|^2 + \|\tilde{x}\|^2$. By assumption we have $\|x\| = \|\tilde{x}\|$ and consequently $\|x - \tilde{x}\| = 0$ which implies that $x = \tilde{x}$. Hence we have established that (1) is true.

Suppose we have that $\{e_n : n \in \mathbb{N}\}$ is an orthonormal basis of X. Then any x is of the form

$$x = \sum_{k=1}^{\infty} \langle x, e_k \rangle e_k.$$

Take the innerproduct with x and you have (2). □

We summarize all these results in the following theorem.

THEOREM 6.4. *Let $\{e_n : n \in \mathbb{N}\}$ be an orthonormal sequence in a Hilbert space X. The following are equivalent:*

(1) *$\{e_n : n \in \mathbb{N}\}$ is an orthonormal basis of X.*
(2) *$\sum_{k=1}^{\infty} |\langle x, e_k \rangle|^2 = \|x\|^2$ for any $x \in X$.*
(3) *$x = \sum_{k=1}^{\infty} \langle x, e_k \rangle e_k$ holds for any $x \in X$.*
(4) *$\{e_n : n \in \mathbb{N}\}$ is a maximal orthonormal sequence of X, i.e. $\langle x, e_n \rangle = 0$ for all $n \in \mathbb{N}$ implies $x = 0$.*
(5) *The linear span of $\{e_n : n \in \mathbb{N}\}$ is dense in X.*

The existence of a series expansion with respect to an orthonormal basis for a separable Hilbert space implies that the elements of the Hilbert space are uniquely determined by its Fourier coefficients. Hence in some sense any separable Hilbert space looks like ℓ^2. Let us turn this observation into some rigorous statement.

DEFINITION 6.2.8. Two Banach spaces X and Y are called *isometrically isomorphic* if there exists a surjective isometry $T : X \to Y$.

Observe that if $T : X \to Y$ is a surjective isometry, then its inverse T^{-1} is a surjective isometry as well. If X and Y are Hilbert spaces, this implies that the surjective isometry is actually a unitary operator.

THEOREM 6.5 (Riesz-Fischer). *Every infinite dimensional separable Hilbert space is isometrically isomorphic to ℓ^2.*

PROOF. Since any infinite dimensional separable Hilbert space has a countable orthonormal basis $\{e_n : n \in \mathbb{N}\}$ and we can uniquely express every $x \in X$ as $x = \sum_{k=1}^{\infty} \langle x, e_k \rangle e_k$. We define a map $T : X \to \ell^2$

$$Tx = (\langle x, e_n \rangle)_{n \in \mathbb{N}}.$$

By Parseval's identity we have $\|Tx\| = \|x\|$ for any $x \in X$ and T is a surjective isometry. □

Note that this isomorphism theorem is due to the choice of an orthonormal basis for X and that the uniqueness of the coefficients and the special nature of orthonormal series.

CHAPTER 7

Some topics in linear algebra

We review some facts about spanning sets, basis and linear transformations in finite-dimensional vector spaces.

7.1. Spanning sets and bases

Recall that a set of vectors $\{x_1, ..., x_n\} \subset X$ is *linearly independent* if for all $\alpha_1, ..., \alpha$ the equation

$$\alpha_1 x_1 + \cdots + \alpha_n x_n = 0$$

has only $\alpha_1 = \cdots = \alpha_n = 0$ as solution. If there exists a non-trivial linear combination of the x_i's, then we call the $\{x_1, ..., x_n\}$ *linearly dependent*.

Here are a few elementary observations: $\{x_1, ..., x_n\} \subset X$ is linearly independent if and only if every $x \in \text{span}\{x_1, ..., x_n\}$ can be written uniquely as a linear combination of elements of $\{x_1, ..., x_n\}$.

There are two central notions in the theory of vector spaces:

DEFINITION 7.1.1. Let X be a vector space.

(1) If there exists a set $S \subseteq X$ with $\text{span}(S) = X$, then we call S a *spanning set*. In case that S consists of finitely many elements $\{x_1, ..., x_n\}$, then we say that X is *finite-dimensional*. Finally, if there exists no finite spanning set for X, then we call the vector space *infinite-dimensional*.
(2) If there exists a linearly independent spanning set B for X, then we call B a basis for X.

We revisit some vector spaces from this point of view.

EXAMPLE 7.1.2. (1) The space of polynomials of degree at most n is finite-dimensional, because the set of monomials $\{1, x, x^2, ..., x^n\}$ is a spanning set and even a basis for $\mathcal{P}_n$.
(2) The space of all polynomials $\mathcal{P}$ is infinite dimensional.

Let us present the argument for this fact. We have to show that for any n there is only just the trivial linear combination of monomials $\{x_0(t) = 1, x_1(t) = t^2, ..., x_n(t) = t^n\}$ that represents the zero function. We use induction: For $n = 0$ we have $\alpha_0 = 0$ if and only if $\alpha = 0$.
Suppose for n we know that

$$\alpha_0 x_0(t) + \cdots + \alpha_n x_n(t) = 0 \quad \text{for all } t \in \mathbb{R}$$

only holds for $\alpha_0 = \alpha_1 = \cdots = \alpha_n = 0$. Then we want to show that this is also true for $n + 1$. We reduce the latter case to the case n by

differentiation. Suppose that

$$f(t) = \alpha_0 x_0(t) + \cdots + \alpha_n x_n(t) + a_{n+1} x_{n+1}(t) = 0 \quad \text{for all } t \in \mathbb{R}.$$

Then

$$f'(t) = \alpha_1 t + \cdots + n\alpha_n t^{n-1} + (n+1) a_{n+1} t^n = 0 \quad \text{for all } t \in \mathbb{R}.$$

Now the induction hypothesis implies that $\alpha_1 = \cdots \alpha_{n+1} = 0$ and by the induction base we get $a_0 = 0$. Hence $f(t)$ is identically zero. Hence the set of monomials is a linearly independent set of $\mathcal{P}$ and it spans the space of polynomials by definition. Hence it is even a basis of infinite cardinality.

(3) The space of continuous functions on the real-line, or the space of continuously differentiable function, or the space of infinitely often differentiable functions are infinite-dimensional vector spaces.

PROPOSITION 7.1.3. *Every finite-dimensional vector space has a basis.*

The proof is based on an extension principle.

PROPOSITION 7.1.4 (Basis Extension Principle). *Let X be a finite-dimensional vector space. Then any linearly independent subset of X can be extended to a basis.*

Let X be a finite-dimensional vector space of dimension n. Then any set $\{x_1, ..., x_n\}$ of n linearly independent vectors is a basis of X. In other words, any set of vectors $\{x_1, ..., x_m\}$ with $m > n$ is linearly dependent.
Any two bases of a finite-dimensional vector space have the same number of elements. These observations motivate

DEFINITION 7.1.5. Suppose X has a basis $\{x_1, ..., x_n\}$. Then we call the number of elements of this basis the *dimension* of X, denoted by $\dim(X)$. If X is infinite-dimensional, then we write $\dim(X) = \infty$.

We have that $\dim(\mathbb{C}^n) = n, \dim(\mathcal{P}_n) = n + 1$ and $\dim(\mathcal{P}) = \infty$.

There is a relation between the dimensions of two subspaces and the dimensions of their intersection and sum.

PROPOSITION 7.1.6. *Let M, N be subspaces of a finite-dimensional vector space X. Then*

$$\dim(M + N) + \dim(M \cap N) = \dim(M) + \dim(N).$$

Understanding the structure of linear transformation between finite-dimensional vector spaces is one of the main goals of linear algebra. As a first step we discuss the link between matrices and linear transformations. On the one hand a $m \times n$ matrix A defines a linear transformation from $\mathbb{C}^n$ to $\mathbb{C}^m$ by $T(x) = Ax$. Suppose $a_1, a_2, ..., a_n$ are the columns of A. Then we may denote A by $A = (a_1|a_2|\cdots|a_n)$ if the knowledge of columns is of relevance for the argument. For example, the action of a matrix A on a vector $x = (x_1, ..., x_n)^T \in \mathbb{F}^n$ is in terms of columns: $Ax = a_1 x_1 + \cdots + a_n x_n$, i.e. it amounts to taking linear combinations of the columns scaled by the coordinates of x.

On the other hand any linear transformation on finite-dimensional vector spaces can by represented in matrix form relative to a choice of bases.

We present the details for this assertion. Let $\mathcal{B} = \{x_1, ..., x_n\}$ be a basis of X and $\mathcal{C} = \{y_1, ..., y_m\}$ be a basis of Y. Suppose T is a linear transformation $T : X \to Y$ Then

$$x = \sum_{i=1}^{n} \alpha_i x_i$$

yields

$$T(x) = \sum_{i=1}^{n} \alpha_i T(x_i)$$

and thus

$$[T(x)]_{\mathcal{C}} = \sum_{i=1}^{n} \alpha_i [T(x_i)]_{\mathcal{C}}.$$

We define a $m \times n$ matrix A which has as its j-th column $[[T(x_j)]_{\mathcal{C}}]$. Then we have

$$[Tx]_{\mathcal{C}} = A[x]_{\mathcal{B}}.$$

The matrix A represents T with respect to the bases $\mathcal{B}$ and $\mathcal{C}$. Sometimes, we denote this A sometimes by $[T]_{\mathcal{B}}^{\mathcal{C}}$.

We address now the relation between the matrix representation of T depending on the change of bases. Suppose we have two bases $\mathcal{B} = \{x_1, ..., x_n\}$ and $\mathcal{R} = \{y_1, ..., y_n\}$ for X. Let $x = \sum_{j=1}^{n} \alpha_i x_i$. Then

$$[x]_{\mathcal{R}} = \sum_{j=1}^{n} \alpha_i \vec{x_i}_{\mathcal{R}}.$$

Define the $n \times n$ matrix P with j-th column $\vec{x_j}_{\mathcal{R}}$, and we call P the *change of bases matrix*:

$$[x]_{\mathcal{R}} = P[x]_{\mathcal{B}}$$

and by the invertibility of P we also have

$$[x]_{\mathcal{B}} = P^{-1}[x]_{\mathcal{R}}.$$

Let now $\mathcal{C}$ and $\mathcal{S}$ be two bases for Y. Then a linear transformation $T : X \to Y$ has two matrix representations:

$$A = [T]_{\mathcal{B}}^{\mathcal{C}} \text{ and } B = [T]_{\mathcal{R}}^{\mathcal{S}}.$$

In other words we have

$$[Tx]_{\mathcal{C}} = A[x]_{\mathcal{B}} \quad , [Tx]_{\mathcal{S}} = B[x]_{\mathcal{R}}$$

for any $x \in X$. Let P be the change of bases matrix of size $n \times n$ such that $[x]_{\mathcal{R}} = P[x]_{\mathcal{B}}$ for any $x \in X$ and let Q be the invertible $m \times m$ matrix such that $[y]_{\mathcal{S}} = Q[y]_{\mathcal{C}}$.
Hence we get that

$$[Tx]_{\mathcal{S}} = BP[x]_{\mathcal{B}}$$

and

$$[y]_{\mathcal{S}} = [Tx]_{\mathcal{S}} = Q[Tx]_{\mathcal{C}} = QA[x]_{\mathcal{B}}$$

for any $x \in X$. Hence we get that

$$B = QAP^{P} - 1 \text{ and } A = Q^{-1}BP.$$

In the case $X = Y$ we have $P = Q$ and we set $S = Q^{-1}$ to get $B = S^{-1}AS$. Then the matrices A and B represent the same linear transformation T on V with respect to different bases.
These observation motivate the definition.

DEFINITION 7.1.7. Two $m \times n$ matrices A and B are called *equivalent* if there exists an invertible matrix S such that $B = QAP^{-1}$. Furthermore, Two $n \times n$ matrices A and B are called *similar* if there exists an invertible matrix S such that $B = S^{-1}AS$.

A way to look at equivalent matrices is that this is precisely the case, when these matrices have the same rank.

Given a general $n \times n$ matrix A. Two similar matrices are "essentially the same". The notion of similarity is of utmost importance for linear algebra. It allows one to classify matrices.

We close our summary of elementary facts with the rank-nullity theorem. Suppose X and Y are finite dimensional vector spaces. Then one can construct bases for $\ker(T)$ and $\operatorname{ran}(T)$. We call the dimension of the $\ker(T)$ the *nullity* of T and the dimension of $\operatorname{ran}(T)$ the *rank* of T.

PROPOSITION 7.1.8 (rank-nullity theorem). *Let X and Y be finite dimensional vector spaces. For a linear mapping $T : X \to Y$ we have*

$$\dim(X) = \dim(\ker(T)) + \dim(\operatorname{ran}(T)).$$

PROOF. The idea is to use the dimension formula for the sum of vector spaces stated in 7.1.6.
Let X be a n-dimensional vector space. Suppose $\{b_1, ..., b_k\}$ is a basis for $\ker(T)$. Then there exist $b_{k+1}, ..., b_n$ in X such that $\{b_1, ..., b_k, ..., b_n\}$ is a basis for X. We denote by $S = \operatorname{span}\{x_{k+1}, ..., x_n\}$. Then by construction we have

$$\ker(T) \cap S = \{0\}$$

and by the dimension formula for subspaces we have

$$\dim(\ker(T) \cap S) + \dim(() \ker(T) + S) = \dim(\ker(T)) + \dim(S).$$

Since $\dim(\ker(T) \cap S) = 0$ and $\dim(() \ker(T) + S) = \dim(()X)$ we have

$$\dim(()X) = \dim(\ker(T)) + \dim(S).$$

Note that $\operatorname{ran}(T) = T(S)$ and the restriction of T to S is injective. Hence $\dim(\operatorname{ran}(T(S))) = \dim(S) = \dim(\operatorname{ran}(T))$. Thus we have the desired assertion. □

COROLLARY 7.1.9. *Let X and Y be finite dimensional vector spaces. For a linear mapping $T : X \to Y$ we have that T is injective if and only if T is surjective if and only if T is bijective.*

7.2. Invariant subspaces and Schur's form

In this section we start to view linear transformations from a more conceptual point of view. Invariance of a class of objects under some structures is an integral part of mathematics. In the case of linear transformations between vector spaces the invariance of a subspace under a linear transformation is one of the crucial notions. Since it allows one to address the main problem of linear algebra: Show

that given a linear transformation on a vector space X, there exists a basis of X with respect to which T has a reasonable simple matrix representation.

DEFINITION 7.2.1. Suppose T is a linear transformation on a vector space. A subspace M of X is called *invariant* under T if $x \in M$ implies $Tx \in M$. We will also refer to M as T-invariant subspace.

Here are some examples of invariant subspaces.

EXAMPLES 7.2.2. Let T be a linear transformation on a vector space X.
(1) $\{0\}$ and V;
(2) The kernel and the range of T.

A question of interest is if a linear operator on a vector space has an invariant subspace. We will later demonstrate that any linear transformation on a complex vector space has an invariant subspace.

Let us investigate one-dimensional invariant subspaces.

PROPOSITION 7.2.3. *A linear transformation on a finite-dimensional vector space has a one-dimensional invariant subspace if and only if T has an eigenvector.*

PROOF. ($\bullet$) Suppose M is invariant under T, then $Tx \in M$ and hence there is a scalar $\lambda \in \mathbb{F}$ such that $Tx = \lambda x$.

($\bullet$) If $Tx = \lambda x$ for some $\lambda \in \mathbb{F}$ and some non-zero $x \in X$, then the span(x) is a one-dimensional subspace. This subspace is invariant under T. □

We restrict our discussion to complex vector spaces, i.e. the scalars in our linear combinations are complex numbers.

DEFINITION 7.2.4. A scalar λ is called an *eigenvalue* of a linear transformation $T : X \to X$ if there exists a non-zero $x \in X$ such that $Tx = \lambda x$. The set $\sigma(T)$ of $\mathbb{C}$

$$\sigma(T) = \{z \in \mathbb{C} : T - zI \text{ is not invertible}\}$$

is known as the spectrum of T.

In other words, x is an eigenvector of T if and only if $x \in \ker T - \lambda I$. For finite-dimensional vector spaces $\sigma(T)$ is the set of all eigenvalues counting multiplicities of T.

THEOREM 7.1. *Suppose T is a linear transformation on a finite-dimensional complex vector space. Then there exists an eigenvalue $\lambda \in \mathbb{C}$ for an eigenvector x of T.*

PROOF. We assume that $\dim(()X) = n$ and choose any non-zero vector x in X. Consider the following set of $n + 1$ vectors in X:

$$\{x, Tx, T^2x, ..., T^nx\}.$$

Since $n + 1$ vectors in an n-dimensional vector space X are linearly independent, there exists a non-trivial linear combination:

$$a_0x + a_1Tx + \cdots + a_nT^nx = (a_0I + a_1T + \cdots + a_nT^n)x = 0.$$

Note that not all $a_1, ..., a_n$ are zero. If they were all zero, then $a_0x = 0$ which would imply that $a_0 = 0$. Hence that the linear combination is trivial.

Let us denote by $p(z) = a_0 + a_1 z + \cdots + a_n z^n$ the polynomial associated to the linear transformation T. Powers of numbers correspond to powers of T by the corresponding iterates of T and $T^0 = I$.

Then the non-trivial linear combination among the vectors turns into a polynomial equation in T:

$$p(T) = 0.$$

By the Fundamental Theorem of Algebra any polynomial can be written as a product of linear factors:

$$p(t) = c(t - \lambda_1)(t - \lambda_2) \cdots (t - \lambda_n), \ \lambda_i \in \mathbb{C}, c \neq 0.$$

Hence $p(T)$ has a factorization of the form:

$$p(T) = c(T - \lambda_1 I)(T - \lambda_2 I) \cdots (T - \lambda_m I).$$

Hence $p(T)$ is a product of linear mappings $T - \lambda_j I$ for $j = 1, ..., m$. We know that $p(T)x = 0$ for a non-zero $x \neq 0$, which implies that at least one of these linear mappings is not invertible. Thus it has to have a non-trivial kernel, let's say $y \in \ker(T - \lambda_i I)$, which yields that y is an eigenvector for the eigenvalue λ_i. Consequently, we have shown the desired assertion. □

The assumptions of the above statement are crucial: (i) Since there are linear transformations on a real vector space, do not need to have eigenvalues. For example, the rotation by 90 degrees in the plane $\mathbb{R}^2$. (ii) The left shift on ℓ^2, $(x_1, x_2, ...) \mapsto (0, x_1, x_2, ...)$ does not have any eigenvalues.

Matrix representations of a linear transformation are of a nice form if they are of upper-triangular form. We connect these upper-triangular matrices with invariant subspaces.

Proposition 7.2.5. *Suppose T is a linear transformation on a vector space X with a basis $\mathcal{B} = \{b_1, ..., b_n\}$. The following are equivalent:*

(1) *The matrix representation $[T]_{\mathcal{B}}$ of T with respect to $\mathcal{B}$ is upper-triangular.*
(2) *$Tb_j \in \operatorname{span}(b_1, ..., b_j)$ for $j = 1, ..., n$.*
(3) *$\operatorname{span}(b_1, ..., b_j)$ is T-invariant for $j = 1, ..., n$.*

The proof is part of the next problem set.

As an important consequence of the existence of an eigenvector for linear mappings between complex finite-dimensional vector spaces we prove Schur's triangularization theorem, our first classification theorem. Before we introduce a refined version of similarity. Namely, if the matrix S in the definition of similar matrices may be chosen as a chosen as a unitary matrix, then we call the matrices A and B *unitarily equivalent.*

Theorem 7.2 (Triangularization Theorem). *For any $A \in \mathcal{M}_n(\mathbb{C})$ there exists a unitary matrix U such that*

$$A = UTU^*,$$

where T is an upper triangular matrix with the eigenvalues (counted with their multiplicities) on the diagonal.
In other words, there exists an orthonormal basis $x_1, ..., x_n$ for $\mathbb{C}^n$ such that for each $k = 1, ..., n$ the vector Ax_k is a linear combination of $x_1, ..., x_k$.

We refer to the upper triangular matrix as the *Schur form* of the matrix A.

PROOF. We proceed by induction on n. For $n = 1$, there is nothing to show. Suppose that the result is true up to matrices of size $n - 1$.
Let x_1 be a normalized eigenvector of the eigenvalue λ_1 of the $n \times n$ matrix A Let $M = \text{span}(x_1)$ and let $M^\perp$ be its orthogonal complement, i.e. $\mathbb{C}^n = M \oplus M^\perp$. We denote by $P_{M^\perp}$ the orthogonal projection of with range $M^\perp$. For $x \in M^\perp$ we set $A_{M^\perp} x := P_{M^\perp} Ax$. Then $A_{M^\perp}$ is a linear operator on the $(n-1)$-dimensional space $M^\perp$.
By the induction hypothesis, there exists an orthonormal basis $x_2, ..., x_n$ of $M^\perp$ such that $A_{M^\perp} x_k$ for $k = 2, ..., n$ is a linear combination of $x_2, ..., x_n$. Hence, $\{x_1, ..., x_n\}$ is an orthonormal basis for $\mathbb{C}^n$ and each Ax_k is a linear combination of $x_1, ..., x_k$. □

Based on our observation that invariant subspaces of a linear transformation are equivalent to upper-triangular matrix representations of the transformation, we are able to answer the question about the existence of invariant subspaces of linear transformations.

COROLLARY 7.2.6. *Any linear transformation T on a complex finite-dimensional vector space has an invariant subspace.*

EXAMPLE 7.2.7. Find the Schur form of $A = \begin{pmatrix} 5 & 7 \\ -2 & -4 \end{pmatrix}$.
First step: Find an eigenvalue of A and associated eigenvector. The eigenvalues of A are $\lambda_1 = -2$ and $\lambda_2 = 3$. An eigenvector for $\lambda_1 = -2$ is $x_1 = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$.
The second step is to complete it to a basis of $\mathbb{C}^2$. In our case we take the eigenvector to the second eigenvalue and note that the corresponding set of vectors is linearly independent: $x_2 = \begin{pmatrix} 7 \\ -2 \end{pmatrix}$.
Third step: Use a orthonormalization procedure, e.g. Gram-Schmidt, to turn the system $\{x_1, x_2\}$ into a basis $\{u_1 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -1 \end{pmatrix}, u_2 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix}\}$.
Final step: Form the matrix $U = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix}$. Computation of $U^* AU = \begin{pmatrix} 2 & 9 \\ 0 & 3 \end{pmatrix}$, which has the eigenvalues of A on its diagonal and is upper triangular.

7.3. Schur form and spectral theorem

DEFINITION 7.3.1. Let $T : X \to Y$ be a linear transformation from X to Y. The subspace $E_\lambda = \ker T - \lambda I$ is called the *eigenspace* of T for the eigenvalue λ. The dimension of E_λ is called the *geometric multiplicity* of λ.

Note that E_λ consists of the eigenvectors of T and the zero vector 0.

DEFINITION 7.3.2. A $n \times n$ matrix A is called *diagonalizable* if it has n linearly independent eigenvectors. Hence the eigenvectors form a basis of $\mathbb{C}^n$. If the basis of eigenvectors is orthonormal, then we say that A is *unitarliy diagonalizable*

By definition a diagonalizable $n \times n$ matrix A has eigenvalues $\lambda_1, ..., \lambda_n$ and associated eigenvectors $u_1, ..., u_n$ satisfying:

$$Au_1 = \lambda u_1$$
$$\vdots$$
$$Au_n = \lambda u_n.$$

Collect the eigenvectors of A into one matrix: $P = (u_1|u_2|\cdots|u_n)$; and the eigenvalues of A into the diagonal matrix

$$D = \begin{pmatrix} \lambda_1 & 0 \cdots & \cdots & 0 & \\ \vdots & \lambda_2 & 0 & \cdots & 0 \\ \vdots & 0 & \ddots & \ddots & \lambda_n \end{pmatrix}.$$

Then the eigenvalue equations turn into a matrix equation:

$$AU = UD.$$

Since A is diagonalizable, the eigenvectors are a basis for $\mathbb{C}^n$. Hence P is invertible and we have

$$A = PDP^{-1}.$$

Sometimes U is an unitary matrix, i.e. the eigenvectors yield an orthonormal basis for $\mathbb{C}^n$. Then we have $A = UDU^*$. In this case we say that A is *unitarily diagonalizable.*

On several occasions we are going to rely on a basic fact about non-zero eigenvalues of the product of matrices.

LEMMA 7.3. *For a $m \times n$-matrix A and a $n \times m$-matrix B we have that AB and BA have the same non-zero eigenvalues.*

LEMMA 7.4. PROOF. A non-zero scalar λ is an eigenvalue of AB when $AB - \lambda I$ is not invertible. By a rescaling of A we can restrict our discussion to $\lambda = 1$. Let X be the inverse of $(I - AB)$, i.e. $(I - AB)X = I = X(I - AB)$ which is equivalent to ABX=XAB. Then $(I - BA)$ is also invertible and $I + BXA$ is its inverse:

$$(I - BA)(I + BXA) = I - BA + BXA - BAXA = I + BXA - BA - BABXA$$

Note that $BXA - BXABA = B(I - AB)XA = BA$ which yields that

$$(I - BA)(I + BXA) = I - BA + BA = I.$$

□

REMARK 7.3.3. The inverse of $I - BA$ in terms of the inverse of $I - AB$ might seem like a magic trick. Suppose we can use Neumann series to express the inverse of $I - AB$, i.e. $\|AB\| < 1$. Then a rewriting of the geometric series for the inverse of $I - BA$ gives the relation $I + BXA$.

We present an interplay on the structure of diagonalizable matrices and the notions from our discussion of normed spaces. Let $\mathcal{M}_n(\mathbb{C})$ denote the vector space of complex $n \times n$ matrices, and by $\mathcal{D}$ the set of diagonalizable $n \times n$ matrices.

LEMMA 7.5. *Let U be a unitary $n \times n$ matrix. Then $\mathrm{tr}(A) = \mathrm{tr}(UA)$. Furthermore, we have $\mathrm{tr}(AB) = \mathrm{tr}(BA)$ for any $n \times n$ matrices A and B.*

Recall that

$$\mathrm{tr}(A^*A) = \sum_{i,j=1}^{n} |a_{ij}|^2 = \|A\|.$$

PROPOSITION 7.3.4. *The set of diagonalizable matrices $\mathcal{D}$ is dense in $\mathcal{M}_n(\mathbb{C})$ with respect to the Frobenius norm. More explicitly, given $A \in \mathcal{M}_n(\mathbb{C})$ and $\varepsilon > 0$. There exists a diagonalizable matrix $\tilde{A} \in \mathcal{M}_n(\mathbb{C})$ such that*

$$\sum_{i,j=1}^{n} |a_{ij} - \tilde{a}_{ij}|^2 < \varepsilon.$$

Since all norms are equivalent on $\mathcal{M}_n(\mathbb{C})$, the preceding statements holds for any unitarily invariant norm on $\mathcal{M}_n(\mathbb{C})$.

PROOF. We have the Schur form for A

$$A = U \begin{pmatrix} \lambda_1 & x & \cdots & x \\ 0 & \lambda_2 & \ddots & x \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \dots & \dots & \lambda_n \end{pmatrix} U^*,$$

for a unitary matrix and eigenvalues $\lambda_1, ..., \lambda_n$ counting multiplicities. Define small perturbations of these eigenvalues λ_j such that these new numbers $\tilde{\lambda}_1, ..., \tilde{\lambda}_n$ are all distinct. We add multiples of a number η to the λ_j's:

$$\tilde{\lambda}_j = \lambda_j + j\eta, \quad \eta > 0$$

and fixed at the end of the proof. Set $\tilde{A}$

$$U \begin{pmatrix} \tilde{\lambda}_1 & x & \cdots & x \\ 0 & \tilde{\lambda}_2 & \ddots & x \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \dots & \dots & \tilde{\lambda}_n \end{pmatrix} U^*,$$

where we only change the diagonal entries of the upper triangular matrix. Now $\tilde{A}$ is diagonalizable and we have

$$\mathrm{tr}((A - \tilde{A})^*(A - \tilde{A})) = \sum_{i,j=1}^{n} |a_{ij} - \tilde{a}_{ij}|^2$$

Since the diagonal matrix with entries $\lambda_1 - \tilde{\lambda}_1, ..., \lambda_n - \tilde{\lambda}_n$ is unitarily equivalent to $A - \tilde{A}$ we deduce that

$$\mathrm{tr}((A - \tilde{A})^*(A - \tilde{A})) = \sum_{j=1}^{n} |\lambda_j - \tilde{\lambda}_j|^2.$$

By the definition of $\tilde{la}_j$ this gives

$$\sum_{j=1}^{n} |\lambda_j - \tilde{\lambda}_j|^2 = \eta^2 \sum_{j=1}^{n} j^2 = \eta^2 n(n+1)/2.$$

Consequently,

$$\sum_{j=1}^{n} |\lambda_j - \tilde{\lambda}_j|^2 \leq \varepsilon$$

for $\eta \leq 2\varepsilon/(n(n+1))$. □

A well-known criterion for the non-invertiblity of a matrix is the vanishing of its determinant. Hence eigenvalues are the zeros of the polynomial $p_A(z) = \det(zI - A)$, known as the characteristic polynomial of A. Since the eigenvalues are intrinsic to the linear transformation and is independent of its matrix represenation.

LEMMA 7.6. *Similar matrices have the same characteristic equation.*

PROOF. Let A and B be similar matrices. Thus there exists an invertible matrix S such that $B = S^{-1}AS$.

$$p_B(z) = \det(zI - S^{-1}AS) = \det(zS^{-1}S - S^{-1}AS) = \det(S^{-1}(zI - A)S) = p_A(z).$$

□

The question about the diagonalizability is of utmost importance and its answer is known as the Spectral Theorem.

THEOREM 7.7 (Spectral theorem). *Given $A \in \mathcal{M}_n(\mathbb{C})$. Then the following statements are equivalent:*

(1) *A is normal.*
(2) *A is unitarily diagonalizable. Hence there exists a unitary matrix U such that $A = UDU^*$, where D is a diagonal matrix with the eigenvalues of A as entries of the diagonal, the columns of U are the corresponding eigenvectors of A.*
(3) *$\sum_{i,j=1}^{n} |a_{ij}|^2 = \sum_{i,j=1}^{n} |\lambda_i|^2$, where $\lambda_1, ..., \lambda_n$ are the eigenvalues of A counting multiplicities.*

In the proof we make use of two useful statements. An elementary computation yields the following fact.

LEMMA 7.8. *Suppose A and B are unitarily equivalent. Then A is normal if and only if B is normal, i.e. A is normal if and only if UAU^* is normal for some unitary matrix U.*

PROOF. Elementary computations yield the assertion. □

LEMMA 7.9. *An upper triangular matrix is normal if and only if it is diagonal.*

PROOF. ($\Rightarrow$) Suppose T is an upper triangular matrix. Then the n, n-th entry of TT^* is $|t_{nn}|^2$ while the n, n-th entry of T^*T is $|t_{nn}|^2 + \sum_{i=1}^{n-1} |t_{in}|^2$. If T is normal, then these two entries have to be the same. Hence $t_{in} = 0$ for $i = 1, ..., n-1$. Repeating this argument for the entries $n-1, n-1, ..., 1$ gives that T is diagonal. ($\Leftarrow$) If T is diagonal, then T is certainly normal. □

SPECTRAL THEOREM. $(i) \Leftrightarrow (ii)$ By Schur's theorem A is unitarily equivalent to an upper triangular matrix T. Then we know that A is normal if and only if T is normal, which is normal if and only if T is diagonal. In other words, A is unitarily equivalent to a diagonal matrix.

$(ii) \Leftrightarrow (iii)$ Suppose A is unitarily equivalent to a diagonal matrix D where the diagonal entries of D are the eigenvalues $\lambda_1, ..., \lambda_n$ of A. Then

$$\sum_{i,j=1}^{n} |a_{ij}|^2 = \mathrm{tr}(A^*A) = \mathrm{tr}(D^*D) = \sum_{i=1}^{n} |\lambda_i|^2.$$

$(ii) \Leftrightarrow (ii)$ By Schur's theorem A is unitarily equivalent to a triangular matrix T:

$$\sum_{i=1}^{n} |\lambda_i|^2 = \sum_{i,j=1}^{n} |a_{ij}|^2 = \mathrm{tr}(A^*A) = \mathrm{tr}(T^*T) = \sum_{i=1}^{n} |t_{ii}|^2 + \sum_{i,j=1, i\neq j}^{n} |t_{ij}|^2.$$

Since the diagonal entries of T are the eigenvalues of A we have that

$$\sum_{i=1}^{n} |\lambda_i|^2 = \sum_{i=1}^{n} |t_{ii}|^2.$$

Hence $t_{ij} = 0$ for $i \neq j$, i.e. T is diagonal and A is unitarily equivalent to a diagonal matrix. □

The matrix $\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$ *is not normal. This matrix and its higher-dimensional analogs are going to play a crucial role in the Jordan Normal Form.*

Recall that selfadjoint matrices, $A = A^*$, are normal. Consequently our spectral theorem for normal matrices implies the spectral theorem for selfadjoint matrices.

THEOREM 7.10. *Suppose A is a selfadjoint $n \times n$ matrix. Then A is unitarily equivalent to a diagonal matrix, and the eigenvalues of A are real.*

PROOF. The fact about the diagonalizability follows from the Spectral Theorem for unitary matrices. Now let U be the unitary matrix implementing this similarity: $A = UDU^*$. Then we have $A^* = U\overline{D}U^*$. Hence A is selfadjoint if and only if the diagonal entries of D are real. Since these entries are the eigenvalues of A, we have to show that eigenvalues of a selfadjoint matrix are real numbers.
Let $\lambda \in \mathbb{C}$ be an eigenvalue of A and x an eigenvector. Then $\langle Ax, x\rangle = \langle \lambda x, x\rangle = \lambda xx$ but A is selfadjoint and thus $\langle Ax, x\rangle = \langle x, Ax\rangle = \langle x, \lambda x\rangle = \overline{\lambda} xx$. Since $x \neq 0$ we deduce that $\lambda = \overline{\lambda}$, which is the desired assertion. □

COROLLARY 7.3.5. *Let A be a $n \times m$ matrix. Then the $m \times m$ matrix A^*A and the $n \times n$ matrix AA^* are selfadjoint with non-negative eigenvalues and the positive eigenvalues coincide for these two matrices.*

PROOF. Note that $\|Ax\|^2 = \langle Ax, Ax\rangle = \langle A^*Ax, x\rangle \geq 0$. The matrices A^*A and AA^* are selfadjoint and have the same non-zero eigenvalues. Suppose λ is an eigenvalue and x an eigenvector of A^*A. Then $\langle A^*Ax, x\rangle = \langle \lambda x, x\rangle = \lambda \|x\|^2 \geq 0$ and thus $\lambda > 0$. □

In the case of unitary matrices we can also use the spectral theorem to deduce some information about the eigenvalues.

PROPOSITION 7.3.6. *A matrix A is unitary if and only if all of the eigenvalues of A have modulus one.*

DEFINITION 7.3.7. A complex selfadjoint matrix A on an n-dimensional innerproduct space $(X, \langle ., , \rangle .)$ is said to be *positive definite* if $\langle Ax, x\rangle > 0$ for all non-zero vectors $x \in X$. If A satisfies the weaker condition $\langle Ax, x\rangle \geq 0$ for all non-zero vectors $x \in X$, then we call A *semi-positive definite.*

The notion of positivity is also of interest in the infinite dimensional setting, where it lies at the heart of the theory of operator algebras. We restrict our discussion to mappings between finite-dimensional vector spaces.

REMARK 7.3.8. The matrix $\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$ is not positive definite. Hence one cannot deduce from the positivity of the matrix entries its positive definiteness.

For 2×2 matrices there is a way to state some explicit conditions on the matrix entries by just examining the quadratic form $\langle Ax, x\rangle$. Completion the squares yields that A is positive definite if and only if its pivots are positive. A good way to think about positive definite matrices is to understand its relation with the spectrum.

LEMMA 7.11. *A complex selfadjoint $n \times n$ matrix A is positive if and only if all its eigenvalues $\lambda_1, ..., \lambda_n$ are positive.*

PROOF. ($\Leftarrow$) Suppose A is positive definite. Then $\langle Ax, x\rangle$ is positive for all non-zero vectors. In particular, also for eigenvectors. Let x be an eigenvector of A. Then $\langle Ax, x\rangle = \langle \lambda x, x\rangle = \lambda \|x\|^2 > 0$ and thus $\lambda > 0$.
($\Rightarrow$) By the spectral theorem A is unitarily equivalent to a diagonal matrix given by its eigenvalues. Hence $\langle Ax, x\rangle$ is positive for all non-zero vectors. □

7.4. Singular Value Decomposition

We present a way to factorize an arbitrary matrix with complex entries, the singular value decomposition (SVD). The SVD has various applications in signal analysis, statistics, mathematics and other areas, for example the principal component analysis, data compression, identifying structures in higher-dimensional data etc.

DEFINITION 7.4.1. Given an $m \times n$ matrix A of rank r. Let $\sigma_1^2 \geq \cdots \geq \sigma_r^2$ be the positive eigenvalues of A^*A. The numbers $\sigma_1, ..., \sigma_r$ the *singular values* of A.

Since the matrix A^*A is of size $n \times n$, it has n eigenvalues and so we define the singular values to the $n-r$ zero eigenvalues to be 0, i.e. $\sigma_j := 0$ for $j = r+1, ..., n$. Since the non-zero eigenvalues of A^*A and AA^* are the same, one might want to pick either one to determine the singular values of A.

THEOREM 7.12 (Singular Value Decomposition – SVD). *Given an $m\times n$ matrix A of rank r. Let $\sigma_1 \geq \cdots \geq \sigma_r$ be the positive singular values of A. Let Σ be the $m \times n$ diagonal matrix with $\sigma_1, ..., \sigma_r$ in the first r diagonal entries and zeros elsewhere. Then there exist unitary matrices U and V, of sizes $m \times m$ and $n \times n$, respectively, such that*

$$A = U\Sigma V^*.$$

The decomposition in the theorem is often called the full SVD. The columns of the $m \times m$ matrix U are the eigenvectors of AA^*, and the columns of the $n \times n$ matrix V are the eigenvectors of A^*A

PROOF. Note that $D = \Sigma^*\Sigma$ is a real $n \times n$ diagonal matrix with $\sigma_1^2 \geq \sigma_2^2 \geq \cdots \geq \sigma_r^2$ and zeros everywhere. The matrix A^*A is a selfadjoint matrix with r positive eigenvalues $\sigma_1^2 \geq \sigma_2^2 \geq \cdots \geq \sigma_r^2$ and $n - r$ eigenvalues equal to zero. The spectral theorem yields that there exists a unitary matrix V such that

$$V^*A^*AV = D.$$

The ijth entry of V^*A^*AV is the innerproduct of columns j and i of AV. Hence the preceding equation yields that the columns of AV are pairwise orthogonal. Furthermore, when $1 \leq i \leq r$ then the length of column j is σ_j. Let U_r denote the $m \times r$ matrix with $\frac{1}{\sigma_j}$(column j of AV) as its jth column. The r columns of U_r are then an orthonormal set. Now complete U_r to an $m \times m$ matrix by using an orthonormal basis for the orthogonal complement of the column space of U_r for the remaining $m - r$ columns. Hence

$$AV = U\Sigma$$

and hence $AV = U\Sigma V^*$. □

There is other ways to write the SVD. Since only the first r diagonal entries of Σ are non-zero, the last $m - r$ columns of U and the last $n - r$ columns of V are superfluous. Let $\tilde{\Sigma}$ be the $r \times r$ matrix $\operatorname{diag}(\sigma_1, ..., \sigma_r)$. *Replace the $n \times n$ matrix U and the $m \times m$ matrix V by the $(m - r) \times (m - r)$ matrix U_r and by the $r \times n$ matrix V_r consisting of the first r rows, respectively. Hence,*

$$A = U_r\Sigma_r.$$

Summary: *Any matrix A has an SVD with a unique diagonal matrix Σ, but the unitary matrices U and V are not uniquely determined by the matrix A. It is just the way these unitaries are used that is specified: Namely, A(column j of$V$) = σ_j(column j of$U$), or in matrix form:*

$$AV = U\Sigma V^*.$$

DEFINITION 7.4.2. The vectors $u_1, u_2, .., , u_m$ and $v_1, ..., v_n$ are called the *left* and *right singular vectors*. Based on our results on the kernel of T and the range of T^* the properties of singular vectors is not surprising:

PROPOSITION 7.4.3. *Let A be a $m \times n$ matrix of rank r. Then*

$$\operatorname{ran}(A) = \operatorname{span}\{u_1, ..., u_r\}, \ker(A^*) = \operatorname{span}\{u_{r+1}, ..., u_m\}$$
$$\operatorname{ran}(A^*) = \operatorname{span}\{v_1, ..., v_r\}, \ker(A) = \operatorname{span}\{v_{r+1}, ..., v_n\}.$$

Hence we have

$$\operatorname{ran}(A) \oplus \ker(A^*) = \mathbb{C}^m$$

and

$$\operatorname{ran}(A^*) \oplus \ker(A) = \mathbb{C}^n.$$

Or in terms of basis: The columns of V^* are an orthonormal basis for $\mathbb{C}^n$ and the columns of U are an orthonormal basis for $\mathbb{C}^m$. Then A maps the jth basis vector of $\mathbb{C}^n$ to a multiple of the jth basis vector of $\mathbb{C}^m$, where the multiplier is given by the singular value σ_j. If we order the singular values decreasingly, then σ_1 is the largest factor by which the length of a basis vector is multiplied. We now show that this is the largest factor by which the length of any vector is multiplied. In other words, the operator norm of the linear transformation induced by A is equal

to the largest singular value. The operator norm of a matrix is often known as the spectral norm.

PROPOSITION 7.4.4. *Let A be a $m \times n$ matrix with singular values $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0$. Then the operator norm of A equals σ_1:*

$$\|A\| = \sigma_1.$$

PROOF. The equation $AV = U\Sigma$ gives for the first column vector v_1 of V that $\|Av_1\| = \sigma_1$. Let x be a vector of length one in $\mathbb{C}^n$. Then the SVD gives $Ax = U\Sigma V^* x$. Since V is unitary, also V^* is unitary and hence an isometry. Let us denote $V^*x = y$. Then $\|y\| = 1$ and the vector Σy is the vector where the jth component gets multiplied by σ_j. Hence $\|\Sigma y\| \leq \sigma_1 \|y\|$. Since U is unitary

$$\|Ax\| = \|U\Sigma y\| \leq \sigma_1.$$

□

A complex number may be written in polar form $z = |z|e^{2\pi i\varphi}$. The polar decomposition of a matrix A decomposes it as a product of a unitary matrix and a positive definite matrix. If one looks at the eigenvalues of these matrices, then the first one has only eigenvalues of modulus one and the other has only positive eigenvalues. Hence in terms of the spectrum of the matrices the polar decomposition is a natural generalization of the one for complex numbers.

PROPOSITION 7.4.5 (Polar decomposition). *Given a $n \times n$ matrix A. There exist a unitary matrix U and a positive definite matrix R such that*

$$A = UR.$$

PROOF. The SVD decomposition gives us unitary $n \times n$ matrices U and V such that

$$A = U\Sigma V^* = UV^*V\Sigma V^*.$$

Note that UV^* is unitary as a product of two unitary matrices and $V\Sigma V^*$ is positive definite, since Σ is positive definite. Hence $V\Sigma V^*$ is the replacement of the length of a complex number and UV^* the one for the phase factor. □

Consequently, the SVD gives in an elementary manner a polar decomposition of a matrix.

EXAMPLE 7.4.6. Determine the singular value decomposition of $\begin{pmatrix} 3 & 2 & 2 \\ 2 & 3 & -2 \end{pmatrix}$.

Write

$$A = \begin{pmatrix} 3 & 2 & 2 \\ 2 & 3 & -2. \end{pmatrix}$$

We follow the procedure for singular value decomposition. We have

$$A^*A = \begin{pmatrix} 3 & 2 \\ 2 & 3 \\ 2 & -2 \end{pmatrix} \begin{pmatrix} 3 & 2 & 2 \\ 2 & 3 & -2 \end{pmatrix} = \begin{pmatrix} 13 & 12 & 2 \\ 12 & 13 & -2 \\ 2 & -2 & 8 \end{pmatrix}$$

The non-zero eigenvalues of A^*A are $\sigma_1^2 = 25$ and $\sigma_2^2 = 9$. The normalized eigenvectors for the singular values σ_1 and σ_2 are given by:

$\underline{\sigma_1^2 = 25}$

$$\begin{pmatrix} 13-25 & 12 & 2 \\ 12 & 13-25 & -2 \\ 2 & -2 & 8-25 \end{pmatrix} \sim \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 1 & -1 & -\frac{17}{2} \end{pmatrix}$$

$v_1 = \begin{pmatrix} \frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} \\ 0 \end{pmatrix}$ is a normalized eigenvector for $\lambda_1 = 25$.

$\underline{\lambda_2 = 9}$

$$\begin{pmatrix} 13-9 & 12 & 2 \\ 12 & 13-9 & -2 \\ 2 & -2 & 8-9 \end{pmatrix} \sim \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & \frac{1}{4} \\ 1 & 0 & -\frac{1}{4} \end{pmatrix}$$

$v_2 = \begin{pmatrix} \frac{\sqrt{2}}{6} \\ -\frac{\sqrt{2}}{6} \\ \frac{2\sqrt{2}}{3} \end{pmatrix}$ is a normalized eigenvector for $\lambda_2 = 9$.

$\underline{\lambda_3 = 0}$

$$\begin{pmatrix} 13 & 12 & 2 \\ 12 & 13 & -2 \\ 2 & -2 & 8 \end{pmatrix} \sim \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & -2 \\ 1 & 0 & 2 \end{pmatrix}$$

$v_3 = \begin{pmatrix} \frac{2}{3} \\ -\frac{2}{3} \\ -\frac{1}{3} \end{pmatrix}$ is a normalized eigenvector for $\lambda_3 = 0$.

We get the singular value decomposition $A = U\Sigma V^*$ for

$$V = \big(v_1|v_2|v_3\big) = \begin{pmatrix} \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{6} & \frac{2}{3} \\ \frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{6} & -\frac{2}{3} \\ 0 & \frac{2\sqrt{2}}{3} & -\frac{1}{3} \end{pmatrix},$$

$$\Sigma = \begin{pmatrix} \sigma_1 & 0 & 0 \\ 0 & \sigma_2 & 0 \end{pmatrix} = \begin{pmatrix} \sqrt{\lambda_1} & 0 & 0 \\ 0 & \sqrt{\lambda_2} & 0 \end{pmatrix} = \begin{pmatrix} 5 & 0 & 0 \\ 0 & 3 & 0 \end{pmatrix},$$

and

$$U = \big(U_1|U_2\big) = \Big(\tfrac{Av_1}{\|Av_1\|}|\tfrac{Av_2}{\|Av_2\|}\Big) = \begin{pmatrix} \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} \end{pmatrix}.$$

Explicitly, we have

$$\begin{pmatrix} 3 & 2 & 2 \\ 2 & 3 & -2 \end{pmatrix} = \begin{pmatrix} \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} \end{pmatrix} \begin{pmatrix} 5 & 0 & 0 \\ 0 & 3 & 0 \end{pmatrix} \begin{pmatrix} \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} & 0 \\ \frac{\sqrt{2}}{6} & -\frac{\sqrt{2}}{6} & \frac{2\sqrt{2}}{3} \\ \frac{2}{3} & -\frac{2}{3} & -\frac{1}{3} \end{pmatrix}$$

Given a $m \times n$ matrix A and a vector $b \in \mathbb{C}^m$. Then we are interested in solutions of

$$Ax = b.$$

There will be a solution, if b lies in the range space of A, i.e. A has full rank. If that is not the case, there exists no solution, but we still find a best approximation of b

from the range of A. These best approximations of $Ax = b$ that minimize $\|Ax - b\|$ among all the the elements of the range of A are known as least squares solutions. The characterization of best approximation in terms of the orthogonal complement gives in this case that the least square solution with minimal norm, denoted by x^+, is the vector x such that

$$(b - Ax) \perp \operatorname{ran}(A)$$

which is equivalent to $(b-Ax)^*A = 0$, or equivalently stated in terms of the normal equations

$$A^*Ax = A^*b.$$

The matrix that produces x^+ is called the pseudoinverse of A, denoted by A^+ and we have $A^+b = x^+$. If A has full column rank, meaning $m > n$, then $A^+ = (A^*A)^{-1}A^*$. For general matrices the SVD provides a convenient way for finding A^+:

$$A^+ = V\Sigma^+U^*,$$

where Σ^+ is the $n \times m$ matrix which is the transpose of Σ, where the singular values σ_i of A are replaced by σ_i^{-1}.

Example 7.4.7. Solve the equation

$$-x_1 + 2x_2 + 2x_3 = b, \quad \text{for } b \in \mathbb{R},$$

and explain in which sense your result has to be interpreted.

We let $A = \begin{pmatrix} -1 & 2 & 2 \end{pmatrix}$ and rewrite the equation as $Ax = b$. The Singular Value Decomposition gives that

$$A = U\Sigma V^*,$$

where

$$U = \begin{pmatrix} 1 \end{pmatrix}, \quad \Sigma = \begin{pmatrix} 3 & 0 & 0 \end{pmatrix}, \quad V = \begin{pmatrix} -\frac{1}{3} & \frac{2}{\sqrt{5}} & \frac{2}{3\sqrt{5}} \\ \frac{2}{3} & 0 & \frac{\sqrt{5}}{3} \\ \frac{2}{3} & \frac{1}{\sqrt{5}} & \frac{4}{3\sqrt{5}} \end{pmatrix}.$$

The pseudoinverse of A is

$$A^+ = V\Sigma^+U^* = \begin{pmatrix} -\frac{1}{3} & \frac{2}{\sqrt{5}} & \frac{2}{3\sqrt{5}} \\ \frac{2}{3} & 0 & \frac{\sqrt{5}}{3} \\ \frac{2}{3} & \frac{1}{\sqrt{5}} & \frac{4}{3\sqrt{5}} \end{pmatrix} \begin{pmatrix} \frac{1}{3} \\ 0 \\ 0 \end{pmatrix} \begin{pmatrix} 1 \end{pmatrix} = \begin{pmatrix} -\frac{1}{9} \\ \frac{2}{9} \\ \frac{2}{9} \end{pmatrix}$$

The solutions of the equation $Ax = b$ are given by

$$x = A^+b + \ker A = \begin{pmatrix} -\frac{1}{9} \\ \frac{2}{9} \\ \frac{2}{9} \end{pmatrix} b + \ker A.$$

7.5. Generalized eigenspaces and Jordan normal form

Disclaimer: Work in progress *In this section we describe a modification of the Schur form of a matrix that is a consequence of a particular choice of bases for particular subspaces associated to the distinct eigenvalues, the so-called generalized eigenspaces.*

Suppose T is a linear operator on a finite-dimensional vector space X and let λ be an eigenvalue of T.

(1) *If the characteristic polynomial contains a factor of the form* $(x-\lambda)^a$. *Then we call* a *the algebraic multiplicity of the eigenvalue* λ.
(2) *The geometric multiplicity* g *of the eigenvalue* λ *equals the dimension of the eigenspace associated with* λ: $g := \dim(\ker(T-\lambda I))$.

The algebraic multiplicity of an eigenvalue equals the number of times λ *appears on the diagonal of the upper-triangular matrix in the Schur form.*
Note that the geometric multiplicity of an eigenvalue is always less than or equal to the algebraic multiplicity. In case the sum of the geometric multiplicities is less than the sum of the algebraic multiplicities, then T *has not enough eigenvectors to form a basis for* X *and the* T *is not invertible.*

DEFINITION 7.5.1. Let T be a linear transformation on a vector space X.

(1) A non-zero vector $x \in X$ is called a *generalized eigenvector* of T corresponding to a scalar λ if
$$(T-\lambda I)^k x = 0$$
for some positive integer k.
(2) Suppose X is n dimensional and A is the matrix representation of T for a basis of X. A non-zero vector $x \in \mathbb{C}^n$ is called a *generalized eigenvector* of degree k of the $n \times n$ matrix A corresponding to λ if $(T-\lambda I)^k x = 0$ for some positive integer k.
(3) The *generalized eigenspace* $\mathcal{GE}(T,\lambda)$ corresponding to λ is
$$\mathcal{GE}(T,\lambda) = \{x \in X : (A-\lambda I)^p x = 0 \text{ for some positive integer } p\}.$$

Note that
mathcalGE(T,λ) *consists of the zero vector and all generalized eigenvectors corresponding to* λ, *and any generalized eigenvector of degree 1 is an eigenvalue. In numerical analysis are generalized eigenspaces often called Krylov spaces.*

DEFINITION 7.5.2. Let T be a linear transformation on a finite-dimensional vector space X. We say that T is *nilpotent* if there exists a power of the matrix such that $T^k = 0$. The minimal exponent, e, such that $T^e = 0$ and $T^{e-1} \neq 0$, is the *index of nilpotency* of T.

The matrix T_p *defined by*
$$T_p = \begin{pmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \ddots & \ddots & 1 & \vdots \\ 0 & 0 & 0 & \vdots & 1 \\ 0 & 0 & 0 & \vdots & 0 \end{pmatrix}$$
is a nilpotent matrix of index $p-1$.

PROPOSITION 7.5.3. *Let* T *be a linear transformation on a finite-dimensional vector space* X. *Then* T *is nilpotent if and only if* 0 *is the only eigenvalues of* T.

PROOF. ($\Leftarrow$) Suppose T is nilpotent and λ is an eigenvalue of T. Then there exists a non-zero $x \in X$ such that $Tx = \lambda x$. Then there exists a p such that
$$0 = T^p x = \lambda^p x$$

and hence $T^p = 0$ implies that $\lambda = 0$.
($\Rightarrow$) Suppose $\sigma(T) = \{0\}$. Then T is similar to a triangular matrix with all zeros on the diagonal. The powers of an upper-triangular matrix become eventually the zero matrix. Hence T is nilpotent. □

Furthermore, let p be the smallest positive integer such that $(T - \lambda I)^p x = 0$, then $(T - \lambda I)^{p-1} x \neq 0$ and thus $T - \lambda I$ is a nilpotent operator of exponent p with eigenvalue λ.

EXAMPLE 7.5.4. Let T be the differentiation operator on the space $\mathcal{P}_n$ of polynomials of degree at most n. Then T is a nilpotent operator of index $n+1$. This is just a rephrasing of the well-known fact that the (n+1)-th derivative of x^n vanishes identically.
Suppose you pick as basis of $\mathcal{P}_n$ the monomial basis $\{1, x, x^2, ..., x^n\}$. Then the coefficients of $p(x) = a_0 + a_1 x + \cdots + a_n x^n$ are the coordinates wrt the monomial basis. The matrix of the differentiation operator is

$$T = \begin{pmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 2 & \cdots & 0 \\ \vdots & \ddots & \ddots & n-1 & \vdots \\ 0 & 0 & 0 & \vdots & n \\ 0 & 0 & 0 & \vdots & 0 \end{pmatrix}.$$

The basis representation of T wrt the normalized monomial basis $\{1, x, \frac{x^2}{2!}, ..., \frac{x^n}{n!}\}$ is

$$T = \begin{pmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \ddots & \ddots & 1 & \vdots \\ 0 & 0 & 0 & \vdots & 1 \\ 0 & 0 & 0 & \vdots & 0 \end{pmatrix}.$$

We say that a polynomial p annihilates a linear transformation T if the evaluation of p at T vanishes: $p(T) = 0$.

PROPOSITION 7.5.5. *Let T be a linar transformation on a finite-dimensional vector space X. Then there exists a non-zero annihilating polynomial for T.*

PROOF. Suppose that X is n-dimensional. Then the space of linear transformations on X is of dimension n^2. The set $\{I, T, T^2, ..., T^{n^2}\}$ is linearly dependent, since it contains more than element than the dimension of the space of linear transformation. Hence there exists a polynomial of degree n^2+1 satisfying $p(T) = 0$. □

Equivalently, let us pick a basis for the finite-dimensional vector space X. Then there exists a non-zero polynomial p of degree at most n^2 annihilating this matrix. Are there other annihilating polynomials that have degree less than n^2?

DEFINITION 7.5.6. Let A be a $n \times n$-matrix. We call the polynomial m of least degree satisfying $m(A) = 0$, the *minimal polynomial* for A.

A basic fact about the set of annihilating polynomials is that the minimal polynomial is a divisor.

PROPOSITION 7.5.7. *Let A be a $n \times n$-matrix and p a non-zero polynomial p such that $p(A) = 0$. Then there exists a polynomial q with degree less than the degree of p such that $p(z) = m(z)q(z)$.*

PROOF. Suppose p annihilates A. Then there exists q such that $\deg(q) < \deg(p)$ and a polynomial r of degree less than the one of q or $r = 0$ such that

$$p(z) = m(z)q(z) + r(z).$$

The claim is equivalent to $r = 0$. By assumption we have

$$0 = p(A) = m(A)q(A) + r(A) = 0\, q(A) + r(A)$$

and thus $r(A) = 0$. Since the degree of r is less than the one of m, this is only possible if r is the zero polynomial. □

A well-known theorem by Cayley and Hamilton is providing us with a polynomial of degree at most n that annihilates a $n \times n$-matrix: the characteristic polynomial p_A.

THEOREM 7.13 (Cayley-Hamilton). *Let A be an $n \times n$-matrix. Then $p_A(A) = 0$.*

PROOF. Our argument is based on the density of the set of diagonalizable matrices in $(\mathcal{M}_n(\mathbb{C}), \|.\|)$. The theorem is definitely true for diagonal matrices D. Thus it is also true for all diagonalizable matrices, i.e. for all matrices $A = P^{-1}DP$ where P runs through all invertible matrices P. Since $p_A(A) = P^{-1}p_D(D)P$ due to $A^k = P^{-1}D^kP$ for any positive k. We have $p_A(A) = 0$ if and only if $p_D(D) = 0$. The characteristic polynomial is a continuous function on $(\mathcal{M}_n(\mathbb{C}), \|.\|)$, since $p(z) = \det(A - zI)$ is a polynomial in the entries of A. We know that there exists a sequence (A_k) of diagonalizable matrices such that $\|A_k - A\| \to 0$ as $k \to \infty$ and that $p_{A_k}(A_k) = 0$ for $k = 1, 2, \ldots$. By continuity of the characteristic polynomial we have that $p_A(A) = 0$. □

In other words, the characteristic polynomial and the minimal polynomial have the same set of zeros but the multiplicities of the factors in m might be less than the ones for the factors of the characteristic polynomial. Thus an $n \times n$-matrix A is diagonalizable if and only if m_A has n distinct roots.

In contrast to the characteristic polynomial, there is no algorithm for computing the minimal polynomial of a matrix. Consequently, one determines the characteristic polynomial and then computes if there is a polynomial of degree less than of the characteristic polynomial with the same zeros, annihilates the matrix.

EXAMPLE 7.5.8. The minimal polynomial for $A = \begin{pmatrix} 2 & 1 & 0 \\ 0 & 2 & 1 \\ 0 & 0 & 1 \end{pmatrix}$ is equal to the characteristic polynomial $p_A(z) = (A - 2I)^3$, i.e. $\lambda = 2$ has algebraic multiplicity 3 and the $e_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$ is an eigenvector. The vectors $e_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$ and $e_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$ are generalized eigenvectors for A, since we have $(A - 2I)e_2 = 0$ and $(A - 2I)e_3 = 0$ which gives $(A - 2I)^3 e_3 = 0$.

In general, if a linear transformation has the form of a Jordan block of size n

$$T = \begin{pmatrix} \lambda & 1 & 0 & \cdots & 0 \\ 0 & \lambda & 1 & \cdots & 0 \\ \vdots & \ddots & \ddots & 1 & \vdots \\ 0 & 0 & 0 & \vdots & 1 \\ 0 & 0 & 0 & \vdots & \lambda \end{pmatrix}.$$

for a basis $\{x_0, x_1, ..., x_{n-1}\}$, then x_0 is an eigenvector for λ and $\{x_1, ..., x_{n-1}\}$ are generalized eigenvectors for λ. More explicitly, $x_0 \neq 0$ and

$$Ax_0 = \lambda x_0 \quad \text{and } Ax_j = \lambda x_j + x_{j-1}, \quad j = 1, 2, ..., n-1.$$

The set $\{x_0, x_1, ..., x_{n-1}\}$ is also known as Jordan chain. By backward substitution we have that the elements of a Jordan chain are of the form: $x_{n-j} = (A - \lambda I)^j x_{n-1}$ for $j = 1, 2, ..., n-1$ and so

$$\{x_0, x_1, ..., x_{n-1}\} = \{(A - \lambda I)^j x_{n-1} : j = 1, ..., n-1\}.$$

Let us determine the generalized eigenspace for the differentiation operator D on $\mathcal{P}_n$. The operator D is nilpotent of exponent $n+1$ and the eigenvalue $\lambda = 0$ has multiplicity $n+1$. The eigenvector of D is the constant function 1 and the generalized eigenvectors are $\{x, \frac{x^2}{2!}, ..., \frac{x^n}{n!}\} = \{D^j(\frac{x^n}{n!}) : j = 1, ..., n\}$.

We add to the nilpotent operator of differentiation D a multiple of the identity. Then the eigenvector of $D - \lambda I$ is $e^{\lambda x}$ and a generalized eigenvector of degree 2 is the function $xe^{\lambda x}$ as we have $D(xe^{\lambda x}) = e^{\lambda x} + x\lambda e^{\lambda x}$, i.e. $(D - \lambda I)(xe^{\lambda x}) = e^{\lambda x}$. More generally, a generalized eigenvector of degree k is of the form $x^k e^{\lambda x}$. Thus the generalized eigenspace $\mathcal{GE}(\lambda, D - \lambda I)$ is spanned by $\{e^{\lambda x}, xe^{\lambda x}, ..., x^n e^{\lambda x}\}$.

We state that generalized eigengspaces of the form are of the aforementinoned type.

Proposition 7.5.9. *Let T be a nilpotent linear transformation on a finite-dimensional vector space X of dimension n. Then the generalized eigenspace $\mathcal{GE}(\lambda, T)$ is spanned by the Jordan chain $\{x_{n-1}, Tx_{n-1}, ..., T^{n-1}x_{n-1}\}$ for a vector x_{n-1} satisfying $T^{n-1}x_{n-1}x_{n-1} \neq 0$.*

There is a way to determine generalized eigenspaces of T based on the knowledge of the minimal polynomial m_T. Suppose the minimal polynomial factorizes as $m_T(z) = (z - \lambda_1)^{k_1} \cdots (z - \lambda_m)^{k_m}$. Define $p_l(z) = m_T(z)(z - \lambda_l)^{k_l}$ for $l = 1, ..., m$, Then

$$\mathcal{GE}(T, \lambda_l) = ran(p_l(T)).$$

The Jordan Normal Form of a linear transformation T on a finite-dimensional vector space uses Jordan chain for each eigenvalue to break up the matrix in smaller pieces where each transformation behaves like a Jordan block.

Theorem 7.14 (Jordan Normal Form). *Let T be a linear transformation on a finite-dimensional vector space X of dimension n with distinct eigenvalues $\lambda_1, ..., \lambda_m$. Then $X = \mathcal{GE}(T, \lambda_1) \oplus \cdots \oplus \mathcal{GE}(T, \lambda_m)$ and if we pick for each generalized eigenspace*

a Jordan chain, then the matrix of T is of the form

$$\begin{pmatrix} J_1 & & & \\ & J_2 & & \\ & & \ddots & \\ & & & J_m \end{pmatrix},$$

where $J_k = \begin{pmatrix} \lambda_k & 1 & 0 & \cdots & 0 \\ 0 & \lambda_k & 1 & \cdots & 0 \\ \vdots & \ddots & \ddots & 1 & \vdots \\ 0 & 0 & 0 & \vdots & 1 \\ 0 & 0 & 0 & \vdots & \lambda_k \end{pmatrix}$ *for* $k = 1, ..., m$.

In other words, any complex $n \times n$-matrix is similar to $J = \begin{pmatrix} J_1 & & & \\ & J_2 & & \\ & & \ddots & \\ & & & J_m \end{pmatrix}$ *where J_k is a Jordan block for λ_k, $k = 1, ..., m$. The columns of the change of basis P are the Jordan chains for $\lambda_1, ..., \lambda_m$ and we have $A = P^{-1}JP$.*

A standard application of the Jordan Normal Form is the computation of the exponetial of a matrix:

$$e^A = I + \frac{A}{1!} + \frac{A^2}{2!} + \cdots + ...,$$

which is a bounded operator $\|e^A\| \leq e^{\|A\|}$. Like in our discussion of geometric series of bounded operators, one deduces that the exponetional of an $n \times n$-matrix A exists, i.e. the partial sums converge. For $B = P^{-1}AP$ we have $e^B = P^{-1}e^AP$ and thus the computation of e^A can be reduced to Jordan blocks J. Hence we have to compute

$$e^J = \sum_{k=0}^{\infty} \frac{(\lambda I + N)^k}{k!} = I + (I + N) + \frac{(I + N)^2}{2!} + \cdots \frac{(I + N)^n}{n!},$$

and use that $(I + N)^k = \sum_{k=0}^{n} \binom{n}{k} \lambda^{n-k} N^k$. Thus

$$e^J = \begin{pmatrix} e^\lambda & \lambda & \lambda^2/2! & \cdots & \lambda^n/n! \\ 0 & e^\lambda & \lambda & \cdots & \\ \vdots & \ddots & \ddots & \lambda & \vdots \\ 0 & 0 & 0 & \vdots & \lambda \\ 0 & 0 & 0 & \vdots & e^\lambda \end{pmatrix}$$

For a matrix A we have that $e^{(s+t)A} = e^{sA}e^{tA}$ for any $s, t \in \mathbb{R}$. Consequently, $(e^{tA})^{-1} = e^{-tA}$. We have also that the derivative of e^{tA} is

$$(e^{tA})' = A + tA + \cdots = Ae^{tA}.$$

Like in the scalar case this allows us to solve differential equations for vectors in $\mathbb{C}^n$:

PROPOSITION 7.5.10. *Given an $n \times n$-matrix A and a differentiable vector-valued function x on $\mathbb{R}$. Let $x(t_0) = x_0$ be a fixed vector in $\mathbb{C}^n$. Then the initial value propblem $(x(t))' = Ax(t)$ with $x(t_0) = x_0$ has as solution*

$$x(t) = e^{tA}x_0.$$

Bibliography

[1] C. Heil. *A Basis Theory Primer. Expanded ed.* Applied and Numerical Harmonic Analysis. Basel: Birkhäuser, 2011.