Contents

1 What kind of science is functional analysis? Linear spaces and linear operators

What does functional analysis study? What objects does he work with? What problems does it solve? What sciences is it related to?

Functional analysis is associated with many branches of mathematics:

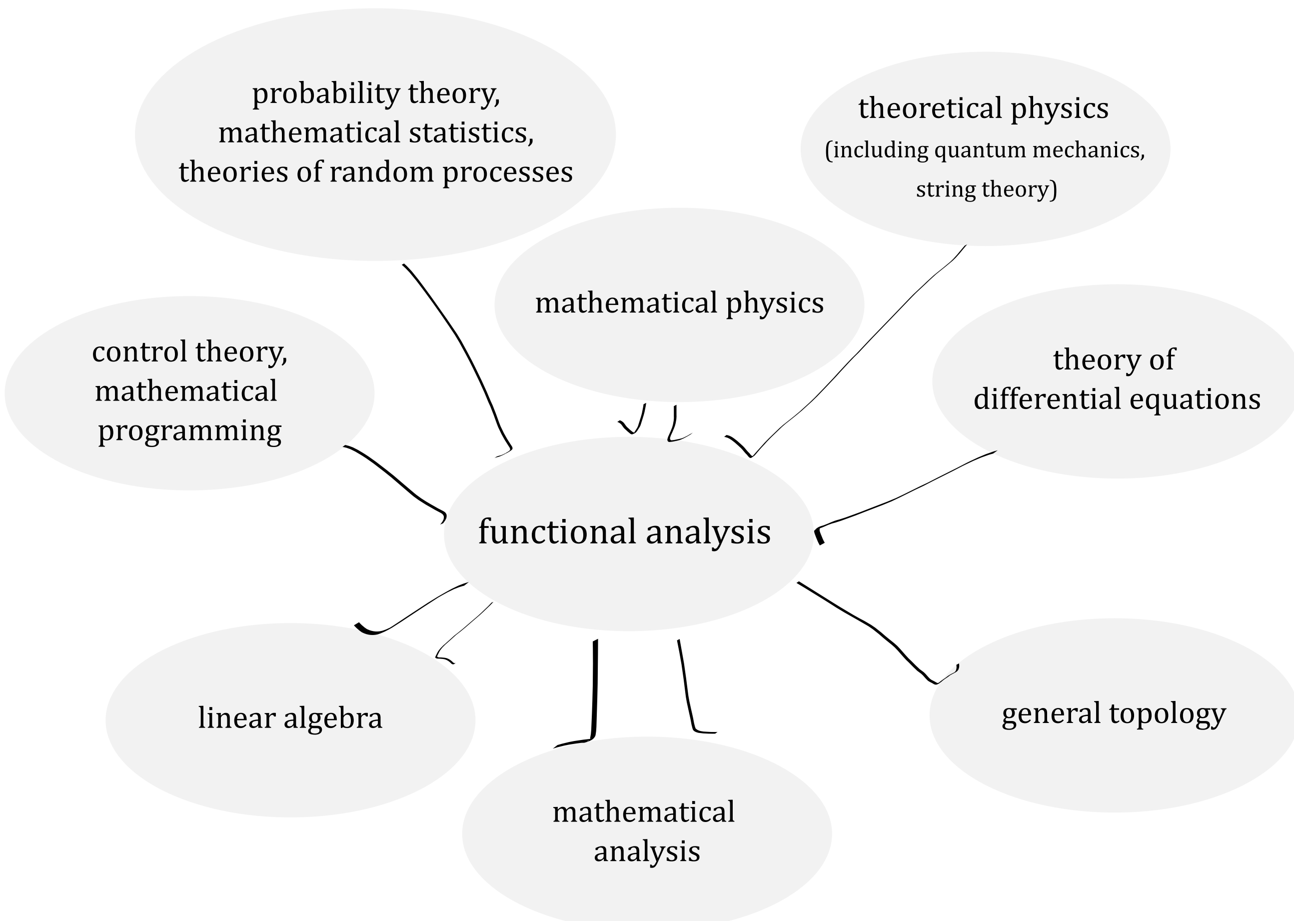


Figure 1: Scheme of connections between functional analysis and other scientific disciplines.

In the second half of XIX century, the interest of mathematicians shifted from studying individual functions (their differentiability, integrability) to function spaces. One may say that a function space wraps up functions of interest into one geometric object. The geometry of the function space reflects important properties of functions. This lead to the development of the field of functional analysis. In functional analysis, we view functions as points or vectors in a function space.

Let's start with most important examples of functional spaces (hence the name "functional analysis") that are linear vector spaces (check!):

Examples of linear vector spaces.

1.0.1 Examples (function spaces)**.**

1. $\{$all functions $\mathbb{R} \to \mathbb{R}\}$ (or $\{$all functions $\mathbb{C} \to \mathbb{C}\}$). This space is too large, and are seldom studied.

2. $\{$all solutions of a linear homogeneous partial differential equation$\}$.

3. $\mathcal{L}^1[a, b] = \{$all Lebesgue integrable functions on $[a, b]\}$ (real or complex-valued, it depends on the context, the designation is the same for both cases).

4. $\mathcal{L}^\infty[a, b] = \{$all bounded almost everywhere functions on $[a, b]\}$.

5. $C[a, b] = \{$all continuous functions on $[a, b]\}$.

6. $C^1[a, b] = \{$all continuously differentiable functions on $[a, b]\}$.

7. $C^\infty[a, b] = \{$all infinitely differentiable functions on $[a, b]\}$.

8. $C(K) = \{$all continuous functions on a compact $K\}$.

9. $\mathcal{P}(x) = \{$all algebraic univariate polynomials$\}$.

10. $\mathcal{P}^n(x) = \{$all algebraic univariate polynomials of degree at most $n\}$.

11. $\mathcal{H}(O) = \{$all $\mathbb{C}$-analytic functions on $O\}$, where O is an open set in $\mathbb{C}$.

Also there are many natural examples of sequence spaces that are linear vector spaces (check!):

1.0.2 Examples (sequence spaces)**.**

1. $s = \{$all numerical sequences $(a_k)_{k=1}^\infty\}$. This space is too large, and is seldom studied.

2. For $p \in [1, +\infty)$

$$l^p = \{\text{all absolutely summable in } p\text{-power sequences}\},$$

(that is $l^p = \{(a_k)_{k=1}^\infty \ : \ \sum_{n=1}^\infty |a_k|^p < \infty\}$.

3. $l^\infty = \{$all bounded sequences of real (or complex) numbers$\}$.

4. $c = \{$all convergent sequences of real numbers$\}$.

5. $c_0 = \{$all sequences of real numbers converging to zero$\}$.

6. $c_{00} = \{$all sequences of real numbers with finite support$\}$ (that is, only a finite number of coordinates of an element of this space are different from zero).

We plan to study the theory of linear operators acting in infinite–dimensional spaces (functional spaces), and we also plan to consider the topological properties of such spaces. From algebra and mathematical analysis, the definitions of this section should be familiar to you, but because of their great importance for this course, let's remind ourselves of them.

1.1 Subspaces

Let's remember that a subset of a vector space, which is itself a vector space, is called a (linear) subspace of this space. More precisely, if X_0 and X are two linear spaces over the same field, and $X_0 \subseteq X$, then X_0 is called a ***(linear) subspace*** of X; notation: $X_0 \leq X$. Any subspace is closed under the operations of addition of vectors and multiplication by scalars.

Remark. One can check that the following set-theoretic inclusions hold:

$$\begin{gathered}\mathcal{P}^n(x) \leq \mathcal{P}(x) \leq C^\infty[a,b] \leq C[a,b] \subseteq \mathcal{L}^\infty[a,b] \leq \mathcal{L}^2[a,b] \leq \mathcal{L}^1[a,b], \\ c_{00} \leq l^1 \leq c_0 \leq c \leq l^\infty.\end{gathered} \tag{1.1}$$

For consistency of these inclusions, we restrict the functions in $\mathcal{P}_n(x)$ and $\mathcal{P}(x)$ onto $[a,b]$.

1.1.1 Exercise (easy)**.** Let X be a linear vector space. Show that $\{\mathbf{0}\}$ and X are subspaces of X. Show that the intersection of an arbitrary collection of subspaces of X is again a subspace of X.

If for a linear space X there exists a maximum number n of linearly independent vectors of this space, then this number is called the ***dimension*** of the space X (and write $\dim X = n$). If such a number does not exist, then the space X is called an ***infinite-dimensional*** space (and write $\dim X = \infty$).

Obviously, $X_0 \leq X$ implies $\dim X_0 \leq \dim X$. Thus, all the spaces in (1.1), excluding $\mathcal{P}_n(x)$, are infinite-dimensional, because $\dim \mathcal{P}_n(x) = n+1$ (why?) and for all n

$$\dim c_{00} \geq \dim\big(\mathbb{R}^n \times \{(0,0,\dots)\}\big) = \dim \mathbb{R}^n = n$$

1.2 Quotient spaces

The notion of quotient space allows one easily to collapse some directions in linear vector spaces. One reason for doing this is when one has unimportant directions and would like to neglect them (a vivid and important example is L^1).

Definition. *Let X_0 be a subspace of a linear vector space X. Consider an equivalence relation on X defined as*

$$x \sim y \quad \text{if} \quad x - y \in X_0.$$

The ***quotient space*** *X/X_0 is then defined as the set of equivalence classes (cosets) $[x]$ for all $x \in X$, with operations defined as*

$$[x] + [y] := [x+y], \quad a[x] := [ax] \quad \text{for } x, y \in X, a \in \mathbb{R}.$$

1.2.1 Exercise (easy)**.** Prove that the operations above are well defined, and that quotient space is indeed a linear space.

The dimension of the quotient space is called the ***codimension*** of X_0, thus

$$\operatorname{codim}(X_0) := \dim(X/X_0).$$

1.2.2 Remarks. 1). Observe that $[x]$ *is an affine subspace:*

$$[x] = x + X_0 := \{x + h \ : \ h \in X_0\}.$$

2). From undergraduate linear algebra we know that *if X is finite dimensional then all of its subspaces X_0 satisfy*

$$\dim X_0 + \operatorname{codim} X_0 = \dim X.$$

1.2.3 Exercise. Let X be the real l_1, $f : X \to \mathbb{R}$, $f(x) = \sum_{k=1}^{\infty} x_k$ if $x = (x_k)_{k=1}^{\infty}$, and $\operatorname{Ker} f = \{x \in l_1 \ : \ f(x) = 0\}$ is the ***kernel*** of f. Calculate $\operatorname{codim}(\operatorname{Ker} f)$. What can be said about the case when l_1 is complex and $f : X \to \mathbb{C}$?

1.2.4 Example (space L^p for $p \in [1, \infty)$)**.** The concept of quotient space arises naturally when we define the space $L^p(X, \mathcal{A}, \mu)$ of all functions integrable to the p-th degree on $(E, \mathcal{A}, \mu)$, where $(E, \mathcal{A}, \mu)$ is an arbitrary measure space. We first consider

$$X = \mathcal{L}^p(E, \mathcal{A}, \mu) :=$$
$$\{f \text{ is measurable on } (E, \mathcal{A}, \mu)\} \cap \{|f|^p \text{ is integrable on } (E, \mathcal{A}, \mu)\},$$

and

$$X_0 = \{\text{all functions equal to zero } \mu\text{-almost everywhere on } \ E\}. \quad (1.2)$$

Then we define

$$L^p(E, \mathcal{A}, \mu) = X/X_0.$$

This way, the elements of L^p are, strictly speaking, not functions but classes of equivalences. But in practice, one thinks of an $f \in L^p$ as a function, keeping in mind that functions that coincide μ-almost everywhere are the same. In fact, thus we expand the concept of "function".

1.2.5 Example (Space L^∞)**.** A similar procedure is used to define the space of essentially bounded functions $L^\infty = L^\infty(E, \mathcal{A}, \mu)$. A real valued (or complex valued) function f on E is called ***essentially bounded*** if there exists a bounded measurable function g on E such that $f = g$ μ-almost everywhere. Similar to the previous example, we consider the linear vector space

$$X = \mathcal{L}^\infty(E, \mathcal{A}, \mu) = \{\text{all essentially bounded functions on } (E, \mathcal{A}, \mu)\},$$

and the subspace X_0 defined by (1.2), which we would like to neglect, and then we define

$$L^\infty(E, \mathcal{A}, \mu) = X/X_0.$$

1.2.6 Example (codim c_0 in c)**.** As we know, the space c_0 of sequences converging to zero is a subspace of the space c of all convergent sequences. Let us observe that $\operatorname{codim} c_0 = 1$.

Indeed, every sequence $x \in c$ can be uniquely represented as

$$x = a\mathbf{1} + z \quad \text{for some} \;\; a \in \mathbb{R}, \;\; z \in c_0,$$

where $\mathbf{1} = (1, 1, \dots)$. (How do we choose the value of a?). Hence

$$[x] = a[\mathbf{1}] + [z] = a[1].$$

It follows that every element $[x] \in c/c_0$ is a constant multiple of the element $[1]$. Therefore, $\dim(c/c_0) = \dim\{a \in \mathbb{R}\} = 1$ as claimed. This example shows that the space c_0 makes up almost the whole space c, except for one dimension given by the constant sequences. This explains why the space c is rarely used in practice; one prefers to work with c_0 which is almost the same as c but has the advantage that we know the limits of all sequences there (zero).

1.3 Metric spaces and normed spaces

1.3.1 Definition. *A* ***metric space*** *is a pair (X, ρ) consisting of a set X and a* ***metric*** *or a* ***distance function*** *$\rho : X \times X \to \mathbb{R}$ that satisfies the following axioms.*

(I) $\rho(x,y) \geq 0$ *for all* $x,y \in X$, *with equality if and only if* $x = y$ *(identity axiom).*

(II) $\rho(x,y) = \rho(y,x)$ *for all* $x,y \in X$ *(symmetry axiom).*

(III) $\rho(x,z) \leq \rho(x,y)+\rho(y,z)$ *for all* $x,y,z \in X$ *(triangle inequality).*

A subset $O \subseteq X$ *of a metric space* (X,ρ) *is called* ***open*** *(or* ρ*-open) if, for every* $x \in O$, *there exists a constant* $\varepsilon > 0$ *such that the* ***open ball***

$$B(x,\varepsilon) := B_\rho(x,\varepsilon) := \{y \in X \quad : \quad \rho(x,y) < \varepsilon\}$$

(centered at x *with radius* ε*) is contained in* O. *The set of* ρ*-open subsets of* X *will be denoted by*

$$\Omega(X,\rho) := \{O \subseteq X \ : \ O \quad \text{is } \rho\text{−open}\}.$$

It follows directly from the definitions that the collection $\Omega(X,\rho) \subseteq 2^X$ of ρ-open sets in a metric space (X,ρ) satisfies the axioms of a ***topology*** (i.e. the empty set and the set X are open, arbitrary unions of open sets are open, and finite intersections of open sets are open). A subset F of a metric space (X,ρ) is ***closed*** (i.e. its complement is open) if and only if the limit point of every convergent sequence in F is itself contained in F. Recall that a sequence $(x_n)_{n\in\mathbb{N}}$ in a metric space (X,ρ) is called a ***Cauchy sequence*** (or ***converging on itself*** or ***fundamental*** sequence) if, for every $\varepsilon > 0$, there exists an $n_0 \in \mathbb{N}$ such that any two integers $n,m \geq n_0$ satisfy the inequality $\rho(x_n,x_m) < \varepsilon$. Recall also that a metric space (X,ρ) is called ***complete*** if every Cauchy sequence in X converges.

1.3.2 Definition. *A* ***normed*** *vector space is a pair* $(X,\|\cdot\|)$ *consisting of a real or complex vector space* X *and a function*

$$X \to \mathbb{R} \ : \ x \to \|x\|$$

satisfying the following axioms.

(I) $\|x\| \geq 0$ *for all* $x \in X$, *with equality if and only if* $x = 0$ *(identity axiom).*

(II) $\|tx\| = |t|\|x\|$ *for all* $x \in X$ *and* $t \in \mathbb{R}$ *(positive homogeneity).*
(III) $\|x + y\| \leq \|x\| + \|y\|$ *for all* $x, y \in X$ *(triangle inequality).*

Let $(X, \|\cdot\|)$ be a normed vector space. Then the formula

$$\rho(x, y) := \|x - y\| \tag{1.3}$$

for $x, y \in X$ defines the ***natural metric*** or the ***metric generated by the norm*** on X. The resulting topology is denoted by $\Omega(X) := \Omega(X, \|\cdot\|) := \Omega(X, \rho)$.

But is every metric on a linear space generated by some kind of norm?

The answer is obviously negative: it is easy to find a metric for which the function given by equality $\|x\| = \rho(x, 0)$ will be bounded and hence not be positive homogeneous (for example, one can consider $\rho(x, y) = |\operatorname{arctg} x - \operatorname{arctg} y\,|$ on $\mathbb{R} \times \mathbb{R}$.)

Such a question is more interesting.

Is a linear topological (or at least linear metric) space always normable? That is, is it true that the topology of space X is generated by some norm?

The following exercise gives the answer.

1.3.3 Exercise (not very easy. More likely, a little complicated. Maybe we'll consider its decision later)**.** Let s is the space of all real sequences, and

$$\rho(a, b) = \sum_{k=1}^{\infty} \frac{1}{2^k} \cdot \frac{|a_k - b_k|}{1 + |a_k - b_k|} \quad \text{for} \quad a = (a_k)_{k=1}^{\infty} \text{ and } \; b = (b_k)_{k=1}^{\infty}.$$

Show that 1) ρ is a metric on s, and 2) the topology generated by the metric ρ is not generated by any norm, that is, the metric space (X, ρ) is not normable.

1.3.4 Exercise (not very easy. More likely, a little complicated. Maybe we'll consider its decision later)**.** Let s is the space of all real sequences, and

$$\rho(a,b) = \sum_{k=1}^{\infty} \frac{1}{2^k} \cdot \frac{|a_k - b_k|}{1 + |a_k - b_k|} \quad \text{for} \quad a = (a_k)_{k=1}^{\infty} \text{ and } \quad b = (b_k)_{k=1}^{\infty}.$$

Show that 1) ρ is a metric on s, and 2) the topology generated by the metric ρ is not generated by any norm, that is, the metric space (X, ρ) is not normable.

As explained to you in the course of mathematical analysis, the following functions set norms on linear spaces, which we consider as the main ("native") norms on these spaces.

1. $X = \mathbb{R}^n$ or $X = \mathbb{C}^n$, $\quad \|x\| := \|x\|_2 = \sqrt{\sum\limits_{k=1}^{n} |\, x_k \,|^2}$

2. $X = \mathbb{R}^n$ or $X = \mathbb{C}^n$, $1 \le p < \infty$, $\|x\|_p = \left(\sum\limits_{k=1}^{n} |\, x_k \,|^p \right)^{1/p}$.

3. $X = \mathbb{R}^n$ or $X = \mathbb{C}^n$. $\quad \|x\|_\infty = \max\limits_{1 \le k \le n} |\, x_k \,|$.

4. $X = l^p$, $1 \le p < \infty$ (real or complex), $\quad \|x\|_p = \left(\sum_{k=1}^{\infty} |\, x_k \,|^p \right)^{1/p}$.

5. $X = \ell^\infty$ (real or complex), $\quad \|x\|_\infty = \sup\limits_{k \ge 1} |\, x_k \,|$.

6. $X = C(K)$, where K is a compact, $\quad \|f\| = \max\{|\, f(x) \,|,\ x \in K\}$.

7. $X = L^p(E, \mathcal{A}, \mu)$, $1 \le p < \infty$, $\|f\|_p = \left(\int\limits_E |\, f \,|^p \, d\mu \right)^{1/p}$.

8. $X = L^\infty(E, \mathcal{A}, \mu)$, $\|f\|_\infty = \operatorname{ess\,sup}_E |f|$, where

$$\operatorname{ess\,sup}_E g := \inf\{\sup_{E\setminus e} g \ : \ e \subseteq E,\ \mu(e) = 0\} =$$
$$\inf\{\sup_E h : \ h = g \ a.e.\}.$$

If it is known from the context which norm is being discussed, the designation $(X, \|\cdot\|)$ is usually shortened to X.

1.3.5 Exercise (L^∞ as the limit of L^p)**.** This exercise explains the index ∞ in the name of the spaces L^∞ and l^∞.

1. Consider the space $L^\infty = L^\infty(E, \mathcal{A}, \mu)$ with finite total measure $\mu(E)$. Show that if $f \in L^\infty$ then $\|f\|_p \to \|f\|_\infty$ as $p \to \infty$.

2. Show that if $x \in l^{p_0}$ for some $p_0 \geq 1$ then $\|x\|_p \to \|x\|_\infty$ as $p \to \infty$.

1.3.6 Definition. *Let X be a linear vector space. Two norms $\|\cdot\|$ and $\|\cdot\|'$ on X are called* ***equivalent*** *if there are two positive constants m, M such that*

$$\forall\, x \in X \qquad m\|x\| \leq \|x\|' \leq M\|x\|.$$

1.3.7 Exercises. 1) Definition 1.3.6 yields an equivalence relation on the set of all norm functions on X.

2) Let X be a linear space, $\|\cdot\|$ and $\|\cdot\|'$ are two norms on X. Then the following statements are equivalent:

(I) $\|\cdot\| \sim \|\cdot\|'$.

(II) there is a constant $C \geq 1$ such that

$$\forall\, x \in X \qquad \frac{1}{C}\|x\| \leq \|x\|' \leq C\|x\|.$$

(III) $\|\cdot\|$ and $\|\cdot\|'$ induce the same topologies on X, i.e. $\Omega(X, \|\cdot\|) = \Omega(X, \|\cdot\|')$.

3) Let $\|\cdot\|$ and $\|\cdot\|'$ be equivalent norms on X. Show that $(X, \|\cdot\|)$ is complete if and only if $(X, \|\cdot\|')$ is complete.

4) Let $X = l^1$. Are the norms $\|\cdot\|_p$ and $\|\cdot\|_q$ equivalent in this space for $p \neq q$, $p, q \in [1; +\infty]$?

5) Are the norms $\|\cdot\|_{C[a;b]}$ and $\|\cdot\|_{L[a,b]}$ equivalent in the $C[a, b]$?

1.4 Finite-Dimensional Normed Vector Spaces.

1.4.1 Theorem. *Let X be a finite-dimensional linear vector space (real or complex). Then any two norms on X are equivalent.*

The proof of this theorem for case $X = \mathbb{R}^n$ was presented in the course of mathematical analysis. There exists an algebraic isomorphism between an arbitrary real n-dimensional space and a $\mathbb{R}^n$ (If $E_1, \ldots, E_n$ is an arbitrary basis of X, and $e_1 = (1, 0, \ldots, 0), \ldots, e_n = (0, \ldots, 0, 1)$ form the standard basis of $\mathbb{R}^n$, then assuming $F(\sum_{k=1}^n x_k e_k) = \sum_{k=1}^n x_k E_k$ we obtain an explicit expression for such an isomorphism). An isomorphism allows you to transfer the norm from one space to another and back, so the theorem is true in general case too.

Or you can consider a direct proof (if you foget the old one).

Proof. Choose a basis $e_1, ..., e_n$ on X and define for all $x \in X$

$$\|x\|_2 = \sqrt{\sum_{k=1}^{n} |x_k|^2} \tag{1.4}$$

if $x = \sum_{k=1}^n x_k e_k$, $x_k \in \mathbb{R}$ (or $\mathbb{C}$ respectively).

It is enough to prove that every norm on X is equivalent to $\|x\|_2$.

Fix any norm function $\|\cdot\|$ on X.

1) Upper estimate of the $\|\cdot\|$-norm. Define

$$M = \sqrt{\sum_{k=1}^{n} |e_k|^2}.$$

Then for any $x \in X$, $x = \sum_{k=1}^{n} x_k e_k$ by the triangle inequality for $\|\cdot\|$ and the Cauchy inequality on $\mathbb{R}^n$, we have

$$\|x\| \leq \sum_{k=1}^{n} |x_k| \cdot \|e_k\| \leq \sqrt{\sum_{k=1}^{n} |x_k|^2} \sqrt{\sum_{k=1}^{n} \|e_k\|^2} = M\|x\|_2.$$

2) Lower estimate of the $\|\cdot\|$-norm. The $\|\cdot\|_2$-sphere $S := \{x \in X : \|x\|_2 = 1\}$ is compact with respect to $\|\cdot\|_2$ by the Heine–Borel Theorem, and the function $x \to \|x\|$ is continuous (even Lipschitz continuous) respectively $\|\cdot\|_2$ by step 1). Hence (by the Weierstrass's theorem) there exists an element $m > 0$ such that $\|x\| \geq m$ for all $x \in S$. Then every nonzero vector $x \in X$ satisfies $\frac{x}{\|x\|_2} \in S$, hence $\left\|\frac{x}{\|x\|_2}\right\| \geq m$ and hence $\|x\| \geq m\|x\|_2$. $\square$

1.5 Banach spaces

1.5.1 Definition. *A normed vector space $(X, \|\cdot\|)$ is called a* ***Banach space*** *if it is complete as the metric space (with natural metric on it).*

What was proved in the course of mathematical analysis? That the spaces $\mathbb{R}^n$, $\mathbb{C}^n$ and $C(K)$ (with usual metric) is Banach.

1.5.2 Exercises. 1. Show that l^∞ and L^∞ are Banach spaces. (Hint: modify the proof of $C(K)$-completeness theorem).

2. Show that c_0 is a Banach space (Hint: the set of sequences converging to zero c_0 are closed subspaces of l_∞ (why?)).

1.6 Series in Banach spaces. Completeness of L^p

We are going to give a useful criterion of completeness of normed spaces in terms of convergence of series rather than sequences. We shall use this criterion to prove the completeness of the L^p spaces.

1.6.1 Definition. Let $(x_k)_{k=1}^{\infty}$ be a sequence of vectors in a normed space X. As before we say that *the series* $\sum_{k=1}^{\infty} x_k$ *converges in* X *to* x *and we write*

$$\sum_{k=1}^{\infty} x_k = x,$$

if $x \in X$ *and the partial sums* $s_N := \sum_{k=1}^{N} x_k$ *converge to* x.

A series $\sum_{k=1}^{\infty} x_k$ *is called* ***absolutely convergent*** *if*

$$\sum_{k=1}^{\infty} \|x_k\| < \infty. \tag{1.5}$$

Recall that in the scalar case, where $X = \mathbb{R}$ or $\mathbb{C}$, absolute convergence of series implies convergence (but not vice versa). As the following theorem shows, this happens precisely because of the completeness of $\mathbb{R}$ and $\mathbb{C}$.

1.6.2 Theorem (Completeness criterion)**.** *A normed space* X *is a Banach space if and only if every absolutely convergent series in* X *converges in* X.

Proof. 1. Necessity. Let X be a Banach space, and consider an absolutely convergent series $\sum_{k=1}^{\infty} x_k$, i.e. let (1.5) holds.

We want to prove that the series $\sum_{k=1}^{\infty} x_k$ converges. By completeness of X, it sufficient to show that the partial sums of this series are Cauchy, i.e. that $\|s_n - s_m\| \to 0$ as $n > m \to 0$. To this end, we use triangle inequality and our assumption (1.5) to obtain $\|s_n - s_m\| = \|\sum_{k=n+1}^{m} x_k\| \leq \sum_{k=n+1}^{m} \|x_k\| \to 0$. This completes the proof of necessity.

2. Sufficiency. Assume that X is incomplete; we want to construct a divergent series which is absolutely convergent. By incompleteness, there exists a Cauchy sequence (v_k) in X which diverges.

Every subsequence of (v_k) diverges (check it!). Therefore, there exists a subsequence (w_k) of (v_k) which diverges but which is «rapidly Cauchy», i.e.

$$\|w_2 - w_1\| \le \frac{1}{2}, \quad \dots \quad \|w_{k+1} - w_k\| \le \frac{1}{2^k}, \quad \dots \tag{1.6}$$

(Construct it!) It follows that the sequence (x_k) defined as

$$\forall\, k \in \mathbb{N} \quad x_k = w_{k+1} - w_k$$

forms the terms of an absolutely convergent series:

$$\sum_{k=1}^{\infty} \|x_k\| \le \sum_{k=1}^{\infty} \frac{1}{2^k} = 1 < +\infty.$$

Nevertheless, the partial sums $\sum_{k=1}^{n} x_k = w_{n+1} - w_1$ diverge. So we have constructed an absolutely convergent series in X which diverges. This completes the proof. $\square$

1.6.3 Exercise. Show that if a Cauchy sequence (v_k) in some normed space X diverges, than every subsequence of (v_k) diverges.

1.6.4 Remark. *The conditions of convergence of a series $\sum_{k=1}^{\infty} x_k(t)$ familiar from real analysis, necessary ($\|x_k\| \to 0$) and sufficient (absolutely convergence), are valid in any Banach space.*

1.6.5 Theorem. *For every $p \in [1; +\infty]$, the space $L^p = L^p(E, \mathcal{A}, \mu)$ is a Banach space.*

Proof. We prove the theorem only for $1 \le p < \infty$. The case of $p = \infty$ is some kind of exercise.

Further in this proof $\|\cdot\| := \|\cdot\|_{L^p}$.

Let's use the completeness criterion (1.6.2). Let functions (x_k) in L^p form the terms of an absolutely convergence series, i.e.

$$\sum_{k=1}^{\infty} \|x_k\| =: M < \infty.$$

By the completeness criterion, it suffices to show that the series $\sum_{k=1}^{\infty} x_k$ converges in L^p. Let's first show that this series converges almost everywhere. Let

$$b(t) := \sum_{k=1}^{\infty} |x_k(t)|, \quad b_m(t) = \sum_{k=1}^{m} |x_k(t)| \text{ for all } m \in \mathbb{N}.$$

$b(t)$ is a function defined almost everywhere, and on the domain of its definition it is holds $b(t) \in [0; \infty]$. By Fatou's theorem (because $(b_m(t))$ is a sequence of measurable positive functions, in addition, it is increasing) for all $m \in \mathbb{N}$ we have:

$$\int_E b^p d\mu = \int_E \lim_{m\to\infty} b_m^p d\mu \leq \lim_{m\to\infty} \int_E b_m^p d\mu =$$
$$\lim_{m\to\infty} \|b_m\|^p \leq \lim_{m\to\infty} \left(\sum_{k=1}^{m} \|x_k\| \right)^p \leq M^p.$$

So,

$$b \in L^p(E, \mathcal{A}, \mu) \;\Leftrightarrow\; b^p \in L(E, \mathcal{A}, \mu) \;\Rightarrow\; b \text{ is finite almost everywhere.}$$

It means that the series $\sum_{k=1}^{\infty} x_k(t)$ absolutely converges almost everywhere. Hence it converges almost everywhere. It remains to check that this series converges not only a.e., but also in L^p. The tails of this series

$$r_m(t) = \sum_{k=m+1}^{\infty} x_k(t)$$

form a sequence of functions in L^p that a.e. converge to 0, and the p-th degree of which satisfies Lebesgue's theorem on major convergence:

$$|r_m(t)|^p \leq \left(\sum_{k=m+1}^{\infty} |x_k(t)| \right)^p \leq b^p(t) \in L,$$

so by Lebesgue's theorem $\|r_m\|^p = \int_E |r_m|^p \to 0$ as $m \to \infty$, and we conclude that the series $\sum_{k=1}^{\infty} x_k(t)$ converges in L^p as required. $\square$

From the topology course you probably know that every *metric space* (X,ρ) *has a **completion***, that is is a pair consisting of a complete metric space $(\widetilde{X},\widetilde{\rho})$ and an isometry $\varphi : X \to \widetilde{X}$ such that $\varphi(X)$ is dense in $\widetilde{X}$. *The completion of X is unique up to an isometry.*

To prove the completion statement next ideas are usefull. Consider the space X_0 of all Cauchy sequences (x_k) in X, equipped with the semi-norm (semi-norm is «almost norm» without condition $\|x\| = 0 \Rightarrow x = \mathbf{0}$)

$$\|(x_k)\| := \lim_{k\to\infty} \|x_k\|$$

Then turn this space into a normed space $\widetilde{X}$ by taking quotient over the kernel of the semi-norm.

1.6.6 Exercises (Space of continuously differentiable functions)**.**

1) Show that the space $C^k[a,b]$ of k-times differentiable functions is not a Banach space with respect to the sup-norm.

2) Show that $C^k[a,b]$ is a Banach space with respect to the norm

$$\|f\|_{C^k} := \|f\|_\infty + \|f'\|_\infty + \cdots + \left\|f^{(k)}\right\|_\infty .$$

1.7 Inner product spaces

Hilbert spaces form an important and simplest class of Banach spaces. Speaking imprecisely, Hilbert spaces are those Banach spaces where the concept of orthogonality of vectors is defined. Hilbert spaces will arise as complete inner product spaces.

1.7.1 Definition. *Let X be a linear space over field $\mathcal{K}$, $\mathcal{K} = \mathbb{C}$ or $\mathbb{R}$. An **inner product** on X is a function $\langle\cdot,\cdot\rangle : X \times X \to \mathcal{K}$ which satisfies the following three axioms:*

(i) $\langle x,x\rangle \geq 0$ *for all* $x \in X$*;* $\langle x,x\rangle = 0$ *if and only if* $x = 0$*;* (positivity)

(ii) $\langle \alpha x + \beta y, z\rangle = \alpha\langle x, z\rangle + \beta\langle y, z\rangle$ *for all* $x, y, z \in X$ *and* $\alpha, \beta \in \mathcal{K}$ (linearity in the first variable)*;*

(iii) $\langle x, y\rangle = \overline{\langle y, x\rangle}$ *for all* $x, y \in X$ (conjugate symmetry for $\mathbb{C}$-case and symmetry for $\mathbb{R}$-case.)

*The space X with an inner product is called an **inner product space**. The **norm associated to an inner product** is the function*

$$X \to \mathbb{R}_+ \ : \ x \to \|x\| := \sqrt{\langle x, x\rangle}. \tag{1.7}$$

Remark. The inner product is congjugate linear in the second argument:

$$\langle x, \alpha y + \beta z\rangle = \overline{\alpha}\langle x, y\rangle + \overline{\beta}\langle x, z\rangle.$$

1.7.2 Definition. *If $\langle x, y\rangle = 0$ we say that vectors x and y are orthogonal and write $x \perp y$.*

Exercise. The canonical example of an inner product space is the space $\mathbb{C}^n = \{z = (z_1, \ldots, z_n) \ : \ z_1, \ldots, z_n \in \mathbb{C}\}$ equipped with the inner product

$$\langle z, w\rangle = \sum_{k=1}^{n} z_k \overline{w_k}.$$

1.7.3 Theorem (Elementary inequalities for the norm generated by the inner product)**.** *Let X be an inner product space equipped with the norm (1.7). The inner product and norm satisfy the Cauchy-Bunyakovsky-Schwarz (CBS) inequality*

$$|\langle x, y\rangle| \le \|x\| \cdot \|y\| \tag{1.8}$$

and the triangle inequality

$$\|x + y\| \le \|x\| + \|y\|. \tag{1.9}$$

for all $x, y \in X$. Thus the function defined by (1.7) is a norm on X.

Proof. The CBS inequality is obvious when $x = 0$ or $y = 0$. So assume $x \neq 0$ and $y \neq 0$ and define $X := \frac{x}{\|x\|}$ and $Y := \frac{y}{\|y\|}$. Then $\|X\| = \|Y\| = 1$. Hence

$$0 \leq \|X - \langle X, Y\rangle Y\|^2 =$$
$$\|X\|^2 - \langle\langle X, Y\rangle Y, X\rangle - \langle X, \langle X, Y\rangle Y\rangle + \|\langle X, Y\rangle Y\|^2 =$$
$$1 - \langle X, Y\rangle\overline{\langle X, Y\rangle} - \overline{\langle X, Y\rangle}\langle X, Y\rangle + |\langle X, Y\rangle|^2 = 1 - |\langle X, Y\rangle|^2$$

This implies $|\langle X, Y\rangle| \leq 1$ and hence $|\langle x, y\rangle| = |\langle\|x\|X, \|y\|Y\rangle| = \|x\|\|y\||\langle X, Y\rangle| \leq \|x\|\|y\|$. In turn it follows from the CBS inequality that

$$\|x + y\|^2 = \langle x + y, x + y\rangle = \langle x, x\rangle + \langle x, y\rangle + \langle y, x\rangle + \langle y, y\rangle =$$
$$\|x\|^2 + \langle x, y\rangle + \overline{\langle x, y\rangle} + \|y\|^2 = \|x\|^2 + 2\operatorname{Re}\langle x, y\rangle + \|y\|^2 \leq$$
$$\|x\|^2 + 2|\langle x, y\rangle| + \|y\|^2 \leq \|x\|^2 + 2\|x\| \cdot \|y\| + \|y\|^2$$
$$\leq (\|x\| + \|y\|)^2 \ \Rightarrow (1.9).$$

□

The calculation above clearly implies Pythagorean theorem.

1.7.4 Reminder (Pythagorean theorem). *if X is an inner product space, and $x, y, x_1, \dots x_n \in X$, $x \perp y$ and $x_i \perp x_j$ as $i \neq j$. Then*

$$\begin{gathered}\|x + y\|^2 = \|x\|^2 + \|y\|^2, \\ \|x_1 + \cdots + x_n\|^2 = \|x\|_1^2 + \cdots + \|x\|_n^2.\end{gathered} \tag{1.10}$$

The space L^2. The basic example of an inner product space famous from the course of mathematical analysis is $L^2 = L^2(E, \mathcal{A}, \mu)$:

Proposition (Canonical inner product on L^2). *For any $f, g \in L^2$, the quantity*

$$\langle f, g\rangle := \int_E f\overline{g}\, d\mu$$

is finite, and it defines an inner product on L^2. This inner product obviously agrees with the L^2 norm, i. e. $\|f\|_2 = \sqrt{\langle f, f\rangle}$.

Recall that the space of square-summable sequences l^2 is a particular case of $L^2 = L^2(\mathbb{N}, 2^{\mathbb{N}}, \mu)$ for the counting measure μ. Therefore, l^2 is also an inner product space. The inner product formula reads as

$$\langle x, y\rangle := \sum_{k=1}^{\infty} x_k \overline{y_k} \quad \text{for } x = (x_k),\ y = (y_k) \in l^2.$$

We can recast Cauchy-Schwarz inequality in this specific space L^2 and l^2 as follows.

1.7.5 Corollary (Cauchy-Bunyakovsky-Schwarz inequality in L^2 and l^2). *For every $f, g \in L^2$ and every $x, y \in L^2$ one has*

$$\Big|\int_E f\bar{g}\, d\mu\Big| \leq \sqrt{\int_E |f|^2\, d\mu} \cdot \sqrt{\int_E |g|^2\, d\mu},$$

$$\Big|\sum_{k=1}^{\infty} x_k \bar{y}_k\Big| \leq \sqrt{\sum_{k=1}^{\infty} |x_k|^2} \cdot \sqrt{\sum_{k=1}^{\infty} |y_k|^2}.$$

Maybe all norms are generated by some kind of scalar products?

It turns out that not all norms satisfy the parallelogram low. The parallelogram law in planar geometry states that for every parallelogram, the sums of squares of the diagonals equals the sum of squares of the sides. This statement remains to be true in all inner product spaces:

1.7.6 Proposition (parallelogram law). *Let X be an inner product space. Then for every $x, y \in X$ one has*

$$\|x+y\|^2 + \|x-y\|^2 = 2\left(\|x\|^2 + \|y\|^2\right) \tag{1.11}$$

Proof. The result follows once we recall that

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

and similarly

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

$\square$

1.7.7 Example. Let's consider $\mathbb{R}^2$ with p-norm,

$$\|(x_1, x_2)\|_p = (|x_1|^p + |x_2|^p)^{1/p}, \quad p > 1.$$

Let's consider the vectors $x = (1, 0)$ and $y = (0, 1)$. Then $x + y = (1, 1)$, $x - y = (1, -1)$, $\|x\|_p = \|y\|_p = 1$, $\|x + y\|_p = \|x - y\|_p = 2^{1/p}$ (for $p = \infty$ it means $2^0 = 1$), and the parallelogram law for these vectors reduces to equality $2 = 2^{2/p}$ which is holds only when $p = 2$.

Conclusion: $\|\cdot\|_p$ on the space $\mathbb{R}^2$ does't satisfy the parallelogram low, hence *it is not generated by an inner product.*

1.7.8 Exercise. Prove that $\mathbb{R}^2_p = (\mathbb{R}^2, \|\cdot\|_p)$ isometrically embedded in both L^2 space and l^2 space.

1.7.9 Corollary. *For any $p \in [1, \infty]$, $p \neq 2$ the spaces L^p and l^p are not generated by any inner product.*

Proposition (Relations between L^p spaces). *Let $1 \leq p \leq P \leq \infty$. Then the following statements are true.*

(I) *If $(E, \mathcal{A}, \mu)$ be a measure space with a finite measure ($\mu(E) < \infty$). Then $L^P \leq L^p$.*

(II) $\|x\|_P \leq \|x\|_p$ *for all $x \in l^p$. In particular, we have the inclusion $l^p \leq l^P$.*

(III) *In the case of an arbitrary space with a measure, generally speaking, it is not true that one of the spaces L^p and L^P is a subspace of the other.*

Proof. (I) was proved in Math. an. curse.

(II) In case $\|x\|_p = 1$ one has $\sum_k |x_k|^p = 1 \Rightarrow$ for all k $|x_k|^p \leq 1 \Rightarrow$ for all k $|x_k|^P \leq |x_k|^p \Rightarrow$

$$\sum_k |x_k|^P \leq \sum_k |x_k|^p = 1 \ \Rightarrow \|x\|_P \leq 1 = \|x\|_p .$$

(III) In case $E = (0, +\infty)$, and $\mu = \lambda_1$ (one-dimensional Lebesgue measure) let's consider

$$f_1(t) = \min\{1, 1/t\}, \quad f_2(t) = \frac{1}{\sqrt{t}} \cdot \chi_{(0,1)}.$$

Than
$f_1 \in L^2(0, +\infty)$, $f_1 \notin L^1(0, +\infty)$, hence $L^2(0, +\infty) \not\leq L^1(0, +\infty)$ and $f_2 \in L^1(0, +\infty)$, $f_1 \notin L^2(0, +\infty)$, hence $L^1(0, +\infty) \not\leq L^2(0, +\infty)$.

Then in case $x \neq \mathbf{0}$ we get $\|x\|_P = \left\| \|x\|_p \cdot \frac{x}{\|x\|_p} \right\|_P = \|x\|_p \cdot \left\| \frac{x}{\|x\|_p} \right\|_P \leq \|x\|_p \cdot 1$ (because $\left\| \frac{x}{\|x\|_p} \right\|_P = 1$). The case $x = \mathbf{0}$ is obvious. $\square$

1.7.10 Example. For arbitrary fixed dimensions $m, n \in \mathbb{N}$, we consider the space of matrices

$$M_{m,n} := \{\text{all } m \times n \text{ matrices with complex entries}\}.$$

One can turn $M_{m,n}$ into an inner product space by defining the trace inner product as

$$\langle A, B \rangle := \operatorname{tr}(AB^*) = \sum_{i=1}^{m} \sum_{j=1}^{n} a_{ij} \overline{b_{ij}}. \tag{1.12}$$

This is clearly an inner product. One way to see this is to identify $M_{m,n}$ with $\mathbb{C}^{mn}$ by concatenating the rows of a matrix $A \in M_{m,n}$ into a long vector in $\mathbb{C}^{mn}$. Then the canonical inner product in $\mathbb{C}^{mn}$ is the same as the right hand side of (1.12). The norm defined by the inner product on $M_{m,n}$ is called *Hilbert-Schmidt* or *Frobenius* norm of matrices:

$$\|A\|_{\text{HS}} = \sqrt{\langle A, A \rangle} = \sqrt{\sum_{i=1}^{m} \sum_{j=1}^{n} |a_{ij}|^2}. \tag{1.13}$$

Note some similarity between the forms of the inner product in L^2, which is $\langle f, g \rangle = \int_E f \overline{g} \, d\mu$ and in $M_{m,n}$, which is $\langle A, B \rangle := \operatorname{tr}(AB^*)$

— the integral is replaced by the trace, functions by matrices, complex conjugation by transposition, and product of functions by product of matrices.

1.7.11 Example (A space of random variables)**.** The space L^2 arises in probability theory in a natural way as a *space of random variables* with finite variance. Indeed, consider a probability space $(\Omega, \Sigma, \mathbb{P})$. Recall that a random variable X is a measurable real-valued function defined on Ω. The expectation of X is, by definition, the integral of X:

$$\mathbb{E}X = \int_\Omega X(\omega)\, d\mathbb{P}(\omega).$$

Therefore, the space $L^2(\Omega, \Sigma, \mathbb{P})$ consists of all random variables X with *finite second moment*:

$$\|X\|_2 = \sqrt{\mathbb{E}X^2} < \infty.$$

1.7.12 Exercise. Show that $X \in L^2(\Omega, \Sigma, \mathbb{P})$ if and only if X has *finite variance*:

$$\mathrm{Var}(X) = \mathbb{E}(X - \mathbb{E}X)^2 = \mathbb{E}X^2 - (\mathbb{E}X)^2 < \infty.$$

The concepts of covariance and correlation coefficient have some geometric meaning, too. Consider two random variables X and Y, and for simplicity assume that they have mean zero, i.e. $\mathbb{E}X = \mathbb{E}Y = 0$. Then the covariance of X and Y is nothing else than the inner product in L^2:

$$\mathrm{cov}(X, Y) := \mathbb{E}(X - \mathbb{E}X)(Y - \mathbb{E}Y) = \mathbb{E}XY = \langle X, Y \rangle.$$

Similarly, the correlation coefficient between X and Y is

$$\mathrm{corr}(X, Y) := \frac{\mathrm{cov}(X, Y)}{\sqrt{\mathrm{Var}(X)}\sqrt{\mathrm{Var}(Y)}} = \frac{\mathbb{E}XY}{\sqrt{\mathbb{E}X^2}\sqrt{\mathbb{E}Y^2}} = \frac{\langle X, Y \rangle}{\|X\|_2 \|Y\|_2}.$$

What is the ratio of the inner product to the product of the norms of the multipliers?

This is the cosine of the angle between the multiplier vectors!

Hence the correlation coeffcient is nothing else as the cosine of the angle between random variables X and Y considered as vectors in L^2. This demonstrates the geometric meaning of correlation the more random variables X and Y are correlated, the less the angle between them, and vice versa.

1.7.13 Exercise. Show that functions $\phi_1(\cdot) = \|\cdot\|$ defined on a normed space X and $\phi_2 = \langle \cdot, y \rangle$, defined on an inner product space X with fixed $y \in X$ are Lipschitz continuous.

The geometry of inner product spaces is dominated by the concept of orthogonality.

1.7.14 Definition. *Let Y be a subset of an inner product space X. The **orthogonal complement** of Y is defines as*

$$Y^{\perp} = \{x \in X \ : \ \langle x, y \rangle = 0 \quad \textit{for all} \quad y \in Y\}.$$

1.7.15 Proposition. *Let Y be a subset of an inner product space X. Then $Y^{\perp}$ is a closed subspace of X is a closed linear subspace of X. Moreover,*

$$Y \cap Y^{\perp} \leq \{0\}. \tag{1.14}$$

Inclusion of (1.14) means that if both intersecting spaces are not empty, then the intersection is $\{0\}$.

Proof. It is easy to check that $Y^{\perp}$ is a linear subspace of X. To show that $Y^{\perp}$ is a closed set, express it as

$$Y^{\perp} = \bigcap_{y \in Y} \{y\}^{\perp}.$$

And $\{y\}^{\perp}$ is a closed set for every $y \in Y$ because this set is a preimage of a closed set $\{0\}$ under continuous mapping $\phi(\cdot) = \langle \cdot, y \rangle$ (see 1.7.13).

Finally, to show (1.14), consider $Y \cap Y^{\perp}$; it follows that $\langle x, x \rangle = 0$ which implies $x = 0$. □

1.8 Hilbert spaces

1.8.1 Definition. *A complete inner product space is called a* ***Hilbert space****.*

1.8.2 Example. As we know, $L^2(E, \mathcal{A}, \mu)$ is a Hilbert space. In particular, l^2 is a Hilbert space. Also, $\mathbb{C}^n$ and $\mathbb{R}^n$ are Hilbert spaces; therefore the space of matrices $M_{m,n}$ is also a Hilbert space.

1.8.3 Exercise. Show that any closed subspace of a Hilbert space is Hilbert space itself.

1.8.4 Theorem (Orthogonality principle)**.** *Let Y be a closed linear subspace of a Hilbert space X, and let $x \in X$. Then the following holds.*

(i) *There exists a unique closest point $y \in Y$ to x, i.e. such that*

$$\|x - y\| = \min_{\widetilde{y} \in Y} \|x - \widetilde{y}\| . \tag{1.15}$$

(ii) *The point y is the unique vector in Y such that*

$$x - y \perp Y. \tag{1.16}$$

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

The point y with 1.16 propertie is called the ***orthogonal projection*** of x onto the subspace Y.

Proof. (i). *Existence.* Denote the distance by

$$d := \min_{\widetilde{y} \in Y} \|x - \widetilde{y}\| .$$

Let us choose a sequence $(y_n) \in Y$ which satisfies

$$\|x - y_k\| \to d. \tag{1.17}$$

Since Y is closed, it is a Banach space. Therefore it suffices to show that $(y_k)_k$ is a Cauchy sequence. (By continuity of the norm it would follow for $y = \lim_k y_k$ that (1.15) holds).

To bound $\|y_n - y_m\|$ we use parallelogram law. We apply it for the parallelogram with vertices x, y_n, y_m (and whose fourth vertex is determined by these three).

Parallelogram law then yields

$$\|y_n - y_m\|^2 + 4\left\|x - \frac{1}{2}(y_n + y_m)\right\|^2 = 2\big(\|x - y_n\|^2 + \|x - y_m\|^2\big).$$

By definition of d, we have $\left\|x - \frac{1}{2}(y_n + y_m)\right\| \geq d$, and by construction we have $\|x - y_n\| \to d$, $\|x - y_m\| \to d$. With this, we conclude that

$$0 \leq \liminf \|y_n - y_m\|^2 \leq 2(d^2 + d^2) - 4d^2 = 0.$$

Therefore (y_k) is a Cauchy sequence as required.

Uniqueness. Suppose there existed two different closest points $y, \widetilde{y}$ for x. Then by the alternating sequence $y, \widetilde{y}, y, \widetilde{y}, \ldots$ fulfil (1.17) but it would not be Cauchy, contradicting the argument above. Part (i) of Theorem is proved.

(ii) *Orthogonality*. Assume that $x - y \notin Y^\perp$, so

$$\langle x - y, \widehat{y}\rangle \neq 0 \text{ for some } \widehat{y} \in Y.$$

By multiplying $\widehat{y}$ by an appropriate complex scalar, we can assume that $\langle x-y, \widehat{y}\rangle$ is a real number. We will show that by moving y in the direction of $\widehat{y}$, one can improve the distance from x, which will contradict the definition of y. Namely, for every $t \in \mathbb{R}$ the definition of y implies that

$$\|x - y\|^2 \leq \|x - y + t\widehat{y}\|^2 = \|x - y\|^2 + 2t\langle x - y, \widehat{y}\rangle + t^2 \|\widehat{y}\|^2.$$

This implies that the quadratic polynomial in t satisfies

$$2t\langle x - y, \widehat{y}\rangle + t^2 \|\widehat{y}\|^2 \geq 0 \quad \text{for all} \quad t \in \mathbb{R}.$$

This can only happens if $\langle x-y, \widehat{y}\rangle = 0$ which contradicts our assumption.

Uniqueness. Suppose there are two vectors $y', y'' \in Y$ which satisfy $x - y' \in Y^\perp$, $x - y'' \in Y^\perp$. Since $Y^\perp$ is a linear subspace, subtracting yields that $y1-y2 \in Y^{\perp}$. But Y is also a linear subspace, so $y'-y'' \in Y$. Since $Y \cap Y^{\perp} \subseteq \{0\}$, it follows that $y'-y'' \in \{0\}$, hence $y' = y''$.

(iii) Since $x - y \perp Y \;\Rightarrow x - y \perp y$, from Pythagorean theorem one have $\|x\|^2 = \|x-y\|^2 + \|y\|^2 \geq \|y\|^2$, and the last inequality gives the required. $\square$

1.8.5 Definition. *A sequence (x_k) in a Hilbert space X is called an* ***orthogonal system*** *if*

$$\langle x_k, x_j \rangle = 0 \quad \text{for all} \;\; k \neq j.$$

If additionally $\|x_k\| = 1$ for all k, the sequence (x_k) is called an ***orthonormal system****.*

Equivalently, (x_k) is an orthonormal system if

$$\langle x_k, x_j \rangle = \delta_{kj},$$

where δ_{kj} equals 1 if $k = j$ and 0 otherwise (it is called Kronecker's delta function).

1.8.6 Example (Canonical basis of l^2)**.** In the space l^2, consider the vectors

$$x_k = (0, \ldots, 0, 1, 0, \ldots)$$

whose all coordinates are zero except the k-th equals 1. The sequence $(x_k)_{k=1}^{\infty}$ is clearly an orthonormal system in in l^2.

1.8.7 Example (Fourier basis in L^2)**.** In the space $L^2[-\pi, \pi]$, consider the exponentials

$$e_k(t) = \frac{1}{\sqrt{2\pi}} e^{ikt}, \quad t \in [-\pi, \pi]. \tag{1.18}$$

Then $(e_k)_{k \in \mathbb{Z}}$ is an orthonormal system in $L^2[-\pi, \pi]$ (is it a new example or not?).

1.8.8 Exercise. Prove that an orthogonal system is a linearly independent set.

The main interest in orthogonal systems is that they allow us to form orthogonal expansions of every vector $x \in X$. Such expansions are infinite series. So our first task will be to clarify when orthogonal series converge.

1.8.9 Theorem (Convergence of orthogonal series)**.** *Let (x_k) be an orthogonal system in a Hilbert space X. Then the following are equivalent:*

(i) $\sum_k x_k$ *converges in* X*;*

(ii) $\sum_k \|x_k\|^2 < \infty$*;*

(iii) $\sum_k x_k$ *converges unconditionally in* X*, i.e. for every reordering of terms. In case of convergence, one have*

$$\Big\|\sum_k x_k\Big\|^2 = \sum_k \|x_k\|^2. \tag{1.19}$$

Proof. (i) $\Leftrightarrow$ (ii). By the Cauchy criterion, the series $\sum_k x_k$ converges if and only if its partial sums form a Cauchy sequence in X, i.e.

$$\Big\|\sum_{k=n}^{m} x_k\Big\|^2 \to 0 \quad n, n \to 0. \tag{1.20}$$

Note that by Pythagorean theorem 1.7.4, the quantity in (1.20) equals to $\sum_{k=n}^{m} \|x_k\|^2$. So using Bolzano-Cauchy criterion again we see that (1.20) is equivalent to the convergence of the series $\sum_k \|x_k\|^2$, as required.

(ii)$\Rightarrow$(iii). The scalar series $\sum_k \|x_k\|^2$ converges absolutely, therefore also unconditionally (as we know from the analysis course). Hence, by the equivalence of (i) and (ii) proved above, the series $\sum_k x_k$ converges unconditionally.

(iii)$\Rightarrow$ (i) is trivial.

The last part of the theorem, identity (1.17), follows by taking limit in Pythagorean identity (1.10).

$\square$

1.8.10 Exercise. For what coefficients a_k does the series $\sum_{k\in\mathbb{Z}} a_k e^{ikt}$ converge in $L^2[-\pi,\pi]$?

1.9 Definition and examples of linear operators in infinite-dimensional spaces

1.9.1 Definition. *Let X and Y be linear spaces. The mapping $F : X \to Y$ is called **linear** if*

$$\forall\, x, \widetilde{x} \in X \qquad \forall\, \alpha, \beta \in \mathbb{R}(\mathbb{C}) \qquad F(\alpha x + \beta \widetilde{x}) = \alpha F(x) + \beta F(\widetilde{x}).$$

*If $Y = \mathbb{R}(\mathbb{C})$ is the scalar field, than F is also called a (linear) **functional**.*

Linear maps are often (especially in algebra) called ***linear operators***.

1.9.2 Examples. 1) Let $X = C[a,b]$, $y(s) = \int_a^b x(t)\cos(ts)dt$ for $x = x(t) \in C[a,b]$ and $s \in [a,b]$. Then, by the properties of the integral depending on the parameter, $y = y(s) \in C[a,b]$. Let $F(x) := y$, or $F(x)(s) = \int_a^b x(t)\cos(ts)dt$, so $F : X \to X$ and, as easy to see, F is linear (linearly depends on x).

2) Let $X = C[a,b]$, $f_1(x) = x(a)$, $f_2(x) = \int_a^b x(t)dt$, $f_3(x) = \int_a^b x(t)e^t dt$ are linear functionals.

3) Let $X = C^1[a,b]$, $Y = C[a,b]$, $F(x) = x'$, that is $F(x)(t) = x'(t)$ for all $t \in [a,b]$. Than $F : X \to Y$ is linear.

4) Let $X = L^p(E,\mathcal{A},\mu)$, $p \in [1,+\infty]$, $\frac{1}{p} + \frac{1}{q} = 1$ (if $p = 1$, than $q = \infty$ and vice versa), and $g \in L^q(E,\mathcal{A},\mu)$ is fixed. Than due to Hölder inequality xg is a integrable function, and the mapping $F(x)(t) = x(t)\cdot g(t)$, $F : L^p(E,\mathcal{A},\mu) \to L^1(E,\mathcal{A},\mu)$ is linear.

5) Let $X = L^p(E,\mathcal{A},\mu)$, $\frac{1}{p} + \frac{1}{q} = 1$, and $g \in L^q(E,\mathcal{A},\mu)$ is fixed. Then $f(x) = \int_E x(t)g(t)d\mu(t)$ is a linear functional.

1.9.3 Reminder. *If linear normed spaces X, Y are finite-dimensional, and in these spaces the bases $\{e_1, \ldots, e_k\}$ and $\{E_1, \ldots, E_m\}$ are fixed, than any linear operator $F : X \to Y$ can be represented as an operator of multiplication by the matrix*

$$[F] := (a_{ij})_{\substack{i=1:m \\ j=1:k}},$$

where a_{ij} is i-th coordinate of the vector $F(e_j)$. And any such operator is continuous.

But what about the continuity of a linear operator in the case of infinite-dimensional spaces? In this case the situation is much more complicated.

1.9.4 Definition. *Let X and Y be normed spaces. The linear mapping $F : X \to Y$ is called* ***bounded*** *if there exists a number C such that*

$$\|F(x)\| \leq C\|x\| \quad \textit{for all} \quad x \in X.$$

1.9.5 Remark. A bounded linear operator is *usually* not a bounded mapping *(in the general sense). The exception is the constant null operator.*

1.9.6 Reminder. *1) $F : X \to Y$ is called* ***Lipschitz continuous*** *if there exists a constante $c \geq 0$ such that*

$$\forall\ x, \widetilde{x} \in X \ : \ \|F(x) - F(\widetilde{x})\| \leq c\|x - \widetilde{x}\|.$$

2) For mappings of metric spaces, «ε-δ-continuity» (continuity in terms of neighborhoods) is equivalent to sequential continuity (in terms of sequences).

1.9.7 Proposition. *Let X, Y are normed spaces, $F : X \to Y$ be a linear operator. The following are equivalent.*

(I) *F is bounded.*

(II) *F is Lipschitz continuous.*

(III) *F is continuous.*

(IV) *F is continuous in some point $x_0 \in X$.*

Proof. Obviously, (II)⇒(III)⇒(IV). Let's check that (I)⇒(II) and (IV)⇒(I).

If (I) holds, then as c from the definition of Lipschitz continuity one can take C from the definition of boundedness: $\|F(x) - F(\widetilde{x})\| = \|F(x - \widetilde{x})\| \leq C\|x - \widetilde{x}\|$, so (II) is true.

To prove (IV)⇒(I), we first show that from (IV) it follows that the operator F is (sequentially) continuous at zero. If $x_k \to \mathbf{0}$, than $x_k + x_0 \to x_0$, so $F(x_k) = F(x_k + x_0 - x_0) = F(x_k + x_0) - F(x_0) \to F(x_0) - F(x_0) = \mathbf{0}$, and this is continuity at zero. In terms of neighborhoods this means that for $\varepsilon = 1$ there exists $\delta > 0$ such that $\|x\| \leq \delta$ implies $\|F(x)\| \leq \varepsilon = 1$. Then for arbitrary x we get using linearity of F and positive homogeneity of the norm: $\|F(x)\| = \|F\left(\frac{\|x\|}{\delta}\frac{\delta x}{\|x\|}\right)\| = \|\frac{\|x\|}{\delta}F\left(\frac{\delta x}{\|x\|}\right)\| = \frac{\|x\|}{\delta}\|F\left(\frac{\delta x}{\|x\|}\right)\| \leq \frac{\|x\|}{\delta} \cdot 1 = \frac{1}{\delta}\|x\|$, and thus for C from the definition of boundednes one can take number $\frac{1}{\delta}$. □

1.9.8 Definition. *Let $(X, \|\cdot\|_X)$ and $(Y, \|\cdot\|_Y)$ be normed vector spaces. The **space of bounded linear operators** from X to Y is denoted by*

$$\mathcal{L}(X, Y) := \{F : X \to Y \ : F \textit{ is linear and bounded.}\}$$

And

$$\mathcal{L}(X) := \mathcal{L}(X, X).$$

For a bounded linear operator $F : X \to Y$ a non-negative number

$$\|F\| := \|F\|_{\mathcal{L}(X,Y)} := \inf\{C \ : \ \|F(x)\|_Y \leq C\|x\|_X \quad \text{for all } x \in X\}$$

*is called the the **operator norm** of F.*

Next, to shorten the notation, instead of the $\|\cdot\|_X$ and $\|\cdot\|_Y$ we write $\|\cdot\|$, assuming that it is clear from the context in which space we are talking about the norm.

1.9.9 Theorem (Alternative expressions for the operator norm)**.** *Let X, Y be normed vector spaces, $X \neq \{\mathbf{0}\}$, $F : X \to Y$ is linear. Then*

$$\|F\| = \sup_{x \neq \mathbf{0}} \frac{\|F(x)\|}{\|x\|} = \sup_{\|x\| \leq 1} \|F(x)\| = \sup_{\|x\| < 1} \|F(x)\| = \sup_{\|x\| = 1} \|F(x)\|.$$

Proof. Checking these equalities is not too difficult, and we will actually use only some of them, so we will check only the first three, most frequently used, equalities. Let us denote the four quantities whose equality we plan to prove by N_1, N_2, N_3.

In case $N_2 < +\infty$ from the definition of N_2 it immediately follows $\forall x \in X\ \|F(x)\| \leq N_2\|x\|$, then by definition of N_1 we have $N_1 \leq N_2$. In case $N_2 = +\infty$ the last inequality is obviously also true.

For all $x \neq \mathbf{0}$ due to the homogeneity of F and the positive homogeneity of the norm, we have

$$\frac{\|F(x)\|}{\|x\|} = \frac{1}{\|x\|}\left\|F\left(\|x\| \cdot \frac{x}{\|x\|}\right)\right\| = \frac{\|x\|}{\|x\|}\left\|F\left(\frac{x}{\|x\|}\right)\right\| = \left\|F\left(\frac{x}{\|x\|}\right)\right\| \leq N_3.$$

Passing to the supremum in the final inequality we obtain $N_2 \leq N_3$.

In case $N_1 < +\infty$ for any $C > N_1$ for all $x \in X$ it holds $\|F(x)\| \leq C\|x\|$. Then for all x from the closed unit ball (i.e. $\|x\| \leq 1$) it turns $\|F(x)\| \leq C\|x\| \leq C$. Then the inequality is also true for the supremum over all x from the ball: $N_3 \leq C$. Due to the arbitrariness of C, it follows $N_3 \leq N_1$. In case $N_1 = +\infty$ the last inequality is obviously also true.

So we get $N_1 \leq N_2 \leq N_3 \leq N_1$, it means that $N_1 = N_2 = N_3 = N_1$. □

1.9.10 Remark. For any bounded operator, the infimum in the definition of its norm is a minimum, *that is, it is realized as a value (at least) at one of the points of the closed unit ball:*

$$\|F\| = \min\{C \ : \ \|F(x)\| \leq C\|x\| \quad \text{for all } x \in X\}$$

Indeed, by definition of the operator norm $\|F(x)\| \leq C\|x\|$ for any $C > \|F\|$ and any $x \in X$. Than for any (fixed) $x \in X$ passing to the infimum in the previous inequality we obtain

$$\|F(x)\| \leq \|F\| \cdot \|x\|. \tag{1.21}$$

This means $N_1 = N_2$.

1.9.11 Exercise (a little difficult)**.** Show that the suprema in the formulas for the norm, generally speaking, are not maxima.

2 Compactness and separability.

2.1 Compactness in normed spaces

2.1.1 Definition. *A subset E of a topological space X is* ***compact*** *if every open cover of E has* ***a finite subcover*** *of E.*

The subset E of a topological space X is ***relatively compact*** *(or* ***precompact****)* ***in*** *X if its closure $\overline{E} = \mathrm{Cl}_X E$ in X is compact.*

2.1.2 Examples.

1) A closed segment $[a, b]$ is compact (by Heine-Borel theorem). An open interval (a, b) is relatively compact in $\mathbb{R}$, and it is not relatively compact in itself. Thus, the *property of relative compactness* is conditional, it *depends on the enclosing space.*

2) Any finite set, any set consisting of a convergent sequence and its limit are compacts in any topological space.

2.1.3 Reminder. Some basic properties of compact sets are:

(i) *a compact subset of a Hausdorff space is closed;*

Proof. *Let p be a limit point of K such that $p \notin K$. Then $K \subseteq X \setminus \{p\} = \bigcup_{n=1}^{\infty} X \setminus \overline{B}(p, 1/n)$. The open cover $\{X \setminus \overline{B}(p, 1/n)\}$ contains a finite subcover and, consequently, $X \setminus \overline{B}(p, 1/N)$ for some N and $B(p, 1/N) \cap K = \varnothing$ and p is not a limit point of K.* □

(ii) *a closed subset of any compact set is compact;*

(iii) *image of a compact set under any continuous map is compact;*

(iv) *every continuous function on a compact set is uniformly continuous and it attains their maximum and minimum.*

In metric spaces X, a useful description of compact sets E can be given in terms of ε-nets.

2.1.4 Definition. *In a metric space X a $N \subseteq X$ is ε-**net** of $E \subseteq X$ for some $\varepsilon > 0$ if for every $x \in E$ there exists $y \in N$ such that $\rho(x, y) < \varepsilon$.*

Equivalently, N is an ε-net of E if E can be covered by balls of radius ε centered at points in N.

2.1.5 Examples. 1) For any $\varepsilon > 0$ $\mathbb{Q}$ is an ε-net for $\mathbb{R}$.

2) for any open ball $B_r(a)$ in an arbitrary metric space the single-point set $\{a\}$ is a r-net.

2.1.6 Definition. *A subset E of a metric space X is called **totally bounded** if for every $\varepsilon > 0$, there exists a finite ε-net of E.*

Reminder. *A set K in metric space is compact if and only if K is sequentially compact (i.e. if every sequence of points of K has a convergent in K subsequence, $\forall \{x_n\} \subseteq K \quad \exists (n_k) \ : \ x_{n_k} \to x \in K$.)*

2.1.7 Exercise. Show that a *set E in metric space X is relatively compact if and only if every sequence of points of E has a convergent in X subsequence,* $\forall \{x_n\} \subseteq E \quad \exists (n_k) \ : \ x_{n_k} \to x \in X$.)

As a consequence, precompact sets in metric spaces are always totally bounded. And totally bounded sets in metric spaces are always bounded. The converse is true in all finite dimensional normed spaces:

2.1.8 Theorem (Heine-Borel)**.**

Let E be a subset of finite dimensional normed space X. Then

(I) *E is compact if and only if E is closed and bounded;*

(I) *E is precompact if and only if E is bounded;*

Actually, the classical Heine-Borel theorem is the statement for the specific space $X = \mathbb{R}^n$. But as we know, all finite dimensional normed spaces X are isomorphic to $\mathbb{R}^n$, so the general result is also true.

The first main question.

Is the compactness criterion for $\mathbb{R}^n$ still true for infinite-dimensional spaces? (Is it possible to choose a convergent subsequence from any bounded sequence? Are closed balls compact in normed spaces?)

In infinite-dimensional normed spaces, Heine-Borel theorem fails. For example, an orthonormal basis $\{e_k\}$ of l^2 is a closed bounded set but it is not precompact, because it does not have a convergent subsequence (as $\|x_k - x_j\| = \sqrt{2}$ for $k \neq j$).

2.1.9 Lemma («the almost perpendicular» lemma, Riesz Lemma)**.** *Let X be a normed vector space and let $Y \leq X$ be a closed linear subspace that is not equal to X. For any $\delta \in (0,1)$ there exists a vector $x \in X$ such that*

$$\|x\| = 1, \quad \inf_{y \in Y} \|x - y\| \geq 1 - \delta.$$

Proof. Let $x_0 \in X \setminus Y$. Then $d := \inf_{y \in Y} \|x_0 - y\| > 0$ because Y is closed (and so $x_0 \in \operatorname{Int}(X \setminus Y)$). Choose $y_0 \in Y$ such that

$$\|x_0 - y_0\| \leq \frac{d}{1 - \delta},$$

and define $x := \|x_0 - y_0\|^{-1} (x_0 - y_0)$. Then $\|x\| = 1$ and

$$\|x - y\| = \frac{\|x_0 - y_0 - \|x_0 - y_0\| \cdot y\|}{\|x_0 - y_0\|} =$$
$$\frac{\|x_0 - (y_0 + \|x_0 - y_0\| \cdot y)\|}{\|x_0 - y_0\|} \geq \frac{d}{\|x_0 - y_0\|} \geq 1 - \delta,$$

for all $y \in Y$. $\square$

2.1.10 Theorem (Riesz)**.** *Let X be a normed vector space and let $B = \{x \in X \ : \ \|x\| \leq 1\}$ be closed unit ball in X. Then the following are equivalent.*

(i) $\dim X < \infty$.

(ii) *B is compact.*

Proof. (i) implies (ii) due to Heine-Borel theorem.

(ii)$\Rightarrow$ (i) ?

Let $\dim X = \infty$. We consider any $x_1 \in X$ such that $\|x\|_1 = 1$, and let $X_1 = \operatorname{span}\{x_1\}$. X_1 is finite dimensional hence it is closed subspace of X and $X_1 \neq X$ because $\dim X_1 < \dim X = \infty$. So lemma 2.1.9 guarantees the existence of such x_2 that

$$x_2 \in X \setminus X_1, \quad \|x_2\| = 1, \quad \mathrm{dist}(x_2, X_1) \geq \frac{1}{2}.$$

Taking $X_2 := \mathrm{span}\{x_1, x_2\}$ we get a finite dimensional (hence, closed) subspace of X. As a result of repeating the actions at the next step, we have $X_k := \mathrm{span}\{x_1, \dots x_k\}$ that is a finite dimensional (hence, closed) subspace of X, and by lemma 2.1.9 there exists such x_{k+1} that

$$x_{k+1} \in X \setminus X_k, \quad \|x_{k+1}\| = 1, \quad \mathrm{dist}(x_{k+1}, X_k) \geq \frac{1}{2}.$$

As a result of this procedure, we get a sequence $(x_k) \subseteq B$ such that $\mathrm{dist}(x_k, \mathrm{span}\{x_1, \dots, x_k\}) \geq \frac{1}{2}$. So

$$\|x_k - x_m\| \geq \frac{1}{2} \quad \text{for any} \quad k, m \in \mathbb{N},\ k \neq m.$$

The latter inequality is also true for an arbitrary subsequence of (x_k) and it is incompatible with its convergence in X. $\square$

An analogue of the Heine-Borel theorem in an arbitrary metric space is the following Hausdorff criterion. Before formulating the criterion, let's note an almost obvious fact.

2.1.11 Lemma. *If the Cauchy sequence (x_k) in a metric space X contains a convergent subsequence $x_{k_j} \to x$, then $x_k \to x$.*

Proof. Since (x_k) is a Cauchy sequence, then for every $\varepsilon > 0$ there exists N such that $\rho(x_k, x_m) < \varepsilon$ at $k, m \geq N$. For any fixed $k \geq N$ and $j \geq N$ (so $k_j \geq j \geq N$) we have $\rho(x_k, x_{k_j}) < \varepsilon$. Because $x_{k_j} \to x$, going to the limit (as $j \to \infty$) we get $\rho(x_k, x) \leq \varepsilon$, which proves that $x_k \to x$. $\square$

Remark. *For an arbitrary Cauchy sequence (x_k) in arbitrary metric space X, the following statements are equivalent.*

(i) $x_k \to x$;
(ii) $x_{k_j} \to x$ *for some subsequence* x_{k_j};
(iii) $x_{k_j} \to x$ *for any subsequence* x_{k_j}.

(Implication (i)⇒(iii) is a well-known fact, implication (iii)⇒(ii) is trivial, implication (ii)⇒(i) we have just proved).

2.1.12 Theorem (Hausdorff compactness criterion)**.** *Let* (X, ρ) *be a metric space and let* $E \subseteq X$*. Then the following are equivalent.*
(i) E *is compact.*
(ii) E *is complete and totally bounded.*

Proof. (i)⇒ (ii). *Completeness.* Let (x_k) be a Cauchy sequence in a compact set E. Due to the sequential compactness of E, (x_k) has a convergent subsequence. By lemma 2.1.11 (x_k) converges too.

Totally boundedness. Due to compactness, for any $\varepsilon > 0$, the cover $\{B_\varepsilon(x)\}_{x\in E}$ of the set E with open balls contains a finite subcover $\{B_\varepsilon(x_1), \ldots, B_\varepsilon(x_n)\}_{x\in E}$. Then $\{x_1, \ldots, x_n\}$ is a finite ε-net for E.

(ii)⇒ (i) *Sequential compactness.* let E be completely and totally bounded. Let's consider an arbitrary sequence (x_k) and prove that a convergent subsequence can be extracted from it. To do this, let's take some positive sequence $\varepsilon_k \to 0$ and consider a finite ε_1-net N_1 (according to the condition it exists). If you build balls with centers in N_1 and radius ε_1, then each point of the set E falls into at least one of these balls. Since there are a finite number of balls, there are infinitely many members of the sequence (x_k) in one of them. Let's denote its center y_1. Next, let's take the ε_2-net N_2 and consider balls of radius ε_2 with centers at its points. As before, one of these balls contains infinitely many elements of the sequence (x_k) contained in $B_{\varepsilon_1}(y_1)$, let the center of this ball be at y_2. Continuing in this way, we get a sequence (y_k) of centers of balls $(B_{\varepsilon_k}(y_k))$. An infinite number of elements of the (x_k) fall into the intersection of any finite number of the balls. So one can select

$$x_{k_1} \in B_{\varepsilon_1}(y_1), \qquad x_{k_2} \in B_{\varepsilon_2}(y_2) \cap B_{\varepsilon_1}(y_1),\ k_2 > k_1,$$

and in general

$$x_{k_j} \in \bigcap_{i=1}^{j} B_{\varepsilon_i}(y_i), \quad k_j > k_{j-1} > \cdots > k_1.$$

Since under the condition $j \leq l$ both elements x_{k_j}, x_{k_l} belong to the ball $B_j(y_j)$ then

$$\rho(x_{k_j}, x_{k_l}) \leq \rho(x_{k_j}, y_{k_j}) + \rho(y_{k_j}, x_{k_l}) < 2\varepsilon_j.$$

Hence, the subsequence (x_{k_j}) is a Cauchy sequence, and because of the completeness of E, this sequence converges to some $x \in E$. So sequential compactness is proved. □

2.1.13 Corollary (relative compactness criterion in complete spaces, Hausdorff precompactness criterion)**.**

Let (X, ρ) be a complete metric space and let $E \subseteq X$. Then the following are equivalent.

(i) *E is precompact.*

(ii) *E totally bounded.*

Compact sets are almost finite dimensional. This heuristics, which is made rigorous in the following result, underlies many arguments in analysis:

2.1.14 Theorem (Compactness criterion for infinite-dimensional spaces, approximation by finite dimensional subspaces)**.**

A subset E of a Banach space X is totally bounded (precompact) if and only if E is bounded and, for every $\varepsilon > 0$, there exists a finite dimensional subspace Y of X which forms an ε-net of E.

Proof. *Necessity.* Let E be totally bounded. Choose a finite ε-net N_ε of E; then the subspace $Y = \operatorname{span} N_\varepsilon$ is finite-dimensional and forms an ε-net of E.

Sufficiency. Let B_X, B_Y are balls with centers at zero of unit radius in the spaces X and Y, respectively. Since E is bounded, $E \subseteq rB_X$ for some (finite) radius r. Since Y is an ε-net of E, it follows that $(r+\varepsilon)B_Y$ is also an ε-net of E ($\forall\, x \in E \quad \|x\| \leq r$, and $\exists\, y \in Y$ such that $\|x - y\| < \varepsilon$, so $\|y\| \leq \|x\| + \|y - x\| < r + \varepsilon, \;\Rightarrow\; y \in (r+\varepsilon)B_Y$). Further, since Y is finite-dimensional, the set $(r+\varepsilon)B_Y$ is totally bounded by Heine-Borel theorem. So we have found a totally bounded net of E. Therefore E itself is totally bounded (and precompact). (Why?) $\square$

Now we want to get the criterion of compactness in space $C(K)$. One remark will be usefull.

2.1.15 Remark. Let K be any compact. Let consider the space all bounded functions on K:

$$l^\infty(K) := \{f : K \to \mathbb{C} \;:\; f \text{ is bounded}\},$$

and a norm on it:

$$\|f - g\|_{\sup} = \sup_{t\in K} |\, f(t) - g(t)\,|\,.$$

Then 1) $l^\infty(K)$ *is complete normed space,* and

2) $C(K)$ *is a closed subspace of* $l^\infty(K)$*.*

The completeness of space $l^\infty(K)$ can be proved as well as the completeness of space $C(K)$. The closeness of $C(K)$ in $l^\infty(K)$ follows from completeness of $C(K)$.

2.1.16 Definition. *A subset* $\mathfrak{F} \subseteq C(X, Y)$ *is called* ***equi-continuous*** *if, for every* $\varepsilon > 0$*, there exists a constant* $\delta > 0$ *such that, for all* $x, \widetilde{x} \in X$ *and all* $f \in \mathfrak{F}$*,*

$$\rho_X(x, \widetilde{x}) < \delta \quad \Rightarrow \quad \rho_Y(f(x), f(\widetilde{x})) < \varepsilon$$

2.1.17 Theorem (Arzela-Ascoli, a compactness criteria for $C(K)$). *For any metric compact K a subset $E \subseteq C(K)$ is precompact if and only if E is bounded and equicontinuous.*

So, *for $E \subseteq C(K)$*

$$\left(E \quad is \quad precompact \quad \Leftrightarrow \quad E \quad is \quad \begin{cases} bounded, \\ equicontinuous. \end{cases} \right)$$

Proof. In this proof $\|\cdot\| := \|\cdot\|_{\sup}$, and ρ is the metric of the compact K.

$\Rightarrow$. Thus assume E is precompact. Let's fix $\varepsilon = 1$. By corollary 2.1.13 E is totally bounded, so for E there exists a finite ε-net $\{x_1, \ldots, x_n\} \subset C(K)$, which means that for every $x \in E$ there is an x_j, with $\|x - x_j\| < \varepsilon = 1$, so for any $t \in K$

$$|\, x(t) \,| \leq |\, x_j(t) \,| + |\, x(t) - x_j(t) \,| \leq \|x_j\| + \|x - x_j\| < M_0 + 1,$$

where $M_0 = \max\{\|x_1\|, \ldots, \|x_n\|\}$. and hence, E is *bounded* (by a number $M_0 + 1$).

Now let the number $\varepsilon > 0$ be arbitrary, and $\{x_1, \ldots, x_n\}$ is still ε-net for E in $C(K)$. For each of the functions $x_1, \ldots, x_n$, there is such a number $\delta_j > 0$ that

$$|\, x_j(t') - x_j(t) \,| < \varepsilon \quad \text{if} \quad \rho_K(t', t) < \delta_j.$$

Lets take $\delta := \min\{\delta_1, \ldots, \delta_n\}$. Then if x is an arbitrary element of E, and x_j is such that $\|x - x_j\| < \varepsilon$. Then

$$\begin{gathered} |\, x(t') - x(t) \,| \leq \\ |\, x(t') - x_j(t') \,| + |\, x_j(t') - x_j(t) \,| + |\, x_j(t) - x(t) \,| < \\ \varepsilon + |\, x_j(t') - x_j(t) \,| + \varepsilon. \end{gathered}$$

If $\rho_K(t, t') < \delta$, then one has $|\, x(t') - x(t) \,| < 3\varepsilon$. In that way E is *equicontinuous*.

$\Leftarrow$ Taking into account remark (2.1.15), it is sufficient to prove that E is precompact in $l^\infty(K)$.

As the space $l^\infty(K)$ is complete, by corollary (2.1.13) it is sufficient to check the total boundedness of the set E in $l^\infty(K)$. Let's take an $\varepsilon > 0$ and find for it a $\delta > 0$ from the definition of equicontinuity. Since K is a metric compact (which means it is total bounded by the Hausdorff criterion 2.1.12), then there is a finite $\frac{\delta}{3}$-net $\{t_i\}_{i=1}^m$ for K. let

$$K_1 = B_{\frac{\delta}{3}}(t_1),$$
$$K_2 = B_{\frac{\delta}{3}}(t_2) \setminus K_1,$$
$$\dots$$
$$K_m = B_{\frac{\delta}{3}}(t_m) \setminus (K_1 \cup \dots \cup K_{m-1}).$$

Then $K = \cup_{i=1}^m K_i$ (because $K = \cup_{i=1}^m B_{\frac{\delta}{3}}(t_i)$, besides $K_1, \dots, K_m$ are disjunct, and $\rho(t,t') < \frac{2\delta}{3}$ for any i and every $t, t' \in K_i$. Therefore, by choice δ, if $t, t' \in K_i$, than $|\, x(t) - x(t') \,| < \varepsilon$ for all $x \in E$. Let χ_i be the characteristic function of K_i (so $\chi_i(t)$ =1, if $t \in K_i$, and $\chi_i(t)$ =0, if $t \notin K_i$). Without loss of generality, all sets $K_1, \dots, K_m$ are not empty (otherwise we will simply renumber them, throwing out the extra ones). The space $Y = \operatorname{span}\{\chi_1, \dots, \chi_m\}$ is finite-dimensional. By the theorem (2.1.14), the proof will be complete if we show that Y forms a 2ε-net for E. In each K_i, we fix an arbitrary point t_i. Let's define $y(t) := \sum_{i=1}^m x(t_i)\chi_i(t) \in Y$. Let $t \in K$ be arbitrary. Then t belongs to some K_{i*}, and we get

$$|\, x(t) - y(t) \,| = |\, x(t) - x(t_{i*})\chi_{i*}(t) \,| = |\, x(t) - x(t_{i*}) \,| < \varepsilon.$$

Due to the arbitrariness of t, we get $\|x - y\| \leq \varepsilon < 2\varepsilon$.

□

Arzela-Ascoli theorem implies that the set of differentiable functions f with $\|f'\| \leq 1$ is compact in $C[0,1]$.

2.1.18 Proposition (Compactness in l^p). *A subset $E \subseteq l^p$, $p \in [1, +\infty)$ is precompact if and only if E is bounded and has uniformly decaying tails, i.e.*

$$\sum_{k>n} |x_k|^p \le \varepsilon_n \to 0, \quad \textit{for all } x = (x_k) \in E,$$

where $\varepsilon_n \ge 0$ is some sequence of numbers (that does not depend on x).

2.1.19 Exercise. Prove compactness in l^p using theorem 2.1.14.

2.1.20 Example. The *Hilbert cube* is the following subset of l^2:

$$Q = \{(x_k) \in l^2 \ : \ |x_k| \le \frac{1}{k} \quad \text{for all} \quad k \in \mathbb{N}\}.$$

By 2.1.18, the Hilbert cube is compact.

There are several variants of compactness criteria in L^p spaces: Kolmogorov criterion, Lebesgue-Vitali criterion, and others. The topic of searching for compactness criteria in this space is quite popular, and research continues at the present time. We will mention one of these criteria for L^1 without proof.

2.1.21 Proposition (Compactness in L^1). *A subset $E \subseteq L^1[0,1]$, is precompact if and only if E is bounded and* ***uniformly bounded on average****, i.e. for every $\varepsilon > 0$ there exists $\delta > 0$ such that*

$$|\tau| \le \delta \quad \Rightarrow \quad \int_0^1 |f(t+\tau) - f(t)| dt \le \varepsilon \quad \textit{for all} \quad f \in E.$$

2.1.22 Exercises (Stability of compactness and precompactness).
Prove that compactness in normed spaces is stable under linear operations:

(i) *If A, B are precompact sets in a normed space, then Minkowski sum $A + B$ is precompact;*

(ii) *If A is a precompact subset of X and $F \in \mathcal{L}(X \to Y)$ then $F(A)$ is a precompact set in Y.*

2.1.23 Reminder. *Let X be a linear vector space. A set $E \subseteq X$ is* ***convex*** *if for all $x, y \in E$, and for all $\lambda \in [0,1]$ one has*

$$\lambda x + (1-\lambda)y \in E$$

*(*or for the Minkowski sum, the inclusion is true $\lambda E + (1-\lambda)E \subseteq E$*).*

A function $f : X \to \mathbb{R}$ is ***convex*** *if for all $x, y \in X$, and for all $\lambda \in [0,1]$ one has*

$$f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda)f(y).$$

The ***convex hull*** *of a subset E of a linear vector space X is defined to be the smallest convex set that contains E. The convex hull of E is denoted* $\operatorname{conv}(E)$*.*

A ***convex combination*** *of vectors $x_1, \ldots, x_n$ in a linear vector space X is any vector of the form*

$$x = \sum_{k=1}^{n} \lambda_k x_k$$

where $\lambda_k \geq 0$ are some numbers such that

$$\sum_{k=1}^{n} \lambda_k = 1.$$

Remark. $\operatorname{conv} E$ *coincides with the set of all convex combinations of a finite number of vectors from E.*

2.1.24 Exercises (Convex hull of compact sets)**.**

(i) Show that *the convex hull of a precompact set in a normed space is a precompact set.*

(ii) Show that *the closure of the convex hull of a compact set in a normed space is compact.*

(iii) Construct an example showing that the convex hull of a compact set in a normed space does not need to be compact.

An important characteristic of a compact is its epsilon-entropy.

2.1.25 Definition. *Let's K be a metric compact. The function $\mathcal{N}_k$: $(0,+\infty) \to \mathbb{N}$,*

$$\mathcal{N}_K(\varepsilon) = \min\{j : \ \exists \ \varepsilon\text{-net of } j \text{ points for } K\}$$

*is called the **epsilon-entropy** of the compact K.*

2.1.26 Examples. 1) $K = [0,1]$, $\mathcal{N}_K(\varepsilon) = \lfloor \frac{1}{2\varepsilon} \rfloor + 1$ and so $\mathcal{N}_K(\varepsilon) \sim \frac{C_1}{\varepsilon}$ as $\varepsilon \to 0+$.

2) $K = [0,1]^2 \big(= [0,1] \times [0,1]\big)$. Then $\mathcal{N}_K(\varepsilon) \sim \frac{C_2}{\varepsilon^2}$ as $\varepsilon \to 0+$.

Let for some compact K one has $\mathcal{N}_K(\varepsilon) \sim \frac{C}{\varepsilon^p}$, then $p \geq 0$ and

$$\frac{\ln \mathcal{N}_K(\varepsilon)}{\ln \frac{1}{\varepsilon}} = \frac{\ln \left(\frac{C}{\varepsilon^p} + o\left(\frac{C}{\varepsilon^p} \right) \right)}{\ln \frac{1}{\varepsilon}} = \frac{\ln \frac{1}{\varepsilon^p} + \ln \left(C + o(1) \right)}{\ln \frac{1}{\varepsilon}} \sim p \quad (\ \varepsilon \to 0+).$$

if there is a limit of the ratio $\frac{\ln \mathcal{N}_K(\varepsilon)}{\ln \frac{1}{\varepsilon}}$, then this limit is called the Minkowski dimension (or, sometimes, fractal dimension) of K:

$$\dim_M(K) = \lim_{\varepsilon \to 0+} \frac{\ln \mathcal{N}_K(\varepsilon)}{\ln \frac{1}{\varepsilon}}.$$

So, $\dim_M\left([0,1]^n\right) = n$.

2.1.27 Exercises (rather complicated)**.**

1) For a smooth k-dimensional manifold S $\quad \dim_M(S) = k$.

2) For the Cantor set

$$C = \left\{ \sum_{i=1}^{\infty} \frac{p_i}{3^i} \ : \ p_i \in \{0,2\} \right\}$$

$\dim_M(C) = \ln_3 2$.

3) Find or estimate $\mathcal{N}_E(\varepsilon)$ for a compact E in $C[0,1]$,

$$E = \{ f \in C[0,1] \ : \ \|f\| = 1, \ \forall x, y \in [0,1] \quad |f(x) - f(y)| \leq |x - y| \}.$$

2.2 Stone–Weierstrass Theorem

The classical Weierstrass theorem is as follows.

2.2.1 Theorem (Weierstrass)**.**
The set of all algebraic univariate polynomials is dense in $C[a,b]$.

This theorem is not directly transferred even to the case of complex-valued continuous functions in a closed disk: if $(P_k)_k \in \mathcal{P}(z)$ and $P_k \to f$ in $C(\overline{D})$, where $D = \{z \in \mathbb{C} \ : \ |z| < 1\})$, then $P_k \rightrightarrows f$ on D that implies f is analytic in D. So $\overline{\mathcal{P}(z)} \neq C(\overline{D})$.

Let's for $x, y \in C(K)$:

$$(xy)(t) := x(t) \cdot y(t),$$
$$(x \vee y)(t) := \max\{x(t), y(t)\}, \quad (x \wedge y)(t) := \min\{x(t), y(t)\}$$

and denote $\mathbf{1}$ function that is identically equal to one by K.

2.2.2 Definition. *A linear subspace Y of $C(K)$ is called a **subalgebra** if from $x, y \in Y$ follows $x \cdot y \in Y$.*

*A subset $Z \subseteq C(K)$ is called a **sublattice** if from $x, y \in Z$ follows $x \vee y, x \wedge y \in Z$.*

2.2.3 Definition. *For $X \subseteq C(K)$ it is said that X **separates the points** if for all $t, s \in K$ such that $t \neq s$ there exists a function $x \in X$ such that $x(t) \neq x(s)$.*

*For $X \subseteq C_{\mathbb{R}}(K)$ it is said that X **strongly separates the points** if for all $t, s \in K$ such that $t \neq s$ and for all $a, b \in \mathbb{R}$ there exists a function $x \in X$ such that $x(t) = a$ and $x(s) = b$.*

Remark. *If Z a subspace with $\mathbf{1}$ of $C_{\mathbb{R}}(K)$ and Z separates the points of K, than it strongly separates the points of K.*

Proof. Let $t, s \in K$, $t \neq s$, $a, b \in \mathbb{R}$ and $x_0 \in C(K)$ such that $x_0(t) \neq x_0(s)$. A function $x \in C(K)$ for which $x(t) = a$ and $x(s) = b$ we search in the form $x = A \cdot x_0 + B \cdot \mathbf{1}$. So we have a system:

$$\begin{cases} A \cdot x_0(t) + B \cdot \mathbf{1} = a, \\ A \cdot x_0(s) + B \cdot \mathbf{1} = b \end{cases} \tag{2.1}$$

for A and B. Its determinant is $x_0(t) - x_0(s) \neq 0$, so the system is solvable, the required coefficients A and B exist. $\square$

2.2.4 Theorem (Kakutani, Crane). *A strongly separating sublattice is dense in $C_{\mathbb{R}}(K)$.*

Proof. Let Z be a strongly separating sublattice of $C(K) := C_{\mathbb{R}}(K)$. $x \in C(K)$ and $\varepsilon > 0$ are arbitrary. For any $t, s \in K$ (including matching ones) let $x_{ts} \in Z$ such that

$$x_{ts}(t) = x(t) \quad \text{and} \quad x_{ts}(s) = x(s). \tag{2.2}$$

Due to the continuity of all the functions under consideration for any $t, s \in K$ there is a neighborhood B_{st} of the point t such that

$$\forall\ u \in B_{st} \quad |\, x_{st}(u) - x(u)\,| < \varepsilon. \tag{2.3}$$

The collection $\{B_{st}\}_{t \in K}$ forms an open cover of K that has a finite subcover $\{B_{st_1}, \ldots, B_{st_n}\}$. Let

$$y_s := x_{st_1} \vee \cdots \vee x_{st_n} \in Z,$$

moreover, from (2.2) we have

$$y_s(s) = x(s).$$

For any $u \in K$ there is such an $i \in \{1, \ldots, n\}$ that $u \in B_{st_i}$, so from (2.3) we get

$$y_s(u) \geq x_{st_i}(u) \geq x(u) - \varepsilon. \tag{2.4}$$

Now we consider a neighborhood B_s of s such that

$$\forall u \in B_s \quad |\, y_s(u) - x(u)\,| < \varepsilon. \tag{2.5}$$

From the open cover $\{B_s\}_{s\in K}$, we select the finite subcover $\{B_{s_1}, \dots, B_{s_m}\}$ and define

$$y := y_{s_1} \wedge \dots \wedge y_{s_m} \in Z.$$

Then for an arbitrary u from K we have:
1) $y(u) \geq x(u) - \varepsilon$ because of (2.4);
2) there exists $j \in \{1, \dots, m\}$ such that $u \in B_{s_j}$, so

$$y(u) \leq y_{s_j}(u) \leq x(u) + \varepsilon.$$

(from (2.5)). 1) and 2) means that $\|y - x\| < \varepsilon$. □

2.2.5 Examples.

1) The set of all algebraic univariate polynomials $\mathcal{P}$ is not a sublattice, because $p, q \in \mathcal{P} \not\Rightarrow p \vee q, p \wedge q \in \mathcal{P}$. For $f(x) = x, g(x) = -x$, $f, g \in \mathcal{P}$, but $f \vee g = |\,x\,|\,, f \wedge g = -\,|\,x\,| \notin \mathcal{P}$.

2) The set Y_1 of all piecewise linear functions,

$$Y_1 = \Big\{ f \;:\; \exists (x_i)_{i=0}^n \;:\; a = x_0 < x_1 < \dots < x_n = b,$$
$$f\big|_{[x_{i-1}, x_i]} \quad \text{is an affine function } \forall i \Big\}$$

is a sublattice, so Y_1 is dense in $C[a, b]$.

2.2.6 Theorem (The Stone–Weierstrass Theorem, $\mathbb{R}$-case)**.**

Let K be a nonempty compact space and let Y be a subalgebra of real $C(K)$. If

i) $\mathbf{1} \in Y$*, and*

ii) Y *separates points,*

then Y is dense in $C(K)$.

Proof. First of all, Y strongly separates points due to remark from 2.2.3 because Y is a subspace. The second, lets check that its closure $\operatorname{Cl} Y$ (in $C[a, b]$) is a sublattice.

1) $\operatorname{Cl} Y$ is a linear subspace: $x, y \in \operatorname{Cl} Y \Rightarrow Ax + By \in \operatorname{Cl} Y$.
2) Is it true that $x \in Y \Rightarrow |\, x \,| \in \operatorname{Cl} Y$?
Or, Is it true that for $\varphi(t) := |t| \in Y \Rightarrow |\, \varphi(x) \,| = |\, x \,| \in \operatorname{Cl} Y$?
$|\, t \,| = \sqrt{1 - (1 - t^2)}$. For $s \in [0, 1]$ from Taylor's formulas

$$\sqrt{1+s} = 1 + \frac{1}{2}s + \cdots + C^n_{\frac{1}{2}} s^n + r_n(s), \quad \text{where}$$
$$C_{\frac{1}{2}}k = \frac{\frac{1}{2}(\frac{1}{2}-1)\dots(\frac{!}{2}-k+1)}{k!}$$

and $r_n(s) \rightrightarrows 0$ on [0,1]. For $x \in Y$ with $\|x\| \leq 1$ then $\varphi_n(x(u))$ is a composition of a algebraic polynomial with x so $\varphi_n(x(u)) \in Y$. Because $\varphi_n(x(t)) \rightrightarrows \varphi(x(t)) \in \operatorname{Cl} Y$. If $x \in Y$ and $\|x\| > 1$, we also have

$$\varphi(x(t)) = \varphi\left(\|x(t)\| \frac{x(t)}{\|x(t)\|}\right) = \|x(t)\| \cdot \varphi\left(\frac{x(t)}{\|x(t)\|}\right) \in \operatorname{Cl} Y.$$

So 2.1) is true. Then for any $x, y \in Y$ $x \vee y = \max\{x, y\} = \frac{|\, x-y \,|+x+y}{2} \in \operatorname{Cl} Y$, and $x \wedge y = \min\{x, y\} = -((-x) \vee (-y)) \in \operatorname{Cl} Y$. Thus, $\operatorname{Cl} Y$ satisfies the conditions of the Kakutani-Crane theorem, hence $\operatorname{Cl} Y$ is dense in $C(K)$, so $C(K) = \operatorname{Cl}\operatorname{Cl}(Y) = \operatorname{Cl}(Y)$, i.e. Y is dense.

$\square$

2.2.7 Theorem (The Stone–Weierstrass Theorem, $\mathbb{C}$-case)**.**

Let K be a nonempty compact space and let Z be a subalgebra of complex $C(K)$. If

i) $\mathbf{1} \in Z$,

ii) *Z separates points,*

iii) $f \in Z \;\Rightarrow\; \overline{f} \in Z$

Then Z is dense in $C(K)$.

Proof. Let consider

$$Y = \{y \in Y \;:\; \forall\, t \in K \; y(t) \in \mathbb{R}\}.$$

Then

$$Y = \{\operatorname{Re} z\}_{z\in Z} = \{\operatorname{Im} z\}_{z\in Z}. \tag{2.6}$$

Why (2.6) is it true? Let's denote $Y_1 = \{\operatorname{Re} z\}_{z\in Z}$. First of all, $y \in Y \Rightarrow y = \operatorname{Re} y \Rightarrow y \in Y_1$. Secondly, $y \in Y_1 \Rightarrow \exists\ z \in Z\ :\ y = \operatorname{Re} z$. From condition (iii) $\bar{z} \in Z$, and Z is an algebra, so $\frac{1}{2}(z+\bar{z}) = y \in Z \ \Rightarrow y \in Y$.

Further, $\{\operatorname{Re} z\}_{z\in Z} = \{\operatorname{Im}(iz)\}_{z\in Z} = \{\operatorname{Im}(z)\}_{z\in Z}$, so (2.6) is it true.

Besides, In addition, we note that

$$\operatorname{span}_{\mathbb{C}} Y = Z, \quad \operatorname{span}_{\mathbb{C}} C_{\mathbb{R}}(K) = C_{\mathbb{C}}(K). \tag{2.7}$$

The set Y_1 is a subalgebra. It contains a unit $\mathbf{1}$ and separates the points (if $s \neq t \in K$, then by (ii) $\exists\ z \in Z\ :\ z(s) \neq z(t) \ \Rightarrow \operatorname{Re} z(s) \neq \operatorname{Re} z(t)$ or $\operatorname{Im} z(s) \neq \operatorname{Im} z(t) \ \Rightarrow \exists\ y \in Y\ : y(t) \neq y(s)$).

So, according to the real theorem of Stone–Weierstrass Y is dense in $C_{\mathbb{R}}(K)$. Hence $\operatorname{span}_{\mathbb{C}} Y = Z$ is dense in $\operatorname{span} C_{\mathbb{R}}(K)$. □

2.3 Dense subsets of Lebesgue spaces

In this section, we investigate dense subsets of spaces $L^p(X,\mu)$, where (X,μ) is a measure space and $1 \le p < \infty$.

2.3.1 Definition. *As a* ***integrable step function*** *we consider here a function $\varphi(\cdot) = \sum_{k=1}^{n} c_k \chi_{E_k}(\cdot)$, where $c_1,\ldots,c_n \in \mathbb{C}$, $E_k \subseteq X$ are measurable, and $\mu(E_k) < +\infty$.*

Remark. *1) An equivalent definition of step functions is obtained if we additionally require the disjunctness of sets E_k.*

2) f is a integrable step function $\Leftrightarrow f \in \operatorname{span}\{\chi_E\ :\ E \subseteq X,\ \mu E < \infty\}$.

2.3.2 Proposition. *The set of step functions is dense in $L^p(X,\mu)$.*

Proof. It is known from the course of mathematical analysis that for any measurable $\mathbb{C}$-valued function f there is such a sequence (f_n) of step functions that

$$f_n \to f \text{ a.e. on } X, \quad \text{and } \forall\, x \in X,\ \forall\, n \in \mathbb{N} \quad |f_n(x)| \le |f(x)|.$$

From the conditions $f \in L^p$, $|\, f_n\,| \le |\, f\,|$ it follows that $f_n \in L^p$. Than by Lebesgue dominated convergence theorem $f_n \to f$ in L^p (since $|\, f_n - f\,|^p \to 0$ a. e. and $|\, f_n - f\,|^p$ is bounded by $(2\,|\, f\,|)^p$, then $\int_X |\, f_n - f\,|^p \to 0 \ \Rightarrow\ \|f_n - f\|_p \to 0 \ \Rightarrow f_n \to f$ in L^p). □

Reminder. For a function f defined on a topological space X the ***support*** of f, denoted by $\operatorname{supp} f$ is the *smallest closed set outside of which the function f vanishes identically,* so

$$\operatorname{supp} f = \mathrm{Cl}\{f \neq 0\}$$

Let's introduce the following notation.

2.3.3 Notation.

$$C_0(\mathbb{R}^n) = \{f \in C(\mathbb{R}^n) \ :\ \operatorname{supp} f \ \text{ is compact}\},$$
$$C_0^\infty(\mathbb{R}^n) = C_0(\mathbb{R}^n) \cap\ C^\infty(\mathbb{R}^n).$$

So C_0^∞ consists of compactly supported (real-)smooth functions.

Reminder. *For any open set $G \subseteq \mathbb{R}^n$ and any compact $K \subseteq G$ there exists $f \in C_0^\infty(\mathbb{R}^n)$ such that*

$$f\Big|_K = 1, \qquad f\Big|_{\mathbb{R}^n \setminus O} = 0, \qquad 0 \le f \le 1 \textit{ everywhere.} \tag{2.8}$$

Functions with such properties are used when constructing a unit partition.

2.3.4 Theorem. *Let X be a subset of $\mathbb{R}^n$, and λ_n is Lebesgue measure with then $C_0^\infty|_X$ is dense in real $L^p(X, \lambda_n)$.*

Proof. 1) We show that the characteristic functions of bounded measurable sets can be approximated in $L^p(\mathbb{R}^n)$ by smooth functions with compact support. Due to the regularity of the Lebesgue measure for arbitrary set E of finite Lebesgue measure for any $\varepsilon > 0$ there exist such a compact K and an open set O, that

$$K \subseteq E \subset O, \quad \lambda_n(O \setminus K) < \varepsilon.$$

Then there exists a function $f \in C_0^\infty(\mathbb{R}^n)$ for satisfying the requirements (2.8) (if you do not require the smoothness function f, but only continuity, then an example of such a function can be found in this way:

$$d := \min_{x \in \mathbb{R}^n \setminus O} \operatorname{dist}(x, K), \quad g(x) := \min\{\operatorname{dist}(x, K), d\}, \quad f(x) := \frac{d - g(x)}{d}).$$

Then we have

$$\|\chi_E - f\|_p^p = \int_X |\, \chi_E - f\,|^p \, d\lambda_n \le \int_{\mathbb{R}^n} |\, \chi_E - f\,|^p \, d\lambda_n =$$
$$\int_{O \setminus K} |\, \chi_E - f\,|^p \, d\lambda_n \le \int_{O \setminus K} 1 d\lambda_n < \varepsilon,$$

Thus, for any bounded measurable E, we obtain $\chi_E \in \mathrm{Cl}_{L^p} C_0^\infty(\mathbb{R}^n)$.

2) If $\lambda_n(E) < \infty$, but E is unbounded, then there is a sequence of bounded measurable sets that splits E:

$$E = \cup_{k=1}^\infty E_k, \quad E_k \text{ are disjoint, measurable.}$$

In this situation $\chi_E = \sum_{k=1}^\infty \chi_{E_k}$, the series converges pointwise and in L^p (since partial sums converge pointwise, and they have a dominating function χ_E in L^p.)

3) Based on the latest estimates and proposition 2.3.2, we obtain

$$\forall\, N \in \mathbb{N} \quad \sum_{k=1}^N \chi_{E_k} \in \mathrm{Cl}\,(C_0|_X) \;\Rightarrow\; \chi_E \in \mathrm{Cl}\,(C_0^\infty|_X) \;\Rightarrow$$
$$\text{any integrable step function} \in \mathrm{Cl}\,(C_0^\infty|_X) \;\Rightarrow L^p(X, \lambda_n) \subseteq \mathrm{Cl}\,(C_0^\infty|_X)\,.$$

□

2.3.5 **Corollary** (on the continuity on average of elements of L^p.). *$p \in [1, +\infty)$, $f \in L^p(\mathbb{R}^n)$, then*

$$\lim_{\delta \to 0+} \left(\sup_{\|\tau\| < \delta} \|f(\cdot + \tau) - f(\cdot)\|_{L^p} \right) = 0.$$

(Compare with (2.1.21))

Proof. Here $\|\cdot\|_p$ means $\|\cdot\|_{L^p(\mathbb{R}^n)}$.

For given $f \in L^p(\mathbb{R}^n)$ and any positive ε there exists $g \in C_0(\mathbb{R}^n)$ such that $\|g - f\|_p < \varepsilon$. The function g is uniformly continuous (see the next exercise), so

$$\forall\, \sigma > 0\ \exists\, \delta > 0\ :\ \forall\, t \in \mathbb{R}^n, \quad \forall\, \tau \in \mathbb{R}^n : \|\tau\| < \delta\ \Rightarrow$$
$$|\, g(t + \tau) - g(t)\, | < \sigma.$$

So if for a fixed τ such that $\|\tau\| < \delta \leq 1$ we define

$$K_\delta := \{ s \in \mathbb{R}^n\ :\ \operatorname{dist}(s, K) < \delta \}$$

than

$$\|g(\cdot + \tau) - g(\cdot)\|_{L^p} = \left(\int_{\mathbb{R}^n} |\, g(t + \tau) - g(t)\,|^p \, d\lambda_n(t) \right)^{1/p} =$$
$$= \left(\int_{K_\delta} |\, g(t + \tau) - g(t)\,|^p \, d\lambda_n(t) \right)^{1/p} \leq \left(\int_{K_\delta} \sigma^p d\lambda_n(t) \right)^{1/p} =$$
$$= \sigma \left(\lambda_n(K_\delta)\right)^{1/p} \leq \sigma \left(\lambda_n(K_1)\right)^{1/p} < \varepsilon \quad \text{for } \sigma < \varepsilon \left(\lambda_n(K_1)\right)^{-1/p}.$$

Thus for the fixed $\varepsilon > 0$

$$\exists\, \delta > 0\ :\ \forall\, \tau \in \mathbb{R}^n : \|\tau\|_p < \delta\ \Rightarrow\ \|g(t + \tau) - g(t)\|_p < \varepsilon.$$

and than

$$\|f(t + \tau) - f(t)\|_p \leq$$
$$\|f(t + \tau) - g(t + \tau)\|_p + \|g(t + \tau) - g(t)\|_p + \|g(t) - f(t)\|_p \leq$$
$$\varepsilon + \varepsilon + \varepsilon = 3\varepsilon$$

$\square$

2.3.6 Exercise. Prove that any $f \in C_0(\mathbb{R}^n)$ (continuous with compact support) is uniformly continuous.

3 Applications of Baire theorem

3.1 The Baire Category Theorem

Is it true that the intersection of at least two dense sets is dense? In the general case, the answer is, of course, negative — just recall the example of representing the real line as a union of disjoint dense sets: $\mathbb{R} = \mathbb{Q} \cup (\mathbb{R} \setminus \mathbb{Q})$. However, it seems that the answer to this question becomes yes if you require from sets, in addition to density, the property of «non-meagrity». The Baire Category Theorem speaks about this. Here are the relevant definitions.

3.1.1 Definition (Baire Category)**.** *Let (X, ρ) be a metric space. A subset $A \subseteq X$ is called*
*(— **dense** in X if its closure is equal to X or, equivalently, every nonempty open subset of X contains an element of A.)*
*— **nowhere dense** if its closure $\overline{A}$ has an empty interior (that is, if there is no ball $B_r(x) \subseteq \overline{A}, r > 0$.*
*— **meagre** or **of the first category** (in the sense of Baire) if it is a countable union of nowhere dense subsets of X.*
*— **nonmeagre** or **of the second category** if it is not meagre.*
*— **residual** or **comeagre** if its complement is meagre.*

Note that the complement of a nowhere dense set is necessarily dense, but the complement of a dense set, generally speaking, is not nowhere dense (as we can see in the case of sets $\mathbb{Q}$ and $\mathbb{R} \setminus \mathbb{Q}$ where both sets are dense in $\mathbb{R}$). But the complement of any open dense set is nowhere dense. The next lemma summarizes some elementary consequences of the last definitions.

3.1.2 Lemma. *Let (X, ρ) be a metric space. Then the following hold.*
(i) A subset $A \subseteq X$ is nowhere dense if and only if its complement $X \setminus A$ contains a dense open subset of X.
(i′) A subset of X is residual if and only if it contains a countable intersection of dense open subsets of X.
(ii) If $B \subseteq X$ is nowhere dense and $A \subseteq B$, then A is nowhere dense.
(ii′) If $B \subseteq X$ is meagre and $A \subseteq B$, then A is meagre.
(ii″) If $A \subseteq X$ is nonmeagre and $A \subseteq B \subseteq X$, then B is nonmeagre.
(iii) Every finite union of nowhere dense sets is again nowhere dense.
(iii′) Every countable union of meagre sets is again meagre.
(iii″) Every countable intersection of residual sets is again residual.

Proof. The complement of the closure of a subset of X is the interior of the complement and vice versa. Thus every subset $A \subseteq X$ satisfies

$$X \setminus \mathrm{Int}(\overline{A}) = \overline{X \setminus \overline{A}} = \overline{\mathrm{Int}(X \setminus A)}.$$

This shows that a subset $A \subseteq X$ is nowhere dense if and only if the interior of $X \setminus A$ is dense in X, i.e. $X \setminus A$ contains a dense open subset of X. This proves (i). The rest of the statements of the lemma follow from the definitions and laws of de Morgan. □

3.1.3 Proposition. *Let X be an arbitrary normed space, and X_0 be its closed proper subspace («proper» means that $X \neq X_0$). Than*
a) X_0 is nowhere dense in X,
b) for any $a \in X$ the affine space $X_0 + a = \{x + a \ : \ x \in X_0\}$ is nowhere dense in X.

Proof. Let's prove a) from the opposite. If $\overline{E} = E$ is not empty, then there is a ball $B \subseteq X_0$. The center of this ball $x \in X_0$, that is, for some $r > 0$ we have $B_r(x) \subseteq X_0$, then, since X_0 is a linear space, $B_r(0) = B_r(x) - x \subseteq X_0$, and $nB_r(0) \subseteq X_0$ for any $n \in \mathbb{N}$, that implies $\cup_{n=1}^{\infty} nB_r(0) \subseteq X_0$. But $\cup_{n=1}^{\infty} nB_r(0) = X$, so $X = X_0$ and we have a contradiction. □

3.1.4 Examples.

1) for $X = \mathbb{R}$, any finite set or any infinite set with a finite set of limit points, the Cantor set are nowhere dense. $\mathbb{Q}$ is meagre and, relatively, $\mathbb{R}\backslash\mathbb{Q}$ is residual.

2) If $X = \varnothing$ than X is both meagre and residual.

3) Let $X = C[0,1]$, $A = \{f \in X \ : \ \int_0^1 f(x)dx = 0\}$, $B = \{f \in X \ : \int_0^1 f(x)dx = f(0)\}$.

We see that nowhere dense sets are in some sense small sets, and the properties of sparse sets are similar to the properties of sets of measure zero. Maybe the meager sets and the sets of measure zero are the same sets? The following sentence answers this question.

3.1.5 Proposition. *The space $\mathbb{R}^n$ admits a representation in the form of a union of two sets, one of which is meagre, and the second has a zero Lebesgue measure.*

Proof. Choose a bijection $\mathbb{N} \to \mathbb{Q}^n : k \to x_k$ and, for all $\varepsilon > 0$, define

$$U_\varepsilon := \bigcup_{k=1}^{\infty} B_{2^{-k}\varepsilon}(x_k).$$

This is a dense open subset of $\mathbb{R}^n$ and its Lebesgue measure is less than $(2\varepsilon)^n$. Hence $G := \bigcap_{j=1}^{\infty} U_{1/j}$ is a residual set of Lebesgue measure zero and its complement $E := \mathbb{R}^n \setminus G = \bigcup_{k=1}^{\infty}(\mathbb{R}^n \setminus U_{1/j})$ is a meagre set of full Lebesgue measure. $\square$

3.1.6 Theorem (Baire Category Theorem). *Let (X, ρ) be a nonempty complete metric space and $E \subseteq X$ is a set with nonempty interior. Then E is nonmeagre (of the second cathegory).*

Proof. From the opposite. Let the set E be represented as a union of a sequence of nowhere dense sets $E = \cup_{k=1}^{\infty} E_k$. Then $(\operatorname{Int} E) \setminus \overline{E}_1$

is open (since this is the difference between open and closed sets) and is not empty (because $\overline{E}_1$ is nowhere dense, so there are no open sets contained in it). Then there exists a ball $B_1 = B_{r_1}(x_1)$, such that

$$\overline{B_1} \subseteq (\operatorname{Int} E) \setminus \overline{E_1}, \qquad 0 < r_1 \leq 1.$$

Than $B_1 \setminus \overline{E}_1 \neq \varnothing$ is open, then there exists a ball $B_2 = B_{r_2}(x_2)$, such that

$$\overline{B_2} \subseteq B_1 \setminus \overline{E_2}, \qquad 0 < r_2 \leq \frac{1}{2}.$$

Having a ball B_{k-1} we take $B_k = B_{r_k}(x_k)$, such that

$$\overline{B_k} \subseteq B_{k-1} \setminus \overline{E_k}, \qquad 0 < r_k \leq \frac{1}{k},$$

Existence of such a ball follows from $\operatorname{Int}(B_{k-1} \setminus \overline{E}_k) = B_{k-1} \setminus \overline{E}_k \neq \varnothing$ since $\overline{E}_k$ is nowhere dense. The centers of the found balls form a fundamental sequence, because for all $k, p \in \mathbb{N}$ we have

$$B_{k+1} \subseteq B_k \;\Rightarrow\; B_{k+p} \subseteq B_k \;\Rightarrow\; x_{k+p} \in B_k \;\Rightarrow\; \rho(x_{k+p}, x_k) < r_k \underset{k\to\infty}{\to} 0.$$

Since the space X is full, there exists a limit of the sequence $x = \lim x_k$, and

$$\forall\, k \quad x \in \overline{B}_k \subseteq X \setminus \overline{E}_k \Rightarrow \; x \notin \cup_{k=1}^{\infty} \overline{E}_k.$$

At the same time $x \in B_1 \subseteq E$ which contradicts $E = \bigcup_{k=1}^{\infty} E_k$. □

3.1.7 Corollary. *Let (X, ρ) be a nonempty complete metric space and $\forall k \in N$ $U_k \subseteq X$ is an open dense set. Then $U = \bigcap_{k=1}^{\infty} U_k$ is dense.*

Proof. For $k \in \mathbb{N}$ define $E_k := X \setminus U_k$. Then $\{E_k\}_k$ is a sequence of closed subsets of X with empty interior. Hence $E := \bigcup_{k=1}^{\infty} E_k$ has empty interior. Therefore, any ball in the space X contains points of $X \setminus E = U$ and this means the density of U. □

3.2 Bounded linear operators

Examples of detection of boundedness and calculation of the operator norm.

3.2.1 Example. *If the spaces X and Y are normed, and X is also finite-dimensional, then any linear operator $F : X \to Y$ is bounded (i.e., continuous).*

Proof. Let $e_1, \ldots, e_n$ be a basis in X, let $\|\cdot\|_2$ be the norm on X defined by the relation (1.4), and $C_1 := \sqrt{\sum_{k=1}^{n} \|F(e_k)\|^2}$. Then for any $x = \sum_{k=1}^{n} x_k e_k \in X$ due to the linearity of F, we obtain from Cauchy's inequality

$$\|F(x)\| \leq \sum_{k=1}^{n} |x_k| \cdot \|F(e_k)\| \leq \sqrt{\sum_{k=1}^{n} |x_k|^2} \sqrt{\sum_{k=1}^{n} \|F(e_k)\|^2} = C_1 \|x\|_2.$$

By Theorem 1.4.1 the norms $\|\cdot\|$ and $\|\cdot\|_2$ are equivalent, hence there exists $C_2 > 0$ such that $\|x\|_2 \leq C_2 \|x\|$ for all $x \in X$. As a result, it turns out that $\|F(x)\| \leq C\|x\|$, where $C = C_1 \cdot C_2$, so F is bounded and $\|F\| \leq C$. □

3.2.2 Example. *Let X be an inner product space linear space $\langle \cdot, \cdot \rangle$ over $\mathbb{R}$ (or $\mathbb{C}$), $a \in X$ is fixed, $f : X \to \mathbb{R}$ (or $\mathbb{C}$), $f(x) := \langle x, a \rangle$. Then f is a bounded functional and $\|f\| = \|a\|$.*

Proof. Due to Cauchy's inequality $|f(x)| = |\langle x, a \rangle| \leq \|x\| \cdot \|a\|$, hence f is bounded (Lipschitz continuous), and $\|f\| \leq \|a\|$.

To proof $\|f\| \geq \|a\|$ we consider $x_0 = \frac{a}{\|a\|}$. Then $\|x_0\| = 1$, so

$$\|f\| = \sup_{\|x\| \leq 1} \|f(x)\| \geq |f(x_0)| = \left| \left\langle \frac{a}{\|a\|}, a \right\rangle \right| = \frac{1}{\|a\|} |\langle a, a \rangle| = \|a\|.$$

Eventually $\|f\| = \|a\|$. □

3.2.3 Example. *Function multiplication operator. Let K is a compact, $X = C(K)$; $g \in C(K)$;*

$$F(x) := g \cdot x$$

(or with more details, $F(x)(t) := g(t)x(t)$ for all $t \in K$). *Then F is linear bounded and $\|F\| = \|g\|_{C(K)}$.*

Proof. It's obvious that

$$\forall\, t \in K \quad |g(t)x(t)| \le \|g\| \cdot |x(t)| \le \|g\| \cdot \|x\| \Rightarrow$$

$$\|g \cdot x\| \le \|g\| \cdot \|x\| \Leftrightarrow \|F(x)\| \le \|g\| \cdot \|x\|.$$

It means that $\|F\| \le \|g\|$.

Why $\|F\| \ge \|g\|$? Let

$$M = \|g\|_{C(K)} = \max_{t \in K} |g(t)| = |g(t_0)|.$$

If one consider $x_0(t) :\equiv 1$ then

$$\Rightarrow \|x_0\| = 1 \Rightarrow \|F\| = \sup_{\|x\| \le 1} \|F(x)\| \ge \|F(x_0)\| \ge |g(t_0)x_0(t_0)| = M = \|g\|.$$

$\square$

3.2.4 Example. *Let $X = L^p(E, \mathcal{A}, \mu)$ with σ-finite measure μ, $p \in [1, +\infty]$, $\frac{1}{p} + \frac{1}{q} = 1$ (if $p = 1$, than $q = \infty$ and vice versa) and $w \in L^q(E, \mathcal{A}, \mu)$ is fixed and*

$$f(x) = \int_E x(t) \cdot w(t)\, d\mu(t).$$

Then f is linear bounded and $\|f\| = \|w\|_q \left(= \|w\|_{L^q(E,\mathcal{A},\mu)}\right)$.

Proof. For case $p > 1$ from Hölder's inequality we get

$$|f(x)| = \left| \int_E x(t)w(t)d\mu(t) \right| \le \|w\|_q \cdot \|x\|_p \Rightarrow \|f\| \le \|w\|_q.$$

Why «$\geq$»? We consider $x(t) := |w(t)|^{q-2}\overline{w}(t)$ (if $w(t) = 0$ for some t, then $x(t) = 0$), so

$$\|x\|_p^p = \int_E |x(t)|^p d\mu(t) = \int_E |w(t)|^{p(q-1)} d\mu(t) = \int_E |w(t)|^q d\mu(t) = \|w\|_q^q.$$

$$f(x) = \int_E |w|^{q-2}\overline{w} \cdot w\, d\mu = \int_E |w|^{q-2} \cdot |w|^2 d\mu = \int_E |w|^q d\mu = \|w\|_q^q.$$

$$\|f\| \geq \frac{|f(x)|}{\|x\|_p} = \frac{\|w\|_q^q}{\|w\|_q^{q/p}} = \|w\|_q^{q-\frac{q}{p}} = \|w\|_q^{q\cdot\frac{1}{q}} = \|w\|_q.$$

$\square$

Exercises. Prove the last statement for the «borderline case» $p = 1$.

3.2.5 Example. As it was established in the course of mathematical analysis, for an operator $F : \mathbb{R}^n \to \mathbb{R}^n$ given by a square matrix $A = (a_{ij})$ the following estimates are true:

$$\|F\| \leq \sqrt{\sum_{i,j=1}^n a_{ij}^2}.$$

$$\|F\| = \max\{\sqrt{\lambda}\ :\ \lambda \text{ is an eigenvalue of the matrix } A^T A\}.$$

3.2.6 Remarks. *1) For any normed spaces X and Y the norm of operators (defined by 1.9.8) is the norm (in the sense of definition 1.3.2) on $\mathcal{L}(X \to Y)$.*

2) Let X, Y be linear normed spaces, $\{F_k\}_{k\in\mathbb{N}} \subset \mathcal{L}(X \to Y)$ be such a bounded linear operator sequence that the sequence $(F_k(x))_{k\in\mathbb{N}}$ converges in Y for every $x \in X$, and $F(x) := \lim_{k\to\infty} F_k(x)$ for all $x \in X$. Then F is linear.

Linearity of F follows from the definition, the fact that the limit of a sum of two sequences is the sum of the limits, and the fact that the limit of a product of a sequence with a scalar is the product of the limit with the scalar.

The next theorem shows that the normed vector space $\mathcal{L}(X{\to}Y)$ is complete whenever the target space Y is complete, even if X is not complete.

3.2.7 Theorem (Completeness of the operator space)**.** *Let X be a normed vector space and let Y be a Banach space. Then $\mathcal{L}(X{\to}Y)$ is a Banach space with respect to the operator norm.*

Proof. Let $(F_n)_{n\in\mathbb{N}}$ be a Cauchy sequence (or fundamental sequence) in $\mathcal{L}(X{\to}Y)$. Then for all $x \in X$ and all $k, l \in \mathbb{N}$

$$\|F_k(x) - F_l(x)\| = \|(F_k - F_l)(x)\| \leq \|F_k - F_l\| \cdot \|x\|.$$

Hence $(F_k(x))_{k\in\mathbb{N}}$ is a Cauchy sequence in Y for every $x \in X$. Since Y is complete, this implies that the limit

$$F(x) := \lim_{k\to\infty} F_k(x) \tag{3.1}$$

exists for all $x \in X$. This defines a map $F : X \to Y$. It is linear by remark 3.2.6 2). It remains to prove that F is bounded and that $\lim_{k\to\infty} \|F_k - F\| = 0$. Any fundamental sequence is bounded (this is true in any metric space, verification is the same as in the case of number sequences). So,

$$\begin{gathered}
\exists\, C > 0 \ :\ \forall\, k \in \mathbb{N} \qquad \|F_k\| \leq C \quad \Leftrightarrow \\
\forall\, k \in \mathbb{N} \quad \forall\, x \in X \ :\ \|x\| \leq 1 \qquad \|F_k(x)\| \leq C \quad \Rightarrow \\
\forall\, x \in X \ :\ \|x\| \leq 1 \qquad \|F(x)\| \leq C
\end{gathered}$$

the first transition is correct by the property of the norm (1.21), next, the limit transition was made. The last obtained inequality means that the operator F is bounded and $\|F\| \leq C$.

Let's evaluate $\|F_k - F\|$. For fixed $x \in X \ :\ \|x\| \leq 1$ we have:

$$\begin{gathered}
\|(F_k - F)(x)\| = \lim_{j\to\infty} \|F_k(x) - F_j(x)\| \leq \\
\sup_{\{j:j\geq k\}} \|F_k(x) - F_j(x)\| \leq \sup_{\{j:j\geq k\}} \|F_k - F_j\| \Rightarrow \\
\|F_k - F\| = \sup_{\{x:\|x\|\leq 1\}} \|(F_k - F)(x)\| \leq \sup_{\{j:j\geq k\}} \|F_k - F_j\| \xrightarrow[k\to\infty]{} 0,
\end{gathered}$$

because the sequence $(F_k)_{k\in\mathbb{N}}$ is fundamental. $\square$

Definition. *For a normed space X over the field K, $K = \mathbb{R}$ or $K = \mathbb{C}$, the space $X^* := \mathcal{L}(X\to K)$ of bounded linear functionals on X is called the **dual space** of X.*

Corollary. *X^* is a Banach space for every normed vector space X.*

3.3 Principle of uniform boundedness

3.3.1 Theorem (Banach-Steinhaus uniform boundedness principle)**.** *Let X be a Banach space and let Y be a normed space. let I be any set, and, for each $i \in I$ let $F_i : X \to Y$ be a bounded linear operator. Assume that the operator family $\{F_i\}_{i\in I}$ is pointwise bounded, i.e.*

$$\forall\, x \in X \quad \sup_{i\in I} \|F_i(x)\| < \infty.$$

Then $\{F_i\}_{i\in I}$ is uniformly bounded, i.e.

$$\sup_{i\in I} \|F_i\| < \infty.$$

Note that the reverse direction is trivially true — uniform boundedness implies pointwise boundedness (Why?).

Proof. For every $x \in X$, define

$$M(x) = \sup_{i\in I} \|F_i(x)\|.$$

Pointwise boundedness of $\{F_i\}$ means that

$$X = \bigcup_{k\in\mathbb{N}} X_k, \qquad \text{where} \quad X_k = \{x \in X \ : \ M(x) \le k\}.$$

All X_k are closed (because $(x_j)_{j\in\mathbb{N}} \in X_k,\ x_j \to x \Rightarrow \forall\, i \in I\ F_i(x_j) \to_{j\to\infty} F_i(x)$ and $\|F_i(x_j)\| \le k \Rightarrow \|F_i(x)\| \le k \Rightarrow x \in X_k$). Baire category theorem (3.1.6) implies that one of X_k is not a nowhere dense subset of

X. For such X_k it holds $\operatorname{Int}\operatorname{Cl} X_k = \operatorname{Int} X_k \neq \varnothing$. Summarizing, we have shown that there exist $k \in \mathbb{N}$, $x_0 \in X$ and $\varepsilon > 0$ such that

$$x_0 + B_\varepsilon(0) \subseteq X_k.$$

X_k is symmetrical, so $-(x_0 + B_\varepsilon(0)) = -x_0 + B_\varepsilon(0) \subseteq X_k$. Hence by convexity of X_k, we have

$$\frac{1}{2}\big(x_0 + B_\varepsilon(0)\big) + \frac{1}{2}\big(-x_0 + B_\varepsilon(0)\big) = B_\varepsilon(0) \subseteq X_k.$$

By definition of X_k this means that for $x \in X$,

$$\|x\| \leq \varepsilon \;\Rightarrow\; \sup_{i \in I} \|F_i(x)\| \leq k.$$

It follows that for every $x \in X$, $\|x\| \leq 1$

$$\sup_{i \in I} \|F_i(x)\| \leq \frac{k}{\varepsilon}.$$

This implies that $\sup_{i \in I} \|F_i\| \leq \frac{k}{\varepsilon} < \infty$ as required. □

3.3.2 Corollary (on the pointwise limit of a sequence of linear operators)**.** *Let X be a Banach space and let Y be a normed space. Let $\{F_k\}_{k \in \mathbb{N}} \subset \mathcal{L}(X \to Y)$ be such a bounded linear operator sequence that the sequence $(F_k(x))_{k \in \mathbb{N}}$ converges in Y for every $x \in X$, and*

$$F(x) := \lim_{k \to \infty} F_k(x) \quad \textit{for all } x \in X.$$

Then
I) $(F_k)_{k \in \mathbb{N}}$ *is uniformly bounded;*
II) $\|F\| \leq \varliminf_{x \to \infty} \|F_k\|$.

Proof. F is linear by remark 3.2.6 2). Pointwise convergence implies pointwise boundedness, so I) follows from theorem 3.3.1. Let's check II). For any $x \in X$ such that $\|x\| \leq 1$ we have

$$\|F(x)\| = \lim_{k \to \infty} \|F_k(x)\| = \varliminf_{k \to \infty} \|F_k(x)\| \leq \varliminf_{k \to \infty} \|F_k\|$$

From here, using one of the formulas for the norm (see 1.9.9), we get $\|F\| \le \varliminf_{k\to\infty} \|F_k\|$. $\square$

Remark. *Pointwise convergence of a sequence of linear operators does not imply convergence in norm (i. e. not every pointwise convergent sequence converges in norm).*

3.3.3 Example. Let H be a Hilbert space, and let $(e_k)_{k\in\mathbb{N}}$ is an infinite orthonormal system. Consider the following sequence of linear functionals $(f_k)_{k\in\mathbb{N}} \subset H^*$:

$$\forall\, k \in \mathbb{N}, \quad \forall\, x \in H \qquad f_k(x) = \langle x, e_k \rangle.$$

Than $(f_k)_{k\in\mathbb{N}}$ pointwise convergent to $\mathbf{0}$, because for all $x \in H$ Bessel's inequality gives

$$\sum_{k=1}^{\infty} |f_k(x)|^2 = \sum_{k=1}^{\infty} \langle x, e_k \rangle^2 \le \|x\|^2 \quad \Rightarrow \quad f_k(x) \xrightarrow[k\to\infty]{} 0.$$

But $\|f_k\| = \|e_k\| = 1$ for all k, so it does not tend to zero by the norm.

3.3.4 Definition. *A set $E \subseteq X$ is called **fundamental** if* $\operatorname{span}(E)$ *is dense in X, i.e.*

$$\overline{\operatorname{span}(E)} = X.$$

Example. Set $E = \{t^k\}_{k\in\mathbb{N}, t\in[a,b]}$ is fundamental in $C[a,b]$ and in $L^p[a,b]$ for $p \in [1, \infty]$.

3.3.5 Corollary (Banach-Steinhaus theorem).
Let X, Y be Banach spaces, $(F_k)_{k\in\mathbb{N}} \subseteq \mathcal{L}(X \to Y)$. Then the following statements are equivalent.
(I) $(F_k)_{k\in\mathbb{N}}$ *converges pointwise on X.*
(II) $(F_k)_{k\in\mathbb{N}}$ *converges pointwise on some fundamental set $E \subseteq X$ and* $\sup_k \|F_k\| < \infty$.

Proof. (I) $\Rightarrow$ (II) is obvious, one can take $E = X$ and $\sup_k \|F_k\| < \infty$ by 3.3.2.

(II) $\Rightarrow$ (I)? Since, by the condition of the theorem, the space Y is complete, it suffices to show that for any x the sequence $(F_k(x))_{k\in\mathbb{N}}$ is a Cauchy sequence. This is true for every $x \in E$ and every $x \in \operatorname{span} E$ too (if $(F_k(x_1))_k$ and $(F_k(x_2))$ converge, then $(F_k(\alpha_1 x_1 + \alpha_2 x_2))$ converges too for all scalars α_1, α_2).

Let $C = \sup_k \|F_k\| > 0$, $x \in X$. For any $\varepsilon > 0$ there exists $x' \in \operatorname{span}(E)$ such that

$$\|x - x'\| < \frac{\varepsilon}{3C}.$$

$(F_k(x'))_k$ is a Cauchy sequence, there exists number N such that $\|F_l(x') - F_m(x')\| < \frac{\varepsilon}{3}$ for any $l, m \geq N$. So

$$\begin{gathered} l, m \geq N \quad \|F_l(x) - F_m(x)\| \leq \\ \|F_l(x) - F_l(x')\| + \|F_l(x') - F_m(x')\| + \|F_m(x') - F_m(x)\| < \\ \|F_l\| \cdot \|x - x'\| + \frac{\varepsilon}{3} + \|F_m\| \cdot \|x - x'\| \leq \\ C \cdot \frac{\varepsilon}{3C} + \frac{\varepsilon}{3} + C \cdot \frac{\varepsilon}{3C} = \varepsilon. \end{gathered}$$

Thus we have shown what we wanted, $(F_k(x))_k$ is a Cauchy sequence. □

3.3.6 Example. This example shows that the hypothesis that X is complete cannot be removed in theorem 3.3.1 and its corollaries 3.3.2, and 3.3.5.

Let's consider

$$X := c_{00} = \{x{=}(x_i)_{i=1}^{\infty} \in l^{\infty} \;:\; \exists\, n \;\in \mathbb{N} \quad \forall\, i \geq n \quad x_i = 0\}$$

with the supremum norm from l^{∞}. This is a normed vector space. It is not complete, but is a linear subspace of l^{∞} whose closure $\operatorname{Cl} X = c_0$ is the subspace of sequences of complex numbers that converge to zero. Define the linear operators $F_n : X \to X$ and $F : X \to X$ by

$$F_n(x) = (x_1, 2x_2, \ldots, nx_n, 0, 0, \ldots), \quad F(x) = (ix_i)_{i\in\mathbb{N}}. \tag{3.2}$$

for $n \in \mathbb{N}$ and $x = (x_i) \in X$. Then $F(x) = \lim_{n\to\infty} F_n(x)$ for every $x \in X$ and $\|F_n\| = n$ for every $n \in \mathbb{N}$. Thus the sequence $(F_n(x))_{n\in\mathbb{N}}$ is bounded for every $x \in X$, the sequence $(F_n)_n$ pointwise converges to F, but the linear operator F is not bounded, and the sequence $\|F_n\|$ is not bounded.

3.4 Open mapping theorem

A small digression about perfectly convex sets.

3.4.1 Definition (Perfectly convex set)**.** *A set K in a Banach space Y is called **perfectly convex** if for every sequence $(x_k)_{k=1}^{\infty}$ in K and every non-negative number sequence $(\lambda_k)_{k=1}^{\infty}$ such that $\sum_{k=1}^{\infty} \lambda_k = 1$, one has $\sum_{k=1}^{\infty} \lambda_k x_k \in K$.*

Convex sets satisfy this property only for finite sequences (x_k). Therefore, every perfectly convex set is convex, but not vice versa.

3.4.2 Example. Let $X = C[0,1]$, $K = \mathcal{P}(x)$ be the set of all algebraic polynomial functions on $[0,1]$. Then K is convex. The functions $x, x^2, x^3, \dots$ belong to K, but the function $\frac{x}{2} + \frac{x^2}{4} + \cdots = \frac{x/2}{1-x/2} = \frac{x}{2-x}$ does not belong to K, so K is not perfectly convex.

Some simple facts concerned the concept of a perfectly convex set are collected in the following exercise.

3.4.3 Exercises. Let K be a subset of a linear normed space X. Prove the following statements.

(a) If K is perfectly convex then it's convex.

(b) If K is perfectly convex then it's bounded.

(c) If X is Banach, and K is convex, bounded and closed then K is perfectly convex.

(d) The interior and closure of a perfect convex set are perfectly convex sets.

(e) The intersection of any family $(K_\alpha)_{\alpha\in A}$ of perfect convex sets is perfect convex.

3.4.4 Reminder. *For a subset K of a metric space (X,ρ) the set*

$$K_\varepsilon := \{x \in X \ : \ \exists \ k \in K \ : \ \rho(x,k) < \varepsilon\}$$

*is called an ε-**neighborhood** of the set K.*

When X is a normed (linear) space ε-neighborhood may be represented in the form

$$K_\varepsilon = K + \varepsilon B_1(\mathbf{0}),$$

In general setting a perfectly convex set may have not inner points but the following fact is true.

3.4.5 Proposition. *Let K be a perfectly convex set of a normed space X, $B = rB_1(\mathbf{0})$ for some $r > 0$. If the condition*

$$B \subseteq K + \varepsilon B \tag{3.3}$$

is satisfied for some $\varepsilon \in (0,1)$, then $(1-\varepsilon)B \subseteq K$.

Proof. Iterating (3.3) we get

$$B \subseteq K + \varepsilon B \subseteq K + \varepsilon\left(K + \varepsilon B\right) \subseteq K + \varepsilon K + \cdots + \varepsilon^n K + \varepsilon^{n+1} B$$

for all $n \in \mathbb{N}$. Above inclusion means that for any $x \in B$ there are two sequences $x_n \in K$ and $y_n \in B$ such that for all $n \in \mathbb{N}$ it holds

$$x = x_0 + \varepsilon x_1 + \ldots \varepsilon^n x_n + \varepsilon^{n+1} y_n \quad \Rightarrow$$

$$(1-\varepsilon)x = \sum_{k=0}^{n} (1-\varepsilon)\varepsilon^k x_k + (1-\varepsilon)\varepsilon^{n+1} y_n. \tag{3.4}$$

Both terms on the right side have limits as $n \to \infty$:

$$\left\|(1-\varepsilon)\varepsilon^{n+1} y_n\right\| \leq (1-\varepsilon)\varepsilon^{n+1} r \quad \Rightarrow \quad (1-\varepsilon)\varepsilon^{n+1} y_n \to 0,$$

and since $\sum_{k=0}^{\infty}(1-\varepsilon)\varepsilon^k = 1$, due to the perfect convexity of K, we have $\sum_{k=0}^{\infty}(1-\varepsilon)\varepsilon^k x_k \in K$, that is this series converges in K. The limit transition in (3.4) gives

$$(1-\varepsilon)x = \sum_{k=0}^{\infty}(1-\varepsilon)\varepsilon^k x_k \in K,$$

which is exactly what was required.

□

3.4.6 Theorem (Open mapping theorem by S. Banach)**.**
Let X, Y be Banach spaces. Then every surjective bounded linear operator $F : X \to Y$ is an open map, i.e. F maps open sets in X to open sets in Y.

Proof. To prove the theorem, it suffices to find $\varepsilon > 0$ such that

$$F(B_X) \supseteq \varepsilon B_Y, \tag{3.5}$$

where B_X and B_Y are open balls centred at zero radius 1 in spaces X and Y respectively, because if we find it, then for every open $U \subseteq X$ and all $y \in F(X)$ for any $x \in F^{-1}(y)$ there exists $\delta > 0$ such that

$$U \supseteq x + \delta B_X,$$

(since U is open). So

$$F(U) \supseteq F(x + \delta B_X) = F(x) + F(\delta B_X) = y + \delta F(B_X) \supseteq y + \delta\varepsilon B_Y,$$

and y is an interior point of $F(U)$.

Now we find ε. In view of application of Baire category theorem, we represent

$$X = \bigcup_{k\in\mathbb{N}} kB_X.$$

Therefore

$$Y = F(X) = \bigcup_{k\in\mathbb{N}} kF(B_X).$$

By Baire category theorem, there exists $k \in \mathbb{N}$ such that $kF(B_X)$ is not a nowhere dense set. Thus $F(B_X)$ is not a nowhere dense set, i.e. its closure has nonempty interior. So there exist $y \in Y$ and $r > 0$ such that

$$\overline{F(B_X)} \supseteq y + rB_Y.$$

By symmetry, $\overline{F(B_X)} \supseteq -(y + rB_Y) = -y + rB_Y$. Hence by convexity (check this!) we have

$$\overline{F(B_X)} \supseteq \frac{1}{2}\big(y + rB_Y\big) + \frac{1}{2}\big(-y + rB_Y\big) = rB_Y.$$

3.4.7 Exercise. $\overline{K} \supseteq D$ does not imply $K \supseteq D$ even for convex and symmetric sets K, D in a Banach space.

Since $\overline{F(B_X)} \subseteq F(B_X) + \frac{r}{2}B_Y$, by proposition 3.4.5, it suffices to show that $K = F(B_X)$ is perfectly convex (then $\frac{r}{2}B_Y \subseteq K$ and (3.5) holds for $\varepsilon = \frac{r}{2}$). Perfect convexity of $F(B_X)$ is easy to check. Indeed, consider any sequence $(F(x_k)) \subseteq F(B_X)$ with $x_k \in B_X$, and numbers λ_k such that $\sum_{k=1}^{\infty} \lambda_k = 1$. Then

$$\sum_{k=1}^{\infty} \lambda_k F(x_k) = F\left(\sum_{k=1}^{\infty} \lambda_k x_k\right) \tag{3.6}$$

provided that the series $\sum_k \lambda_k x_k$ converges. It indeed converges absolutely:

$$\left\|\sum_k \lambda_k x_k\right\| \le \sum_k \lambda_k \|x_k\| < \sum_k \lambda_k = 1.$$

By completeness of X, the series $\sum_k \lambda_k x_k$ converges to a vector in B_X. It follows that the right side of (3.6) belongs to $F(B_X) = K$, as required. This completes the proof of the open mapping theorem. $\square$

As an immediate consequence of the open mapping theorem, we obtain:

3.4.8 Corollary (Inverse Operator Theorem)**.**
Let X, Y be Banach spaces. Then every bijective linear operator $F \in \mathcal{L}(X \to Y)$ is a homeomorphism, i.e. $F^{-1} \in \mathcal{L}(Y \to X)$.

Proof. The open mapping theorem states that the preimages of open sets under F^{-1} are open, hence F^{-1} is a continuous map. □

The inverse mapping theorem is often used to establish stability of solutions of linear equations.

3.4.9 Example. Consider a linear equation in x in Banach space:

$$\mathcal{F}(x) = b \tag{3.7}$$

with $F \in \mathcal{L}(X \to Y)$ and $b \in Y$. Assume a solution x *exists and is unique* for every right hand side b. Then, by inverse operator theorem, the solution $x = x(b)$ is *continuous* with respect to b. In other words, the solution is stable under perturbations of the right hand side of (3.7).

The following two examples show that the hypothesis that X and Y are *complete cannot be removed* in Theorems 3.4.6 and 3.4.8.

3.4.10 Exercises. 1) As in Example 3.3.6, let $X = c_{00}$ be the subspace of number sequences that vanish for a sufficiently large number, equipped with the supremum norm. Thus X is a normed vector space but is not a Banach space. Define the operator $G : X \to X$, by $G(x) = (k^{-1}x_k)_k$ for $(x_k)_k \in X$. Then G is a bijective bounded linear operator but its inverse is unbounded.

2) Here is an example where X is complete and Y is not. Let $X = Y = C[0, 1]$ be the space of continuous functions equipped with the different norms, the usual norm on X and the norm induced from $L[0, 1]$ on Y. Then X is a Banach space, Y is a non-Banach space, and the identity mapping $\mathrm{id} : X \to Y$ is a bijective bounded linear operator with an unbounded inverse.

In case F is not injective (but is surjective) in the inverse mapping theorem, one can still apply inverse mapping theorem to the injectivization of F. What is injectivization?

3.4.11 Exercise (*Injectivization*)**.** This is a linear version of the fundamental theorem on homomorphisms for groups. Consider a linear operator $F : X \to Y$ acting between linear spaces X and Y. The operator F may not be injective; we would like to make it into an injective operator. To this end, we consider the map $\tilde{F} : X/\operatorname{Ker} F \to Y$ which sends every coset $[x]$ into a vector $F(x)$, i.e. $\tilde{F}([x]) := F(x)$.

(i) Prove that $\tilde{F}$ is well defined, i.e. $[x_1] = [x_2]$ implies $F(x_1) = F(x_2)$.

(ii) Check that $\tilde{F}$ is a linear and injective operator.

(iii) Check that F is surjective then $\tilde{F}$ is also surjective, and thus $\tilde{F}$ is a linear isomorphism between $X/\operatorname{Ker} F$ and Y.

3.4.12 Definition. *Given a subspace X_0 of a linear vector space E, there is a canonical linear operator associated with the quotient space X/X_0,* ***quotient map*** *$q : X \to X_0$, which acts as $q(x) = [x]$.*

3.4.13 Corollary (Surjective operators are essentially quotient maps)**.** *Let X, Y be Banach spaces. Then every surjective linear operator $F \in \mathcal{L}(X{\to}Y)$ is a composition of a quotient map and a homeomorphism. Specifically,*

$$F = \tilde{F}q,$$

where $q : X \to X/\operatorname{Ker} F$ is the quotient map, $\tilde{F} : X/\operatorname{Ker} F \to Y$ is an homeomorphism.

Proof. Let $\tilde{F}$ be the injectivization of F constructed in Example 3.4.11. By construction, $F = \tilde{F}q$ and F is injective. Since F is surjective, $\tilde{F}$ is also surjective. Therefore, by inverse mapping theorem $\tilde{F}$ is an homeomorphism, completing the proof. $\square$

3.5 Closed graph theorem

Closed graph theorem is an alternative way to check whether a given linear operator is bounded. This result characterizes bounded operators in terms of their graphs.

3.5.1 Definition. *Let $(X, \|\cdot\|_X)$ and $(Y, \|\cdot\|_Y)$ be normed spaces and $p \in [1, +\infty]$. Define the direct sum of $X \oplus_p Y$ as the Cartesian product $X \times Y$ equipped with the norm*

$$\|(x, y)\|_p := \begin{cases} \left(\|x\|_X^p + \|y\|_Y^p\right)^{1/p}, & \text{if } p \in [1, +\infty), \\ \max\{\|x\|_X, \|y\|_Y\}, & \text{if } p = \infty, \end{cases}$$

and

$$\|(x, y)\| := \|(x, y)\|_1, \qquad X \oplus Y := X \oplus_1 Y. \tag{3.8}$$

3.5.2 Exercises. 1) Show that $X \oplus_p Y$ is a normed space, and all norms $\|(x, y)\|_p$, $p \in [1, +\infty]$, are equivalent to each other.

2) Let X and Y be Banach spaces. Show that the direct sum $X \oplus Y$ (therefore, $X \oplus_p Y$ for any $p \in [1, +\infty]$ too) is a Banach space.

3.5.3 Definition. *Let $F : X \to Y$ be a mapping between normed spaces X and Y. The* ***graph*** *of F is the following subset of the direct sum $X \oplus Y$:*

$$\Gamma(F) := \{(x, F(x)) \ : \ x \in X\}.$$

If F is linear than $\Gamma(F)$ is a linear subspace of the normed space $X \oplus Y$.

The main result of this section is that for linear F its graph $\Gamma(F)$ is closed if and only if F is bounded.

Let us compare these two notions, continuity of F and having closed graph. F is continuous if and only if

$$x_n \to x \in X \quad \text{implies} \quad F(x_n) \to F(x). \tag{3.9}$$

In contrast, $\Gamma(F)$ is closed if and only if

$$(x_n \to x \in X \quad \text{and} \quad F(x_n) \to y) \quad \text{implies} \quad y = F(x). \tag{3.10}$$

It is clear from these two lines that continuity always implies the closed graph property:

3.5.4 Proposition. *For any continuous mapping $F : X \to Y$ between normed spaces X and Y its graph $\Gamma(F)$ is closed.*

The opposite statement even in the linear case is nontrivial and requires completeness of both spaces X and Y:

3.5.5 Theorem (Closed graph theorem)**.** *Let $F : X \to Y$ be a linear operator between Banach spaces X and Y. Then F is continuous if and only if its graph is a closed linear subspace of $X \oplus Y$.*

Proof. Taking into account proposition 3.5.4, it is sufficient to check that the boundedness of operator follows from the closeness of the graph.

The direct sum $X \oplus Y$ is a Banach space (Exercise 3.5.2). The graph $\Gamma(F)$ is a closed linear subspace of $X \oplus Y$ hence $\Gamma(F)$ is a Banach space itself.

Consider the linear operator

$$\pi : \Gamma(F) \to X, \quad \pi(x, F(x)) := x.$$

Then π is a bounded, surjective and injective linear operator between two Banach spaces. By the open mapping theorem, π^{-1} is bounded. This means that there exists a number C such that

$$\|x\| + \|F(x)\| = \|(x, F(x))\| \leq C\,\|x\| \quad \text{for all } x \in X.$$

The inequality $\|F(x)\| \leq C\,\|x\|$ implies that F is bounded. $\square$

Remark. For nonlinear mappings, continuity does not follow from the closeness of the graph. As an example, we can consider a discontinuous at zero function $f : \mathbb{R} \to \mathbb{R}$,

$$f(x) = \begin{cases} 1/x, & \text{if } x \neq 0, \\ 0, & \text{if } x = 0, \end{cases}$$

the graph of which is closed in $\mathbb{R}^2$ (since its complement is open).

Interpretation and an example

3.5.6 Example. The hypothesis that X is complete cannot be removed in Theorem 3.5.5.

Let $X := C^1[0,1]$ and $Y := C[0,1]$, both equipped with the supremum norm, and define $F : X \to Y$ by $F(f) := f'$. Then F is unbounded. And it has a closed graph: if $f_n \in C^1[0,1]$ is a sequence of continuously differentiable functions such that the pair $(f_n, F(f_n))$ converges to (f, g) in $X \times X$, then f_n converges uniformly to f and f_n converges uniformly to g, and hence f is continuously differentiable with $f' = g$ by the fundamental theorem of calculus. Nevertheless, as we know the differential operator is unbounded. This does not contradict the closed graph theorem, because $C^1[0,1]$ is not complete under the sup-norm. If we consider $C^1[0,1]$ under its natural norm $\|f\|_\infty + \|f'\|_\infty$ in which it is a Banach space, then the differential operator will obviously be bounded.

A remarkable application of closed graph theorem is that the symmetry property of an operator always implies boundedness:

3.5.7 Theorem (Hellinger-Toeplitz)**.** *Let $F : H \to H$ be a linear operator on a Hilbert space H. Suppose that*

$$\langle x, F(y) \rangle = \langle F(x), y \rangle \quad \text{for all } x, y \in H. \tag{3.11}$$

Then F is bounded.

Proof. By the closed graph theorem, it suffices to check that the graph of F is closed. To this end we choose convergent sequences $x_n \to x$, $F(x_n) \to y$ in H. We would like to show that $y = F(x)$. It suffices to show that

$$\langle z, y \rangle = \langle z, F(x) \rangle \quad \text{for all } x, y \in H.$$

(Why?) This follows by using continuity of the inner product and (3.11) twice:

$$\langle z, y \rangle = \lim_n \langle z, F(x_n) \rangle = \lim_n \langle F(z), x_n \rangle = \langle F(z), x \rangle = \langle z, F(x) \rangle.$$

The proof is complete. □

Hellinger-Toeplitz theorem identifies the source of considerable difficulties in mathematical physics. Many natural operators, such as differential, satisfy the symmetry condition (3.11) but are unbounded. Hellinger-Toeplitz theorem declares that such operators can not be defined everywhere on the Hilbert space. For example, one can never define a useful notion of differentiation that would make all functions in L^2 differentiable.

This explains that working with unbounded operators one has to always keep track of their domains. For example, a linear operator F on a Hilbert space H is called *symmetric* if the domain of F is dense in H, and (3.11) holds. An example of a symmetric operator is the differential operator on $L^2[0,1]$

$$\mathcal{F} = i\frac{d}{dt}$$

with domain

$$\operatorname{Dom} \mathcal{F} = \{f \in C^1[0,1], \quad f(0) = f(1) = 1\}.$$

4 Bounded linear functionals

4.1 Dual space. Linear functionals and hyperspaces

4.1.1 Reminder. *For a normed space X over the field K, $K = \mathbb{R}$ or $K = \mathbb{C}$, a linear mapping from X to K is called a linear functional, and the space $X^* := \mathcal{L}(X \to K)$ of bounded linear functionals on X is called the* ***dual space*** *of X.*

Some examples of linear functionals on functional spaces one can find in 1.9.2.

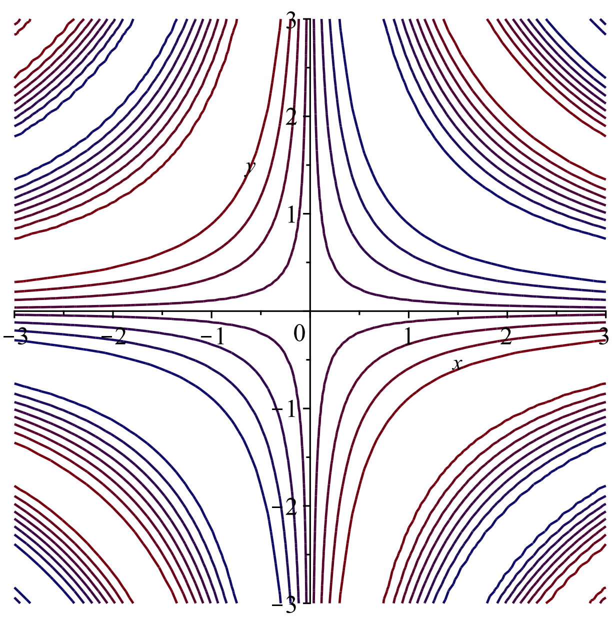


Figure 2: An example of a level sets of for some function of two variables ($f(x,y) = \sin(xy)$).

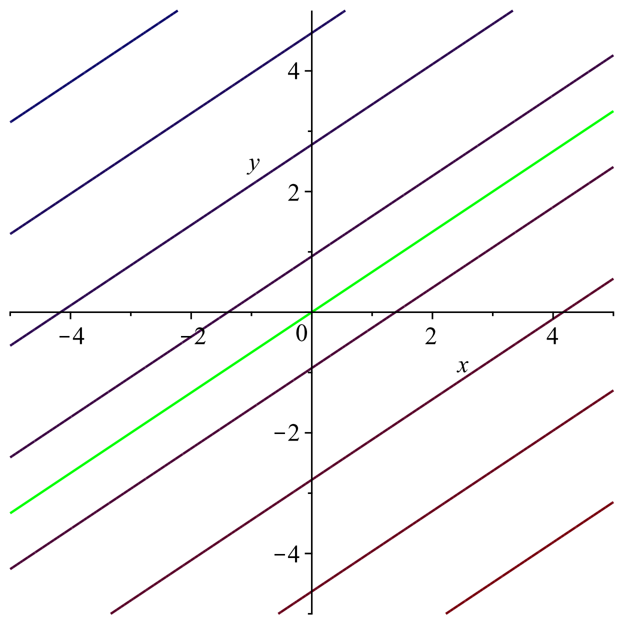


Figure 3: Level sets of a linear functional f. The 0-level set (i. e. the $\operatorname{Ker} f$) is highlighted in green.

General functions (see fig. 2), and in particular linear functionals f, on a linear vector space X may be visualized by describing their *level sets*

$$\{x \in X \ : \ f(x) = c\}$$

for various values $c \in \mathbb{C}$. The level set corresponding to $c = 0$ is the kernel of f, $\operatorname{Ker} f$. It turns out that $\operatorname{Ker} f$ is a hyperplane, i.e. a subspace of X of codimension 1. All other level sets of f are obviously the translates of $\operatorname{Ker} f$ (see fig. 3). Moreover, there is a canonical correspondence between the linear functionals and the hyperplanes in X. This is clarified in the following proposition.

4.1.2 Proposition (Linear functionals and hyperplanes)**.**
Let X be a linear vectors space.

(i) *For every linear functional f on X,* $\operatorname{Ker} f$ *is a hyperplane in X, i.e.* $\operatorname{codim} \operatorname{Ker} f = 1$.

(ii) *If $f, g \neq 0$ are linear functionals on X such that* $\operatorname{Ker} f = \operatorname{Ker} g$, *then $f = a \cdot g$ for some (constant) scalar $a \neq 0$.*

(iii) *For every hyperplane $X_0 \in X$ there exists a linear functional $f \neq 0$ such that* $\operatorname{Ker} f = X_0$.

Proof. (i) Follows from a linear version of the fundamental theorem on homomorphisms, Exercise 3.4.11. Indeed, the injectivization

$$\tilde{f} : X/\operatorname{Ker} f \to \mathbb{C}$$

of f establishes a linear bijection (isomorphism) between $X/\operatorname{Ker} f$ and the range $\mathbb{C}$ of f. Thus $\dim(X/\operatorname{Ker} f) = \dim \mathbb{C} = 1$, so $\operatorname{Ker} f$ is a hyperplane in X.

(ii) Since $\operatorname{Ker} f = \operatorname{Ker} g =: X_0$, the injectivizations $\tilde{f}, \tilde{g} : X/X_0 \to \mathbb{C}$ are linear functionals on the *one-dimensional* space X/X_0. A moment's thought yields that such linear functionals must be equal up to some constant factor a, i.e. $\tilde{f}[x] = a\tilde{g}[x]$ for all $x \in X$. On the other hand, by construction of injectivization, $f(x) = \tilde{f}[x]$ and $g(x) = \tilde{g}[x]$. Therefore $f(x) = a \cdot g(x)$ as required.

(iii) Since $\dim(X/X_0)$, we have

$$X_0 = \{k[x_1]\}_{k \in \mathbb{C}}.$$

for some $x_1 \in X \setminus X_0$. Let $x \in X$ be arbitrary; then $[x] = k[x_1]$ for some $k = k(x) \in \mathbb{C}$, which implies $x = kx_1 + h$ for some $h \in X_0$. Let us define f on X by $f(x) := k(x)$. Then f is a linear functional (why?), and clearly $\operatorname{Ker} f = X_0$. $\square$

4.1.3 Proposition. *A linear functional on a normed space is continuous if and only if its kernel is closed.*

Proof. It immediately follows from the continuity of the functional f that the preimage $\ker f$ of a closed set $\{0\}$ must be closed.

It remains to prove (from the opposite) that the closeness of $\ker f$ implies the boundedness of f. So let's $\ker f = \overline{\ker f}$, and f is not continues. By proposition 1.9.7, the continuity of a functional is equivalent to its continuity at zero, therefore, for a discontinuous functional f there is such a sequence of elements $x_n \to \mathbf{0}$, that $|\, f(x_n)\,| \geq C$ for some $C > 0$. But then $y_n := \frac{Cx_n}{f(x_n)} \to \mathbf{0}$ too, and $f(y_n) \equiv C$, and $z_n := y_1 - y_n \to y_1$, and $z_n \in \ker f$ for all n. But $\lim z_n = y_1 \notin \ker f$, as $f(y_1) = C > 0$, so $\ker f$ is not closed. The resulting contradiction proves the required statement. $\square$

4.1.4 Exercise. Let X be a normed space. The kernel of a linear functional f on X is either closed or dense in X.

4.2 Representation theorems for linear functionals

In concrete Banach spaces, the bounded linear functionals usually have a specific and useful form. Generally speaking, all linear functionals on function spaces (such as L^p and $C(K)$) act by integration of the function (with respect to some weight or measure). Similarly, linear functionals on sequence spaces (such as l^p and c_0) act by summation with weights.

Dual of a Hilbert space: Riesz representation theorem

We start by characterizing bounded linear functionals on a Hilbert space H.

The following remark will be useful to the next theorem.

4.2.1 Remarks. 1) *Let H be a Hilbert space and Y be a closed subspace. Then every vector $x \in H$ can be uniquely represented as*

$$x = y + z, \quad y \in Y, \quad z \in Y^{\perp}. \tag{4.1}$$

This orthogonal decomposition is usually abbreviated as

$$H = Y \oplus Y^{\perp}.$$

2) *For an arbitrary subspace $Y \subseteq H$, $Y^{\perp}$ is closed.*
3) *For a closed $Y \subseteq H$ is true $\left(Y^{\perp}\right)^{\perp} = Y$.*

here, statement 1) is a reformulation of the orthogonality principle 1.8.4 2). Statement 2) follows from the continuity of the inner product, and 3) from 2) and (4.1) (from definition of $Y^{\perp}$ we have that $Y \subseteq \left(Y^{\perp}\right)^{\perp}$, then

$$x = y + z = \tilde{z} + \tilde{y}, \quad \tilde{z} \in Y^{\perp}, \quad \tilde{y} \in \left(Y^{\perp}\right)^{\perp} \quad z = \tilde{z},\ y = \tilde{y}$$

by representation uniqueness for $Y^{\perp}$.

The following theorem says that every functional f acts as an inner product with some vector in H.

4.2.2 Theorem (Riesz representation theorem)**.**
Let H be a Hilbert space.
(i) *For every $v \in H$, the function*

$$f(x) := \langle x, v \rangle \tag{4.2}$$

is a bounded linear functional on H, and its norm is $\|f\| = \|v\|$.
(ii) *For every bounded linear functional $f \in X^*$ there exists a unique vector $v \in H$ such that (4.2) holds.*

Proof. (i) has been proven in example 4.2.2.

(ii) Let $f \in X^*$. By Proposition 4.1.2, $\ker f$ is a hyperplane in H. Since f is bounded, $\ker f$ is closed (see Proposition 4.1.3). Therefore, H can be represented by the orthogonal decomposition

$$X = \ker f \oplus \operatorname{span}(v_0) \quad \text{for some} \quad v_0 \in (\ker f)^{\perp}.$$

(see remark 4.2.1). Consider the map

$$g(x) := \langle x, v_0 \rangle, \quad x \in H$$

We have $g \in X^*$ by part (i). Moreover, by 4.2.1 3)

$$\ker g = \{v_0\}^{\perp} = \ker f.$$

Therefore, by 4.1.2, the functionals f and g are equal up to some constant factor k, that is $f = kg$. It follows that

$$f(x) = k\langle x, v_0 \rangle = \langle kx, v_0 \rangle$$

and the conclusion follows with $v = kv_0$. Checking uniqueness is an easy exercise. □

4.2.3 Corollary. *An arbitrary Hilbert space H is isometrically isomorphic to its conjugate space H^* i.e. , there exists such a one-to-one linear map $L : H \to H^*$ that for any $v \in H$*

$$\|L(v)\|_{H^*} = \|v\|_H .$$

In a concise form, the last statement can be expressed as

$$H^* = H.$$

The following statement is actually a special case of the Riesz representation theorem 4.2.2 for the most important example of a Hilbert space, the L^p space.

4.2.4 Corollary ($(L^2)^* = L^2$). *Consider the space $L^2 = L^2(E, \mathcal{A}, \mu)$.*

(i) *For every weight function $v \in L^2$, integration with weight*

$$f(x) := f_v(x) := \int_E xv\, d\mu, \quad x \in L^2$$

is a bounded linear functional on L^2, and its norm is

$$\|f\| = \|v\|_2 = \left(\int_E |\, v \,|^2 \, d\mu \right)^{1/2}.$$

(ii) *Conversely, every bounded linear functional $f \in L^{2^*}$ can be represented as integration with weight for some unique weight function $v \in L^2$.*

Application: proof of Radon-Nikodym theorem.

Riesz representation theorem can be used to give a «soft» proof of Radon-Nikodym theorem in measure theory. This argument is due to von Neumann (1940). Consider two measures on the same σ-algebra. Recall that ν *is called* ***absolutely continuous*** *with respect to* μ, abbreviated $\nu \ll \mu$, *if*

$$\mu(A) = 0 \quad \text{implies} \quad \nu(A) = 0$$

for measurable sets A.

4.2.5 Theorem (Radon-Nikodym theorem for finite measures)**.**

Let μ and ν be two finite measures both defined on one σ-algebra $\mathcal{A} \subseteq 2^X$ such that $\nu \ll \mu$. Then there exists a non-negative μ-integrable function g such that

$$\forall \; A \in \mathcal{A} \qquad \nu(A) = \int_A g\, d\mu. \tag{4.3}$$

Moreover, g is uniquely determined μ-a.e.

A *function g satisfying (4.12) is called the* ***Radon-Nikodym derivative*** and is denoted $\frac{d\nu}{d\mu} := g$.

Proof. Existence. Let $\vartheta := \mu + \nu$, obviously, ϑ is countably additive, and it defined on $\mathcal{A}$. and for any $\mathcal{A}$-measurable $x \geq 0$ it holds

$$\int_X x\, d\,\vartheta = \int_X x\, d\,\mu + \int_X x\, d\,\nu,$$

as for $x = \chi_E$ it holds by definition of ϑ, for step x it holds by linearity, then it is true for other non-negative measurable ones, since according to Levy's theorem, going to the limit in equality is possible for an approximating sequence of step functions. Hence

$$x \in L^2(\vartheta) \;\Leftrightarrow\; |\,x\,|^2 \in L(\vartheta) \;\Leftrightarrow$$
$$|\,x\,|^2 \in L(\mu) \cap L(\nu) \;\Leftrightarrow\; x \in L^2(\mu) \cap L^2(\nu),$$

and (4.2) is also holds. So the linear functional

$$F(x) := \int_X x\, d\mu, \quad \text{for} \quad x \in L^2(X,\vartheta) \subseteq L^2(X,\mu).$$

is well defined (on $L^2(X,\vartheta)$) and by Riesz representation theorem, there exists $h \in L^2(\vartheta)$ such that for all $x \in L^2(\vartheta)$

$$F(x) = \int_X xh\, d\vartheta = \int_X xh\, d\mu + \int_X xh\, d\nu \tag{4.4}$$

Rearranging the terms (all the integrals involved are finite), we obtain

$$\int_X xh\, d\nu = \int_X x(1-h)\, d\mu \quad \text{for all} \quad x \in L^2(\vartheta). \tag{4.5}$$

We claim that

$$0 < h \leq 1 \quad \mu\text{−a. e.} \tag{4.6}$$

Indeed, consider the set $A = \{h \leq 0\}$ and the indicator function $x_0 := \chi_A$. In this case using the first equality of (4.4) one will get

$$\mu(A) = \int_X x_0\, d\mu = F(x_0) = \int_A h\, d\vartheta \leq 0,$$

hence $\mu(A) = 0$. Similarly, consider the set $B := \{h > 1\}$ and the indicator function $x_1 = \chi_B$. If $\mu(B) > 0$ then

$$\mu(B) = \int_X x_1 \, d\mu = F(x_1) = \int_B h \, d\vartheta > \mu(B),$$

a contradiction. This proves (4.6).

Since $\nu \ll \mu$, we moreover have

$$0 < h \leq 1 \quad \vartheta\text{−a. e.} \tag{4.7}$$

Now, given a measurable set A, we choose x so that $xh = \chi_A$. In other words, we consider

$$x := \frac{\chi_A}{h}$$

and apply the identity in (4.5). We obtain

$$\nu(A) = \int_A \frac{1-h}{h} d\mu. \tag{4.8}$$

The proof of existence is complete with $g := \frac{1-h}{h}$. □

4.2.6 Exercise. Prove uniqueness up to almost everywhere in the Radon-Nicodym theorem.

Below we will use the symbol

$$\{X_i\}_i \vdots X$$

if $\{X_i\}_i$ is a partition of X, that is X is a disjunct union of measurable sets $\{X_i\}_i$.

4.2.7 Corollary (Radon-Nikodym theorem for σ-finite measures)**.** *Let μ and ν be two σ-finite measures both defined on one σ-algebra $\mathcal{A} \subseteq 2^X$ such that $\nu \ll \mu$. Then there exists a non-negative measurable function g such that*

$$\forall \ A \in \mathcal{A} \qquad \nu(A) = \int_A g \, d\mu. \tag{4.9}$$

Moreover, g is uniquely determined μ-a.e.

Proof. By definition of σ-finiteness, there are two countable partitions $\{X_i\}_i$ and $\{X'_j\}_j$ of X such that for all indexes $\mu(X_i) < \infty$ and $\nu(X'_j) < \infty$, then $\{X_i \cap X_j\}_{i,j}$ is a partition and both measures μ and ν are finite on its elements. Applying the Radon-Nikodym theorem to each of the sets, we find non-negative functions g_{ij} defined and measurable on $X_i \cap X_j$ such that

$$\forall\; E \in \mathcal{A} \;:\; E \subseteq X_i \cap X_j \qquad \nu(E) = \int_E g_{ij}\, d\mu.$$

Let define the function g on X by

$$g(x) := g_{ij}(x), \quad \text{if} \quad x \in X_i \cap X_j.$$

So for any $E \in A$

$$\nu(E) = \sum_{i,j} \nu(E \cap X_i \cap X'_j) = \sum_{i,j} \int_{E\cap X_i \cap X'_j} g d\mu = \int_E g\, d\mu.$$

$\square$

4.2.8 Definition. *Let $(X, \mathcal{A})$ be a measure space. An arbitrary countably additive function $\varphi : \mathcal{A} \to \mathbb{C}$ is called a **charge**, or **signed measure** (on X). A charge is called **finite** if its value at any $E \in \mathcal{A}$ is expressed as a number ($\pm\infty$ is not allowed).*

If φ is a real charge then $\varphi(A)$ can be imagined as a total electric charge enclosed in volume A.

4.2.9 Examples. 1) The difference of measures. If μ, ν are two finite measure on one measure space (X, μ). Then

$$\varphi(E) := \mu(E) - \nu(E) \quad \text{for} \quad E \in \mathcal{A}$$

is a finite charge.

2) Let $(X, \mathcal{A}, \mu)$ be any measure space, $g \in L(X, \mu)$. Then

$$\varphi(E) := \int_E g(x) d\mu \quad \text{for} \quad E \in \mathcal{A} \tag{4.10}$$

Do the last two examples describe different or identical charges? Are there finite charges other than 1) and 2)?

Our goal will be to prove that not only every integrable function generates a charge, but, on the contrary, under certain conditions every charge can be represented in the form (4.10) with some function $g(x)$.

4.2.10 Proposition (Continuity of a charge)**.**
Any finite charge φ, defined on a measure space $(X, \mathcal{A})$ like a finite measure, has the properties of lower continuity and upper continuity, that is if

$$\forall\ k \in \mathbb{N} \quad A_k \in \mathcal{A}$$

and

1. *If $A_k \nearrow A$ then $\varphi A_k \underset{k\to\infty}{\longrightarrow} \varphi A$.*
2. *If $A_k \searrow A$ then $\varphi A_k \underset{k\to\infty}{\longrightarrow} \varphi A$.*

Proof. 1. If $A_k \nearrow A$, let's define $A := \cup_{k=1}^{\infty} A_k$, then

$$A = A_1 \cup \bigcup_{k=1}^{\infty} (A_{k+1} \setminus A_k),$$

which is a disjunct union, hence, by virtue of countable additivity, we have

$$\varphi(A) = \varphi(A_1) + \sum_{k=1}^{\infty} \varphi(A_{k+1} \setminus A_k) =$$

$$\varphi(A_1) + \lim_{n\to\infty} \sum_{k=1}^{n} \varphi(A_{k+1} \setminus A_k) = \lim_{n\to\infty} \varphi(A_n).$$

2. $A_k \searrow A$ then

$$A_1 = \bigcup_{k=1}^{\infty} (A_k \setminus A_{k+1}) \cup \left(\bigcap_{k=1}^{\infty} A_k \right), \quad A_n = \bigcup_{k=n}^{\infty} (A_k \setminus A_{k+1}) \cup \left(\bigcap_{k=1}^{\infty} A_k \right)$$

which are again disjunct unions, hence, by virtue of countable additivity, we have

$$\varphi(A_1) = \sum_{k=1}^{\infty} \varphi(A_k \setminus A_{k+1}) + \varphi\left(\bigcap_{k=1}^{\infty} A_k\right) =$$

The series converges, which means that its remainder tends to zero, and we get

$$\varphi(A_n) = \sum_{k=n}^{\infty} \varphi(A_k \setminus A_{k+1}) + \varphi\left(\bigcap_{k=1}^{\infty} A_k\right) \to \varphi\left(\bigcap_{k=1}^{\infty} A_k\right)$$

□

In this subsection, we consider only measurable sets, so we won't mention this every time.

4.2.11 Theorem. *Every finite charge is bounded. That is if $\varphi : \mathcal{A} \to \mathbb{C}$ is a finite charge then there is some $C \geq 0$ that $|\varphi(E)| \leq C$ for any $E \in \mathcal{A}$.*

Proof. We prove the opposite without loss of generality by assuming the charge to be real and upper unbounded. Thus

$$\sup_{E \in \mathcal{A}} \varphi(E) = +\infty. \tag{4.11}$$

We will prove that there is a disjunct or decreasing sequence $(E_k)_k$ such that $\varphi(E_k) \to +\infty$.

Consider the following statement $S(A, a)$, the truth of which may depend on the set $A \in \mathcal{A}$ and the real parameter a.

$$\begin{cases} \sup \varphi(A) = +\infty, \\ \forall\ E \subseteq A\ :\ \varphi(E) \geq a \quad \exists\ E' \subseteq E, \quad \varphi(E') > \varphi(E) + 1 \end{cases} \tag{$S(A,a)$}$$

1) If $(S(A, a))$ is true for some $A \in \mathcal{A}$ and some $a \in \mathbb{R}$, then there is a decreasing sequence $(E_k)_k$ such that $\varphi(E_{k+1}) > \varphi(E_k) + 1$ for all $k \in \mathbb{N}$ thus $\varphi(E_k) \to +\infty$.

Indeed, due to unboundedness φ on A, there is some $E_1 \subseteq A$: $\varphi(E) \geq a$, then by virtue of $(S(A,a))$ there is $E_2 := E'$, such that $\varphi(E_2) > \varphi(E_1) + 1 \geq a + 1 \geq a$, so we can find $E_3 \subseteq E_2$, such that $\varphi(E_3) > \varphi(E_2) + 1 \geq a$ and so on.

2) Let $(S(A,a))$ be not true for any $A \in \mathcal{A}$ and any $a \in \mathbb{R}$. Let $X_1 := X$. By assumption φ is upper unbounded and $S(X_1, 1)$ is not true. So

$$\exists\ E_1 \subseteq X_1 \quad : \quad \varphi(E_1) \geq 1, \quad \forall\ E' \subseteq E_1$$
$$\varphi(E') \leq \varphi(E_1) + 1 \quad \Rightarrow \quad \varphi\Big|_{E_1} \text{ is upper bounded.}$$

Let $X_2 := X_1 \setminus E_1$. $S(X_2, 2)$ is not true and $\varphi\Big|_{X_2}$ is upper unbounded (else φ is upper bounded on $E_1 \cup X_2 = X$). Then

$$\exists\ E_2 \subseteq X_2 \quad : \quad \varphi(E_2) \geq 2, \quad \forall\ E' \subseteq E_2$$
$$\varphi(E') \leq \varphi(E_2) + 1, \quad \varphi\Big|_{E_2} \text{ is upper bounded.}$$

Let $X_3 := X_2 \setminus E_2 \setminus E_1 = X_2 \setminus (E_2 \cup E_1)$. $S(X_3, 3)$ is not true and $\varphi\Big|_{X_3}$ is upper unbounded (else φ is upper bounded on $E_1 \cup E_2 \cup X_3 = X$). Then

$$\exists\ E_3 \subseteq X_3 \quad : \quad \varphi(E_3) \geq 3, \quad \forall\ E' \subseteq E_3$$
$$\varphi(E') \leq \varphi(E_3) + 1, \quad \varphi\Big|_{E_3} \text{ is upper bounded.}$$

And so on. As a result, there is a sequence $(E_k)_{k=1}^{\infty}$, such that for all

$$\forall\ k \in \mathbb{N} \quad \varphi(E_k) \geq k, \quad \varphi\Big|_{E_k} \text{ is upper bounded,}$$
$$(E_k) \text{ is disjunct.}$$

But the existence of the found sequence $(E_k)_{k=1}^{\infty}$ in both case 1) and case 2) contradicts the countable additivity of the charge because:

1) (decreasing) $\varphi(E_k) \to \varphi\left(\cap_{i=1}^{\infty} E_i\right) = \infty$, by upper continuity of φ (see proposition 4.2.10).

2) (dis) $\varphi\left(\cup_{k=1}^{\infty} E_k\right) = \sum_{k=1}^{\infty} \varphi(E_k) \geq \sum_{k=1}^{\infty} k = \infty$.

The contradiction in both cases. □

4.2.12 Definition. *Given charge φ and measurable set A, A is* ***positive*** *relative to the charge if every φ-measurable $B \subseteq A$ has $\varphi(B) \geq 0$. A is* ***negative*** *relative to the charge if every φ-measurable $B \subseteq A$ has $\varphi(B) \leq 0$.*

4.2.13 Definition. *Let φ be a real charge on a measure space $(X, \mathcal{A})$. A measurable partition of X consisting of one positive and one negative relative to the charge φ sets, is called* ***Hahn decomposition*** *for φ.*

4.2.14 Theorem (Hahn decomposition)**.** *For any real finite charge there exists its Hahn decomposition. The Hahn decomposition is unique up to subsets of zero charge (if X_+, X'_+ and X_-, X'_- are positive and negative sets of a finite charge φ respectively, $\{X_+, X_-\}$ and $\{X_+, X_-\}$ form partitions of X, then $\varphi(X_+ \triangle X_+) = 0$).*

Proof. By the previous theorem

$$\sup_{E \in \mathcal{A}} \varphi(E) = c < \infty.$$

For any $n \in \mathbb{N}$ let A_k be such a subset that $\varphi(A_k) > c - \frac{1}{2^k}$,

$$B_n := \bigcup_{k=n}^{\infty} A_k, \quad X_+ := \bigcap_{n=1}^{\infty} B_n, \quad X_- := X \setminus X_+.$$

To verify that $\{X_+, X_-\}$ is the desired decomposition, calculate $\varphi(X_+)$.

For any measurable A, B one has

$$\varphi(A \cup B) = \varphi(A) + \varphi(B) - \varphi(A \cap B).$$

Hence,

$$\varphi(A_i \cap A_j) \leq C \quad \Rightarrow$$
$$\varphi(A_i \cup A_j) = \varphi(A_i) + \varphi(A_j) - \varphi(A_i \cap A_j) >$$
$$> c - \frac{1}{2^i} + c - \frac{1}{2^j} - c = c - \left(\frac{1}{2^i} + \frac{1}{2^j}\right) \quad \Rightarrow$$
$$\varphi\left(\bigcup_{i=1}^{n} A_i\right) \geq c - \sum_{i=1}^{n} \frac{1}{2^i}, \quad \Rightarrow$$
$$\varphi\left(\bigcup_{i=1}^{\infty} A_i\right) = \lim_{n\to\infty} \varphi\left(\bigcup_{i=1}^{n} A_i\right) \leq c - 1$$
$$\varphi(B_n) = \varphi\left(\bigcup_{i=n}^{\infty} A_i\right) \geq c - \sum_{i=n}^{\infty} \frac{1}{2^i} = c - \frac{1}{2^{n-1}},$$

and $\varphi(X_+) = \lim_{n\to\infty} \varphi(B_n) \geq c \quad \Rightarrow \quad \varphi(X_+) = c.$

If there is $E \subseteq X_+$, such that $\varphi(E) < 0$, then $\varphi(X_+ \setminus E) = \varphi(X_+) - \varphi(E) > c$, that is impossible. Similarly, for any $E \subseteq X_- \quad \varphi(E) \leq 0$ else $\varphi(X_+ \cup E) = \varphi(X_+) + \varphi(E) > c$. Hence X_+, X_- are the ones we are looking for.

If X'_+, X'_- is another pair forming the Hahn decomposition, then $E \subseteq X_+ \setminus X'_+ \qquad E \subseteq X_+, E \subseteq X'_-$ from which $\varphi(E) = 0$. □

4.2.15 Definition. *Two measures μ_1 and μ_2 are* ***mutually singular*** *if there is a set A with $\mu_1(A) = 0$ and $\mu_2(A^c) = 0$ In this case we say μ_1 is* ***singular with respect to*** *μ_2 and write*

$$\mu_1 \perp \mu_2.$$

Exercise. $\mu_1 \perp \mu_2$ if and only if for any measurable E there is such its partition E_1, E_2, that $\mu_1(E_i) \cdot \mu_2(E_i) = 0$.

Example. 1) The Dirac delta measure, uniform measure on the Cantor set and the Lebesgue measure are pairwise mutually singular.

2) The uniform measure on the Cantor set is singular with respect to Lebesgue measure.

4.2.16 Corollary (Jordan decomposition of charges)**.** *Let* $(X, \mathcal{A})$ *be a measure space. The following statements are equivalent.*

(I) φ *is finite real charge on* $\mathcal{A}$.

(II) *There are mutually singular finite measures* φ_+ *and* φ_- *so that*

$$\varphi = \varphi_+ - \varphi_-.$$

Moreover, the pair of measures defined by condition (II) *is unique.*

Proof. Let $X = X_+ \cup X_-$ be a Hahn decomposition. Define

$$\varphi_+(E) := \varphi(E \cap X_+), \quad \varphi_-(E) = \varphi(E \cap X_-).$$

This gives a decomposition as required. To prove the uniqueness, let ν_+ and ν_- be singular measures, such that $\varphi = \nu_+ - \nu_-$. Due to their mutually singularity there is such a set A that $\nu_-(D) = 0$ and $\nu_+(D^c) = 0$. Then D and D^c are positive and negative sets of φ. By the uniqueness of the Hahn decomposition, X_+ and D differ on a null set, so that

$$\varphi_+(E) = \varphi(E \cap X_+) = \varphi(E \cap D) =$$
$$= \nu_+(E \cap D) - \nu_-(E \cap D) = \nu_+(E \cap D) = \nu_+(E),$$

then $\nu_-(E) = \nu_+(E) - \varphi(E) = \varphi_+(E) - \varphi(E) = \varphi_-(E)$, which concludes the proof. □

4.2.17 Theorem (Radon-Nikodym theorem for charges)**.**

Let μ *be a* σ*-finite measure and let* φ *be a finite real or complex charge both defined on one* σ*-algebra* $\mathcal{A} \subseteq 2^X$ *such that* $\varphi \ll \mu$. *Then there exists a* μ*-integrable function* g *such that*

$$\forall\ A \in \mathcal{A} \qquad \varphi(A) = \int_A g\, d\mu. \tag{4.12}$$

Moreover, g *is uniquely determined* μ*-a.e.*

Proof. Let $\nu := \operatorname{Re}\varphi$ be a real charge and let $X = X_+ \cup X_-$ be a Hahn decomposition for this charge ν. Then $\nu\big|_{X_+}$ is a finite and absolutely continuous with respect to $\mu\big|_{X+}$ measure. Then by theorem 4.2.7 there is a defined on X_+ non-negative measurable derivative $g_+ := \frac{d\nu}{d\mu}$. Let's continue g_+ on X with zero, $g_+\big|_{X_-} := 0$. Then for any $E \in \mathcal{A}$ we get

$$\nu(E \cap X_+) = \int_{E\cap X_+} g_+ d\mu = \int_E g_+ d\mu.$$

Similarly, there is a non-negative everywhere function g_- such that for any $E \in \mathcal{A}$ we get

$$-\nu(E \cap X_-) = \int_E g_- d\mu.$$

Let $g := g_+ - g_-$. Then for any $E \in \mathcal{A}$

$$\int_E g\, d\mu = \nu(E \cap X_+) + \nu(E \cap X_-) = \nu(E),$$

by Radon-Nikodym theorem for σ-finite measures 4.2.7 such function g is «almost unique». And $g \in L^1(\mu)$, as

$$|\, g\,| = g_+ + g_-, \quad \int_X |\, g\,|\, d\mu = \nu(X_+) - \nu(X_-) < \infty.$$

Completing the proof is a simple exercise. □

4.2.18 Definition. *Let φ be a real charge on a measure space $(X, \mathcal{A})$, and let $\{X_+, X_-\}$ form a Hahn decomposition for this charge. Then defined on $\mathcal{A}$ function* $\operatorname{var}\varphi$

$$\operatorname{var}\varphi(E) := \varphi(E \cap X_+) - \varphi(E \cap X_-), \quad E \in \mathcal{A}$$

is called the ***variation of the charge*** *φ. The number* $\operatorname{var}\varphi(X)$ *is called the* ***total variation of the charge*** *φ.*

4.2.19 Proposition (elementary variation's of the charge properties)**.** *Let φ be a finite charge on a measure space $(X, \mathcal{A})$. Then*

(I) $\operatorname{var}(\varphi)$ *is a measure.*

(II) *if for some σ-finite measure μ, defined on $\mathcal{A}$ it holds $g = \frac{d\varphi}{d\mu}$, then $|\, g \,|$ is a $\frac{d \operatorname{var} \varphi}{d\mu}$.*

The latter explains another designation of total variation:

$$\|\varphi\| := \operatorname{var} \varphi(X).$$

4.2.20 Definition. *Let $(X, \mathcal{A})$ be a space with a measure space, φ is a finite charge on $\mathcal{A}$*

$$\int\limits_X x(t) d\varphi := \int\limits_{X_+} x(t) d\varphi - \int\limits_{X_-} x(t) d(-\varphi) \quad \text{if } \varphi \text{ is real,}$$

if x integrable on E, that is $\int_{E \cap X_+} |\, x \,|\, d\varphi < +\infty$ and $\int_{E \cap X_-} |\, x \,|\, d(-\varphi) < +\infty$. And

$$\int\limits_X x(t) d\varphi := \int\limits_X x(t) d \operatorname{Re} \varphi + i \int\limits_X x(t) d \operatorname{Im} \varphi \quad \text{if } \varphi \text{ is complex,}$$

if both integrals on the right side of the equality exist and are finite.

A fact that we may come back to discuss later:

4.2.21 Exercises. For any φ-integrable on X function y, the following function $\psi : \mathcal{A} \to \mathbb{C}$,

$$\psi(E) := \int_E y \, d\varphi, \quad E \in \mathcal{A}$$

with fixed y is a charge.

2) The set of all ψ-integrable functions coincides with $L^1(X, \operatorname{var} \varphi)$.

The dual of L^p.

A modification of the representation theorem for L^2, Corollary 4.2.4, holds in fact for all L^p spaces (except what p?). In short, it states that $(L^p)^* = L^q$, where p and q are conjugate exponents as in Hölders inequality, i.e.

$$\frac{1}{p} + \frac{1}{q} = 1, \quad 1 < p, q < \infty,$$

and $q = \infty$ if $p = 1$. The rigorous statement is the following:

4.2.22 Theorem $((L^p)^* = L^q)$**.**

Let $L^p := L^p(X, \mathcal{A}, \mu)$ with σ-finite measure, and $p \in [1, +\infty)$. Let q be the conjugate exponent of p.

(i) *For every weight function $y \in L^q$, integration with weight*

$$f(x) := f_y(x) := \int\limits_E xy \, d\mu, \quad x \in L^p$$

is a bounded linear functional on L^p, and its norm is

$$\|f\| = \|y\|_q = \left(\int\limits_E |\, y \,|^q \, d\mu \right)^{1/q}.$$

(ii) *Conversely, every bounded linear functional $f \in (L^p)^*$ can be represented as integration with weight for some unique weight function $y \in L^q$.*

So the mapping $y \mapsto f_y$ is an isometric isomorphism between (L^q) and $(L^p)^*$

Proof.

(i) follows from Hölders inequality. example 3.2.4

(ii) *Existence* for case of finite μ.

Step 1) of finite μ. Let's define

$$\varphi(E) := f(\chi_E) \quad E \in \mathcal{A}.$$

φ is defined correctly, since $\chi_E \in L^\infty \subseteq L^p$ due to the finiteness of μ. Let's check that φ is a finite charge that is absolutely continuous with respect to μ.

Let $(E_k)_{k=1}^{\infty}⁝E$, all measurable. As a positive series $\chi_E = \sum_{k=1}^{\infty} \chi_{E_k}$ converges pointwise and lists partial sums are limited by the integrable constant 1, according to by Lebesgue's theorem on dominated convergence

$$\int_X \sum_{k=1}^{n} \chi_{E_k} d\mu \to \int_X \chi_E d\mu \quad \Rightarrow \quad \int_X \sum_{k=n+1}^{\infty} \chi_{E_k} d\mu \to 0 \quad \Rightarrow$$

$$\left(\int_X \sum_{k=n+1}^{\infty} \chi_{E_k} d\mu \right)^{(1/p)} = \left(\int_X \left(\sum_{k=n+1}^{\infty} \chi_{E_k} \right)^p d\mu \right)^{(1/p)} \to 0$$

The last means that $\sum_{k=1}^{\infty} \chi_{E_k}$ converges in L^p. Then using the continuity of the functional f we obtain

$$\varphi(E) = f(\chi_E) = f\left(\sum_{k=1}^{\infty} \chi_{E_k} \right) = \sum_{k=1}^{\infty} f\left(\chi_{E_k}\right) = \sum_{k=1}^{\infty} \varphi(E_k).$$

Thus, φ is countably additive, so it is a charge, and it's finite by its definition. And $\mu(E) = 0 \Rightarrow \chi_E = \mathbf{0}$ in $L^p \;\Rightarrow f(\chi_E) = 0 = \varphi(E)$. That is $\varphi \ll \mu$. By Radon–Nikodym theorem there exists $y \in L^1(\mu)$ such that

$$f(\chi_E) = \varphi(E) = \int_X \chi_E \cdot y \, d\mu = \int_E y \, d\mu.$$

Consider a linear bounded functional $g(x) := \int_X x \cdot y \, d\mu$ on L^∞. It coincides with f on a dense subset of the L^∞ space, on the linear span of the set of all characteristic functions. And $f \in (L^p)^* \subseteq (L^\infty)^*$. Due to the continuity of f and g on L^∞, this means that these functions coincide on whole L^∞ space.

But why $y \in L^q(\mu)$? The Radon-Nikodym theorem only guarantees that $y \in L^1$, but generally speaking, $L^1 \nsubseteq L^q$, actually $L^q \subseteq L^1$.

1) If $p \in (1, +\infty)$ let's define

$$A_n = \{t \;:\; |\, y(t) \,| \leq n\}, \quad A_n^+ = \{t \;:\; 0 < |\, y(t) \,| \leq n\}$$

Consider

$$y_n(t) := |\, y(t) \,|^{q/p} \cdot e^{-i \arg y(t)} \cdot \chi_{A_n^+}(t)$$

Then, on the one hand, for all $n \in \mathbb{N}$

$$f(y_n) = \int_{A_n^+} |\, y \,|^{q/p} \,|\, y \,|\, d\mu = \int_{A_n} |\, y \,|^q \, d\mu,$$

as $1 + \frac{q}{p} = 1 + q(1 - \frac{1}{q}) = q$.

On the other hand, $y_n \in L^\infty \subseteq L^p$, so for all $n \in \mathbb{N}$

$$|\, f(y_n) \,| \leq \|f\| \cdot \|y_n\|_p \,,$$

So there is an inequality

$$\int_{A_n} |\, y \,|^q \, d\mu \leq \|f\| \left(\int_{A_n} |\, y_n \,|^p \right)^{1/p} = \|f\| \left(\int_{A_n} |\, y \,|^q \right)^{1/p}$$

hence, for all $n \in \mathbb{N}$

$$\left(\int_X |\, y \,|^q \, \chi_{A_n} d\mu \right)^{1/q} \leq \|f\| \,.$$

By virtue of Levy's theorem, we conclude that $y \in L^q(X, \mu)$.

2) For $p = 1$ let's check that $y \in L^\infty$. Consider

$$A := \{t \;:\; |\, y(t) \,| > \|f\|\}.$$

It is enough for us to show that this set has a zero μ measure. Let's assume that $\mu(A) > 0$. Then define

$$x_0 := \frac{1}{\mu(A)} \cdot \chi_A \cdot e^{-i \arg y},$$

then $\|x_0\|_1 = 1$ and

$$f\,(x_0) = \frac{1}{\mu(A)} \int_A |\, y \,|\, d\mu > \frac{1}{\mu(A)} \int_A \|f\| \, d\mu = \|f\| \,.$$

Then we have

$$\|f\| = \sup_{\|x\|_1 \le 1} |\, f(x)\,| \ge f(x_0) > \|f\|\,.$$

The resulting contradiction proves that $\mu(A) = 0$. And so, almost everywhere $|\, y(t)\,| \le \|f\| \quad \Rightarrow \quad y \in L^\infty(\mu)$.

So we got the following result. For an arbitrary linear, continuous functional $f \in (L^p)^*$ for $p \in [1, +\infty)$ there is such a function $y(t) \in L^q(\mu)$ that there is equality $f(x) = \int\limits_E xy\, d\mu$ valid for all simple functions x, but, as is known, the set of all simple functions are dense in $L^p(\mu)$ at $p \in [1, +\infty)$. Therefore, we come to the statement of the theorem for the case $\mu(X) < \infty$.

$\square$

4.2.23 Exercise. * Prove the part (II) of the previous theorem for σ-finite measures.

4.2.24 Corollary $((l^p)^* = l^q)$**.**
Let $p \in [1, +\infty)$ and let q be the conjugate exponent of p.
(I) *For every $y \in l^q$, summation with weight*

$$f_y(x) := \sum_{k=1}^{\infty} x_k y_k$$

is a bounded linear functional on l^p, and its norm is $\|f_y\| = \|y\|_q$.
(ii) Every bounded linear functional $f \in (l^p)^*$ can be represented as summation with weight for some unique weight $y \in l^q$.

4.2.25 Exercise (*)**.** Prove that $c_0^* = l^1$. The meaning of this is the same as in previous corollary, i.e. the functionals on c_0 are given by summation with weight from l^1.

The dual of $C(K)$. Finally, we state without proof the following characterization of bounded linear functionals on $C(K)$.

4.2.26 Theorem ($(C(K))^*$, Riesz-Markov theorem)**.**

Let K be a compact topological space.

(i) *For every Borel regular charge φ on K, integration*

$$f(x) := f_\varphi(x) := \int_E x \, d\varphi, \quad x \in C(K)$$

is a bounded linear functional on $C(K)$, and its norm is the total variation

$$\|f\| = \operatorname{var} \varphi(K).$$

(ii) *Conversely, every bounded linear functional $f \in (C(K))^*$ can be represented as integration with respect to a unique Borel regular charge φ on K.*

4.3 Hahn–Banach theorem

Hahn-Banach theorem allows one to extend continuous linear functional f from a subspace to the whole normed space, while preserving the continuity of f. Hahn–Banach theorem is a major tool in functional analysis. Together with its variants and consequences, this result has applications in various areas of mathematics, computer science, economics and engineering.

4.3.1 Reminder. Let X be a normed space, and let X_0 be a subspace of X. Consider a bounded linear functional f_0 defined on X_0, i.e. $f_0 \in X_0^*$. An ***extension*** of f_0 to the whole space X is a bounded linear functional $f \in X^*$ whose restriction on X_0 coincides with f_0, i.e.

$$f\big|_{X_0} = f_0, \quad \text{meaning that} \quad f_0(x) = f(x) \quad \text{for all} \quad x \in X_0.$$

Constructing extensions is a nontrivial problem because of the continuity (=boundedness) requirement for f.

Extension by continuity

Before we state Hahn–Banach theorem, let us address the simpler problem of extending a continuous linear functional from a dense subspace to the whole space.

4.3.2 Proposition (Extension by continuity)**.**

Let X_0 be a dense subspace of a normed space X. Then every bounded functional $f_0 \in X_0^*$ admits a unique extension $f \in X^*$. Moreover,

$$\|f\| = \|f_0\|\,.$$

Proof. Let $x \in X$ be arbitrary. By density, we can find a sequence $(x_n) \subseteq X_0$ such that $x_n \to x$. Then $(f(x_n))$ is a Cauchy sequence, since

$$|\, f_0(x_n) - f_0(x_m)\,| \le \|f_0\|\, \|x_n - x_m\| \to 0, \quad \text{as} \quad n, m \ge N \to \infty.$$

By completeness of $\mathbb{R}$ or $\mathbb{C}$, this sequence converges. So we define

$$f(x) := \lim_{n\to\infty} f_0(x_n), \tag{4.13}$$

Let's check that f is the one we are looking for.

First of all, since the limit (4.13) exists for any Cauchy sequence (x_n), this limit it does not depend on the choice of of such a sequence, so f is well defined. Due to the linearity of the functional f_0 as well as the limit, the functional f is linear too. Finally, f is a bounded linear functional. Indeed, for $x_n \to x$ we have

$$|\, f(x)\,| = \lim_{n\to\infty} |\, f_0(x_n)\,| \le \|f_0\| \lim_{n\to\infty} \|x_n\| = \|f_0\|\, \|x\|\,.$$

This shows that $f \in X^*$ and $\|f\| \le \|f_0\|$. Note that the reverse inequality $\|f\| \ge \|f_0\|$ trivially holds (for any extension f). □

Now we address the more difficult problem of extending linear functionals from arbitrary subspaces. The result is the same as for dense subspaces, except the extensions need not be unique.

The Hahn-Banach theorem is based on Zorn's lemma, we will give the necessary definitions.

The Lemma of Zorn

4.3.3 Definition. *Arelation $\preceq$ on a set P is called a* ***partial order*** *if it is reflexive, anti-symmetric, and transitive, i.e. if it satisfies the conditions*

$$\begin{aligned} p &\preceq p, \\ p \preceq q,\ p \preceq q \quad &\Rightarrow \quad p = q, \\ p \preceq q,\ q \preceq r \quad &\Rightarrow \quad p \preceq r \end{aligned}$$

for all $p, q, r \in P$. A ***partially ordered set*** *is a pair $(P, \preceq)$ consisting of a set P and a partial order $\preceq$ on P.*

An element m of partially ordered set P is called ***maximal*** *if $m \not\preceq p$ for all $p \in P \setminus \{m\}$.*

A ***chain*** *in P is a* ***totally ordered subset*** *$C \subseteq P$, i.e. any two distinct elements $p, q \in C$ satisfy either $p \preceq q$ or $q \preceq p$.*

Let $C \subseteq P$ be a nonempty chain. An element $a \in P$ is called an ***upper bound*** *of C if every element $p \in C$ satisfies $p \preceq a$. It is called a* ***supremum*** *of C if it is an upper bound of C and every upper bound $b \in P$ of C satisfies $a \preceq b$. The supremum, if it exists, is unique and denoted by $\sup C$.*

4.3.4 Lemma (Zorn's lemma)**.** *Let $(P, \preceq)$ be a partially ordered set such that every nonempty chain $C \subseteq P$ admits an upper bound. Let $p \in P$. Then there exists a maximal element $m \in P$ such that $p \preceq m$.*

The Axiom of Choice. *Let I and X be two nonempty sets and, for each element $i \in I$, let $X_i \subseteq X$ be a nonempty subset. Then there exists a map $g : I \to X$ such that every $i \in I$ satisfies $g(i) \in X_i$.*

4.3.5 Theorem. *The axiom of choice is equivalent to the Lemma of Zorn.*

4.3.6 Definition. *Let X be a real vector space. A function $p : X \to \mathbb{R}$ is called a* ***quasi-seminorm*** *or* ***sublinear functions*** *if it satisfies*

$$p(x+y) \leq p(x) + p(y), \quad p(\lambda x) = \lambda p(x) \tag{4.14}$$

for all $x, y \in X$ *and all* $\lambda \geq 0$. *It is called a* ***seminorm*** *if it is a quasi-seminorm and*

$$p(\lambda x) = |\,\lambda\,|\, p(x)$$

for all $x \in X$ *and all* $\lambda \in R$.

A seminorm has nonnegative values, because $2p(x) = p(x)+p(-x) \geq p(0) = 0$ for all $x \in X$. Thus a seminorm satisfies all the axioms of a norm except nondegeneracy (i.e. there may be nonzero elements $x \in X$ such that $p(x) = 0$).

4.3.7 Theorem (Hahn–Banach)**.**

Let X *be a normed vector space and let* $p : X \to \mathbb{R}$ *be a quasi-seminorm. Let* $X_0 \subseteq X$ *be a linear subspace and let* $f_0 : X_0 \to \mathbb{R}$ *be a linear functional such that*

$$f_0(x) \leq p(x) \quad \textit{for all} \quad x \in X_0.$$

Then there exists a linear extension $f : X \to \mathbb{R}$ *of* f_0 *such that*

$$f(x) \leq p(x) \quad \textit{for all} \quad x \in X.$$

4.3.8 Lemma. *Let* $X, X_0, p,$ *and* f_0 *be as in Theorem 4.3.7. Let* $x_0 \in X \setminus X_0$ *and define* $X_1 := X_0 \oplus \mathbb{R}x_0$. *Then there exists a linear extension* $f_1 : X_1 \to \mathbb{R}$ *of* f *(on* X_1*) such that*

$$f_1(x) \leq p(x) \quad \textit{for all} \quad x \in X_1.$$

Let's prove the lemma first.

Proof. An extension $f_1 : X_0 \to \mathbb{R}$ of the linear functional $f_0 : X_0 \to R$ is uniquely determined by its value $a := f_1(x_0) \in \mathbb{R}$ on x_0, as

$$f_1(x + \lambda x_0) = f_0(x) + \lambda a \quad \text{for all} \quad x \in X_0.$$

This extension satisfies the required condition $f_1(x_1) \leq p(x_1)$ for all $x_1 \in X_1$ if and only if

$$f_0(x) + \lambda a \leq p(x + \lambda x_0) \quad \text{for all} \quad x \in X_0 \quad \text{and all } \lambda \in \mathbb{R}. \tag{4.15}$$

If this holds, then

$$f_0(x) \pm a \leq p(x \pm x_0) \quad \text{for all} \quad x \in X_0. \tag{4.16}$$

Conversely, if (4.16) holds and $\lambda > 0$, then

$$\begin{aligned} f_0(x) + \lambda a = \lambda(f_0(\lambda^{-1}x) + a) \leq \lambda p(\lambda^{-1}x + x_0) = p(x + \lambda x_0), \\ f_0(x) - \lambda a = \lambda(f_0(\lambda^{-1}x) - a) \leq \lambda p(\lambda^{-1}x - x_0) = p(x - \lambda x_0), \end{aligned}$$

This shows that (4.15) is equivalent to (4.16). Thus it remains to find a real number $a \in \mathbb{R}$ that satisfies (4.16). Equivalently, a must satisfy

$$f_0(x) - p(x - x_0) \leq a \leq p(x + x_0) - f_0(x)$$

for all $x \in X_0$. To see that such a number exists, fix two vectors $y, y' \in X_0$. Then

$$\begin{gathered} f_0(y) + f_0(y') = f_0(y + y') \leq p(y + y') = \\ p(y + x_0 + y' - x_0) \leq p(y + x_0) + p(y' - x_0). \end{gathered}$$

Thus

$$f_0(y') - p(y' - x_0) \leq p(y + x_0) - f_0(y)$$

for all $y, y \in X_0$ and this implies

$$\sup_{y' \in X_0} (f_0(y') - p(y' - x_0)) \leq \inf_{y \in X_0} (p(y + x_0) - f_0(y))$$

Hence there exists a real number $a \in \mathbb{R}$ that satisfies (4.3) and this proves this lemma. □

Now let's prove the theorem.

Proof. Let's define $\mathcal{P}$ as a set of all pairs $\{(Z, g)\}$ where Z is a linear subspace of X and $X_0 \leq Z \leq X$ and $g : Z \to \mathbb{R}$ is a linear functional such that

$$g\big|_{X_0} = f_0, \quad g(x) \leq p(x) \quad \text{for all} \quad x \in Z.$$

This set is partially ordered by the relation

$$(Z, g) \preceq (Z', g') \quad \Leftrightarrow \quad Z \leq Z' \quad \text{and} \quad g'\big|_Z = g$$

for $(Z, g), (Z', g') \in \mathcal{P}$. A chain in $\mathcal{P}$ is a totally ordered subset $\mathcal{C} \subseteq P$. Every nonempty chain $\mathcal{C} \subseteq P$ has a supremum (Z^*, g^*) given by

$$Z^* = \bigcup_{(Z,g)\in\mathcal{C}} Z, \quad g^*(x) := g(x) \quad \text{for all} \quad (Z, g) \in \mathcal{C} \quad \text{and all} \quad x \in Z.$$

Hence it follows from the Lemma of Zorn that $\mathcal{P}$ has a maximal element (Z, g). By Lemma 4.3.8 every such maximal element satisfies $Z = X$ and this proves Theorem. □

One of the popular formulations of the Hahn-Banach theorem concerns the case when the seminorm is a norm:

4.3.9 Theorem. *Let X_0 be a subspace of a normed space X. Then every functional $f_0 \in X_0^*$ admits an extension $f \in X^*$ such that $\|f\| = \|f_0\|$.*

Supporting functionals

Hahn-Banach theorem has a variety of consequences, both analytic and geometric. One of the basic tools guaranteed by Hahn-Banach theorem is the existence of a supporting functional $f \in X^*$ for every vector $x \in X$.

4.3.10 Proposition (Supporting functional)**.**

Let X be a normed space. For every $x \in X$ there exists $f \in X^*$ such that

$$\|f\| = 1, \quad f(x) = \|x\|\,. \tag{4.17}$$

The functional f satisfying 4.17 is called the ***supporting functional*** of x.

Proof. Consider the one-dimensional subspace $X_0 := \operatorname{span}(x)$, and define a functional $f_0 \in X_0^*$ by

$$f_0(tx) := t\,\|x\|\,.$$

Then $\|f_0\| = 1$. An extension $f \in X$ of f_0 guaranteed by Hahn-Banach theorem clearly satisfies the conclusion. $\square$

4.3.11 Exercises. 1) Consider a normed space $X = \mathbb{R}^2$ and a unit vector $x_0 \in X$. Let f be a supporting functional of x_0. Interpret geometrically the level set $\{x \ : \ f(x) = 1\}$ as a tangent hyperplane for the unit ball $B_1(\mathbf{0})$ at point x_0. Construct an example of a normed space for which the supporting functional of x is not unique.

2) (a variation of the exercise 1.9.11) Construct a bounded linear functional on $C[0,1]$ which does not attain its norm.

Recall that the norm of a functional $f \in X^*$ equals (see theorem 1.9.9)

$$\|f\| = \sup_{x \neq \mathbf{0}} \frac{|\, f(x) \,|}{\|x\|}$$

As the last exercise shows, generally it is not true that every functional f attains its norm on some vector x, i.e. that the supremum above can be replaced by the maximum.

However, every vector x does attain its norm on some functional $f \in X^*$, namely the supporting functional. This immediately follows from Proposition 4.3.10.

4.3.12 Corollary. *For every vector x in a normed space X, one has*

$$\|x\| = \sup_{f \neq \mathbf{0}} \frac{|\, f(x) \,|}{\|f\|}$$

Hahn-Banach theorem implies that there are enough bounded linear functionals $f \in X^*$ on every space X. One manifestation of this is the following:

4.3.13 Corollary (X^* separates the points of X)**.**

For every two vectors $x_1 \neq x_2$ in a normed space X, there exists a functional $f \in X^*$ such that $f(x_1) \neq f(x_2)$.

Proof. The supporting functional $f \in X^*$ of the vector $x := x_1 - x_2$ must satisfy $f(x_1 - x_2) = \|x_1 - x_2\| \neq 0$, as required. $\square$

Second dual space

Let X be a normed space as usual. The functionals f are designed to act on vectors $x \in X$ via $f : x \mapsto f(x)$. Vice versa, we can say that vectors $x \in X$ ***act on functionals*** $f \in X^*$ via

$$x : f \mapsto f(x), \quad f \in X^*. \tag{4.18}$$

Thus a vector $x \in X$ can itself be considered as a function from X^* to $\mathbb{R}$. This function (4.18) is clearly linear, so we may consider x as a linear functional on X^*. Also, the inequality

$$|\, f(x)\,| \leq \|x\| \cdot \|f\|$$

shows that this functional is bounded, so

$$x \in X^{**}$$

and the norm of x as a functional is $\|x\|_{X^{**}} \leq \|x\|$. Considering the supporting functional $f \in X^*$ of x we see that actually

$$\|x\|_{X^{**}} = \|x\|$$

We demonstrated that there exists a ***canonical embedding*** of X into into X^{**}. We summarize this as follows.

4.3.14 Theorem (Second dual space)**.**
Let X be a normed vector space. Then the linear map

$$I : X \mapsto X^{**}, \qquad I(x)(f) := f(x) \quad x \in X, f \in X^* \tag{4.19}$$

is an isometric embedding.

4.3.15 Definition. *A normed space X is called **reflexive** if $X^{**} = X$ (under the canonical embedding).*

4.3.16 Proposition. *Let X be a reflexive space. Then every functional $f \in X^*$ attains its norm on X.*

Proof. By reflexivity, the supporting functional of f is a vector $x \in X$ thus $\|x\| = 1$ and $f(x) = \|f\|$, as required. □

The converse of Proposition 4.3.16 is also true. If every functional $f \in X^*$ on a Banach space X attains its norm, then X is reflexive. This is James' theorem (1971).

4.3.17 Examples.

1) As we know from exisise 4.2.25, $c_0^* = l_1$ while $l_1^* = l_\infty$, so

$$c_0^{**} = l_\infty.$$

The space c_0 of sequences converging to zero is not reflexive and it is indeed canonically embedded into the larger space l^∞ of all bounded sequences (and with the same sup-norm).

2) As we know from Theorem 4.2.22, $(L^p)^* = L^q$ where $p \in [1, +\infty)$ and q is the conjugate index of p. Therefore, L^p is a reflexive space for $p \in (1, +\infty)$. One can show that L^1 and L^∞ are not reflexive spaces.

3) As we know from exisise 4.3.11 and proposition 4.3.16, $C[a, b]$ is not reflexive.

Separation of convex sets

4.3.18 Definition. *A subset E of a linear vector space X is called **absorbing** if*

$$X = \bigcup_{t \geq 0} tE$$

*(where $tE = \{tx \ : \ x \in E\}$). The **Minkowski functional** on X generated by the absorbing set E is*

$$p_E(x) = \inf\{t > 0 \ : \ x \in tE\} \quad x \in X.$$

4.3.19 Exercise (easy)**.** Let E be a subset of a normed space X such that $0 \in \operatorname{Int} E$. Then E is an absorbing set.

4.3.20 Proposition (Minkowski functional)**.**

(I) *Let E be a absorbing convex subset of a linear vector space X such that $\mathbf{0} \in E$. Then Minkowski functional generated by E is a quasi-seminorm* (see definition 4.3.6)*.*

(II) *For any quasi-seminorm p on a linear vector space X the sub-level set*

$$E = \{x \in X \ : \ p(x) \leq 1\}$$

is an absorbing convex set, and $\mathbf{0} \in E$.

Exercise. Prove Proposition 4.3.20.

Hahn-Banach theorem has some remarkable geometric implications, which are grouped together under the name of separation theorems. Under some mild topological requirements, these results guarantee that two convex sets A, B can always be separated by a hyperplane. As we know from 4.1.2, the hyperplanes correspond to the level sets of linear functionals f. Therefore, we expect that a separation theorem for A, B would give us a linear functional f and a number C such that

$$f(a) \leq C \leq f(b), \quad a \in A, \ b \in B$$

In this case, the sets A and B get separated by the hyperplane $\{x \ : \ f(x) = C\}$.

Let us start from the simpler case when one of the two sets is a point.

4.3.21 Theorem (Separating a point from a convex set)**.**

Let G be an open convex subset of a normed space X, and consider a point $x_0 \notin G$. Then there exists a functional $f \in X^*$, $f \neq 0$, such that

$$\forall \ x \in G \quad f(x) \leq f(x_0).$$

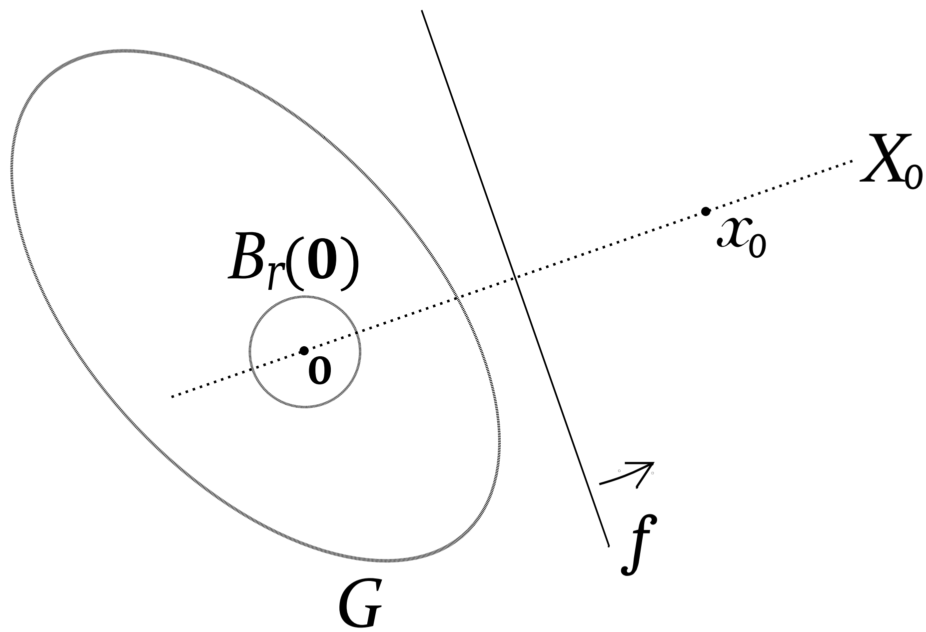


Figure 4: Separation of a point x_0 from the set G by a functional f

Proof. Translating G if necessary, we can assume without loss of generality that $\mathbf{0} \in G$ (why?). By Exercise 4.3.19, G is an absorbing set. Therefore, by Proposition 4.3.20, Minkowski functional p_G of G is a sublinear functional (or a quasi-seminorm) on X. Since $0 \in G$ and G is open, it contains a centered ball $B_r(\mathbf{0})$ for some radius $r > 0$ (see the figure 4). The set inclusion $B_r(\mathbf{0}) \subseteq G$ implies the inequality for the Minkowski functionals:

$$\forall\, x \in G \quad \|x\| \geq r \cdot p_G(x)$$

(why?)

Consider the one-dimensional subspace

$$X_0 = \operatorname{span}(x_0)$$

and define a linear functional f_0 on X_0 by

$$f_0(tx_0) := t\,\|x_0\| \quad t \in \mathbb{R}.$$

Then f_0 is dominated by p_G on X_0, since for $t \geq 0$ we have

$$f_0(tx_0) = p_G(tx_0)$$
$$f_0(-tx_0) = -t \cdot f_0(x_0) = -t \cdot p_G(x_0) \leq 0 \leq p_G(tx_0).$$

By Hahn-Banach theorem for quasi-seminorm (Theorem 4.3.7), f_0 admits an extension f onto the whole space X such that the domination is preserved, i.e.

$$\forall\, x \in X \qquad f(x) \le p_G(x).$$

To finish the proof, we need to check that f is bounded and that it separates x_0 from G as required. The boundedness follows from the inequality

$$f(x) \le p_G(x) \le \frac{1}{r}\,\|x\| \quad \text{for all } x \in X.$$

so $f \in X_0$. To check the separation, consider $x \in G$. Since $x_0 \notin G$, we have

$$f(x) \le p_G(x) \le 1 \le \|x_0\| = f_0(x_0) = f(x_0) \quad \text{for all } x \in X.$$

This completes the proof. □

4.3.22 Theorem (Separation of open convex sets)**.**

Let A, B be disjoint convex subsets of a normed space X.

(i) *If A is open, then there exists a functional $f \in X^*$ and a number $C \in \mathbb{R}$ such that*

$$f(a) < C \le f(b) \quad \textit{for all } a \in A,\ b \in B.$$

(ii) *If both A and B are open, then the stronger inequality holds:*

$$f(a) < C < f(b) \quad \textit{for all } a \in A,\ b \in B.$$

Proof. (i) Consider the Minkowski difference set

$$G := A - B := \{a - b \ :\ a \in A,\ b \in B\}.$$

The set G is open and convex (why?). Since A and B are disjoint, $0 \notin G$. Using the previous theorem (4.3.21), we obtain a functional $f \in X^*$, $f \neq 0$ such that

$$f(a - b) \le f(0) = 0 \quad \text{for all } a \in A,\ b \in B.$$

Hence $f(a) \leq f(b)$ for all $a \in A, b \in B$, so letting $C := \sup_{a \in A} f(a)$ we obtain

$$f(a) \leq C \leq f(b) \quad \text{for all } a \in A, \; b \in B.$$

Since A is open, by considering a small neighborhood of a in A we obtain the strict inequality

$$f(a) < C \leq f(b) \quad \text{for all } a \in A, \; b \in B.$$

(ii) This part is similar, and follows by considering small neighborhoods of a in A and of b in B. □

4.3.23 Corollary (Separation of closed convex sets)**.**

Let A, B be disjoint closed convex subsets of a normed space X. Assume B is compact. Then there exists a functional $f \in X^*$ such that

$$\sup_{a \in A} f(a) < \inf_{b \in B} f(b).$$

Proof. Let

$$r := \operatorname{dist}(A, B) := \inf_{a \in A, b \in B} \|a - b\|$$

By the assumptions, $r > 0$ (why?). Therefore, the open $r/3$-neighborhoods $A_{r/3}$ of A and $B_{r/3}$ are disjoint, open and convex sets. Applying Theorem 4.3.22, we obtain a functional $f \in X$ that separates the neighborhoods:

$$\sup_{a \in A_{r/3}} f(a) \leq \inf_{b \in B_{r/3}} f(b).$$

From this the conclusion easily follows. (How?) □

4.3.24 Remark. *Suppose that in Theorem 4.3.22, the set E is either open (i.e. as stated) or closed. Then the strict separation holds:*

$$f(x) < f(x_0), \quad x \in E.$$

Indeed, for open sets this follows from Theorem 4.3.22, while for closed sets this follows from Corollary 4.3.23.

Convex sets are intersections of half-spaces

4.3.25 Corollary. *Every closed convex subset K of a normed space X is the intersection of all (closed) half-spaces that contain K.*

Recall that the half-space is what lies on one side of a hyperplane; therefore half-spaces have the form

$$\{x \in X \ : \ f(x) \leq a\}.$$

See the picture illustrating Corollary.

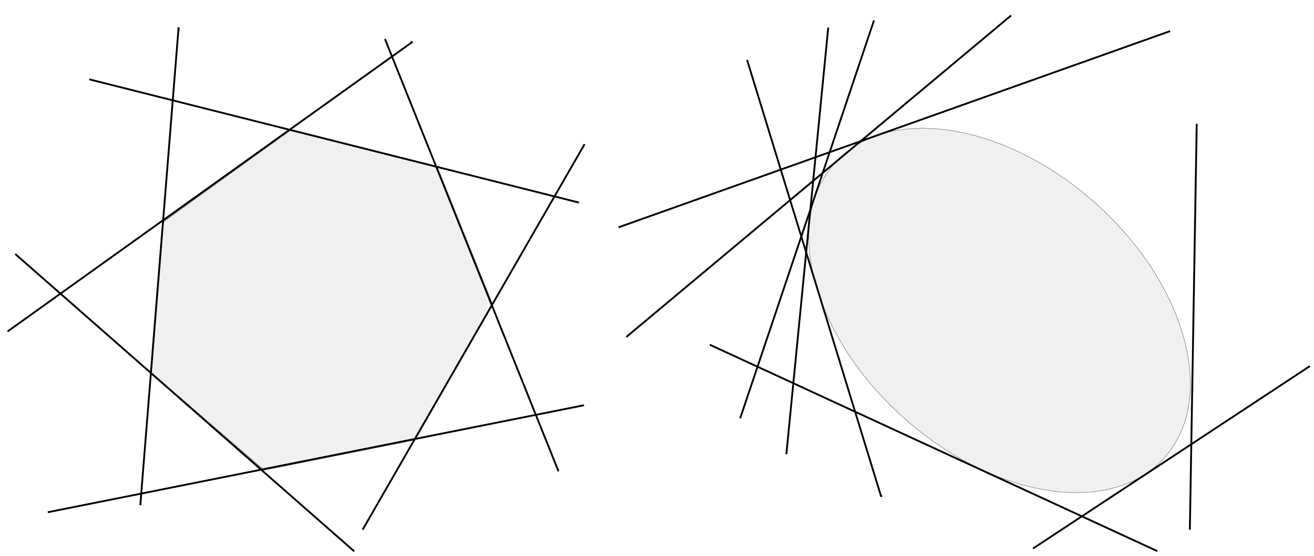


Figure 5: Convex sets in normed spaces are intersections of half-spaces

Proof. K is trivially contained in the intersection of the half-spaces that contain K. To prove the reverse inclusion, choose a point $x_0 \notin K$ and use Separation Theorem 4.3.23 for $A = K$ and $B = \{x_0\}$. We thus obtain a functional $f \in X^*$ such that

$$a := \sup_{x \in K} f(x) < f(x_0).$$

It follows that the half-space $\{x \in X \ : \ f(x) \leq a\}$ contains K but not x_0. This completes the proof. $\square$

4.3.26 Exercise (Closed convex sets that can not be strictly separated)**.** Show that the compactness assumption in Corollary 4.3.23 is essential. Construct two closed convex sets on the plane $\mathbb{R}^2$ that can not be strictly separated.

4.3.27 Exercise (Convex sets that can not be separated)**.** Consider the Hilbert space $H = l^2$ and define

$$A := \left\{ x \in l^2 \ : \ \exists\, n \in \mathbb{N} \quad \forall\, i \in \mathbb{N} \quad \begin{matrix} i < n \Rightarrow x_i > 0 \\ i \geq n \Rightarrow x_i = 0 \end{matrix} \right\},$$

$$B := \left\{ x \in l^2 \ : \ \exists\, n \in \mathbb{N} \quad \forall\, i \in \mathbb{N} \quad \begin{matrix} i < n \Rightarrow x_i = 0 \\ i \geq n \Rightarrow x_i > 0 \end{matrix} \right\}.$$

Show that A, B are nonempty disjoint convex subsets of l^2 with empty interior whose closures agree. If $f : l^2 \to \mathbb{R}$ is a bounded linear functional and c is a real number such that $f(x) \geq c$ for all $x \in A$ and $f(x) \leq c$ for all $x \in B$, show that $f = 0$ and $c = 0$.

4.3.28 Exercise (Functionals that annihilate a subspace)**.** Let X_0 be a closed subspace of a normed space X. Prove that there exists a functional $f \in X^*$ such that

$$f(x) = 0 \quad \text{for all} \quad x \in X_0.$$

You may deduce this from Hahn-Banach theorem directly or from a separation theorem.

5 The Weak and Weak* Topologies

5.1 Topological Vector Spaces

5.1.1 Definition. *A* ***topological vector space*** *is a vector space that is also a topological space with the property that the vector space operations (vector addition and scalar multiplication) are also continuous functions.*

To say that addition is continuous means, by definition, that the mapping

$$(x, y) \to x + y$$

of the cartesian product $X \times X$ into X is continuous: If $x_1, x_2 \in X$, and if V is a neighborhood of $x_1 + x_2$, there should exist neighborhoods V_i of x_i; such that

$$V_1 + V_2 \subseteq V.$$

Similarly, the assumption that scalar multiplication that is continuous means that the mapping

$$(\alpha, x) \to \alpha x$$

of $\mathbb{C} \times X$ (or $\mathbb{R} \times X$) into X is continuous: if $x \in X$, α is a scalar, and V is a neighborhood of αx, then for some $r > 0$ and some neighborhood W of x we have $\beta W \subseteq V$ whenever $|\,\beta - \alpha\,| < r$.

A subset E of a topological vector space is said to be ***bounded*** if to every neighborhood V of $\mathbf{0}$ in X corresponds a number $s > 0$ such that $E \subseteq tV$ for every $t > s$.

A topological vector space (X, Ω) is called ***locally convex*** if, for every open set $U \subseteq X$ and every $x \in U$, there is an open set $V \subseteq X$ such that

$$x \in V \subseteq U, \quad V \text{ is convex.}$$

5.1.2 Examples. 1) (Strong Topology) A normed vector space $(X, \|\cdot\|)$ is a topological vector space with the topology $\Omega_s := \Omega(X, \|\cdot\|)$ induced by the norm. This is sometimes called the ***strong topology*** or ***norm topology*** to distinguish it from other weaker topologies discussed below.

2) (Smooth Functions). The space $X := C^\infty(O)$ of smooth functions on an open subset $O \subseteq \mathbb{R}^n$ is a locally convex Hausdorff topological vector space. The topology is given by uniform convergence with all derivatives on compact sets and is induced by the complete metric

$$d(f,g) := \sum_{l=1}^{\infty} \frac{1}{2^l} \cdot \frac{\|f-g\|_{C^l(K_l)}}{1+\|f-g\|_{C^l(K_l)}}$$

Here $K_l \subseteq \Omega$ is an exhausting sequence of compact sets.

3) Let X be a vector space. Then (X, Ω) is a topological vector space with antidiscrete topology $\Omega := \{\varnothing, X\}$, but not with the discrete topology.

4) (Convergence in Measure). Let $(X, \mathcal{A}, \mu)$ be a measure space such that $\mu(X) < +\infty$, denote by $\mathcal{L}^0(\mu)$ the vector space of all measurable functions on X, and define

$$L^0(\mu) := \mathcal{L}^0(\mu)/\sim,$$

where the equivalence relation is given by equality almost everywhere. Define a metric on $L^0(\mu)$ by

$$d(f,g) := \int_X \frac{|\,f-g\,|}{1+|\,f-g\,|} d\mu \quad \text{for } f,g \in L^0(\mu).$$

Then $L^0(\mu)$ is a topological vector space with the topology induced by d. A sequence $f_n \in L^0(\mu)$ converges to $f \in L^0(\mu)$ in this topology if and only if it ***converges in measure***, i.e.

$$\lim_{n\to\infty} \mu\,\{x \in X \;:\; |\,f_n(x) - f(x)\,| > \varepsilon\} = 0 \quad \text{for all } \varepsilon > 0.$$

The topological vector space $L^0(\mu)$ with the topology of convergence in measure is not locally convex, in general.

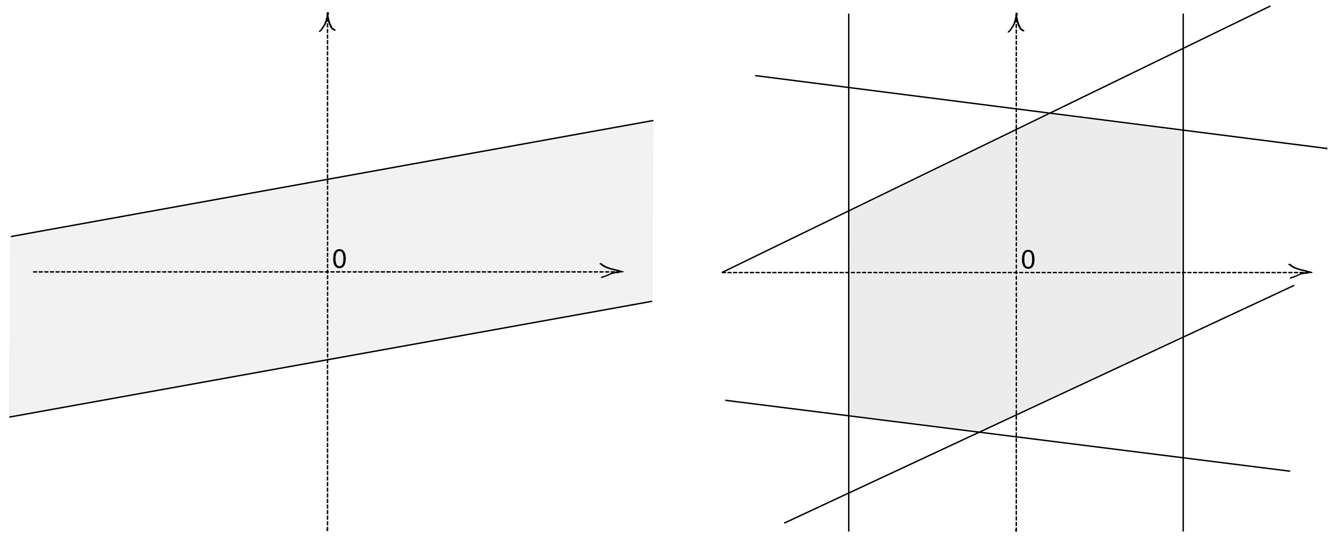


Figure 6: Examples of cylinders in $\mathbb{R}^2$ (type $\bigcap_{i=1}^{m} \{x \ : |\, f_i(x)\,| < \varepsilon\}$ for $m = 1$ and $m = 3$).

An important class of topological vector spaces is determined by sets of linear functionals as follows. Fix a vector space X and let

$$\mathcal{F} \subseteq \{f : X \to \mathbb{C} \text{ (or } \mathbb{R}) \ : \ f \text{ is linear}\}$$

is required to be nonempty. Define $\Omega_{\mathcal{F}} \subseteq 2^X$ to be the weakest topology on X such that every linear functional $f \in \mathcal{F}$ is continuous. Then the pre-image of an open circle under any of the linear functionals $f \in \mathcal{F}$ is an open subset of X. Hence so are the ***cylinders***, which are the sets of the form

$$V := V_{f_1,\dots,f_m}(x_0, \varepsilon) := \{x \in X \ : \ |\, f_i(x) - f_i(x_0)\,| < \varepsilon \text{ for } i = 1, \dots, m\} \tag{5.1}$$

for all integers $m \in \mathbb{N}$, all $f_1, \dots, f_m \in \mathcal{F}$, all $x_0 \in X$ and all $\varepsilon > 0$. Denote the collection of all cylinders in X by

$$\mathcal{V}_{\mathcal{F}} := \left\{ V_{f_1,\dots,f_m}(x_0, \varepsilon) \ : \ \begin{array}{c} m \in \mathbb{N}, \quad f_1, \dots, f_m \in \mathcal{F}, \\ x_0 \in X, \quad \varepsilon > 0. \end{array} \right\} \tag{5.2}$$

5.1.3 Proposition. *Let X be a vector space, let $\mathcal{F} \subseteq \mathbb{C}^X$ be a set of linear functionals on X, and let $\Omega_{\mathcal{F}} \subseteq 2^X$ be the weakest topology on X such that all elements of $\mathcal{F}$ are continuous. Then the following hold.*

(i) *The collection $\mathcal{V}_{\mathcal{F}}$ in (5.2) of all cylinders is a basis for the topology $\Omega_{\mathcal{F}}$, i.e.*

$$\Omega_{\mathcal{F}} = \{U \subseteq X \;:\; \forall\, x \in U \quad \exists\, V \in \mathcal{V}_{\mathcal{F}} \text{ such that } x \in V \subseteq U\} \tag{5.3}$$

(ii) *$(X, \Omega_{\mathcal{F}})$ is a locally convex topological vector space.*

(iii) *A sequence $x_n \in X$ converges to an element $x_0 \in X$ with respect to the topology $\Omega_{\mathcal{F}}$ if and only if $f(x_0) = \lim_{n\to\infty} f(x_n)$ for all $f \in \mathcal{F}$.*

(iv) *The topological space $(X, \Omega_{\mathcal{F}})$ is Hausdorff if and only if $\mathcal{F}$ separates points (i.e. for every nonzero vector $x \in X$ there exists a linear functional $f \in \mathcal{F}$ such that $f(x) \neq 0$).*

Proof. (I) It follows directly from the definitions that $\Omega_{\mathcal{F}} \subseteq 2^X$ defined by (5.3) is a topology, that every linear function $f \in \mathcal{F}$ is continuous with respect to this topology, and that every other topology $\Omega \subseteq 2^X$ with respect to which each element of $\mathcal{F}$ is continuous must contain $\mathcal{V}_{\mathcal{F}}$ and hence also $\Omega_{\mathcal{F}}$.

(II) Why is addition continuous with respect to $\Omega_{\mathcal{F}}$? It is continuous because for any $\varepsilon > 0$, any $x_1, x_2 \in X$ and every $f_1, \dots, f_m \in \mathcal{F}$ it holds

$$V_{f_1,\dots,f_m}\left(x_1, \frac{\varepsilon}{2}\right) + V_{f_1,\dots,f_m}\left(x_2, \frac{\varepsilon}{2}\right) \subseteq V_{f_1,\dots,f_m}(x_1 + x_2, \varepsilon)$$

Why is scalar multiplication continuous with respect to $\Omega_{\mathcal{F}}$? It is enough to check that for any $x_0 \in X$, any scalar α and an arbitrary cylindrical neighborhood $V_{f_1,\dots,f_m}(\alpha x_0, \varepsilon)$ of the point αx_0 there are some positive δ and r, such that

$$\beta \cdot V_{f_1,\dots,f_m}(x_0, \delta) \subseteq V_{f_1,\dots,f_m}(\alpha x_0, \varepsilon) \quad \text{if } |\,\beta - \alpha\,| < r$$

(see 5.1.1.) Let

$$M := \max\{|\, f_1(x_0)\,|, \dots, |\, f_m(x_0)\,|, 1\} > 0,$$
$$r := \frac{\varepsilon}{3M}, \quad \delta := \frac{\varepsilon}{3(r + |\,\alpha\,|)}.$$

Then for $i = 1, \ldots, m$ and $x \in V_{f_1,\ldots,f_m}(x_0, \delta)$ we have

$$| f_i(\beta x) - f_i(\alpha x_0) | \leq | f_i(\beta x) - f_i(\beta x_0) | + | f_i(\beta x_0) - f_i(\alpha x_0) | \leq$$
$$| \beta | \cdot \delta + | \beta - \alpha | \cdot M \leq (r + | \alpha |) \cdot \delta + r \cdot M = \frac{\varepsilon}{3} + \frac{\varepsilon}{3} < \varepsilon,$$

so $\beta x \in V_{f_1,\ldots,f_m}(\alpha x_0, \varepsilon)$.

Hence $(X, \Omega_{\mathcal{F}})$ is a topological vector space. That $(X, \Omega_{\mathcal{F}})$ is locally convex follows from the fact that the cylinders are all convex sets.

(iii) Fix a sequence $(x_n)_{n\in\mathbb{N}}$ in X and an element $x_0 \in X$. Assume x_n converges to x_0 with respect to the topology $\Omega_{\mathcal{F}}$. Let $f \in \mathcal{F}$ and fix a constant $\varepsilon > 0$. Then the set

$$U := \{x \in X \ : \ | f(x) - f(x_0) | < \varepsilon\}$$

is a cylinder and hence an element of $\Omega_{\mathcal{F}}$. Since $x_0 \in U$, there exists a positive integer n_0 such that $x_n \in U$ for every integer $n \geq n_0$. Thus we have proved

$$\forall\, f \in \mathcal{F} \quad \forall\, \varepsilon > 0 \quad \exists\, n_0 \in \mathbb{N} \quad \forall\, n \in \mathbb{N} \ :$$
$$n \geq n_0 \Rightarrow \ | f(x_n) - f(x_0) | < \varepsilon.$$

This means that $\lim_{n\to\infty} f(x_n) = f(x_0)$ for all $f \in \mathcal{F}$.

Conversely suppose that $\lim_{n\to\infty} f(x_n) = f(x_0)$ for all $f \in \mathcal{F}$. Fix a set $U \in \Omega_{\mathcal{F}}$ such that $x_0 \in U$. Then there exists a cylinder

$$V_{f_1,\ldots,f_m}(x_0, \varepsilon) = \{x \in X \ : \ | f_i(x) - f_i(x_0) | < \varepsilon\} \ \subseteq U$$

Since $\lim\limits_{n\to\infty} f_i(x_n) = f_i(x_0)$ for every $i \in \{1, \ldots, m\}$, there is a positive integer n_0 such that $| f_i(x_n) - f_i(x_0) | < \varepsilon$ for every integer $n \geq n_0$ and every $i \in \{1, \ldots, m\}$. Thus $x_n \in V \subseteq U$ for every integer $n \geq n_0$ and this proves part (iii).

(iv) Assume first that $(X, \Omega_{\mathcal{F}})$ is Hausdorff and let $x \in X \setminus \{\mathbf{0}\}$. Then there exists an open set $U \in \Omega_{\mathcal{F}}$ such that $\mathbf{0} \in U$, $x \notin U$. Choose a cylinder

$$V = V_{f_1,\ldots,f_m}(\mathbf{0}, \varepsilon) \quad : \quad V \subseteq U.$$

Since $x \notin V$, there exists index $i \in \{1, \dots, m\}$ such that $|\, f_i(x)\,| > \varepsilon$ and so $f_i(x) \neq 0$.

Conversely suppose that, for every $x \in X$, there exists an element $f \in \mathcal{F}$ such that $f(x) \neq 0$. Let $x_0, x_1 \in X$ such that $x_0 \neq x_1$ and choose $f \in \mathcal{F}$ such that $f(x_1 - x_0) \neq 0$. Choose $\varepsilon > 0$ such that $2\varepsilon < |\, f(x_1 - x_0)\,|$ and consider the cylinders

$$U_i := \{x \in X \;:\; |\, f(x - x_i)\,| < \varepsilon\}$$

for $i = 0, 1$. Then $U_0, U_1 \in \mathcal{V}_\mathcal{F} \subseteq \Omega_\mathcal{F}$, $x_0 \in U_0$, $x_1 \in U_1$ and $U_0 \cap U_1 = \varnothing$. □

5.1.4 Example (Product Topology)**.** Let I be any set and consider the space $X := \mathbb{R}^I$ of all functions $x : I \to \mathbb{R}$. This is a real vector space. For $i \in I$ denote the evaluation map at i by $\pi_i : \mathbb{R}^I \to \mathbb{R}$, i.e. $\pi_i(x) := x(i)$ for $x \in \mathbb{R}^I$. Then $\pi_i : X \to \mathbb{R}$ is a linear functional for every $i \in I$. Let

$$\pi := \{\pi_i \;:\; i \in I\}$$

be the collection of all these evaluation maps and denote by Ω_π the weakest topology on X such that the projection π_i is continuous for every $i \in I$. By Proposition 5.1.3 this topology is given by (5.3). It is called the ***product topology*** on $\mathbb{R}^I$. Thus $\mathbb{R}^I$ is a locally convex Hausdorff topological vector space with the product topology.

5.1.5 Example (Weak Topology)**.** Let X be a normed vector space. *The* ***weak topology*** *on X is the weakest topology $\Omega^w \subseteq 2^X$ (or $\sigma(X, X^*)$) with respect to which every bounded linear functional $f : X \to \mathbb{C}$ is continuous.* It is the special case of the topology $\Omega_\mathcal{F} \subseteq 2^X$ in Proposition 5.1.3, where $\mathcal{F} := X^*$ is the dual space. By Proposition 4.3.10 the dual space separates points, i.e. for every $x \in X \setminus \{\mathbf{0}\}$ there is an supporting functional $f \in X^*$ such that $f(x) \neq 0$. Hence Proposition 5.1.3 asserts that (X, Ω^w) is a locally convex Hausdorff topological vector space.

Let $(x_n)_{n\in\mathbb{N}}$ be a sequence in X and let $x \in X$. Then Proposition 5.1.3 asserts that x_n ***converges weakly*** to x (i.e. in the weak topology) if and only if

$$f(x) = \lim_{n\to\infty} f(x_n) \quad \text{for all} \quad f \in X^*. \tag{5.4}$$

In this case they write

$$x_n \xrightarrow{w} x \quad \text{or} \quad x = w\text{-}\lim_{n\to\infty} x_n.$$

Every bounded linear functional is continuous with respect to the strong topology $\Omega^s := \Omega(X, \|\cdot\|)$. Hence

$$\Omega^w \subseteq \Omega^s. \tag{5.5}$$

5.1.6 Example (Weak* Topology)**.** Let X be a normed vector space and let X^* be its dual space. *The* ***weak* topology*** *on* X^* *is the weakest topology* $\Omega^{w^*} \subseteq 2^{X^*}$ (or $\sigma(X^*, X)$) *with respect to which linear functional* $I(x) : X^* \to \mathbb{C}$ (see 4.19) *is continuous for all* $x \in X$. It is the special case of the topology $\Omega_{\mathcal{F}} \subseteq 2^{X^*}$ in Proposition 5.1.3, where $\mathcal{F} := I(X) \subseteq X^{**}$. This collection of linear functionals separates points, i.e. for every $f \in X^* \setminus \{\mathbf{0}\}$ there is an element $x \in X$ such that $I(x)(f) = f(x) \neq 0$. Hence Proposition 5.1.3 asserts that (X^*, Ω^{w^*}) is a locally convex Hausdorff topological vector space.

Let $(f_n)_{n\in\mathbb{N}}$ be a sequence in X^* and let $f \in X^*$. Then Proposition 5.1.3 asserts that f_n ***weak* converges*** to x (i.e. in the weak* topology) if and only if

$$f_k(x) \to f(x) \quad \text{for every} \quad x \in X.$$

In this case they write

$$f_k \xrightarrow{w^*} f \quad \text{or} \quad f = w^*\text{-}\lim_{k\to\infty} f_k.$$

If $\Omega^s \subseteq 2^{X^*}$ is the strong topology induced by the norm, and $\Omega^w \subseteq 2^{X^*}$ is the weak topology in Example 5.1.5. Then

$$\Omega^{w^*} \subseteq \Omega^w \subseteq \Omega^s \tag{5.6}$$

These weak and weak* topologies agree when X is a reflexive Banach space.

5.2 Weak convergence and weak topology

Weak convergence has already been defined by (5.4).

Strong convergence (convergence by norm) clearly implies weak convergence. The converse is generally not true:

5.2.1 Example. An orthonormal system (e_k) in a Hilbert space H converges weakly to zero, and it clearly does not converge strongly. Indeed, by Riesz representation theorem the weak convergence to zero is equivalent to

$$\langle e_k, x\rangle \to 0 \quad \text{for every} \quad x \in H.$$

This indeed follows from Bessels inequality

$$\sum_{k=1}^{\infty} |\langle e_k, x\rangle|^2 \leq \|x\|^2$$

5.2.2 Proposition (Weak and strong boundedness). *Let A be a subset of a Banach space X. Assume that A is weakly bounded, i.e.*

$$\sup_{x\in A} |f(x)| < \infty \quad \textit{for every } f \in X^*.$$

Then A is (strongly) bounded, i.e.

$$\sup_{x\in A} \|x\| < \infty$$

Here again the reverse statement is trivially true — (strong) boundedness trivially implies weak boundeness.

Proof. We embed X into X^{**} using the canonical embedding (see (4.19)). So we consider vectors $x \in A$ as bounded linear functionals on X^* acting as $x(f) := f(x)$, $f \in X^*$. Rewriting the weak boundedness assumption as $\sup_{x \in A} |\, x(f)\,| < \infty$ for every $f \in X^*$, we may understand this assumption as point-wise boundedness of the family $A \subseteq X^{**}$. The principle of uniform boundedness (3.3.1) implies that $\sup_{x \in A} \|x\|_{X^{**}} = \sup_{x \in A} \|x\|_X < \infty$, as required. □

Weak convergence clearly implies weak boundedness, so the following statement is true.

5.2.3 Corollary. *Weakly convergent sequences in Banach spaces are bounded.*

Moreover, we have a good control of the weak limit, given in the next two results.

5.2.4 Proposition. *If $x_n \xrightarrow{w} x$ in a normed space then*

$$\|x\| \le \varliminf_{n \to \infty} \|x_n\|.$$

Proof. Let $f \in X^*$ be a supporting functional of x, i.e. $\|f\| = 1$, $f(x) = \|x\|$. Then $|\, f(x_n)\,| \le \|x_n\|$ for all n. Taking lim inf of both sides, we conclude that

$$\varliminf_{n \to \infty} \|x_n\| \ge \varliminf_{n \to \infty} |\, f(x_n)\,| = |\, f(x)\,| = \|x\|$$

as required. □

5.2.5 Lemma (Mazur). *Let $x_n \xrightarrow{w} x$ in a normed space then*

$$x \in \overline{\operatorname{conv}\{x_k\}_k}.$$

Proof. Suppose $x \notin K := \overline{\text{conv}\{x_k\}_k}$. Using a separation theorem (Corollary 4.3.23), we can separate the closed convex set K from the one-point set $\{x\}$. Namely, there exists a functional $f \in X^*$ such that

$$\sup_{y \in K} f(y) < f(x).$$

Since $x_k \in K$ for all $k \in \mathbb{N}$ this implies that

$$\sup_k f(x_k) < f(x),$$

which contradicts weak convergence. □

Criteria of weak convergence

Some known criteria of weak convergence in classical normed spaces rely on the following tool.

5.2.6 Lemma (Testing weak convergence on a dense set)**.**
Let X be a normed space and $A \subseteq X^*$ be a (strong) dense set. Then $x_n \xrightarrow{w} x$ if and only if (x_k) is bounded and

$$f(x_k) \to f(x) \quad \textit{for every} \quad f \in A.$$

Proof. *Nesessity* follows by Corollary 5.2.3.

Sufficiency will be proved by a standard approximation argument. Consider arbitrary $g \in X^*$ and $\varepsilon > 0$. By density, we can choose $f \in A$ such that $\|g - f\| \leq \varepsilon$. Then

$$\overline{\lim_{k\to\infty}} \,|\, g(x_k - x) \,|\, \leq \overline{\lim_{k\to\infty}} \,|\, f(x_k - x) \,|\, + \overline{\lim_{k\to\infty}} \,|\, (g - f)(x_k - x) \,|\, \leq$$
$$0 + \|g - f\| \overline{\lim_{k\to\infty}} (\|x_k\| + \|x\|) \leq M\varepsilon.$$

where $M := \|x_k\| + \|x\| < \infty$ by the boundedness assumption. Since $\varepsilon > 0$ is arbitrary, we conclude that $\overline{\lim_{k\to\infty}} \,|\, g(x_k - x) \,|\, = 0$. Hence $g(x_k) \to g(x)$ as required. □

5.2.7 Theorem (Weak convergence in c_0 and l^p). *Let $X = c_0$ or $X = l^p$, $p \in (p, \infty)$. Then $x_k \xrightarrow{w} x$ in X if and only if the sequence (x_k) is bounded and converges to x pointwise, i.e.*

$$x_k(i) \to x(i) \quad \textit{for every } i \in \mathbb{N}.$$

Proof. *Necessity.* If $x_n \xrightarrow{w} x$ then by applying coordinate functionals $e_i^* \in X^*$ (i.e. those acting as $e_i^*(x) = x(i)$) we see that $x_k(i) \to x(i)$ as required.

Sufficiency. We are given that (x_k) is bounded and that $f(x_k) \to f(x)$ for every coordinate functional $f = e_i^*$. By linearity, we get $f(x_k) \to f(x)$ for every $f \in \operatorname{span}\{e_i\}_{i=1}^{\infty}$.

On the other hand, the representation theorems state that $X^* = l^1$ if $X = c_0$ and $X^* = l^q$ if $X = l^p$. The functionals $e_i^* \in X^*$ get identified with the coordinate vectors $(0, \dots, 0, 1, 0, \dots)$ that $\operatorname{span}\{e_i\}_{i=1}^{\infty}$ is dense in X^*. (Why?)

The proof is finished by applying Lemma 5.2.6 to $A := \operatorname{span}\{e_i\}_{i=1}^{\infty}$.
□

5.2.8 Exercise. Consider the sequence $x_k = (1, \dots, 1, 0, 0, \dots)$ (with k ones) in l^∞. Use Mazur's lemma to show that x_k does not converge weakly. Deduce that the criterion of weak convergence in l^p, $p \in (1, \infty)$, given in Theorem 5.2.7 fails for l^∞. (There is no useful criterion of weak convergence in l^∞.

Asimilar criterion of weak convergence holds in spaces of continuous functions.

5.2.9 Theorem (Weak convergence in $C(K)$). *Let K be a compact topological space. Then $x_k \xrightarrow{w} x$ in $C(K)$ if and only if the sequence of functions $x_k(t)$ is uniformly bounded and converges to $x(t)$ pointwise (i.e.*

$$x_k(t) \to x(t) \textit{ for every } t \in K).$$

Proof. *Necessity.* Boundedness of (x_k) follows from weak convergence as before. Pointwise convergence follows by applying point evaluation functionals $\delta_t \in (C(K))^*$ (acting as $\delta_t(x) = x(t)$).

Sufficiency. We need to show that $f(x_k) \to f(x)$ for all $f \in (C(K))^*$. By representation Theorem 4.2.26, this is equivalent to claiming that

$$\int_K x_k d\varphi \to \int_K x d\varphi \tag{5.7}$$

for every Borel regular charge φ. On the other hand, our assumptions are that the sequence of functions $x_n(t)$ is uniformly bounded and it converges to $x(t)$ pointwise. The Lebesgue dominated convergence theorem implies (5.8). $\square$

A similar criterion of weak convergence holds in L^p spaces. However, it does not make sense to consider the values of functions $x \in L^p$ in individual points. Instead, we shall consider integrals of $x(t)$ over short intervals.

5.2.10 Theorem (Weak convergence in L^p)**.** *Let $p \in (1, \infty)$. A sequence $x_k \xrightarrow{w} x$ in $L^p[0,1]$ if and only if the sequence (x_k) is bounded in L^p and*

$$\int_a^b x_k dt \to \int_a^b x dt \quad \textit{for every interval } [a,b] \subseteq [0,1]. \tag{5.8}$$

Proof. One notices that the set of characteristic functions $\chi_{[a,b]}$ for $[a,b] \subseteq [0,1]$ spans the set of step functions, which is dense in $L^{p*} = L^q$. (Why?) The argument is finished similarly to Theorem 5.2.7. $\square$

Remark. *The same criterion holds for $L^p(\mathbb{R})$.* (Why?)

5.2.11 Example (Sliding bumps)**.** A good example of weakly convergent but strongly divergent sequence of functions is formed by a sliding bump in $L^p(\mathbb{R})$, $p \in (1, \infty)$. Consider a function $x \in L^p(\mathbb{R})$ with compact support. Then the sequence $x_k(t) = x(t-k)$ converges weakly to zero by Theorem 5.2.10.

Some properties of a weak topology

5.2.12 Remark. In infinite-dimensional spaces X, the cylinders are rather large as they contain subspaces of finite codimension $\{x \in X : f_k(x - x_0) = 0,\ k = 1, \dots, m\}$. This shows that in infinite dimensions, *weakly open sets are unbounded.*

5.2.13 Exercise. Prove that *in an infinite dimensional normed space X, weak topology is strictly weaker than the strong topology.*

Nevertheless, some weak and strong properties are equivalent. For example, weak boundedness and strong boundedness are equivalent. This follows from the principle of uniform boundedness, see Proposition 5.2.2. Also, weak closedness and strong closedness are equivalent for convex sets:

5.2.14 Proposition. *Let K be a convex set in a normed space X. Then K is weakly closed if and only if K is (strongly) closed.*

Proof. *Necessity* is trivial.

Sufficiency. Assume K is closed and convex. By Corollary 4.3.25 to Hahn-Banach theorem, K is the intersection of the closed half-spaces that contain K. Each closed half-space has the form

$$A_{f,a} = \{x \in X \ : \ f(x) \le a\}$$

for some $f \in X^*$ and $a \in \mathbb{R}$. Hence $A_{f,a}$ is weakly closed. The intersection K of the closed half-spaces is therefore automatically weakly closed. $\square$

Remark. Convexity assumption is critical in Proposition 5.2.14. Otherwise the result would claim that the weak and strong topologies are equivalent, which is false.

5.2.15 Lemma (Weak Closure of the Unit Sphere)**.** *Let X be an infinite-dimensional normed vector space and define*

$$S := \{x \in X \ : \ \|x\| = 1\}, \quad B := \{x \in X \ : \ \|x\| \le 1\}. \tag{5.9}$$

Then B is the weak closure of S.

Proof. The set B is weakly closed by Proposition 5.2.14 and hence contains the weak closure of S. We prove that B is contained in the weak closure of S. To see this, let $x_0 \in B$ and let $U \subseteq X$ be a weakly open set containing x_0. Then there exist a cylinder $V_{f_1,\dots,f_n}(x_0,\varepsilon)$ such that $x \in V \subseteq U$. Since X is infinite-dimensional, there is a nonzero vector $x_1 \in \operatorname{Ker} f_1 \cap \dots \cap \operatorname{Ker} f_n$ such that

$$f_i(x_1) = 0 \quad \text{for } i = 1, \dots, n.$$

Since $\|x_0\| \leq 1$ there exists a real number t such that $\|x_0 + tx_1\| = 1$. Hence $x_0 + tx_1 \in V \cap S$ and so $U \cap S \neq \varnothing$. Thus x_0 belongs to the weak closure of S and this completes the proof of the Lemma. □

In view of the last Lemma one might ask whether every element of B is the limit of a weakly convergent sequence in S. The answer is negative in general. For example, the next exercise shows that a sequence in l^1 converges weakly if and only if it converges strongly. Thus the limit of every weakly convergent sequence of norm one in l^1 has again norm one. The upshot is that the weak closure of a subset of a Banach space is in general much bigger than the set of all limits of weakly convergent sequences in that subset.

5.2.16 Exercise (Schur's Theorem)**.** Let $x_n = (x_{n,i})_{i\in\mathbb{N}}$ for $n \in \mathbb{N}$ be a sequence in l^1 that converges weakly to an element $x = (x_i)_{i\in\mathbb{N}} \in l^1$. Prove that $\lim_{n\to\infty} \|x_n - x\|_1 = 0$.

Thus, *the weak and strong convergence l^1 space are equivalent.* This result is called *Schur property* of l^1.

Why does not this contradict the discrepancy between the strong and weak topologies in an infinite-dimensional space mentioned in 5.2.13?

5.2.17 Exercise (Weak convergence in finite dimensional spaces)**.** Show that *all finite-dimensional normed spaces X have Schur property, so the weak and strong convergence in X coincide.*

5.2.18 Exercise. Let X be a Banach space and suppose X^* is separable. Let $E \subseteq X$ be a bounded set and let $x \in X$ be an element in the weak closure of E. Prove that there is a sequence (x_n) in E that converges weakly to x.

5.3 Weak* topology. Banach–Alauglu's theorem

The definitions of weak topology* and weak* convergence were given in 5.1.6.

While weak convergence of functionals $f_k \in X^*$ is tested on all functionals from X^{**}, weak* convergence of f_k is tested on the subset $X \subseteq X^{**}$. Therefore, weak convergence implies weak* convergence in X^*. Of course, for reflexive spaces, weak and weak* convergence coincide.

5.3.1 Example (Weak convergence of measures)**.** In probability theory, one says that a sequence of regular Borel measures μ_n on $\mathbb{R}$ converges weakly to a Borel regular measure μ if

$$\int_{\mathbb{R}} f d\mu_n \to \int_{\mathbb{R}} f d\mu \quad \text{for every } f \in C(\mathbb{R}).$$

Assume that the measures μ_n and μ are compactly supported, say on an interval $[a, b]$. By the representation theorem for $(C[a, b])^*$, Theorem 4.2.26, this convergence is nothing different from

$$\mu_n \xrightarrow{w^*} \mu$$

in $(C[a, b])^*$. Summarizing, the weak convergence of measures in probability theory is actually the weak* convergence of measures acting as linear functionals on $C[a, b]$.

5.3.2 Example (Dirac delta function)**.** Recall that we understand Dirac delta function δ_0 as the point evaluation functional at zero. Equivalently, Dirac delta function may be identified with the probability measure on $\mathbb{R}$ with the only atom at the origin. Therefore Dirac delta function is

Figure 7: Approximation of Dirac delta function δ by functions $\delta_n(t)$ $(n = 1 \ldots 5)$.

the weak limit of uniform measures on $\left[-\frac{1}{n}, \frac{1}{n}\right]$ as $n \to \infty$. This gives a natural way to approximate Dirac delta function δ (which does not exists as a function on $\mathbb{R}$) by genuine functions $\delta_n(t)$, which are the probability distribution functions of the uniform measures on $\left[-\frac{1}{n}, \frac{1}{n}\right]$, see the picture.

Recall that the base of the weak* topology is given by the cylinders, which are the sets of the form

$$V_{x_1,\ldots,x_m}(f_0, \varepsilon) = \{f \in X^* \ : \ |\, f(x_i) - f_0(x_i)\,| < \varepsilon \quad i = 1, \ldots, m\}.$$

So, these cylinders form a local base of weak* topology at f_0.

The main result on weak topology is Alaoglu's theorem. It allows one to bring back to life compactness arguments in infinite-dimensional normed spaces X, even though the unit ball of such X is always not compact.

5.3.3 Theorem (Banach-Alaoglu). *For every normed space X, the closed unit ball $B^* := \{f \in X^* \ : \ \|f\| \leq 1\}$ in the dual space X^* is weak* compact.*

The proof will be based on *Tychonoff's theorem* that states that the product of any collection of compact topological spaces is compact. Let us briefly recall this result.

Consider a collection of $(K_j)_{i \in J}$ of any number (countable or uncountable) of topological spaces K_j. The Cartesian product can be equipped with the product topology (see 5.1.4) whose base is formed by the sets of the form

$$\left\{ \prod_{j \in J} K_j \quad : \quad \begin{array}{c} K_j \text{ is open in } X_j; \\ \text{all but finitely many of } K_j \text{ equal } X_j \end{array} \right\}$$

Tychonoff's theorem states that *if each* X_j *is compact then* $\prod_{j \in J} K_j$ *is compact in the product topology.*

Proof of Banach-Alaoglu's theorem. We shall embed B into the product space of intervals

$$K := \prod_{x \in X} \Big[- \|x\| \, , \|x\| \Big] = \Big\{ f : X \to \mathbb{R} \ : \ |\, f(x) \,| \le \|x\| \quad \text{for all } x \in X \Big\}$$

equipped with the product topology. This is the weakest topology in which the point evaluation maps $f \mapsto f(x)$ from K to $\mathbb{R}$ are continuous for all $x \in X$ (see 5.1.4). We identify a functional $f \in B^*$ with the element of the product space $(f(x))_{x \in X} \in K$. With this identification, the weak* topology on B^* coincides with the product topology on K. (Why?) Therefore, this identification is a homeomorphic embedding of B^* into K.

It remains to check that B^* is a weak* closed subset of K; the proof will then be finished by Tychonoff's theorem. This is simple. Indeed, note that B^* consists of the *linear* functions in K. So we can represent

$$B^* = \bigcap_{x,y \in X, \ a,b \in \mathbb{R}} B_{x,y,a,b}, \quad \text{where}$$

$$B_{x,y,a,b} = \{ f \in K \ : \ f(ax + by) = af(x) + bf(y) \}.$$

Each set $B_{x,y,a,b}$ is the preimage of the weak* closed set $\{0\}$ under the map $f \mapsto f(ax + by) - af(x) - bf(y)$ which, as we know, is continuous in the product topology (recall that the point evaluation maps are continuous in the product topology). Therefore all sets $B_{x,y,a,b}$ are weak* closed, and so is their intersection B^*. This completes the proof.

$\square$

Universality of space $C(K)$. As an application of Banach-Alaoglu's theorem, we will show that the space of continuous functions $C(K)$ is universal in the sense that it contains every Banach space X as a subspace (a little disclaimer is that the compact topological space K may depend on X; otherwise the result is false for spaces X of too large cardinality).

5.3.4 Theorem (Universality of $C(K)$). *Every Banach space X can be isometrically embedded into $C(K)$ for some compact topological space K.*

Proof. Let $K := B_{X^*}$ equipped with weak* topology. By Banach-Alaoglu's theorem, K is indeed compact. We define the embedding $X \to C(K)$ by associating every $x \in X$ the point evaluation function

$$x(f) := f(x), \quad f \in K. \tag{5.10}$$

Recall that the point evaluation function is indeed in $C(K)$ by the definition of weak* topology. The map defined by (5.10) is linear by construction. Finally, this map is an isometric embedding; indeed

$$\|x\|_{C(K)} = \max_{f \in B_{X^*}} |\, f(x) \,| = \|x\|_X \,,$$

where the last inequality uses a consequence of Hahn-Banach theorem, Corollary 4.3.12. $\square$

5.3.5 Exercise (Universality of l^∞). Show that l^∞ is a universal space for all separable Banach spaces. In other words, show that every separable Banach space X isometrically embeds into l^∞.

Hint: Consider a dense subset $(x_k)_k$ of the unit sphere S_X, choose supporting functionals $f_k \in S_{X^*}$ of x_k, and define the embedding $X \to l^\infty$ by $x \mapsto (f_k(x))_{k=1}^\infty$.

5.4 The Krein–Milman Theorem

The Krein–Milman Theorem is a general result about compact convex subsets of a locally convex Hausdorff topological vector space. It asserts that every such convex subset is the closed convex hull of its set of extremal points. In particular, the result applies to the dual space of a Banach space, equipped with the weak* topology. Here are the relevant definitions.

5.4.1 Definition. *Let X be a real vector space and let $K \subseteq X$ be a nonempty convex subset. A subset $F \subseteq K$ is called a* ***face*** *of K if F is a nonempty convex subset of K and*

$$\begin{array}{c} x_0, x_1 \in K,\ 0 < \lambda < 1 \\ (1-\lambda)x_0 + \lambda x_1 = x \end{array} \quad \Rightarrow \quad x_0, x_1 \in F \tag{5.11}$$

An element $x \in K$ is called an ***extremal point*** *of K if*

$$\begin{array}{c} x_0, x_1 \in K,\ 0 < \lambda < 1 \\ (1-\lambda)x_0 + \lambda x_1 = x \end{array} \quad \Rightarrow \quad x_0 = x_1 = x \tag{5.12}$$

Denote the set of extremal points of K by $\mathcal{E}(K)$.

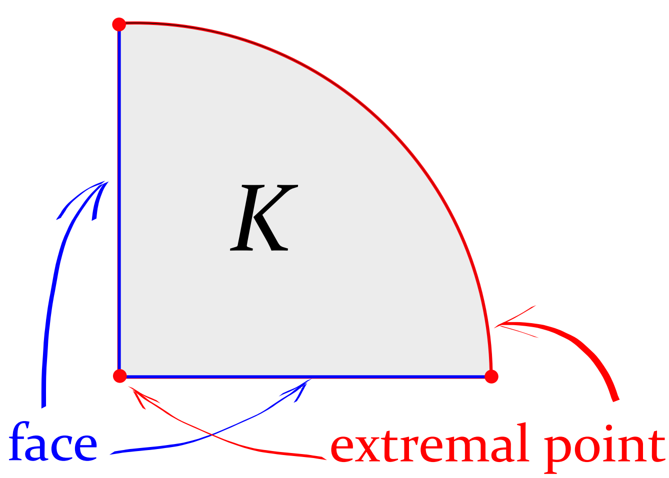


Figure 8: Extremal points and faces.

Remark. (5.12) means that the *singleton* $F := \{x\}$ *is a face of* K or, equivalently, that *there is no open line segment in* K *that contains* x (see Figure).

5.4.2 Theorem (Krein–Milman)**.** *Let* X *be a locally convex Hausdorff topological vector space and let* $K \subseteq X$ *be a nonempty compact convex set. Then* K *is the closed convex hull of its extremal points, i.e.*

$$K = \overline{\operatorname{conv}(\mathcal{E}(K))}. \tag{5.13}$$

In particular, K *admits an extremal point, i.e.* $\mathcal{E}(K) \neq \varnothing$*.*

5.4.3 Lemma. *For any element* $f \in X^*$ *and any nonempty compact convex set* K *a set*

$$F_f = \{y \in K \ : \ f(y) = \max f(K)\}$$

is a face.

Proof. Firstly, note that since K is compact and f is continuous therefore F_f is nonempty. Suppose $\lambda x + (1-\lambda)y = z \in F_f$. Then

$$\begin{gathered}\max f(K) = f(z) = \lambda f(x) + (1-\lambda) f(y) \leq \\ \lambda \max f(K) + (1-\lambda) \max f(K) = \max f(K).\end{gathered}$$

On the other hand we always have the equality which can be fullfield if and only if $f(y) = \max f(K)$ and $f(x) = \max f(K)$. Thus $x, y \in F_f$. □

Proof of the theorem for the case of normed space.

1) The *existence of an extreme point.* By Zorn's lemma we do the following procedure: if K consists with only one point then we are done. If there are two distinct points say $x \neq y$ both belonging to K, then by Hahn-Banach theorem there exists $f \in X^*$ which strictly separates these two points, in other words $f(x) > f(y)$ (see, for example 4.3.23). Now we construct the face F_f surely it does not contain the point y. Then we look at F_f and make the same procedure. Thus we obtain

the sequence of faces $(F_f)_f$. It is linearly ordered (ordered by inclusion) set. They are compact (As a closed (indeed) subset of compact set) so they have an upper bound, for example intersecion of compacts is not empty. We choose the minimal element. Note that a minimal element is a face (why?). If it contains more than 1 point then we make the same procedure which will bring us to the contradiction with minimality. Thus we obtain the extreme point.

2) *Equality 5.13 check.* Firstly note that $K_1 := \overline{\operatorname{conv}(\mathcal{E}(K))} \subseteq K$. We need to prove the convers inclusion. From contrary, let $x \in K \setminus K_1$. Then we use the Corollary 4.3.23 of Hahn-Banach theorem to the point x (as a compact set) and closed convex K_1. There exists $f \in X^*$ such that we have $\sup f(K_1) < f(y)$. Then we construct the face F_f. Surely it does not intersect the set K_1 and by the first part of the theorem it has an extreme point. This extreme point of face F_f is an extreme point of K too (why?). So we obtain the contradiction.

□

5.4.4 Example (Hilbert Cube)**.** The Hilbert cube (see 2.1.20) is a compact convex subset of $\mathbb{R}^{\mathbb{N}}$ with respect to the product topology. Its set of extremal points is the compact set

$$\mathcal{E}(Q) = \{x = (x_i)_{i\in\mathbb{N}} \in \mathbb{R}^{\mathbb{N}} \ : \ x_i \in \{0, 1/i\}\}.$$

The convex hull of any finite subset of $\mathcal{E}(Q)$ is nowhere dense in Q. Hence

$$\operatorname{conv}(\mathcal{E}(Q)) \subset Q \quad (\text{i.e. } \operatorname{conv}(\mathcal{E}(Q)) \neq Q)$$

by the Baire Category Theorem (3.1.6).

Exercise. The product topology on the Hilbert cube agrees with the topology induced by the l^2 norm.

5.5 Ergodic Theory

This section establishes the existence of an ergodic measure for any homeomorphism of a compact metric space. The proof is a fairly straightforward consequence of the Banach–Alaoglu Theorem 5.3.3 and the Krein–Milman Theorem 5.4.2. We also show that the ergodic measures are precisely the extremal points of the convex set of all invariant measures. The proof that every ergodic measure is extremal requires von Neumann's Mean Ergodic Theorem, the proof of which will in turn be based on an abstract ergodic theorem for operators on Banach spaces.

We will need an addition to the Banach-Alaoglu theorem: the fact of sequential compactness of a weak* closed ball weak* topology, generally speaking, is not metrizable).

5.5.1 Theorem (Banach–Alaoglu: The Separable Case)**.** *Let X be a separable real normed vector space. Then every bounded sequence in the dual space X^* has a weak* convergent subsequence.*

Proof. Let $D = \{x_1, x_2, x_3, \dots\} \subseteq X$ be a countable dense subset and let $(f_n)_{n\in\mathbb{N}}$ be a bounded sequence in X^*. Then the standard diagonal sequence argument shows that there is a subsequence $(f_{n_i})_{i\in\mathbb{N}}$ such that the *sequence of real numbers* $(f_{n_i}(x_k))_{i\in\mathbb{N}}$ converges for every $k \in \mathbb{N}$. More precisely, the sequence $(f_n(x_1))_{n\in\mathbb{N}}$ is bounded and hence has a convergent subsequence $(f_{n_{i,1}}(x_1))_{i\in\mathbb{N}}$. Since the sequence $(f_{n_{i,1}}(x_2))_{i\in\mathbb{N}}$ is bounded it has a convergent subsequence $(f_{n_{i,2}}(x_2))_{i\in\mathbb{N}}$. Continue by induction to construct a sequence of subsequences $(f_{n_{i,k}})_{i\in\mathbb{N}}$ such that, for every $k \in \mathbb{N}$, $(f_{n_{i,k+1}})_{i\in\mathbb{N}}$ is a subsequence of $(f_{n_{i,k}})_{i\in\mathbb{N}}$ and the sequence $(f_{n_{i,k}}(x_k))_{i\in\mathbb{N}}$ converges. Now consider the diagonal subsequence $f_{n_i} := f_{n_{i,i}}$. Then the sequence $(f_{n_i}(x_k))_{i\in\mathbb{N}}$ con verges for every $k \in \mathbb{N}$ as claimed.

With this understood, it follows from the Banach-Steinhaus Theorem (3.3.5), that there exists an element $f \in X^*$ such that

$$f(x) = \lim_{i\to\infty} f_{n_i}(x)$$

for all $x \in X$. Hence the sequence $(f_{n_i})_{i\in\mathbb{N}}$ converges to f in the weak* topology as claimed. $\square$

5.5.2 Example. This example shows that the hypothesis that X is separable cannot be removed in Theorem 5.5.1. The Banach space $X = l^\infty$ with the supremum norm is not separable. For $n \in \mathbb{N}$ define the bounded linear functional $f_n : l^\infty \to \mathbb{R}$ by $f_n(x) := x_n$ for $x = (x_i)_{i\in\mathbb{N}} \in l^\infty$. Then the sequence $(f_n)_{n\in\mathbb{N}} \in X$ does not have a weak* convergent subsequence. To see this, let $n_1 < n_2 < n_3 < \dots$ be any sequence of positive integers and define the sequence $x = (x_i)_{i\in\mathbb{N}} \in l^\infty$ by $x_i := 1$ for $i = n_{2k}$ with $k \in \mathbb{N}$ and by $x_i := -1$ otherwise. Then $f_{n_k}(x) = x_{n_k} = (-1)^k$ and hence the sequence of real numbers $(f_{n_k}(x))_{k\in\mathbb{N}}$ does not converge. Thus the subsequence $(f_{n_k})_{k\in\mathbb{N}}$ does not converge in the weak* topology.

Invariant Measures. Let (K, ρ) be a compact metric space and let $\Phi : K \to K$ be a homeomorphism. Denote by $\mathcal{B} \subseteq 2^M$ the Borel σ-algebra. Recall that the space $C(K)$ of all continuous functions $f : K \to \mathbb{R}$ with the supremum norm is a separable Banach space and its dual space is isomorphic to the space $\mathcal{M}(K)$ of signed Borel measures $\mu : \mathcal{B} \to \mathbb{R}$, (see 4.2.26), equipped with the norm function

$$\|\mu\| = \operatorname{var}\mu(K) = \sup_{E\in B}\left(\mu(E) - \mu(K \setminus E)\right).$$

for $\mu \in \mathcal{M}(K)$.

5.5.3 Definition. *A probability measure $\mu : \mathcal{B} \to [0,1]$ is called Φ-**invariant** if*

$$\int_K f \circ \Phi \, d\mu = \int_K f d\mu \quad \textit{for all} \;\; f \in C(K). \tag{5.14}$$

The *set $\mathcal{M}(\Phi)$ of Φ-invariant Borel probability measures* is a weak* closed convex subset of the unit sphere in $\mathcal{M}(K)$. The next lemma shows that it is nonempty.

5.5.4 Lemma. *Every homeomorphism of a compact metric space admits an invariant Borel probability measure.*

Proof. Let $\Phi : K \to K$ be a homeomorphism of a compact metric space. Fix an element $t_0 \in K$ and, for every integer $n \geq 1$, define the Borel probability measure $\mu_n : \mathcal{B} \to [0, 1]$ by

$$\int_K f \, d\mu_n := \frac{1}{n} \sum_{k=0}^{n-1} f\left(\Phi^k(t_0)\right) \quad \text{for} \quad f \in C(K).$$

Here $\Phi^0 := \mathrm{id} : K \to K$ and $\Phi^k := \Phi \circ \cdots \circ \Phi$ denotes the kth iterate of Φ for $k \in \mathbb{N}$. By Theorem 5.5.1, the sequence μ_n has a weak* convergent subsequence $(\mu_{n_i})_{i\in\mathbb{N}}$. Its weak* limit is a Borel measure $\mu : \mathcal{B} \to [0, \infty)$ such that

$$\|\mu\| = \int_K 1 \, d\mu = \lim_{i\to\infty} \int_K 1 \, d\mu_{n_i} = 1$$

and

$$\int_K f \circ \Phi \, d\mu = \lim_{i\to\infty} \frac{1}{n_i} \sum_{k=1}^{n_i} f\left(\Phi^k(t_0)\right) =$$

$$\lim_{i\to\infty} \frac{1}{n_i} \sum_{k=0}^{n_i-1} f\left(\Phi^k(t_0)\right) = \int_K f \, d\mu.$$

for all $f \in C(K)$. Hence $\mu \in \mathcal{M}(\Phi)$. $\square$

5.5.5 Definition. *Let K be a metric compact and let $\Phi : K \to K$ be a homeomorphism. A Φ-invariant Borel probability measure $\mu : \mathcal{B} \to [0, 1]$ is called Φ-**ergodic** if, for every Borel set $E \subseteq K$,*

$$\Phi(E) = E \quad \Rightarrow \quad \mu(E) \in \{0, 1\}. \tag{5.15}$$

*The homeomorphism Φ is called μ-**ergodic** if μ is an ergodic measure for Φ.*

Example. If $t \in K$ is a fixed point of Φ, then the Dirac measure $\mu = \delta_t$ is ergodic for Φ. If $\Phi = \mathrm{id}$, then the Dirac measure at each point of K is ergodic for Φ and there are no other ergodic measures.

5.5.6 Theorem (Ergodic Measures are Extremal). *Let $\mu : \mathcal{B} \to [0,1]$ be a Φ-invariant Borel probability measure. Then the following are equivalent.*

(i) *μ is an ergodic measure for Φ.*

(ii) *μ is an extremal point of $\mathcal{M}(\Phi)$.*

Proof (ii) $\Rightarrow$ (i). We prove this implication by an indirect argument. Assume that μ is not ergodic for Φ. Then there exists a Borel set $E_0 \subseteq K$ such that

$$\Phi(E_0) = E_0, \quad 0 < \mu(E_0) < 1.$$

Define $\mu_0, \mu_1 : B \to [0,1]$ by

$$\mu_0(B) := \frac{\mu(B \setminus E_0)}{1 - \mu(E_0)}, \quad \mu_1(B) := \frac{\mu(B \cap E_0)}{\mu(E_0)}$$

for $B \in \mathcal{B}$. These are Φ-invariant Borel probability measures and they are not equal because $\mu_0(E_0) = 0$ and $\mu_1(E_0) = 1$. Moreover, $\mu = (1-\lambda)\mu_0 + \lambda\mu_1$ where $\lambda := \mu(E_0)$. Hence μ is not an extremal point of $\mathcal{M}(\Phi)$. This shows that (ii) implies (i). The converse will be proved a little later.

5.5.7 Corollary (Existence of Ergodic Measures). *Every homeomorphism of a compact metric space admits an ergodic measure.*

Proof. The set $\mathcal{M}(\Phi)$ of Φ-invariant Borel probability measures on $\mathcal{M}$ is nonempty by Lemma 5.5.4 and it is a weak* compact convex subset of $\mathcal{M}(K)$. Hence $\mathcal{M}(\Phi)$ has an extremal point μ by Krein-Milman Theorem (5.4.2). Thus μ is an ergodic measure by (ii) $\Rightarrow$ (i) in Theorem 5.5.6. □

Space and Time Averages

Given a homeomorphism $\Phi : K \to K$ of a compact metric space K, a Φ-ergodic measure $\mu : \mathcal{B} \to [0,1]$ on the Borel σ-algebra $\mathcal{B} \subseteq 2^K$, a continuous function $f : K \to \mathbb{R}$, and an element $t \in K$, one can ask the question of whether the sequence of averages $\frac{1}{n}\sum_{k=0}^{n-1} f(\Phi^k(t))$ converges.

A theorem of Birkhoff answers this question in the affirmative for almost every $t \in K$. This is Birkhoff's Ergodic Theorem. It asserts that, *if μ is a Φ-ergodic measure, then for every continuous function $f : K \to \mathbb{R}$, there exists a Borel set $E_0 \subseteq K$ such that*

$$\Phi(E_0) = E_0, \quad \mu(E_0) = 1,$$

$$\int_K f\,d\mu = \lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} f(\Phi^k(t)) \quad \textit{for all} \;\; t \in E_0. \tag{5.16}$$

In other words, the *time average* of f agrees with the *space average* for almost every orbit of the dynamical system. If Φ is uniquely ergodic, i.e. Φ admits only one ergodic measure or, equivalently, only one Φ-invariant Borel probability measure, then equation (5.16) actually holds for all $t \in K$. Birkhoff's Ergodic Theorem extends to μ-integrable functions and asserts that the sequence of measurable functions $\frac{1}{n}\sum_{k=0}^{n-1} f \circ \Phi^k$ converges pointwise almost everywhere to the mean value of f. A particularly interesting case is where f is the characteristic function of a Borel set $B \subseteq K$. Then the integral of f is the measure of B and it follows from Birkhoff's Ergodic Theorem that

$$\mu(B) = \lim_{n\to\infty} \frac{1}{n} \cdot \textit{number}\big\{k \in \{0, \dots, n-1\} \;:\; \Phi^k(t) \in B\big\}$$

for μ-almost all $t \in K$. A weaker result is the following von Neumann's Mean Ergodic Theorem. It asserts that the sequence $\frac{1}{n}\sum_{k=0}^{n-1} f \circ \Phi^k$ converges to the mean value of f in $L^p(\mu)$ for $1 < p < \infty$. This implies pointwise almost everywhere convergence for a suitable subsequence.

5.5.8 Theorem (Von Neumann's Mean Ergodic Theorem)**.**
Let $\Phi : K \to K$ be a homeomorphism of a compact metric space K, let $\mu \in \mathcal{M}(\Phi)$ be a Φ-ergodic measure, let $1 < p < \infty$, and let $f \in L^p(\mu)$. Then

$$\lim_{n\to\infty} \left\| \frac{1}{n} \sum_{k=0}^{n-1} f \circ \Phi^k - \int_K f\,d\mu \right\|_{L^p} = 0. \tag{5.17}$$

Theorem 5.5.8 follows from the following «big» Ergodic Theorem 5.5.15.

Theorem 5.5.8 implies Theorem 5.5.6. The proof uses the following fact linking two types of convergence of functional sequences.

5.5.9 Lemma. *If a sequence $(x_n)_n$ converges in a space $L^p(E, \mu)$ for $1 \le p \le \infty$, then there is a subsequence $(x_{n_k})_k$ converging almost everywhere.*

Proof. The statement of the lemma is meaningful when $1 \le p < +\infty$.

Consider a subsequence $(x_{n_k})_k$ such that $\|x_{n_{k+1}} - x_{n_k}\| < 2^{-k}$ for all $k \in \mathbb{N}$ (taking $k = 1$ we find a number n_1 such that for all $n > n_1$ it holds $\|x_n - x_{n_1}\| < 2^{-1}$. Next, for $k = 2$ we find a number $n_2 > n_1$ such that for all $n > n_2$ it holds $\|x_n - x_{n_2}\| < 2^{-2}$, so $\|x_{n_1} - x_{n_2}\| < 2^{-1}$ by choice n_1. Continuing the process indefinitely, we will get the desired sequence $(x_{n_k})_k$). Then

$$\sum_{k=1}^{\infty} \|x_{n_{k+1}} - x_{n_k}\| < 1.$$

We prove that this subsequence converges almost everywhere. Consider the series

$$\sum_{k=1}^{\infty} |\, x_{n_{k+1}}(t) - x_{n_k}(t)\,| \tag{5.18}$$

Let $S_l(t)$ be its partial sums and $S(t) \in [0, +\infty]$ it's sum. Then for all $l \in \mathbb{N}$ we have

$$\|S_l\| \le \sum_{k=1}^{l} \|x_{n_{k+1}} - x_{n_k}\| < 1.$$

By Fatou's theorem

$$\int_E S^p d\mu = \int_E \lim_{l\to\infty} S_l^p d\mu \le \lim_{l\to\infty} \|S_l\|^p \le 1.$$

Therefore, $S^p \in L^p(E,\mu)$ and, therefore, S is almost everywhere finite, that is, the series (5.18) converges for almost all t. Moreover, the series

$$x_{n_1} + \sum_{k=1}^{\infty} \big(x_{n_{k+1}}(t) - x_{n_k}(t)\big)$$

converges for almost all t to some sum x_0, and $x_0 \in L^p(E,\mu)$ since $|\,x_0\,| \le |\,x_{n_1}\,| + |\,S\,|$. But the convergence of this series is convergence $x_{n_k} \to x_0$ almost everywhere. □

5.5.10 Lemma. *Let $\mu_0, \mu_1 \in \mathcal{M}(\Phi)$ be ergodic measures such that $\mu_0(E) = \mu_1(E)$ for every Φ-invariant Borel set $E \subseteq K$. Then $\mu_0 = \mu_1$.*

Proof. Fix a continuous function $f : K \to \mathbb{R}$. Then it follows from Theorem 5.5.8 and lemma 5.5.9 that there exist Borel sets $B_0, B_1 \subseteq K$ and an increasing sequence of integers $(n_j)_j$ such that $\mu_i(B_i) = 1$ and

$$\int_K f\, d\mu_i = \lim_{j\to\infty} \frac{1}{n_j} \sum_{k=0}^{n_j-1} f(\Phi^k(t)) \quad \text{for all} \quad t \in B_i \quad \text{and} \quad i = 0, 1. \tag{5.19}$$

For $i = 0, 1$ define $E_i := \bigcap_{n\in\mathbb{Z}} \Phi^n(B_i)$. So E_i is a Φ-invariant Borel set such that $\mu_i(E_i) = 1$. Thus $\mu_1(E_0) = \mu_0(E_0) = 1$ and $\mu_0(E_1) = \mu_1(E_1) = 1$ by assumption. This implies that the Φ-invariant Borel set $E := E_0 \cap E_1$ is nonempty. Since $E \subseteq B_0 \cap B_1$, it follows from (5.19) that

$$\int_K f\, d\mu_0 = \lim_{j\to\infty} \frac{1}{n_j} \sum_{k=0}^{n_j-1} f(\Phi^k(t)) = \int_K f\, d\mu_1 \quad \text{for all} \quad t \in E.$$

Thus the integrals of f with respect to μ_0 and μ_1 agree for every continuous function $f : K \to \mathbb{R}$. Hence $\mu_0 = \mu_1$ by uniqueness in the Riesz-Markov Representation Theorem 4.2.26. □

Completion of the **proof** of the theorem 5.5.6 (part (i) $\Rightarrow$ (ii)).

Let $\mu \in \mathcal{M}(\Phi)$ be ergodic. We need to show that μ is an extremal point of $\mathcal{M}(\Phi)$ (the definition of $\mathcal{M}(\Phi)$ see 5.5.3).

Let $\mu_0, \mu_1 \in \mathcal{M}(\Phi)$ and $0 < \lambda < 1$ such that $\mu = (1-\lambda)\mu_0 + \lambda\mu_1$.

If $B \subseteq K$ is a Borel set such that $\mu(B) = 0$, then $(1-\lambda)\mu_0(B) + \lambda\mu_1(B) = 0$, and hence $\mu_0(B) = \mu_1(B) = 0$ because $0 < \lambda < 1$.

If $B \subseteq K$ is a Borel set such that $\mu(B) = 1$, then $\mu(K \setminus B) = 0$, hence $\mu_0(K \setminus B) = \mu_1(K \setminus B) = 0$, and therefore $\mu_0(B) = \mu_1(B) = 1$.

Now let $E \subseteq K$ be a Φ-invariant Borel set. Then $\mu(E) \in \{0, 1\}$ because μ is Φ-ergodic, and hence $\mu_0(E) = \mu_1(E) = \mu(E)$. Thus μ_0 and μ_1 are Φ-ergodic measures that agree on all Φ-invariant Borel sets. Hence $\mu_0 = \mu_1 = \mu$ by Lemma 5.5.10.

An Abstract Ergodic Theorem

Theorem 5.5.8 translates into a theorem about the iterates of a bounded linear operator from a Banach space to itself provided that these iterates are uniformly bounded. The ergodic theorem in functional analysis asserts that, if $F : X \to X$ is a bounded linear operator on a reflexive Banach space whose iterates F^n form a bounded sequence of bounded linear operators, then its averages $S_n := \frac{1}{n}\sum_{k=0}^{n-1} F^k$ form a sequence of bounded linear operators that converge strongly to a projection onto the kernel of the operator $\mathbf{1} - F$. Here is the relevant definition.

5.5.11 Definition. *Let X be a real normed vector space. A bounded linear operator $P : X \to X$ is called a **projection** if $P^2 = P$.*

5.5.12 Lemma. *Let X be a Banach space and let $X_1, X_2 \subseteq X$ be two closed linear subspaces such that $X = X_1 \oplus X_2$, i.e. $X_1 \cap X_2 = \{\mathbf{0}\}$ and every vector $x \in X$ can be written as $x = x_1 + x_2$ with $x_1 \in X_1$ and $x_2 \in X_2$. Then there exists a constant $c \geq 0$ such that*

$$\|x_1\| + \|x_2\| \leq c\,\|x_1 + x_2\| \tag{5.20}$$

for all $x_1 \in X_1$ and all $x_2 \in X_2$.

Proof. The vector space $X_1 \times X_2$ is a Banach space with the norm function $\|(x_1, x_2)\| := \|x_1\| + \|x_2\|$, and the linear operator $A : X_1 \times X_2 \to X$, defined by $A(x_1, x_2) := x_1 + x_2$ for $(x_1, x_2) \in X_1 \times X_2$, is bijective by assumption and bounded by the triangle inequality. Hence its inverse is bounded by the Inverse Operator Theorem 3.4.8. □

5.5.13 Lemma. *Let X be a real normed vector space and let $P : X \to X$ be a bounded linear operator. Then the following are equivalent.*

(I) *P is a projection.*

(II) *There exist closed linear subspaces $X_0, X_1 \subseteq X$ such that*

$$X_0 \cap X_1 = \{\mathbf{0}\}, \quad X = X_0 \oplus X_1$$
$$\forall x_0 \in X_0, \quad \forall x_1 \in X_1 \quad P(x_0 + x_1) = x_1.$$

Proof. If P is a projection, then $P^2 = P$ and hence the linear subspaces

$$X_0 := \ker(P), \quad X_1 := \mathrm{im}(P) = \ker(\mathbf{1} - P)$$

satisfy the requirements of part (ii). If P is as in (ii), then $P^2 = P$ by definition and $P : X \to X$ is a bounded linear operator by lemma 5.5.12. □

5.5.14 Example. The direct sum of two closed linear subspaces of a Banach space need not be closed. For example, let $X := C[0, 1]$ with the supremum norm. Then the linear subspaces

$$Y := \{(f, g) \in X \times X \ : \ f = 0\},$$
$$Z := \{(f, g) \in X \times X \ : \ f \in C^1[0, 1], f = g\}$$

of $X \times X$ are closed, their intersection $Y \cap Z$ is trivial, and their direct sum $Y \oplus Z = \{(f, g) \in X \times X \ : f \in C^1[0, 1]\}$ is not closed.

5.5.15 Theorem (Ergodic Theorem)**.** *Let $F : X \to X$ be a bounded linear operator on a Banach space X, and $S_n := \frac{1}{n}\sum_{k=0}^{n-1} F^k$ for $n \in \mathbb{N}$. Assume that there is a constant c such that*

$$\forall\, n \in \mathbb{N} \quad \|F^n\| \le c. \tag{5.21}$$

Then the following hold.

(i) *For any* $x \in X$ *the sequence* $(S_n(x))_{n\in\mathbb{N}}$ *converges if and only if it has a weakly convergent subsequence.*

(ii) *The set* $Z := \{x \in X \ : \ \text{the sequence} \ (S_n(x))_{n\in\mathbb{N}} \ \text{converges}\}$ *is a closed* F*-invariant linear subspace of* X *and*

$$Z = \ker(\mathbf{1} - F) \oplus \overline{\operatorname{im}(\mathbf{1} - F)}.$$

Moreover, if X *is reflexive, then* $Z = X$.

(iii) *Let* $S : Z \to Z$, $S(x_0 + x_1) := x_0$ *for* $x_0 \in \ker(\mathbf{1} - F)$ *and* $x_1 \in \overline{\operatorname{im}(\mathbf{1} - \mathrm{F})}$. *Then*

$$\forall \ x \in Z \quad \lim_{n\to\infty} S_n(x) = S(x), \tag{5.22}$$

$$\begin{gathered} SF = FS = S^2 = S, \\ \|S\| \leq c. \end{gathered} \tag{5.23}$$

Theorem 5.5.15 implies Theorem 5.5.8

Let $\Phi : K \to K$ be a homeomorphism of a metric compact K and let $\mu \in \mathcal{M}(\Phi)$ be an ergodic Φ-invariant Borel probability measure on $\mathcal{M}$. Define the bounded linear operator $F : L^p(\mu) \to L^p(\mu)$ by

$$F(x) := x \circ \Phi \quad \text{for all } x \in L^p(\mu).$$

Then $\|Fx\|_p = \|x\|_p$ for all $f \in L^p(\mu)$, by the Φ-invariance of μ, and so $\|F\| = 1$. Thus F satisfies the requirement of Theorem 5.5.15. Let $x \in L^p(\mu)$. Since $L^p(\mu)$ is reflexive, Theorem 5.5.15 asserts that the sequence

$$S_n(x) = \frac{1}{n}\sum_{k=0}^{n-1} F^k(x) = \frac{1}{n}\sum_{k=0}^{n-1}(x \circ \Phi^k)$$

converges in $L^p(\mu)$ to a function $S(x) \in \ker(\mathbf{1} - F)$. It remains to prove that $S(x)$ is equal to the constant $C := \int_K f d\mu$ almost everywhere. The key to the proof is the fact that every function in the kernel of the operator $\mathbf{1} - F$ is constant (almost everywhere). Once this is understood, it follows that there exists a constant $C \in \mathbb{R}$ such that $S(x) = C$ almost everywhere, and hence

$$C = \int_K S(x)\, d\mu = \lim_{n\to\infty} \int_K S_n(x)\, d\mu = \lim_{n\to\infty} \sum_{k=0}^{n-1} \int_K x \circ \Phi^k\, d\mu = \int_K x\, d\mu.$$

Thus it remains to prove that every function in the kernel of $\mathbf{1} - F$ is constant. Let $y \in L^p(\mu)$ and suppose that $F(y) = y$. Then $y(t) = y(\Phi(t))$ for almost all $t \in K$. Define

$$E_0 := \{t \in K \ : \ y(t) \neq y(\Phi(t))\}, \quad E := \bigcup_{k\in\mathbb{Z}} \Phi^k(E_0).$$

Then $E \subseteq K$ is a Borel set with $\Phi(E) = E$, $\mu(E) = 0$, and $y(\Phi(t)) = y(t)$ for every $t \in K \setminus E$. Let $C := \int_K y d\mu$ and define $B_-, B_0, B_+ \subseteq K$ by

$$B_0 := \{t \in K \setminus E \ : \ y(t) = C\}, \quad B_\pm := \{t \in K \setminus E \ : \ \pm y(t) > C\}.$$

Each of these three Borel sets is invariant under Φ and hence has measure either zero or one. Moreover, $B_- \cup B_0 \cup B_+ = K \setminus E$ and this implies

$$\mu(B_-) + \mu(B_0) + \mu(B_+) = 1.$$

Hence one of the three sets has measure one and the other two have measure zero. This implies that $\mu(B_0) = 1$, because otherwise either $\int_K y d\mu < C$ or $\int_K y d\mu > C$. Thus y is equal to its mean value almost everywhere. This proves Theorem 5.5.8.

Proof of Theorem 5.5.15.

1) Estimations norms of S_n and $S_n(1-F)$. From 5.21 we get $\|S_n\| \leq c$, and

$$S_n(1-F) = \frac{1}{n}\sum_{k=0}^{n-1} F^k - \frac{1}{n}\sum_{k=1}^{n} F^k = \frac{1}{n}(1-F^n) \quad \Rightarrow \quad \|S_n(1-F)\| \leq \frac{1}{n}(\|1\| + \|F^n\|) \leq \frac{1+c}{n}. \tag{5.24}$$

2) Estimation of the operator's fixed point norm. Let $x \in X$ such that $F(x) = x$. Then $S_n(x) = x$ for all $n \in \mathbb{N}$ and for all $\xi \in X$ by 5.24 $\lim_{n\to\infty} \|S_n(\xi - F(\xi))\| = 0$ and hence

$$\|x\| = \lim_{n\to\infty} \|x + S_n(\xi - F(\xi))\| = \|S_n(x + \xi - F(\xi))\| \leq c \|x + \xi - F(\xi)\|$$

3) *If* $x \in X_0 := \ker(\mathbf{1} - F)$ *and* $y \in X_1 := \overline{\mathrm{im}(\mathbf{1} - F)}$, *then*

$$\|x\| \leq c \|x + y\| \tag{5.25}$$

Choose a sequence $\xi_n \in X$ such that $y = \lim_{n\to\infty}(\xi_n - F(\xi_n))$. Then, by (5.25), we have $\|x\| \leq c \|x + \xi_n - F(\xi_n)\|$ for all $n \in \mathbb{N}$. Take the limit $n \to \infty$ to obtain $\|x\| \leq c \|x + y\|$.

4) $X_0 \cap X_1 = \mathbf{0}$ *and the direct sum* $Z := X_0 \oplus X_1$ *is a closed linear subspace of* X.

Let $x \in X_0 \cap X_1$ and define $y := -x$. Then $\|x\| \leq c \|x + y\| = 0$ by 3) and hence $x = 0$. Why is Z closed? Let $x_n \in X_0$ and $y_n \in X_1$ be sequences whose sum $z_n := x_n + y_n$ converges to some element $z \in X$. Then $(z_n)_n$ is a Cauchy sequence and hence $(x_n)_n$ is a Cauchy sequence by (5.25). This implies that $y_n = z_n - x_n$ is a Cauchy sequence and hence $z = x + y$, where $x := \lim_{n\to\infty} x_n \in X_0$ and $y := \lim_{n\to\infty} y_n \in X_1$. This proves 4).

5) *If* $z \in Z$ *then* $F(z) \in Z$.

For $z \in Z$ we have $z = x + y$, $x \in X_0$, $y \in X_1$. Choose a sequence $\xi_i \in X$ such that $y = \lim_{i\to\infty}(\xi - F(\xi_i))$. Then $F(y) = \lim_{i\to\infty} F(\xi_i - F(\xi_i)) = \lim_{i\to\infty}(\mathbf{1} - F)F(\xi_i)) \in X_1$. Hence $F(z) = F(x) + F(y) = x + F(y) \in Z$.

6) *Let* $x \in X_0$ *and* $y \in X_1$. *Then* $\lim_{n\to\infty} S_n(x + y) = x$.

By (5.24) for every $\xi \in X$ the sequence $\|S_n(1 - F)\xi\| \leq \frac{1+c}{n} \|\xi\|$ converges to zero as n tends to infinity. Hence it follows from the estimate $\|S\|_n \leq c$ in 1) and the Banach–Steinhaus Theorem 3.3.5 that $\lim_{n\to\infty} S_n(y) = 0$ for all $y \in X_1$. Moreover, $S_n(x) = x$ forall $n \in \mathbb{N}$ by 2). Hence $\lim_{n\to\infty} S_n(x + y) = \lim_{n\to\infty} S_n(x) = x$.

5.5.16 Lemma. *Let $x, z \in X$. Then the following are equivalent.*

(a) *$F(x) = x$ and $z - x \in X_1$ (see 3) for the definition).*

(b) $\lim_{n\to\infty} S_n(z) - x = 0$.

(c) *There is a subsequence S_{n_j} such that* $\lim_{j\to\infty} f(S_{n_j}(z)) = f(x)$ *for all $f \in X^*$.*

Proof. That (a) implies (b) follows immediately from Step 6) and that (b) implies (c) is obvious. We prove that (c) implies (a). Thus assume (c) and fix a bounded linear functional $f \in X^*$. Then $F^*f := f \circ F : X \to \mathbb{R}$ is a bounded linear functional and $f(x - F(x)) = (f - F^*f)x = \lim_{j\to\infty}(f - F^*f)(S_{n_j}(z)) = \lim_{j\to\infty} f(\mathbf{1} - F)(S_{n_j}(z)) = 0$. Here the last equation follows from Step 1). Taking into account the Proposition 4.3.10 we have $F(x) = x$.

Why $z - x \in X_1$? Assume, by contradiction, that $z - x \notin X_1$. Then, using the Hahn-Banach theorem, we obtain, there exists an element $f \in X^*$ such that

$$f(z - x) = 1, \quad f(\xi - F(\xi)) = 0 \text{ for all } \xi \in X. \tag{5.26}$$

This implies $\forall\ k \in \mathbb{N},\ \forall\ \xi \in X \quad f(F^k(\xi) - F^{k+1}(\xi) = 0$. Hence, by induction, $\forall\ k \in \mathbb{N},\ \forall\ \xi \in X \quad f(\xi) = f(F^{k+1}(\xi))$. Thus $\forall\ k \in \mathbb{N},\quad f(S_k(z)) = f(z)$. Hence it follows from (c) that

$$f(z - x) = \lim_{j\to\infty} f(S_{n_j}(z) - x) = 0.$$

This contradicts (5.26). Thus $z - x \in X_1$. □

Completion of the proof of Theorem 5.5.15. The subspace Z is closed by Step 4) and is F-invariant by Step 5. Moreover, Lemma 5.5.16 asserts that an element $z \in X$ belongs to Z if and only if the sequence $S_n(z)$ converges in the norm topology if and only if $S_n(z)$ has a weakly convergent subsequence. If X is reflexive this holds for all $z \in X$ by Step 1). This proves (i) and (ii).

For the operator $S : Z \to Z$ defined in (III) it holds $\|S\| \le c$ by Step 3), the equation $\lim_{n\to\infty} S_n(z) = S(z)$ for $z \in Z$ follows from Step

6), and $S^2 = S$ by definition. The equation $SF = FS = S$ follows from the fact that S commutes with $T|_Z$ and vanishes on the image of the operator $\mathbf{1} - F$. □

5.5.17 Exercise (Strict Convexity and Extremal Points)**.** A normed vector space is ***strictly convex*** (that is for all $x, y \in X$, $\|x + y\| = 2\|x\| = 2\|y\| \Rightarrow x = y$) if and only if the unit sphere is equal to the set of extremal points of the closed unit ball.

5.5.18 Exercise (Extremal Points of Unit Balls)**.** Determine the extremal points of the closed unit balls in the Banach spaces c_0, c, $C[0,1]$, l^1, l^p, l^∞, $L^1[0,1]$, $L^p[0,1]$, $L^\infty[0,1]$.

5.5.19 Exercise (Birkhoff–von Neumann Theorem)**.** An $n \times n$-matrix $M = (m_{ij})$ with nonnegative coefficients $m_{ij} \geq 0$ is called ***doubly stochastic*** if its row sums and column sums are all equal to one. The Birkhoff–von Neumann Theorem asserts that *every doubly stochastic matrix is a convex combination of permutation matrices.* Thus the doubly stochastic matrices form a convex set whose extremal points are the permutation matrices.

This can be proved as follows. Let M be a doubly stochastic matrix and denote by $\nu(M)$ the number of positive entries. If $\nu(M) > n$ find a permutation matrix P and a constant $0 < \lambda < 1$ such that the matrix $N := M - \lambda P_1$ has nonnegative entries and strictly fewer positive entries than M. In the case $N \neq 0$ the matrix $M_1 := (1-\lambda)^{-1}N$ is again doubly stochastic with $\nu(M_1) < \nu(M)$, and $M = \lambda P_1 + (1-\lambda)M_1$. Continue by induction until $\nu(M_k) = n$ and so M_k is a permutation matrix. Here is a method to find P_1 and λ.

6 Compact operators. Elements of spectral theory

6.1 Compact operators

Compact operators form an important class of bounded linear operators. On the one hand, they are almost finite rank operators (in the same way as compact sets are almost finite dimensional). So compact operators do share some properties of finite rank operators, which facilitates their study. On the other hand, the class of compact operators is wide enough to include integral and Hilbert-Schmidt operators, which are important in many applications.

Throughout this section, X, Y will denote normed spaces.

6.1.1 Definition. *A* ***linear operator*** $T : X \to Y$ *is called* ***compact*** *if it maps bounded sets in* X *to precompact sets in* Y*. The set of compact operators is denoted* $\mathcal{K}(X{\to}Y)$*.*

6.1.2 Exercise. (easy) Show that T is compact if and only if it maps the ball B_X to a precompact set in Y.

Since precompact sets are bounded, compact operators are always bounded, i.e. $\mathcal{K}(X{\to}Y) \le \mathcal{L}(X{\to}Y)$.

An operator is called a ***finite rank operator*** *if its image is finite-dimensional.*

6.1.3 Example. Every finite rank operator $T \in \mathcal{L}(X{\to}Y)$ is compact. Indeed, $T(B_X)$ is a bounded subset of a finite dimensional normed space $\mathrm{im}T \le Y$, so $T(B_X)$ is precompact by Heine-Borel theorem.

The next example is one of the main motivation to study compact operators.

6.1.4 Proposition (Compactness of the operator with a continuous kernel)**.** *The **integral operator** $T : C[0,1] \to C[0,1]$ defined as*

$$Tx\,(t) := \int_0^1 k(t,s)x(s)\,ds$$

*with **kernel** $k(t,s) \in C([0,1]^2)$. Then T is a compact operator.*

Proof. We need to show that $K := T(B)$ is a precompact subset of $C[0,1]$ (see Exercise 6.1.2). By Arzela-Ascoli Theorem, this would follow from (uniform) boundedness and equicontinuoity of the set K. The (uniform) boundedness of K follows from the boundedness of T. (Why?) To prove equicontinuity, we let $\varepsilon > 0$ and choose $\delta > 0$ such that

$$|\,t_1 - t_2\,| \le \delta \quad \Rightarrow \quad \forall\ s \in [0,1] \quad |\,k(t_1,s) - k(t_2,s)\,| \le \varepsilon.$$

Now, for every $x \in B_{C[0,1]}$, we obtain by triangle inequality that

$$|\,Tx(t_1) - Tx(t_2)\,| \le \int_0^1 |\,k(t_1,s){-}k(t_2,s)\,| \cdot |\,x(s)\,|\;ds \le \varepsilon.$$

as $|x(s)| \le 1$ for all s. This shows that the set K is equicontinuous, and therefore precompact. □

6.1.5 Exercise. ***Volterra operator*** T is defined as

$$Tx(t) := \int_0^t x(s)ds.$$

Show that Volterra operator is compact on $C[0,1]$, even though its kernel $k(t,s) = \chi_{\{s \le t\}}$ is discontinuous.

6.1.6 Proposition (Properties of $\mathcal{K}(X{\to}Y)$)**.**

(i) *The set of compact opera tors $\mathcal{K}(X{\to}Y)$ is a closed linear subspace of $\mathcal{L}(X{\to}Y)$.*

(ii) *$\mathcal{K}(X{\to}Y)$ is an operator ideal. This means that if $T \in \mathcal{K}(X{\to}Y)$ then the compositions ST and TS are both compact for every bounded linear operator S.*

Proof. (i) *Linearity* follows from the observation that the Minkowski sum of two precompact sets is precompact (see exercise below).

Closedness. Consider a sequence $T_n \in \mathcal{K}(X \to Y)$ such that $T_n \to T$ in $\mathcal{L}(X \to Y)$; we want to prove that $T \in \mathcal{K}(X \to Y)$. Let $\varepsilon > 0$ and choose $n \in N$ such that $\|T_n - T\| \le \varepsilon$. This means that

$$\|T_n(x) - T(x)\| \le \varepsilon$$

for every $x \in B_X$. This shows that $T_n(B_X)$ is a precompact-net of $T(B_X)$. Since ε is arbitrary, $T(B_X)$ is itself precompact. (Why?)

(ii) is straightforward and is left as an exercise. $\square$

Exercise. Prove that Minkowski sum $A+B$ of two precompact subsets A, B of a normed space is a precompact set.

6.1.7 Corollary (Isomorphisms are not compact)**.** *Let X be an infinite dimensional normed space. Then the identity operator on X is not compact. More generally, any isomorphism $T : X \to Y$ is not compact.*

Proof. For the identity operator on X, the result follows from F. Rieszs Theorem on non-compactness of B_X. As for the general statement, if an isomorphism $T : X \to Y$ were compact then the identity operator $T^{-1}T$ would also be compact by Proposition 6.1.6, which would be a contradiction. $\square$

As we know, finite rank operators are compact (6.1.3). More generally, since $\mathcal{K}(X \to Y)$ is closed, it follows that any operator that can be approximated by finite rank operators is also compact:

6.1.8 Corollary (Almost finite rank operators are compact)**.** *Suppose a linear operator $T : X \to Y$ can be approximated by finite rank operators $T_n \in \mathcal{L}(X \to Y)$, i.e.*

$$\|T_n - T\| \to 0 \quad \text{as } n \to \infty.$$

Then T is compact.

6.1.9 Exercise. Fix a sequence of real numbers $(v_k)_{k=1}^{\infty}$, and define the linear operator $T : l^2 \to l^2$ by

$$Tx = (v_k x_k)_{k=1}^{\infty}$$

For what multiplier sequences (v_k) is the operator T
(a) well defined? (b) bounded? (c) compact?

A digression on the continuation of linear operators

Similar to Section 4.3, we consider a normed space X and its subspace X_0. Consider an operator $T_0 \in \mathcal{L}(X_0 \to Y)$ where Y is some normed space. An *extension* of T_0 to the whole space X is an operator $T \in \mathcal{L}(X \to Y)$ whose restriction on X_0 coincides with T_0. As we know from Section 4.3, every bounded linear functional can be extended from either dense or closed subspace to the whole space. For dense subspaces, we can extend by continuity, while for closed subspaces the extension is guaranteed by Hahn–Banach theorem. For general linear operators, extension by continuity holds with the same proof as in Proposition 4.3.2:

6.1.10 Proposition (Extension by continuity)**.**

Let X_0 be a dense subspace of a normed space X, and Y be a Banach space. Then every operator $T_0 \in \mathcal{L}(X_0 \to Y)$ admits a unique extension $T \in \mathcal{L}(X \to Y)$. Moreover, $\|T\| = \|T_0\|$.

Unfortunately, extension from a closed subspace is not always possible, and Hahn-Banach theorem does not generalize to bounded linear operators.

Hilbert-Schmidt operators.

This is a most frequently used class of compact operators in Hilbert spaces.

6.1.11 Definition. *Let H be a separable Hilbert space, and let (e_k) be an orthonormal basis of H. A linear operator $T : H \to H$ is called a*

Hilbert-Schmidt operator *if*

$$\sum_{k=1}^{\infty} \|T(e_k)\|^2 < \infty$$

The quantity

$$\|T\|_{\mathrm{HS}} := \left(\sum_{k=1}^{\infty} \|T(e_k)\|^2 \right)^{1/2}$$

is called the ***Hilbert-Schmidt norm*** *of T.*

6.1.12 Proposition. *The definition of Hilbert–Schmidt operator and of the Hilber–Schmidt norm does not depend on the choice of an orthonormal basis of H.*

We will check this statement a little later.

6.1.13 Proposition. *A Hilbert–Schmidt operators are bounded. Specifically,*

$$\|T\| \le \|T\|_{\mathrm{HS}}\,.$$

Proof. Let T be a Hilbert–Schmidt operator on a Hilbert space H, and (e_k) be an orthonormal basis of H. It suffices to prove that the restriction of T on the dense subspace $\mathrm{span}(e_k)$ of H is bounded and has norm at most $\|T\|_{\mathrm{HS}}$; the result would then follow by extension by continuity (Proposition 6.1.10).

So let $x \in \mathrm{span}(e_k)$, which means that $x = \sum_k a_k e_k$ for some scalars a_k (finite sum). Then using triangle inequality and Cauchy–Schwarz inequality, we obtain

$$\|Tx\| = \left\| \sum_k a_k T(e_k) \right\| \le \sum_k |\, a_k \,| \, \|T(e_k)\| \le$$

$$\left(\sum_k |\, a_k \,|^2 \right)^{1/2} \cdot \left(\sum_k \|T(e_k)\|^2 \right)^{1/2} = \|x\| \cdot \|T\|_{\mathrm{HS}}\,.$$

□

A similar construction in function spaces L^2 leads to the notion of Hilbert-Schmidt integral operators. To this end, consider a function $k(t,s) \in L^2([0,1]^2)$ which we call the kernel. Define a linear operator $T : L^2[0,1] \to L^2[0,1]$ as

$$Tx\,(t) := \int_0^1 k(t,s)x(s)\,ds \tag{6.1}$$

We can view this definition as a continuous version of (2.10), where kernel $k(t,s)$ can be considered as a continuous version of matrix. The operator T defined this way is called ***Hilbert-Schmidt integral operator*** with kernel $k(t,s)$.

6.1.14 Proposition (Hilbert-Schmidt integral operators)**.** *A Hilbert–Schmidt integral operator $T : L^2[0,1] \to [0,1]$ with kernel*

$$k(t,s) \in L^2([0,1]^2)$$

is a Hilbert-Schmidt operator. Specifically, it is bounded and

$$\|T\| \le \|T\|_{\mathrm{HS}} = \|k\|_2 \,.$$

Proof. We will view the integral in the definition of T as the inner product of x with the kernel k. Specifically, consider the function $K_t(s) = k(t,s)$; then

$$(Tx)(t) = \langle K_t(s), x\rangle \quad \text{for every} \quad t \in [0,1].$$

Let us fix some orthonormal basis (e_k) of $L^2[0,1]$. Then

$$\|T\|_{\mathrm{HS}}^2 = \sum_k \|Te_k\|_2^2 = \sum_k \int_0^1 |\,Te_k(t)\,|^2\,dt = \sum_k \int_0^1 |\,\langle K_t, e_k\rangle\,|^2\,dt =$$

$$= \int_0^1 \sum_k |\,\langle K_t, e_k\rangle\,|^2\,dt \quad \text{(by monotone convergence theorem)}$$

$$= \int_0^1 \|K_t\|_2^2\,dt \quad \text{(by Parseval's identity)}$$

$$= \|k\|_2^2 \,. \quad \text{by definition of } K_t \text{ and Fubini theorem.}$$

□

Equality $\|T\|_{\mathrm{HS}} = \|k\|_2$ proves Proposition 6.1.12 for the integral case of Hilbert-Schmidt operator.

Adjoint operators. Compactness of the adjoint operator.

The concept of adjoint operator is a generalization of matrix transpose in linear algebra. Recall that if $A = (a_{ij})$ is an $n \times n$ matrix with complex entries, then the Hermitian transpose of A is the $n \times n$ matrix A^* (or A^T) $A^* = (a_{ji})$. The transpose thus satisfies the identity

$$\langle A^* x, y\rangle = \langle x, Ay\rangle, \quad x, y \in \mathbb{C}^n. \tag{6.2}$$

Now we would like to extend this to a general definition of the adjoint T^* for a linear operator $T : X \to Y$ acting between normed spaces X and Y.

6.1.15 Definition. *Let $T \in \mathcal{L}(X\to Y)$. The **adjoint** operator $T^* \in \mathcal{L}(Y^*\to X^*)$ is defined by*

$$(T^* f)(x) := f(Tx), \quad f \in Y^*, \quad x \in X.$$

In order to see a similarity with (7.1), we adopt the following alternative notation for the action of functionals, one that resembles the inner product:

$$\langle f, x\rangle := f(x), \quad f \in X^*, \quad x \in X. \tag{6.3}$$

Notice that if X is a Hilbert space, this notation agrees with the inner product by Riesz representation theorem (up to complex conjugation). In general, $\langle f, x\rangle$ does not define an inner product since the arguments are taken from different spaces. Then the definition of the adjoint reads as

$$\langle T^* f, x\rangle = \langle f, Tx\rangle, \quad f \in Y^*, \quad x \in X.$$

and we see that this is a general form of (7.1).

6.1.16 Remark. *(Adjoint operators on Hilbert spaces)* For operators T on a Hilbert space $X{=}Y{=}H$, the definition (6.1.15) of the adjoint operator T takes place with $\langle\cdot,\cdot\rangle$ denoting the inner product on H. This makes a small difference — the inner product is *conjugate linear* in the second argument, while the action of functionals (6.3) is just *linear*. So, for operators on a Hilbert space one has $(aT)^* = \overline{a}T^*$ for $a \in \mathbb{C}$, while the general definition of adjoint for Banach spaces incurs $(aT)^* = aT^*$.

One point has not been justified in Definition 6.1.15, why T is a bounded linear operator. We shall prove this now:

6.1.17 Proposition. *For every $T \in \mathcal{L}(X{\to}Y)$ it holds $T^* \in \mathcal{L}(Y^*{\to}X^*)$, and*

$$\|T^*\| = \|T\|.$$

Proof. Denoting S_X the unit sphere of X, and using notation (6.3), we have

$$\begin{aligned}\|T^*\| &= \sup_{f\in S_{Y^*}} \|T^*f\|_{X^*} = \sup_{f\in S_{Y^*}} \sup_{x\in S_X} |\langle T^*f, x\rangle| = \sup_{f\in S_{Y^*}} \sup_{x\in S_X} |\langle f, Tx\rangle| = \\ &= \sup_{x\in S_X} \|Tx\|_Y \qquad \text{(choosing } f \text{ as a supporting functional of } Tx) \\ &= \|T\|.\end{aligned}$$

as required. □

6.1.18 Reminder. *If (e_k) is a basis of a Hilbert space H, that is for each element $x \in H$ it holds*

$$x = \sum_{k=1}^{\infty} \langle x, e_k\rangle e_k,$$

than the following Parseval's identity is valid:

$$\|x\|^2 = \sum_k \|\langle x, e_k\rangle\|^2. \tag{6.4}$$

It follows from 1.19.

Using the conjugate operator makes it easy to prove Prop 6.1.12 in the general case.

Proof. Assume that $\sum_k \|Te_k\|^2 < \infty$ for some orthonormal basis (e_k) of H. Using Parseval's identity twice, we obtain

$$\sum_k \|Te_k\|^2 = \sum_{k,j} |\langle Te_k, e_j\rangle|^2 = \sum_{k,j} |\langle e_k, T^*e_j\rangle|^2 = \sum_j \|T^*e_j\|^2. \tag{6.5}$$

Let $\widetilde{e}_k$ be another orthonormal basis of H. Then a similar argument gives

$$\sum_j \|T^*e_j\|^2 = \sum_{k,j} |\langle \widetilde{e}_k, T^*e_j\rangle|^2 = \sum_{k,j} |\langle T\widetilde{e}_k, e_j\rangle|^2 = \sum_k \|T\widetilde{e}_k\|^2. \tag{6.6}$$

This completes the proof. $\square$

6.1.19 Remark. As a byproduct of the proof, we obtained in (6.5) that

$$\|T^*\|_{\mathrm{HS}} = \|T\|_{\mathrm{HS}}$$

6.1.20 Exercise. Let T be the Hilbert-Schmidt integral operator with kernel $k(t,s)$. Show that T^* is also the Hilbert-Schmidt integral operator with kernel $\overline{k(s,t)}$.

6.1.21 Exercise. Let R and L denote the right and left shift operators on l^2:

$$R(x) = (0, x_1, x_2, \dots); \quad L(x) = (x_2, x_3, x_4, \dots) \quad \text{for} \quad x = (x_1, x_2, \dots).$$

Prove that $R^* = L$.

6.1.22 Exercise. (I) Let $T \in \mathcal{L}(X \to Y)$ and $S \in \mathcal{L}(Y \to Z)$. Show that

$$(ST)^* = T^*S^*.$$

(II) Let $S, T \in \mathcal{L}(X \to Y)$ and $a, b \in \mathbb{C}$. Show that

$$(aS + bT)^* = \overline{a}S^* + \overline{b}T^*.$$

(III) Let $T \in \mathcal{L}(X \to Y)$ be such that $T^{-1} \in \mathcal{L}(Y \to X)$. Show that

$$(T^{-1})^* = (T^*)^{-1}.$$

The kernel and image of bounded linear operators are in a duality relation. To state it, we consider a generalization of the notion of orthogonal complement, which we studied in Section 1.8 for Hilbert spaces.

A similar duality principle holds for compact operators:

6.1.23 Definition. *An **annihilator** of a subset A of a normed space X is the set $A^\perp \subseteq X^*$ defined as*

$$A^\perp = \{f \in X^* \ : \langle f, x\rangle = 0 \quad \textit{for all } x \in A\}.$$

6.1.24 Proposition (Duality of kernel and image)**.** *Let $T \in \mathcal{L}(X \to Y)$. Then*

$$(\operatorname{im} T)^\perp = \ker T^*.$$

Proof. Let $f \in Y^*$. Then $f \in \ker T^*$ means that $T^* f = 0$, which is equivalent to $\langle T^* f, x\rangle = \langle f, Tx\rangle = 0$ for all $x \in X$, which means that $f \in (\operatorname{im} T)^\perp$. □

6.1.25 Corollary. *Let H be a Hilbert space, and $T \in \mathcal{L}(H \to H)$. Then the orthogonal decomposition holds:*

$$H = \overline{\operatorname{im} T} \oplus \ker T^*.$$

Proof. By Proposition 6.1.24, we have $(\overline{\operatorname{im} T})^\perp = (\operatorname{im} T)^\perp = \ker T^*$ (why does the first identity hold?). By Remark 4.2.1, the proof is complete. □

6.1.26 Exercise (Duality of kernel and image II)**.** For a subset $B \subseteq X^*$, define the «***pre-anihilator***» as

$$B_\perp = \{x \in X \ : \langle f, x\rangle = 0 \quad \text{for all } f \in B\}.$$

Let $T \in \mathcal{L}(X \to Y)$. Prove that

$$\ker T = (\operatorname{im} T^*)_\perp.$$

Deduce that

$$(\ker T)^\perp = \operatorname{im} T^*.$$

Give an example of a linear operator T such that $(\ker T)^\perp \neq \operatorname{im} T^*$.

Recall the basic duality property for bounded linear operators: if $T \in \mathcal{L}(X \to Y)$ then $T^* \in \mathcal{L}(Y^* \to X^*)$ and $\|T\|^* = \|T\|$ (see 6.1.17). A similar duality principle holds for compact operators:

6.1.27 Theorem (Schauder). *Let X and Y be Banach spaces. If $T \in \mathcal{K}(X \to Y)$ then $T^* \in \mathcal{K}(Y^* \to X^*)$.*

Proof. Given $f \in Y^*$, we are seeking a bound on

$$\|T^* f\|_{X^*} = \sup_{x \in B_X} |(T^* f)x| = \sup_{x \in B_X} |f(Tx)| = \sup_{y \in K} |f(y)|, \tag{6.7}$$

where $K := \overline{T(B_X)}$. (Taking the closure here is justified by continuity of f). We shall interpret the identity (6.7) in topological terms. Indeed, we know that K is compact, and we need to prove that $E := T^*(B_{Y^*})$ is precompact in X^*. Let us embed the subset $E \subseteq X^*$ into $C(K)$ and use Arzela-Ascoli theorem. Namely, we define the embedding $J : E \to C(K)$ by

$$J(T^* f) := f|_K.$$

Then identity (6.7) implies that

$$\|T^* f\|_{X^*} = \big\| f\big|_K \big\|_{C(K)} \quad \text{for every } f \in Y^*.$$

which shows that J is an isometric (thus homeomorphic) embedding.

Now, $J(E)$ is (uniformly) bounded in $C(K)$ as

$$\|T^* f\|_{X^*} = \|T^*\| \, \|f\|_{Y^*} \leq \|T\| \quad \text{for every } f \in B_{Y^*}.$$

Moreover, $J(E)$ is equicontinuous. Indeed, for every $f \in B_{Y^*}$ and for $y_1, y_2 \in K$ we have

$$\left| f\big|_K(y_1) - f\big|_K(y_2) \right| = |\, f(y_1 - y_2)\,| \leq \|f\|_{X^*} \|y_1 - y_2\| \leq \|y_1 - y_2\| .$$

Arzela-Ascoli theorem completes the proof. $\square$

6.1.28 Exercise (Compactness of integral operators)**.**
Consider an intergal operator T with kernel $k(t,s) : [0,1]^2 \to \mathbb{R}$ which satisfies the following:
(i) for each $s \in [0,1]$, the function $k_s(t) = k(t,s)$ is integrable in t;
(ii) the map $s \mapsto k_s$ is a continuous map from $[0,1]$ to $L^1[0,1]$.
Show that the integral operator T is compact in $C[0,1]$.

6.2 Fredholm theory

Fredholm theory studies operators of the form «identity plus compact». They are conveniently put in the form $I - T$ where I is the identity operator on some Banach space X and $T \in \mathcal{K}(X \to X)$. Fredholm theory is motivated by two applications. One is for solving linear equations $\lambda x - Tx = b$, and in particular integral equations (T being an integral operator). Another related application is in spectral theory, where the spectrum of T consists of numbers for which the operator $\lambda I - T$ is invertible. We will discuss both applications later.

We start with a small digression about isomorphic embeddings. As we know, the kernel of every operator $T \in \mathcal{L}(X \to Y)$ is always a closed subspace. The image of T may or may not be closed. The following result characterizes operators with closed images.

6.2.1 Proposition (Isomorphic embeddings)**.** *Let $T \in \mathcal{L}(X \to Y)$ be an operator between Banach spaces X and Y. Then the following are equivalent:*

(I) *T is an* ***isomorphic embedding****, i.e. T acts as an isomorphism between spaces X and $\operatorname{im} T \subseteq Y$;*

(ii) *T is injective and it has closed image;*
(iii) *T is bounded below, i.e. one can find $c > 0$ such that*

$$\|Tx\| \geq c\,\|x\| \quad \textit{for all} \quad x \in X.$$

Proof. (i) $\Rightarrow$ (ii). Recall that isomorphisms preserve completeness of spaces. Since X is complete, it follows that the subspace $\operatorname{im} T$ is complete, so it is closed. Injectivity of T is an obvious consequence of the isomorphic embedding property.

(ii) $\Rightarrow$ (iii). Considering T as an operator from X to $\operatorname{im} T$ we see that T is injective and surjective. By inverse mapping theorem, T is an homeomorphism.

(iii) $\Rightarrow$ (i). We have $c\,\|x\| \leq \|Tx\| \leq C\,\|x\|$ for all $x \in X$, where $C = \|T\|$. It follows that T is an isomorphic embedding (how?). □

In Theorem 6.2.5, we will need the elementary properties of the quotient space as a Banach space.

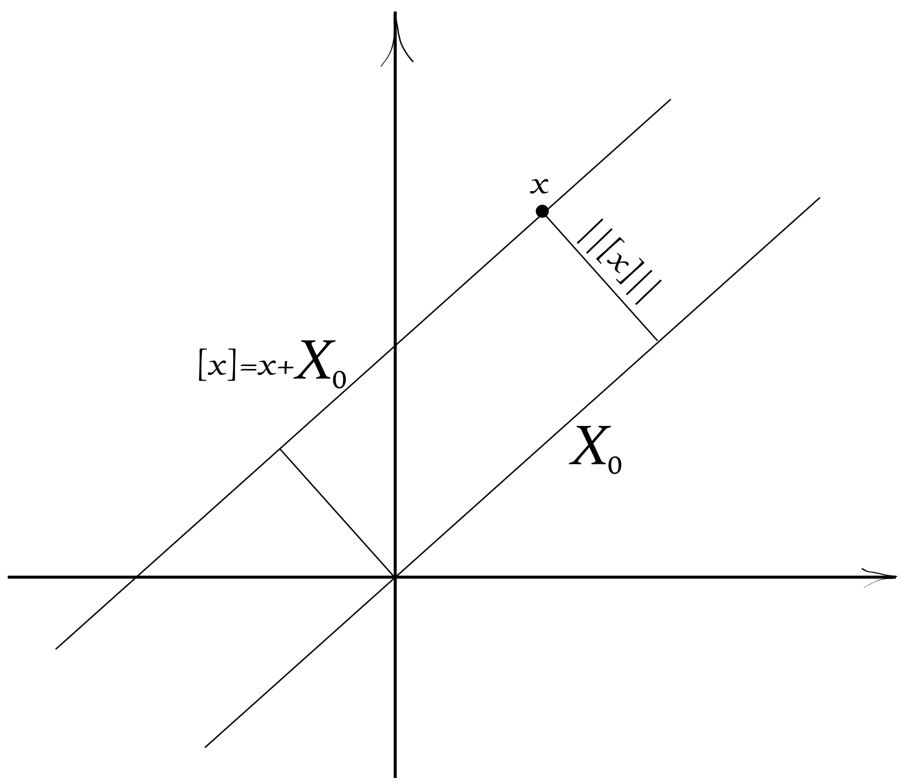


Figure 9: $\|\,[x]\,\| = \operatorname{dist}(0, [x]) = \operatorname{dist}(x, X_0)$.

6.2.2 Definition. *Let X be a normed space and X_0 be a closed subspace of X. We define a* ***norm on*** *X/X_0 as follows. For every coset $[x] = x + X_0$, we put*

$$\| [x] \| := \inf_{y \in X_0} \|x + y\| .$$

It is easy to understand the norm in the quotient space geometrically as the distance from the origin to the coset $[x]$, or from x to X_0:

$$\| [x] \| = \operatorname{dist}(0, [x]) = \operatorname{dist}(x, X_0) \tag{6.8}$$

6.2.3 Proposition. *The definition of $\| [\,\cdot\,] \|$ above indeed produces a norm on X/X_0.*

Proof. First we observe that the number $\| [x] \|$ is well defined, i.e. it does not depend on a choice of a representative x in the coset $[x]$. This clearly follows from the geometric definition (6.8). Next, we have to check the three norm axioms.

(i) Assume that $\| [x] \| = 0$. Then, from the geometric definition (6.8) we see that 0 is a limit point of X_0. Since X_0 is closed, so is $x \in X_0$. Therefore $0 = -x + x \in [x]$. Hence $[x] = [0]$, which verifies norm axiom (i).

(ii) Let $x \in X$ and $\lambda \in \mathbb{R}$. Then

$$\| [\lambda x] \| = \inf_{y \in X_0} \|\lambda x + y\| = \inf_{y \in X_0} \|\lambda x + \lambda y\| = |\,\lambda\,| \inf_{y \in X_0} \|x + y\| = |\,\lambda\,| \cdot \| [x] \|$$

This verifies norm axiom (ii).

(iii) Let us $x_1, x_2 \in X$; we want to show that $\|[x_1 + x_2]\| \leq \|[x_1]\| + \|[x_2]\|$. To this end, fix an arbitrary $\varepsilon > 0$. By the definition of the quotient norm there exist $y_1, y_2 \in Y$ so that

$$\|x_1 + y_1\| \leq \|[x_1]\| + \varepsilon, \quad \|x_2 + y_2\| \leq \|[x_2]\| + \varepsilon.$$

Using triangle inequality for the norm in X, we obtain

$$\|x_1 + x_2 + y_1 + y_2\| \leq \|x_1 + y_1\| + \|x_2 + y_2\| \leq \|[x_1]\| + \|[x_2]\| + 2\varepsilon$$

We conclude that

$$\|[x_1 + x_2]\| = \inf_{y \in X_0} \|x_1 + x_2 + y\| \leq$$
$$\|x_1 + x_2 + y_1 + y_2\| \leq \|[x_1]\| + \|[x_2]\| + 2\varepsilon.$$

Since $\varepsilon > 0$ is arbitrary, this completes the proof of norm axiom (iii). □

6.2.4 Exercise. Let X be a Banach space and X_0 be a (linear) subspace of X. Show that if X_0 is closed, then X/X_0 is a Banach space (recall that X_0 has to be closed in order for X/X_0 to be well defined).

6.2.5 Theorem. *Let X be a Banach space and $T \in \mathcal{K}(X \to X)$. Then operator $I - T$ has closed image.*

Proof. The argument relies on the characterization of injective operators with closed image, Proposition 6.2.1. So we consider the operator $A = I - T$ and its injectivization $\widetilde{A} : X/\ker A \to X$ (see 3.4.11). Since $\operatorname{im} A = \operatorname{im} \widetilde{A}$, it suffices to show that $\widetilde{A}$ is bounded below.

Suppose the contrary, that there exists cosets $[x_k]$ with $\|\,[x_k]\,\| = 1$ and such

$$\widetilde{A}[x_k] \to 0.$$

We can find representatives $x_k \to [x_k]$ with, say, $\|x_k\| \leq 2$ and such that

$$\operatorname{dist}(x_k, \ker A) = \|\,[x_k]\,\| = 1 \quad \text{but} \quad Ax_k = \widetilde{A}[x_k] \to 0.$$

So we have $Ax_k = x_k - Tx_k \to 0$. By compactness of T, we can assume (passing to a subsequence if necessary) that $Tx_k \to z$ for some $z \in X$. It follows that $x_k \to z$. Therefore $Ax_k \to Az$; but we already know that $Ax_k \to 0$, thus $Az = 0$. So $z \in \ker A$. Furthermore, since $x_k \to z$ it follows that $\operatorname{dist}(x_k, \ker A) \to 0$. This contradiction completes the proof.
□

Fredholm alternative

We now state and prove a partial case of the so-called Fredholm alternative.

6.2.6 Theorem (Fredholm alternative). *Let X be a Banach space and $T \in \mathcal{K}(X \to X)$. Then operator $A = I - T$ is injective if and only if A is surjective.*

Proof. *Necessity.* Assume that A is injective but not surjective. Consider the subspaces of X

$$X_n := \operatorname{im}(A^n), \quad n = 0, 1, \dots$$

Then

$$X_0 \supset X_1 \supset X_2 \supset \dots$$

is a chain of *proper* inclusions. Indeed, the first inclusion $X \supset \operatorname{im}(A)$ is proper by assumption; the claim follows by induction (why?).

Furthermore, X_n are *closed* subspaces of X. Indeed, by Newtons binomial expansion we see that $A^n = (I - T)^n$ has the form $A = I - T_1$ for some compact operator T_1, so the claim follows from Theorem 6.2.5. By Hahn-Banach theorem (see Exercise 4.3.28) we can find functionals

$$f_n \in X_n^* \quad \text{such that } \|f_n\| = 1, \quad f_n \in X_{n+1}^{\perp}.$$

We can extend f_n, again by Hahn-Banach theorem, so that $f_n \in X^*$. We are going to show that the sequence $(T^* f_n)$ has no convergent subsequences. This will contradict the compactness of T and, by Schauders Theorem 6.1.27, the compactness of T. To this end, let us fix $n > m$ and compute the pairwise distances

$$d_{n,m} := \|T^* f_n - T^* f_m\| = \|T^*(f_n - f_m)\| = \|(I - T)^*(f_n - f_m) + f_n - f_m\|$$

So

$$d_{n,m} \geq \sup_{x \in B_{X_n}} |\langle T^* f_n - T^* f_m, x \rangle| =$$

$$\sup_{x \in B_{X_n}} |\langle f_n - f_m, (I - T)x \rangle + \langle f_n - f_m, x \rangle|$$

Now, $(I - T)x = Ax \in X_{n+1}$ while $f_n - f_m \in X_{n+1}^{\perp}$ by construction, so $\langle f_n - f_m, (I - T)x \rangle = 0$. Further, $x \in X_n$ while $f_m \in X_n^{\perp}$, so $\langle f_m, x \rangle = 0$.

Therefore

$$d_{n,m} \geq \sup_{x \in B_{X_n}} |\langle f_n, x \rangle| = 1.$$

by construction. It follows that the terms of the sequence $(T^* f_n)$ are pairwise separated, so there can not be any convergent subsequence. This completes the proof of the necessity direction.

Suffciency will follow from a duality argument. We use the relations

$$(\operatorname{im} A)^\perp = \ker A^*, \quad (\ker A)^\perp \subseteq \operatorname{im} A^* \tag{6.9}$$

which we proved in Proposition 6.1.24 and Exercise 6.1.26. So, assume that $A = I - T$ is surjective. Then $A^* = I^* - T^*$ is injective by (6.9). Since T^* is compact by Schauders theorem, the first part of the proof gives that A^* is surjective. This implies that A is injective by (6.9). □

6.2.7 Remark (Compactness is essential)**.** Fredholm alternative does not hold for non-compact operators in general. For example, the right shift operator in l^2 is injective but not surjective; the left shift operator in l^2 is surjective but not injective.

The name «*Fredholm alternative*» is explained by the following application to solving linear equations of the form

$$\lambda x - Tx = b$$

where $T \in \mathcal{K}(X \to X)$, $\lambda \in \mathbb{C}$, $b \in X$. One is interested in existence and uniqueness of solution. Theorem 6.2.6 states that **exactly one** the following statements holds for every $\lambda \neq 0$:
either the homogeneous equation $\lambda x - Tx = 0$ has a nontrivial solution,
or the inhomogeneous equation $\lambda x - Tx = b$ has a solution for every b; this solution is automatically unique.

This alternative is particularly useful for studying integral equations, since for the integral operator $Tx\,(t) = \int_0^1 k(t,s)x(s)\,ds$ the homogeneous

Fredholm equation is

$$\lambda\, x(t) - \int_0^1 k(t,s)x(s)\, ds = 0,$$

while the inhomogeneous Fredholm equation («of second kind») is

$$\lambda\, x(t) - \int_0^1 k(t,s)x(s)\, ds = b(t).$$

6.2.8 Exercise (General Fredholm alternative)**.** *Let X be a Banach space and $T \in \mathcal{K}(X \to X)$. Show that operator $A = I - T$ satisfies*

$$\dim \ker A = \dim \ker A^* = \operatorname{codim} \operatorname{im} A = \operatorname{codim} \operatorname{im} A^*.$$

6.3 Spectrum of a bounded linear operator

Studying linear operators through their spectral properties is a powerful approach in analysis and mathematical physics. Recall from linear algebra that the *spectrum* of a linear operator T acting on $\mathbb{C}^n$ consists of the *eigenvalues* of T, which are the numbers $\lambda \in \mathbb{C}$ such that $Tx = x$ for some nonzero vector $x \in \mathbb{C}^n$; such x are called the *eigenvectors* of T. Eigenvalues always exist by the fundamental theorem of algebra, as they are the roots of the characteristic polynomial $\det(T - \lambda I) = 0$. There are at most n eigenvalues of T, or one can say exactly n counting multiplicities. Eigenvectors corresponding to different eigenvalues are linearly independent.

Infinnite-dimensional normed spaces, the spectrum is a richer concept than in finite-dimensional spaces. Let us illustrate the difference on two examples.

6.3.1 Examples. 1) (*Uncoutnable number of eigenvalues*) Consider the differential operator

$$T = \frac{d}{dt}$$

acting, for example, on $C^\infty(\mathbb{R})$ (the space of complex-valued functions). To compute the spectrum of T, we solve the ordinary differential equation $u' = \lambda u$. The solution has the form $u(t) = Ce^{\lambda t}$. Therefore, every $\lambda \in \mathbb{C}$ is an eigenvalue of T.

2)(*No eigenvalues*). Consider a multiplication operator on $L^2[0,1]$ acting as

$$(Tx)(t) = tx(t).$$

Suppose λ is an eigenvalue of T with eigenvector $x \in L^2[0,1]$. This means that the following identity holds

$$t\,x(t) = \lambda\,x(t) \quad \text{for all} \quad t \in [0,1].$$

It follows that $x = 0$. Therefore, T has no eigenvalues.

6.3.2 Definition. *Let X be a normed space and $T \in \mathcal{L}(X \to X)$. A number $\lambda \in \mathbb{C}$ is called a **regular point** if $T - \lambda I$ is a homeomorphism (i.e. $(T - \lambda I)^{-1} \in \mathcal{L}(X \to X)$). All other λ are called **spectrum points**. The set of all regular points is denoted $\rho(T)$ and is called the **resolvent set** of T. The set of all spectrum points is denoted $\sigma(T)$ and is called the **spectrum of** T.*

For operators T acting on a finite dimensional space, the spectrum consists of eigenvalues of T. In infinite dimensions, this is not true, as there are various reasons why $T - \lambda I$ may be non-invertible. These reasons are listed in the following definition:

6.3.3 Definition. *Let X be a normed space and $T \in \mathcal{L}(X \to X)$.*

(I) *The **point spectrum** $\sigma_p(T)$ is the set of all eigenvalues of T, i.e. the numbers $\lambda \in \mathbb{C}$ satisfying*

$$\ker(T - \lambda I) \neq \mathbf{0}.$$

(II) *The **continuous spectrum** $\sigma_c(T)$ is the set of all $\lambda \in \mathbb{C}$ such that $\lambda \in \mathbb{C}$ satisfying*

$$\ker(T - \lambda I) = \mathbf{0} \quad \text{and} \quad \operatorname{im}(T - \lambda I) \text{ is dense in } X.$$

(III) *The* ***residual spectrum*** $\sigma_r(T)$ *is the set of all* $\lambda \in \mathbb{C}$ *such that* $\lambda \in \mathbb{C}$ *satisfying*

$$\ker(T-\lambda I)=\mathbf{0} \quad \textit{and} \quad \operatorname{im}(T-\lambda I) \textit{ is not dense in } X.$$

So the spectrum of T can be expressed as a disjoint union

$$\sigma(T)=\sigma_p(T)\,\vee\,\sigma_c(T)\,\vee\,\sigma_r(T).$$

Let us now compute and classify the spectrum of come basic linear operators.

6.3.4 Examples. 1) (The *spectrum of diagonal operator* on l^2). Let us fix some sequence $\lambda_k \to 0$ in $\mathbb{C}\setminus\{0\}$, and consider the operator T on l^2 defined as

$$T(x_k)_{k=1}^{\infty}=(\lambda_k x_k)_{k=1}^{\infty}.$$

As $(T-\lambda I)x=((\lambda_k-\lambda)x_k)_{k=1}^{\infty}$, we have $(T-\lambda I)^{-1}y=\left(\frac{y_k}{\lambda_k-\lambda}\right)_{k=1}^{\infty}$. It follows that $(T-\lambda I)^{-1}$ is a bounded operator if and only if λ is not in the closure of $\{\lambda_k\}_{k=1}^{\infty}\cup\{0\}$.

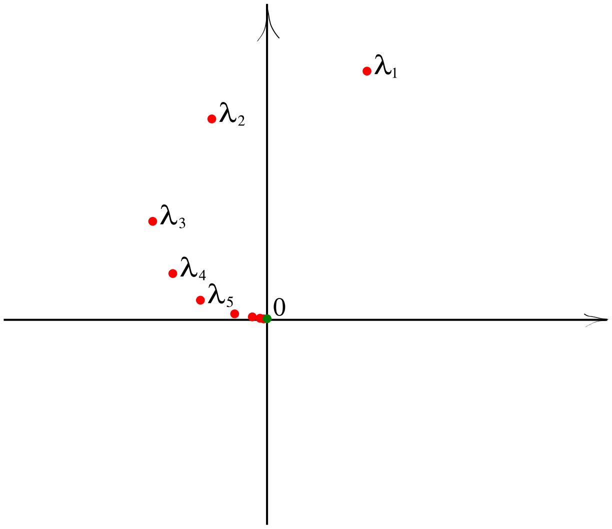
All λ_k are clearly the eigenvalues of T as $Te_k=\lambda_k e_k$ for the canonical basis (e_k) of l^2. 0 is not an eigenvalue since T is injective (as all $\lambda_k \neq 0$). So 0 is either in continuous or residual spectrum. Now, $\operatorname{im} T$ is dense in l^2 (why?), so 0 is in the continuous spectrum. Our conclusion is (see fig.10 a)):

$$\sigma_p(T)=\{\lambda_k\}_{k=1}^{\infty}, \quad \sigma_c(T)=\{0\}, \quad \sigma_r(T)=\varnothing.$$

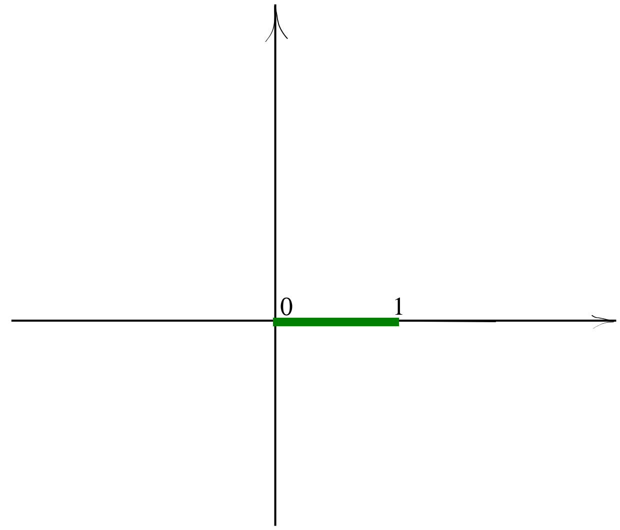
2) (The *spectrum of the multiplication operator* on L^2). Let's explore the spectrum of the operator from the 6.3.1 2), i.e. $(Tx)(t)=tx(t)$. As $(T-\lambda I)x(t)=(t-\lambda)x(t)$, we have

$$(T-\lambda I)^{-1}y(t)=\frac{1}{(t-\lambda)}y(t). \tag{6.10}$$

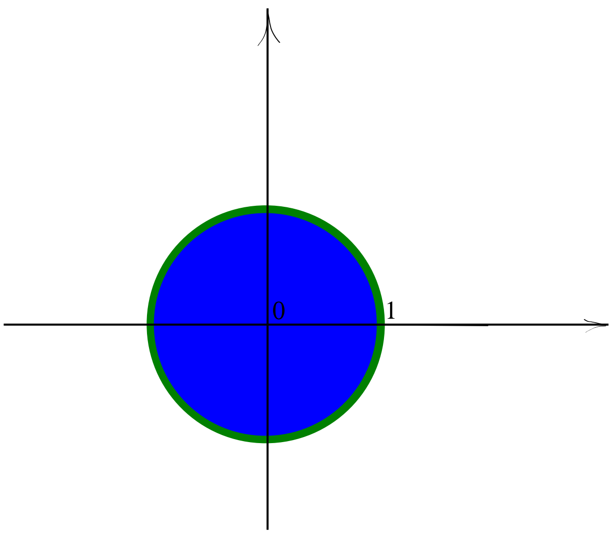
If $\lambda \notin [0,1]$ then the function $\frac{1}{(t-\lambda)}$ is bounded, thus $(T-\lambda I)^{-1}$ is a bounded operator. Therefore such λ are regular points. Conversely, if



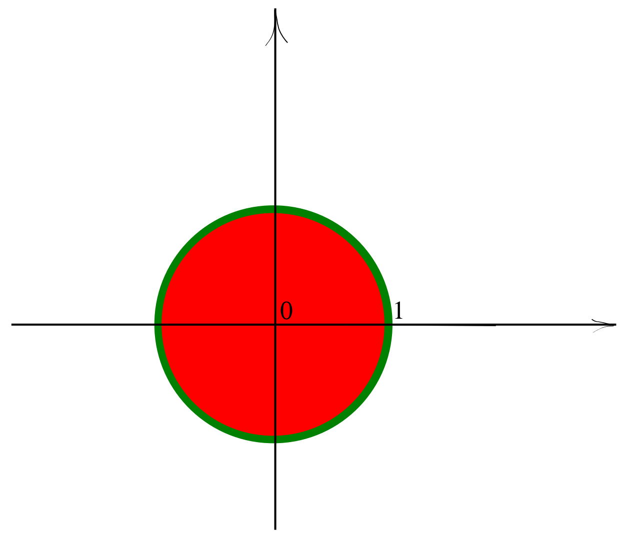
a) The spectrum of diagonal operator $T(x_k)_{k=1}^{\infty} = (\lambda_k x_k)_{k=1}^{\infty}$ on l^2



b) The spectrum of the multiplication operator $(Tx)(t){=}tx(t)$ on $L^2[0,1]$.



c) The spectrum of the right shift operator $R(x){=}(0, x_1, x_2, \ldots)$ on l^2



d) The spectrum of the left shift operator $L(x) = (x_2, x_3, \ldots)$ on l^2.

Figure 10: Examples of Point, continuous and residual spectra of operators.

$\lambda \in [0,1]$ then $\frac{1}{(t-\lambda)} \notin L^2[0,1]$ because of the non-integrable singularity at 0. Hence $T - \lambda I$ is not invertible (at $y(t) \equiv 1$) Hence all such λ are spectrum points. Therefore, $\sigma(T) = [0,1]$.

As we noticed in Example 6.3.1 2), T has no eigenvalues. It follows from (6.10) that $\operatorname{im}(T - \lambda I)$ is dense in $L^2[0,1]$ (why?) Our conclusion is (see fig.10 b)):

$$\sigma_p(T) = \varnothing, \quad \sigma_c(T) = [0, 1], \quad \sigma_r(T) = \varnothing.$$

6.3.5 Exercise. (The *spectrum of the shift operators*) Consider the right and left shift operators on l^2, acting on a vector $x = (x_1, x_2, \dots)$ (see Example 6.1.21)

$$R(x) = (0, x_1, x_2, \dots); \quad L(x) = (x_2, x_3, x_4, \dots).$$

Since R is clearly injective but $\operatorname{im} R$ is not dense in l^2 (why?), 0 is in the residual spectrum of R. Show that

$$\sigma_p(R) = \varnothing, \quad \sigma_c(R) = \{\lambda \in \mathbb{C} : |\lambda| = 1\}, \quad \sigma_r(R) = \{\lambda \in \mathbb{C} : |\lambda| < 1\}.$$

$$\sigma_p(L) = \{\lambda \in \mathbb{C} : |\lambda| < 1\}, \quad \sigma_c(L) = \{\lambda \in \mathbb{C} : |\lambda| = 1\}, \quad \sigma_r(L) = \varnothing.$$

Studying the spectrum of an operator is convenient via the so-called resolvent operator:

6.3.6 Definition. *Let X be a Banach space and $T \in \mathcal{L}(X \to X)$. For $\lambda \in \rho(T)$ define the* ***resolvent operator*** $R_\lambda \in \mathcal{L}(X \to X)$ *by*

$$R_\lambda := (T - \lambda I)^{-1}.$$

So the resolvent is a function $R : \rho(T) \to \mathcal{L}(X \to X)$. The resolvent operator can be computed in terms of series expansion involving T. This technique is based on the following simple lemma:

6.3.7 Lemma (Von Neumann)**.** *Let X be a Banach space and $F \in \mathcal{L}(X \to X)$ such that $\|F\| < 1$. Then $I - F$ is invertible, and its inverse can be expressed as a convergent series in $\mathcal{L}(X \to X)$:*

$$(I - F)^{-1} = \sum_{k=0}^{\infty} F^k, \tag{6.11}$$

$$\left\|(I - F)^{-1}\right\| \leq \frac{1}{1 - \|F\|}.$$

Proof. The series $\sum_{k=0}^{\infty} F^k$ converges absolutely because $\left\|F^k\right\| \leq \|F\|^k$ while $\|F\| < 1$. Furthermore,

$$(I-F)\sum_{k=0}^{\infty} F^k = \sum_{k=0}^{\infty} F^k (I-F) = I$$

as telescoping series (why?). Finally,

$$\left\|(I-F)^{-1}\right\| \leq \sum_{k=0}^{\infty} \left\|F^k\right\| \leq \sum_{k=0}^{\infty} \|F\|^k = \frac{1}{1-\|F\|}.$$

This completes the proof. □

6.3.8 Definition. *Let $O \subseteq \mathbb{C}$ be an open set, let X be a complex Banach space, and let $F : O \to X$ be a function.*

(i) *The function F is called **(strongly) holomorphic** if the limit*

$$F'(z) := \lim_{h\to 0} \frac{F(z+h)-F(z)}{h}$$

exists for all $z \in O$.

6.3.9 Theorem (Resolvent set and resolvent operators properties)**.** *Let X be a complex Banach space and $T \in \mathcal{L}(X \to X)$. Then*

(I) *the regular set $\rho(T)$ is an open set;*

(II) *the resolvent $R : \rho(T) \to \mathcal{L}(X \to X) : \lambda \mapsto R_\lambda$ is holomorphic and satisfies*

$$R_\lambda - R_\nu = (\lambda - \nu) R_\lambda R_\nu \tag{6.12}$$

*for all $\lambda, \nu \in \rho(T)$. Equation (6.12) is called the **resolvent identity**.*

(III) *The resolvent $R : \lambda \mapsto R_\lambda$ is an **analytic** operator-valued function on its domain $\rho(T)$. Specifcally, for any $\lambda_0 \in \rho(T)$ the operator $R(\lambda)$ can be expressed as a power series convergent in the circle $B\left(\lambda_0, \|R_{\lambda_0}\|^{-1}\right)$:*

$$R_\lambda = \sum_{k=1}^{\infty} (\lambda-\lambda_0)^{k-1} R_{\lambda_0}^k. \tag{6.13}$$

(IV) $\{\lambda \,:\, |\,\lambda\,| > \|T\|\} \subseteq \rho(T)$, *and if* $|\,\lambda\,| > \|T\|$ *then*

$$R_\lambda = -\sum_{k=1}^{\infty} \lambda^{-k-1} T^k, \quad \|R_\lambda\| \leq \frac{1}{|\,\lambda\,| - \|T\|}. \tag{6.14}$$

Proof. (I) Let $\lambda_0 \in \rho(T)$, so $R_{\lambda_0} = (T - \lambda_0 I)^{-1} \in \mathcal{L}(X \to X)$. Then for an arbitrary λ it holds

$$T - \lambda I = T - \lambda_0 I - (\lambda - \lambda_0) I = (T - \lambda_0 I)\,(I - (\lambda - \lambda_0) R_{\lambda_0}) \tag{6.15}$$

According to Lemma 6.3.7, the second factor is invertible if

$$\|(\lambda - \lambda_0) R_{\lambda_0}\| < 1,$$

so it's holds for any λ such that $|\,\lambda - \lambda_0\,| < \frac{1}{\|R_{\lambda_0}\|}$, hence

$$B\left(\lambda_0, \|R_{\lambda_0}\|^{-1}\right) \subseteq \rho(T),$$

hence $\rho(T)$ is open.

(III) Applying (6.11) in (6.15) we get (6.13):

$$R_\lambda = (T - \lambda I)^{-1} = (T - \lambda_0 I)^{-1}\,(I - (\lambda - \lambda_0) R_{\lambda_0})^{-1} =$$
$$= R_{\lambda_0} \sum_{k=0}^{\infty} ((\lambda - \lambda_0) R_{\lambda_0})^k = \sum_{k=1}^{\infty} (\lambda - \lambda_0)^{k-1} R_{\lambda_0}^k.$$

(II) We prove the resolvent identity. Let $\lambda, \nu \in \rho(T)$. Then

$$(T - \lambda I)(R_\lambda - R_\nu)(T - \nu I) = (T - \nu I) - (T - \lambda I) = (\lambda - \nu).$$

Multiply by R_λ on the left and by R_ν on the right to obtain the resolvent identity (6.12). That the map $R : \rho(T) \to \mathcal{L}(X \to X) : \lambda \mapsto R_\lambda$ it is holomorphic follows from the equation

$$\lim_{\nu \to \lambda} \frac{R_\nu - R_\lambda}{\nu - \lambda} = \lim_{\nu \to \lambda} R_\nu R_\lambda = R_\lambda^2.$$

Let $|\lambda| \geq \|T\|$. Then $T-\lambda I = -\lambda\left(I - \frac{1}{\lambda}T\right)$ and as $\left\|\frac{1}{\lambda}T\right\| = \frac{\|T\|}{|\lambda|} < 1$ then by Lemma 6.3.7 $I - \frac{1}{\lambda}T \in \mathcal{L}(X \to X) \Rightarrow \lambda \in \rho(T)$. Writing the series expansion of the inverse of $T - \lambda I = -\lambda(I - \frac{1}{\lambda}T)$ according to von Neumann's lemma, we immediately obtain (IV). □

6.3.10 Corollary. *Let X be a Banach space and $T \in \mathcal{L}(X \to X)$. Then the spectrum $\sigma(T)$ is a compact subset of the closed circle*

$$B(0, \|T\|) = \{\lambda \in \mathbb{C} \;:\; |\lambda| \leq \|T\|\}.$$

6.3.11 Lemma. *Let X be a Banach space and $T \in \mathcal{L}(X \to X)$. Then the spectrum $\sigma(T)$ is not empty.*

Proof. We shall deduce this result from Liouville's theorem in complex analysis. To this end, assume that $\sigma(T) = \varnothing$, hence $\rho(T) = \mathbb{C}$ and the resolvent $R : \lambda \mapsto R_\lambda$ is an *entire* function (i.e. analytic on the whole complex plane).

R_λ is also *bounded* function on $\mathbb{C}$ with $R_\lambda \to 0$ as $\lambda \to \infty$. Indeed, by Theorem 6.3.9 (IV), R_λ is a bounded in the annulus $\{|\lambda| \geq 2\|T\|\}$ and vanishes at infinity. Since R_λ is a continuous function by Theorem 6.3.9 (III) R_λ is also bounded in the disc $\{|\lambda| \leq 2\|T\|\}$.

By Liouville's theorem, $R_\lambda = 0$ everywhere.

Indeed, we fix a functional $f \in \mathcal{L}^*(X \to X)$ and apply the usual Liouville's theorem for the bounded entire function $f(R_\lambda)$. It follows that $f(R_\lambda)$ is constant, and since it must vanish at infinity it is zero everywhere. The last claim contradicts the fact that $f(R_\lambda)$ is an invertible operator. □

Summarizing our findings, we can state that the *spectrum of every bounded linear operator is a nonempty compact subset of* $\mathbb{C}$.

6.3.12 Definition. *The* ***spectral radius*** *of an operator $T \in \mathcal{L}(X \to X)$ is defined as*

$$r(T) := \max\{|\lambda| \;:\; \lambda \in \sigma(T)\}.$$

6.3.13 Theorem (Gelfand's formula). *For every operator $T \in \mathcal{L}(X\to X)$ acting on a Banach space X, one has*

$$r(T) = \lim_{n\to\infty} \|T^n\|^{1/n} = \inf_{n\in\mathbb{N}} \|T^n\|^{1/n}.$$

Exercise. Using 6.13 check that $r(T) \le \|T^n\|^{1/n} \le \|T\|$. Give an example where $r(T) < \|T\|$.

Remark. Gelfands formula gives an asymptotics for the growth of operator powers $\|T^n\| \sim r^n(T)$.

Spectrum of compact operators As compact operators are proxies of finite rank operators, one is able to fully classify their spectrum. First of all, for every $T \in \mathcal{K}(X\to X)$ one has

$$0 \in \sigma(T) \tag{6.16}$$

since T is not invertible by Corollary 6.1.7.

6.3.14 Exercise. Construct three examples of compact operators for which 0 is in the point, continuous, and residual spectrum respectively.

6.3.15 Theorem (Point spectrum of compact operators). *Let $T \in \mathcal{K}(X\to X)$ be a compact operator on a normed space X. For every $\varepsilon > 0$ there exists a finite number of linearly independent eigenvectors corresponding to eigenvalues λ_k with $|\,\lambda_k\,| > \varepsilon$.*

Consequently, the point spectrum $\sigma_p(T)$ is at most countable, and it lies in a sequence that converges to zero.

It also follows that each eigenvalue λ_k of T has finite multiplicity, i.e.

$$\dim\ker(T - \lambda_k I) < \infty.$$

Proof. Clearly, the second and third claims of the theorem follow from the first one. So, assume the contrary, that there exist $\varepsilon > 0$ and an infinite sequence of linearly independent vectors $(x_k)_{k=1}^{\infty}$ such that

$$Tx_k = \lambda_k x_k \quad \text{where} \quad |\,\lambda\,|_k > \varepsilon.$$

Consider the subspaces $E_n = \text{span}(x_k)_{k=1}^n$; then $T(E_k) \subseteq E_k$ for all k and $E_1 \subseteq E_2 \subseteq \dots$ is a sequence of proper inclusions. Therefore we can choose vectors

$$y_n \in E_n, \quad |\, y_n \,| = 1, \quad \text{dist}(y_n, E_{n-1}) \geq \frac{1}{2}$$

(Why? Think about E_n / E_{n-1}). We will show that the sequence $(Ty_n)_{n=1}^\infty$ contains no Cauchy subsequences, which will contradict compactness of T. To this end, we express y_n as a linear combination

$$y_n = a_n x_n + u_{n-1}, \quad \text{where } u_{n-1} \in E_{n-1}.$$

Then

$$Ty_n = \lambda_n a_n x_n + v_{n-1}, \quad \text{where } v_{n-1} = Tu_{n-1} \in E_{n-1}.$$

Now we are ready to estimate $\|Ty_n - Ty_m\|$ for $n > m$. Since $Ty_m \in E_m \subseteq E_{n-1}$, we obtain

$$\begin{aligned} \|Ty_n - Ty_m\| &= \|\lambda_n a_n x_n + w_{n-1}\|, \quad \text{where } w_{n-1} \in E_{n-1} \\ &= \|\lambda_n y_n + w_{n-1}\| \geq |\, \lambda_n \,| \, \text{dist}(y_n, E_{n-1}) \geq \frac{\varepsilon}{2}. \end{aligned}$$

It follows that $(Ty_n)_{n=1}^\infty$ contains no Cauchy subsequences as claimed. The proof is complete. $\square$

6.3.16 Corollary (Classification of spectrum of compact operators). *Let $T \in \mathcal{K}(X \to X)$ be a compact operator on a Banach space X. Then*

$$\sigma(T) = \sigma_p(T) \cup \{0\}.$$

Thus, the situation with the spectrum of the compact operator almost corresponds to the figure 10 a), the only possible difference is the type of spectrum at point 0 (see exercise 6.3.14).

Proof. As we already noticed in 6.16, $0 \in \sigma(T)$. Let now $\lambda \in \sigma(T)$, $\lambda \neq 0$. Fredholm alternative (Theorem 6.2.6) states that either $T - \lambda I$

is not injective (in which case $\lambda \in \sigma_p(T)$) or $T - \lambda I$ is both injective and surjective. In the latter case $T - \lambda I$ is invertible by the inverse mapping theorem, which means that $\lambda \notin \sigma(T)$. The proof is complete. $\square$

Spectrum of unitary operators

6.3.17 Definition. *Let H be a Hilbert space. An operator $U \in \mathcal{L}(H \to H)$ is called **unitary** if U*

$$U^*U = UU^* = I. \tag{6.17}$$

6.3.18 Proposition. *Let H be a Hilbert space and let $U \in \mathcal{L}(H \to H)$. Then the following are equivalent.*

(I) *U is an unitary operator.*

(II) *U is a bijective isometry on H (U is bijective and $\|Ux\| = \|x\|$ for all $x \in H$.)*

(III) *$U(H)$ is dense in H and U preserves the inner product $\langle Ux, Uy\rangle = \langle x, y\rangle$ for all $x, y \in H$).*

Proof. (I)$\Rightarrow$(II). $UU^* = I \Rightarrow U$ is surjective. $U^*U = I \Rightarrow U$ is injective, so (6.17) $\Rightarrow U$ is bijective. And $\|Ux\|^2 = \langle Ux, Ux\rangle = \langle U^*Ux, x\rangle = \langle x, x\rangle = \|x\|^2$.

(II)$\Rightarrow$(III) by the identity $\mathrm{Re}\langle x, y\rangle = \frac{1}{2}(\|x+y\|^2 - \|x\|^2 - \|y\|^2)$, and therefore $\mathrm{Im}\langle x, y\rangle = \mathrm{Re}(-i\langle x, y\rangle) = \mathrm{Re}\langle -ix, y\rangle = \frac{1}{2}(\|-ix+y\|^2 - \|x\|^2 - \|y\|^2)$ (see 1.11).

(III)$\Rightarrow$(I) Since U preserves the inner product then it preserves the norm too. Then U is an injection. The image of a complete space under an isometry is complete, therefore, it is closed. Then density of the image $U(H)$ actually means that $U(H) = H$, so U is surjective, as a result, it is a bijection, hence it has an inverse $U^{-1} : H \to H$.

From preserving the inner product we get $\langle x, y\rangle = \langle Ux, Uy\rangle = \langle U^*Ux, y\rangle$ for any $x, y \in H$. Due to the uniqueness of the representing vector for the functional $f(y) = \langle x, y\rangle$ (see Riesz representation theorem, 4.2.2) this means $U^*Ux = x$ for any x, so $U^*U = I$, then $U^{-1} = U^*UU^{-1} = U^*$, then $UU^* = UU^{-1} = I$ too. $\square$

6.3.19 Remark. In infinite-dimensional spaces (unlike finite-dimensional ones), not every isometric linear operator is a bijection, and therefore unitary. So the right shift operator R in l^2 is not a surjection, and the left shift operator L is not an injection (see 6.1.21), therefore, both of them are not unitary, $LR = I$ but $RL \neq I$.

Examples. The following operators are unitary.

1) Operators on $\mathbb{C}^n$ and $\mathbb{R}^n$ given by $n \times n$ unitary complex matrices and orthogonal real matrices; in particular rotations, symmetries, and permutations of coordinates in $\mathbb{C}^n$ and $\mathbb{R}^n$.

2) The operator $T(x_k)_{k=1}^{\infty} = (x_2, x_1, x_4, x_3, x_6, \dots)$ on l^2.

3) The operator $T(x)(t) = x(t)e^{i\varphi(t)}$, where φ is any measurable real-valued function, on $L^2(a, b)$, $-\infty \leq a < b \leq +\infty$.

6.3.20 Theorem (Spectrum of unitary operators)**.** *Let H be a Hilbert space. The spectrum of a unitary operator $U \in \mathcal{L}(H \to H)$ lies on the unit circle:*

$$\sigma(U) \subseteq \{\lambda \in \mathbb{C} \; : \; |\lambda| = 1\}.$$

Proof. The isometry property implies that $\|U\| = \left\|U^{-1}\right\| = 1$. Therefore the spectral radius is $r(U) \leq \|U\| = 1$ by Corollary 6.3.10. On the other hand, if $|\lambda| < 1$ then the operator $U^{-1}(U - \lambda I) = I - \lambda U^{-1}$ is invertible by von Neumann's Lemma 6.3.7 since $\left\|\lambda U^{-1}\right\| = |\lambda| < 1$. It follows that $U - \lambda I$ is invertible too. □

6.3.21 Exercise. Show that eigenvectors of a unitary operator U that correspond to distinct eigenvalues are orthogonal.

7 Self-adjoint operators on Hilbert space

Throughout this section, H will denote a Hilbert space,

7.1 Spectrum of self-adjoint operators

Let T be a bounded linear operator on a Hilbert space H, i.e. $T \in \mathcal{L}(H \to H)$. Recall from 6.1.15 that the adjoint operator $T^* \in \mathcal{L}(H \to H)$ is defined by $\langle T^* x, y \rangle = \langle x, Ty \rangle$ for $x, y \in H$.

7.1.1 Definition. *An operator $T \in \mathcal{L}(H \to H)$ is called **self-adjoint** if*

$$\langle Tx, y \rangle = \langle x, Ty \rangle, \quad x, y \in H. \tag{7.1}$$

Example. Examples of self-adjoint operators include:

(i) linear operators on $\mathbb{C}^n$ given by Hermitian matrices (a_{ij}), i.e. such that $a_{ij} = \overline{a_{ji}}$;

(ii) integral operators $Tx(t) = \int_0^1 k(s,t)x(s)\,ds$ on $L^2[0,1]$ with Hermitian symmetric kernels, i.e. such that $k(s,t) = \overline{k(t,s)}$;

(iii) orthogonal projections P on H. (Why?)

Every bounded linear operator can be decomposed into two self-adjoint operators:

7.1.2 Lemma. *Every operator $F \in \mathcal{L}(H \to H)$ can be uniquely represented as*

$$F = T + iS$$

where $T, S \in \mathcal{L}(H \to H)$ are self-adjoint operators.

Proof. The lemma holds with $T = \frac{1}{2}(F + F^*)$ and $S = \frac{1}{2i}(F - F^*)$. □

7.1.3 Exercise. Prove that the set of self-adjoint operators forms a closed linear subspace in $\mathcal{L}(H \to H)$.

The quadratic form and the norm of a self-adjoint operator

It is convenient to study self-adjoint operators $T \in \mathcal{L}(H \to H)$ through the quadratic form

$$Q(x) := Q_T(x) := \langle Tx, x \rangle, \quad x, x \in H.$$

One sees immediately that this quadratic form is real-valued, i.e. $Q(x) \in \mathbb{R}$ for all $x \in H$, this follows from the identity $\overline{\langle Tx, x \rangle} = \langle x, Tx \rangle = \langle Tx, x \rangle$.

Furthermore, the quadratic form Q determines the operator T uniquely. This follows from the following ***polarization identity***:

$$\langle Tx, y \rangle = \frac{1}{4}\Big(Q(x+y) - Q(x-y) + iQ(x+iy) - iQ(x-iy) \Big). \qquad (7.2)$$

To check 7.2, note that $Q(x+y) - Q(x-y) = \langle Tx + Ty, x + y \rangle - \langle Tx - Ty, x - y \rangle = 2\langle Tx, y \rangle + 2\langle Ty, x \rangle = 4 \operatorname{Re}\langle Tx, y \rangle$, so $\operatorname{Im}\langle Tx, y \rangle = \operatorname{Re}(-i\langle Tx, y \rangle) = 4 \operatorname{Re}\langle Tx, iy \rangle = \frac{1}{4}(Q(x+iy) - Q(x-iy))$. Then $\langle Tx, y \rangle = \operatorname{Re}\langle Tx, y \rangle + i \operatorname{Im}\langle Tx, y \rangle = \frac{1}{4}\Big(Q(x+y) - Q(x-y) + iQ(x+iy) - iQ(x-iy) \Big)$.

7.1.4 Remark. Taking $T = I$, we obtain an important special case of the polarization identity:

$$\langle x, y \rangle = \frac{1}{4}\left(\|x+y\|^2 - \|x-y\|^2 + i\,\|x+iy\|^2 - i\,\|x-iy\|^2 \right). \qquad (7.3)$$

The norm of T can be conveniently computed from the quadratic form Q_T:

7.1.5 Proposition (Norm of a self-adjoint operator)**.** *Let H be a Hilbert space, $S_H := \{x \in H \ : \ \|x\| = 1\}$. For every self-adjoint operator $T \in \mathcal{L}(H \to H)$, one has*

$$\|T\| = \sup_{x \in S_H} |\langle Tx, x \rangle|. \qquad (7.4)$$

Proof. The lower bound follows by definition of the operator norm:

$$\|T\| = \sup_{x\in S_H} \|Tx\| = \sup_{x,y\in S_H} |\langle Tx, y\rangle| \geq \sup_{x\in S_H} |\langle Tx, x\rangle| =: M.$$

It remains to show that the inequality here is actually the identity. To this end, we note that

$$\sup_{x,y\in S_H} |\langle Tx, y\rangle| = \sup_{x,y\in S_H} \operatorname{Re}\langle Tx, y\rangle.$$

and use the real part of polarization identity (7.2)

$$\begin{aligned}
\operatorname{Re}\langle Tx, y\rangle &= \frac{1}{4}\Big(\langle Tx + Ty, x + y\rangle - \langle Tx - Ty, x - y\rangle\Big) \\
&\leq \frac{1}{4}\Big(\|x + y\|^2 + \|x - y\|^2\Big) \quad \text{(by the definition of } M) \\
&\leq \frac{1}{2}\Big(\|x\|^2 + \|y\|^2\Big) \quad \text{(by the parallelogram law)} \\
&\leq M \quad (\text{as } \|x\| = \|y\| = 1).
\end{aligned}$$

This completes the proof. $\square$

Criterion of spectrum points of a self-adjoint operator

We would like to study the spectrum of self-adjoint operators $T \in \mathcal{L}(H\to H)$. An easy observation is that all eigenvalues of T must be real, that is

$$\sigma_p(T) \subset \mathbb{R}.$$

Indeed, if λ is an eigenvalue with an eigenvector x then $\langle Tx, x\rangle = \langle \lambda x, x\rangle = \lambda\langle x, x\rangle$ which must be the same as $\langle x, Tx\rangle = \langle x, \lambda x\rangle = \overline{\lambda}\langle x, x\rangle$, so $\lambda = \overline{\lambda} \in \mathbb{R}$.

We will soon prove that the whole spectrum of T is real, i.e. $\sigma(T) \subset \mathbb{R}$, and moreover we will compute the smallest interval containing $\sigma(T)$. Let us start with ruling out the residual spectrum:

7.1.6 Proposition (No residual spectrum). *Let $T \in \mathcal{L}(H \to H)$ be a self-adjoint operator. Then*

$$\sigma_r(T) = \varnothing.$$

Proof. Let $\lambda \in \sigma_r(T)$. This means that $\ker(T - \lambda I) = \mathbf{0}$ while $\operatorname{im}(T - \lambda I)$ is not dense in H. Since λ is not an eigenvalue, $\overline{\lambda}$ is not an eigenvalue either (recall that all eigenvalues of T must be real). Using this and the duality relation (6.1.24), we obtain that $(\ker(T - \lambda I))^{\perp} = \ker(T - \lambda I)^* = \ker(T - \overline{\lambda} I) = \mathbf{0}$ is dense in H, which is a contradiction. □

7.1.7 Proposition (A criterion of homeomorphism for self-adjoint operators). *Let $T \in \mathcal{L}(H \to H)$ is a homeomorphism if and only if T is bounded below, i.e. there exists $c > 0$ such that*

$$\|Tx\| \geq c \|x\| \quad \textit{for all } x \in H.$$

Proof. If T is invertible then T is bounded below with $c = \left\|T^{-1}\right\|^{-1}$. To prove the converse, assume that T is bounded below. Then by the criterion of isomorphic embedding (6.2.1), T is injective and $\operatorname{im} T$ is closed in H. On the other hand, since $0 \notin \sigma_r(T) = \varnothing$, injectivity of T implies that $\operatorname{im} T$ is dense in H. It follows that $\operatorname{im} T = H$. So T is injective and surjective, thus T is invertible by the inverse mapping theorem. □

Applying this result for the operator $T - \lambda I$, we immediately obtain

7.1.8 Corollary (Criterion of spectrum points). *Let $T \in \mathcal{L}(H \to H)$ be a self adjoint operator. Then $\lambda \in \sigma(T)$ if and only if the operator $T - \lambda I$ is not bounded below.*

7.1.9 Remark. A number $\lambda \in \sigma(T)$ for which $\lambda \in \sigma(T)$ is not bounded below is called an ***approximate eigenvalue*** of T, and the set of all approximate eigenvalues is called the ***approximate point spectrum***

of T. Corollary 7.1.8 states that for self-adjoint operators, the whole spectrum is the approximate point spectrum.

The reason for the name «approximate» is the following. If λ is an eigenvalue then $(T-\lambda I)x = 0$ for some $x \in S_H$ (i.e. $\|x\| = 1$). If λ is an approximate eigenvalue then $(T-\lambda I)x$ can be made *arbitrarily close* to zero for some $x \in S_H$. So, eigenvalues of T form the point spectrum $\sigma_p(T)$ while the approximate eigenvalues of T form the continuous spectrum $\sigma_c(T)$.

The spectrum interval

Now we compute the tightest interval that contains the spectrum of a self-adjoint operator T. This interval can be computed from the quadratic form of T.

7.1.10 Theorem. *Let $T \in \mathcal{L}(H \to H)$ be a self adjoint operator. Then*

(I) *The spectrum of T is real, and moreover $\sigma(T) \subseteq [m, M]$ where*

$$m = \inf_{x \in S_H} \langle Tx, x\rangle, \quad M = \sup_{x \in S_H} \langle Tx, x\rangle.$$

(II) *The endpoints $m, M \in \sigma(T)$.*

Proof. (I) Let $\lambda \in \mathbb{C} \setminus [m, M]$; since the interval is compact we have $d := \operatorname{dist}(\lambda, [m, M]) > 0$. Given $x \in S_H$, we use Cauchy–Schwarz to obtain the lower bound

$$\|(T-\lambda I)x\| \geq |\,\langle (T-\lambda I)x, x\rangle\,| = |\,\langle Tx, x\rangle - \lambda\,| \geq d,$$

where the last inequality follows because $\langle Tx, x\rangle \in [m, M]$ by definition. We have shown that $T - \lambda I$ is bounded below. By the criterion of spectrum points (Corollary 7.1.8), we conclude that $\lambda \notin \sigma(T)$.

(II) Let us show that $M \in \sigma(T)$; the claim for m can be proved similarly. Without loss of generality we can assume that $0 \leq m \leq M$ (this follows by a translation argument, namely by considering $T - mI$ instead of T. So $\|T\| = M$.

Let us choose a sequence of vectors $x_k \in S_H$ so that $\langle Tx_k, x_k \rangle \to M$. Then

$$\|(T - MI)x_k\|^2 = \langle (T - MI)x_k, (T - MI)x_k \rangle =$$
$$\|Tx_k\|^2 - 2M\langle Tx_k, x_k \rangle + M^2 \|x_k\|^2$$

Now $\|Tx_k\|^2 \leq \|T\|^2 = M^2$, $\langle Tx_k, x_k \rangle \to M$ and $\|x_k\|^2 = 1$. It follows that

$$\varlimsup_{k\to\infty} \|(T - MI)x_k\|^2 \leq M^2 - 2M + M^2 = 0$$

so $T - MI$ is not bounded below. Therefore $M \in \sigma(T)$. The proof is complete. □

As a consequence of this result, the spectral radius $r(T)$ of a self-adjoint operator equals $\|T\|$, so Gelfands formula is useless for self-adjoint operators:

7.1.11 Corollary. *Let $T \in \mathcal{L}(H \to H)$ be a self adjoint operator. Then*

$$r(T) = \max_{\lambda \in \sigma(T)} |\,\lambda\,| = \|T\|\,.$$

Proof. By the properties of the spectrum interval in Theorem 7.1.10, $r(T) = \max\{|\,m\,|\,, |\,M\,|\} = \|T\|$ as claimed. □

7.1.12 Exercise. Consider a self-adjoint operator $P \in \mathcal{L}(H \to H)$ such that $P^2 = P$. Prove that P is an orthogonal projection.

7.2 Spectral theorem for compact self-adjoint operators

Compact self-adjoint operators on a Hilbert space H are proxies of Hermitian matrices. As we know from linear algebra, every Hermitian matrix has diagonal form in some orthonormal basis of $\mathbb{C}^n$. Equivalently, such for such a matrix there exists an orthonormal basis of $\mathbb{C}^n$ consisting of the eigenvectors. In this section, we generalize this fact to infinite dimensions, for all compact self-adjoint operators on H.

First of all, let's establish the fact that any separable Hilbert spaces are isomorphic.

A digression: existence of orthogonal bases and isometry of all separate Hilbert spaces.

7.2.1 Lemma. *A Banach space X is separable if and only if it contains a linearly independent system of vectors $(x_k)_{k=1}^{\infty}$ whose linear span is dense in X, i.e.*

$$\overline{\operatorname{span}(x_k)_{k=1}^{\infty}} = X.$$

Proof. *Necessity.* If X is separable, it contains a system of vectors $(z_k)_{k=1}^{\infty}$ whose linear span is dense in X. We construct (x_k) inductively as a subset of (z_k). Namely, we include z_1, and if $z_{n+1} \notin \operatorname{span}(z_1, \ldots, z_n)$ for $n = 1, 2, \ldots$. By construction, (x_k) is linearly independent and $\operatorname{span}(x_1, \ldots, x_n) = \operatorname{span}(z_1, \ldots, z_n)$ for all $n = 1, 2, \ldots$ Letting $n \to \infty$ we conclude that $\overline{\operatorname{span}(x_k)} = \overline{\operatorname{span}(z_k)} = X$ as required.

Sufficiency. If $\operatorname{span}(x_k)_{k=1}^{\infty}$ is dense in X, so is the set of all finite linear combination $\sum_{k=1}^{n} a_k x_k$ with $a_k \in \mathbb{Q}$, which is a countable set. The lemma is proved. □

7.2.2 Theorem. *Every separable Hilbert space has an orthonormal basis.*

Proof. Let X be a separable Hilbert space. By Lemma 7.2.1, there is a linearly independent system of vectors (x_k) in X such that $\overline{\operatorname{span}(x_k)_{k=1}^{\infty}} = X$. Applying the Gram-Schmidt orthogonalization, we obtain an orthonormal system $(h_k)_{k=1}^{\infty}$ in X, such that $\operatorname{span}(x_k)_{k=1}^{n} = \operatorname{span}(h_k)_{k=1}^{n}$ for all $n \in \mathbb{N}$. Letting $n \to \infty$ we conclude that

$$\overline{\operatorname{span}(x_k)_{k=1}^{\infty}} = \overline{\operatorname{span}(h_k)_{k=1}^{\infty}} = X.$$

This proves the completeness of (h_k). Hence (h_k) is an orthonormal basis in X. □

7.2.3 Exercise (Haar system)**.** Consider the function

$$h(t) = \begin{cases} 1, & t \in [0, 1/2) \\ -1, & t \in [1/2, 1) \\ 0, & \text{otherwise.} \end{cases}$$

and define the functions

$$h_{k\,l}(t) = 2^{k/2} h(2^k t - l), \quad k, l \in \mathbb{Z}, \quad t \in \mathbb{R}.$$

Together with the characteristic function of the segment $[0, 1]$, this system of functions is called the *Haar system*; the individual functions is called the *Haar wavelets* and the function $h(t)$ is called the Haar *mother wavelet*. Show that, the Haar system $\mathcal{H}$ is an orthonormal basis in $L^2[0, 1]$. (Hint: first show that for any $x \in C[0, 1]$ it holds $x \in \mathrm{Cl}_{L^2[0,1]} \mathcal{H}$.)

We are ready to show that all separable Hilbert spaces have «the same geometry»:

7.2.4 Theorem. *All infinite-dimensional separable Hilbert spaces are isometric to each other. Precisely, for every such spaces H and H', one can find a linear bijective map $\Theta : H \to H'$ which preserves the inner product, i.e.*

$$\langle \Theta(x_1), \Theta(x_2) \rangle = \langle x_1, x_2 \rangle \quad \textit{for all } x_1, x_2 \in H. \tag{7.5}$$

Proof. Let $(e_k)_{k=1}^{\infty}$ and $(e'_k)_{k=1}^{\infty}$ be orthonormal bases of spaces H and H' respectively. Let Θ be the map that takes (e_k) to (e'_k). More precisely, define Θ by

$$\Theta\left(\sum_k a_k e_k \right) = \sum_k a_k e'_k.$$

By Parseval's identity (6.4),

$$\left\| \sum_k a_k e_k \right\|^2 = \left\| \sum_k a_k \right\|^2 = \left\| \sum_k a_k e'_k \right\|^2 \tag{7.6}$$

Therefore, Θ is well defined on X, its inverse is also well defined as

$$\Theta^{-1}\left(\sum_k a_k e'_k\right) = \sum_k a_k e_k,$$

so Θ is bijective and clearly linear. Additionally, (7.6) shows that $\|\Theta x\| = \|x\|$ for all $x \in H$. So Θ preserves the norm. Since by polarization formula (7.3), the inner product is uniquely determined by the norm, Θ must also preserve the inner product, i.e. (7.5) holds. This completes the proof. □

7.2.5 Proposition (Eigenvectors orthogonal)**.** *Let $T \in \mathcal{L}(H \to H)$ be a self adjoint operator. Then its eigenvectors corresponding to distinct eigenvalues are orthogonal.*

Proof. If $Tx_1 = \lambda_1 x_1$ and $Tx_2 = \lambda_2 x_2$ then $\lambda_1\langle x_1, x_2\rangle = \langle Tx_1, x_2\rangle = \langle x_1, Tx_2\rangle\langle x_1, Tx_2\rangle = \lambda_2\langle x_1, x_2\rangle$ (in the last identity we used that λ_2 is always real, so there is no conjugation). □

7.2.6 Definition. *A subspace H_0 of H is called an* ***invariant subspace*** *of T if $T(H_0) \subseteq H_0$.*

Example. Every eigenspace of T is invariant. More generally, the linear span of any subset of eigenvectors of T is an invariant subspace.

One of the most well known open problems in functional analysis is the *invariant subspace problem*. It asks whether every operator $T \in \mathcal{L}(H \to H)$ has a proper invariant subspace (i.e. different from $\{\mathbf{0}\}$ and H).

7.2.7 Proposition. *Let $T \in \mathcal{L}(H \to H)$ be self-adjoint. If $H_0 \subseteq H$ is an invariant subspace of T then ${H_0}^\perp$ is also an invariant subspace of T.*

Proof. Let $x \in {H_0}^\perp$; we need check that $Tx \in {H_0}^\perp$. So let us choose $y \in H_0$ arbitrarily. Then $\langle Tx, y\rangle = \langle Tx, y\rangle = 0$ since $x \in {H_0}^\perp$ and $y \in H_0$ so $Ty \in H_0$. □

7.2.8 Theorem (Spectral theorem for compact self-adjoint operators, Hilbert-Schmidt theorem)**.** *Let T be a compact self-adjoint linear operator on a separable Hilbert space H. Then there exists an orthonormal basis of H consisting of eigenvectors of H.*

Proof. Let us first prove that T has at least one eigenvector. By Corrolary 6.3.16, $\sigma(T) = \sigma_p(T) \cup \{\mathbf{0}\}$.

If $\sigma(T) \neq \{\mathbf{0}\}$ then $\sigma_p(T) \neq \varnothing$, so T has an eigenvector. If $\sigma(T) = \{\mathbf{0}\}$ then by Corollary 7.1.11 we have $\|T\| = r(T) = 0$, so $T = \mathbf{0}$ and every vector in H is an eigenvector of T.

We will complete the proof by induction. Consider the family of all orthonormal sets in H consisting of eigenvectors of T. All such sets are at most countable since H is separable. By Zorn's lemma, this family has a maximal element $(v_k)_{k=1}^{\infty}$ (why?). It remains to show that $H_0 := \operatorname{span}(v_k) = H$.

Suppose $H_0 \neq H$. Since H_0 is an invariant subspace of T (why?), ${H_0}^{\perp}$ is also an invariant subspace of T by Proposition 7.2.7. So we can use the first part of the proof for the restriction $T|_{{H_0}^{\perp}}$ which is a compact self-adjoint operator on ${H_0}^{\perp}$. It follows that $T|_{{H_0}^{\perp}}$ (and thus T itself) has an eigenvector in ${H_0}^{\perp}$. This contradicts the maximality of (v_k). $\square$

We illustrate Spectral Theorem 7.2.8 with a purely analytic consequence. The following result shows how one can separate variables of a general function $k(t,s)$. It is also due to Hilbert and Schmidt.

7.2.9 Theorem (Separation of variables)**.** *For any $k(t,s) \in L^2([0,1]^2)$ such that $k(s,t) \equiv \overline{k(t,s)}$ there exists an orthonormal basis (ϕ_n) of $L^2[0,1]$ and number sequence $\lambda_n \to 0$ such that*

$$k(t,s) = \sum_{n=1}^{\infty} \lambda_n \phi_n(t) \overline{\phi_n(s)}.$$

The convergence of this series is understood in $L^2([0,1]^2)$.

Proof. Consider the integral operator $(Tx)(t) = \int_0^1 k(t,s)x(s)ds$ on $L^2[0,1]$. Let (ϕ_n) be an orthonormal basis of its eigenvectors. Then the functions

$$\psi_{m,n}(t,s) = \phi_m(t)\overline{\phi_n(s)}, \quad m,n \in \mathbb{N}$$

form an orthonormal basis of $L^2([0,1]^2)$ (the orthonormality of this system follows from Fubini's theorem. In addition, also by virtue of Fubini's theorem, if $g \in L^2([0,1]^2)$ is orthogonal to each of the functions of this set, then this function g is the zero element in $L^2([0,1]^2)$).

Let us write the basis expansion of our function in $L^2([0,1]^2)$:

$$k = \sum_{m,n} \langle k, \psi_{m,n}\rangle \psi_{m,n}.$$

Now we compute the coefficients.

$$\begin{aligned}
\langle k, \psi_{m,n}\rangle &= \iint\limits_{[0,1]\times[0,1]} k(t,s)\overline{\phi_m(t)}\phi_n(s)dsdt \\
&= \int_0^1 \left(\int_0^1 k(t,s)\phi_n(s)ds\right)\overline{\phi_m(t)}dt \quad \text{(by Fubini theorem)} \\
&= \int_0^1 (T\phi_n)(t)\overline{\phi_m(t)}dt = \langle T\phi_n, \phi_m\rangle = \lambda_n\langle \phi_n, \phi_m\rangle \\
&= \lambda_n \delta_{n,m}.
\end{aligned}$$

Therefore $k = \sum_{m,n}\langle k, \psi_{m,n}\rangle\psi_{m,n} = \sum_{n=1}^{\infty} \lambda_n \phi_n(t)\overline{\phi_n(s)}$. □

7.3 Positive operators

7.3.1 Definition. *A self-adjoint operator $T \in \mathcal{L}(H \to H)$ is called* ***positive*** *if*

$$\langle Tx, x\rangle \geq 0 \quad \textit{for all } x \in H.$$

Positive operators are generalizations of non-negative numbers (which correspond to operators on one-dimensional space $\mathbb{C}$). Note that in linear algebra, positive operators are called *positive semidefinite.*

Exercises. Examples of positive operators include:

(i) T^2 for every self-adjoint $T \in \mathcal{L}(H \to H)$, as $\langle T^2 x, x \rangle = \langle Tx, Tx \rangle \geq 0$;

(ii) Hermitian matrices with non-negative eigenvalues;

(iii) More generally, compact self-adjoint operators on H with non-negative eigenvalues. (why?)

7.3.2 Definition (Partial order)**.** *For self-adjoint operators $S, T \in \mathcal{L}(H \to H)$, we shall say that $S \leq T$ if $T - S \geq 0$.*

This defines a partial order on $\mathcal{L}(H \to H)$.

Let us restate Theorem 7.1.10 on the spectrum interval in these new terms:

7.3.3 Theorem (Spectrum interval)**.** *Let $T \in \mathcal{L}(H \to H)$ be a self adjoint operator. Let m, M be the smallest and the largest numbers such that*

$$mI \leq T \leq MI$$

Then $\sigma(T) \subseteq [m, M]$ and $m, M \in \sigma(T)$.

As an immediate corollary, T is positive if and only if its spectrum is positive:

7.3.4 Corollary. *Let $T \in \mathcal{L}(H \to H)$ be a self adjoint operator. Then $T \geq 0$ if and only if $\sigma(T) \subseteq [0, +\infty)$.*

Polynomials of an operator

Working with polynomials is straightforward, and the result of this subsection remain valid for every bounded linear operator T on a general Banach space X.

7.3.5 Definition (Polynomials of an operator)**.** *Consider a polynomial $p(t) = a_0 + a_1 t + \ldots a_n t^n$ with complex coefficients. For an operator $T \in \mathcal{L}(H \to H)$, we define*

$$p(T) := a_0 I + a_1 T + \ldots a_n T^n.$$

If T is self-adjoint operator then $p(T)$ is also self-adjoint (why?). Moreover, for two polynomials f and g, one has

$$(af + bg)(T) = a \cdot f(T) + b \cdot g(T),$$
$$(fg)(T) = f(T) \cdot g(T)$$

7.3.6 Exercise. Let T be a self-adjoint linear operator on an n-dimensional Hilbert space. In an orthonormal basis of eigenvectors, T can be identified with the $n \times n$ diagonal matrix

$$T = \operatorname{diag}(\lambda_1, \dots, \lambda_n),$$

where λ_k are the eigenvalues of T. Check that for every polynomial $p(t)$ it holds

$$p(T) = \operatorname{diag}(p(\lambda_1), \dots, p(\lambda_n)).$$

Check that this statement can be generalized for all compact self-adjoint operators T on a general Hilbert space H.

7.3.7 Lemma (Invertibility)**.** *Let $p(t)$ be a polynomial and $T \in \mathcal{L}(H \to H)$. Then the operator $p(T)$ is invertible if and only if $p(t) \neq 0$ for all $t \in \sigma(T)$.*

Proof. Let us factorize $p(t) = a_n(t - t_1) \dots (t - t_n)$ where t_k are the roots of $p(t)$. Then $p(T) = a_n(T - t_1 I) \dots (T - t_n I)$. It's easy to see that $P(T)$ is invertible if and only if all factors $T - t_k I$ are invertible. This in turn is equivalent to the fact that all roots $t_k \notin \sigma(T)$. □

The spectrum of a polynomial $p(T)$ can be easily computed from the spectrum of T:

7.3.8 Theorem. *Let $p(t)$ be a polynomial and $T \in \mathcal{L}(H \to H)$. Then*

$$\sigma(p(T)) = p(\sigma(T)).$$

Proof. For every complex number λ, we have $\lambda \in \sigma(p(T))$ if and only if the operator $p(T) - \lambda I = (p - \lambda)T$ is not invertible. By the invertibility

Lemma 7.3.7, this is equivalent to the condition that $(p - \lambda)T = 0$ for some $t \in \sigma(T)$, which means that $p(t) = \lambda$ for some $t \in \sigma(T)$. The latter is equivalent to $\lambda \in p(\sigma(T))$. $\square$

Using the spectral mapping theorem, one can in particular easily compute the norms of operator polynomials:

7.3.9 Corollary. *Let $p(t)$ be a polynomial and $T \in \mathcal{L}(H \to H)$ be a self-adjoint operator. Then*

$$\|p(T)\| = \max_{t \in \sigma(T)} |\, p(t)\,| \,.$$

This result generalizes the identity $r(T) = \|T\|$ for the spectral radius of self-adjoint operators T proved in Corollary 7.1.11.

Proof. Let us apply Corollary 7.1.11 for the operator $p(T)$. Then spectral mapping theorem yields

$$\|p(T)\| = r(p(T)) = \max_{t \in \sigma(p(T))} |\, t\,| = \max_{t \in p(\sigma(T))} |\, t\,| = \max_{s \in \sigma(T)} |\, p(s)\,|$$

as claimed. $\square$