Functional Analysis
An Introduction

Henk de Snoo & Alef Sterk

Bernoulli Institute for Mathematics, Computer Science & Artificial Intelligence
University of Groningen

January 2023

Contents

Chapter 1

Linear spaces and linear operators

The building blocks of functional analysis are linear spaces and the linear maps between them. A short review of the definitions and properties of these notions is presented. Moreover, a number of examples of infinite-dimensional linear spaces is provided; these examples will play a major role in the development of the theory when a topological structure is added to these spaces. The last section is devoted to the infinite-dimensional case; it is shown how Zorn's lemma, which is really an axiom, is used to prove a number of useful properties. This last section can be skipped upon a first reading.

1.1 Linear spaces

Definition 1.1. Let X be a set and let $\mathbb{K} = \mathbb{R}$ or $\mathbb{C}$. Assume that X is provided with two operations: addition and scalar multiplication, i.e., maps from $X \times X$ to X and from $\mathbb{K} \times X$ to X, denoted by

$$(x,y) \mapsto x+y, \quad (\lambda,x) \mapsto \lambda x, \quad x,y \in X, \quad \lambda \in \mathbb{K},$$

respectively. Then X is said to be a *linear space* over $\mathbb{K}$ if for all $x,y,z \in X$ and $\lambda,\mu \in \mathbb{K}$ the following axioms are satisfied:

1. $x+y = y+x$;
2. $(x+y)+z = x+(y+z)$;
3. there exists an element $0 \in X$ such that $x+0 = x$;
4. there exists an element $-x \in X$ such that $x+(-x) = 0$;
5. $\lambda(\mu x) = (\lambda\mu)x$;
6. $1x = x$;
7. $\lambda(x+y) = \lambda x + \lambda y$;
8. $(\lambda+\mu)x = \lambda x + \mu x$.

A subset V of a linear space X is called a *linear subspace* when it is a linear space itself with the given operations.

Example 1.2. The set $\mathbb{K}^n$ is the collection of all finite sequences with n entries

$$x = (x_1, x_2, \ldots, x_n), \quad x_i \in \mathbb{K}.$$

Provide $\mathbb{K}^n$ with the usual pointwise addition and scalar multiplication:

$$x+y = (x_1+y_1, \ldots, x_n+y_n), \quad \lambda x = (\lambda x_1, \ldots, \lambda x_n).$$

Then it is clear that $\mathbb{K}^n$ with this structure is a linear space. □

Example 1.3. As an extension of $\mathbb{K}^n$ consider the set $\mathbb{K}^\infty$ which consists of all sequences of the form

$$x = (x_1, x_2, x_3, \dots), \quad x_i \in \mathbb{K}.$$

Provide the collection $\mathbb{K}^\infty$ with pointwise addition and scalar multiplication:

$$x + y = (x_1 + y_1, x_2 + y_2, \dots), \quad \lambda x = (\lambda x_1, \lambda x_2, \dots).$$

It is clear that $\mathbb{K}^\infty$ with this structure is a linear space. In Chapter 2 linear spaces will be provided with a topology. To that end, the following subsets of $\mathbb{K}^\infty$ will be of interest. For $p \geq 1$ the set ℓ^p is defined by

$$\ell^p = \left\{ x = (x_1, x_2, \dots) : x_i \in \mathbb{K}, \sum_{i=1}^{\infty} |x_i|^p < \infty \right\},$$

and for $p = \infty$ the set ℓ^∞ is defined by

$$\ell^\infty = \left\{ x = (x_1, x_2, \dots) : x_i \in \mathbb{K}, \sup_{i \in \mathbb{N}} |x_i| < \infty \right\}.$$

The sets ℓ^p with $1 \leq p \leq \infty$ are linear subspaces of $\mathbb{K}^\infty$. This is clear for $p = \infty$. For $1 \leq p < \infty$ the linearity of ℓ^p follows from $|x_i + y_i| \leq 2 \max(|x_i|, |y_i|)$, so that

$$|x_i + y_i|^p \leq 2^p \max(|x_i|^p, |y_i|^p) \leq 2^p (|x_i|^p + |y_i|^p).$$

The set of all convergent sequences is given by

$$c = \left\{ x = (x_1, x_2, \dots) : x_i \in \mathbb{K}, \lim_{i \to \infty} x_i \text{ exists} \right\},$$

and the set of all sequences converging to zero is given by

$$c_0 = \left\{ x = (x_1, x_2, \dots) : x_i \in \mathbb{K}, \lim_{i \to \infty} x_i = 0 \right\}.$$

Finally, let s be the set of all finitely supported sequences:

$$s = \left\{ x = (x_1, x_2, \dots) : \text{there exists } N_x \in \mathbb{N} \text{ such that } x_i = 0 \text{ for all } i \geq N_x \right\}.$$

Note that all these sets are linear subspaces of $\mathbb{K}^\infty$ and that for $1 \leq p \leq q < \infty$

$$s \subset \ell^1 \subset \ell^p \subset \ell^q \subset c_0 \subset c \subset \ell^\infty.$$

It will be clear from the context if these spaces are taken over $\mathbb{K} = \mathbb{C}$ or $\mathbb{K} = \mathbb{R}$. □

Example 1.4. Let S be a nonempty set and let X be a linear space. The set $\mathcal{F}(S, X)$ consisting of all functions $f : S \to X$ is provided with the usual pointwise addition and scalar multiplication:

$$(f + g)(s) = f(s) + g(s), \quad (\lambda f)(s) = \lambda f(s), \quad s \in S.$$

It is clear that $\mathcal{F}(S, X)$ with this structure is a linear space. In particular, $\mathcal{F}(S, \mathbb{K})$ is a linear space. Note that we have the special cases $\mathcal{F}(\{1, 2, \dots, n\}, \mathbb{K}) = \mathbb{K}^n$ and $\mathcal{F}(\mathbb{N}, \mathbb{K}) = \mathbb{K}^\infty$. Useful topologies can only be defined on certain subsets of $\mathcal{F}(S, \mathbb{K})$. A particularly important example is

$$\mathcal{B}(S, \mathbb{K}) = \left\{ f : S \to \mathbb{K} : \sup_{s \in S} |f(s)| < \infty \right\},$$

which is clearly a linear subspace of $\mathcal{F}(S, \mathbb{K})$. Note that we have the special case $\mathcal{B}(\mathbb{N}, \mathbb{K}) = \ell^\infty$. If S is a metric space, then $\mathcal{C}(S, \mathbb{K})$ denotes the set of continuous functions from S to $\mathbb{K}$ which is clearly a linear subspace of $\mathcal{F}(S, \mathbb{K})$. If S is a compact metric space, then $\mathcal{C}(S, \mathbb{K}) \subset \mathcal{B}(S, \mathbb{K})$. □

Definition 1.5. Let X be a linear space. The *sum* of two linear subspaces $V,W \subset X$ is defined as

$$V+W = \{x+y : x \in V,\, y \in W\}.$$

The sum is called *direct* if $V \cap W = \{0\}$.

It is clear that the sum $V+W$ is a linear subspace of X. Moreover, the sum is direct if and only if each element in $V+W$ can be written as unique sum of elements in V and W.

Note that an arbitrary intersection of linear subspaces is again a linear subspace. This leads to the following definition of a linear span.

Definition 1.6. Let X be a linear space and let $E \subset X$ be a set. The *linear span* of the set E is defined by

$$\operatorname{span}(E) = \bigcap \{H \subset X : E \subset H,\, H \text{ linear subspace}\}.$$

Proposition 1.7. The linear span of E is the unique linear subspace of X which contains E and is contained in every linear subspace which contains E. In fact,

$$\operatorname{span}(E) = \left\{ \sum_{i=1}^{n} \lambda_i e_i : n \in \mathbb{N}, \lambda_i \in \mathbb{K}, e_i \in E,\, i = 1,\dots,n \right\}. \tag{1.1}$$

Proof. Let Y be a linear subspace of X with $E \subset Y$, such that $Y \subset H$ for every linear subspace H with $E \subset H$. Then clearly $Y \subset \operatorname{span}(E)$. However by definition $\operatorname{span}(E) \subset Y$. Hence Y is uniquely determined and $Y = \operatorname{span}(E)$.

Denote the righthand side of (1.1) by F. Then F is a linear subspace of X and it contains E. Furthermore, if H is a linear subspace of X with $E \subset H$, then clearly $F \subset H$. Consequently, $F = \operatorname{span}(E)$. ☺

Definition 1.8. Let X be a linear space over $\mathbb{K}$. A nonempty finite set $M = \{e_1,\dots,e_n\} \subset X$ is called *linearly independent* if with $\lambda_1,\dots,\lambda_n \in \mathbb{K}$ one has

$$\sum_{i=1}^{n} \lambda_i e_i = 0 \quad \Rightarrow \quad \lambda_1 = \cdots = \lambda_n = 0.$$

A nonempty subset $M \subset X$ is called linearly independent if every finite subset of M is linearly independent. The *dimension* of X is defined by

$$\dim X = \begin{cases} 0 & \text{if } X = \{0\}, \\ n & \text{if } X \text{ is spanned by } n \text{ linearly independent vectors}, \\ \infty & \text{if } X \text{ has an infinite linearly independent subset.} \end{cases}$$

A set of n linearly independent vectors which span X is called a *basis* for X.

1.2 Linear operators

Recall that the Cartesian product of any two sets X and Y is given by

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

where (x,y) indicates an ordered pair of $x \in X$ and $y \in Y$.

Definition 1.9. A *relation* T from X to Y is a set $T \subset X \times Y$. If $Y = X$ one speaks of a relation on the set X. The *domain* and *range* of T are defined by

$$\begin{aligned} \operatorname{dom} T &= \{x \in X : (x,y) \in T \text{ for some } y \in Y\}, \\ \operatorname{ran} T &= \{y \in Y : (x,y) \in T \text{ for some } x \in X\}, \end{aligned}$$

respectively. A relation is called a *map* $T : X \to Y$ if it satisfies the following property:

$$(x,y) \in T \quad \text{and} \quad (x,z) \in T \quad \Rightarrow \quad y = z,$$

in which case one uses the notation $y = Tx$. A map $T : X \to Y$ is called
1. *injective* if for every $y \in Y$ there is at most one $x \in X$ with $y = Tx$;
2. *surjective* if for every $y \in Y$ there is at least one $x \in X$ with $y = Tx$;
3. *bijective* if for every $x \in X$ there is $y \in Y$ with $y = Tx$ and for every $y \in Y$ there is precisely one $x \in X$ with $y = Tx$.

Definition 1.10. Let X and Y be linear spaces over $\mathbb{K}$. A map $T : X \to Y$ is called *linear* if T is everywhere defined on X and if for all $x, y \in X$ and $\lambda \in \mathbb{K}$
1. $T(x+y) = Tx + Ty$;
2. $T(\lambda x) = \lambda(Tx)$.

The collection of all linear maps from X to Y is denoted by $L(X,Y)$.

Elements in $L(X,Y)$ are also referred to as *linear operators*. Clearly, $L(X,Y)$ is a linear subspace of $\mathcal{F}(X,Y)$.

Let X and Y be linear spaces over $\mathbb{K}$ and let $T : X \to Y$ be a linear map. Then clearly $\operatorname{dom} T$ is a linear subspace of X and $\operatorname{ran} T$ is a linear subspace of Y. Moreover the *kernel* or *null space* of T, defined by

$$\ker T = \{x \in X : Tx = 0\},$$

is a linear subspace of X. It is not difficult to see that T is injective if and only if $\ker T = \{0\}$, and that T is surjective if and only if $\operatorname{ran} T = Y$;

Lemma 1.11. Let X and Y be linear spaces over $\mathbb{K}$ and let $T : X \to Y$ be a linear map. Then T is bijective if and only if there exists a unique linear map $S : Y \to X$ such that $ST = I_X$ and $TS = I_Y$.

Proof. First assume that T is bijective. If $y \in Y$ there is precisely one $x \in X$ with $y = Tx$. Define $S : Y \to X$ by $Sy = x$. Then S is a linear map with $TSy = y$. Furthermore, if $x \in X$ and $y = Tx$ then $STx = Sy = x$.

Conversely, assume that there exists a linear map $S : Y \to X$ such that $ST = I_X$ and $TS = I_Y$. Then it is clear that T is bijective.

It is clear that S is uniquely determined. ☺

Definition 1.12. Let X be a linear space and let $P : X \to X$ be a linear map. Then P is called a *projection* if $P^2 = P$.

Lemma 1.13. A linear map $P : X \to X$ is a projection if and only if $I - P$ is a projection. In this case:

$$\operatorname{ran} P = \ker (I - P), \quad \ker P = \operatorname{ran} (I - P).$$

Moreover, $X = \operatorname{ran} P + \ker P$ is a direct sum.

Proof. The first statement follows from the identity

$$(I-P)^2 = I - 2P + P^2.$$

For the next statements it suffices to only check that $\operatorname{ran} P = \ker(I-P)$ as $\operatorname{ran}(I-P) = \ker P$ then follows. If $x \in \operatorname{ran} P$, then $x = Py$. Hence $Px = P^2 y = Py = x$, i.e., $(I-P)x = 0$ or $x \in \ker(I-P)$. Conversely, if $x \in \ker(I-P)$. Then $(I-P)x = 0$ or $x = Px \in \operatorname{ran} P$.

To see that $\operatorname{ran} P \cap \ker P = \{0\}$, assume that $x \in \operatorname{ran} P \cap \ker P$. Then $x = Pu$ for some $u \in X$, while $0 = Px = P^2 u = Pu$. It follows that $x = 0$. ☺

Definition 1.14. Let X be a linear space and let $V, W \subset X$ be linear subspaces. Then V and W are called *complementary* if

$$X = V + W, \quad \text{direct sum.}$$

The subspaces induce corresponding linear projections denoted by P_V and P_W with $P_V + P_W = I$.

It is clear that a linear subspace $V \subset X$ has a complementary subspace $W \subset X$ if and only if there exists a linear projection P_V such that $\operatorname{ran} P_V = V$. It will be shown that for every linear subspace in X there exists such a linear projection; see Corollary 1.43.

Definition 1.15. Let X be a linear space over $\mathbb{K}$ and let $T : X \to X$ be a linear map. The *point spectrum* $\sigma_{\mathrm{p}}(T)$ of T is the set of all eigenvalues of T:

$$\sigma_{\mathrm{p}}(T) = \{\lambda \in \mathbb{K} : Tx = \lambda x \text{ for some } x \neq 0\}.$$

The *geometric multiplicity* of $\lambda \in \sigma_{\mathrm{p}}(T)$ is the dimension of the corresponding eigenspace $\ker(T - \lambda)$.

Theorem 1.16. Let X be a linear space and let $T : X \to X$ be a linear map. Eigenvectors corresponding to different eigenvalues are linearly independent.

Proof. Assume that $x_1, \ldots, x_n$ are eigenvectors corresponding to eigenvalues $\lambda_1, \ldots, \lambda_n$ and that all eigenvalues are different. Let x_m, $1 < m \leq n$, be the first vector which can be expressed as

$$x_m = \mu_1 x_1 + \cdots + \mu_{m-1} x_{m-1},$$

with the elements $x_1, \ldots, x_{m-1}$ being linearly independent. Then apply the operator $T - \lambda_m$ to both sides, which leads to

$$\begin{aligned} 0 = (T - \lambda_m) x_m &= [T - \lambda_m](\mu_1 x_1 + \cdots + \mu_{m-1} x_{m-1}) \\ &= \mu_1(\lambda_1 - \lambda_m) x_1 + \cdots + \mu_{m-1}(\lambda_{m-1} - \lambda_m) x_{m-1}. \end{aligned}$$

Since $x_1, \ldots, x_{m-1}$ are linearly independent while $\lambda_1 - \lambda_m, \ldots \lambda_{m-1} - \lambda_m$ are all nonzero, it follows that $\mu_1 = \cdots = \mu_{m-1} = 0$, which is a contradiction since x_m is nontrivial. ☺

Linear spaces and the linear maps between them are of great interest when the spaces are provided with a topological structure; see Chapters 2–8.

1.3 Quotient spaces of linear spaces

The following definition and theorem are of a general nature. They are included for the convenience of the reader; they will be frequently used in the context of linear spaces.

Definition 1.17. A relation R on a set X is called an *equivalence relation* if
1. for each $x \in X$ one has $(x,x) \in R$ (reflexivity);
2. if $(x,y) \in R$, then $(y,x) \in R$ (symmetry);
3. if $(x,y) \in R$ and $(y,z) \in R$, then $(x,z) \in R$ (transitivity).

The statement $(x,y) \in R$ is denoted by $x \sim y$ and the equivalence relation is denoted by $\sim$. For $x \in X$ the *equivalence class* $[x]$ of x is defined as

$$[x] = \{y \in X \, : \, x \sim y\}.$$

The set of all equivalence classes in X is denoted by $X/\sim$ and the map $\pi : X \to X/\sim$ given by $\pi(x) = [x]$ is called the *quotient map*.

Thus in terms of the notation $\sim$ the above conditions read as
1. for each $x \in X$ one has $x \sim x$;
2. if $x \sim y$, then $y \sim x$;
3. if $x \sim y$ and $y \sim z$, then $x \sim z$.

The notation $\sim$ may be a little easier in practice.

Theorem 1.18. Let X be a set with an equivalence relation $\sim$. Let $x, y \in X$, then one has the following statements:
1. $x \in [x]$;
2. $[x] = [y] \Leftrightarrow x \sim y$;
3. $[x] \cap [y] \neq \emptyset \Rightarrow [x] = [y]$;
4. $X = \bigcup_{x \in X} [x]$, the disjoint union of all equivalence classes.

Proof. (1) This follows directly from reflexivity.

(2) ($\Rightarrow$) Let $[x] = [y]$. Then $y \in [y] = [x]$, so that $x \sim y$. ($\Leftarrow$) Let $x \sim y$ and $z \in [x]$. Then $y \sim x$ and $x \sim z$ so that $y \sim z$. Hence $z \in [y]$, which leads to $[x] \subset [y]$. The other inclusion is proved similarly.

(3) Let $z \in [x] \cap [y]$. Then $x \sim z$ and $y \sim z$, so that $x \sim y$ and hence $[x] = [y]$ by the equivalence in (2).

(4) Let $x \in X$. Then $x \in [x]$ and hence $X = \bigcup_{x \in X} [x]$. It follows from (2) and (3) that these equivalence classes are either equal or disjoint. ☺

Let X be a set with an equivalence relation $\sim$. Then according to Theorem 1.18 the set X can be written as the disjoint union of all equivalence classes. In fact, if X is any set with a *partition*, i.e., a disjoint class of nonempty subsets of X whose union gives all of X, then one may define $x \sim y$ when x and y belong to the same subset in the partition. It is straightforward to see that $\sim$ is an equivalence relation on X which produces the given partition.

Example 1.19. Let $X = \mathbb{R}$ and define $x \sim y$ if $x - y \in \mathbb{Z}$. Show that $\sim$ is an equivalence relation. Moreover one sees that for any $x \in \mathbb{R}$ the equivalence class $[x]$ is given by

$$[x] = \{y \in \mathbb{R} \, : \, x \sim y\} = \{x + n \, : \, n \in \mathbb{Z}\}.$$

It is clear that $\mathbb{R} = \bigcup_{x \in [0,1)} [x]$. □

Definition 1.20. Let X be a linear space and let $V \subset X$ be a linear subspace. Then V induces an equivalence relation on X by

$$x \sim y \quad \Leftrightarrow \quad y - x \in V.$$

The equivalence class to which $x \in X$ belongs is denoted by $x+V$:

$$x+V = \{y \in X \,:\, y-x \in V\}.$$

The set of equivalence classes is denoted by X/V.

Proposition 1.21. The space X/V provided with the following sum and scalar multiplication:
1. $(x+V)+(y+V) = x+y+V$, $x,y \in X$;
2. $\lambda(x+V) = \lambda x+V$; $x \in X$, $\lambda \in \mathbb{K}$,

is a linear space over $\mathbb{K}$.

Proof. Assume that $x+V = x'+V$ and $y+V = y'+V$. Then $x-x' \in V$ and $y-y' \in V$, so that $x+y-(x'+y') \in V$, which implies

$$x'+y'+V = x+y+V$$

Hence the sum is well-defined, i.e., it does not depend on the chosen representatives. Likewise, scalar multiplication is well-defined. Next observe that

$$(x+V)+(0+V) = x+V,$$

while, similarly one sees that

$$((x+V)+(y+V))+(z+V) = (x+y+V)+(z+V) = (x+y)+z+V,$$

and

$$(x+V)+((y+V)+(z+V)) = (x+V)+(y+z+V) = x+(y+z)+V.$$

Hence associativity in X/V follows from associativity in X. By checking similar identities it follows that X/V is a linear space where $0+V = V$ is the zero element of X/V. ☺

Definition 1.22. Let X be a linear space and let $V \subset X$ be a linear subspace. The *quotient map* $\pi : X \to X/V$ is defined by

$$\pi(x) = x+V, \quad x \in X.$$

Lemma 1.23. The map π is linear, surjective, and $\ker \pi = V$.

Proof. The map π is linear as follows by direct verification. It is clear that $V \subset \ker \pi$. Conversely, if $\pi(x) = 0$ then $x+V = 0$ or equivalently $x \in V$. Hence $\ker \pi \subset V$. Moreover, it is trivial to see that π is surjective. ☺

1.4 Isomorphisms between linear spaces

Theorem 1.24. Let X and Y be linear spaces and let $V \subset X$ be a linear subspace.
1. Let $T : X \to Y$ be a linear map such that $V \subset \ker T$. Then T induces a well-defined linear map
$$\widehat{T} : X/V \to Y, \quad x+V \mapsto T(x),$$
such that $T = \widehat{T} \circ \pi$.
2. Let $S : X/V \to Y$ be a linear map. Then $T = S \circ \pi : X \to Y$ is a linear map with $V \subset \ker T$.

Proof. (1) To see that $\widehat{T}$ is well-defined, suppose that $x+V = x'+V$ in X/V. Then $x-x' \in V \subset \ker T$ so that $T(x-x') = 0$ and $Tx = Tx'$. This shows that $\widehat{T}$ is well defined. It is trivial to see that $T = \widehat{T} \circ \pi$.

(2) Assume that $S : X/V \to Y$ is a linear map. Clearly $T = S \circ \pi : X \to Y$ is a linear map. For $x \in V$ one has $Tx = S(x+V) = 0$, which shows that $V \subset \ker T$. ☺

Corollary 1.25. Let X and Y be linear spaces and let $T : X \to Y$ be a linear map. Then

$$\widehat{T} : X/\ker T \to Y, \quad x + \ker T \mapsto T(x),$$

is an injective linear map, so that $X/\ker T$ isomorphic to $\operatorname{ran} T$. If in addition T is surjective, $\widehat{T} : X/\ker T \to Y$ is an isomorphism of linear spaces.

Proof. Apply Theorem 1.24 with $V = \ker T$, so that $\widehat{T}$ is a well-defined linear map. To see that $\widehat{T}$ is injective, let $x+V \in \ker \widehat{T}$. Then $0 = \widehat{T}(x+V) = Tx$, so that $x \in \ker T = V$. This proves that $\widehat{T}$ is injective. Furthermore, if T is surjective, it follows from $\operatorname{ran} \widehat{T} = \operatorname{ran} T$ that $\widehat{T}$ is surjective. ☺

Theorem 1.26. Let X be a linear space with $V \subset X$ a linear subspace. If $\dim X < \infty$, then $\dim X/V < \infty$ and

$$\dim X/V = \dim X - \dim V.$$

Proof. Let $n = \dim X$ and $k = \dim V$. Let $\{e_1, \dots, e_k\}$ be a basis for V and extend this basis with vectors $e_{k+1}, \dots, e_n$ to a basis $\{e_1, \dots, e_n\}$ for X. Define the map $T : X \to X$ by

$$\lambda_1 e_1 + \cdots + \lambda_n e_n \mapsto \lambda_{k+1} e_{k+1} + \cdots + \lambda_n e_n.$$

Clearly, T is a linear map with $\ker T = V$ and $\operatorname{ran} T = \operatorname{span}\{e_{k+1}, \dots, e_n\}$. Hence X/V and $\operatorname{ran} T$ are isomorphic, so in particular $\dim X/V = \dim \operatorname{ran} T = n - k$. ☺

Corollary 1.27. Let $T : X \to Y$ be a linear map with $\dim X < \infty$. Then

$$\dim \ker T + \dim \operatorname{ran} T = \dim X.$$

1.5 Dual spaces of linear spaces

The collection of all linear $\mathbb{K}$-valued maps on a linear space X is called the dual space of X. This concept is important in analysis as it used to describe for instance distributions, measures, and Hilbert spaces. Chapter 7 returns to this concept when the spaces and the maps are provided with a topological structure.

Definition 1.28. Let X be a linear space over $\mathbb{K}$. The *dual space* of X is defined as $X' = L(X, \mathbb{K})$. The elements of X' called *functionals* on X.

Lemma 1.29. Let X be a finite-dimensional linear space. Then X' is a finite-dimensional linear space and $\dim X' = \dim X$.

Proof. Let $\dim X = n$ and let $\{x_1, \dots, x_n\}$ be a basis for X. Then define the linear maps $f_j : X \to \mathbb{K}$ by $f_j(x) = \lambda_j$ whenever $x = \sum_{i=1}^n \lambda_i x_i$, so that

$$f_j(x_i) = \delta_{ij}.$$

Then the functionals $\{f_1,\dots,f_n\}$ are linearly independent in X'. To see this assume

$$\sum_{j=1}^{n} \mu_j f_j = 0,$$

which implies that

$$\mu_i = \sum_{j=1}^{n} \mu_j \delta_{ij} = \sum_{j=1}^{n} \mu_j f_j(x_i) = 0, \quad i = 1,\dots,n.$$

In order to show that $\dim X' = n$ let $f \in X'$. Define $\mu_i = f(x_i)$ and observe that

$$\sum_{j=1}^{n} \mu_j f_j(x_i) = \mu_i = f(x_i), \quad i = 1,\dots,n.$$

Hence f and $\sum_{j=1}^{n} \mu_j f_j$ coincide on the elements of the basis, which leads to $f = \sum_{j=1}^{n} \mu_j f_j$. Therefore, $\dim X' = n$. ☺

Definition 1.30. Let X be a linear space over $\mathbb{K}$. The *second-dual space* of X is defined as $X'' = L(X', \mathbb{K})$. The *natural map* $J : X \to X''$ is given by

$$J(x)(f) = f(x), \quad x \in X, \quad f \in X'. \tag{1.2}$$

Indeed, the map J is linear. To see this, note that for $x, y \in X$

$$J(x+y)(f) = f(x+y) = f(x) + f(y) = J(x)(f) + J(y)(f), \quad f \in X'.$$

This shows $J(x+y) = J(x) + J(y)$. Similarly, it follows that $J(\lambda x) = \lambda J(x)$ for all $x \in X$ and $\lambda \in \mathbb{K}$.

The natural map $J : X \to X''$ has an interesting property. If the space X is finite-dimensional, then J is a bijection between X and X''; see Lemma 1.31. Moreover, in general the map J is injective, so that X is embedded in X''; see Corollary 1.44.

Lemma 1.31. Let X be a finite-dimensional linear space. Then $J : X \to X''$ is a bijection.

Proof. First of all observe that $\dim X'' = \dim X' = \dim X$ by Lemma 1.29. Hence it suffices to show that J is injective; cf. Corollary 1.27. Let $\{x_1,\dots,x_n\}$ be a basis for X and let $\{f_1,\dots,f_n\}$ be the dual basis in X', defined by $f_j(x_i) = \delta_{ij}$. Let $x \in X$ and assume that $J(x) = 0$. According to (1.2) this means that $f(x) = 0$ for all $f \in X'$. Write x in terms of a basis $x = \sum_{i=1}^{n} \lambda_i x_i$, and choose for $f = f_j$, $1 \le j \le n$. Then $\lambda_j = 0$ for $1 \le j \le n$ and, hence, $x = 0$. ☺

Definition 1.32. Let X and Y be linear spaces and let $T : X \to Y$ be a linear operator. Then the *conjugate operator* $T^\times : Y' \to X'$ is the linear operator defined by

$$(T^\times f)(x) = f(Tx), \quad f \in Y', \quad x \in X.$$

Indeed, the conjugate operator $T^\times$ in Definition 1.32 is linear. To see this, just observe for example that

$$(T^\times(f+g))(x) = (f+g)(Tx) = f(Tx) + g(Tx) = (T^\times f)(x) + (T^\times f)(x),$$

for all $f, g \in Y'$ and $x \in X$.

In finite-dimensional spaces one can always write linear operators in terms of matrices. Here is a brief review of frequently used facts.

Let X and Y be linear spaces with $\dim X = n$ and $\dim Y = m$ with bases $\{x_1, \dots, x_n\}$ and $\{y_1, \dots, y_m\}$, respectively. Let $T : X \to Y$ be a linear map and let the matrix (t_{ji}) be defined by

$$Tx_i = \sum_{j=1}^{m} t_{ji} y_j, \quad i = 1, \dots, n. \tag{1.3}$$

Let $x \in X$ and $Tx \in Y$ be expressed as

$$x = \sum_{i=1}^{n} \lambda_i x_i, \quad Tx = \sum_{j=1}^{m} \mu_j y_j. \tag{1.4}$$

It follows from the representation of x in (1.4) that

$$Tx = \sum_{i=1}^{n} \lambda_i T x_i.$$

Substitution of the representation of Tx_i as in (1.3) leads to

$$Tx = \sum_{i=1}^{n} \lambda_i \left(\sum_{j=1}^{m} t_{ji} y_j \right) = \sum_{j=1}^{m} \left(\sum_{i=1}^{n} t_{ji} \lambda_i \right) y_j.$$

Comparing with the representation of Tx in (1.4) shows that the coefficients in (1.4) are connected by the matrix transformation

$$\mu_j = \sum_{i=1}^{n} t_{ji} \lambda_i, \quad j = 1, \dots, m, \tag{1.5}$$

In particular, when $X = Y$, $\zeta \in \mathbb{K}$ is an eigenvalue of T with corresponding eigenvector $x = \sum_{i=1}^{n} \lambda_i x_i$ if and only if

$$\begin{pmatrix} t_{11} & \dots & t_{1n} \\ \vdots & & \vdots \\ t_{n1} & \dots & t_{nn} \end{pmatrix} \begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_n \end{pmatrix} = \zeta \begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_n \end{pmatrix}$$

has a nontrivial solution or, equivalently, $\det(t_{ij} - \zeta) = 0$.

For the conjugate operator $T^\times$ there is a similar observation. Let X' and Y' have the dual bases $\{f_1, \dots, f_n\}$ and $\{g_1, \dots, g_m\}$, respectively. Let $g \in Y'$ and $T^\times g \in X'$ be expressed as

$$g = \sum_{k=1}^{m} \beta_k g_k, \quad T^\times g = \sum_{\ell=1}^{n} \gamma_\ell f_\ell. \tag{1.6}$$

Apply $T^\times g$ to $x = \sum_{i=1}^{n} \lambda_i x_i$ as in (1.4). Then the duality of the bases gives

$$T^\times g(x) = \left(\sum_{\ell=1}^{n} \gamma_\ell f_\ell \right) \left(\sum_{i=1}^{n} \lambda_i x_i \right) = \sum_{i=1}^{n} \gamma_i \lambda_i.$$

Apply g to $Tx = \sum_{j=1}^{m} \mu_j y_j$ as in (1.4). Then the duality of the bases gives

$$g(Tx) = \left(\sum_{k=1}^{m} \beta_k g_k \right) \left(\sum_{j=1}^{m} \mu_j y_j \right) = \sum_{j=1}^{m} \beta_j \mu_j.$$

Now use (1.5), then

$$g(Tx) = \sum_{j=1}^{m} \beta_j \left(\sum_{i=1}^{n} t_{ji} \lambda_i \right) = \sum_{i=1}^{n} \left(\sum_{j=1}^{m} t_{ji} \beta_j \right) \lambda_i.$$

Recall that $(T^\times g)(x) = g(Tx)$, $x \in X$. Hence the coefficients in (1.6) are connected by the transpose matrix transformation

$$\gamma_i = \sum_{j=1}^{m} t_{ji} \beta_j, \quad i = 1, \dots, n. \tag{1.7}$$

1.6 Infinite-dimensional linear spaces

In a finite-dimensional linear space a number of results is easy to prove, such as the existence of a basis, the extension of a linear operator to the whole space if it is only defined on a subspace, and the existence of sufficiently many functionals. In order to prove such facts in an infinite-dimensional linear space one needs Zorn's lemma. This section is devoted to that lemma and its consequences. It may be skipped upon first reading. Only one of the results in this section, the Hahn-Banach theorem for linear spaces, will be used later, namely in Chapter 7.

Definition 1.33. A relation R on a set X is called a *partial order* if
1. for each $x \in X$ one has $(x,x) \in R$ (reflexivity);
2. if $(x,y) \in R$ and $(y,x) \in R$, then $x = y$ (anti-symmetry);
3. if $(x,y) \in R$ and $(y,z) \in R$, then $(x,z) \in R$ (transitivity).

The statement $(x,y) \in R$ will be written as $x \preceq y$ and the partial order will be indicated by $\preceq$. A set X with a partial order will be called a partially ordered set.

If for any pair $x,y \in X$ one has either $x \preceq y$ or $y \preceq x$ then $\preceq$ is called a *total order*. A set X with a total order will be called a totally ordered set.

Thus in terms of the notation $\preceq$ the above definitions read as
1. for each $x \in X$ one has $x \preceq x$;
2. if $x \preceq y$ and $y \preceq x$, then $x = y$;
3. if $x \preceq y$ and $y \preceq z$, then $x \preceq z$.

This notation emphasizes the ordering between the various elements.

Example 1.34. Let X be a set and let $P(X)$ be its power set, the collection of all subsets of X. Then for $A,B \in P(X)$ one may define $A \preceq B$ if $A \subset B$. It follows that $\preceq$ is a partial order on $P(X)$ which is not a total order. $\square$

Example 1.35. Let X be a set and let P be the class of all functions from X to $\mathbb{R}$. Define $f \preceq g$ if $f(x) \leq g(x)$ for all $x \in X$. Then P provided with $\preceq$ is a partially ordered set, which is not totally ordered. $\square$

Example 1.36. Let $X = \mathbb{R}$ and define $x \preceq y$ if $x \leq y$. Then X provided with $\preceq$ is a partially ordered set, which is in fact totally ordered. $\square$

If $V \subset X$ and $\preceq$ is a partial (total) order on X, then the restriction to V gives a partial (total) order on V.

Definition 1.37. Let X be a partially ordered set, and let $V \subset X$ be a subset. An element $y \in X$ is called:
1. an *upper bound* for V if $x \preceq y$ for all $x \in V$;
2. a *maximal element* of X if $y \preceq x$ implies that $y = x$.

The following statement, Zorn's lemma, is actually an axiom. It is equivalent to the axiom of choice.

Lemma 1.38 (Zorn's lemma)**.** Let X be a nonempty set with a partial order. If every totally ordered subset of X has an upper bound in X, then X contains a maximal element.

Definition 1.39. Let X be a nontrivial linear space. A subset $V \subset X$ is called a *Hamel basis* of X if V is linearly independent and $\operatorname{span}(V) = X$.

Theorem 1.40. Every nontrivial linear space has a Hamel basis.

Proof. Let P be the class of linearly independent subsets of X and let it be partially ordered by set inclusion; cf. Example 1.34. Observe that the class P is not empty as the subset consisting of $x \in X$, $x \neq 0$, is in P.

Let Q be any totally ordered subset of P. The union of all subsets of X which belong to Q is in P and is an upper bound for Q. Hence the conditions of Zorn's lemma are satisfied and P has a maximal element, say V.

The maximal element V of P is a Hamel basis for X. To see this observe that V is linearly independent since it belongs to P. Moreover $\mathrm{span}\,(V) = X$, for if there exists $x \in X$ and $x \notin \mathrm{span}\,(V)$ then the set $V \cup \{x\}$ is linearly independent and contains V as a proper subset, contradicting that V is maximal. ☺

Let X and Y be linear spaces, let $V \subset X$ be a proper linear subspace of X, and let $f : V \to Y$ be a linear operator. The interest is in extending f to a linear operator whose domain contains V. The following simple construction of a one-dimensional extension is helpful. Let $x_0 \in X$ be such that $x_0 \notin V$ and let the linear subspace V_0 be spanned by V and x_0: $V_0 = \mathrm{span}\,\{x_0, V\}$. Define the map $F_0 : V_0 \to Y$ by

$$F_0(x + \lambda x_0) = f(x) + \lambda y_0, \quad x \in V, \quad \lambda \in \mathbb{K}, \tag{1.8}$$

where y_0 is a fixed vector in Y. Note that the map F_0 is linear as, for instance,

$$\begin{aligned} F_0((x + \lambda x_0) + (y + \mu x_0)) &= F_0(x + y + (\lambda + \mu)x_0) \\ &= f(x + y) + (\lambda + \mu)y_0 \\ &= f(x) + \lambda y_0 + f(y) + \mu y_0 \\ &= F_0(x + \lambda x_0) + F_0(y + \mu x_0). \end{aligned}$$

Clearly the linear operator F_0, defined on $V_0 \supset V$, properly extends f. This one-dimensional extension plays a role in the following theorem which shows that there is always an extension of f defined on all of X, via Zorn's lemma. The one-dimensional extension in (1.8) will also appear in the proof of Theorem 1.45, where a more special assertion can be found.

Theorem 1.41. Let X and Y be linear spaces and let $V \subset X$ be a proper linear subspace of X. Let $f : V \to Y$ be a linear operator. Then there exists a linear operator $F : X \to Y$ defined on all of X which extends f.

Proof. This theorem will be proved with Zorn's lemma. In order to apply Zorn's lemma, let P be the class of all linear operators $g : X \to Y$ with

$$V \subset \mathrm{dom}\, g,$$

where V is a proper subset of $\mathrm{dom}\, g$ and g is an extension of f. For linear operators $g, h \in P$ define $g \preceq h$ if h is an extension of g, i.e., $\mathrm{dom}\, g \subset \mathrm{dom}\, h$ and $g(x) = h(x)$ for all $x \in \mathrm{dom}\, g$. Then the set P with $\preceq$ is partially ordered.

The class P is nonempty. To see this, note that V is assumed to be a proper subset of X. Hence choose $x_0 \in X \setminus V$ and define the linear operator F_0 as in (1.8) for some $y_0 \in Y$. Since F_0 is a linear operator which properly extends f, it follows that $F_0 \in P$.

Now let Q be any totally ordered subset of P. It will be shown that there exists an upper bound $G \in P$, thus that $g \preceq G$ for all $g \in Q$. Define the set $\mathrm{dom}\, G$ by

$$\mathrm{dom}\, G = \bigcup_{g \in Q} \mathrm{dom}\, g.$$

Then $\operatorname{dom} G$ is a linear subspace of X. To see this, let $x_1, x_2 \in \operatorname{dom} G$. Then there exist $g_1, g_2 \in Q$ with

$$x_1 \in \operatorname{dom} g_1, \quad x_2 \in \operatorname{dom} g_2.$$

Since Q is totally ordered one may assume that $g_1 \preceq g_2$, so that $\operatorname{dom} g_1 \subset \operatorname{dom} g_2$. Therefore it follows that $x_1 + x_2 \in \operatorname{dom} g_2 \subset \operatorname{dom} G$. Similarly it can be shown that $\operatorname{dom} G$ is closed under multiplication by elements from $\mathbb{K}$.

Let $x \in \operatorname{dom} G$, then $x \in \operatorname{dom} g$ for some $g \in Q$, and define $G(x) = g(x)$. This definition does not depend on the particular choice of $g \in Q$. If $x \in \operatorname{dom} g$ and $x \in \operatorname{dom} g'$ for $g, g' \in Q$. Since Q is totally ordered it follows that $g \preceq g'$ or $g' \preceq g$ and in each case $g(x) = g'(x)$. The same reasoning that was used to show that $\operatorname{dom} G$ is a linear subspace can be used to show that the operator G is linear:

$$\begin{aligned} G(x_1) + G(x_2) &= g_1(x_1) + g_2(x_2) = g_2(x_1) + g_2(x_2) \\ &= g_2(x_1 + x_2) = G(x_1 + x_2), \quad x_1, x_2 \in \operatorname{dom} G, \end{aligned}$$

when $g_1 \preceq g_2$. Note that $G \in P$ is an upper bound of Q. Hence the conditions of Zorn's lemma are satisfied and P has a maximal element, say F.

Note that by definition of the set P the maximal element F is a linear operator. Assume that $\operatorname{dom} F \neq X$. Then choose $x_0 \in X \setminus \operatorname{dom} F$ and the procedure as in (1.8) shows that F allows a linear extension, contradicting the maximality of F. Therefore $\operatorname{dom} F = X$ and the theorem has been proved. ☺

Corollary 1.42. Let X be a linear space and let $V \subset X$ be a proper linear subspace of X. Let $x_0 \in X$ such that $x_0 \notin V$. Then there exists an element $f \in X'$ such that $f(x_0) = 1$ and $f(x) = 0$ for all $x \in V$.

Proof. Define the linear map f_0 on the space V_0 spanned by V and x_0 as follows:

$$f_0(x + \lambda x_0) = \lambda, \quad x \in V, \quad \lambda \in \mathbb{K}.$$

By Theorem 1.41 with $Y = \mathbb{K}$ there exists an element $f \in X'$ extending f_0. Hence f has the desired properties. ☺

Corollary 1.43. Let X be a linear space and let $V \subset X$ be a linear subspace. Then there exists a linear subspace $W \subset X$ such that V and W are complementary subspaces.

Proof. If $V = \{0\}$ or $V = X$ the situation is clear. Now assume that $V \subset X$ is a proper subspace of X. It suffices to show a projection P in X such that $\operatorname{ran} P = V$; cf. Lemma 1.13. Let $f(x) = x$ be defined on V, so that f is a linear operator from X to V with $\operatorname{dom} f = V$. Let $F : X \to V \subset X$ be a linear extension of f defined on all of X; cf. Theorem 1.41. Clearly $\operatorname{ran} F = V$, and note that

$$F^2(x) = F(F(x)) = f(F(x)) = F(x), \quad x \in V,$$

since $F(x) \in V$ for all $x \in V$. Thus F is a projection in X with $\operatorname{ran} F = V$. ☺

Corollary 1.44. Let X be a linear space. Then the natural map $J : X \to X''$ is injective.

Proof. Let $x \in X$ and recall that

$$J(x)(f) = f(x), \quad x \in X, \quad f \in X'.$$

Assume that $J(x) = 0$ for some $x \in X$. By definition this means that $f(x) = 0$ for all $f \in X'$. This implies $x = 0$, because otherwise there exists some $f \in X'$ such that $f(x) = 1$; cf. Corollary 1.42. ☺

Theorem 1.45 (Hahn-Banach theorem for linear spaces)**.** Let X be a linear space over $\mathbb{K}$ and let $p : X \to \mathbb{R}$ be a map such that

1. $p(x+y) \leq p(x) + p(y)$, $x, y \in X$;
2. $p(\lambda x) = |\lambda|\, p(x)$, $x \in X$, $\lambda \in \mathbb{K}$.

Assume that $V \subset X$ is a proper linear subspace and that $f : V \to \mathbb{K}$ is a linear map with

$$|f(x)| \leq p(x), \quad x \in V.$$

Then there exists a linear functional $F : X \to \mathbb{K}$, extending f to all of X, such that

$$|F(x)| \leq p(x), \quad x \in X.$$

Proof. The proof consists of two steps. In Step 1 the case $\mathbb{K} = \mathbb{R}$ is treated, similar to the proof of Theorem 1.41. In Step 2 the case $\mathbb{K} = \mathbb{C}$ will be proved by a reduction to Step 1.

Step 1. Assume that X is a linear space over $\mathbb{K} = \mathbb{R}$. As a first result it will be shown that there is a one-dimensional extension of f which is dominated by p. Since V is a proper linear subspace of X one may choose $x_0 \in X \setminus V$. Then the linear operator F_0, defined on $V_0 = \operatorname{span}\{x_0, V\} \supset V$ by

$$F_0(x + \lambda x_0) = f(x) + \lambda y_0, \quad x \in V, \quad \lambda \in \mathbb{R}, \tag{1.9}$$

properly extends f for any $y_0 \in \mathbb{R}$; see the discussion around (1.8). Next it will be shown that $y_0 \in \mathbb{R}$ in (1.9) can be chosen such that the inequality

$$|F_0(x + \lambda x_0)| = |f(x) + \lambda y_0| \leq p(x + \lambda x_0), \quad x \in V, \quad \lambda \in \mathbb{R}, \tag{1.10}$$

is satisfied. Observe that the conditions (1) and (2) lead to

$$\begin{aligned} f(x) - f(y) = f(x-y) &\leq p(x-y) \\ &= p(x + x_0 - y - x_0) \leq p(x + x_0) + p(y + x_0), \quad x, y \in V, \end{aligned}$$

or, equivalently, to

$$p(x + x_0) - f(x) \geq -p(y + x_0) - f(y), \quad x, y \in V.$$

It follows from this last inequality that y_0 defined by

$$y_0 = \inf\{\, p(x + x_0) - f(x) : x \in V\} \tag{1.11}$$

satisfies

$$y_0 \geq -p(y + x_0) - f(y), \quad y \in V, \tag{1.12}$$

and, in particular, $y_0 > -\infty$. Now use y_0 in (1.11) in the definition (1.9) of the linear extension F_0 of f. Then the definition (1.9) together with (1.11) and (1.12) lead to

$$\begin{aligned} F_0(x + x_0) &= f(x) + y_0 \leq p(x + x_0), \quad x \in V, \\ F_0(-y - x_0) &= -f(y) - y_0 \leq p(y + x_0), \quad y \in V. \end{aligned} \tag{1.13}$$

This implies that

$$F_0(x + \lambda x_0) \leq p(x + \lambda x_0), \quad x \in V, \quad \lambda \in \mathbb{R}. \tag{1.14}$$

This is clear for $\lambda = 0$, while for $\lambda > 0$ this follows from (1.13) and

$$F_0(x + \lambda x_0) = \lambda F_0(x/\lambda + x_0) \leq \lambda p(x/\lambda + x_0) = p(x + \lambda x_0),$$

and for $\lambda < 0$ this follows from (1.13) and

$$F_0(x + \lambda x_0) = -\lambda F_0(-x/\lambda - x_0) \leq -\lambda p(-x/\lambda - x_0) = p(x + \lambda x_0).$$

Finally note that (1.14) and condition (2) imply that

$$\begin{aligned} -F_0(x+\lambda x_0) &= F_0(-x-\lambda x_0) \\ &\leq p(-x-\lambda x_0) = p(x+\lambda x_0), \quad x \in V, \quad \lambda \in \mathbb{R}. \end{aligned} \tag{1.15}$$

Hence, due to (1.14) and (1.15), the linear extension F_0 in (1.9) satisfies (1.10).

In order to prove the assertion of the theorem for the case $\mathbb{K} = \mathbb{R}$, let P be the class of all linear $\mathbb{R}$-valued maps g whose domain $\operatorname{dom} g$ satisfies

$$V \subset \operatorname{dom} g,$$

where V is a proper subset of $\operatorname{dom} g$, and which have the property that

$$g(x) = f(x), \quad x \in V, \quad \text{and} \quad |g(x)| \leq p(x), \quad x \in \operatorname{dom} g.$$

For linear maps $g, h \in P$ define $g \preceq h$ when h extends g. Then the set P with $\preceq$ is partially ordered.

The class P is nonempty. Recall that if $x_0 \notin V$ then according to the one-dimensional extension procedure above there exists a linear operator $F_0 : V_0 \to \mathbb{R}$ which belongs to P.

Now let Q be any totally ordered subset of P. It will be shown that there exists an upper bound $G \in P$, thus that $g \preceq G$ for all $g \in Q$. Define the linear map G by

$$\operatorname{dom} G = \bigcup_{g \in Q} \operatorname{dom} g, \quad G(x) = g(x), \quad x \in \operatorname{dom} g.$$

Since Q is totally ordered, this definition is correct; see also the similar reasoning in the proof of Theorem 1.41. Note that $G \in P$ is an upper bound of Q. Hence the conditions of Zorn's lemma are satisfied and P has a maximal element, say F.

Note that by definition of the set P the maximal element F is a linear operator. Assume that $\operatorname{dom} F \neq X$. Then choose $x_0 \in X \setminus \operatorname{dom} F$ and the one-dimensional extension procedure above shows there exists a proper linear extension in P to $\operatorname{span}\{\operatorname{dom} F, x_0\}$, contradicting the maximality of F. Therefore $\operatorname{dom} F = X$ and the theorem has been proved for the case $\mathbb{K} = \mathbb{R}$.

Step 2. Assume that X is a linear space over $\mathbb{K} = \mathbb{C}$. Then the linear operator $f : V \to \mathbb{C}$ can be written as

$$f(x) = \operatorname{Re} f(x) + i \operatorname{Im} f(x), \quad x \in V,$$

or, since $\operatorname{Im} f(x) = -\operatorname{Re} f(ix)$, as

$$f(x) = \operatorname{Re} f(x) - i \operatorname{Re} f(ix), \quad x \in V.$$

By definition the operator f is $\mathbb{C}$-linear and satisfies $|f(x)| \leq p(x)$, $x \in V$. Note that $\operatorname{Re} f : V \to \mathbb{R}$ is an $\mathbb{R}$-linear operator on V, which satisfies

$$|\operatorname{Re} f(x)| \leq |f(x)| \leq p(x), \quad x \in V.$$

Since the space V is closed under addition and scalar multiplication by elements in $\mathbb{C}$ it may be viewed as a linear space over $\mathbb{R}$ and one may apply Step 1 to the $\mathbb{R}$-linear operator $\operatorname{Re} f$. Hence there exists an $\mathbb{R}$-linear extension $\Phi : X \to \mathbb{R}$ of $\operatorname{Re} f$ such that $|\Phi(x)| \leq p(x)$ for all $x \in X$. Now define $F : X \to \mathbb{C}$ by

$$F(x) = \Phi(x) - i\Phi(ix), \quad x \in X.$$

Then it is easily seen that F is a $\mathbb{C}$-linear operator on X and that F extends f, as

$$F(x) = \operatorname{Re} f(x) - i \operatorname{Re} f(ix) = f(x), \quad x \in V.$$

Finally note that for any $x \in X$

$$F(x) = |F(x)|e^{i\theta},$$

with some $\theta \in \mathbb{R}$, so that, using condition (2),

$$|F(x)| = F(e^{-i\theta}x) = \operatorname{Re} F(e^{-i\theta}x) = \Phi(e^{-i\theta}x) \le p(e^{-i\theta}x) = p(x).$$

Thus F is a linear extension of f, defined on X, which satisfies $|F(x)| \le p(x)$, $x \in X$. Therefore the theorem has been proved for the case $\mathbb{K} = \mathbb{C}$. ☺

Chapter 2

Normed linear spaces and inner product spaces

The notion of a norm on a linear space turns the linear space into a metric space whose distance is produced by the norm. The definition of norm respects in some sense the notion of linearity. This singles out the normed linear spaces from all metric spaces. A number of examples of normed linear spaces is provided. An important result says that the unit ball in a normed linear space is compact if and only if that space is finite-dimensional. When the linear space is given the structure of an inner product it turns automatically into a normed linear space. It is interesting and useful that the orthogonal complement of a linear subspace of an inner product space can be described in terms of the corresponding norm. For the sake of reference the Gram-Schmidt orthogonalisation procedure is included and a corresponding best approximation property is explained.

2.1 Linear spaces with a norm

Definition 2.1. Let X be a linear space over $\mathbb{K} = \mathbb{R}$ or $\mathbb{K} = \mathbb{C}$. The map $\|\cdot\| : X \to \mathbb{R}$ is called a *norm* if for all $x, y \in X$ and $\lambda \in \mathbb{K}$ the following axioms are satisfied:

1. $\|x\| \geq 0$;
2. $\|x\| = 0 \Leftrightarrow x = 0$;
3. $\|\lambda x\| = |\lambda| \|x\|$;
4. $\|x+y\| \leq \|x\| + \|y\|$.

The pair $(X, \|\cdot\|)$ is called a normed linear space over $\mathbb{K}$. By abuse of language X itself will often be called a normed linear space. If in (2) only the implication $(\Leftarrow)$ holds, then $\|\cdot\|$ is called a *semi-norm* and X is called a semi-normed linear space over $\mathbb{K}$.

Remark 2.2. The following observations are useful.

1. If $(X, \|\cdot\|)$ is a normed linear space, then clearly $d(x,y) = \|x-y\|$ defines a metric on X.
2. If $V \subset X$ is a linear subspace of a normed linear space $(X, \|\cdot\|)$, then the restriction $(V, \|\cdot\|)$ is a normed linear space.
3. If $(X, \|\cdot\|)$ is a semi-normed linear space, then it is clear that the set $N = \{x \in X : \|x\| = 0\}$ is a linear subspace of X.

$\square$

In order to prove that certain nonnegative maps on special linear spaces give rise to a norm one needs Young's inequality. First make the trivial observation that for $1 < p, q < \infty$ one has

$$\frac{1}{p} + \frac{1}{q} = 1 \quad \Leftrightarrow \quad (p-1)(q-1) = 1.$$

Lemma 2.3 (Young's inequality)**.** Let $1/p+1/q=1$ with $1<p<\infty$. If $a,b\geq 0$, then

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

Proof. For every strictly increasing function $f:[0,\infty)\to\mathbb{R}$ with $f(0)=0$ it follows that

$$ab \leq \int_0^a f(x)\,dx + \int_0^b f^{-1}(y)\,dy,$$

where equality holds if and only if $f(a)=b$. In particular, Young's inequality follows by taking $f(x)=x^{p-1}$ so that $f^{-1}(y)=y^{1/(p-1)}=y^{q-1}$. ☺

Lemma 2.4 (Hölder's and Minkowski's inequality)**.** Let $x,y\in\mathbb{K}^n$ and let $1/p+1/q=1$. For $1<p<\infty$ Hölder's inequality is valid:

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

Moreover, for $1\leq p<\infty$ Minkowski's inequality is valid:

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

Proof. For $1<p<\infty$ an application of Lemma 2.3 with

$$a=\frac{|x_i|}{\left(\sum_{i=1}^n |x_i|^p\right)^{1/p}},\quad b=\frac{|y_i|}{\left(\sum_{i=1}^n |y_i|^q\right)^{1/q}},$$

leads to Hölder's inequality.

Minkowski's inequality is clear for $p=1$. To see it for $1<p<\infty$ observe that

$$|x_i+y_i|^p = |x_i+y_i||x_i+y_i|^{p-1} \leq |x_i||x_i+y_i|^{p-1}+|y_i||x_i+y_i|^{p-1},$$

which leads to

$$\sum_{i=1}^n |x_i+y_i|^p \leq \sum_{i=1}^n |x_i||x_i+y_i|^{p-1}+\sum_{i=1}^n |y_i||x_i+y_i|^{p-1};$$

Apply Hölder's inequality to each of the sums in the right-hand side; note that $q(p-1)=p$. ☺

Corollary 2.5 (Cauchy-Schwarz inequality on $\mathbb{K}^n$)**.** For $x,y\in\mathbb{K}^n$ the following inequality holds:

$$\sum_{i=1}^n |x_i y_i| \leq \sqrt{\sum_{i=1}^n |x_i|^2}\sqrt{\sum_{i=1}^n |y_i|^2}.$$

Example 2.6. For $n\in\mathbb{N}$ and $1\leq p\leq\infty$ define $\|\cdot\|:\mathbb{K}^n\to\mathbb{R}$ by

$$\|x\| = \|(x_1,\dots,x_n)\| = \begin{cases} \left(\sum_{i=1}^n |x_i|^p\right)^{1/p}, & \text{if } 1\leq p<\infty,\\ \max(|x_1|,\dots,|x_n|), & \text{if } p=\infty,\end{cases}$$

when $x\in\mathbb{K}^n$. Then the space $\mathbb{K}^n$ is a normed linear space over $\mathbb{K}$. This is clear for $p=\infty$ and for $1\leq p<\infty$ it follows from the Minkowski inequality in Lemma 2.4. □

Example 2.7. Let X be a finite-dimensional linear space over $\mathbb{K}$ with basis $\{e_1,\dots,e_n\}$. For any $x \in X$ of the form $x = \lambda_1 e_1 + \cdots + \lambda_n e_n$ define

$$\|x\|_+ = \begin{cases} \left(\sum_{i=1}^n |\lambda_i|^p\right)^{1/p}, & \text{if } 1 \leq p < \infty, \\ \max(|\lambda_1|,\dots,|\lambda_n|), & \text{if } p = \infty. \end{cases}$$

Then X is a normed linear space; see Example 2.6. □

Recall the definition of the linear spaces ℓ^p with $1 \leq p \leq \infty$:

$$\ell^p = \left\{ x = (x_1, x_2, \dots) : x_i \in \mathbb{K}, \ \sum_{i=1}^\infty |x_i|^p < \infty \right\}, \quad 1 \leq p < \infty,$$

$$\ell^\infty = \left\{ x = (x_1, x_2, \dots) : x_i \in \mathbb{K}, \ \sup_{i \in \mathbb{N}} |x_i| < \infty \right\},$$

cf. Example 1.3. The following lemma provides an extension of the inequalities of Hölder and Minkowski to the setting of the space ℓ^p.

Lemma 2.8 (Hölder's and Minkowski's inequality)**.** Let $x \in \ell^p$, $y \in \ell^q$, and let $1/p + 1/q = 1$. For $1 < p < \infty$ Hölder's inequality is valid:

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

Moreover, for $1 \leq p < \infty$ Minkowski's inequality is valid:

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

Proof. Assume that $1 < p < \infty$ and let $z \in \ell^p$. Then for any $n \in \mathbb{N}$

$$\left(\sum_{i=1}^n |z_i|^p \right)^{1/p} \leq \left(\sum_{i=1}^\infty |z_i|^p \right)^{1/p}.$$

Hence for $x, y \in \ell^p$ and any $n \in \mathbb{N}$ the Hölder inequality in Lemma 2.4 shows that

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

Taking the supremum for $n \in \mathbb{N}$ leads to the desired result. In a similar way the Minkowski inequality in Lemma 2.4 gives the present Minkowski inequality. ☺

Example 2.9. For $1 \leq p \leq \infty$ define $\|\cdot\| : \ell^p \to \mathbb{R}$ by

$$\|x\|_p = \begin{cases} \left(\sum_{i=1}^\infty |x_i|^p\right)^{1/p}, & \text{if } 1 \leq p < \infty, \\ \sup_{i \in \mathbb{N}} |x_i|, & \text{if } p = \infty, \end{cases}$$

when $x \in \ell^p$. Then the space ℓ^p is a normed linear space over $\mathbb{K}$. This is clear for $p = \infty$ and for $1 \leq p < \infty$ it follows from the Minkowski inequality in Lemma 2.8. The space c of all convergent sequences x_i is a linear subspace of ℓ^∞. It is a normed linear space with the norm inherited from the normed linear space ℓ^∞. The space c_0 of all convergent sequences x_i which converge to 0 is a linear subspace of ℓ^∞. It is a normed linear space with the norm inherited from the normed linear space ℓ^∞. Likewise, the space s of all finitely supported sequences is a linear subspace of ℓ^∞. It is a normed linear space with the norm inherited from the normed linear space ℓ^∞. Recall that $\ell^p \subset \ell^\infty$ and one has

$$\|x\|_\infty \leq \|x\|_p, \quad x \in \ell^p.$$

□

Example 2.10. Let S be a nonempty set. The linear space $\mathcal{B}(S,\mathbb{K})$ has a natural norm given by

$$\|f\|_\infty = \sup_{s\in S} |f(s)|, \quad f \in \mathcal{B}(S,\mathbb{K}).$$

If S is a compact metric space, then $\mathcal{C}(S,\mathbb{K}) \subset \mathcal{B}(S,\mathbb{K})$, which implies that $\mathcal{C}(S,\mathbb{K})$ becomes a normed linear space by inheriting there norm from $\mathcal{B}(S,\mathbb{K})$. In particular, the space $\mathcal{C}([a,b],\mathbb{K})$ is a normed linear space when provided with the sup-norm. □

Example 2.11. Let S be a nonempty set, and let $(X,\|\cdot\|)$ be a normed linear space. The set

$$\mathcal{B}(S,X) = \left\{ f \in \mathcal{F}(S,X) \,:\, \sup_{s\in S} \|f(s)\| < \infty \right\}$$

is clearly a linear subspace of $\mathcal{F}(S,X)$, and

$$\|f\|_\infty = \sup_{s\in S} \|f(s)\|, \quad f \in \mathcal{B}(S,X),$$

provides a norm on $\mathcal{B}(S,X)$; see also Example 3.11. □

Example 2.12. For functions $f,g \in \mathcal{C}([a,b],\mathbb{K})$ there is the Hölder inequality with $1 < p < \infty$:

$$\int_a^b |f(s)g(s)|\,ds \leq \left(\int_a^b |f(s)|^p\,ds \right)^{1/p} \left(\int_a^b |g(s)|^q\,ds \right)^{1/q},$$

(use Young's inequality), and the Minkowski inequality with $1 \leq p < \infty$:

$$\left(\int_a^b |f(s)+g(s)|^p\,ds \right)^{1/p} \leq \left(\int_a^b |f(s)|^p\,ds \right)^{1/p} + \left(\int_a^b |g(s)|^p\,ds \right)^{1/p},$$

cf. Example 2.6. Here the integrals are understood in the sense of Riemann. For $1 \leq p < \infty$ define

$$\|f\|_p = \left(\int_a^b |f(t)|^p\,dt \right)^{1/p}, \quad f \in \mathcal{C}([a,b],\mathbb{K}),$$

then $\|\cdot\|_p$ is a norm on $\mathcal{C}([a,b],\mathbb{K})$, the so-called L^p-norm. Clearly, the arguments in this example remain valid for classes of more general Riemann integrable functions; for instance for functions which are continuous on $[a,b]$ except for a jump at $a < c < b$. □

2.2 Properties of norms

Since a normed linear space $(X,\|\cdot\|)$ is a metric space with $d(x,y) = \|x-y\|$, convergence of sequences in $(X,\|\cdot\|)$ is understood in the sense of the metric space.

Definition 2.13. Let X be a normed linear space. A sequence (x_n) in X is said to *converge* to $x \in X$ if $\|x_n - x\| \to 0$. In this case we also write $x_n \to x$ in X.

Clearly the limit of a convergent sequence is uniquely determined. This convergence in the norm of the space is also referred to as *strong convergence*.

Proposition 2.14. Let $(X, \|\cdot\|)$ be a normed linear space. Then

$$|\|x\| - \|y\|| \leq \|x - y\|, \quad x, y \in X. \tag{2.1}$$

Assume that $x_n \to x$, $y_n \to y$ in X, and $\lambda_n \to \lambda$ in $\mathbb{K}$. Then
1. $\|x_n\| \to \|x\|$ and, in particular, $\|x_n\|$ is bounded;
2. $x_n + y_n \to x + y$;
3. $\lambda_n x_n \to \lambda x$.

Proof. It is clear that for all $x, y \in X$

$$\|x\| = \|x - y + y\| \leq \|x - y\| + \|y\|, \quad \|y\| = \|y - x + x\| \leq \|x - y\| + \|x\|,$$

which leads to (2.1). Hence if $x_n \to x$ and $y_n \to y$ then

$$|\|x_n\| - \|x\|| \leq \|x_n - x\| \to 0,$$

and

$$\|x_n + y_n - (x + y)\| \leq \|x_n - x\| + \|y_n - y\| \to 0.$$

Moreover if in addition $\lambda_n \to \lambda$, then

$$\begin{aligned}\|\lambda_n x_n - \lambda x\| &\leq \|\lambda_n x_n - \lambda x_n\| + \|\lambda x_n - \lambda x\| \\ &= |\lambda_n - \lambda| \|x_n\| + |\lambda| \|x_n - x\| \\ &\leq |\lambda_n - \lambda| M + |\lambda| \|x_n - x\| \to 0;\end{aligned}$$

recall that the convergent sequence $\|x_n\|$ is bounded. ☺

Assume that the linear space X has two norms $\|\cdot\|$ and $\|\cdot\|_+$, which satisfy

$$\|x\| \leq M \|x\|_+, \quad x \in X.$$

In this case $x_n \to x$ in $(X, \|\cdot\|_+)$ implies that $x_n \to x$ in $(X, \|\cdot\|)$.

Example 2.15. Provide $X = \mathcal{C}([a,b], \mathbb{K})$ with the sup-norm $\|\cdot\|_\infty$ and with the L^2-norm $\|\cdot\|_p$, $1 \leq p < \infty$, respectively. Then

$$\|f\|_p = \left(\int_a^b |f(t)|^p \, dt \right)^{1/p} \leq (b-a)^{1/p} \sup_{t \in [a,b]} |f(t)| = (b-a)^{1/p} \|f\|_\infty, \quad f \in X;$$

see Example 2.12. □

Definition 2.16. Let X be a linear space over $\mathbb{K}$ with norms $\|\cdot\|$ and $\|\cdot\|_+$. These norms are said to be *equivalent* if there exist $0 < m \leq M$ such that

$$m \|x\|_+ \leq \|x\| \leq M \|x\|_+, \quad x \in X.$$

Example 2.17. The following inequalities are easy to verify for $\mathbb{R}^n$ or $\mathbb{C}^n$:

$$\begin{aligned}\frac{1}{n} \sum_{i=1}^n |x_i| \leq \max |x_i| &\leq \sqrt[p]{\sum_{i=1}^n |x_i|^p} \leq \sum_{i=1}^n |x_i| \\ &\leq \sqrt[q]{n} \sqrt[p]{\sum_{i=1}^n |x_i|^p} \leq n \max |x_i| \leq n \sum_{i=1}^n |x_i|.\end{aligned}$$

The third inequality follows by taking the p-th power and the fourth inequality follows from Hölder's inequality. These inequalities show that the norms in Example 2.7 are equivalent. □

Theorem 2.18. Let X be a finite-dimensional linear space. Then all norms on X are equivalent.

Proof. Let X be the finite-dimensional linear space with norm $\|\cdot\|$. Let $\{e_1,\dots,e_n\}$ be a basis for X and with $x=\sum_{i=1}^n \lambda_i e_i$ define the norm

$$\|x\|_+ = \sqrt{\sum_{i=1}^n |\lambda_i|^2};$$

see Example 2.7. It suffices to show that $\|\cdot\|$ and $\|\cdot\|_+$ are equivalent.

First observe that an upper estimate is obtained by Corollary 2.5 as follows:

$$\|x\| \le \sum_{i=1}^n |\lambda_i|\|e_i\| \le \sqrt{\sum_{i=1}^n |\lambda_i|^2}\sqrt{\sum_{i=1}^n \|e_i\|^2} = M\|x\|_+, \quad M=\sqrt{\sum_{i=1}^n \|e_i\|^2}.$$

To obtain a lower estimate note that the function $f:\mathbb{K}^n \to [0,\infty)$ defined by

$$f(\lambda)=f(\lambda_1,\dots,\lambda_n)=\left\|\sum_{i=1}^n \lambda_i e_i\right\| = \|x\|, \quad x=\sum_{i=1}^n \lambda_i e_i,$$

is continuous, since

$$\begin{aligned}|f(\lambda)-f(\mu)| &= |\|x\|-\|y\||\\ &\le \|x-y\| = \left\|\sum_{i=1}^n \lambda_i e_i - \sum_{i=1}^n \mu_i e_i\right\|\\ &\le \sum_{i=1}^n |\lambda_i-\mu_i|\|e_i\| \le M\sqrt{\sum_{i=1}^n |\lambda_i-\mu_i|^2}.\end{aligned}$$

The closed unit sphere $\mathbb{S}$ in $\mathbb{K}^n$ is bounded and hence compact. Therefore there exists a point $\mu\in\mathbb{S}$ such that for all $\lambda\in\mathbb{S}$

$$0 \le m = f(\mu) \le f(\lambda).$$

If $m=0$ then $\|\sum_{i=1}^n \mu_i e_i\|=0$, so that $\sum_{i=1}^n \mu_i e_i = 0$ with $\mu\in\mathbb{S}$, which contradicts the fact that $\{e_1,\dots,e_n\}$ is a basis for X. Thus $\|x\|=f(\lambda)\ge m>0$ if $\|x\|_+=1$. Hence for every $x\in X$, $x\neq 0$, the above implies that

$$\left\|\frac{x}{\|x\|_+}\right\|_+ = 1 \quad \Rightarrow \quad \left\|\frac{x}{\|x\|_+}\right\| \ge m.$$

Therefore it follows that $m\|x\|_+ \le \|x\|$ for all $x\in X$.

With the above positive upper and lower estimates m and M one concludes that

$$m\|x\|_+ \le \|x\| \le M\|x\|_+, \quad x\in X.$$

Therefore the norms $\|\cdot\|$ and $\|\cdot\|_+$ are equivalent. ☺

Example 2.19. Provide the linear space $\mathcal{C}([a,b],\mathbb{K})$ with the sup-norm $\|\cdot\|_\infty$ and with the L^p-norm $\|\cdot\|_p$, $1\le p<\infty$. Then

$$\|f\|_p \le (b-a)^{1/p}\|f\|_\infty, \quad f\in\mathcal{C}([a,b],\mathbb{K}),$$

see Example 2.15. However, the sup-norm and the L^p-norm on $\mathcal{C}([a,b],\mathbb{K})$ are not equivalent, as there is no $m>0$ such that

$$m\|f\|_\infty \le \|f\|_p, \quad f\in\mathcal{C}([a,b],\mathbb{K}).$$

To see that no such $m > 0$ exists, take for simplicity $a = 0$ and $b = 1$ and define the functions $f_n \in \mathcal{C}([0,1],\mathbb{K})$ by

$$f(x) = \begin{cases} nx, & \text{if } 0 \le x \le 1/n, \\ 1, & \text{if } 1/n \le x \le 1. \end{cases}$$

It is clear that $\|f_n - f\|_p \to 0$ where $f \in \mathcal{C}([0,1],\mathbb{K})$ is defined by $f(x) = 1$, $x \in [0,1]$. However, note that $|f_n(0) - f(0)| = 1$, so that $\|f_n - f\|_\infty = 1$. □

Lemma 2.20. Let X and Y be normed linear spaces. Then the function $\|(\cdot,\cdot)\|_p$ with $1 \le p \le \infty$ defined by

$$\|(x,y)\|_p = \begin{cases} (\|x\|_X^p + \|y\|_Y^p)^{1/p}, & 1 \le p < \infty, \\ \max(\|x\|_X, \|y\|_Y), & p = \infty, \end{cases} \tag{2.2}$$

is a norm on the product space $X \times Y$, and one has

$$(x_n, y_n) \to (x,y) \text{ in } X \times Y \quad \Leftrightarrow \quad x_n \to x \text{ in } X,\ y_n \to y \text{ in } Y.$$

Moreover, all the norms for $p = 1$, $1 < p < \infty$, and $p = \infty$ are equivalent.

2.3 Closed linear subspaces

Let V be any set in a normed linear space $(X, \|\cdot\|)$. Then the distance between x and V is given by

$$d(x,V) = \inf\{\|x - v\| : v \in V\}.$$

Hence the *closure* of V in X can be described as follows:

$$\overline{V} = \{x \in X : d(x,V) = 0\}. \tag{2.3}$$

A set $V \subset X$ is *closed* in X if $V = \overline{V}$.

Example 2.21. Let $s \subset \ell^p$, $1 \le p \le \infty$, be the linear subspace of all finitely supported sequences (entries equal to 0 after a certain index depending on the sequence). Then s is not closed. □

Example 2.22. Let $X = \mathcal{C}([0,1],\mathbb{K})$ with the sup-norm and let

$$V = \{f \in X : f(0) = 0\}. \tag{2.4}$$

Then the linear subspace V is closed in X: if (f_n) in V converges to $f \in X$ then

$$\sup_{x \in [0,1]} |f_n(x) - f(x)| \to 0,$$

and, in particular $f_n(0) \to f(0)$, so that $f \in V$. Note that in this case also

$$\left(\int_0^1 |f_n(x) - f(x)|^p \, dx \right)^{1/p} \to 0,$$

since $\|f\|_p \le \|f\|_\infty$, $f \in \mathcal{C}([0,1],\mathbb{K})$; see Example 2.19.

Now consider the space $X = \mathcal{C}([0,1],\mathbb{K})$ with the L^p-norm, $p \ge 1$, and let V be as in (2.4). Define the functions $f_n \in V$ and $f \in \mathcal{C}([0,1],\mathbb{K})$ as in Example 2.19. Since $\|f_n \to f\|_p \to 0$ and $f \notin V$, it follows that V is not closed in $X = \mathcal{C}([0,1],\mathbb{K})$ with the L^p-norm. □

The result in the following lemma will be revisited in Proposition 3.3 and Proposition 3.6.

Lemma 2.23. Let $(X, \|\cdot\|)$ be a normed linear space and let $V \subset X$ be a finite-dimensional linear subspace. Then V is closed.

Proof. Let (x_n) be a sequence in V such that $x_n \to x$ with $x \in X$. Let V be d-dimensional and choose a basis $\{e_1, \dots, e_d\}$ in V and define the norm $\|\cdot\|_+$ as in Example 2.7 with $p = 1$. Each x_n can be written in terms of this basis:

$$x_n = \lambda_{n,1} e_1 + \cdots + \lambda_{n,d} e_d, \tag{2.5}$$

and therefore one sees that

$$\|x_n - x_m\|_+ = \sum_{i=1}^{d} |\lambda_{n,i} - \lambda_{m,i}|. \tag{2.6}$$

Clearly $x_n - x_m = (x_n - x) - (x_m - x) \to 0$ in X, where $x_n - x_m \in V$. By the equivalence of the norms on V one gets $\|x_n - x_m\|_+ \to 0$. By (2.6) this implies that $(\lambda_{n,i})$ is a Cauchy sequence in $\mathbb{K}$ for each i, hence $\lambda_{n,i} \to \lambda_i$ for some $\lambda_i \in \mathbb{K}$ since $\mathbb{K}$ is complete. Now take $n \to \infty$ in (2.5), then by Proposition 2.14 it follows that

$$x = \lambda_1 e_1 + \cdots + \lambda_d e_d.$$

Thus $x \in V$ and V is closed. ☺

Lemma 2.24. Let $(X, \|\cdot\|)$ be a normed linear space and let $V \subset X$ be a linear subspace. Then the closure $\overline{V}$ is a closed linear subspace.

Proof. Assume that $x, y \in \overline{V}$. Then there exists sequences x_n and y_n in V such that $x_n \to x$ and $y_n \to y$. By Proposition 2.14 this implies that $x_n + y_n \in V$ converges to $x + y \in \overline{V}$. Similarly, $\overline{V}$ is closed under scalar multiplication. ☺

Proposition 2.25. Let $(X, \|\cdot\|)$ be a normed linear space and let $V \subset X$ be a linear subspace. Then

$$\|x + V\| = d(x, V) = \inf\{\|x - y\| : y \in V\} \tag{2.7}$$

is a semi-norm on X/V, and $\|x + V\| \le \|x\|$. This semi-norm is a norm if and only if V is closed.

Proof. It follows from the definition that $\|x + V\| \ge 0$ for $x \in V$. Furthermore, note

$$\|\lambda(x + V)\| = \|\lambda x + V\| = d(\lambda x, V) = |\lambda| d(x, V), \quad x \in X,$$

and that

$$\begin{aligned}
\|(x + V) + (y + V)\| &= \|(x + y) + V\| = d(x + y, V) \\
&= \inf\{\|x + y - z\| : z \in V\} = \inf\{\|x + y - (u + v)\| : u, v \in V\} \\
&\le \inf\{\|x - u\| : u \in V\} + \inf\{\|y - v\| : v \in V\} = \|x + V\| + \|y + V\|.
\end{aligned}$$

Hence (2.7) defines a semi-norm. It follows from (2.3) that

$$\|x + V\| = 0 \quad \Leftrightarrow \quad x \in \overline{V}.$$

Thus if V is closed and $\|x + V\| = 0$, then $x \in V$ and (2.7) defines a norm. Conversely, let (2.7) define a norm. If $x \in \overline{V}$, then $\|x + V\| = 0$, which implies that $x \in V$. Hence V is closed. ☺

Lemma 2.26. Let $(X, \|\cdot\|)$ be a normed linear space and let $E \subset X$ be a set. The closure of the linear span of the set E is given by

$$\overline{\operatorname{span}}\,(E) = \bigcap \{ H \subset X : E \subset H, H \text{ closed linear subspace} \}.$$

Proof. Denote the righthand side by F, so that F is a closed linear subspace. Observe that $E \subset F$, thus $\operatorname{span}(E) \subset F$ and since F is closed, also

$$\overline{\operatorname{span}}\,(E) \subset F.$$

Since $\overline{\operatorname{span}}\,(E)$ is a closed linear subspace of X which contains E, it is clear that

$$F \subset \overline{\operatorname{span}}\,(E).$$

Therefore it follows that $\overline{\operatorname{span}}\,(E) = F$. ☺

Definition 2.27. Let X be a metric space. A subset $E \subset X$ is *dense* in X if $\overline{E} = X$. The space X is *separable* if X has a countable dense subset.

Let $(X, \|\cdot\|)$ be a normed linear space and let $E \subset X$ be a subset. Then

$$E \text{ dense} \quad \Rightarrow \quad \overline{\operatorname{span}}\,(E) = X, \tag{2.8}$$

as follows from $E \subset \operatorname{span}(E)$.

Lemma 2.28. Let $(X, \|\cdot\|)$ be a normed linear space. Then the following statements are equivalent:

1. X is separable;
2. X has a countable subset E with $X = \overline{\operatorname{span}}\,(E)$.

Proof. (1) $\Rightarrow$ (2) There exists a countable set $E \subset X$ such that $\overline{E} = X$. Now apply (2.8).

(2) $\Rightarrow$ (1) First assume that $\mathbb{K} = \mathbb{R}$. Let $E \subset X$ be a countable set such that $\overline{\operatorname{span}}\,(E) = X$ and define the set

$$F = \left\{ \sum_{i=1}^{n} \lambda_i e_i : n \in \mathbb{N}, \lambda_i \in \mathbb{Q}, e_i \in E \right\}.$$

Then F is a countable subset of X. It will be shown that $\overline{F} = X$.

For any $x \in X$ and $\varepsilon > 0$ there exists $y \in \operatorname{span}(E)$ such that $\|x - y\| < \varepsilon/2$. If $y = \sum_{i=1}^{n} \lambda_i e_i$, there exist $\lambda_i' \in \mathbb{Q}$ such that $y' = \sum_{i=1}^{n} \lambda_i' e_i \in F$ with $\|y - y'\| < \varepsilon/2$. Then $\|x - y'\| \le \|x - y\| + \|y - y'\| < \varepsilon$. Hence $\overline{F} = X$ and since F is countable, it follows that X is separable. The case $\mathbb{K} = \mathbb{C}$ follows similarly. ☺

Example 2.29. Here are some examples of separable and nonseparable normed linear spaces.

1. The normed linear space $X = \mathcal{C}([a,b], \mathbb{K})$ with the sup-norm is separable. This is a consequence of the Stone-Weierstrass theorem: take for E the subset in $\mathcal{C}([a,b], \mathbb{K})$ of all polynomials with rational coefficients.
2. The normed linear space $X = \ell^p$, $1 \le p < \infty$, is separable. Take for E the finitely supported sequences with rational coefficients.
3. The normed linear space ℓ^∞ is not separable. To see this, assume that ℓ^∞ is separable. Then there exists a countable subset E of ℓ^∞ which is dense. Denote the elements of E by

$$x_n = (x_{n,1}, x_{n,2}, x_{n,3}, \dots).$$

The element $y = (y_1, y_2, y_3, \dots)$ defined by

$$y_n = \begin{cases} x_{n,n} + 1 & \text{if } |x_{n,n}| \le 1 \\ 0 & \text{if } |x_{n,n}| > 1 \end{cases}$$

clearly belongs to ℓ^∞ as $|y_n| \le 2$ for each $n \in \mathbb{N}$. Furthermore the element y has the property that $|y_n - x_{n,n}| \ge 1$ for each $n \in \mathbb{N}$, so that

$$\|y - x_n\|_\infty \ge 1.$$

This is a contradiction and, hence, ℓ^∞ is not separable.

$\square$

Example 2.30. The unit ball in ℓ^p, $1 \le p \le \infty$, is not compact. To see this observe

$$\|e_i - e_j\| = \begin{cases} 2^{1/p} & \text{if } 1 \le p < \infty \\ 2 & \text{if } p = \infty \end{cases}$$

where e_i stands for the i-th unit vector $e_i = (0, 0, \cdots, 1, 0, 0, \cdots)$. Thus (e_i) is a sequence which does not contain a convergent subsequence. $\square$

Lemma 2.31 (Riesz's lemma)**.** Let $(X, \|\cdot\|)$ be a normed linear space and let $V \subset X$ be a closed linear subspace with $V \ne X$. For every λ with $0 < \lambda < 1$ there exists an element $x_\lambda \in X$ with $\|x_\lambda\| = 1$ and

$$\|x_\lambda - v\| > \lambda \text{ for all } v \in V.$$

Proof. Since V is a closed linear subspace and $V \ne X$ there exists $x \in X \setminus V$ such that

$$0 < d(x, V) = \inf\{\|x - v\| : v \in V\}.$$

Since $d(x, V) < d(x, V)/\lambda$ there exists $w \in V$ such that

$$d(x, V) \le \|x - w\| < \frac{d(x, V)}{\lambda}, \quad \text{so that} \quad \frac{1}{\|x - w\|} > \frac{\lambda}{d(x, V)}.$$

Define the element x_λ by

$$x_\lambda = \frac{x - w}{\|x - w\|},$$

so that $\|x_\lambda\| = 1$ and for all $v \in V$:

$$\begin{aligned} \|x_\lambda - v\| = \left\| \frac{x - w}{\|x - w\|} - v \right\| &= \frac{1}{\|x - w\|} \|x - (w + \|x - w\| v)\| \\ &> \frac{\lambda}{d(x, V)} d(x, V) = \lambda, \end{aligned}$$

since $w + \|x - w\| v \in V$. ☺

Theorem 2.32. Let $(X, \|\cdot\|)$ be a normed linear space for which the unit ball

$$D = \{x \in X : \|x\| \le 1\}$$

is compact. Then X is finite-dimensional.

Proof. Assume that X is infinite-dimensional. It will be shown that then there is a sequence in D which has no convergent subsequence, which leads to a contradiction.

Choose any $e_1 \in X$ with $\|e_1\| = 1$, then span$\{e_1\}$ is a one dimensional closed linear subspace of X, while span$\{e_1\} \neq X$. By the Riesz lemma with $V = \text{span}\{e_1\}$ and $\lambda = 1/2$ there exists $e_2 \in X$ with $\|e_2\| = 1$ and

$$d(e_2, \text{span}\{e_1\}) \geq 1/2.$$

Therefore span$\{e_1, e_2\}$ is a two dimensional closed linear subspace of X, while span$\{e_1, e_2\} \neq X$. Hence by the Riesz lemma with $V = \text{span}\{e_1, e_2\}$ and $\lambda = 1/2$ there exists $e_3 \in X$ with $\|e_3\| = 1$ and

$$d(e_3, \text{span}\{e_1, e_2\}) \geq 1/2.$$

Clearly span$\{e_1, e_2, e_3\}$ is a three dimensional closed linear subspace of X. Now comes the induction step. Assume that there are linearly independent vectors $\{e_1, \dots, e_n\}$ with $\|e_i\| = 1$ and

$$d(e_i, \text{span}\{e_1, \dots, e_{i-1}\}) \geq 1/2,$$

for $i = 1, \dots, n$. Hence by the Riesz lemma with $V = \text{span}\{e_1, \dots, e_n\}$ and $\lambda = 1/2$ there exists $e_{n+1} \in X$ with $\|e_{n+1}\| = 1$ and

$$d(e_{n+1}, \text{span}\{e_1, \dots, e_n\}) \geq 1/2.$$

It is straightforward to see that the vectors $\{e_1, \dots, e_{n+1}\}$ are linearly independent.

By induction it has been shown there exists a sequence of linearly independent elements (e_i) in X with $\|e_i\| = 1$, $i \in \mathbb{N}$, with

$$d(e_{i+1}, \text{span}\{e_1, \dots, e_i\}) \geq 1/2.$$

Hence the sequence (e_i) has the property that $\|e_i\| = 1$, while

$$\|e_i - e_j\| \geq 1/2, \quad i \neq j.$$

This implies that there is no convergent subsequence. Thus the unit ball D in X is not compact, which is a contradiction. ☺

2.4 Inner product spaces

Definition 2.33. Let X be a linear space over $\mathbb{K}$. The map $\langle \cdot, \cdot \rangle : X \times X \to \mathbb{K}$ is called an *inner product* if

1. $\langle x, x \rangle \geq 0$, $x \in X$;
2. $\langle x, x \rangle = 0 \Leftrightarrow x = 0$;
3. $\langle \lambda x + \mu y, z \rangle = \lambda \langle x, z \rangle + \mu \langle y, z \rangle$, $x, y, z \in X$, $\lambda, \mu \in \mathbb{K}$;
4. $\langle x, y \rangle = \overline{\langle y, x \rangle}$ if $\mathbb{K} = \mathbb{C}$ or $\langle x, y \rangle = \langle y, x \rangle$ if $\mathbb{K} = \mathbb{R}$, $x, y \in X$.

The pair $(X, \langle \cdot, \cdot \rangle)$ is a called an inner product space over $\mathbb{K}$. By abuse of language X itself will often be called an inner product space. If in (2) only the implication ($\Leftarrow$) holds then $\langle \cdot, \cdot \rangle$ is called a *semi-definite inner product* and $(X, \langle \cdot, \cdot \rangle)$ is a called an semi-definite inner product space.

Remark 2.34. If $V \subset X$ is a linear subspace of an inner product space $(X, \langle \cdot, \cdot \rangle)$, then the restriction $(V, \langle \cdot, \cdot \rangle)$ is an inner product space. □

Example 2.35. The linear $\mathbb{K}^n$ is an inner product space when it is provided with

$$\langle x, y \rangle = \sum_{i=1}^{n} x_i \bar{y}_i, \quad x, y \in \mathbb{K}^n.$$

□

Example 2.36. The linear space ℓ^2 is an inner product space when it is provided with

$$\langle x,y\rangle = \sum_{i=1}^{\infty} x_i \bar{y}_i, \quad x,y \in \ell^2.$$

The convergence of the infinite sum is guaranteed by Hölder's inequality. □

Example 2.37. The linear space $\mathcal{C}([a,b],\mathbb{K})$ can be seen as an inner product space when it is provided with

$$\langle f,g\rangle = \int_a^b f(t)\overline{g(t)}\,dt, \quad f,g \in \mathcal{C}([a,b],\mathbb{K}).$$

Here the integral is understood in the sense of Riemann. Note that $\langle\cdot,\cdot\rangle$ is an inner product on $\mathcal{C}([a,b],\mathbb{K})$, the so-called L^2-inner product and that this inner product generates the L^2-norm. □

Example 2.38. Define on $\mathcal{C}^1([a,b],\mathbb{K})$ the product

$$\langle f,g\rangle = \int_a^b f(t)\overline{g(t)}\,dt + \int_a^b f'(t)\overline{g'(t)}\,dt, \quad f,g \in \mathcal{C}^1([a,b]).$$

Then this defines an inner product. □

Example 2.39. The $m \times n$ matrices over $\mathbb{K}$ form a linear space under the usual operations, which can be identified with $\mathbb{K}^{m\times n}$. Define for $m \times n$ matrices A and B the product

$$\langle A,B\rangle = \operatorname{tr}(B^*A).$$

Then $\langle\cdot,\cdot\rangle$ is an inner product. □

Lemma 2.40. Let $(X,\langle\cdot,\cdot\rangle)$ and $(Y,\langle\cdot,\cdot\rangle)$ be inner product spaces. Then the product space $X \times Y$ provided with

$$\langle (x,y),(u,v)\rangle = \langle x,u\rangle + \langle y,v\rangle, \quad (x,y),(u,v) \in X \times Y, \tag{2.9}$$

is an inner product space. The induced norm is equivalent to any of the norms $\|(\cdot,\cdot)\|_p$, $1 \le p \le \infty$, as in (2.2).

Lemma 2.41 (Cauchy-Schwarz inequality)**.** Let $\langle\cdot,\cdot\rangle$ be an inner product on the linear space X over $\mathbb{K}$. Then

$$|\langle x,y\rangle|^2 \le \langle x,x\rangle\langle y,y\rangle, \quad x,y \in X. \tag{2.10}$$

In particular,

$$\|x\| = \sqrt{\langle x,x\rangle}, \quad x \in X,$$

defines a norm on X. The Cauchy-Schwarz inequality becomes an equality if and only if x and y are linearly dependent.

Proof. It is clear that for all $\lambda \in \mathbb{K}$ one has

$$0 \le \langle x-\lambda y, x-\lambda y\rangle = \langle x,x\rangle - \lambda\langle y,x\rangle - \bar{\lambda}\langle x,y\rangle + |\lambda|^2\langle y,y\rangle.$$

With $\lambda = t\langle x,y\rangle$ and $t \in \mathbb{R}$ this becomes

$$0 \le \langle x-\lambda y, x-\lambda y\rangle = \langle x,x\rangle - 2t|\langle x,y\rangle|^2 + t^2|\langle x,y\rangle|^2\langle y,y\rangle. \tag{2.11}$$

Since this quadratic polynomial in t has at most one zero, it follows that the discriminant must be non-positive, which gives (2.10).

Assume that equality holds in (2.10). If $y = 0$, then x and y are clearly linearly dependent. If $y \neq 0$, then the right-hand side of (2.11) is zero for $t = 1/\langle y,y\rangle$ in which case it follows that $x - \lambda y = 0$ with $\lambda = \langle x,y\rangle/\langle y,y\rangle$. Conversely, if $x = \lambda y$, then clearly equality holds in (2.10).

Finally observe that

$$\begin{aligned}\|x+y\|^2 &= \|x\|^2 + 2\mathrm{Re}\,\langle x,y\rangle + \|y\|^2 \\ &\leq \|x\|^2 + 2\|x\|\|y\| + \|y\|^2 = (\|x\| + \|y\|)^2,\end{aligned}$$

which completes the proof. ☺

An inspection of the proof of Lemma 2.41 shows that only the semi-definiteness of the inner product has been used to prove the inequality. This leads to the following useful observation; cf. Definition 2.1.

Corollary 2.42 (Cauchy-Schwarz inequality for the semi-definite case)**.** Let $\langle\cdot,\cdot\rangle$ be a semi-definite inner product on a linear space X over $\mathbb{K}$. Then

$$|\langle x,y\rangle|^2 \leq \langle x,x\rangle\langle y,y\rangle, \quad x,y \in X,$$

In particular,

$$\|x\| = \sqrt{\langle x,x\rangle}, \quad x \in X,$$

defines a semi-norm on X.

Remark 2.43. For a norm induced by an inner product there are some specific observations. For obvious reasons the following identity is called

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

the *parallelogram law* (as in the usual two dimensional case). Note that the sup-norm on $\mathcal{C}([a,b],\mathbb{K})$ is not inherited from an inner product as in this context the parallelogram law is not satisfied (take $a = 0$, $b = 1$, $f(t) = 1$, and $g(t) = t$).

Furthermore there are the so-called *polarization identities*:

$$4\langle x,y\rangle = \begin{cases} \|x+y\|^2 - \|x-y\|^2 & \text{if } \mathbb{K} = \mathbb{R}, \\ \|x+y\|^2 - \|x-y\|^2 + i\|x+iy\|^2 - i\|x-iy\|^2 & \text{if } \mathbb{K} = \mathbb{C}. \end{cases}$$

These formulas allow one to recuperate an inner product from the norm. □

Lemma 2.44. Let $(X, \langle\cdot,\cdot\rangle)$ be an inner product space. If $x_n \to x$ and $y_n \to y$, then

$$\langle x_n, y_n\rangle \to \langle x,y\rangle.$$

Proof. Observe that the sequence $(\|y_n\|)$ is bounded (as it is convergent). Hence

$$\begin{aligned}|\langle x_n,y_n\rangle - \langle x,y\rangle| &\leq |\langle x_n,y_n\rangle - \langle x,y_n\rangle| + |\langle x,y_n\rangle - \langle x,y\rangle| \\ &\leq \|y_n\|\|x_n - x\| + \|x\|\|y - y_n\| \\ &\leq M\|x_n - x\| + \|x\|\|y - y_n\|,\end{aligned}$$

which shows the result. ☺

Definition 2.45. Let $(X, \langle\cdot,\cdot\rangle)$ be an inner product space. The vectors $x, y \in X$ are called *orthogonal*, denoted by $x \perp y$, if $\langle x,y\rangle = 0$. Let $E \subset X$ be a subset, then $x \in X$ is orthogonal to E, denoted by $x \perp E$, if $x \perp e$ for all $e \in E$. The *orthogonal complement* of E is defined by

$$E^{\perp} = \{x \in X : x \perp E\}.$$

The subsets $E \subset X$ and $F \subset X$ are orthogonal if $\langle x,y\rangle = 0$ for all $x \in E$, $y \in F$.

Remark 2.46. Let $(X, \langle\cdot,\cdot\rangle)$ be an inner product space and let $x, y \in X$. If $x \perp y$, then the Pythagoras identity follows:

$$\|x+y\|^2 = (x+y, x+y) = \|x\|^2 + \|y\|^2.$$

□

Lemma 2.47 (Properties of orthogonal complements)**.** Let $(X, \langle\cdot,\cdot\rangle)$ be an inner product space and let E, F be subsets of X. Then

1. $E^{\perp}$ is a closed linear subspace;
2. $E \subset F \Rightarrow F^{\perp} \subset E^{\perp}$;
3. $(\overline{\mathrm{span}}\,(E))^{\perp} = E^{\perp}$;
4. $E \subset (E^{\perp})^{\perp}$.

Proof. (1) Note that $x \perp E$ and $y \perp E$ imply $x + y \perp E$ and $\lambda x \perp E$. Likewise $x_n \perp E$ and $x_n \to x$ imply $x \perp E$.

(2) This is trivial.

(3) It follows from $E \subset \mathrm{span}\,(E) \subset \overline{\mathrm{span}}\,(E)$ that $(\overline{\mathrm{span}}\,E)^{\perp} \subset E^{\perp}$. The converse is straightforward: if $x \perp E$, then $x \perp \mathrm{span}\,(E)$ and $x \perp \overline{\mathrm{span}}\,(E)$.

(4) If $e \in E$, then $\langle e,x\rangle = 0$ for all $x \in E^{\perp}$. This implies $e \in (E^{\perp})^{\perp}$. ☺

Lemma 2.48 (Characterization of best approximations in inner product spaces)**.** Let $(X, \langle\cdot,\cdot\rangle)$ be an inner product space and let $V \subset X$ be a linear subspace. If $x \in X$ and $v \in V$, then

$$\|x - v\| = d(x,V) \quad \Leftrightarrow \quad x - v \in V^{\perp}.$$

Proof. $(\Rightarrow)$ For all $z \in V$ and $\lambda \in \mathbb{K}$ we have

$$\begin{aligned}\|x-v\|^2 &\le \|x - v - \lambda z\|^2 \\ &= \|x-v\|^2 - \bar{\lambda}\langle x-v, z\rangle - \lambda\langle z, x-v\rangle + |\lambda|^2\|z\|^2.\end{aligned}$$

For $\lambda = t\langle x-v,z\rangle$ with $t > 0$ this gives

$$2|\langle x-v,z\rangle|^2 \le t|\langle x-v,z\rangle|^2\|z\|^2.$$

Letting $t \to 0$ implies that $x - v \in V^{\perp}$.

$(\Leftarrow)$ If $z \in V$, then $x - v \perp v - z$ so that

$$\|x-z\|^2 = \|x-v+v-z\|^2 = \|x-v\|^2 + \|v-z\|^2 \ge \|x-v\|^2.$$

Taking the infimum over all $z \in V$ implies that $d(x,V) \ge \|x - v\|$. The reverse inequality is trivial. ☺

2.5 Orthonormal systems and the Gram-Schmidt procedure

Definition 2.49. Let $(X, \langle\cdot,\cdot\rangle)$ be an inner product space. The elements e_i, $i \in I \subset \mathbb{N}$, in X are called *orthonormal* if

$$\langle e_i, e_j\rangle = \delta_{ij}, \quad i, j \in I.$$

Remark 2.50. Orthonormal vectors e_i, $i \in \mathbb{N}$, in X are linearly independent. To see this, note that

$$\left\| \sum_{i=1}^{m} \lambda_i e_i \right\|^2 = \left\langle \sum_{i=1}^{m} \lambda_i e_i, \sum_{j=1}^{m} \lambda_j e_j \right\rangle = \sum_{i,j=1}^{m} \lambda_i \bar{\lambda}_j \langle e_i, e_j\rangle = \sum_{i=1}^{m} |\lambda_i|^2.$$

In particular, if $\lambda_1 e_1 + \cdots + \lambda_m e_m = 0$, then $\lambda_i = 0$ for all $i = 1, \ldots, m$. □

Theorem 2.51 (Gram-Schmidt)**.** Let X be an inner product space and let $f_1, \ldots, f_n$ be linearly independent vectors in X. Then there exist orthonormal vectors $e_1, \ldots, e_n$, such that

$$\operatorname{span}\{e_1, \ldots, e_k\} = \operatorname{span}\{f_1, \ldots, f_k\}$$

for every $k = 1, \ldots, n$.

Proof. The proof is by induction on n. Define

$$e_1 = \frac{f_1}{\|f_1\|} \quad \text{so that} \quad \|e_1\| = 1,$$

and observe that

$$\operatorname{span}\{e_1\} = \operatorname{span}\{f_1\},$$

which completes the proof for the case $n = 1$.

For the induction step, assume for some n that the vectors $e_1, \ldots, e_n$ are orthonormal and that

$$\operatorname{span}\{e_1, \ldots, e_k\} = \operatorname{span}\{f_1, \ldots, f_k\}$$

for all $k = 1, \ldots, n$. Then introduce the vector $\widetilde{e}_{n+1}$ by

$$\widetilde{e}_{n+1} = f_{n+1} - \sum_{i=1}^{n} \langle f_{n+1}, e_i\rangle e_i,$$

and observe that $\langle \widetilde{e}_{n+1}, e_j\rangle = 0$ for $j = 1, \ldots, n$, and that $\widetilde{e}_{n+1} \neq 0$. To see the last statement, note that otherwise $f_{n+1} \in \operatorname{span}\{e_1, \ldots, e_n\} = \operatorname{span}\{f_1, \ldots, f_n\}$, a contradiction. Next normalise the vector $\widetilde{e}_{n+1}$:

$$e_{n+1} = \frac{\widetilde{e}_{n+1}}{\|\widetilde{e}_{n+1}\|} \quad \text{so that} \quad \|e_{n+1}\| = 1.$$

Then the vectors $e_1, \ldots, e_{n+1}$ are orthonormal and

$$\operatorname{span}\{e_1, \ldots, e_{n+1}\} = \operatorname{span}\{f_1, \ldots, f_{n+1}\}.$$

For the last statement observe that f_{n+1} is a linear combination of $e_1, \ldots, e_{n+1}$. ☺

Corollary 2.52. Let X be an inner product space with $\dim X = n$. Let $V \subset X$ be a linear subspace. Then we have the following *orthogonal sum* of subspaces:

$$X = V \oplus V^{\perp}.$$

Proof. Let V have a basis $\{f_1,\dots,f_m\}$ and complete this to a basis $\{f_1,\dots,f_m,f_{m+1},\dots,f_n\}$ for X. Applying the Gram-Schmidt procedure results in an orthonormal basis $\{e_1,\dots,e_m,e_{m+1},\dots,e_n\}$ for X, such that $V=\operatorname{span}\{e_1,\dots,e_m\}$ and $V^\perp=\operatorname{span}\{e_{m+1},\dots,e_n\}$. ☺

Corollary 2.53. Let X be an inner product space and let $V\subset X$ and $W\subset X$ be linear subspaces. Assume for some $n\in\mathbb{N}$ that

$$\dim V=n-1,\quad \dim W=n. \tag{2.12}$$

Then

$$V^\perp\cap W\neq\{0\}. \tag{2.13}$$

Proof. Due to the assumption $\dim V=n-1$, there are linearly independent vectors $f_1,\dots,f_{n-1}$ such that $V=\operatorname{span}\{f_1,\dots,f_{n-1}\}$. Due to the assumption $\dim W=n$, there are n orthonormal vectors $e_1,\dots,e_n$ which span W; cf. Theorem 2.51. It is clear that the linear combination $\sum_{i=1}^n c_ie_i$ belongs to $V^\perp$ if and only if

$$\begin{pmatrix}\langle e_1,f_1\rangle & \dots & \langle e_n,f_1\rangle\\ \vdots & & \vdots\\ \langle e_1,f_{n-1}\rangle & \dots & \langle e_n,f_{n-1}\rangle\end{pmatrix}\begin{pmatrix}c_1\\ \vdots\\ c_n\end{pmatrix}=0. \tag{2.14}$$

The matrix in the left-hand side of (2.14) maps $\mathbb{C}^n$ to $\mathbb{C}^{n-1}$. Hence the equation (2.14) has a nontrivial solution or, equivalently, the statement in (2.13) holds. ☺

Corollary 2.54. Let X be an inner product space and let f_i, $i\in\mathbb{N}$, be linearly independent vectors in X. Then there exist an orthonormal system e_i, $i\in\mathbb{N}$, such that for every $k\in\mathbb{N}$:

$$\operatorname{span}\{e_1,\dots,e_k\}=\operatorname{span}\{f_1,\dots,f_k\}.$$

In particular,

$$\overline{\operatorname{span}}\{e_n:n\in\mathbb{N}\}=\overline{\operatorname{span}}\{f_n:n\in\mathbb{N}\}.$$

Lemma 2.55 (Best approximation)**.** Let X be an inner product space and let $V\subset X$ be a finite-dimensional subspace. For each $x\in X$ there exists precisely one *best approximation* in V, namely there is a unique element $v\in V$ such that

$$\|x-v\|=d(x,V)=\inf\{\|x-z\|:z\in V\}.$$

In terms of an orthonormal basis $\{e_1,\dots,e_n\}$ for V, one has

$$v=\sum_{i=1}^n\langle x,e_i\rangle e_i\quad\text{and}\quad \|x-v\|=\sqrt{\|x\|^2-\sum_{i=1}^n|\langle x,e_i\rangle|^2}.$$

Proof. If V is n-dimensional, then V is spanned by linearly independent vectors $f_1,\dots,f_n$. The Gram-Schmidt procedure gives orthonormal vectors $e_1,\dots,e_n$ such that

$$V=\operatorname{span}\{f_1,\dots,f_n\}=\operatorname{span}\{e_1,\dots,e_n\}.$$

Then it is clear that

$$\inf\{\|x-z\|:z\in V\}=\inf\left\{\left\|x-\sum_{k=1}^n\lambda_ke_k\right\|:\lambda_1,\dots,\lambda_n\in\mathbb{K}\right\}.$$

Denoting $c_k = \langle x, e_k \rangle$ for $k = 1, \ldots, n$, it follows from completing squares in the following identity

$$\begin{aligned}\left\| x - \sum_{k=1}^{n} \lambda_k e_k \right\|^2 &= \|x\|^2 - \sum_{k=1}^{n} \bar{\lambda}_k c_k - \sum_{k=1}^{n} \lambda_k \bar{c}_k + \sum_{k=1}^{n} |\lambda_k|^2 \\ &= \|x\|^2 + \sum_{k=1}^{n} |\lambda_k - c_k|^2 - \sum_{k=1}^{n} |c_k|^2,\end{aligned}$$

that the infimum of the left-hand side is a minimum which is attained precisely when $\lambda_k = c_k$, in which case

$$\inf\{\|x - z\| : z \in V\} = \sqrt{\|x\|^2 - \sum_{k=1}^{n} |c_k|^2}.$$

This completes the proof. ☺

Example 2.56. Provide $\mathcal{C}([0,1], \mathbb{K})$ with the L^2-inner product. Let V be the linear subspace of first order polynomials. By means of the Gram-Schmidt procedure an orthonormal basis for V is given by the functions

$$e_1(t) = 1, \quad e_2(t) = \sqrt{3}\,(2t - 1), \quad t \in [0,1].$$

The best approximation in V of the function $x(t) = t^2$ in the sense of the L^2-norm and the corresponding distance are given by

$$v(t) = t - \frac{1}{6} \quad \text{and} \quad \|x - v\| = \frac{1}{6\sqrt{5}}.$$

□

Chapter 3

Banach spaces and Hilbert spaces

Banach spaces are normed linear spaces which are complete, i.e., each Cauchy sequence converges to an element in the space. Likewise Hilbert spaces are inner product spaces which are complete. The completeness of a space sometimes guarantees nice properties: think of the best approximation problem in a Hilbert space. Many normed linear spaces will be shown to be complete. However, when a normed linear is not complete one can construct a completion by "identifying Cauchy sequences with limit elements". Another and sometimes more satisfactory way to complete for instance $\mathcal{C}([a,b],\mathbb{K})$ with L^p-norm is via the theory of measure and integration. This procedure is briefly explained. As an application the abstract theory of complete orthonormal systems in Hilbert spaces is applied to the theory of Fourier series.

3.1 Banach spaces

Definition 3.1. Let $(X,\|\cdot\|)$ be a normed linear space. A sequence (x_n) of elements in X is called a *Cauchy sequence* if for every $\varepsilon > 0$ there exists $N \in \mathbb{N}$ such that $n,m \geq N$ imply that $\|x_n - x_m\| < \varepsilon$. A normed linear space $(X,\|\cdot\|)$ is called complete if every Cauchy sequence in X converges in X. A complete normed linear space is called a *Banach space*.

Remark 3.2. The following properties are often useful.

1. The normed linear space $\mathbb{K}$ is complete.
2. Let (x_i) be a Cauchy sequence in a normed linear space X. Then the sequence $(\|x_i\|)$ is convergent in $\mathbb{R}$, due to
$$\big|\|x_i\| - \|x_j\|\big| \leq \|x_i - x_j\|,$$
so that $(\|x_i\|)$ is a Cauchy sequence in $\mathbb{R}$ and $\mathbb{R}$ is complete.
3. Let (x_i) be a Cauchy sequence in a normed linear space X. Then, in particular, the sequence (x_i) is bounded: there exists $M > 0$ such that $\|x_n\| \leq M$ for all $n \in \mathbb{N}$, as the sequence $(\|x_i\|)$ converges in $\mathbb{R}$.
4. Let the Cauchy sequence (x_i) have a convergent subsequence. Then the Cauchy sequence itself converges to the same limit. To see this observe the following. For every $\varepsilon > 0$ there exists $N \in \mathbb{N}$ such that $\|x_n - x_m\| < \varepsilon/2$ for $m,n \geq N$. Let (x_{n_k}) be the subsequence of (x_n) which converges to $x \in X$. Hence by definition of convergence there exists an index $n_k > N$ such that $\|x_{n_k} - x\| < \varepsilon/2$. With $n \geq N$ it now follows that
$$\|x - x_n\| \leq \|x - x_{n_k}\| + \|x_{n_k} - x_n\| < \varepsilon.$$
In other words the Cauchy sequence (x_n) converges to $x \in X$.

□

Proposition 3.3. Let X be a finite-dimensional normed linear space. Then X is a Banach space.

Proof. Recall that all norms on a finite-dimensional normed linear space are equivalent; cf. Theorem 2.18. Hence it suffices to show the assertion for some norm. Choose a basis $\{e_1,\dots,e_d\}$ for X and define for $x=\sum_{i=1}^d \lambda_i e_i$ the norm

$$\|x\|_+ = \sum_{i=1}^d |\lambda_i|.$$

Let $x_n=\sum_{i=1}^d \lambda_{n,i}e_i$, then the sequence (x_n) is Cauchy if and only if for every $i=1,\dots,d$ the sequence $(\lambda_{n,i})$ is Cauchy in $\mathbb{K}$. Therefore $\lambda_{n,i}\to\lambda_i$ as $n\to\infty$ and for $x=\sum_{i=1}^d \lambda_i e_i$ it follows that $x_n\to x$ in X. ☺

Example 3.4. The space $\mathbb{K}^n$ provided with the ℓ^p norm for $1\le p\le\infty$ is complete; cf. Example 2.6. This is clear from Proposition 3.3 and Theorem 2.18. □

Lemma 3.5. Let X and Y be Banach spaces. Then the normed linear space $X\times Y$ provided with the norm $\|(\cdot,\cdot)\|_p$ as in (2.2) is complete for $p=1$, $1<p<\infty$, and $p=\infty$.

Proposition 3.6. Let X be a normed linear space and let $V\subset X$ be a linear subspace. Then the following statements are valid:

1. If X is a Banach space and V is closed, then V is a Banach space.
2. If V is a Banach space, then V is closed.

In particular, any finite-dimensional subspace of X is closed.

Proof. (1) Let (v_n) be a Cauchy sequence in V. Then (v_n) is Cauchy in X and hence $v_n\to x$ for some $x\in X$. Since V is closed, it follows that $x\in V$. Hence V is a Banach space.

(2) Let the sequence (v_n) in V converge to $x\in X$. Then (v_n) is Cauchy in V and hence converges to $v\in V$. As the limit is unique it follows that $x=v\in V$. Hence V is closed. ☺

How to check that a normed linear space is complete? Roughly speaking there are two situations to consider.

1. If one is given a space X with a norm an option is to choose a Cauchy sequence x_n in X. Due to the structure of the norm on X it is often possible, by a reduction to the case $\mathbb{K}$, to identify a candidate for the limit, say x. Then it remains to show that $x\in X$ and that actually $x_n\to x$ in X.
2. If one is given a normed linear space V which lies inside a Banach space X then it suffices to show that V is closed; see Proposition 3.6.

Example 3.7. For $1\le p<\infty$ the normed linear space ℓ^p with norm

$$\|x\| = \left(\sum_{i=1}^\infty |x_i|^p\right)^{1/p}, \quad x\in\ell^p,$$

is complete. To see this, let $x_n=(x_{n,1},x_{n,2},\dots)$ be a Cauchy sequence in ℓ^p. Then with $\varepsilon>0$ there exists $N\in\mathbb{N}$ such that for all $n,m\ge N$

$$\sum_{i=1}^\infty |x_{n,i}-x_{m,i}|^p = \|x_n-x_m\|^p < \varepsilon^p.$$

For each $i\in\mathbb{N}$ it follows from $|x_{n,i}-x_{m,i}|<\varepsilon$ that $x_{m,i}\to x_i$ for some $x_i\in\mathbb{K}$ as $m\to\infty$ since $\mathbb{K}$ is complete. Define the formal element $x=(x_1,x_2,\dots)$. It remains to show that $x\in\ell^p$ and $x^n\to x$ in

ℓ^p. To do this, observe that for $n,m \geq N$ one has for arbitrary $M \in \mathbb{N}$:

$$\sum_{i=1}^{M} |x_{n,i} - x_{m,i}|^p < \varepsilon^p \quad \underset{m\to\infty}{\Longrightarrow} \quad \sum_{i=1}^{M} |x_{n,i} - x_i|^p \leq \varepsilon^p \quad \underset{M\to\infty}{\Longrightarrow} \quad \sum_{i=1}^{\infty} |x_{n,i} - x_i|^p \leq \varepsilon^p.$$

Thus $x = x^n - (x^n - x) \in \ell^p$ and $x^n \to x$ in ℓ^p. Hence the normed linear space ℓ^p is complete.

Likewise the normed linear space ℓ^∞ with norm

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

is complete. The normed linear space c of all convergent sequences is a closed linear subspace of ℓ^∞ and hence is a Banach space in its own right. The space c_0 of all sequences which converge to 0 is a closed linear subspace of ℓ^∞, and hence is a Banach space in its own right. The space s of all finitely supported sequences is not a closed linear subspace of ℓ^∞. It is not complete in its own right, but it is dense in each ℓ^p. □

Example 3.8. Let S be a nonempty set. The space $\mathcal{B}(S,\mathbb{K})$ provided with the sup-norm is a Banach space. The space $\mathcal{C}([a,b],\mathbb{K})$ is closed in $\mathcal{B}([a,b],\mathbb{K})$ and hence a Banach space. □

Example 3.9. For $1 \leq p < \infty$ the normed linear space $\mathcal{C}([a,b],\mathbb{K})$ provided with the L^p-norm

$$\|f\| = \left(\int_a^b |f(t)|^p \, dt \right)^{1/p}, \quad f \in \mathcal{C}([a,b],\mathbb{K}),$$

is not complete. For simplicity consider the case $a = -1$ and $b = 1$ and define the functions $\varphi_n \in \mathcal{C}([-1,1],\mathbb{K})$ by

$$\varphi_n(t) = \begin{cases} 0 & \text{if } -1 \leq t < 0, \\ nt & \text{if } 0 \leq t < 1/n, \\ 1 & \text{if } 1/n \leq t \leq 1. \end{cases}$$

Define the function φ by

$$\varphi(t) = \begin{cases} 0 & \text{if } -1 \leq t \leq 0 \\ 1 & \text{if } 0 < t \leq 1. \end{cases}$$

and observe that $\varphi \notin \mathcal{C}([-1,1],\mathbb{K})$. Let X be the linear space defined by

$$X = \{ f + \lambda\varphi : f \in \mathcal{C}([-1,1],\mathbb{K}),\ \lambda \in \mathbb{K} \},$$

and define for $1 \leq p < \infty$

$$\|f + \lambda\varphi\|_p = \left(\int_{-1}^{1} |f(t) + \lambda\varphi(t)|^p \, dt \right)^{1/p}.$$

Then one sees that $\|\cdot\|_p$ is a well-defined norm on X, cf. Example 2.12, and that $\|\varphi_n - \varphi\|_p \to 0$. Hence the linear subspace $V = (\mathcal{C}([-1,1],\mathbb{K})$ is not closed in X, and by Proposition 3.6 the space $V = (\mathcal{C}([-1,1],\mathbb{K})$ with the norm $\|\cdot\|_p$ is not complete. □

Example 3.10. For $1 \leq p < \infty$ the normed linear space $\mathcal{C}^1([a,b],\mathbb{K})$ provided with the L^p-norm

$$\|f\| = \left(\int_a^b (|f(t)|^p + |f'(t)|^p) \, dt \right)^{1/p}, \quad f \in \mathcal{C}^1([a,b],\mathbb{K}),$$

is not complete. □

Example 3.11. Let S be a nonempty set and let X be a Banach space. Then the linear space $\mathcal{B}(S,X)$ provided with the sup-norm

$$\|f\|_\infty = \sup_{s\in S} \|f(s)\|, \quad f \in \mathcal{B}(S,X),$$

is a Banach space.

If the set S is a metric space, then $\mathcal{BC}(S,X) \subset \mathcal{B}(S,X)$ denotes the linear subspace of bounded continuous functions from S to X with the supremum norm. Then $\mathcal{BC}(S,X)$ is itself a Banach space; cf. Proposition 3.6. Of course when S is compact then $\mathcal{C}(S,X) = \mathcal{BC}(S,X)$. □

Definition 3.12. Let X be a normed linear space and let (x_n) be a sequence in X. Then the *infinite series* $\sum_{i=1}^\infty x_i$ is said to converge when the sequence of partial sums

$$s_n = \sum_{i=1}^n x_i$$

has a limit in X, in which case the infinite sum is defined as

$$\sum_{i=1}^\infty x_i = \lim_{n\to\infty} s_n.$$

Theorem 3.13 (Absolute convergence implies convergence)**.** Let X be a Banach space and let (x_i) be a sequence in X for which the series $\sum_{i=1}^\infty x_i$ converges absolutely, i.e.,

$$\sum_{i=1}^\infty \|x_i\| < \infty. \tag{3.1}$$

Then the infinite series $\sum_{i=1}^\infty x_i$ converges and

$$\left\|\sum_{i=1}^\infty x_i\right\| \le \sum_{i=1}^\infty \|x_i\|. \tag{3.2}$$

Proof. Let $n > m$ and consider the difference $s_n - s_m$ for the partial sums $s_n = \sum_{i=1}^n x_i$, then

$$\|s_n - s_m\| = \left\|\sum_{i=m+1}^n x_i\right\| \le \sum_{i=m+1}^n \|x_i\|.$$

The assumption (3.1) of absolute convergence implies that the sequence (s_n) is Cauchy in X, and hence converges since X is a Banach space. Taking the limit $n \to \infty$ in

$$\left\|\sum_{i=1}^n x_i\right\| \le \sum_{i=1}^n \|x_i\|,$$

gives the desired result. ☺

Proposition 3.14. Let X be a normed linear space with the property

$$\sum_{i=1}^\infty \|x_i\| < \infty \quad \Rightarrow \quad \sum_{i=1}^\infty x_i \quad \text{converges in} \quad X.$$

Then X is a Banach space.

Proof. Let (s_n) be a Cauchy sequence in X. Then there exists $N_1 \in \mathbb{N}$ such that for $m,n \geq N_1$ one has $\|s_n - s_m\| \leq 1/2$. Then there exists $N_2 \in \mathbb{N}$ which will be chosen such that $N_2 > N_1$ such that for $m,n \geq N_2$ one has $\|s_n - s_m\| \leq 1/2^2$. Continue by induction to obtain $N_i \in \mathbb{N}$ with $N_i > N_{i-1}$, such that for $m,n \geq N_i$ one has $\|s_n - s_m\| \leq 1/2^i$. Then define the elements

$$x_1 = s_{N_1}, \quad x_i = s_{N_i} - s_{N_{i-1}},$$

so that

$$s_{N_j} = \sum_{i=1}^{j} x_i \quad \text{with} \quad \|x_i\| \leq \frac{1}{2^{i-1}}.$$

Hence, by assumption, the subsequence (s_{N_j}) converges in X. Therefore the Cauchy sequence (s_n) has a convergent subsequence, which implies that the sequence (s_n) converges; cf. Remark 3.2. ☺

Proposition 3.15. Let $(X, \|\cdot\|)$ be a Banach space and let $V \subset X$ be a closed linear subspace. Then X/V is a Banach space.

Proof. Since X is a normed linear space and V is closed it follows from Proposition 2.25 that (2.7) defines a norm on X/V. It will be shown that X/V is complete, when X is a Banach space. Assume that the sequence $(x_i + V)$ of elements in X/V is absolutely convergent:

$$\sum_{i=1}^{\infty} \|(x_i + V)\| < \infty.$$

By Proposition 3.14 it suffices to show that $\sum_{i=1}^{\infty} (x_i + V)$ converges in X/V.

To verify this, observe that the norm $\|(x_i + V)\|$ is given by

$$\|x_i + V\| = \inf\{\|x_i - y\| : y \in V\}.$$

Hence there exists $y_i \in x_i + V$ so that

$$\|y_i\| \leq \|x_i + V\| + \frac{1}{2^i},$$

and it therefore follows that

$$\sum_{i=1}^{\infty} \|y_i\| < \infty.$$

Since X is a Banach space it follows from Theorem 3.13 that there is an element $x \in X$ such that

$$x = \sum_{i=1}^{\infty} y_i.$$

Now note that $y_i + V = x_i + V$ and that

$$\begin{aligned}
\left\|(x+V) - \sum_{i=1}^{n} (x_i + V)\right\| &= \left\|(x+V) - \sum_{i=1}^{n} (y_i + V)\right\| \\
&= \left\|\left(x - \sum_{i=1}^{n} y_i\right) + V\right\| \\
&\leq \left\|x - \sum_{i=1}^{n} y_i\right\| \to 0,
\end{aligned}$$

as $n \to \infty$, which completes the proof. ☺

3.2 Hilbert spaces

Definition 3.16. An inner product space $(X, \langle\cdot,\cdot\rangle)$ is called complete if the corresponding normed linear space is complete. A complete inner product space is called a *Hilbert space*.

Example 3.17. The linear space $\mathbb{K}^n$ with the inner product

$$\langle x,y\rangle = \sum_{i=1}^{n} x_i \bar{y}_i, \quad x,y \in \mathbb{K}^n.$$

is a Hilbert space. □

Example 3.18. The linear space ℓ^2 with the inner product

$$\langle x,y\rangle = \sum_{i=1}^{\infty} x_i \bar{y}_i, \quad x,y \in \ell^2.$$

is a Hilbert space. □

Example 3.19. The linear space $\mathcal{C}([a,b],\mathbb{K})$ with L^2-inner product

$$\langle f,g\rangle = \int_a^b f(t)\overline{g(t)}\,dt, \quad f,g \in \mathcal{C}([a,b],\mathbb{K}),$$

is not complete. □

Example 3.20. The linear space $\mathcal{C}^1([a,b],\mathbb{K})$ with L^2-inner product

$$\langle f,g\rangle = \int_a^b f(t)\overline{g(t)} + f'(t)\overline{g'(t)}\,dt, \quad f,g \in \mathcal{C}^1([a,b],\mathbb{K}),$$

is not complete. □

Lemma 3.21. Let X be a Hilbert space and let $V \subset X$ be a linear subspace. Then V is a Hilbert space if and only if V is closed in X.

Lemma 3.22. Let X and Y Hilbert spaces. Then the product space $X \times Y$ provided with the inner product (2.9) is a Hilbert space.

Theorem 3.23 (Existence and uniqueness of best approximations)**.** Let X be a Hilbert space and let $V \subset X$ be a nonempty subset which is closed and convex. For every $x \in X$ there exists a unique *best approximation* in V, namely there exists a unique $v \in V$ such that

$$\|x-v\| = d(x,V) = \inf\{\|x-z\| : z \in V\}.$$

Proof. *Existence*. Let $d = \inf\{\|x-z\| : z \in V\}$, so that $d \geq 0$. For each $n \in \mathbb{N}$ there exists $z_n \in V$ with

$$d^2 \leq \|x-z_n\|^2 < d^2 + \frac{1}{n}.$$

The parallelogram identity gives

$$\|(x-z_n)+(x-z_m)\|^2 + \|(x-z_n)-(x-z_m)\|^2 = 2\|x-z_n\|^2 + 2\|x-z_m\|^2,$$

which shows

$$\|2x-(z_n+z_m)\|^2 + \|z_n-z_m\|^2 < 4d^2 + \frac{2}{n} + \frac{2}{m}.$$

Now observe that
$$\|2x-(z_n+z_m)\| = 2\left\|x-\frac{z_n+z_m}{2}\right\| \geq 2d,$$
since $(z_n+z_m)/2 \in V$, as V is convex. One therefore concludes
$$\|z_n-z_m\|^2 < \frac{2}{n}+\frac{2}{m}.$$
Hence (z_n) is a Cauchy sequence so that $z_n \to v$ for some $v \in X$. As V is closed this shows that $v \in V$. It follows from the construction that $\|x-v\| = d$.

Uniqueness. Assume that $v' \in V$ has the property $\|x-v'\| = d$. Then the parallelogram identity gives
$$\|(x-v)+(x-v')\|^2 + \|(x-v)-(x-v')\|^2 = 2\|x-v\|^2 + 2\|x-v'\|^2 = 4d^2,$$
which shows
$$\|v-v'\|^2 = 4d^2 - \|(x-v)+(x-v')\|^2.$$
Observe that
$$\|(x-v)+(x-v')\| = 2\left\|x-\frac{v+v'}{2}\right\| \geq 2d,$$
since $(v+v')/2 \in V$, as V is convex. One therefore concludes
$$\|v-v'\|^2 \leq 0,$$
so that $v = v'$. ☺

Remark 3.24 (Best approximation)**.** Theorem 3.23 guarantees the existence of a best approximation to V, when V is a nonempty closed convex subset in a Hilbert space X. Recall from Lemma 2.55 that in the context of an inner product space X there is a unique best approximation to V when V is a finite-dimensional linear subspace. In a general normed linear space X both the existence and the uniqueness of best approximations to a closed linear subspace $V \subset X$ are not guaranteed. □

Theorem 3.25 (Orthogonal decomposition)**.** Let X be a Hilbert space and let $V \subset X$ be a closed linear subspace. For every $x \in X$ there exist unique elements $v \in V$ and $w \in V^\perp$ such that
$$x = v + w,$$
in which case the Pythagoras identity holds:
$$\|x\|^2 = \|v\|^2 + \|w\|^2.$$

Proof. The subspace V is linear and closed, so in particular closed and convex. Hence Theorem 3.23 may be applied.

Existence. For $x \in X$ there exists a unique $v \in V$ such that $\|x-v\| = d(x,V)$. Define $w = x-v$, then $x = v+w$ and $w \in V^\perp$ by Lemma 2.48.

Uniqueness. Assume that $x = v+w$ and that $x = v'+w'$ with $v, v' \in V$ and $w, w' \in V^\perp$. Then $v - v' = w' - w$ with $v-v' \in V$ and $w-w' \in V^\perp$. This shows $v = v'$ and $w = w'$.

Pythagoras identity. Note that with $x = v+w$, $v \in V$, $w \in V^\perp$ it follows that
$$\|x\|^2 = \langle v+w, v+w\rangle = \langle v,v\rangle + \langle v,w\rangle + \langle w,v\rangle + \langle w,w\rangle = \|v\|^2 + \|w\|^2.$$
☺

Corollary 3.26. Let X be a Hilbert space and let $V \subset X$ be a closed linear subspace. Then

$$V = (V^{\perp})^{\perp}.$$

In particular, if $E \subset X$ is any subset, then

$$\overline{\operatorname{span}}\,(E) = (E^{\perp})^{\perp}.$$

Proof. It is clear that $V \subset (V^{\perp})^{\perp}$. For the converse inclusion let $x \in (V^{\perp})^{\perp}$ and decompose $x = v + w$, $v \in V$, $w \in V^{\perp}$. Then $\langle x, w\rangle = 0$ and $\langle v, w\rangle = 0$, so that

$$0 = \langle x, w\rangle = \langle v + w, w\rangle = \langle w, w\rangle,$$

which shows $w = 0$ and thus $x = v \in V$. Hence $(V^{\perp})^{\perp} \subset V$.

For any subset $E \subset X$ one has $E^{\perp} = (\overline{\operatorname{span}}\,(E))^{\perp}$; cf. Lemma 2.47. This implies that

$$(E^{\perp})^{\perp} = ((\overline{\operatorname{span}}\,(E))^{\perp})^{\perp} = \overline{\operatorname{span}}\,(E),$$

since $\overline{\operatorname{span}}\,(E)$ is a closed linear subspace of X. ☺

Remark 3.27. Let V and W be linear subspaces of a Hilbert space X. Then

$$(V + W)^{\perp} = V^{\perp} \cap W^{\perp},$$

and, hence,

$$\overline{\operatorname{span}}\,(V + W) = (V^{\perp} \cap W^{\perp})^{\perp}.$$

Even if both V and W are closed, their linear span need not be closed; see for instance Example 5.23. □

3.3 Completion of normed linear spaces

Definition 3.28. Let X and Y be normed linear spaces. A linear map $T : X \to Y$ is called *isometric* if $\|Tx\| = \|x\|$ for all $x \in X$. The map T is called an *isometric isomorphism* from X to Y if T maps X isometrically onto Y.

If X and Y are inner product spaces, then a linear map $T : X \to Y$ is isometric if and only if $\langle Tx, Ty\rangle = \langle x, y\rangle$ for all $x, y \in X$. This is a straightforward consequence of the polarization identity.

Theorem 3.29. Let X and Y be Banach spaces and let $V \subset X$ and $W \subset Y$ be dense linear subspaces. Let T be an isometric isomorphism from V onto W. Then there is a unique isometric isomorphism $\dot{T}$ from X onto Y extending T.

Proof. Let $x \in X$ and let (v_n) be a sequence in V with $v_n \to x$. Then (v_n) is Cauchy in V and, hence (Tv_n) is Cauchy in Y. Since Y is complete it follows that there is $y \in Y$ such that $Tv_n \to y$. Define the operator T from X to Y by $Tx = y$. To see that T is well-defined, i.e., independent of the sequence (v_n), let (v_n') be another sequence such that $v_n' \to x$, in which case $Tv_n' \to y'$. Then $\|T(v_n - v_n')\| = \|v_n - v_n'\|$ and

$$T(v_n - v_n') \to y - y' \quad \text{and} \quad v_n - v_n' \to 0,$$

so that $y = y'$. It is then clear that T is linear. It follows from

$$\|v_n\| \to \|x\| \quad \text{and} \quad \|v_n\| = \|Tv_n\| \to \|y\|,$$

that T is isometric. To see that T is onto, let $y \in Y$. Then there is a sequence (w_n) in W with $w_n \to y$. For each $w_n \in W$ there is a unique $v_n \in V$ with $Tv_n = w_n$. The sequence (w_n) is Cauchy in Y and it follows that (v_n) is Cauchy in X. Since X is a Banach space there is $x \in X$ with $v_n \to x$. It is clear that $Tx = y$. ☺

Theorem 3.30 (Completion)**.** Let X be a normed linear space. Then there exists a complete normed linear space $\dot{X}$ and a linear map $\imath : X \to \dot{X}$, such that the normed linear spaces X and $\imath(X) \subset \dot{X}$ are isometrically isomorphic and $\imath(X)$ is dense in $\dot{X}$.

Proof. Construction of normed linear space. Let $\mathcal{X}$ be the collection of all Cauchy sequences $\mathbf{x} = (x_i)_{i=1}^{\infty}$ in X, and provide $\mathcal{X}$ with pointwise addition and scalar multiplication:

$$\mathbf{x}+\mathbf{y} = (x_1+y_1, x_2+y_2, \dots), \quad \lambda\mathbf{x} = (\lambda x_1, \lambda x_2, \dots).$$

Then it is clear that $\mathcal{X}$ is a linear space. Let $\mathcal{V}$ be the collection of all sequences $\mathbf{x} = (x_i)_{i=1}^{\infty}$ in X such that $x_i \to 0$. Then clearly $\mathcal{V}$ is a linear subspace of $\mathcal{X}$. With the above linear structure the quotient space $\mathcal{X}/\mathcal{V}$, consisting of all equivalence classes of the form $\mathbf{x}+\mathcal{V}$ is a linear space with

$$(\mathbf{x}+\mathcal{V}) + (\mathbf{y}+\mathcal{V}) = (\mathbf{x}+\mathbf{y}) + \mathcal{V}, \quad \lambda(\mathbf{x}+\mathcal{V}) = \lambda\mathbf{x}+\mathcal{V}.$$

Note that if $\mathbf{x} = (x_i)_{i=1}^{\infty}$ is a Cauchy sequence in X, then the sequence $(\|x_i\|)_{i=1}^{\infty}$ is Cauchy in $\mathbb{K}$ and hence converges in $\mathbb{K}$. Therefore

$$\|\mathbf{x}+\mathcal{V}\| = \lim_{i\to\infty} \|x_i\|$$

is well-defined and independent of the chosen representative. In fact, it is straightforward to check that $\|\cdot\|$ is a norm on $\mathcal{X}/\mathcal{V}$.

Embedding in normed linear space. Define the map $\imath : X \to \mathcal{X}/\mathcal{V}$ by

$$\imath(x) = (x,x,x,\dots) + \mathcal{V}, \quad x \in X.$$

This map is well-defined as the constant sequence involving $x \in X$ is a Cauchy sequence in X, which actually converges to x. Note that the map $\imath$ is linear, and isometric since

$$\|\imath(x)\| = \|(x,x,x,\dots)+\mathcal{V}\| = \|x\|.$$

Thus X and $\imath(X) \subset \mathcal{X}/\mathcal{V}$ are isometrically isomorphic.

Denseness of embedding. The range $\imath(X)$ is dense in $\mathcal{X}/\mathcal{V}$. To see this, let $\mathbf{x}+\mathcal{V}$ be an element of $\mathcal{X}/\mathcal{V}$ and let $\varepsilon > 0$. Since $\mathbf{x} = (x_i)_{i=1}^{\infty}$ is a Cauchy sequence in X, there exists $N \in \mathbb{N}$ such that $\|x_i - x_N\| < \varepsilon/2$ for all $i \geq N$. Let $y = x_N$ and consider the element

$$\imath(y) = (y,y,y,\dots) + \mathcal{V}, \quad y = x_N \in X.$$

Observe that $\imath(y) \in \imath(X)$ satisfies

$$\|(\mathbf{x}+\mathcal{V}) - \imath(y)\| = \lim_{i\to\infty} \|x_i - y\| < \varepsilon.$$

Hence it has been shown that $\imath(X)$ is dense in $\mathcal{X}/\mathcal{V}$.

Completeness of normed linear space. The normed linear space $\mathcal{X}/\mathcal{V}$ is complete. To see that, let $\mathbf{x}^n + \mathcal{V}$, $n \in \mathbb{N}$, be a Cauchy sequence in $\mathcal{X}/\mathcal{V}$. It suffices to show that there exists an element $\mathbf{z}+\mathcal{V}$ in $\mathcal{X}/\mathcal{V}$ such that

$$\|(\mathbf{x}^n+\mathcal{V}) - (\mathbf{z}+\mathcal{V})\| \to 0 \quad \text{as} \quad n \to \infty.$$

Since $\iota(X)$ is dense in $\mathcal{X}/\mathcal{V}$ there exists an element $z_n \in X$ for every $n \in \mathbb{N}$, such that

$$\|(\mathbf{x}^n + \mathcal{V}) - \iota(z_n)\| \leq \frac{1}{n}.$$

In terms of the elements $\iota(z_n)$ observe that

$$\begin{aligned}\|\iota(z_n) - \iota(z_m)\| \leq \|\iota(z_n) - (\mathbf{x}^n + \mathcal{V})\| \\ + \|(\mathbf{x}^n + \mathcal{V}) - (\mathbf{x}^m + \mathcal{V})\| + \|(\mathbf{x}^m + \mathcal{V}) - \iota(z_m)\|.\end{aligned}$$

The construction of the limit $\mathbf{z} + \mathcal{V}$ takes place via the elements $z_n \in X$.

Choose $\varepsilon > 0$. Then there exists $M \in \mathbb{N}$ such that for $n, m \geq M$ each term in the righthand side of the above inequality is bounded by $\varepsilon/6$. Hence the sequence $(\iota(z_n))$ is also Cauchy in $\mathcal{X}/\mathcal{V}$. Now recall that

$$\|z_n - z_m\| = \|\iota(z_n) - \iota(z_m)\|,$$

which is $< \varepsilon/2$ for $n, m \geq M$. Therefore the sequence (z_n) is Cauchy in X. Define the element $\mathbf{z}$ by

$$\mathbf{z} = (z_1, z_2, \ldots),$$

then $\mathbf{z} \in \mathcal{X}$ and consider the inequality

$$\|(\mathbf{x}^n + \mathcal{V}) - (\mathbf{z} + \mathcal{V})\| \leq \|(\mathbf{x}^n + \mathcal{V}) - \iota(z_n)\| + \|\iota(z_n) - (\mathbf{z} + \mathcal{V})\|.$$

For $n \geq M$ the first term on the righthand side is bounded by $\varepsilon/6$, while

$$\|\iota(z_n) - (\mathbf{z} + \mathcal{V})\| = \lim_{i \to \infty} \|z_n - z_i\| \leq \varepsilon/2.$$

Hence for $n \geq M$ one obtains

$$\|(\mathbf{x}^n + \mathcal{V}) - (\mathbf{z} + \mathcal{V})\| < \varepsilon.$$

In other words, it has been shown that $(\mathbf{x}^n + \mathcal{V}) \to (\mathbf{z} + \mathcal{V})$.

It is clear that the space $\dot{X} = \mathcal{X}/\mathcal{V}$ and the map $\iota : X \to \dot{X} = \mathcal{X}/\mathcal{V}$ have all the announced properties. ☺

Remark 3.31. The uniqueness of the construction of a completion will now be addressed. Assume that there is a normed linear space X and that there are two Banach spaces $\dot{X}_1$ and $\dot{X}_2$ and two isometries ι_1 and ι_2 such that

$$\iota_1 : X \to \iota_1(X) \subset \dot{X}_1, \quad \iota_2 : X \to \iota_2(X) \subset \dot{X}_2,$$

such that $\iota_1(X)$ is dense in $\dot{X}_1$ and $\iota_2(X)$ is dense in $\dot{X}_2$. Then the map $\iota_1 \circ (\iota_2)^{-1}$ is an isometry from $\iota_2(X)$ onto $\iota_1(X)$. This isometry can be uniquely extended to an isometric isomorphism from $\dot{X}_2$ onto $\dot{X}_1$; cf. Theorem 3.29. □

Corollary 3.32. Let X be an inner product space. Then there exists a Hilbert space $\dot{X}$ and a linear map $\iota : X \to \dot{X}$, such that the inner product spaces X and $\iota(X) \subset \dot{X}$ are isometrically isomorphic and $\iota(X)$ is dense in $\dot{X}$.

Proof. Let $\mathbf{x} = (x_i)_{i=1}^{\infty}$ and $\mathbf{y} = (y_i)_{i=1}^{\infty}$ be Cauchy sequences in X. Then the sequence $\langle x_i, y_i \rangle$ converges in $\mathbb{K}$. To see this, note that $\|x_i\|$ and $\|y_i\|$ are Cauchy sequences; therefore they converge and, in particular, they are bounded. Therefore the estimate

$$|\langle x_i, y_i \rangle - \langle x_j, y_j \rangle| \leq \|x_i\| \|y_i - y_j\| + \|x_i - x_j\| \|y_j\|,$$

shows that $(\langle x_i, y_i\rangle)$ is a Cauchy sequence in $\mathbb{K}$, so that it converges in $\mathbb{K}$. Therefore

$$\langle \mathbf{x}+\mathcal{V}, \mathbf{y}+\mathcal{V}\rangle = \lim_{i\to\infty}\langle x_i, y_i\rangle,$$

is well-defined and independent of the chosen representatives. In fact, it is straightforward to check that $\langle\cdot,\cdot\rangle$ is an inner product on $\mathcal{X}/\mathcal{V}$, and it induces the norm $\|\cdot\|$ on $\mathcal{X}/\mathcal{V}$. The map $\iota : X \to \mathcal{X}/\mathcal{V}$ by

$$\iota(x) = (x,x,x,\dots)+\mathcal{V}, \quad x \in X.$$

is isometric, since

$$\langle \iota(x), \iota(y)\rangle = \langle (x,x,x,\dots)+\mathcal{V}, (y,y,y,\dots)+\mathcal{V}\rangle = \langle x, y\rangle.$$

Thus the inner product spaces X and $\iota(X) \subset \mathcal{X}/\mathcal{V}$ are isometrically isomorphic. ☺

Corollary 3.33. Let X be a normed linear space and let $E \subset X$ be a dense subset. Let $\dot{X}$ be a completion of X with a linear map $\iota : X \to \dot{X}$, such that the normed linear spaces X and $\iota(X) \subset \dot{X}$ are isometrically isomorphic. Then $\iota(E)$ is dense in $\dot{X}$.

Remark 3.34. In Example 2.12 it has been argued that the linear space $\mathcal{C}([a,b],\mathbb{K})$ is a normed linear space when it is provided with the L^p-norm. However, this normed linear space is not complete. The completion procedure provides for each $p \geq 1$ a Banach space completion and for $p = 2$ a Hilbert space completion of the space $\mathcal{C}([a,b],\mathbb{K})$. These completions are uniquely defined up to isometric isomorphisms. In the theory of measure and integration one is able to give a concrete identification for the completion. This particular identification shows, roughly speaking, that the completion $L^p(a,b)$ consists of (equivalence classes of) Lebesgue integrable functions whose norm is calculated by the same formulas as above but now in terms of the Lebesgue integral (which coincides with the Riemann integral on $\mathcal{C}([a,b],\mathbb{K})$). □

3.4 A brief encounter with measure and integration

The purpose of this section is to give a very brief introduction to integration on measure spaces. At the end of this section it will be shown how this applies to the context of real-valued functions defined on (subsets of) $\mathbb{R}^d$.

Definition 3.35. Let Ω be an arbitrary set. Then a collection $\mathcal{A}$ of subsets of a set Ω is called a *σ-algebra* if

1. $\Omega \in \mathcal{A}$;
2. $A \in \mathcal{A} \Rightarrow A^c \in \mathcal{A}$;
3. $A_n \in \mathcal{A}$, $n \in \mathbb{N} \Rightarrow \bigcup_{n=1}^{\infty} A_n \in \mathcal{A}$.

A *measurable space* $(\Omega, \mathcal{A})$ is a set Ω provided with a σ-algebra $\mathcal{A}$ on Ω. The elements of $\mathcal{A}$ are called *measurable sets*. A *measure* μ on a σ-algebra $\mathcal{A}$ is a map $\mu : \mathcal{A} \to [0,\infty]$ which satisfies

1. $\mu(\emptyset) = 0$;
2. $A_n \in \mathcal{A}$, $n \in \mathbb{N}$, pairwise disjoint $\Rightarrow \mu\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \mu(A_n)$.

The notions of σ-algebra and measure are fundamental in the development of the integration theory. The null sets of $\mathcal{A}$ are those $A \in \mathcal{A}$ for which $\mu(A) = 0$ (sets of measure 0). It is clear that countable unions of sets of measure 0 are measurable and have measure 0.

In the following the notation $\overline{\mathbb{R}} = \mathbb{R} \cup \{\infty\} \cup \{-\infty\}$ (the extended real line) will be standard.

Definition 3.36. Let $(\Omega, \mathcal{A}, \mu)$ be a measurable space. A map $f : (\Omega, \mathcal{A}) \to \overline{\mathbb{R}}$ is *measurable* if and only if for each $c \in \mathbb{R}$ one of the following (equivalent) assertions are satisfied:

1. $\{\omega \in \Omega : f(\omega) < c\} \in \mathcal{A}$;
2. $\{\omega \in \Omega : f(\omega) \leq c\} \in \mathcal{A}$;
3. $\{\omega \in \Omega : f(\omega) > c\} \in \mathcal{A}$;
4. $\{\omega \in \Omega : f(\omega) \geq c\} \in \mathcal{A}$.

The measurable functions form the core of integration theory. Many operations stay closed under measurability. For instance, if the function f is measurable, then also its positive and negative parts

$$f^+ = \max(f, 0), \quad f^- = \max(-f, 0),$$

are measurable functions.

A function $(\Omega, \mathcal{A}) \to \mathbb{R}$ is called *simple* if it is measurable and takes on only a finite number of values. If f is simple and $\alpha_1, \ldots, \alpha_r \in \mathbb{R}$ are its values then the sets

$$A_k = f^{-1}(\alpha_k), \quad k = 1, \ldots, r,$$

are measurable and pairwise disjoint. The function f can be written as

$$f = \sum_{k=1}^{r} \alpha_k \mathbf{1}_{A_k}.$$

The integral $\int_\Omega f \, d\mu$ of f over Ω with respect to μ, is defined by

$$\int_\Omega f \, d\mu = \sum_{k=1}^{r} \alpha_k \mu(A_k).$$

Definition 3.37. Let $(\Omega, \mathcal{A}, \mu)$ be a measure space and let $f : (\Omega, \mathcal{A}) \to \overline{\mathbb{R}}$ be measurable. If f is nonnegative, the integral $\int_\Omega f \, d\mu$ is defined by

$$\int_\Omega f \, d\mu = \sup \left\{ \int_\Omega g \, d\mu : 0 \leq g \leq f, \, g \text{ simple} \right\} \in [0, \infty].$$

In general the *integral* $\int_\Omega f \, d\mu$ is defined by

$$\int_\Omega f \, d\mu = \int_\Omega f^+ \, d\mu - \int_\Omega f^- \, d\mu,$$

if at least one of the integrals in the right-hand side is finite. The function f is said to be integrable if both integrals in the right-hand side are finite, or, equivalently, if

$$\int_\Omega |f| \, d\mu < \infty.$$

The usual operations are all justified: if f and g are integrable then the sum $f + g$ is integrable and $\int_\Omega (f + g) \, d\mu = \int_\Omega f \, d\mu + \int_\Omega g \, d\mu$ and, furthermore, the important inequality $|\int_\Omega f \, d\mu| \leq \int_\Omega |f| \, d\mu$.

Theorem 3.38 (Monotone convergence theorem)**.** Let $(\Omega, \mathcal{A}, \mu)$ be a measure space. Let f_n, $n \in \mathbb{N}$, and f be nonnegative measurable functions from Ω to $[0, \infty]$, such that

$$f_n(\omega) \to f(\omega), \quad f_n(\omega) \leq f_{n+1}(\omega), \quad \omega \in \Omega.$$

Then

$$\lim_{n\to\infty}\int_\Omega f_n\,d\mu = \int_\Omega f\,d\mu.$$

A property is said to hold *almost everywhere* (denoted by a.e.), if the complement of the set where the property does not hold is measurable and has measure zero. For instance, $f = g$ a.e. means that the set

$$\{\omega \in \Omega : f(\omega) \neq g(\omega)\}$$

is measurable and has measure zero.

Theorem 3.39 (Dominated convergence theorem)**.** Let $(\Omega, \mathcal{A}, \mu)$ be a measure space. Let f_n, $n \in \mathbb{N}$, and f be measurable functions from Ω to $\overline{\mathbb{R}}$, and let the function $g : \Omega \to [0,\infty]$ be integrable, such that

$$f_n(\omega) \to f(\omega), \quad |f_n(\omega)| \leq g(\omega), \quad \text{for almost all} \quad \omega \in \Omega. \tag{3.3}$$

Then f_n and f are integrable and

$$\lim_{n\to\infty}\int_\Omega |f_n - f|\,d\mu = 0, \quad \lim_{n\to\infty}\int_\Omega f_n\,d\mu = \int_\Omega f\,d\mu. \tag{3.4}$$

In the present context one has again Hölder and Minkowski inequalities. They ultimately lead to the following definition (which in the case of counting measure on $\Omega = \mathbb{N}$ gives the space ℓ^p).

Definition 3.40. Let $(\Omega, \mathcal{A}, \mu)$ be a measure space and let $1 \leq p < \infty$. The space $\mathcal{L}^p(\Omega)$ is the set of all complex measurable functions f for which

$$\int_\Omega |f|^p\,d\mu < \infty. \tag{3.5}$$

Furthermore, one defines for $f \in \mathcal{L}^p(\Omega)$

$$\|f\|_p = \left(\int_\Omega |f|^p\,d\mu\right)^{1/p}. \tag{3.6}$$

Note that $\|\cdot\|_p$ is not a norm on the linear space $\mathcal{L}^p(\Omega)$, but a semi-norm. To deal with this situation define the set $\mathcal{N} = \{f \in \mathcal{L}^p(\Omega) : \|f\|_p = 0\}$. Then $\mathcal{N}$ is a linear subspace and

$$\|f\|_p = 0 \quad \Leftrightarrow \quad f(\omega) = 0 \text{ for almost all } \omega \in \Omega.$$

The linear space $\mathcal{N}$ sets up an equivalence relation in $\mathcal{L}^p(\Omega)$; actually by definition the semi-norm induces a norm on each equivalence class (the norm of the equivalence class is the semi-norm of any element in the equivalence class). In every day practice one does not distinguish between an equivalence class and any of its elements.

Theorem 3.41. The quotient space $L^p(\Omega) = \mathcal{L}^p(\Omega)/\mathcal{N}$ is a Banach space.

Very often one needs measure spaces with the following useful property. A measure space $(\Omega, \mathcal{A}, \mu)$ is called *σ-finite* if there is a sequence $\Omega_n \in \mathcal{A}$ such that

$$\mu(\Omega_n) < \infty, \quad \Omega = \bigcup_{n=1}^\infty \Omega_n. \tag{3.7}$$

One place where these spaces appear is in the construction of product measures. Let $(\Omega_1, \mathcal{A}_1, \mu_1)$ and $(\Omega_2, \mathcal{A}_2, \mu_2)$ be σ-finite measure spaces. Then there exists a measurable space $\mathcal{A}_1 \otimes \mathcal{A}_2$

containing the products $A_1 \times A_2$, where $A_1 \in \mathcal{A}_1, A_1 \in \mathcal{A}_1$. If a function f on $\Omega_1 \times \Omega_2$ is measurable with respect to $\mathcal{A}_1 \otimes \mathcal{A}_2$, then its sections f_{ω_1} and f^{ω_1} defined by

$$f_{\omega_1}(\omega_2) = f(\omega_1, \omega_2), \quad f^{\omega_2}(\omega_1) = f(\omega_1, \omega_2),$$

are measurable with respect to $\mathcal{A}_2$ and $\mathcal{A}_1$, respectively. Moreover there is a so-called *product measure* $\mu_1 \otimes \mu_2$ on $\mathcal{A}_1 \otimes \mathcal{A}_2$ such that

$$\mu_1 \otimes \mu_2(A_1 \times A_2) = \mu_1(A_1)\mu_2(A_2)$$

when $A_1 \in \mathcal{A}_1, A_2 \in \mathcal{A}_2$. Assume that $f : \Omega_1 \times \Omega_2 \to \overline{\mathbb{R}}$ measurable. Then the following integrals are all equal:

$$\int_{\Omega_1 \times \Omega_2} |f| \, d(\mu_1 \otimes \mu_2), \quad \int_{\Omega_1} \left(\int_{\Omega_2} |f_{\omega_1}| \, d\mu_2 \right) d\mu_1(\omega_1), \quad \int_{\Omega_2} \left(\int_{\Omega_1} |f^{\omega_2}| \, d\mu_1 \right) d\mu_2(\omega_2).$$

In particular, they are simultaneously finite. Now the following Fubini theorem is available (stated with some abuse of language since the integrals over the sections are only defined almost everywhere).

Theorem 3.42 (Fubini)**.** Assume that $f : (\Omega_1 \times \Omega_2, \mathcal{A}_1 \otimes \mathcal{A}_2) \to \overline{\mathbb{R}}$ is an integrable function. Then

$$\begin{aligned} \int_{\Omega_1 \times \Omega_2} f \, d(\mu_1 \otimes \mu_2) &= \int_{\Omega_1} \left(\int_{\Omega_2} f_{\omega_1} \, d\mu_2 \right) d\mu_1(\omega_1) \\ &= \int_{\Omega_2} \left(\int_{\Omega_1} f^{\omega_2} \, d\mu_1 \right) d\mu_2(\omega_2). \end{aligned}$$

With the abstract theory taken care of, now specialize to $\Omega = \mathbb{R}^d$. The σ-algebra $\mathcal{B}^d$ of Borel sets of $\mathbb{R}^d$ is the smallest σ-algebra which contains the open blocks (products of intervals) of $\mathbb{R}^d$. The Borel measure m_d is the measure on $\mathcal{B}^d$ which coincides with volume (length when $d = 1$) on open blocks. It can be shown that there exist subsets of $\mathbb{R}^d$ which are not Borel measurable. However most useful subsets of $\mathbb{R}^d$ are Borel measurable. For instance, every closed block in $\mathbb{R}^d$ is Borel measurable. The continuous real-valued functions on $\mathbb{R}^d$ are Borel measurable. But, for instance, also products of continuous functions and characteristic functions of Borel measurable subsets are Borel measurable. As to the Fubini theorem: it can be shown that $\mathcal{B}^{p+q} = \mathcal{B}^p \otimes \mathcal{B}^q$ and that $m_{p+q} = m_p \otimes m_q$.

First the one-dimensional case will be discussed. Let $\Omega = [a,b]$ be a compact interval. In a similar way as above one can define Borel measurable sets on $[a,b]$ and Borel measurable functions $f : [a,b] \to \mathbb{R}$. The above abstract integral coincides with the Riemann integral for step functions, continuous functions, and continuous functions with a finite number of discontinuities. To see this, let $f : [a,b] \to \mathbb{R}$ be a continuous function. Then the functions

$$\sum_{i=1}^{n} f(t_{i-1}) \mathbf{1}_{[t_{i-1}, t_i]}$$

tend to f pointwise on $[a,b]$ when the partition $P = \{a = t_0 < t_1 < \cdots < t_n = b\}$ goes to zero. Since f is bounded on $[a,b]$ the corresponding Riemann sums

$$\sum_{i=1}^{n} f(t_i)(t_i - t_{i-1}) = \int_a^b \left(\sum_{i=1}^{n} f(t_{i-1}) \mathbf{1}_{[t_{i-1}, t_i]}(t) \right) dt$$

are bounded and converge to $\int_a^b f(t) \, dt$ when the partition P goes to zero. This is a direct consequence of the dominated convergence theorem.

Let the linear space $\mathcal{C}([a,b],\mathbb{R})$ be provided with the L^p-norm

$$\|f\|_p = \left(\int_a^b |f(t)|^p \, dt \right)^{1/p}, \quad f \in \mathcal{C}([a,b],\mathbb{K}), \quad p \geq 1,$$

and with the L^2-inner product

$$\langle f, g \rangle = \int_a^b f(t)\overline{g(t)} \, dt, \quad f, g \in \mathcal{C}([a,b],\mathbb{K}), \quad p = 2.$$

These expressions define a norm on $\mathcal{C}([a,b],\mathbb{R})$, and for $p=2$ an inner product on $\mathcal{C}([a,b],\mathbb{R})$. The linear space $\mathcal{C}([a,b],\mathbb{R})$ provided with such a norm is not complete. The space $\mathcal{C}([a,b],\mathbb{K})$, provided with the L^p-norm, is dense in the Banach space $L^p([a,b])$, which is a Hilbert space for $p=2$.

Next the case $\Omega = \mathbb{R}^d$ will be discussed briefly. A function $f : \mathbb{R}^d \to \mathbb{R}$ is said to have *compact support* if the support, i.e. the closure of the set $\{x \in \mathbb{R}^d : f(x) \neq 0\}$ is bounded. Define $\mathcal{C}_c(\mathbb{R}^d,\mathbb{R})$ as the collection of all functions on $\mathbb{R}^d$ which are continuous and which have compact support. It is clear that $\mathcal{C}_c(\mathbb{R}^d,\mathbb{R})$ is a linear space. The abstract integral defined above coincides with the Riemann integral on $\mathcal{C}_c(\mathbb{R}^d,\mathbb{R})$. Similar to the one-dimensional case one sees that the space $\mathcal{C}_c(\mathbb{R}^d,\mathbb{R})$, provided with the L^p-norm, is dense in the Banach space $L^p(\mathbb{R}^d)$, which is a Hilbert space for $p=2$.

3.5 Orthonormal bases in Hilbert spaces

Lemma 3.43 (Bessel's inequality)**.** Let X be an inner product space with an orthonormal system e_i, $i \in \mathbb{N}$. Then

$$\sum_{i=1}^{\infty} |\langle x, e_i \rangle|^2 \leq \|x\|^2, \quad x \in X.$$

Proof. Let $n \in \mathbb{N}$ be arbitrary. Expanding the norm in terms of the inner product gives

$$0 \leq \left\| x - \sum_{i=1}^{n} \langle x, e_i \rangle e_i \right\|^2 = \|x\|^2 - \sum_{i=1}^{n} |\langle x, e_i \rangle|^2.$$

The assertion now follows by taking $n \to \infty$. ☺

Theorem 3.44. Let X be a Hilbert space with an orthonormal system e_i, $i \in \mathbb{N}$. Then $\sum_{i=1}^{\infty} \lambda_i e_i$ converges if and only if

$$\sum_{i=1}^{\infty} |\lambda_i|^2 < \infty.$$

In this case

$$\left\| \sum_{i=1}^{\infty} \lambda_i e_i \right\|^2 = \sum_{i=1}^{\infty} |\lambda_i|^2.$$

Proof. ($\Rightarrow$) Assume that $\sum_{i=1}^{\infty} \lambda_i e_i$ converges to $x \in X$. Then

$$\langle x, e_m \rangle = \lim_{n \to \infty} \left\langle \sum_{i=1}^{n} \lambda_i e_i, e_m \right\rangle = \lambda_m,$$

which leads to

$$\sum_{i=1}^{\infty} |\lambda_i|^2 = \sum_{i=1}^{\infty} |\langle x, e_i \rangle|^2 \leq \|x\|^2$$

by Bessel's inequality.

($\Leftarrow$) Assume that $\sum_{i=1}^{\infty} |\lambda_i|^2 < \infty$. Define $x_n = \sum_{i=1}^{n} \lambda_i e_i$, then for $n > m$

$$\|x_n - x_m\|^2 = \left\| \sum_{i=m+1}^{n} \lambda_i e_i \right\|^2 = \sum_{i=m+1}^{n} |\lambda_i|^2.$$

Hence (x_n) is a Cauchy sequence in the Hilbert space X, thus there exists $x \in X$ such that $x_n \to x$. This shows that $\sum_{i=1}^{\infty} \lambda_i e_i$ converges.

If either condition is satisfied one sees

$$\left\| \sum_{i=1}^{\infty} \lambda_i e_i \right\|^2 = \lim_{n\to\infty} \left\| \sum_{i=1}^{n} \lambda_i e_i \right\|^2 = \lim_{n\to\infty} \sum_{i=1}^{n} |\lambda_i|^2 = \sum_{i=1}^{\infty} |\lambda_i|^2,$$

which completes the proof. ☺

Corollary 3.45. Let X be a Hilbert space with an orthonormal system e_i, $i \in \mathbb{N}$. Then

$$\sum_{i=1}^{\infty} \langle x, e_i \rangle e_i$$

converges for all $x \in X$.

Proof. It follows from Bessel's inequality that

$$\sum_{i=1}^{\infty} |\langle x, e_i \rangle|^2 \leq \|x\|^2, \quad x \in X.$$

Hence Theorem 3.44 implies that $\sum_{i=1}^{\infty} \langle x, e_i \rangle e_i$ converges in X. ☺

Theorem 3.46. Let X be a Hilbert space and let e_i, $i \in \mathbb{N}$, be an orthonormal system. Then the following statements are equivalent:

1. $\{e_i : i \in \mathbb{N}\}^{\perp} = \{0\}$;
2. $\overline{\operatorname{span}}\{e_i : i \in \mathbb{N}\} = X$;
3. $\|x\|^2 = \sum_{i=1}^{\infty} |\langle x, e_i \rangle|^2$ for all $x \in X$;
4. $x = \sum_{i=1}^{\infty} \langle x, e_i \rangle e_i$ for all $x \in X$.

Proof. The proof will be given as follows:

$$\begin{array}{ccccc} & & \Rightarrow \ (2) \ \Rightarrow & \\ (1) \ \Rightarrow \ (4) & & & & (1) \\ & & \Rightarrow \ (3) \ \Rightarrow & \end{array}$$

(1) $\Rightarrow$ (4) Let $x \in X$ and $y = x - \sum_{i=1}^{\infty} \langle x, e_i \rangle e_i$. Then

$$\langle y, e_m \rangle = \langle x, e_m \rangle - \langle x, e_m \rangle = 0$$

for all $m \in \mathbb{N}$. Hence it follows that $y = 0$.

(4) $\Rightarrow$ (2) Let $x \in X$ then $x = \lim_{n\to\infty} \sum_{i=1}^{n} \langle x, e_i \rangle e_i$. Thus

$$x \in \overline{\operatorname{span}}\{e_i : i \in \mathbb{N}\}$$

which shows (2).

(4) $\Rightarrow$ (3) For each $x \in X$ one has $x = \lim_{n\to\infty} \sum_{i=1}^{n} \langle x, e_i \rangle e_i$. Thus

$$\|x\|^2 = \lim_{n\to\infty} \left\| \sum_{i=1}^{n} \langle x, e_i \rangle e_i \right\|^2 = \lim_{n\to\infty} \sum_{i=1}^{n} |\langle x, e_i \rangle|^2.$$

(2) $\Rightarrow$ (1) Observe that

$$\{e_i : i \in \mathbb{N}\}^\perp = (\overline{\operatorname{span}}\{e_i : i \in \mathbb{N}\})^\perp = X^\perp = \{0\}.$$

(3) $\Rightarrow$ (1) If $x \perp e_i$ then $\langle x, e_i \rangle = 0$, $i \in \mathbb{N}$. Thus (3) implies that $x = 0$. ☺

Definition 3.47. Let X be a Hilbert space with an orthonormal system e_i, $i \in \mathbb{N}$. The system is called an *orthonormal basis* for X if $\overline{\operatorname{span}}\{e_i : i \in \mathbb{N}\} = X$.

Theorem 3.48. Let X be an infinite-dimensional Hilbert space. Then X has an orthonormal basis if and only if X is separable.

Proof. ($\Rightarrow$) Let e_i, $i \in \mathbb{N}$, be an orthonormal basis. Then it follows from $\overline{\operatorname{span}}\{e_i : i \in \mathbb{N}\} = X$ and Lemma 2.28 that X is separable.

($\Leftarrow$) The space X contains a countable dense subset E, so that $\overline{\operatorname{span}}(E) = X$; see (2.8). From the elements $x_i \in E$ one constructs inductively a subsequence denoted by (y_i) which is linearly independent. Denote the collection of these elements by F. Then $F \subset E$, $\operatorname{span}(F) = \operatorname{span}(E)$, and $\overline{\operatorname{span}}(F) = \overline{\operatorname{span}}(E) = X$. Now apply the Gram-Schmidt procedure to the elements $y_i \in F$. ☺

Remark 3.49. Let X be a Hilbert space with an orthonormal basis e_i, $i \in \mathbb{N}$. Then the linear map $T : X \to \ell^2$ defined by

$$Tx = (\langle x, e_1 \rangle, \langle x, e_2 \rangle, \dots)$$

is an isometric isomorphism from X onto ℓ^2, see Definition 3.28. In particular, according to Theorem 3.48, all separable Hilbert spaces are isometrically isomorphic. □

3.6 Fourier series

The abstract theory of orthonormal bases in Hilbert spaces will now be applied to the theory of Fourier series. A *Fourier series* is a formal expression of the form

$$\frac{a_0}{2} + \sum_{k=1}^{\infty} (a_k \cos kx + b_k \sin kx), \quad x \in [-\pi, \pi], \tag{3.8}$$

where the coefficients a_k and b_k are real. It is clear that if

$$\sum_{k=1}^{\infty} |a_k| < \infty, \quad \sum_{k=1}^{\infty} |b_k| < \infty, \tag{3.9}$$

then the series in (3.8) converges uniformly on $[-\pi, \pi]$ and defines a continuous 2π-periodic function $f : [-\pi, \pi] \to \mathbb{R}$. Moreover, due to the trigonometric identities

$$\int_{-\pi}^{\pi} \cos kt \, \sin nt \, dt = 0, \quad \int_{-\pi}^{\pi} \cos kt \, \cos nt \, dt = \int_{-\pi}^{\pi} \sin kt \, \sin nt \, dt = \pi \delta_{kn},$$

it is clear that a_k and b_k can be expressed in terms of f:

$$a_k = \frac{1}{\pi}\int_{-\pi}^{\pi} f(t)\cos kt\,dt, \quad b_k = \frac{1}{\pi}\int_{-\pi}^{\pi} f(t)\sin kt\,dt. \tag{3.10}$$

Conversely, given a 2π-periodic function f and coefficients a_k and b_k determined by (3.10), one may ask under which conditions the corresponding Fourier series converges to f. Appendix B gives some sufficient conditions for pointwise and uniform convergence in terms of the smoothness properties of f. However, in many applications the conditions on the coefficients (3.9) or the smoothness of f are too restrictive. Alternatively, recalling that $\ell^1 \subset \ell^2$, the more general conditions

$$\sum_{k=1}^{\infty} |a_k|^2 < \infty, \quad \sum_{k=1}^{\infty} |b_k|^2 < \infty, \tag{3.11}$$

lead to a Hilbert space approach. The price to pay for this point of view is that the convergence of the Fourier series then takes place with respect to the norm of the Hilbert space $L^2(-\pi,\pi)$.

Theorem 3.50 (Fourier series; real form)**.** The system e_n, $n \in \mathbb{N} \cup \{0\}$, defined by

$$e_0(t) = \frac{1}{\sqrt{2\pi}}, \quad e_{2n-1}(t) = \frac{1}{\sqrt{\pi}}\cos nt, \quad e_{2n}(t) = \frac{1}{\sqrt{\pi}}\sin nt, \quad n \geq 1,$$

is a complete orthonormal system in the Hilbert space $L^2(-\pi,\pi)$.

Proof. It follows from the trigonometric formulas that the system is orthonormal. Let $f \in L^2(-\pi,\pi)$ and let $\varepsilon > 0$. Then there is $g \in \mathcal{C}([-\pi,\pi],\mathbb{K})$ such that $\|f-g\| < \varepsilon/2$. It may be assumed that $g(-\pi) = g(\pi)$. By the Weierstrass theorem there exists a trigonometric polynomial P such that

$$\sup\{|g(t)-P(t)| : t \in [-\pi,\pi]\} < \frac{\varepsilon}{2\sqrt{2\pi}},$$

which implies in terms of the L^2-norm that $\|g-P\| < \varepsilon/2$; cf. Example 2.19. Therefore

$$\|f-P\| \leq \|f-g\| + \|g-P\| < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon.$$

The result now follows from Theorem 3.46. ☺

Corollary 3.51 (Fourier series; complex form)**.** The system $\widetilde{e}_n$, $n \in \mathbb{Z}$, defined by

$$\widetilde{e}_n(t) = \frac{1}{\sqrt{2\pi}} e^{int}, \quad t \in [-\pi,\pi],$$

is a complete orthonormal system in $L^2(-\pi,\pi)$ over $\mathbb{C}$.

It is clear that every $f \in L^2(-\pi,\pi)$ can be written as the sum of an even and an odd function: $f = f_{\mathrm{e}} + f_{\mathrm{o}}$, where

$$f_{\mathrm{e}}(t) = \frac{f(t)+f(-t)}{2}, \quad f_{\mathrm{o}}(t) = \frac{f(t)-f(-t)}{2}.$$

Hence there is the orthogonal decomposition

$$L^2(-\pi,\pi) = L^2_{\mathrm{e}}(-\pi,\pi) \oplus L^2_{\mathrm{o}}(-\pi,\pi),$$

into even and odd square integrable functions. The Fourier expansion of even and odd functions via Theorem 3.50 involves only cosine functions and only sine functions, respectively. Furthermore, one sees

$$\int_{-\pi}^{\pi} f(t)\frac{1}{\sqrt{\pi}}\cos nt\,dt = \int_{0}^{\pi} f(t)\frac{2}{\sqrt{\pi}}\cos nt\,dt, \quad f \in L^2_{\mathrm{e}}(-\pi,\pi)$$

and

$$\int_{-\pi}^{\pi} f(t)\frac{1}{\sqrt{\pi}}\sin nt\,dt = \int_{0}^{\pi} f(t)\frac{2}{\sqrt{\pi}}\sin nt\,dt, \quad f \in L^2_{\mathrm{o}}(-\pi,\pi)$$

Each of these cases leads to a complete orthonormal system in $L^2(0,\pi)$; they give rise to "half range expansions" as they are sometimes called in mathematical physics texts.

Corollary 3.52. The system c_n, $n \in \mathbb{N} \cup \{0\}$, defined by

$$c_0(t) = \sqrt{\frac{1}{\pi}}, \quad c_n(t) = \sqrt{\frac{2}{\pi}}\cos nt, \quad n \geq 1, \quad t \in [-\pi,\pi],$$

is a complete orthonormal system in $L^2(0,\pi)$.

Corollary 3.53. The system s_n, $n \in \mathbb{N}$, defined by

$$s_n(t) = \sqrt{\frac{2}{\pi}}\sin nt, \quad t \in [-\pi,\pi],$$

is a complete orthonormal system in $L^2(0,\pi)$.

Chapter 4

Bounded linear operators in normed linear spaces

Among all operators between normed linear spaces the linear ones have the useful property that continuity in one point, continuity everywhere, uniform continuity, or boundedness are all equivalent. The spaces of such operators can be given the structure of normed linear spaces which are complete depending on the underlying space. The invertibility of a bounded linear operator means that it is bijective and that the inverse operator is bounded. By means of the geometric series (in the scalar case) one develops useful criteria for the invertibility of an operator which perturbs the identity in some sense. Particular attention is paid to a special class of bounded linear operators, namely the compact linear operators, which are very close to linear operators in finite-dimensional spaces. The classical Fredholm and Volterra integral operators are studied as compact operators.

4.1 Continuity and boundedness of linear operators

Lemma 4.1. Let X and Y be normed linear spaces and let $T : X \to Y$ be a linear operator. Then the following statements are equivalent:

1. T is uniformly continuous;
2. T is continuous;
3. T is continuous at 0;
4. for some $K > 0$ and for all $x \in X$: $\|x\| \leq 1 \Rightarrow \|Tx\| \leq K$;
5. for some $K > 0$ and for all $x \in X$: $\|Tx\| \leq K\|x\|$.

Proof. (1) $\Rightarrow$ (2) $\Rightarrow$ (3) This is clear.

(3) $\Rightarrow$ (4) The continuity at 0 implies that there is $\delta > 0$ such that $\|z\| < \delta$ implies that $\|Tz\| < 1$. Let $\|x\| \leq 1$. Then

$$\left\| \frac{\delta x}{2} \right\| = \frac{\delta}{2}\|x\| \leq \frac{\delta}{2} < \delta \quad \text{and} \quad \left\| T\,\frac{\delta x}{2} \right\| < 1.$$

The linearity of T implies $T((\delta x)/2) = (\delta/2)Tx$, and thus $\|Tx\| \leq 2/\delta$.

(4) $\Rightarrow$ (5) Let $x \in X$ and $x \neq 0$. Then

$$\left\| \frac{x}{\|x\|} \right\| = 1 \quad \text{and} \quad \left\| T\,\frac{x}{\|x\|} \right\| \leq K.$$

The linearity of T implies $T(x/\|x\|) = (1/\|x\|)Tx$, and thus $\|Tx\| \leq K\|x\|$.

(5) $\Rightarrow$ (1) Observe that the linearity of T implies that

$$\|Tx - Ty\| = \|T(x-y)\| \leq K\|x-y\|, \quad x, y \in X.$$

☺

Definition 4.2. Let X and Y be normed linear spaces. A linear operator $T : X \to Y$ is called *bounded* if $\|Tx\| \le K\|x\|$, $x \in X$, for some $K > 0$.

Recall that a not necessarily linear map $T : X \to Y$ is called bounded if $\|Tx\| \le M$ for all $x \in X$. The present definition of bounded is always in conjunction with a linear operator.

Lemma 4.3. Let X and Y be normed linear spaces and let $T : X \to Y$ be a linear operator. If $\dim X < \infty$, then:
1. T is a *finite rank operator*, i.e., $\dim \operatorname{ran} T < \infty$;
2. T is bounded in the sense of Definition 4.2.

Proof. (1) Let $\dim X = n$ and let $\{e_1, \dots, e_n\}$ be a basis for X. Hence

$$x = \sum_{i=1}^{n} \lambda_i e_i \quad \text{and} \quad Tx = \sum_{i=1}^{n} \lambda_i T e_i,$$

which shows that $\dim \operatorname{ran} T < \infty$.

(2) It follows with the Cauchy-Schwarz inequality that

$$\|Tx\| \le \sum_{i=1}^{n} |\lambda_i| \|Te_i\| \le \sqrt{\sum_{i=1}^{n} \|Te_i\|^2} \sqrt{\sum_{i=1}^{n} |\lambda_i|^2} = \sqrt{\sum_{i=1}^{n} \|Te_i\|^2}\ \|x\|_+,$$

where

$$\|x\|_+ = \sqrt{\sum_{i=1}^{n} |\lambda_i|^2}$$

defines a norm on X; cf. Example 2.7. The result now follows as all norms on X are equivalent. ☺

Definition 4.4. Let X and Y be normed linear spaces. Then $B(X,Y)$ is the subset of $L(X,Y)$, consisting of all linear operators from X to Y which are bounded in the sense of Definition 4.2.

Lemma 4.5. Let X and Y be normed linear spaces. Then $B(X,Y)$ with the usual addition and scalar multiplication is a linear space.

Proof. Assume that $\|Tx\| \le M\|x\|$ and $\|Sx\| \le N\|x\|$ for all $x \in X$. Then

$$\|(T+S)x\| = \|Sx + Tx\| \le \|Tx\| + \|Sx\| \le M\|x\| + N\|x\| = (M+N)\|x\|,$$

and

$$\|(\lambda T)(x)\| = \|\lambda Tx\| = |\lambda|\|Tx\| \le |\lambda| N \|x\|,$$

for all $x \in X$. Hence the linear operators $T + S$ and λT are bounded. ☺

Definition 4.6. Let X and Y be normed linear spaces and let $T : X \to Y$ be a bounded linear operator. Then the *operator-norm* $\|T\|$ is defined by

$$\|T\| = \sup_{x \ne 0} \frac{\|Tx\|}{\|x\|},$$

or, equivalently,

$$\|T\| = \inf\{K \ge 0 : \ \|Tx\| \le K\|x\| \ \text{ for all } x \in X\}.$$

Lemma 4.7. Let X and Y be normed linear spaces and let $T \in B(X,Y)$. Then

$$\|T\| = \sup_{\|x\|=1} \|Tx\| = \sup_{\|x\|\leq 1} \|Tx\|.$$

Proof. The first equality is shown as follows. Let $x_0 \in X$ with $x_0 \neq 0$. Then

$$\frac{\|Tx_0\|}{\|x_0\|} = \left\| T\left(\frac{x_0}{\|x_0\|}\right) \right\| \leq \sup_{\|x\|=1} \|Tx\|,$$

which implies that

$$\sup_{x \neq 0} \frac{\|Tx\|}{\|x\|} \leq \sup_{\|x\|=1} \|Tx\|.$$

For the reverse inequality, let $x_0 \in X$ with $\|x_0\| = 1$. Then

$$\|Tx_0\| = \frac{\|Tx_0\|}{\|x_0\|} \leq \sup_{x \neq 0} \frac{\|Tx\|}{\|x\|},$$

which leads to

$$\sup_{\|x\|=1} \|Tx\| \leq \sup_{x \neq 0} \frac{\|Tx\|}{\|x\|}.$$

Now the second equality is shown. Observe that

$$\sup_{\|x\|=1} \|Tx\| \leq \sup_{\|x\|\leq 1} \|Tx\|.$$

For the reverse inequality, let $x_0 \in X$ with $\|x_0\| \leq 1$ and $x_0 \neq 0$. Then

$$\|Tx_0\| = \|x_0\| \left\| T\left(\frac{x_0}{\|x_0\|}\right) \right\| \leq \left\| T\left(\frac{x_0}{\|x_0\|}\right) \right\| \leq \sup_{\|x\|=1} \|Tx\|,$$

which leads to

$$\sup_{\|x\|\leq 1} \|Tx\| \leq \sup_{\|x\|=1} \|Tx\|.$$ ☺

Example 4.8. Let X and Y be normed linear spaces and let $T : X \to Y$ be a linear operator. Then T is called an *isometry* if $\|Tx\| = \|x\|$ for all $x \in X$; see Definition 3.28. Then it is clear that $T \in B(X,Y)$ and $\|T\| = 1$. Likewise, the operator T is called a *contraction* if $\|Tx\| \leq \|x\|$ for all $x \in X$. Then it is clear that $T \in B(X,Y)$ and $\|T\| \leq 1$. Note that if $T \in B(X,Y)$ has the property $\|T\| = 1$, then T is a contraction, which is not necessarily an isometry. □

Example 4.9. Let $X = \mathcal{C}([a,b],\mathbb{K})$ with the sup-norm and let $c \in [a,b]$. Let $T : X \to \mathbb{K}$ be the *evaluation operator* defined by

$$Tf = f(c).$$

Then, indeed, T is a linear rank one map from X into $\mathbb{K}$ and observe that

$$|Tf| = |f(c)| \leq \sup_{x \in [a,b]} |f(x)| = \|f\|_\infty,$$

so that T is a bounded linear operator with $\|T\| \leq 1$. In fact, by considering a constant function it follows that $\|T\| = 1$. □

Example 4.10. Consider $X = \mathcal{C}([a,b],\mathbb{K})$ with the sup-norm. Let $T : X \to \mathbb{K}$ be the *integration operator* defined by

$$Tf = \int_a^b f(t)\,dt.$$

Then, indeed, T is a linear rank one map from X into $\mathbb{K}$ and observe that

$$|Tf(x)| \leq (b-a) \sup_{x\in[a,b]} |f(x)| = (b-a)\|f\|_\infty,$$

so that T is a bounded linear operator with $\|T\| \leq b-a$. In fact, by considering a constant function it follows that $\|T\| = b-a$. □

Example 4.11. Let $X = \ell^p$, $1 \leq p < \infty$, let $\lambda \in \ell^\infty$, and let $T : X \to X$ be the *multiplication operator* defined by

$$Tx = (\lambda_1 x_1, \lambda_2 x_2, \dots), \quad x \in X = \ell^p.$$

Indeed, T is a linear operator in X and, clearly,

$$\|Tx\|^p = \sum_{i=1}^\infty |\lambda_i x_i|^p \leq \|\lambda\|_\infty^p \|x\|^p,$$

so that T is a bounded linear operator, i.e., $T \in B(X)$, and $\|T\| \leq \|\lambda\|_\infty$. In fact, one has

$$\|T\| = \|\lambda\|_\infty.$$

To see this, let $\varepsilon > 0$ then there exists $N \in \mathbb{N}$ such that

$$\|\lambda\|_\infty - \varepsilon < |\lambda_N| \leq \|\lambda\|_\infty.$$

Define $x \in \ell^p$ by $x_i = \delta_{i,N}$, so that $\|x\| = 1$ and

$$\|\lambda\|_\infty - \varepsilon < |\lambda_N| = \frac{\|Tx\|}{\|x\|} \leq \|\lambda\|_\infty,$$

from which the claim follows. □

Example 4.12. Let $X = \mathcal{C}([a,b],\mathbb{K})$ with the sup-norm and let $T : X \to X$ be the *multiplication operator* defined by $Tf = \varphi f$ with $\varphi \in \mathcal{C}([a,b],\mathbb{K})$. Clearly T is a linear operator. Assume that $|\varphi|$ takes its maximum in $t_0 \in [a,b]$. Then

$$\|Tf\|_\infty = \sup_{t\in[a,b]} |\varphi(t)f(t)| \leq |\varphi(t_0)| \sup_{t\in[a,b]} |f(t)| = |\varphi(t_0)|\,\|f\|_\infty,$$

which shows that $\|T\| \leq |\varphi(t_0)|$, so that T is a bounded linear operator. In fact one has

$$\|T\| = |\varphi(t_0)|.$$

To see this observe that for $\varepsilon > 0$ there exists $\delta > 0$ such that $t \in (t_0-\delta, t_0+\delta)$ one has $|\varphi(t_0)| - \varepsilon < |\varphi(t)| \leq |\varphi(t_0)|$. Then for any nontrivial $f \in \mathcal{C}([a,b],\mathbb{K})$ with support in $(t_0-\delta, t_0+\delta)$, then

$$(|\varphi(t_0)| - \varepsilon)|f(t)| < |\varphi(t)f(t)| \leq |\varphi(t_0)||f(t)|.$$

Taking suprema leads to

$$|\varphi(t_0)| - \varepsilon \leq \frac{\|Tf\|_\infty}{\|f\|_\infty} \leq |\varphi(t_0)|,$$

from which the claim follows.

When $X = \mathcal{C}([a,b],\mathbb{K})$ with the L^2-norm and $T : X \to X$ is the *multiplication operator* defined by $Tf = \varphi f$ with $\varphi \in \mathcal{C}([a,b],\mathbb{K})$, then

$$\|Tf\|_2^2 = \int_a^b |\varphi(t)f(t)|^2\,dt \le |\varphi(t_0)|^2 \int_a^b |f(t)|^2\,dt = |\varphi(t_0)|^2\,\|f\|_2^2,$$

which shows that $\|T\| \le |\varphi(t_0)|$. In fact again one has $\|T\| = |\varphi(t_0)|$, using a similar argument. □

Example 4.13. Let $X = \mathcal{C}^1([0,1],\mathbb{K})$ provided with the sup-norm. Define the operator $T : X \to \mathbb{K}$ by

$$Tf = f'(1), \quad f \in X.$$

Then clearly T is a linear rank one map. Define the functions $f_n \in X$ by $f_n(t) = t^n$, $n \in \mathbb{N}$, so that $f_n'(t) = nt^{n-1}$, $n \in \mathbb{N}$. Note that

$$\|f_n\|_\infty = 1, \quad |Tf_n| = n,$$

which implies that the operator $T : X \to \mathbb{K}$ is not bounded. For bounded rank one operators and, more generally, bounded finite-rank operators, see Lemma 4.44. □

Example 4.14. Let $X = \mathcal{C}^1([0,1],\mathbb{K})$ and $Y = \mathcal{C}([0,1],\mathbb{K})$, each with the sup-norm. Let $T : X \to Y$ be the *differentiation operator* defined by $Tf = f'$. Then clearly T is linear. Define the functions $f_n \in X$ as in Example 4.13: $f_n(t) = t^n$, $n \in \mathbb{N}$. Note that

$$\|f_n\|_\infty = 1, \quad \|Tf_n\|_\infty = n,$$

which implies that the operator $T : X \to Y$ is not bounded.

However, note that the boundedness or the nonboundedness depends on the choice of norm. For instance, if $X = \mathcal{C}^1([a,b],\mathbb{K})$ is provided with the norm

$$\|f\|_1 = \sup_{t\in[a,b]} |f(t)| + \sup_{t\in[a,b]} |f'(t)|,$$

then the differentiation operator operator $T : X \to Y$ is bounded:

$$\|Tf\|_1 = \sup_{t\in[a,b]} |f'(t)| \le \|f\|_1, \quad f \in X,$$

so that $\|T\| \le 1$. In fact, in this case

$$\frac{\|Tf_n\|_1}{\|f_n\|_\infty} = \frac{n}{n+1} \to 1 \quad \text{as} \quad n \to \infty,$$

so that $\|T\| = 1$. □

Example 4.15. Let $X = \mathcal{C}([a,b],\mathbb{K})$ with the sup-norm. Define on $\mathcal{C}([a,b],\mathbb{K})$ the *integral operator* T by

$$Tf(x) = \int_a^x f(t)\,dt, \quad f \in \mathcal{C}([a,b],\mathbb{K}), \quad x \in [a,b].$$

Then, indeed, T is a linear map from X into X and observe that

$$\begin{aligned}\left|\int_a^x f(t)\,dt\right| &\le \int_a^x |f(t)|\,dt \le \left(\sup_{t\in[a,b]} |f(t)|\right)\int_a^x \mathbf{1}\,dt \\ &= \left(\sup_{t\in[a,b]} |f(t)|\right)(x-a) \le \left(\sup_{t\in[a,b]} |f(t)|\right)(b-a).\end{aligned}$$

It follows from these inequalities that

$$\|Tf\|_\infty = \sup_{x\in[a,b]} |Tf(x)| \le (b-a)\|f\|_\infty,$$

so that T is bounded and $\|T\| \le b-a$. In fact, by taking the function $f = \mathbf{1}$, one sees $\|T\| = b-a$.

Let $X = \mathcal{C}([a,b],\mathbb{K})$ with the L^2-norm, then the integral operator T still is a linear map from X into X, but now observe that

$$\begin{aligned}\left|\int_a^x f(t)\,dt\right|^2 &\le \left(\int_a^x |f(t)|\,dt\right)^2 \\ &\le \left(\int_a^x |f(t)|^2\,dt\right)\left(\int_a^x \mathbf{1}^2\,dt\right) \le \left(\int_a^b |f(t)|^2\,dt\right)(x-a).\end{aligned}$$

It follows from these inequalities that

$$\|Tf\|_2^2 = \int_a^b |Tf(x)|^2\,dx \le \frac{(b-a)^2}{2}\|f\|_2^2 \quad \text{or} \quad \|Tf\|_2 \le \frac{(b-a)}{\sqrt{2}}\|f\|_2,$$

so that T is bounded and $\|T\| \le (b-a)/\sqrt{2}$. The actual determination of $\|T\|$ requires a more sophisticated approach; see Example 6.46. □

Proposition 4.16. Let X and Y be normed linear spaces. Then $B(X,Y)$ provided with the operator-norm is a normed linear space.

Proof. Let $T,S \in B(X,Y)$. It is clear that $\|T\| \ge 0$ and if $T = 0$, then $\|T\| = 0$. Conversely, if $\|T\| = 0$ then $\|Tx\| = 0$ for all $x \in X$. Hence $T = 0$. Moreover,

$$\|(\lambda T)x\| = \|\lambda(Tx)\| \le |\lambda|\|T\|\|x\|, \quad x \in X,$$

so that $\|\lambda T\| \le |\lambda|\|T\|$. Hence,

$$\|T\| = \left\|\frac{1}{\lambda}\lambda T\right\| \le \frac{1}{|\lambda|}\|\lambda T\|,$$

so that $|\lambda|\|T\| \le \|\lambda T\|$. Now let $T,S \in B(X,Y)$, then

$$\|(T+S)x\| = \|Tx+Sx\| \le \|Tx\| + \|Sx\| \le (\|T\|+\|S\|)\|x\|, \quad x \in X.$$

From this it follows immediately $\|T+S\| \le \|T\|+\|S\|$. ☺

Lemma 4.17. Let X, Y, and Z be normed linear spaces. Let $T \in B(X,Y)$ and let $S \in B(Y,Z)$. Then the product ST belongs to $B(X,Z)$ and

$$\|ST\| \le \|S\|\,\|T\|.$$

Proof. It is clear that $ST \in L(X,Z)$. It follows from the definitions that

$$\|(ST)x\| = \|S(Tx)\| \le \|S\|\|Tx\| \le \|S\|\|T\|\|x\|, \quad x \in X.$$

Hence $ST \in B(X,Z)$. ☺

Example 4.18. In the Banach space ℓ^p the *right shift* S_r and the *left shift* S_l are defined by

$$S_r(x_1,x_2,\dots) = (0,x_1,x_2,\dots), \quad S_l(x_1,x_2,\dots) = (x_2,x_3,\dots), \quad x \in \ell^p.$$

The operators S_r and S_l are linear and bounded in ℓ^p; they are contractions with $\|S_l\| = \|S_r\| = 1$, and in fact S_r is an isometry. It is clear that

$$S_l S_r = I \quad \text{and} \quad (S_r S_l)(x) = (0,x_2,x_3,\dots), \quad x \in \ell^p.$$

□

Lemma 4.19. Let X, Y, and Z be normed linear spaces. Let $T_n \in B(X,Y)$, $S_n \in B(Y,Z)$, and assume that

$$T_n \to T \quad \text{and} \quad S_n \to S,$$

with $T \in B(X,Y)$ and $S \in B(Y,Z)$. Then

$$S_n T_n \to ST,$$

and, in particular, $S_n T \to ST$ and $ST_n \to ST$,

Proof. Observe that

$$\begin{aligned} \|S_n T_n - ST\| &\leq \|S_n T_n - ST_n\| + \|ST_n - ST\| \\ &\leq \|S_n - S\| \|T_n\| + \|S\| \|T_n - T\|, \end{aligned}$$

and that $(\|T_n\|)$ is bounded as T_n converges in $B(X,Y)$. ☺

Lemma 4.20. Let X be a normed linear space and let $V \subset X$ be a closed linear subspace. Then the quotient map $\pi : X \to X/V$, defined by

$$\pi(x) = x + V,$$

maps the open unit sphere in X onto the open unit sphere in X/V. In particular, $\|\pi\| = 1$.

Proof. Under the assumption that V is closed one has that

$$\|x+V\| = \inf\{\|x-y\| : y \in V\}$$

is a norm on X/V and that $\|x+V\| \leq \|x\|$; cf. Proposition 2.25. Hence $\|\pi(x)\| \leq \|x\|$ and thus $\|\pi\| \leq 1$.

Now observe that if $x \in X$ and $\|x\| < 1$, then $\|x+V\| \leq \|x\| < 1$; i.e., π maps the open unit sphere into the open unit sphere. Conversely, let $x+V$ be an element in X/V with $\|x+V\| < 1$. Then by definition there exists an element $u \in V$ with $\|x-u\| < 1$ and $\pi(x-u) = x+V$, i.e., π maps the open unit sphere onto the open unit sphere. ☺

Theorem 4.21. Let X and Y be normed linear spaces and and let $V \subset X$ be a closed linear subspace and let π be the quotient map. Let $T : X \to Y$ be a bounded linear operator and let $V \subset \ker T$. Then

$$\widehat{T} : X/V \to Y,$$

where $T = \widehat{T} \circ \pi$ has the property $\|\widehat{T}\| = \|T\|$.

Proof. Recall that $\widehat{T}$ is well-defined by

$$\widehat{T} : X/V \to Y, \quad x+V \mapsto T(x),$$

and satisfies $T = \widehat{T} \circ \pi$. Observe that for $x \in X$ and $y \in V$:

$$\|\widehat{T}(x+V)\| = \|Tx\| = \|T(x+y)\| \leq \|T\|\|x+y\|, \quad y \in V.$$

which shows that $\|\widehat{T}(x+V)\| \leq \|T\|\|x+V\|$. Hence one sees that $\widehat{T}$ is bounded and that $\|\widehat{T}\| \leq \|T\|$. The reverse inequality follows from $T = \widehat{T} \circ \pi$ and $\|\pi\| = 1$, so that $\|T\| \leq \|\widehat{T}\|\|\pi\| = \|\widehat{T}\|$. ☺

For $T \in B(X,Y)$ the null space $\ker T$ is automatically closed. The closedness of $\operatorname{ran} T$ is a completely different matter and will be further discussed in the next chapters.

Definition 4.22. Let X and Y be normed linear spaces and let $T \in B(X,Y)$. The operator T is said to be *invertible* if T is a bijection from X onto Y and if the inverse T^{-1} belongs to $B(Y,X)$.

Lemma 4.23. Let X and Y be normed linear spaces and let $T \in B(X,Y)$. Then T is invertible if and only if there exists an operator $S \in B(Y,X)$ such that

$$ST = I_X, \quad TS = I_Y,$$

in which case $S = T^{-1}$. In particular, if T is invertible, then also T^{-1} is invertible.

Proof. ($\Rightarrow$) It is clear that T is injective and surjective, cf. Lemma 1.11. Moreover $S = T^{-1}$ belongs to $B(Y,X)$.

($\Leftarrow$) The identities $ST = I_X$ and $TS = I_Y$ imply that T is a bijection and that $S = T^{-1}$. ☺

Corollary 4.24. Let X, Y, and Z be normed linear spaces. Let $T \in B(X,Y)$ and $S \in B(Y,Z)$ be invertible. Then ST is invertible and

$$(ST)^{-1} = T^{-1}S^{-1}.$$

Definition 4.25. Let X and Y be normed linear spaces and let $T \in B(X,Y)$. If T is invertible, then T is called an *isomorphism* between X and Y. The spaces X and Y are called *isomorphic* if there is an isomorphism between them.

Example 4.26. The Banach spaces c and c_0 are isomorphic. To see this define $T : c \to c_0$ and $S : c_0 \to c$ by

$$T(x_1,x_2,x_3,\dots) = (\lambda, x_1-\lambda, x_2-\lambda,\dots), \quad x \in c, \quad \lambda = \lim_{i\to\infty} x_i,$$

and

$$S(y_1,y_2,y_3,\dots) = (y_2+y_1, y_3+y_1, y_4+y_1,\dots), \quad y \in c_0.$$

Then $STx = x$, $x \in c$, and $TSy = y$, $y \in c_0$, while

$$\|Tx\|_\infty \leq 2\|x\|_\infty, \quad x \in c, \quad \text{and} \quad \|Sy\|_\infty \leq 2\|y\|_\infty, \quad y \in c_0.$$

For the first inequality note that $x_i \to \lambda$ implies $|x_i| \to |\lambda|$; since $|x_i| \leq \|x\|_\infty$ it follows that $|\lambda| \leq \|x\|_\infty$. Due to these inequalities, one sees that T and S are invertible. □

Remark 4.27. Let X and Y be normed linear spaces. According to Definition 3.28 a linear map T from X to Y is called an *isometric isomorphism* if T maps X bijectively onto Y such that $\|Tx\| = \|x\|$, $x \in X$. Sometimes one speaks of a *unitary map*. Note that in this case the inverse T^{-1} from Y to X is also isometric. In particular, one sees that for an isometric isomorphism one has $T \in B(X,Y)$ and $T^{-1} \in B(Y,X)$. Hence an isometric isomorphism in the sense of Definition 3.28 is an isomorphism in the sense of Definition 4.25. □

Lemma 4.28. Let X and Y be normed linear spaces and let X and Y be isomorphic. Then
1. X is finite-dimensional if and only if Y is finite-dimensional;
2. X is separable if and only if Y is separable;
3. X is complete if and only if Y is complete.

4.2 Spaces of bounded linear operators

Definition 4.29. Let X and Y be normed linear spaces, and let (T_n) be a sequence in $B(X,Y)$. Then
1. T_n converges *uniformly* to $T \in B(X,Y)$ if $\|T_n - T\| \to 0$ where $\|\cdot\|$ denotes the operator norm;
2. T_n converges *strongly* to $T \in B(X,Y)$ if $T_n x \to Tx$ for all $x \in X$.

Note that uniform convergence implies strong convergence, but the converse is not true.

Example 4.30. Define in $X = \ell^p$, $1 \leq p < \infty$, the sequence of linear operators T_n by

$$T_n x = (0, \dots, 0, x_{n+1}, x_{n+2}, \dots) \quad \text{when} \quad x = (x_1, x_2, \dots) \in \ell^p.$$

Then clearly $T_n \in B(X)$,

$$\|T_n x\|_p^p = \sum_{i=n+1}^{\infty} |x_i|^p, \quad x \in X,$$

and $\|T_n\| = 1$. Note that (T_n) converges in the strong sense to the zero operator, but not in the uniform sense. □

Theorem 4.31. Let X be a normed linear space and let Y be a Banach space. Then $B(X,Y)$ is a Banach space.

Proof. Let (T_n) be a Cauchy sequence in $B(X,Y)$. Then $\|T_n\| \leq M$ for all $n \in \mathbb{N}$. Furthermore note that for all $x \in X$:

$$\|T_n x - T_m x\| = \|(T_n - T_m)x\| \leq \|T_n - T_m\| \|x\|.$$

Hence for each $x \in X$ the sequence $(T_n x)$ is Cauchy in Y. Since Y is a Banach space there exists a limit, say $Tx \in Y$, such that $T_n x \to Tx$. It will be shown that $T \in B(X,Y)$ and that $\|T_n - T\| \to 0$.

Linearity. The linearity of T follows from

$$T(x+y) = \lim_{n \to \infty} T_n(x+y) = \lim_{n \to \infty} (T_n x + T_n y) = Tx + Ty, \quad x, y \in X.$$

Similarly, it follows that $T(\lambda x) = \lambda Tx$ for all $x \in X, \lambda \in \mathbb{K}$.

Boundedness. The boundedness of T follows from $\|T_n x\| \leq \|T_n\| \|x\| \leq M\|x\|$ and the fact that $T_n x \to Tx$ implies $\|T_n x\| \to \|Tx\|$. Therefore it follows that $\|Tx\| \leq M\|x\|$. Hence $T \in B(X,Y)$.

Convergence. To see that $\|T_n - T\| \to 0$, observe that for each $\varepsilon > 0$ there exists $N \in \mathbb{N}$ such that $\|T_n - T_m\| < \varepsilon/2$ when $m, n \geq N$. Let $x \in X$ with $\|x\| \leq 1$. Then for $m, n \geq N$:

$$\|T_n x - T_m x\| = \|(T_n - T_m)x\| \leq \|T_n - T_m\| \|x\| < \varepsilon/2.$$

Take $m \to \infty$, which leads to

$$\|T_n x - Tx\| \leq \varepsilon/2,$$

for all $n \geq N$. Hence for all $n \geq N$

$$\|T_n - T\| = \sup_{\|x\| \leq 1} \|T_n x - Tx\| \leq \varepsilon/2 < \varepsilon,$$

which completes the proof. ☺

The following result is a useful tool in practical situations.

Theorem 4.32 (Extension theorem)**.** Let X be a normed linear space and let Y be a Banach space. Let $V \subset X$ be a linear subspace which is dense and assume that $T \in B(V,Y)$. Then there exists a unique $\dot{T} \in B(X,Y)$ with

$$\dot{T}x = Tx, \quad x \in V,$$

so that $\dot{T}$ is an extension of T. Moreover $\|\dot{T}\| = \|T\|$, i.e., the extension preserves the norm.

Proof. Uniqueness. The uniqueness of the extension is clear: if $S \in B(X,Y)$ and $Sv = 0$ for all $v \in V$, then $Sx = 0$ for all $x \in X$.

Existence. The existence of the extension is based on the following argument. Assume that $x \in X$ and let (v_n), (w_n) be sequences in V with $v_n \to x$ and $w_n \to x$. Then (Tv_n) and (Tw_n) are Cauchy sequences in Y as

$$\|Tv_n - Tv_m\| \leq \|T\|\|v_n - v_m\|, \quad \|Tw_n - Tw_m\| \leq \|T\|\|w_n - w_m\|.$$

Since Y is a Banach space it follows (Tv_n) and (Tw_n) converge to y and y' respectively. Then

$$\begin{aligned} \|y - y'\| &= \|y - Tv_n\| + \|Tv_n - Tw_n\| + \|Tw_n - y'\| \\ &\leq \|y - Tv_n\| + \|T\|\|v_n - w_n\| + \|Tw_n - y'\|, \end{aligned}$$

shows that $y = y'$. Hence $\dot{T} : X \to Y$ given by

$$\dot{T}x = \lim_{n\to\infty} Tv_n,$$

is well-defined (independent of the sequence converging to x). It is straightforward to see that $\dot{T}$ is linear. Moreover, by construction $\dot{T}$ is an extension of T.

Preservation of the norm. Let $x \in X$ and let (v_n) in V converge to x. Then it follows from the construction that

$$\|\dot{T}x\| = \lim_{n\to\infty} \|Tv_n\| \leq \lim_{n\to\infty} \|T\|\|v_n\| = \|T\|\|x\|.$$

Hence $\dot{T}$ is bounded and $\|\dot{T}\| \leq \|T\|$. Since $\dot{T}$ is an extension of T, it is clear from

$$\|T\| = \sup\{\|Tx\| : x \in V,\, \|x\| = 1\} \leq \sup\{\|\dot{T}x\| : x \in X,\, \|x\| = 1\} = \|\dot{T}\|,$$

that $\|T\| \leq \|\dot{T}\|$. Hence it has been shown that $\|\dot{T}\| = \|T\|$. ☺

Corollary 4.33. Let X be a normed linear space, let Y be a Banach space, and let $T \in B(X,Y)$. Let $\dot{X}$ be a Banach space completion of X. Then there exists a unique $\dot{T} \in B(\dot{X},Y)$ such that $\dot{T} \restriction X = T$. Moreover $\|\dot{T}\| = \|T\|$.

The particular case of Theorem 4.31 where $Y = \mathbb{K}$ deserves special attention. The following definition plays an important role in functional analysis.

Definition 4.34. Let X be a normed linear space over $\mathbb{K}$. Then the *dual space* of X is the Banach space defined by $X' = B(X,\mathbb{K})$. The elements of X' are the linear functionals $f : X \to \mathbb{K}$ which are bounded and whose norm is given by

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

Dual spaces will be extensively studied in Chapter 7; in particular, a number of examples will be given where dual spaces are identified. In the case of a Hilbert space the dual space can be identified with the space itself, which makes it a popular object of study; see Chapter 6.

4.3 A class of invertible operators

Let X be a Banach space and let $T \in B(X)$. In order to uniquely solve the abstract equation $(I-T)x = y$ one must find for any $y \in X$ precisely one element $x \in X$ such that the equation holds. In other words, $I-T$ must be bijective and the solution has the form $x = (I-T)^{-1}y$. It will turn out later that in this case $I-T$ is automatically an isomorphism, i.e., $(I-T)^{-1} \in B(X)$. Presently it will be shown that there are simple conditions which guarantee that $I-T$ is an isomorphism. The key argument is due to Theorem 3.13: an absolutely convergent series in a Banach space converges.

Theorem 4.35. Let X be a Banach space and let $T \in B(X)$. If T satisfies the property

$$\sum_{i=0}^{\infty} \|T^i\| < \infty, \tag{4.1}$$

then the operator $I-T$ is invertible and its inverse is given by

$$(I-T)^{-1} = \sum_{i=0}^{\infty} T^i, \tag{4.2}$$

with convergence in $B(X)$. Moreover, the following estimates are valid:

$$\|(I-T)^{-1}\| \leq \sum_{i=0}^{\infty} \|T^i\|, \quad \|(I-T)^{-1} - \sum_{i=0}^{n} T^i\| \leq \sum_{i=n+1}^{\infty} \|T^i\|. \tag{4.3}$$

In particular, if $\|T\| < 1$ then (4.1) is satisfied and

$$\|(I-T)^{-1}\| \leq \frac{1}{1-\|T\|}, \quad \|(I-T)^{-1} - \sum_{i=0}^{n} T^i\| \leq \frac{\|T\|^{n+1}}{1-\|T\|}. \tag{4.4}$$

Proof. Note that $B(X)$ is a Banach space, since X is a Banach space. Hence the condition (4.1) about absolute convergence implies the convergence in $B(X)$ of the series $\sum_{i=0}^{\infty} T^i$; cf. Theorem 3.13. This leads to

$$(I-T)\left(\sum_{i=0}^{\infty} T^i\right) = I, \quad \left(\sum_{i=0}^{\infty} T^i\right)(I-T) = I,$$

cf. Lemma 4.19. Since $\sum_{i=0}^{\infty} T^i \in B(X)$ these identities imply that $I-T$ is invertible, and that its inverse is given by (4.2). The estimates in (4.3) follow from Theorem 3.13; cf. (3.2). Observe that $\|T^i\| \leq \|T\|^i$, which leads to

$$\sum_{i=0}^{\infty} \|T^i\| \leq \sum_{i=0}^{\infty} \|T\|^i.$$

Hence if $\|T\| < 1$ the geometric series result implies convergence in (4.1). Likewise the estimates in (4.4) follow from (4.3). ☺

Remark 4.36. The condition (4.1) is satisfied if $\|T\| < 1$ as shown in Theorem 4.35. However, if $\|T^p\| < 1$ for some $p \in \mathbb{N}$ the condition (4.1) is still satisfied, since each of the p sums in the right-hand side of

$$\sum_{i=1}^{\infty} \|T^i\| = \sum_{i=0}^{\infty} \|T^{ip+1}\| + \cdots + \sum_{i=0}^{\infty} \|T^{ip+p}\|$$

converges and then rearranging the nonnegative terms leads to the identity. □

Corollary 4.37. Let X be a Banach space and let $T \in B(X)$ satisfy the condition (4.1). Then for all $x \in X$

$$(I-T)^{-1}x = \sum_{i=0}^{\infty} T^i x,$$

with convergence in X.

Proof. Observe that for all $x \in X$ one has

$$\left\| (I-T)^{-1}x - \sum_{i=0}^{n} T^i x \right\| \leq \left\| (I-T)^{-1} - \sum_{i=0}^{n} T^i \right\| \|x\|,$$

so that the result follows from Theorem 4.35. ☺

Theorem 4.38. Let X,Y be Banach spaces and let $T,S \in B(X,Y)$. Assume that T is invertible and that S is close to T in the sense that

$$\|S-T\| < \frac{1}{\|T^{-1}\|}. \tag{4.5}$$

Then $(T-S)T^{-1} \in B(Y)$ satisfies $\|(T-S)T^{-1}\| < 1$. Furthermore S is invertible and S^{-1} is given by

$$S^{-1} = T^{-1}[I-(T-S)T^{-1}]^{-1}. \tag{4.6}$$

In particular, the set of all invertible operators in $B(X,Y)$ is open.

Proof. Let $T \in B(X,Y)$ be invertible and let $S \in B(X,Y)$. Then one may write S in terms of T as follows:

$$S = T - (T-S) = [I-(T-S)T^{-1}]T, \tag{4.7}$$

where clearly $I-(T-S)T^{-1} \in B(Y)$. Assume that S is close to T in the sense of (4.5), then the operator $(T-S)T^{-1}$ satisfies

$$\|(T-S)T^{-1}\| \leq \|T-S\|\|T^{-1}\| < 1.$$

Therefore the operator $I-(T-S)T^{-1}$ is invertible in $B(Y)$ by Theorem 4.35. Hence in the right-hand side of the identity (4.7) each factor is invertible. One concludes that S is invertible and that (4.6) holds. ☺

Corollary 4.39. In the situation of Theorem 4.38 there is the following bound for the inverse S^{-1}:

$$\|S^{-1}\| \leq \|T^{-1}\|\,(1-\|T^{-1}\|\|T-S\|)^{-1}, \tag{4.8}$$

and a bound for the difference $S^{-1}-T^{-1}$:

$$\|S^{-1}-T^{-1}\| \leq \|T^{-1}\|^2 \, \|T-S\| \, (1-\|T^{-1}\|\|T-S\|)^{-1}. \tag{4.9}$$

4.4 Compact operators in normed linear spaces

Definition 4.40. Let X and Y be normed linear spaces and let $T: X \to Y$ be a linear operator. The operator T is called *compact* if for every bounded subset $V \subset X$ the image TV is relatively compact in Y. The collection of all compact operators is denoted by $K(X,Y)$.

Lemma 4.41. Let X and Y be normed linear spaces and let $T \in K(X,Y)$. Then
1. $T \in B(X,Y)$;
2. $\operatorname{ran} T$ is separable in Y.

Proof. (1) Let B denote the unit ball in X, a bounded subset of X. Since T is compact, TB is relatively compact, and thus contained in a ball:

$$\|x\| \leq 1 \quad \Rightarrow \quad \|Tx\| \leq C.$$

Hence $\|Tx\| \leq C\|x\|$, $x \in X$; in other words, $T \in B(X,Y)$.

(2) Let $B_n = \{x \in X : \|x\| \leq n\}$. Then $B_n \subset X$ is a bounded subset. Hence TB_n is relatively compact, thus totally bounded, thus separable. Observe that

$$X = \bigcup_{n=1}^{\infty} B_n, \quad TX = \bigcup_{n=1}^{\infty} TB_n.$$

Each TB_n has a countable dense subset E_n, and therefore $\bigcup_{n=1}^{\infty} E_n$ is a countable dense subset for $\operatorname{ran} T = TX$. ☺

Lemma 4.42. Let X and Y be normed linear spaces and let $T: X \to Y$ be a linear operator. Then T is compact if and only if for each bounded sequence (x_n) in X there is a subsequence of (Tx_n) which converges in Y.

Proof. ($\Rightarrow$) Let (x_n) in X be a bounded sequence. Then the closure of the set $\{Tx_n : n \in \mathbb{N}\}$ is compact. Hence (Tx_n) has a convergent subsequence.

($\Leftarrow$) Let $V \subset X$ be a bounded subset. It will be shown that $T(V)$ is relatively compact. Let $y_n \in T(V)$, so that $y_n = Tx_n$ with $x_n \in V$. Since V is bounded the sequence (x_n) is bounded. Hence (y_n) contains a convergent subsequence. ☺

Lemma 4.43. Let X, Y, and Z be normed linear spaces.
1. Let $T,S \in K(X,Y)$ and $\lambda, \mu \in \mathbb{K}$, then $\lambda T + \mu S \in K(X,Y)$.
2. Let $T \in B(X,Y)$ and $S \in B(Y,Z)$ and assume one of them is compact, then $ST \in K(X,Z)$.

Proof. (1) Let (x_n) be a bounded subsequence in X. Since T is compact there is a subsequence (x_{n_k}) for which (Tx_{n_k}) converges. Since S is compact the bounded sequence (x_{n_k}) has a subsequence, say (y_n) for which (Sy_n) converges. Note that (Ty_n) converges as a subsequence of the convergent sequence (Tx_{n_k}).

(2) Assume T is compact. Let the sequence (x_n) be bounded in X then there is a subsequence (x_{nk}) in X such that (Tx_{n_k}) converges in Y. Since S is bounded (STx_{n_k}) converges in Z. Hence ST is compact.

Assume S is compact. Let (x_n) be a bounded sequence in X, then (Tx_n) is bounded in Y, since T is bounded. Now there is a subsequence (Tx_{n_k}), for which (STx_{n_k}) converges in Z. Hence ST is compact. ☺

Remember that the condition that the operator is bounded in the following result is essential. There exist finite-rank operators which are not bounded; cf. Example 4.13.

Lemma 4.44. Let X and Y be normed linear spaces and let $T \in B(X,Y)$. Assume that T has finite-rank, i.e., $\dim \operatorname{ran} T < \infty$. Then $T \in K(X,Y)$. In particular, a linear operator $T : X \to Y$ with $\dim X < \infty$ is compact.

Proof. Let (x_n) be a bounded sequence in X. Since $T \in B(X,Y)$ it follows that the sequence (Tx_n) is bounded in Y. Actually (Tx_n) is a bounded sequence in the finite-dimensional subspace $\operatorname{ran} T \subset Y$. By the Bolzano-Weierstrass theorem the sequence (Tx_n) contains a convergent subsequence. Hence the operator T is compact.

As to the last statement, a linear operator $T : X \to Y$ with $\dim X < \infty$ is a finite-rank operator which belongs to $B(X,Y)$; see Lemma 4.3. Therefore $T \in K(X,Y)$. ☺

Remark 4.45. Let X be an infinite-dimensional normed linear space. Then the identity operator I is not compact as the unit ball in X is not compact. As a consequence an operator $T \in K(X)$ is not invertible when X is infinite-dimensionsal. □

Theorem 4.46. Let X be a normed linear space and let Y be a Banach space. The compact operators $K(X,Y)$ form a closed linear subspace of $B(X,Y)$.

Proof. Let (T_n) be a sequence in $K(X,Y)$ with $T_n \to T$ in $B(X,Y)$. To show that $T \in K(X,Y)$, it suffices to show that for any bounded sequence (x_i) the sequence (Tx_i) has a convergent subsequence. Assume that $\|x_i\| \le c$ for some $c > 0$. The selection of the subsequence will be done via the sequence of compact operators (T_n).

Since T_1 is compact, there is a subsequence (x_i^1) of (x_i) such that $(T_1 x_i^1)$ is convergent. Since T_2 is compact, there is a subsequence (x_i^2) of (x_i^1) such that $(T_2 x_i^2)$ is convergent. Thus one gets the array of sequences

$$\begin{array}{cccc} x_1^1, & x_2^1, & x_3^1, & \dots \\ x_1^2, & x_2^2, & x_3^2, & \dots \\ x_1^3, & x_2^3, & x_3^3, & \dots \\ \vdots & \vdots & \vdots & \ddots \end{array}$$

where each sequence is a subsequence of the one above it, while for each $n \in \mathbb{N}$ the sequence

$$T_n x_1^n, T_n x_2^n, T_n x_3^n, \dots$$

converges. Now define the diagonal sequence $z_1, z_2, z_3, \dots$, where

$$z_i = x_i^i.$$

Then (z_i) is a subsequence of (x_i). Also, for each n, apart from the first $n-1$ terms, (z_i) is a subsequence of (x_i^n) and so $(T_n z_i)$ is convergent. It will now be shown that (Tz_i) is convergent by showing that it is a Cauchy sequence. For all $i, j, n \in \mathbb{N}$ one has

$$\begin{aligned} \|Tz_i - Tz_j\| &= \|(T - T_n)(z_i - z_j) + T_n(z_i - z_j)\| \\ &\le \|T - T_n\|(\|z_i\| + \|z_j\|) + \|T_n(z_i - z_j)\|. \end{aligned} \tag{4.10}$$

Let $\varepsilon > 0$. Since $T_n \to T$ there is n_0 such that $\|T - T_n\| < \varepsilon/4c$ for $n > n_0$. Choose one fixed such n. Now, since $(T_n z_i)$ converges, it is a Cauchy sequence and so there is an $i_0 \in \mathbb{N}$ such that for $i, j > i_0$ implies $\|T_n z_i - T_n z_j\| < \varepsilon/2$. Combining these with (4.10) shows that for $i > i_0$, $j > i_0$, $\|Tz_i - Tz_j\| < \varepsilon$ so (Tz_i) is convergent as required. ☺

Corollary 4.47. Let X be a normed linear space and let Y be a Banach space. Let $T_n : X \to Y$ be finite-rank operators such that $T_n \to T$ in $B(X,Y)$. Then $T \in K(X,Y)$.

Example 4.48. Let $X = \ell^p$, $1 \leq p < \infty$, and let $S_n \in B(X)$ be given by

$$S_n x = (x_1, \dots, x_n, 0, 0, \dots), \quad x = (x_1, x_2, \dots) \in \ell^p.$$

Then S_n has finite-rank and $\|S_n x - Ix\| \to 0$ for all $x \in \ell^p$. However S_n does not converge to I in the norm of $B(X)$, as the identity I in the infinite-dimensional space X is not compact. □

Example 4.49. Define in $X = \ell^p$, $1 \leq p < \infty$, the multiplication operator $T \in B(X)$ by

$$Tx = (\lambda_1 x_2, \lambda_2 x_2, \dots), \quad x = (x_1, x_2, \dots) \in \ell^p,$$

where (λ_n) is a bounded sequence in $\mathbb{K}$; cf. Example 4.11. Note that T has finite rank if there is an $N \in \mathbb{N}$ such that $\lambda_n = 0$ for all $n \geq N$. In general one has

$$T \in K(X) \quad \Leftrightarrow \quad \lambda_n \to 0.$$

To see this, assume that $\lambda_n \to 0$, and define the finite-rank operators T_n by

$$T_n x = (\lambda_1 x_1, \dots, \lambda_n x_n, 0, 0, \dots), \quad x = (x_1, x_2, \dots) \in \ell^p.$$

Then

$$\|(T - T_n)x\|^p = \sum_{i=n+1}^{\infty} |\lambda_i|^p |x_i|^p.$$

If $\varepsilon > 0$ then there exists $N \in \mathbb{N}$ such that $|\lambda_i| < \varepsilon$ when $i \geq N$. Hence for $n \geq N$

$$\|(T - T_n)x\|^p = \sum_{i=n+1}^{\infty} |\lambda_i|^p |x_i|^p \leq \varepsilon^p \|x\|^p \quad \text{or} \quad \|T - T_n\| < \varepsilon,$$

i.e., $T_n \to T$ in $B(X)$. Hence $T \in K(X)$ by Corollary 4.47.

For the converse let $T \in K(X)$. Assume that the sequence (λ_n) does not converge to 0. Then there exists $\varepsilon > 0$ and a subsequence, again denoted by (λ_n), such that $|\lambda_n| \geq \varepsilon$. Consider the sequence (e_n) in X with entries δ_{in}. Then (e_n) is a bounded sequence and

$$Te_n = (0, \dots, 0, \lambda_n, 0, \dots).$$

Therefore it follows for $n \neq m$ that

$$\|Te_n - Te_m\|^2 = |\lambda_n|^2 + |\lambda_m|^2 \geq 2\varepsilon^2.$$

Thus the sequence (Te_n) does not contain any subsequence which is convergent. This contradicts the assumption that T is compact. □

The converse in Corollary 4.47 is not true in general. However under certain extra conditions on the space Y there is a converse. Here is an example when the space Y is a Hilbert space. For a further discussion, see Lemma 4.51.

Theorem 4.50. Let X be a normed linear space and let Y be a Hilbert space. Let $T \in K(X,Y)$, then there exists a sequence $T_n : X \to Y$ of finite-rank operators such that $T_n \to T$ in $B(X,Y)$.

Proof. Assume that $\operatorname{ran} T$ is infinite-dimensional, so that also $\overline{\operatorname{ran}}\, T$ is infinite-dimensional. Since $T \in K(X,Y)$ it follows from Lemma 4.41 that $\operatorname{ran} T$ is separable, and hence also $\overline{\operatorname{ran}}\, T$ is separable. Therefore $\overline{\operatorname{ran}}\, T$ is a separable Hilbert space and there exists an orthonormal basis e_n, $n \in \mathbb{N}$, for $\overline{\operatorname{ran}}\, T$; see Theorem 3.48. Thus every $y \in \overline{\operatorname{ran}}\, T$ can be written as

$$y = \sum_{i=1}^{\infty} \langle y, e_i \rangle e_i, \quad y \in \overline{\operatorname{ran}}\, T.$$

Now define the finite-rank operator P_n by

$$P_n y = \sum_{i=1}^{n} \langle y, e_i \rangle e_i, \quad y \in \overline{\operatorname{ran}}\, T.$$

It is clear that $\|P_n\| \le 1$ and, in fact, $\|P_n\| = 1$. Define $T_n : X \to Y$ by $T_n = P_n T$, so that T_n is a finite-rank operator. It will be shown that $\|T - T_n\| \to 0$, which completes the proof.

Assume that one does not have $\|T - T_n\| \to 0$. Then there is $\varepsilon > 0$ and a subsequence of (T_n), again denoted by (T_n), for which

$$\|T - T_n\| \ge \varepsilon.$$

By the definition of $\|T - T_n\|$ there is a corresponding sequence $x_n \in X$ with $\|x_n\| = 1$ for which

$$\|(T - T_n)x_n\| \ge \varepsilon/2. \tag{4.11}$$

Since $T \in K(X,Y)$ there is a subsequence of (x_n), again denoted by (x_n), and an element $y \in Y$, such that

$$T x_n \to y.$$

Now observe the following identity:

$$\begin{aligned}
(T - T_n)x_n &= (I - P_n)T x_n \\
&= (I - P_n)y + (I - P_n)(T x_n - y) \\
&= \sum_{i=n+1}^{\infty} \langle y, e_i \rangle e_i + (I - P_n)(T x_n - y),
\end{aligned}$$

which leads to

$$\|(T - T_n)x_n\| \le \left(\sum_{i=n+1}^{\infty} |\langle y, e_i \rangle|^2 \right)^{1/2} + 2\|T x_n - y\|.$$

Observe that the right-hand side goes to 0 as $n \to \infty$, while the left-hand side satisfies (4.11), which gives a contradiction. ☺

The following lemma is useful to establish more cases where the converse holds. The assertion in it is modeled by the proof of the previous result.

Lemma 4.51. Let X and Y be normed linear spaces and let $T \in K(X,Y)$. Let the sequence $S_n \in B(Y)$ be finite-rank operators which satisfy the following conditions:

1. $\|S_n\| \le C$ for some constant C;
2. for each $y \in Y$ the sequence $\|S_n y - y\| \to 0$.

Then $(S_n T)$ is a sequence of finite-rank operators and $\|S_n T - T\| \to 0$.

Proof. Define $T_n = S_n T$, which is a finite-rank operator from X to Y. The claim is that $\|T_n - T\| \to 0$. Assume that one does not have $\|T - T_n\| \to 0$. Then there is $\varepsilon > 0$ and a subsequence of (T_n), again denoted by (T_n), for which

$$\|T - T_n\| \geq \varepsilon.$$

By the definition of $\|T - T_n\|$ there is a corresponding sequence $x_n \in X$ with $\|x_n\| = 1$ for which

$$\|(T - T_n)x_n\| \geq \varepsilon/2. \tag{4.12}$$

Since $T \in K(X,Y)$ there is a subsequence of (x_n), again denoted by (x_n), and an element $y \in Y$, such that

$$Tx_n \to y.$$

Now observe the following identity:

$$\begin{aligned}(T - T_n)x_n &= (I - S_n)Tx_n \\ &= (I - S_n)y + (I - S_n)(Tx_n - y),\end{aligned}$$

which leads to

$$\|(T - T_n)x_n\| \leq \|(I - S_n)y\| + (1 + C)\|Tx_n - y\|.$$

Observe that the right-hand side goes to 0 as $n \to \infty$, while the left-hand side satisfies (4.12), which gives a contradiction. ☺

Proposition 4.52. Let X be a normed linear space, let Y be a Banach space, and let $T \in K(X,Y)$. Let $\dot{X}$ be a Banach space completion of X. Then the unique $\dot{T} \in B(\dot{X},Y)$ such that

$$\dot{T} \restriction X = T,$$

belongs to $K(\dot{X},Y)$.

Proof. Let the sequence $(\dot{x}_n)$ in $\dot{X}$ be bounded. Since X is dense in $\dot{X}$ there exists a sequence (x_n) in X with $x_n - \dot{x}_n \to 0$. Note that the sequence (x_n) is bounded in X. The operator T is compact; hence there exists a subsequence (x_{n_k}) of x_n such that $Tx_{n_k} \to y$ in Y. Therefore

$$\dot{T}\dot{x}_{n_k} = \dot{T}(\dot{x}_{n_k} - x_{n_k}) + Tx_{n_k} \to y.$$

It follows that $\dot{T} : \dot{X} \to Y$ is compact. ☺

4.5 Fredholm and Volterra integral operators

The development of the theory of linear operators in infinite-dimensional spaces was stimulated by the theory of integral equations around the beginning of the twentieth century. Roughly speaking the idea is to solve the function f from the integral equation

$$f(x) - \mu \int_a^b K(x,y) f(y)\, dy = g(x), \quad x \in [a,b],$$

when the function g is given and to see for what values of the parameter $\mu \in \mathbb{C}$ this is possible. In the present setting one needs to choose a suitable space of functions X and a operator T in $B(X)$ in order to solve the equation $(I - \mu T)f = g$. It is then clear that for $|\mu| < 1/\|T\|$ there exists a unique solution; cf. Theorem 4.35. Therefore it makes sense to look at integral operators and their estimates in some more detail; in fact they turn out to be not only bounded, but also compact.

Definition 4.53. Let $K : [a,b] \times [a,b] \to \mathbb{K}$ be a continuous function. The *Fredholm operator* T is the integral operator defined by

$$Tf(x) = \int_a^b K(x,y) f(y)\, dy, \quad x \in [a,b], \quad f \in \mathcal{C}([a,b]). \tag{4.13}$$

Proposition 4.54. Let T be the Fredholm operator in (4.13) with the function K continuous on $[a,b] \times [a,b]$. Then

1. T is compact from $\mathcal{C}([a,b],\mathbb{K})$ to $\mathcal{C}([a,b],\mathbb{K})$, both with the sup-norm.
2. T is compact from $\mathcal{C}([a,b],\mathbb{K})$ with the L^2-norm to $\mathcal{C}([a,b],\mathbb{K})$ with the sup-norm.
3. T is compact from $\mathcal{C}([a,b],\mathbb{K})$ to $\mathcal{C}([a,b],\mathbb{K})$, both with the L^2-norm.
4. T has a unique extension $\dot{T}$ from $L^2(a,b)$ to $L^2(a,b)$, given by

$$\dot{T}f(x) = \int_a^b K(x,y) f(y)\, dy, \quad x \in [a,b], \quad f \in L^2(a,b),$$

which is compact.

Proof. (1) The continuous function K is uniformly continuous on the square $[a,b] \times [a,b]$. In order to show that Tf is continuous on $[a,b]$ choose $\varepsilon > 0$. Then there is a $\delta > 0$ such that

$$\sqrt{(x-x')^2 + (y-y')^2} < \delta \quad \Rightarrow \quad |K(x,y) - K(x',y')| < \varepsilon.$$

Now let $|x - x'| < \delta$, then

$$|Tf(x) - Tf(x')| \leq \int_a^b |K(x,y) - K(x',y)||f(y)|\, dy < \varepsilon(b-a) \sup_{y \in [a,b]} |f(y)|. \tag{4.14}$$

Hence Tf is uniformly continuous on $[a,b]$. Thus T maps $\mathcal{C}([a,b],\mathbb{K})$ into itself and it is clear that T is linear. Let

$$M = \sup\{|K(x,y)| : a \leq x, y \leq b\},$$

then it follows from

$$|Tf(x)| \leq \int_a^b |K(x,y)||f(y)|\, dy \leq M(b-a) \sup_{y \in [a,b]} |f(y)|$$

for all $x \in [a,b]$ that

$$\sup_{x \in [a,b]} |Tf(x)| \leq M(b-a) \sup_{y \in [a,b]} |f(y)|. \tag{4.15}$$

Thus $\|T\| \leq M(b-a)$ and T is a bounded linear operator in $\mathcal{C}([a,b],\mathbb{K})$.

To see that T is compact, let $V \subset \mathcal{C}([a,b],\mathbb{K})$ be a bounded subset, i.e., there is $C > 0$ such that for all $f \in V$ it follows that $\sup_{x \in [a,b]} |f(x)| \leq C$. It will be shown that TV is relatively compact by means of Theorem A.9. By the above estimate (4.15) it is clear that TV is bounded. Now it will be shown that TV is equicontinous. Choose $\varepsilon > 0$. Then by the estimate (4.14)

$$|Tf(x) - Tf(x')| \leq \int_a^b |K(x,y) - K(x',y)||f(y)|\, dy < \varepsilon(b-a)C.$$

Hence TV is equicontinuous.

The boundedness and equicontinuity of TV imply that TV is relatively compact by Theorem A.9. Hence the operator T is compact.

(2) The Fredholm operator may also be considered as an operator from $\mathcal{C}([a,b],\mathbb{K})$ with the L^2-norm to $\mathcal{C}([a,b],\mathbb{K})$ with the sup-norm. It follows from

$$\left|\int_a^b K(x,y)f(y)\,dy\right| \leq \sqrt{\int_a^b |K(x,y)|^2\,dy}\sqrt{\int_a^b |f(y)|^2\,dy}, \quad f \in \mathcal{C}([a,b],\mathbb{K}),$$

that T is bounded.

Moreover, T in this context is compact. To see this, let $V \subset X$ be a bounded subset, i.e., there is $C > 0$ such that for all $f \in V$ it follows that

$$\left(\int_a^b |f(t)|^2\,dt\right)^{\frac{1}{2}} \leq C.$$

Recall that T maps into $\mathcal{C}([a,b],\mathbb{K})$, so that $TV \subset \mathcal{C}([a,b],\mathbb{K})$. It will be shown that TV is relatively compact in $\mathcal{C}([a,b],\mathbb{K})$ with the sup-norm. It is clear from the above that TV is bounded. Now it will be shown that TV is equicontinous. Recall that the continuous function K is uniformly continuous on the square $[a,b]\times[a,b]$. Choose $\varepsilon > 0$. Then there is a $\delta > 0$ such that

$$\sqrt{(x-x')^2+(y-y')^2} < \delta \quad \Rightarrow \quad |K(x,y)-K(x',y')| < \varepsilon.$$

Now let $|x-x'| < \delta$ and $f \in V$, then

$$\begin{aligned} |Tf(x)-Tf(x')| &\leq \sqrt{\int_a^b |K(x,y)-K(x',y)|^2\,dy}\sqrt{\int_a^b |f(y)|^2\,dy} \\ &\leq \varepsilon\sqrt{b-a}C. \end{aligned}$$

Hence TV is equicontinuous in $\mathcal{C}([a,b],\mathbb{K})$ with the sup-norm. The boundedness and equicontinuity of TV imply that TV is relatively compact by Theorem A.9. Hence the operator T from $\mathcal{C}([a,b],\mathbb{K})$ with the L^2-norm to $\mathcal{C}([a,b],\mathbb{K})$ with the sup-norm is compact.

(3) The Fredholm operator may also be considered as an operator from $\mathcal{C}([a,b],\mathbb{K})$ with the L^2-norm to $\mathcal{C}([a,b],\mathbb{K})$ with the L^2-norm. It follows from

$$\sqrt{\int_a^b \left|\int_a^b K(x,y)f(y)\,dy\right|^2 dx} \leq \sqrt{\int_a^b\int_a^b |K(x,y)|^2\,dxdy}\sqrt{\int_a^b |f(y)|^2\,dy},$$

for all $f \in \mathcal{C}([a,b],\mathbb{K})$ that T is bounded. Moreover, T in this context is compact.

To see this, let $V \subset \mathcal{C}([a,b],\mathbb{K})$ be a bounded subset, i.e., there is $C > 0$ such that for all $f \in V$ it follows that

$$\left(\int_a^b |f(t)|^2\,dt\right)^{\frac{1}{2}} \leq C.$$

It has been shown that TV is relatively compact in $\mathcal{C}([a,b],\mathbb{K})$ with the sup-norm. But then it is clear that TV is relatively compact in $\mathcal{C}([a,b],\mathbb{K})$ with the L^2-norm. Hence also in this context T is compact.

(4) It follows from (3) that the unique bounded extension $\dot{T}$ of T to all of $L^2(a,b)$ is also compact; cf. Proposition 4.52. For the actual form of $\dot{T}$ observe that for $f \in L^2(a,b)$ there exists a sequence (f_n) of elements in $\mathcal{C}([a,b],\mathbb{K})$ such that $f_n \to f$ in $L^2(a,b)$. Then by definition $Tf_n \to \dot{T}f$ in $L^2(a,b)$. To identify $\dot{T}f$ define the function g by the Lebesgue integral

$$g(x) = \int_a^b K(x,y)f(y)\,dy, \quad x \in [a,b], \quad f \in L^2(a,b).$$

It follows from the Fubini theorem that $g \in L^2(a,b)$ and that

$$\int_a^b |g(x) - Tf_n(x)|^2\, dx \leq \left(\int_a^b \int_a^b |K(x,y)|^2\, dxdy \right) \int_a^b |f(y) - f_n(y)|^2\, dy,$$

so that $Tf_n \to g$ in $L^2(a,b)$. This implies that $g = \dot{T} f$. ☺

An interesting class of integral operators is obtained when the kernel K vanishes above the diagonal.

Definition 4.55. Let $K : \{(x,y) : a \leq y \leq x \leq b\} \to \mathbb{K}$ be a continuous function. The *Volterra operator* T is defined by

$$Tf(x) = \int_a^x K(x,y) f(y)\, dy, \quad x \in [a,b], \quad f \in \mathcal{C}([a,b],\mathbb{K}). \tag{4.16}$$

Proposition 4.56. Let T be the Volterra operator in (4.16) with the function K continuous on $\{(x,y) : a \leq y \leq x \leq b\}$. Then

1. T is compact from $\mathcal{C}([a,b],\mathbb{K})$ to $\mathcal{C}([a,b],\mathbb{K})$, both with the sup-norm.
2. T is compact from $\mathcal{C}([a,b],\mathbb{K})$ with the L^2-norm to $\mathcal{C}([a,b],\mathbb{K})$ with the sup-norm.
3. T is compact from $\mathcal{C}([a,b],\mathbb{K})$ to $\mathcal{C}([a,b],\mathbb{K})$, both with the L^2-norm.
4. T has unique extension $\dot{T}$ from $L^2(a,b)$ to $L^2(a,b)$, given by

$$\dot{T} f(x) = \int_a^x K(x,y) f(y)\, dy, \quad x \in [a,b], \quad f \in L^2(a,b),$$

which is compact.

Proof. It suffices to show how to proceed in case (1). The continuous function K is uniformly continuous on the triangle $\{(x,y) : a \leq y \leq x \leq b\}$. In order to show that Tf is continuous on $[a,b]$ choose $\varepsilon > 0$. Then there is a $\delta > 0$ such that

$$\sqrt{(x-x')^2 + (y-y')^2} < \delta \quad \Rightarrow \quad |K(x,y) - K(x',y')| < \varepsilon.$$

Now let $|x - x'| < \delta$, then

$$\begin{aligned} Tf(x) - Tf(x') &= \int_a^x K(x,y) f(y)\, dy - \int_a^{x'} K(x',y) f(y)\, dy \\ &= \int_a^x (K(x,y) - K(x',y)) f(y)\, dy \\ &\quad + \int_a^x K(x',y) f(y)\, dy - \int_a^{x'} K(x',y) f(y)\, dy. \end{aligned}$$

The first term in the righthand side can be estimated as follows

$$\left| \int_a^x (K(x,y) - K(x',y)) f(y)\, dy \right| \leq \varepsilon (b-a) \sup_{y \in [a,b]} |f(y)|,$$

while the remaining terms are estimated as follows

$$\begin{aligned} \left| \int_a^x K(x',y) f(y)\, dy - \int_a^{x'} K(x',y) f(y)\, dy \right| &= \left| \int_{x'}^x K(x',y) f(y)\, dy \right| \\ &\leq M |x - x'| \sup_{y \in [a,b]} |f(y)|, \end{aligned}$$

where

$$M = \sup\{|K(x,y)| \,:\, a \le y \le x \le b\}.$$

Note that $|x-x'| < \delta$. Hence, by additionally requiring $\delta < \varepsilon$ one obtains

$$|Tf(x) - Tf(x')| \le \varepsilon[b-a+M] \sup_{y\in[a,b]} |f(y)|.$$

Hence Tf is uniformly continuous on $[a,b]$. Thus T maps $\mathcal{C}([a,b],\mathbb{K})$ into itself and it is clear that T is linear. It follows from

$$|Tf(x)| \le \int_a^x |K(x,y)||f(y)|\,dy \le M(b-a) \sup_{y\in[a,b]} |f(y)|,$$

for all $x \in [a,b]$ that

$$\sup_{x\in[a,b]} |Tf(x)| \le M(b-a) \sup_{y\in[a,b]} |f(y)|.$$

Thus $\|T\| \le M(b-a)$ and T is a bounded linear operator in $\mathcal{C}([a,b],\mathbb{K})$.

The cases (2), (3), and (4) are proven as in the case of the Fredholm operator. ☺

Remark 4.57. The Fredholm operator T maps $\mathcal{C}([a,b],\mathbb{K})$ into itself. Assume that the continuous function K is bounded on the set $[a,b]\times[a,b]$ by M. Note that

$$|Tf(x)| \le \int_a^b |K(x,y)||f(y)|\,dy \le M(b-a) \sup_{y\in[a,b]} |f(y)|, \quad x \in [a,b],$$

for all $f \in \mathcal{C}([a,b],\mathbb{K})$. Thus it follows that

$$\|T\| \le M(b-a).$$

Similarly, the Volterra operator T maps $\mathcal{C}([a,b],\mathbb{K})$ into itself. Assume that the continuous function K is bounded on the set $D = \{(x,y) : a \le y \le x \le b\}$ by M. Now the special form of the operator leads to the estimate

$$|Tf(x)| \le \int_a^x |K(x,y)||f(y)|\,dy \le M(x-a) \sup_{y\in[a,b]} |f(y)|, \quad x \in [a,b],$$

for all $f \in \mathcal{C}([a,b],\mathbb{K})$. An induction argument gives the estimates

$$|(T^n f)(x)| \le \frac{M^n(x-a)^n}{n!}\|f\|, \quad x\in[a,b], \quad f \in \mathcal{C}([a,b],\mathbb{K}).$$

From this it follows that

$$\|T^n\| \le \frac{M^n(b-a)^n}{n!}, \quad n \in \mathbb{N}.$$

Hence $I-T$ is invertible and

$$\|(I-T)^{-1}\| \le e^{M(b-a)},$$

cf. Theorem 4.35. □

Remark 4.58. It is possible to extend the notion of Fredholm operator to the setting of L^2-spaces by enlarging the class of kernels. Let $K : [a,b]\times[a,b] \to \mathbb{R}$ be *square integrable*, i.e.,

$$\int_a^b \int_a^b |K(x,y)|^2\,dxdy < \infty. \tag{4.17}$$

It is clear from the Fubini theorem that

$$\begin{aligned}\int_a^b \int_a^b |K(x,y)|^2\,dxdy &= \int_a^b \left(\int_a^b |K(x,y)|^2\,dy \right) dx \\ &= \int_a^b \left(\int_a^b |K(x,y)|^2\,dx \right) dy.\end{aligned} \tag{4.18}$$

To be a bit more specific: for almost all $x \in [a,b]$ the section $K_x(\cdot)$ defined by

$$K_x(y) = K(x,y), \quad x \in [c,d],$$

belongs to $L^2(a,b)$, the map $x \mapsto \|K_x\|^2$ is integrable, and $\int_a^b \|K_x\|^2 dx$ is equal to the integral in (4.17). A similar argument exists when x and y are interchanged.

The function K induces the Fredholm operator T, defined by

$$Tf(x) = \int_a^b K(x,y)f(y)\,dy, \quad x \in [a,b], \quad f \in L^2(a,b).$$

In the theory of measure and integration it is shown that the function Tf is measurable. Furthermore, one sees by the Cauchy-Schwarz inequality that for almost all $x \in [a,b]$

$$|Tf(x)|^2 \leq \left(\int_a^b |K(x,y)|^2\,dy \right) \int_a^b |f(y)|^2\,dy,$$

that

$$\int_a^b |Tf(x)|^2\,dx \leq \left(\int_a^b \int_a^b |K(x,y)|^2\,dxdy \right) \int_a^b |f(y)|^2\,dy.$$

Note that here one has used the identity (4.18). Hence the linear operator T takes $L^2(a,b)$ into itself, and is bounded:

$$\|T\| \leq \sqrt{\int_a^b \int_a^b |K(x,y)|^2\,dxdy}.$$

Next it will be shown that the operator $T : L^2(a,b) \to L^2(a,b)$ is compact. For every $\varepsilon > 0$ there exists a continuous function K_ε such that

$$\int_a^b \int_a^b |K(x,y) - K_\varepsilon(x,y)|^2\,dxdy < \varepsilon^2,$$

and define the integral operator T_ε by

$$T_\varepsilon f(x) = \int_a^b K_\varepsilon(x,y)f(y)\,dy, \quad x \in [a,b], \quad f \in L^2(a,b),$$

so that T_ε is a compact linear operator from $L^2(a,b)$ into itself; cf. Proposition 4.54. Now note that

$$|Tf(x) - T_\varepsilon f(x)|^2 \leq \left(\int_a^b |K(x,y) - K_\varepsilon(x,y)|^2\,dy \right) \int_a^b |f(y)|^2\,dy,$$

which leads to

$$\|Tf - T_\varepsilon f\|^2 < \varepsilon^2 \|f\|^2 \quad \text{or} \quad \|T - T_\varepsilon\| \leq \varepsilon.$$

Hence $T \in K(X)$ by Corollary 4.47. Another way of looking at this result can be found in Theorem 6.58.

Likewise it is possible to extend the notion of Volterra operator to the setting of L^2-spaces. Let $D = \{(x,y) : a \leq y \leq x \leq b\}$ and let $K : D \to \mathbb{R}$ be square integrable, i.e.,

$$\iint_D |K(x,y)|^2 \, dxdy < \infty.$$

Then the Volterra operator T, defined by

$$Tf(x) = \int_a^x K(x,y) f(y) \, dy, \quad x \in [a,b], \quad f \in L^2(a,b),$$

is a compact linear operator from $L^2(a,b)$ into itself. $\square$

Chapter 5

Open mappings, closed graphs, and uniform boundedness

The open mapping theorem, the closed graph theorem, and the uniform boundedness principle each guarantee the boundedness of a linear operator. In this sense they are the building blocks of operator theory in Banach spaces. Their proofs will be given in a unified way using Zabreĭko's lemma. The notion of the spectrum of a linear operator in a Banach space extends the notion of eigenvalues in the finite-dimensional case: it facilitates the idea of solving linear equations in infinite-dimensional spaces.

5.1 Baire's category theorem

The results in this chapter depend very much on the Baire category theorem. Here is a brief treatment.

Definition 5.1. Let (X,d) be a metric space and let $M \subset X$ be a subset. Then M is called:
1. *nowhere dense* if $\operatorname{int}\overline{M} = \emptyset$;
2. *meager* if $M = \bigcup_{i=1}^{\infty} M_i$ where all $M_i \subset X$ are nowhere dense;
3. *nonmeager* if M is not meager.

Sometimes meager and nonmeager are called *first category* and *second category*, respectively.

Remark 5.2. The following observations are useful.
1. Let $M \subset X$ be nowhere dense, then for each open ball $B(x;\varepsilon) \subset X$ the open set
$$B(x;\varepsilon) \setminus \overline{M} = B(x;\varepsilon) \cap (\overline{M})^c$$
is nonempty (as otherwise $B(x,\varepsilon) \subset \overline{M}$, a contradiction).
2. If $M \subset X$ is nonmeager and if $M = \bigcup_{i=1}^{\infty} M_i$, then for at least one $i \in \mathbb{N}$ it follows that $\operatorname{int}\overline{M}_i \neq \emptyset$.

□

Theorem 5.3 (Baire's category theorem)**.** Let (X,d) be a complete metric space and let $O \subset X$ be open. Then O is nonmeager. In particular, X is nonmeager.

Proof. Assume that the open set O is meager, so that $O = \bigcup_{i=1}^{\infty} M_i$ where all $M_i \subset X$ are nowhere dense. This will lead to a contradiction. Let $x_0 \in O$. Since O is open there exists $\varepsilon_0 > 0$ such that $B(x_0;\varepsilon_0) \subset O$. The following construction will take place inside the open ball $B(x_0;\varepsilon_0/2)$.

Since M_1 is nowhere dense it follows that $B(x_0;\varepsilon_0/2)\setminus\overline{M}_1$ is a nonempty set and thus there exists $x_1\in X$ such that

$$x_1\in B(x_0;\varepsilon_0/2)\setminus\overline{M}_1.$$

As $B(x_0;\varepsilon_0/2)\setminus\overline{M}_1$ is open there exists $\varepsilon_1>0$ with $\varepsilon_1<\varepsilon_0/2$ such that

$$B(x_1;\varepsilon_1)\subset B(x_0;\varepsilon_0/2)\setminus\overline{M}_1.$$

Since M_2 is nowhere dense it follows that $B(x_1;\varepsilon_1/2)\setminus\overline{M}_2$ is a nonempty set and thus there exists $x_2\in X$ such that

$$x_2\in B(x_1;\varepsilon_1/2)\setminus\overline{M}_2.$$

As $B(x_1;\varepsilon_1/2)\setminus\overline{M}_2$ is open there exists $\varepsilon_2>0$ with $\varepsilon_2<\varepsilon_1/2$, such that

$$B(x_2;\varepsilon_2)\subset B(x_1;\varepsilon_1/2)\setminus\overline{M}_2.$$

By induction one obtains a sequence of open balls $B(x_i;\varepsilon_i)$, $i\in\mathbb{N}$, with $\varepsilon_i<\varepsilon_{i-1}/2$ such that

$$B(x_i;\varepsilon_i)\subset B(x_{i-1};\varepsilon_{i-1}/2)\setminus\overline{M}_i.$$

In particular this implies that the sequence of balls is nested:

$$B(x_i;\varepsilon_i)\subset B(x_{i-1};\varepsilon_{i-1}/2)\subset B(x_{i-1};\varepsilon_{i-1}),$$

and that $B(x_i;\varepsilon_i)\subset(\overline{M}_i)^c$, which shows that

$$B(x_i;\varepsilon_i)\cap M_i\subset B(x_i;\varepsilon_i)\cap\overline{M}_i=\emptyset.$$

It follows from the construction that $\varepsilon_i<2^{-i}\varepsilon_0$ and that

$$d(x_i,x_{i-1})<\frac{1}{2}\varepsilon_{i-1}<\frac{\varepsilon_0}{2^i},\quad i\in\mathbb{N}.$$

Therefore the sequence (x_i) is Cauchy in X and hence $d(x_i,x)\to 0$ for some $x\in X$, since it is assumed that X is complete. Now fix $m\in\mathbb{N}$ and observe that for $n>m$

$$B(x_n;\varepsilon_n)\subset B(x_m;\varepsilon_m/2),$$

so that $d(x_n,x_m)<\varepsilon_m/2$. Therefore the inequality

$$d(x_m,x)\le d(x_m,x_n)+d(x_n,x)$$

with $n\to\infty$ implies that $d(x_m,x)\le\varepsilon_m/2$. Hence $x\in B(x_m;\varepsilon_m)$, which shows that $x\in O$ and also that $x\notin M_m$, $m\in\mathbb{N}$. This contradicts the assumption $O=\bigcup_{i=1}^\infty M_i$. ☺

5.2 Zabreĭko's lemma

Recall that a semi-norm on a linear space X is a map $p:X\to[0,\infty)$ that satisfies the following properties:

$$p(x+y)\le p(x)+p(y),\quad p(\lambda x)=|\lambda|p(x),\quad x,y\in X,\quad \lambda\in\mathbb{K}.$$

If X is a normed linear space, then the semi-norm p is called bounded if there exists a constant $C\ge 0$ such that

$$p(x)\le C\|x\|,\quad x\in X.$$

Moreover, note that $p(x) = p(x-y+y) \le p(x-y) + p(y)$ then leads to the inequality

$$|p(x) - p(y)| \le p(x-y) \le C\|x-y\|, \quad x,y \in X,$$

which shows that the map p is continuous. In this case, the map p is also countably subadditive: if $\sum_{n=1}^{\infty} x_n$ is a convergent series in X, then

$$p\left(\sum_{n=1}^{\infty} x_n\right) \le \sum_{n=1}^{\infty} p(x_n) \le \infty.$$

Conversely, Zabreĭko's lemma states that if X is a Banach space, then a countably subadditive semi-norm is bounded.

Lemma 5.4 (Zabreĭko)**.** Let X be a Banach space and let $p : X \to [0,\infty)$ be a semi-norm. Assume that for all convergent series $\sum_{n=1}^{\infty} x_n$ in X it follows that

$$p\left(\sum_{n=1}^{\infty} x_n\right) \le \sum_{n=1}^{\infty} p(x_n) \le \infty.$$

Then p is bounded.

Proof. The proof will be given in four steps.

Step 1. Define the sets $M_n = \{x \in X \,:\, p(x) \le n\}$. Then

$$X = \bigcup_{n=1}^{\infty} M_n.$$

Since X is nonmeager by Baire's theorem, there exists $n \in \mathbb{N}$ for which $\overline{M}_n$ has a nonempty interior. Hence, there exists an element $x_0 \in X$ and $\varepsilon > 0$ such that $B(x_0;\varepsilon) \subset \overline{M}_n$.

Step 2. Next, it is shown with $\varepsilon > 0$ from step 1 that $B(0;\varepsilon) \subset \overline{M}_n$. To see this, note that the set M_n is symmetric and convex: if $x,y \in M_n$, then $-x \in M_n$ and $\lambda x + (1-\lambda)y \in M_n$ for all $\lambda \in [0,1]$. The same properties hold for the closure $\overline{M}_n$. If $x \in B(0;\varepsilon)$, then

$$x + x_0 \in B(x_0;\varepsilon) \subset \overline{M}_n \quad \text{and} \quad x - x_0 \in B(-x_0;\varepsilon) = -B(x_0;\varepsilon) \subset \overline{M}_n,$$

where the last inclusion follows from the symmetry of $\overline{M}_n$. Finally, the convexity of $\overline{M}_n$ implies that

$$x = \frac{1}{2}(x - x_0) + \frac{1}{2}(x + x_0) \in \overline{M}_n.$$

Step 3. It will be shown that $B(0;\varepsilon) \subset M_{2n}$. Let $x \in B(0;\varepsilon) \subset \overline{M}_n$. There exists an element $x_1 \in M_n$ such that $\|x - x_1\| < \frac{1}{2}\varepsilon$. Since $2(x - x_1) \in B(0;\varepsilon) \subset \overline{M}_n$, there exists an element $x_2 \in M_n$ such that $\|x - x_1 - \frac{1}{2}x_2\| < \frac{1}{4}\varepsilon$. By induction, there exists a sequence (x_j) in M_n such that for every $k \in \mathbb{N}$:

$$\left\| x - \sum_{j=1}^{k} \frac{x_j}{2^{j-1}} \right\| < \frac{\varepsilon}{2^k}.$$

By taking $k \to \infty$ it follows that $x = \sum_{j=1}^{\infty} x_j/2^{j-1}$, so that by hypothesis

$$p(x) \le \sum_{j=1}^{\infty} \frac{p(x_j)}{2^{j-1}} \le \sum_{j=1}^{\infty} \frac{n}{2^{j-1}} = 2n,$$

which implies that $x \in M_{2n}$.

Step 4. Finally, for $x \neq 0$ it follows that $\varepsilon x/2\|x\| \in B(0;\varepsilon) \subset M_{2n}$. This implies that $p(\varepsilon x/2\|x\|) \le 2n$ so that $p(x) \le C\|x\|$ with $C = 4n/\varepsilon$. This completes the proof. ☺

5.3 The open mapping theorem

Definition 5.5. Let X and Y be Banach spaces and let $T \in B(X,Y)$. Then the map T is called *open* if $T(O)$ is open in Y for every open set O in X.

Theorem 5.6 (Open mapping theorem)**.** Let X and Y be Banach spaces and let $T \in B(X,Y)$. If the map T is surjective, then T is an open map.

Proof. The proof will be given in three steps.

Step 1. Since T is surjective, the following map is well-defined:

$$p : Y \to [0,\infty), \quad p(y) = \inf\{\|x\| : x \in X,\, Tx = y\}.$$

For a convergent series $\sum_{n=1}^{\infty} y_n$ in Y it will be shown that

$$p\left(\sum_{n=1}^{\infty} y_n\right) \le \sum_{n=1}^{\infty} p(y_n). \tag{5.1}$$

This is clear if the right-hand side is infinite. Hence, it may be assumed that the right-hand side is finite. For fixed $\varepsilon > 0$ there exists a sequence (x_n) in X such that $Tx_n = y_n$ and $\|x_n\| < p(y_n) + \varepsilon/2^n$, which implies that the series $\sum_{n=1}^{\infty} x_n$ is absolutely convergent and hence convergent by Theorem 3.13. Therefore,

$$T\left(\sum_{n=1}^{\infty} x_n\right) = \sum_{n=1}^{\infty} Tx_n = \sum_{n=1}^{\infty} y_n$$

so that

$$p\left(\sum_{n=1}^{\infty} y_n\right) \le \left\|\sum_{n=1}^{\infty} x_n\right\| \le \sum_{n=1}^{\infty} \|x_n\| < \sum_{n=1}^{\infty} p(y_n) + \varepsilon.$$

Since $\varepsilon > 0$ is arbitrary the inequality in (5.1) follows.

Step 2. In particular, it follows from Step 1 that

$$p(y_1 + y_2) \le p(y_1) + p(y_2).$$

With $\lambda \neq 0$ it follows that

$$\begin{aligned} p(\lambda y) &= \inf\{\|x\| : x \in X,\, T(\lambda^{-1}x) = y\} \\ &= \inf\{\|\lambda x\| : x \in X,\, Tx = y\} \\ &= |\lambda|\, p(y). \end{aligned}$$

Since $p(0) = 0$ by definition, this relation also holds for $\lambda = 0$. Hence p is a semi-norm.

Step 3. Zabreĭko's lemma implies that $p : Y \to [0,\infty)$ is continuous. Hence, the set

$$T(B_X(0;1)) = \{y \in Y : p(y) < 1\}$$

is open. Let $O \subset X$ be nonempty and open, and let $y \in T(O)$, so that $y = Tx$ with $x \in O$. There exist $\delta > 0$ and $\varepsilon > 0$ such that

$$B_X(x;\delta) \subset O \quad \text{and} \quad B_Y(0;\varepsilon) \subset T(B_X(0;1)).$$

This implies that

$$\begin{aligned} B_Y(y;\delta\varepsilon) &= Tx + \delta B_Y(0;\varepsilon) \\ &\subset Tx + \delta T(B_X(0;1)) \\ &= T(B_X(x;\delta)) \\ &\subset T(O), \end{aligned}$$

which shows that $T(O)$ is open. This completes the proof. ☺

Corollary 5.7. Let X and Y be Banach spaces and let $T \in B(X,Y)$. If the map T is bijective, then $T^{-1} \in B(Y,X)$.

Example 5.8. If X and Y are normed linear spaces and $T \in B(X,Y)$ is bijective, then it is not necessarily true that $T^{-1} \in B(Y,X)$. Consider $X = s$, the linear subspace of ℓ^∞ consisting of all sequences $x \in \ell^\infty$ which are of the form

$$x = (x_1, x_2, \dots, x_m, 0, 0, \dots)$$

for some m, depending on x, and let T be defined by

$$T(x_1, x_2, x_3, \dots) = (x_1, \tfrac{1}{2}x_2, \tfrac{1}{3}x_3, \dots), \quad x \in s.$$

Then T is a bounded linear map from s onto s, but T^{-1} is not bounded. □

Corollary 5.9. Let X be a normed linear space with norms $\|\cdot\|_1$ and $\|\cdot\|_2$ for which

$$\|x\|_1 \leq M\|x\|_2, \quad x \in X.$$

If X is a Banach space with respect to each norm, then the norms $\|\cdot\|_1$ and $\|\cdot\|_2$ are equivalent.

Proof. The assumptions imply that the identity operator I taking the Banach space $(X, \|\cdot\|_2)$ into the Banach space $(X, \|\cdot\|_1)$ is bounded. Furthermore I is injective and surjective, implying that I^{-1} is a bounded operator taking $(X, \|\cdot\|_1)$ into $(X, \|\cdot\|_2)$, so that

$$\|x\|_2 \leq m\|x\|_1, \quad x \in X.$$

The two inequalities lead to the norms being equivalent. ☺

Proposition 5.10. Let X be a Banach space, let Y be a normed linear space, and let $T \in B(X,Y)$. If there exists $c > 0$ such that

$$\|Tx\| \geq c\|x\|, \quad x \in X, \tag{5.2}$$

then T is injective and $\operatorname{ran} T$ is closed in Y. Moreover, if Y is Banach space, T is injective, and $\operatorname{ran} T$ is closed in Y, then there exists $c > 0$ such that (5.2) holds.

Proof. Assume that (5.2) holds. Then $Tx = 0$ implies that $x = 0$, which shows that T is injective. Now let $Tx_n \to y$ in Y. Then (Tx_n) is Cauchy in Y and via (5.2) one obtains

$$\|Tx_n - Tx_m\| = \|T(x_n - x_m)\| \geq c\|x_n - x_m\|,$$

so that (x_n) is Cauchy in X. Since X is a Banach space there exists $x \in X$ such that $x_n \to X$. This implies $Tx_n \to Tx$, which shows $y = Tx \in \operatorname{ran} T$. Hence $\operatorname{ran} T$ is closed.

Now assume that Y is Banach space, T is injective, and that $\operatorname{ran} T$ is closed in Y. In fact, it follows that $\operatorname{ran} T$ as a closed linear subspace of a Banach space is itself a Banach space. Thus $T \in B(X,Y)$ induces a bijective bounded map between the Banach space X and the Banach space $\operatorname{ran} T$. Therefore by Corollary 5.7 it follows that $T^{-1} \in B(\operatorname{ran} T, X)$. Hence there exists $\gamma > 0$ such that $\|T^{-1}y\| \leq \gamma\|y\|$ for all $y \in \operatorname{ran} T$, or with $y = Tx$ one gets $\|x\| \leq \gamma\|Tx\|$. ☺

Corollary 5.11. Let X and Y be Banach spaces and let $T \in B(X,Y)$. Then the following statements are equivalent:

1. T is invertible;
2. $\operatorname{ran} T$ is dense in Y and $\|Tx\| \geq c\|x\|$, $x \in X$, for some $c > 0$.

Equivalently, $T \in B(X,Y)$ is not invertible if and only if either $\operatorname{ran} T$ is not dense or there exists a sequence (x_n) in X with $\|x_n\| = 1$ and $Tx_n \to 0$.

5.4 The closed graph theorem

To appreciate the following definition and results remember that the differentiation operator D in Example 4.14 from $X = \mathcal{C}([a,b],\mathbb{K})$ to $Y = \mathcal{C}([a,b],\mathbb{K})$ has a "natural" domain of definition $V = \mathcal{C}^1([a,b],\mathbb{K})$ where $V \subset X$ is a proper linear subspace. A similar situation occurs with the Laplace operator Δ in $X = L^p(\mathbb{R}^d)$. In order to define a linear operator one chooses a "natural" domain of definition $V = C_0^\infty(\mathbb{R}^d)$ or $V = C_0^2(\mathbb{R}^d)$ with $V \subset X$ a proper linear subspace. The basic message is that in many applications linear operators are not defined on the full space, but rather on a proper subspace. However this general line of thinking will not be pursued in the present notes.

Definition 5.12. Let X and Y be normed linear spaces, let $V \subset X$ be a linear subspace, and let $T : V \to Y$ be a linear map. The *graph* $G(T)$ of T is defined as the linear subset of the product space $X \times Y$ by

$$G(T) = \{(x,Tx) : x \in V\}.$$

The map T is called *closed* if its graph is closed in the normed linear space $X \times Y$.

Lemma 5.13. Let X and Y be normed linear spaces, let $V \subset X$ be a closed linear subspace, and let $T \in B(V,Y)$. Then the graph

$$G(T) = \{(x,Tx) : x \in V\}$$

is closed in the product space $X \times Y$.

Proof. Let $((x_n,Tx_n))$ be a sequence in the graph $G(T)$ converging to an element (x,y). Then $x_n \to x$ and $Tx_n \to y$. Since V is closed it follows from $x_n \to x$ that $x \in V$ and since $T \in B(V,Y)$ it follows from $x_n \to x$ that $Tx_n \to Tx$. Hence it follows that $y = Tx$ and $(x,y) = (x,Tx)$ is in the graph $G(T)$ of T. ☺

Theorem 5.14 (Closed graph theorem)**.** Let X and Y be Banach spaces, let $V \subset X$ be a closed linear subspace, and let $T : V \to Y$ be a linear map. If the graph $G(T)$ of T is closed, then $T \in B(V,Y)$.

Proof. Note that V is a Banach space. Define the map

$$p : V \to [0,\infty), \quad p(x) = \|Tx\|, \quad x \in V,$$

which is clearly a semi-norm on V. For a convergent series $\sum_{n=1}^\infty x_n$ in V it will be shown that

$$p\left(\sum_{n=1}^\infty x_n\right) \le \sum_{n=1}^\infty p(x_n). \tag{5.3}$$

Clearly, this is true when the right-hand side is infinite. Hence, it may be assumed that the right-hand side is finite, in which case it follows that the series $\sum_{n=1}^\infty Tx_n$ is absolutely convergent and therefore convergent by Theorem 3.13.

Since

$$\sum_{n=1}^m x_n \to \sum_{n=1}^\infty x_n \quad \text{and} \quad T\left(\sum_{n=1}^m x_n\right) = \sum_{n=1}^m Tx_n \to \sum_{n=1}^\infty Tx_n$$

as $m \to \infty$ it follows from the closedness of $G(T)$ that

$$T\left(\sum_{n=1}^{\infty} x_n\right) = \sum_{n=1}^{\infty} Tx_n.$$

Therefore,

$$\left\| T\left(\sum_{n=1}^{\infty} x_n\right)\right\| = \left\| \sum_{n=1}^{\infty} Tx_n \right\| \le \sum_{n=1}^{\infty} \|Tx_n\|,$$

which implies inequality (5.3). From Zabreĭko's lemma it follows that the semi-norm $p : V \to [0,\infty)$ is bounded, which implies that T is bounded. This completes the proof. ☺

Theorem 5.15 (Extension theorem)**.** Let X be a normed linear space and let Y be a Banach space. Let $V \subset X$ be a linear subspace which is dense and assume that $S \in B(V,Y)$. If the operator S is closed then $V = X$; in other words the continuous extension of S coincides with S.

Proof. Let $x \in X$, then there exists a sequence (v_n) in V such that $v_n \to x$. Since S is bounded

$$\|Sx_n - Sx_m\| = \|S(x_n - x_m)\| \le \|S\|\|x_n - x_m\|,$$

it follows that (Sx_n) is Cauchy in Y. By assumption $Sx_n \to y$ for some $y \in Y$. The assumption that S is closed implies that $(x,y) \in G(S)$ and $x \in V$. Thus $X = V$. ☺

5.5 Complemented subspaces and bounded projections

Lemma 5.16. Let X be a normed linear space and let $P \in B(X)$ be a projection. Then also $I - P \in B(X)$ is a projection. Moreover,
1. either $P = 0$ or $\|P\| \ge 1$;
2. $\ker P$ and $\operatorname{ran} P$ are closed linear subspaces of X;
3. $X = \operatorname{ran} P + \ker P$, direct sum.

Proof. (1) It follows from $\|P\| = \|P^2\| \le \|P\|^2$ that either $\|P\| = 0$ or $\|P\| \ge 1$.
(2) Since $P \in B(X)$ one sees that $\ker P$ is closed in X. Moreover, $\operatorname{ran} P = \ker\,(I-P)$ shows that $\operatorname{ran} P$ is closed.
(3) This is clear. ☺

Theorem 5.17. Let X be a Banach space and let V and W be closed linear subspaces of X for which

$$X = V + W, \quad \text{direct sum.}$$

Then the projections onto V and W are bounded, respectively.

Proof. Assume that V and W are closed linear subspaces of X for which $X = V + W$ and $V \cap W = \{0\}$. For every $x \in X$ there exists unique elements $v \in V$ and $w \in W$ such that $x = v + w$. Define the operator P from X into itself by $Px = v$. It is clear that P is linear and that $P^2 = P$. In order to show that $P \in B(X)$ it suffices to show that P is closed.

Therefore, assume that $(x_n, Px_n) \to (x,y)$. Then $x_n \to x$ and $Px_n \to y$. Since V is closed it follows that $y \in V$. Moreover, due to $x_n - Px_n \in W$ it follows that

$$x - y = \lim_{n\to\infty}(x_n - Px_n) \in W,$$

and $0 = P(x-y) = Px - y$. Hence $y = Px$. Thus P is closed and, hence, by Theorem 5.14 P is bounded. ☺

Definition 5.18. Let X be a normed linear space and let $V \subset X$ be a closed linear subspace. Then V is said to be *complemented* in X if there exists a complement, i.e., a closed linear subspace $W \subset X$ such that

$$X = V + W, \quad \text{direct sum}.$$

Theorem 5.19. Let X be a Banach space and let V be a closed linear subspace of X such that there exists a bounded projection onto V. Then V is complemented in X.

Proof. Let X be a Banach space and let V be a closed linear subspace of X such that there is a bounded projection P onto V. Then $V = \operatorname{ran} P$ and $W = \ker P$ satisfy $X = V + W$ and $V \cap W = \{0\}$; see Lemma 5.16. ☺

Let X be a Banach space and let $V \subset X$ be a closed linear subspace. Then according to Theorem 5.17 and Theorem 5.19 the subspace V is complemented in X if and only if there exists a projection $P \in B(X)$ onto V. In general in a Banach space a closed linear subspace need not be complemented; this does not contradict Corollary 1.43 in which topology plays no role. However in a normed linear space any finite-dimensional subspace is complemented; see Lemma 7.8.

In a Hilbert space the situation is more satisfactory. First consider Theorem 5.17 in the following context. Let X be a Hilbert space and let V and W be closed subspaces of X for which $V \perp W$, i.e., $W \subset V^{\perp}$, and

$$X = V \oplus W,$$

i.e., let the decomposition in Theorem 5.17 be orthogonal. Comparing with Theorem 3.25 one sees that now $W = V^{\perp}$. Let P be the projection from X onto V. Then according to Theorem 3.25 any $x \in X$ decomposes as

$$x = v + w \quad \text{with} \quad \|x\|^2 = \|v\|^2 + \|w\|^2 \quad \text{where} \quad v \in V,\ w \in W.$$

Hence $\|Px\| = \|v\| \leq \|x\|$, so that $P \in B(X)$ and $\|P\| \leq 1$. Recall that if $P \neq 0$ then $\|P\| \geq 1$; cf. Lemma 5.16. Therefore $\|P\| = 1$ if P is not trivial.

Definition 5.20. Let X be a Hilbert space. Then a projection $P \in B(X)$ is *orthogonal* if

$$\operatorname{ran} P \perp \ker P.$$

Lemma 5.21. Let X be a Hilbert space and let V be a closed subspace of X. Then V is orthocomplemented, i.e., there is an orthogonal complement $V^{\perp}$, and there is an orthogonal projection P onto V with norm $\|P\| = 1$, when P is not trivial.

Closely related to the above ideas is the question when the sum of two closed linear subspaces of a Banach space is closed.

Proposition 5.22. Let X be a Banach space and let V and W be closed linear subspaces of X with $V \cap W = \{0\}$. Assume that $V + W$ is closed then for some $c > 0$

$$c\|v\| \leq \|v + w\|, \ \ c\|w\| \leq \|v + w\|, \ \ v \in V,\ w \in W. \tag{5.4}$$

Conversely, if there exists $c > 0$ for which one of the inequalities in (5.4) holds, then $V + W$ is closed. Moreover, in this case, both inequalities in (5.4) hold with c replaced by the smaller positive constant $c/(c+1)$.

Proof. Assume that the subspace $Y = V + W$ is closed in X. Then in fact $Y = V + W$ is a Banach space; see Proposition 3.6. The condition $V \cap W = \{0\}$ implies that the sum $Y = V + W$ is direct. Thus the projection from Y onto V is bounded; cf. Theorem 5.17. In other words, if $x = v + w \in Y$ with $v \in V$ and $w \in W$, then there exists $b > 0$ such that

$$\|v\| \leq b\|x\| = b\|v + w\|, \quad v \in V, \quad w \in W.$$

which gives one of the inequalities of (5.4). The other one is obtained in the same way.

Conversely, assume there is $c > 0$ that $\|v + w\| \geq c\|v\|$ for all $v \in V$ and $w \in W$. Assume that there is sequence $u_n \in V + W$ with $u_n \to u \in X$. There are unique elements $v_n \in V$ and $w_n \in W$ such that $u_n = v_n + w_n$. As (u_n) is Cauchy in $V + W$ it follows that (v_n) is Cauchy in V. Since V is closed one has $v_n \to v$ for some $v \in V$. This means that $w_n = u_n - v_n \to u - v$ and that $u - v \in W$ since W is closed. Hence $u = v + w$ with $v \in V$ and $w \in W$. Therefore $V + W$ is closed and both inequalities hold with a smaller positive constant. To be a bit more specific: if $\|v + w\| \geq c\|v\|$ for all $v \in V$ and $w \in W$, then

$$\|w\| = \|v + w - v\| \leq \|v + w\| + \|v\| \leq \frac{c+1}{c}\|v + w\|,$$

which leads to

$$\frac{c}{c+1}\|v\| \leq c\|v\| \leq \|v + w\| \quad \text{and} \quad \frac{c}{c+1}\|w\| \leq \|v + w\|.$$ ☺

It may happen that V and W are closed subspaces of X with $V \cap W = \{0\}$ but without $V + W$ being closed.

Example 5.23. Let $(e_n)_{n=1}^{\infty}$ be an orthonormal basis in a Hilbert space X. Define the closed linear subspaces V and W of X by

$$V = \overline{\text{span}}\{e_{2n+1} : n \in \mathbb{N} \cup \{0\}\} \quad \text{and} \quad W = \overline{\text{span}}\{w_n : n \in \mathbb{N} \cup \{0\}\},$$

where (w_n) is the orthonormal system defined by

$$w_n = \frac{n}{\sqrt{n^2+1}}\, e_{2n+1} + \frac{1}{\sqrt{n^2+1}} e_{2n+2}, \quad n \in \mathbb{N} \cup \{0\}.$$

Then it is clear that $V + W$ is a direct sum and that $V + W$ is dense in X. However, $V + W$ is not closed, i.e., $V + W \neq X$. To see this, let $x \in X$ be defined by

$$x = \sum_{n=0}^{\infty} \frac{1}{\sqrt{n^2+1}} e_{2n+2}.$$

Assume that $x = v + w$ with $v \in V$ and $w \in W$:

$$\sum_{n=0}^{\infty} \frac{1}{\sqrt{n^2+1}} e_{2n+2} = \sum_{n=0}^{\infty} a_{2n+1} e_{2n+1} + \sum_{n=0}^{\infty} b_n w_n$$

for some sequences $(a_{2n+1}), (b_n) \in \ell^2$. Taking the inner product with e_{2m+2} shows that $b_n = 1$, a contradiction. □

5.6 The uniform boundedness principle

The uniform boundedness principle provides a powerful tool to conclude boundedness of operators in various limiting circumstances. The formulation is rather general, but think of a sequence of bounded operators.

Theorem 5.24 (Uniform boundedness principle)**.** Let X be a Banach space and let Y be a normed linear space. Let $F \subset B(X,Y)$ and assume that

$$\sup_{T \in F} \|Tx\| < \infty \quad \text{for all} \quad x \in X.$$

Then the elements $T \in F$ are uniformly bounded:

$$\sup_{T \in F} \|T\| < \infty.$$

Proof. Define the map

$$p : X \to [0,\infty), \quad p(x) = \sup_{T \in F} \|Tx\|,$$

which is clearly a semi-norm on X. If $\sum_{n=1}^{\infty} x_n$ is a convergent series in X and $T \in F$, then

$$\left\| T\left(\sum_{n=1}^{\infty} x_n \right) \right\| = \left\| \sum_{n=1}^{\infty} Tx_n \right\| \leq \sum_{n=1}^{\infty} \|Tx_n\| \leq \sum_{n=1}^{\infty} p(x_n),$$

so taking the supremum over all $T \in F$ implies that

$$p\left(\sum_{n=1}^{\infty} x_n \right) \leq \sum_{n=1}^{\infty} p(x_n).$$

Zabreĭko's lemma implies that the semi-norm $p : X \to [0,\infty)$ is bounded. Hence, there exists a constant $C \geq 0$ such that for each $T \in F$ it follows that

$$\|Tx\| \leq C\|x\|, \quad x \in X,$$

which implies that $\|T\| \leq C$ for each $T \in F$. This completes the proof. ☺

Corollary 5.25. Let X be a Banach space and let Y be a normed linear space, and let $T_n \in B(X,Y)$. If $\sup_{n\in\mathbb{N}} \|T_n x\| < \infty$ for all $x \in X$, then $\sup_{n\in\mathbb{N}} \|T_n\| < \infty$.

Corollary 5.26. Let X be a Banach space and let Y be a normed linear space, and let $T_n \in B(X,Y)$. If $\sup \|T_n\| = \infty$ then there exists an element $x \in X$ for which

$$\sup_{n\in\mathbb{N}} \|T_n x\| = \infty.$$

Theorem 5.27. Let X be a Banach space and let Y be a normed linear space, and let $T_n \in B(X,Y)$. If $T_n x$ converges in Y for every $x \in X$, then the sequence (T_n) is uniformly bounded: $\sup_n \|T_n\| < \infty$. Moreover,

$$Tx = \lim_{n\to\infty} T_n x$$

defines an operator $T \in B(X,Y)$ and $\|T\| \leq \sup_n \|T_n\|$.

Proof. It is clear that $T : X \to Y$ is a linear map. It will be shown that T is bounded. Since $T_n x$ converges in Y for every $x \in X$, one sees that

$$\sup_n \|T_n x\| < \infty.$$

Hence, by Theorem 5.24, it follows that $\sup_n \|T_n\| < \infty$. Therefore for all $x \in X$:

$$\|Tx\| = \lim_{n\to\infty} \|T_n x\| \leq \left(\sup_n \|T_n\| \right) \|x\|.$$ ☺

Corollary 5.28. Let X and Y be Banach spaces, let $E \subset X$ be a dense subset of X, and let $T_n \in B(X,Y)$. Then the sequence $(T_n x)$ converges in Y for every $x \in X$ if and only if

1. $\sup_n \|T_n\| < \infty$;
2. for each $x \in E$ the sequence $(T_n x)$ is Cauchy in Y.

Proof. ($\Rightarrow$) If $(T_n x)$ converges in Y for every $x \in X$, then (1) follows from Theorem 5.27. The statement (2) is clear.

($\Leftarrow$) For any $x \in X$ and $v \in E$ it follows that

$$\begin{aligned}\|T_n x - T_m x\| &\le \|T_n x - T_n v\| + \|T_n v - T_m v\| + \|T_m v - T_m x\| \\ &\le \|T_n\|\|x - v\| + \|T_n v - T_m v\| + \|T_m\|\|x - v\|.\end{aligned}$$

The assumptions imply that $(T_n x)$ is a Cauchy sequence. Since Y is complete, then there exists $Tx \in Y$ such that $T_n x \to Tx$. ☺

Example 5.29 (Fourier series)**.** The uniform boundedness principle provides a technique to show that there exist continuous functions for which the Fourier series

$$\frac{1}{2}a_0 + \sum_{m=1}^{\infty} (a_n \cos mx + b_n \sin mx), \quad -\pi \le x \le \pi, \tag{5.5}$$

is not convergent in a given point $x \in [-\pi, \pi]$. In order to demonstrate this, let $X = \mathcal{C}([-\pi,\pi],\mathbb{R})$ and $x = 0$. Then the n-th partial sum corresponding to (5.5) with $x = 0$ is given by

$$S_n(f) = a_0 + \sum_{m=1}^{n} a_m = \frac{1}{2\pi}\int_{-\pi}^{\pi} q_n(t) f(t)\,dt, \quad q_n(t) = \frac{\sin(n+\frac{1}{2})t}{\sin\frac{1}{2}t}.$$

It is clear that

$$|S_n(f)| \le \left(\frac{1}{2\pi}\int_{-\pi}^{\pi} |q_n(t)|\,dt\right)\|f\|_\infty,$$

so that S_n is a bounded linear operator from X to $\mathbb{R}$ with

$$\|S_n\| \le \frac{1}{2\pi}\int_{-\pi}^{\pi} |q_n(t)|\,dt$$

In fact there is equality. To see this, consider the function h defined by

$$h(t) = 1 \text{ when } q_n(t) \ge 0, \quad h(t) = -1 \text{ elsewhere.}$$

This function h is not continuous, but for every $\varepsilon > 0$ there is a continuous function f with $\|f\|_\infty = 1$ such that

$$\begin{aligned}\left|S_n(f) - \frac{1}{2\pi}\int_{-\pi}^{\pi} |q_n(t)|\,dt\right| &= \left|\frac{1}{2\pi}\int_{-\pi}^{\pi} q_n(t)f(t)\,dt - \frac{1}{2\pi}\int_{-\pi}^{\pi} q_n(t)h(t)\,dt\right| \\ &= \left|\frac{1}{2\pi}\int_{-\pi}^{\pi} q_n(t)(f(t)-h(t))\,dt\right| \le \frac{1}{2\pi}\int_{-\pi}^{\pi} |q_n(t)||f(t)-h(t)|\,dt < \varepsilon.\end{aligned}$$

As a direct consequence one sees that

$$\|S_n\| = \frac{1}{2\pi}\int_{-\pi}^{\pi} |q_n(t)|\,dt,$$

and it is not difficult to show that therefore $\|S_n\| \to \infty$. Thus according to Theorem 5.27 there exists a function $f \in X$ for which the sequence $(S_n f)$ does not converge. In other words, there exists a function $f \in \mathcal{C}([-\pi,\pi],\mathbb{R})$ for which the Fourier series (5.5) does not converge in $x = 0$. □

5.7 The spectrum of bounded linear operators

Let X be a finite-dimensional linear space over $\mathbb{C}$ of dimension n and let T be a linear operator in X (which is automatically bounded). Then there are precisely n eigenvalues and if $\lambda \in \mathbb{C}$ is not an eigenvalue then $T-\lambda$ is a bijection so that $(T-\lambda)^{-1}$ is an operator which is defined on all of X. In the infinite-dimensional situation things are different.

Definition 5.30. Let X be a Banach space over $\mathbb{K}$ and let $T \in B(X)$. The *resolvent set* $\rho(T)$ of T is defined by

$$\rho(T) = \{\lambda \in \mathbb{K} : T-\lambda \text{ is invertible}\}.$$

The *resolvent operator* $R(\lambda)$ is defined by

$$R(\lambda) = (T-\lambda)^{-1}, \quad \lambda \in \rho(T).$$

The *spectrum* $\sigma(T)$ of T is defined by

$$\sigma(T) = \mathbb{C} \setminus \rho(T) = \{\lambda \in \mathbb{K} : T-\lambda \text{ is not invertible}\}.$$

Lemma 5.31. Let X be a Banach space over $\mathbb{K}$ and let $T \in B(X)$. If $|\lambda| > \|T\|$, then $\lambda \in \rho(T)$, which implies in particular that $\sigma(T) \subset \{z \in \mathbb{C} : |z| \le \|T\|\}$. Moreover,

$$R(\lambda) = -\sum_{n=0}^{\infty} \frac{T^n}{\lambda^{n+1}}, \quad |\lambda| > \|T\|. \tag{5.6}$$

Proof. For $T \in B(X)$ and $\lambda \neq 0$ one writes

$$T-\lambda = -\lambda\left(I - \frac{T}{\lambda}\right). \tag{5.7}$$

Assume that $|\lambda| > \|T\|$, so that $\|T\|/|\lambda| < 1$. Then (5.7) and Theorem 4.35 show that $T-\lambda$ is invertible and

$$(T-\lambda)^{-1} = -\frac{1}{\lambda}\left(I - \frac{T}{\lambda}\right)^{-1},$$

which implies that (5.6) holds. ☺

Lemma 5.32. Let X be a Banach space, let $T \in B(X)$, and let $\mu \in \rho(T)$. Assume that $|\lambda - \mu|\|R(\mu)\| < 1$, then $\lambda \in \rho(T)$. Hence, $\rho(T)$ is open and $\sigma(T)$ is closed. Moreover,

$$R(\lambda) = \sum_{n=0}^{\infty} (\lambda-\mu)^n R(\mu)^{n+1}, \tag{5.8}$$

with convergence in $B(X)$. In particular, for any $x \in X$ and $f \in X'$ the scalar function h, defined by

$$h(\lambda) = f(R(\lambda)x),$$

has the Taylor expansion

$$h(\lambda) = \sum_{n=0}^{\infty} c_n(\lambda-\mu)^n, \quad c_n = f(R(\mu)^{n+1}x).$$

As a consequence, the map $\lambda \mapsto h(\lambda)$ from $\rho(T)$ to $\mathbb{K}$ is holomorphic.

Proof. Assume that $\mu \in \rho(T)$ and write $T-\lambda$ in terms of $T-\mu$ via

$$T-\lambda = T-\mu-(\lambda-\mu)I = [I-(\lambda-\mu)(T-\mu)^{-1}](T-\mu). \tag{5.9}$$

Assume further that $|\lambda-\mu|\|(T-\mu)^{-1}\| < 1$, so that $I-(\lambda-\mu)(T-\mu)^{-1}$ is invertible (cf. Theorem 4.38 and replace S and T by $T-\lambda$ and $T-\mu$, respectively). Then each factor in the right-hand side of (5.9) is invertible and thus the left-hand side of (5.9) is invertible. Hence it follows that $\lambda \in \rho(T)$, so that $\rho(T)$ is open. In particular, $\sigma(T)$ is closed.

Moreover, it follows from (5.9) that

$$(T-\lambda)^{-1} = (T-\mu)^{-1}\sum_{n=0}^{\infty}(\lambda-\mu)^n[(T-\mu)^{-1}]^n,$$

cf. Theorem 4.35. This leads to (5.8). By Corollary 4.37 one sees that for all $x \in X$

$$R(\lambda)x = \sum_{n=0}^{\infty}(\lambda-\mu)^n R(\mu)^{n+1}x,$$

with convergence in X. Hence for any $f \in X'$ it follows that

$$f(R(\lambda)x) = \sum_{n=0}^{\infty}(\lambda-\mu)^n f(R(\mu)^{n+1}x). \qquad ☺$$

Corollary 5.33. Let X be a Banach space and let $T \in B(X)$. Then on $\rho(T)$ one has the resolvent identity

$$R(\lambda)-R(\mu) = (\lambda-\mu)R(\lambda)R(\mu), \quad \lambda,\mu \in \rho(T).$$

Proof. Note that for $|\lambda-\mu|\|(T-\mu)^{-1}\| < 1$ this identity follows from (5.8). In general, observe that with $\lambda,\mu \in \rho(T)$

$$\begin{aligned} R(\lambda)-R(\mu) &= (T-\lambda)^{-1}-(T-\mu)^{-1} \\ &= (T-\lambda)^{-1}[T-\mu-(T-\lambda)](T-\mu)^{-1} \\ &= (\lambda-\mu)R(\lambda)R(\mu). \end{aligned} \qquad ☺$$

Example 5.34. Let X be a finite-dimensional Banach space and let T be a linear operator in X, then automatically $T \in B(X)$. The spectrum of T consists of at most n eigenvalues, where $n = \dim X$. Outside the eigenvalues the resolvent operator is given by $R(\lambda) = (T-\lambda)^{-1}$. □

Example 5.35. Let X be a Hilbert space with an orthonormal basis (e_n), $n \in \mathbb{N}$, and let (λ_n) be a bounded sequence in $\mathbb{C}$. Then the multiplication operator $T \in B(X)$ is defined by

$$Tx = \sum_{n=1}^{\infty}\lambda_n\langle x,e_n\rangle e_n, \quad x \in X.$$

Clearly all λ_n belong to the point spectrum of T. Note that for every $\varepsilon > 0$ the inequality $|\lambda - \lambda_n| \geq \varepsilon$ implies that $\lambda \in \rho(T)$ and that

$$(T-\lambda)^{-1}y = \sum_{n=1}^{\infty} \frac{1}{\lambda_n - \lambda} \langle y, e_n \rangle e_n, \quad y \in X.$$

The spectrum is given by $\sigma(T) = \operatorname{clos}\{\lambda_n : n \in \mathbb{N}\}$. As in Example 4.49 it follows that T is compact if and only if $\lambda_n \to 0$, in which case $0 \in \sigma(T)$. □

Proposition 5.36. Let X be a Banach space over $\mathbb{K}$ and let $T \in B(X)$. The resolvent set $\rho(T)$ is given by all $\lambda \in \mathbb{K}$ for which

1. $\operatorname{ran}(T-\lambda)$ is dense in X;
2. $\|(T-\lambda)x\| \geq c\|x\|$, $x \in X$, for some $c > 0$.

The spectrum $\sigma(T)$ of T is given by all $\lambda \in \mathbb{K}$ for which either $\operatorname{ran}(T-\lambda)$ is not dense in X or for which there exists a sequence (x_n) in X with $\|x_n\| = 1$ and $(T-\lambda)x_n \to 0$.

Proof. This is a direct consequence of Proposition 5.10 and Corollary 5.11. ☺

Definition 5.37. A point $\lambda \in \mathbb{C}$ for which there exists a sequence (x_n) in X with $\|x_n\| = 1$ and $(T-\lambda)x_n \to 0$ is called an *approximate eigenvalue*. The set of all approximate eigenvalues of T is denoted by $\sigma_{\mathrm{a}}(T)$. Clearly, $\sigma_p(T) \subset \sigma_{\mathrm{a}}(T) \subset \sigma(T)$.

Remark 5.38. Let X be a Banach space and let $T \in B(X)$. Assume that T is an isometry, then some interesting facts present themselves. It is clear that $Tx = \lambda x$, $x \neq 0$, now implies that $|\lambda| = 1$. Moreover, for $|\lambda| < 1$ one has

$$\begin{aligned}
\|(T-\lambda)x\| &= \|Tx - \lambda x\| \\
&\geq \big| \|Tx\| - |\lambda|\,\|x\| \big| \\
&= (1-|\lambda|)\|x\|, \quad x \in X.
\end{aligned}$$

This implies that for $|\lambda| < 1$ there are no approximate eigenvalues and, moreover, that $\operatorname{ran}(T-\lambda)$ is closed. Hence for $|\lambda| < 1$ one has

$$\lambda \in \rho(T) \quad \Leftrightarrow \quad \operatorname{ran}(T-\lambda) = X.$$

□

Example 5.39. As an illustration consider the left shift S_l in ℓ^2:

$$S_l(x_1, x_2, x_3, \dots) = (x_2, x_3, x_4, \dots), \quad x \in \ell^2.$$

Recall that S_l is a contraction with $\|S_l\| = 1$; see Example 4.8 and Example 4.18. The point spectrum of S_l is given by all points of the open unit disk. Indeed, $S_l x = \lambda x$ if and only if $x = (x_1, \lambda x_1, \lambda^2 x_1, \dots)$. If $x_1 \neq 0$, then $x \in \ell^2$ if and only if $|\lambda| < 1$. Clearly if $|\lambda| = 1$ then $\lambda \in \sigma(S_l)$ since $\sigma(S_l)$ is closed. Assume in this case that $a = (a_1, a_2, \dots) \in \ell^2$ satisfies

$$a \perp \operatorname{ran}(S_l - \lambda),$$

so that $\langle a, (S_l - \lambda)x \rangle = 0$ for all $x \in \ell^2$. In particular, when taking unit vectors for x, it follows that $a = 0$. Thus $\operatorname{ran}(S_l - \lambda)$ is dense in ℓ^2 and each λ with $|\lambda| = 1$ must be an approximate eigenvalue. Note that in this case $\operatorname{ran}(S_l - \lambda) \neq X$, otherwise $\lambda \in \rho(S_l)$.

Now consider the right shift S_r in ℓ^2:

$$S_r(x_1,x_2,x_3,\dots) = (0,x_1,x_2,\dots), \quad x \in \ell^2.$$

It is clear that S_r is an isometric operator so that $\|S_r\| = 1$. Moreover, it is clear that S_r has no eigenvalues. For $|\lambda| < 1$ one sees that $\overline{\operatorname{ran}}\,(S_r - \lambda) \neq X$, since

$$(S_r - \lambda)x = (-\lambda x_1, x_1 - \lambda x_2, x_2 - \lambda x_3, \dots) \perp (1, \bar{\lambda}, \bar{\lambda}^2, \dots).$$

One concludes that if $|\lambda| < 1$ then $\lambda \in \sigma(S_r)$. Furthermore, if $|\lambda| = 1$ then $\lambda \in \sigma(S_r)$ since $\sigma(S_r)$ is closed. Moreover, if in this case $a = (a_1, a_2, \dots) \in \ell^2$ satisfies

$$a \perp \operatorname{ran}(S_r - \lambda),$$

so that $\langle a, (S_r - \lambda)x \rangle = 0$ for all $x \in \ell^2$. In particular, when taking unit vectors for x, it follows that $a = (a_1, \bar{\lambda} a_1, \bar{\lambda}^2 a_1, \dots)$ which belongs to ℓ^2 if and only if $a = 0$. Hence $\operatorname{ran}(S_r - \lambda)$ is dense. Note that in this case $\operatorname{ran}(S_r - \lambda) \neq X$, otherwise $\lambda \in \rho(S_r)$.

There is an intimate connection between the spectra of S_l and S_r; see Example 6.12. □

Theorem 5.40 (Spectral mapping theorem)**.** Let X be a Banach space over $\mathbb{C}$ and let $T \in B(X)$. Let P be a polynomial, then

$$\sigma(P(T)) = \{P(\lambda) : \lambda \in \sigma(T)\}.$$

Proof. Assume that P is a polynomial of degree n.

($\subset$) Let $\mu \in \sigma(P(T))$ and write $P(\lambda) - \mu$ in terms of its zeros $\gamma_1, \dots, \gamma_n$ (depending on μ) as

$$P(\lambda) - \mu = a(\lambda - \gamma_1) \cdots (\lambda - \gamma_n), \quad a \neq 0.$$

For the operator T this leads to the factorization

$$P(T) - \mu = a(T - \gamma_1) \cdots (T - \gamma_n).$$

Then for some i one must have $\gamma_i \in \sigma(T)$; otherwise $\mu \in \rho(P(T))$. But since $\mu = P(\gamma_i)$, it follows that $\mu \in P(\sigma(T))$.

($\supset$) Let $\mu \in P(\sigma(T))$ so that $\mu = P(\gamma)$ for some $\gamma \in \sigma(T)$. Then one can write

$$P(\lambda) - \mu = (\lambda - \gamma) Q(\lambda)$$

for some polynomial Q (depending on μ) of degree $n-1$. For the operator T this leads to the factorization

$$P(T) - \mu = (T - \gamma) Q(T) = Q(T)(T - \gamma).$$

Since $\gamma \in \sigma(T)$ there are two possibilities: either $\gamma \in \sigma_{\mathrm{p}}(T)$ or $\gamma \in \sigma(T) \setminus \sigma_{\mathrm{p}}(T)$.

Case $\gamma \in \sigma_{\mathrm{p}}(T)$. Then $\ker(T - \gamma) \neq \{0\}$ and thus $\ker(P(T) - \mu) \neq \{0\}$. One concludes that $\mu \in \sigma_{\mathrm{p}}(P(T)) \subset \sigma(P(T))$.

Case $\gamma \in \sigma(T) \setminus \sigma_{\mathrm{p}}(T)$. Then $\operatorname{ran}(T - \gamma) \neq X$ (otherwise $(T - \gamma)^{-1} \in B(X)$ by the open mapping theorem) and thus $\operatorname{ran}(P(T) - \mu) \neq X$. One concludes that $\mu \in \sigma(P(T))$. ☺

Example 5.41. Let X be a Banach space and let $T \in B(X)$. Then T is called *nilpotent* if for some $n \in \mathbb{N}$ one has $T^n = 0$. According to Theorem 5.40 one has $\lambda \in \sigma(T)$ if and only if $\lambda^n \in \sigma(T^n) = \{0\}$, which implies that $\sigma(T) = \{0\}$. If $\sigma(T) = \{0\}$, then T is called *quasi-nilpotent*. Nilpotent operators are quasi-nilpotent, but the converse is only true in finite-dimensional spaces. □

Definition 5.42. Let X be a Banach space and let $T \in B(X)$. The *spectral radius* of T is defined by

$$r_\sigma(T) = \sup_{\lambda \in \sigma(T)} |\lambda|. \tag{5.10}$$

In other words, $r_\sigma(T)$ is the radius of the largest circle around 0 which still contains a point in $\sigma(T)$.

Remark 5.43. The spectral radius is given by the limiting formula

$$r_\sigma(T) = \lim_{n\to\infty} \sqrt[n]{\|T^n\|} \ \ (\leq \|T\|).$$

Moreover, if $|\lambda| > r_\sigma(T)$ then the series in (5.6) still converges and, in fact, $r_\sigma(T)$ is the smallest nonnegative number with this property. Clearly, if $\sqrt[n]{\|T^n\|} \to 0$, then the only point in $\sigma(T)$ is 0. For example, this is the case for the Volterra integral operator; cf. Remark 4.57. □

For practical purposes the following theorem about the spectrum of a compact operator is presented.

Theorem 5.44. Let X be a normed linear space over $\mathbb{K}$ and let $T \in B(X)$ be a compact operator.
1. Let $\lambda \neq 0$ be an eigenvalue of T. Then ker $(T-\lambda)$ is finite-dimensional.
2. For every $\varepsilon > 0$ the number of eigenvalues λ of T with $|\lambda| \geq \varepsilon$ is finite. In other words, the set of eigenvalues of T is at most countable and 0 is the only possible accumulation point.
3. If X is infinite-dimensional, then $0 \in \sigma(T)$.

Proof. (1) Note that $M = \ker(T-\lambda)$ is a closed linear subspace of X. Let x_n be a sequence in the unit ball of M. Since T is compact $x_n = (1/\lambda)Tx_n$ has a convergent subsequence, so that the unit ball of M is compact. By Theorem 2.32 it follows that M is finite-dimensional.

(2) Assume that for some $\varepsilon > 0$ there is a sequence of distinct eigenvalues (λ_n) with $|\lambda_n| \geq \varepsilon$. Then there exist nontrivial $x_n \in X$ with $Tx_n = \lambda_n x_n$, $n \in \mathbb{N}$. This set of eigenvectors is linearly independent; cf. Theorem 1.16. Define the closed linear subspace

$$M_n = \operatorname{span}\{x_1, \ldots, x_n\},$$

and one sees that T maps M_n into M_n. By Lemma 2.31 there is a sequence (y_n) such that

$$y_n \in M_n, \quad \|y_n\| = 1, \quad \|y_n - x\| \geq 1/2 \text{ for all } x \in M_{n-1}.$$

It will be shown that for $n > m$

$$\|Ty_n - Ty_m\| \geq \varepsilon/2,$$

which implies that (Ty_n) has no convergent subsequence, contradicting the compactness of T.

To prove this claim rewrite $Ty_n - Ty_m$ for $n > m$ as

$$Ty_n - Ty_m = \lambda_n y_n - u, \quad u = Ty_m - (T-\lambda_n)y_n.$$

Since $m \leq n-1$ it is clear that $Ty_m \in M_m \subset M_{n-1}$. Furthermore it follows from the definition that $(T-\lambda_n)y_n \subset M_{n-1}$, as $y_n \in M_n$ means $y_n = \sum_{i=1}^n c_i x_i$, which implies

$$\begin{aligned}(T-\lambda_n)y_n &= (T-\lambda_n)\sum_{i=1}^n c_i x_i \\ &= \sum_{i=1}^n c_i (T-\lambda_n)x_i = \sum_{i=1}^{n-1} c_i(\lambda_i - \lambda_n)x_i \in M_{n-1}.\end{aligned}$$

Hence the element $u = Ty_m - (T - \lambda_n)y_n$ belongs to M_{n-1} and thus

$$\|Ty_n - Ty_m\| = \|\lambda_n y_n - u\| = |\lambda_n|\|y_n - u/\lambda_n\| \geq |\lambda_n|/2 \geq \varepsilon/2,$$

which proves the claim.

(3) Assume that T is invertible, then $I = TT^{-1}$ is compact, cf. Lemma 4.43. By Theorem 2.32 this implies that X is finite-dimensional, a contradiction. ☺

There exist compact operators whose spectrum consist of the origin, which need not be an eigenvalue. By similar methods as above it can be shown that each nonzero point in the spectrum of a compact operator is an eigenvalue. In the next chapter it will be shown that in the special case of a nontrivial compact selfadjoint operator in a Hilbert space there are nonzero eigenvalues and that each nonzero point in the spectrum is an eigenvalue.

Chapter 6

Bounded linear operators in Hilbert spaces

Hilbert spaces are inner product spaces which are complete with respect to the norm induced by the inner product. The inner product gives rise to the notion of orthogonality which makes it possible to decompose Hilbert spaces orthogonally. In particular one can identify the dual of a Hilbert space with itself via the Riesz-Fréchet representation theorem. The theory of bounded linear operators has a particularly elegant appearance in Hilbert spaces. The adjoint of a bounded linear operator will be introduced via the Riesz-Fréchet representation theorem. The notion of selfadjointness for bounded linear operators allows to decompose such operators in terms of orthogonal projections; a special case of which is treated in this chapter. The Hilbert spaces in this chapter are always over $\mathbb{C}$. This will be tacitly assumed in the text.

6.1 The dual space of a Hilbert space

Let X be a Hilbert space, so that in particular X is a Banach space. Hence the dual space $X' = B(X,\mathbb{C})$ is well-defined and, in fact, it is a Banach space; see Definition 4.34. Recall that the norm of an element $f \in X'$ is defined by

$$\|f\| = \sup_{x\neq 0} \frac{|f(x)|}{\|x\|}.$$

The Hilbert space structure of X makes it possible to identify the spaces X and X' in a certain way.

Theorem 6.1 (Riesz-Fréchet)**.** Let X be a Hilbert space. For $y \in X$ the functional $f_y : X \to \mathbb{C}$ defined by

$$f_y(x) = \langle x, y\rangle, \quad x \in X,$$

belongs to X', and in addition, $\|f_y\| = \|y\|$. In fact, for each $f \in X'$ there exists a unique $y \in X$ such that $f = f_y$.

Proof. Let $y \in X$ and let f_y be given by $f_y(x) = \langle x, y\rangle$, $x \in X$. Then f_y is a linear mapping from X to $\mathbb{C}$. Furthermore

$$\|f_y\| = \sup_{x\neq 0} \frac{|f_y(x)|}{\|x\|} = \sup_{x\neq 0} \frac{|\langle x, y\rangle|}{\|x\|} \leq \|y\|,$$

which shows in particular that $f_y \in X'$ and $\|f_y\| \leq \|y\|$. In addition, for $y \neq 0$ one sees that

$$\|y\| = \frac{|f_y(y)|}{\|y\|} = \left| f_y\left(\frac{y}{\|y\|}\right)\right| \leq \|f_y\|,$$

which shows $\|y\| \leq \|f_y\|$. This inequality trivially holds for $y = 0$. Hence it follows that $\|f_y\| = \|y\|$.

Now it will be shown that any $f \in X'$ is of the form $f = f_y$ for some unique $y \in X$. It suffices to assume that f is not trivial. Then the closed linear subspace $\ker f$ is a proper subspace of X, i.e., $(\ker f)^\perp \neq \{0\}$. Let $z \in (\ker f)^\perp$ be nontrivial and assume that $f(z) = 1$. Then

$$x - f(x)z \in \ker f,$$

as $f(x - f(x)z) = f(x) - f(x)f(z) = 0$. Due to $z \in (\ker f)^\perp$ this leads to

$$\langle x - f(x)z, z \rangle = 0,$$

or equivalently

$$f(x) = \langle x, y \rangle \quad \text{with} \quad y = \frac{z}{\|z\|^2}.$$

Hence $f = f_y$. For the uniqueness of $y \in X$, assume that also $f = f_z$ for some $z \in X$. Then $\langle x, y \rangle = \langle x, z \rangle$ or $\langle x, y - z \rangle = 0$ for all $x \in X$, which implies that $y = z$. ☺

Recall that the dual space X' of a Hilbert space X is a Banach space. In fact X' can be viewed as a Hilbert space. To see this, first observe that according to Theorem 6.1

$$f_{\lambda y}(x) = \langle x, \lambda y \rangle = \bar{\lambda} \langle x, y \rangle = \bar{\lambda} f_y(x), \quad x \in X.$$

Therefore the mapping $y \mapsto f_y$ from X onto X' is an isometry which is not linear but conjugate linear:

$$\lambda y + \mu z \mapsto f_{\lambda y + \mu z} = \bar{\lambda} f_y + \bar{\mu} f_z, \quad y, z \in X, \ \lambda, \mu \in \mathbb{C}.$$

Therefore the dual space X' can be viewed as an inner product space when the following definition is used:

$$\langle f_y, f_z \rangle = \langle z, y \rangle, \quad y, z \in X,$$

and then X' is a Hilbert space.

The special structure in a Hilbert space X allows a weakening of earlier notions. For instance, a subset $E \subset X$ is called bounded if there exists $M \geq 0$ such that for all $x \in E$ one has $\|x\| \leq M$. However, there is also the notion of weakly bounded.

Definition 6.2. A subset $E \subset X$ is called *weakly bounded* if for each $y \in X$ there exists $M_y \geq 0$ such that for all $x \in E$ one has $|\langle x, y \rangle| \leq M_y$.

Lemma 6.3. Let X be a Hilbert space and let $E \subset X$ be a subset. Then E is bounded if and only if E is weakly bounded.

Proof. Assume that E is bounded. Let $y \in X$ then for all $x \in E$:

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

so take $M_y = M\|y\|$. Hence E is weakly bounded.

Conversely, assume that E is weakly bounded. For each $x \in E$ define the functional f_x by $f_x(y) = \langle y, x \rangle$. Then $f_x \in X'$ and $\|f_x\| = \|x\|$; cf. Theorem 6.1. By assumption one sees that

$$|f_x(y)| = |\langle y, x \rangle| = |\langle x, y \rangle| \leq M_y, \quad y \in X.$$

Via the uniform boundedness principle one obtains from $\sup_{x \in E} |f_x(y)| \leq M_y$, that $\sup_{x \in E} \|x\| = \sup_{x \in E} \|f_x\| < \infty$. Hence E is bounded. ☺

Recall the definitions of convergence and Cauchy sequences in a Banach space. These definitions will now be modified.

Definition 6.4. Let X be a Hilbert space. A sequence (x_n) is said to *weakly converge* to $x \in X$ if $\langle x_n, y\rangle \to \langle x, y\rangle$ for all $y \in X$. A sequence (x_n) in X is said to be a *weak Cauchy sequence* if the sequence $(\langle x_n, y\rangle)$ is a Cauchy sequence in $\mathbb{K}$ for all $y \in X$.

The usual convergence implies weak convergence; however a weakly convergent sequence need not be convergent in the usual sense. Note that a weak Cauchy sequence is weakly bounded.

Lemma 6.5. Let X be a Hilbert space. Then X is weakly complete: if (x_n) is a weak Cauchy sequence in X then there exists $x \in X$ such that

$$\langle x_n, y\rangle \to \langle x, y\rangle$$

for all $y \in X$.

Proof. For each $y \in X$ the sequence $\langle x_n, y\rangle$ converges in $\mathbb{C}$. Define $f_n \in X'$ by

$$f_n(y) = \langle y, x_n\rangle, \quad y \in Y,$$

so that $\|f_n\| = \|x_n\|$. Define $f : X \to \mathbb{C}$ by

$$f(y) = \lim_{n\to\infty} \langle y, x_n\rangle, \quad y \in Y,$$

where the limit exists by assumption. By uniform boundedness $f \in X'$ and by Theorem 6.1 there exists $x \in X$ such that $f(y) = \langle y, x\rangle$. Hence $\lim_{n\to\infty}\langle y, x_n\rangle = \langle y, x\rangle$, which completes the proof. ☺

Lemma 6.6. Let X be a Hilbert space. Then X is weakly compact: if (x_n) is a bounded sequence in X then there exists a subsequence $x_{n_k} \in X$ such that x_{n_k} converges weakly.

Proof. Assume that $\|x_i\| \le M$ for some $M > 0$. The selection of the subsequence will be done as follows. The sequence of inner products $\langle x_1, x_i\rangle$ is bounded in $\mathbb{C}$ since

$$|\langle x_1, x_i\rangle|^2 \le M^2, \quad i \in \mathbb{N}.$$

Hence there exists a subsequence x_{1k} of x_k such that

$$\lim_{k\to\infty} \langle x_1, x_{1k}\rangle \quad \text{exists.}$$

Now the sequence of inner products $\langle x_2, x_{1k}\rangle$ is bounded in $\mathbb{C}$ by M^2. Hence there exists a subsequence x_{2k} of x_{1k} such that

$$\lim_{k\to\infty} \langle x_2, x_{2k}\rangle \quad \text{exists.}$$

Thus one gets the array of sequences

$$\begin{array}{cccc} x_{11}, & x_{12}, & x_{13}, & \dots \\ x_{21}, & x_{22}, & x_{23}, & \dots \\ x_{31}, & x_{32}, & x_{33}, & \dots \\ \vdots & \vdots & \vdots & \ddots \end{array}$$

where each sequence is a subsequence of the one above it, while for each $n \in \mathbb{N}$ the sequence

$$\langle x_n, x_{n1}\rangle, \langle x_n, x_{n2}\rangle, \langle x_n, x_{n3}\rangle, \dots$$

converges. Thus for any x_r the sequence $\langle x_r, x_{rn}\rangle$ converges and thus the sequence $\langle x_r, x_{nn}\rangle$ converges (the first $r-1$ elements do not disturb the convergence). This shows that $\langle x, x_{nn}\rangle$ converges for each $x \in \operatorname{span}\{x_n\}$, which implies that also $\langle x, x_{nn}\rangle$ converges for each $x \in \overline{\operatorname{span}}\{x_n\}$. Since $\langle x, x_{nn}\rangle = 0$ for all $x \perp \operatorname{span}\{x_n\}$, the proof is complete. ☺

The statements in this section can be put in the more general context of Banach spaces; see Chapter 7.

6.2 Bounded operators in Hilbert spaces

Let X and Y be Banach spaces and let $T: X \to Y$ be a linear operator. Recall that the operator T is called bounded, i.e. $T \in B(X,Y)$, when

$$\sup_{x \neq 0} \frac{\|Tx\|}{\|x\|} < \infty,$$

in which case the left-hand side is denoted by $\|T\|$. Let X and Y be Hilbert spaces and let $T \in B(X,Y)$. The existence of an adjoint operator $T^* \in B(Y,X)$ is contained in the following theorem.

Theorem 6.7. Let X and Y be Hilbert spaces and let $T \in B(X,Y)$. Then there is a unique bounded linear operator $T^* \in B(Y,X)$ with

$$\langle Tx, y\rangle = \langle x, T^*y\rangle, \quad x \in X, \quad y \in Y.$$

Proof. For a fixed $y \in Y$ define the linear map $f: X \to \mathbb{C}$ by $f(x) = \langle Tx, y\rangle$. Then

$$|f(x)| = |\langle Tx, y\rangle| \leq \|Tx\|\,\|y\| \leq \|T\|\,\|x\|\,\|y\|,$$

which shows that $\|f\| \leq \|T\|\|y\|$ so that $f \in X'$. By Theorem 6.1, the Riesz-Fréchet representation theorem, there exists a unique element $u_y \in X$ with $f(x) = \langle x, u_y\rangle$ and $\|u_y\| = \|f\|$. Define the mapping $T^*: Y \to X$ by $T^*y = u_y$, then it is clear that T^* is linear. In addition,

$$\|T^*y\| = \|u_y\| = \|f\| \leq \|T\|\,\|y\|,$$

which implies that $\|T^*\| \leq \|T\|$. Hence $T^* \in B(Y,X)$.

To see the uniqueness assume that $S \in B(Y,X)$ satisfies $\langle Tx, y\rangle = \langle x, Sy\rangle$ for all $x \in X$ and $y \in Y$. Then $\langle x, T^*y\rangle = \langle x, Sy\rangle$ for all $x \in X$ and $y \in Y$. It follows that $S = T^*$. ☺

Lemma 6.8. Let X and Y be Hilbert spaces and let $T \in B(X,Y)$. Then
1. $(T^*)^* = T$;
2. $\|T^*\| = \|T\|$;
3. $\|T^*T\| = \|T\|^2$.

Proof. (1) It follows directly from Theorem 6.7 that

$$\langle (T^*)^*x, y\rangle = \langle x, T^*y\rangle = \langle Tx, y\rangle, \quad x \in X, y \in Y.$$

(2) The proof of Theorem 6.7 shows that $\|T^*\| \leq \|T\|$. The reverse inequality follows by replacing T by T^* and using (1).

(3) It is clear that $\|T^*T\| \leq \|T^*\|\,\|T\| = \|T\|^2$. For the converse observe that

$$\|Tx\|^2 = \langle Tx, Tx\rangle = \langle T^*Tx, x\rangle \leq \|T^*Tx\|\,\|x\| \leq \|T^*T\|\,\|x\|^2,$$

which implies $\|T\|^2 \leq \|T^*T\|$. ☺

Lemma 6.9. Assume that X, Y, and Z are Hilbert spaces and let $\lambda, \mu \in \mathbb{K}$. Then the following statements hold:

1. Let $T, S \in B(X,Y)$, then
$$(\lambda T + \mu S)^* = \bar{\lambda} T^* + \bar{\mu} S^*.$$
In particular, if $T \in B(X)$, then
$$(T - \lambda)^* = T^* - \bar{\lambda}.$$
2. Let $T \in B(X,Y)$ and $S \in B(Y,Z)$, then
$$(ST)^* = T^* S^*.$$
3. If $T \in B(X)$ is invertible, then $T^* \in B(X)$ is invertible and
$$(T^*)^{-1} = (T^{-1})^*.$$
In particular, $\lambda \in \sigma(T)$ if and only if $\bar{\lambda} \in \sigma(T^*)$.
4. Let $T \in K(X,Y)$, then $T^* \in K(Y,X)$.

Proof. (1) and (2) follow directly from the definition.

(3) If $T \in B(X)$ is invertible, then there exists $T^{-1} \in B(X)$ such that
$$TT^{-1} = T^{-1}T = I.$$
By (2) it follows that
$$(T^{-1})^* T^* = (TT^{-1})^* = I^* = I.$$
Similarly, it follows that $T^*(T^{-1})^* = I$. In particular, $T - \lambda$ is invertible if and only if $T^* - \bar{\lambda}$ is invertible, which implies that $\lambda \in \sigma(T)$ if and only if $\bar{\lambda} \in \sigma(T^*)$.

(4) Assume that $T \in K(X,Y)$. By Lemma 4.43 also TT^* is compact. To see that T^* is compact, let (y_n) be a bounded sequence in Y with $\|y_n\| \leq M$. Then there is a subsequence, say y_{n_k}, such that $TT^* y_{n_k}$ converges, and in particular that $TT^* y_{n_k}$ is a Cauchy sequence. It follows from
$$\begin{aligned}
\|T^*(y_n - y_m)\|^2 &= \langle T^*(y_n - y_m), T^*(y_n - y_m)\rangle \\
&= \langle TT^*(y_n - y_m), y_n - y_m\rangle \\
&\leq \|TT^*(y_n - y_m)\| \|y_n - y_m\| \\
&\leq 2M \|TT^*(y_n - y_m)\|,
\end{aligned}$$
that $T^* y_{n_k}$ is a Cauchy sequence, which converges since X is a Hilbert space. Thus $T^* \in K(Y,X)$. ☺

Lemma 6.10. Let X and Y be Hilbert spaces, let $T \in B(X,Y)$, and $\lambda \in \mathbb{C}$. Then

1. $(\operatorname{ran}(T - \lambda))^\perp = \ker(T^* - \bar{\lambda})$;
2. $(\operatorname{ran}(T^* - \bar{\lambda}))^\perp = \ker(T - \lambda)$;
3. In particular, if $T \in B(X)$, then X has the orthogonal decomposition
$$X = \ker(T^* - \bar{\lambda}) \oplus \overline{\operatorname{ran}}(T - \lambda) = \ker(T - \lambda) \oplus \overline{\operatorname{ran}}(T^* - \bar{\lambda}).$$

Proof. (1) Observe that for $x \in X$ and $y \in Y$ we have
$$\langle (T - \lambda)x, y\rangle = \langle x, (T^* - \bar{\lambda})y\rangle.$$

Hence, $(T^* - \bar{\lambda})y = 0$ if and only if $y \perp \operatorname{ran}(T-\lambda)$.

(2) Just interchange $T-\lambda$ and $T^*-\bar{\lambda}$ in (1).

(3) This just follows from $(E^\perp)^\perp = \overline{E}$ for any linear subspace $E \subset X$. ☺

The closed graph theorem leads to a result about automatic boundedness for linear operators in a Hilbert space.

Lemma 6.11 (Hellinger-Toeplitz)**.** Let X and Y be Hilbert spaces, let T be a linear operator from X to Y, and let S be a linear operator from Y to X. Assume that

$$\langle Tx, y\rangle = \langle x, Sy\rangle, \quad x \in X, y \in Y.$$

Then $T \in B(X,Y)$, $S \in B(Y,X)$, and, in fact, $S = T^*$.

Proof. Assume that $x_n \to x$ in X and that $Tx_n \to u$ in Y. Then taking limits in

$$\langle Tx_n, y\rangle = \langle x_n, Sy\rangle, \quad y \in Y,$$

gives

$$\langle u, y\rangle = \langle x, Sy\rangle, \quad y \in X.$$

By assumption $\langle x, Sy\rangle = \langle Tx, y\rangle$, which leads to $\langle u, y\rangle = \langle Tx, y\rangle$ for all $y \in Y$. Hence $u = Tx$ and the operator T is closed and thus bounded. Similarly S is bounded. Therefore $T \in B(X,Y)$, $S \in B(Y,X)$, and hence $S = T^*$. ☺

Example 6.12. Recall that the *right shift* S_r and the *left shift* S_l are bounded linear operators in ℓ^2 defined by

$$S_r(x_1,x_2,\dots) = (0,x_1,x_2,\dots), \quad S_l(x_1,x_2,\dots) = (x_2,x_3,\dots), \quad x \in \ell^2.$$

Note that $(S_l)^* = S_r$ and $(S_r)^* = S_l$, which is easily seen from

$$\langle S_r x, y\rangle = \sum_{n=1}^{\infty} x_n \bar{y}_{n+1} = \langle x, S_l y\rangle.$$

By Lemma 6.9 it follows that the spectra of S_r and S_l are related by $\sigma(S_r) = \overline{\sigma(S_l)}$. In fact, S_r and S_l have the same spectrum; cf. Example 5.39. □

Example 6.13. Let the function $K : [a,b] \times [a,b] \to \mathbb{C}$ be square-integrable and let T be the Fredholm operator in $X = L^2(a,b)$ given by

$$Tf(x) = \int_a^b K(x,y) f(y)\, dy, \quad f \in X.$$

Then $T \in B(X)$ and it follows from Definition 6.7 that the adjoint $T^* \in B(X)$ is given by

$$T^* f(x) = \int_a^b \overline{K(y,x)} f(y)\, dy, \quad f \in X.$$

To see this consider (Tf, g) and apply the Fubini theorem. In the case that $K(x,y) = 0$ for $y > x$ we obtain the Volterra operator

$$Tf(x) = \int_a^x K(x,y) f(y)\, dy, \quad f \in X.$$

and its adjoint

$$T^* f(x) = \int_x^b \overline{K(y,x)} f(y)\, dy, \quad f \in X.$$

□

6.3 Special classes of operators

The interplay between an operator $T \in B(X,Y)$ and its adjoint $T^* \in B(Y,X)$ gives rise to many interesting observations.

Definition 6.14. Let X be a Hilbert space and let $T \in B(X)$. Then T is said to be *selfadjoint* if $T^* = T$.

The following characterization is important. Note that it makes essential use of the assumption that $\mathbb{K} = \mathbb{C}$. If $\mathbb{K} = \mathbb{R}$ the stated result is not true: a counter-example is given by a rotation over $\pi/2$ radians in $X = \mathbb{R}^2$.

Lemma 6.15. Let X be a Hilbert space and let $T \in B(X)$. Then T selfadjoint if and only if $\langle Tx, x\rangle \in \mathbb{R}$ for all $x \in X$.

Proof. It is clear that $T = T^*$ implies that $\langle Tx, x\rangle \in \mathbb{R}$ for all $x \in X$. Conversely, assume that $\langle Tx, x\rangle \in \mathbb{R}$ for all $x \in X$. Then

$$\langle Tx, x\rangle = \overline{\langle x, Tx\rangle} = \langle x, Tx\rangle, \quad x \in X.$$

By polarization it follows that

$$4\langle Tx, y\rangle = \sum_{k=0}^{3} \langle T(x+i^k y), x+i^k y\rangle = \sum_{k=0}^{3} \langle x+i^k y, T(x+i^k y)\rangle = 4\langle x, Ty\rangle,$$

which completes the proof. ☺

Definition 6.16. Let X be a Hilbert space. Then $T \in B(X)$ is said to be *nonnegative* if

$$\langle Tx, x\rangle \geq 0, \quad x \in X.$$

In this case we write $T \geq 0$.

Note that nonnegative operators in $B(X)$ are selfadjoint by Lemma 6.15 (still assuming that $\mathbb{K} = \mathbb{C}$). For nonnegative operators the following result shows how the norm $\|T\|$ can be evaluated in terms of inner products.

Lemma 6.17 (Cauchy-Schwarz type inequality for operators)**.** Let X be a Hilbert space and let $T \in B(X)$ be nonnegative. Then

$$\|Tx\|^2 \leq \|T\| \langle Tx, x\rangle, \quad x \in X, \tag{6.1}$$

and

$$\|T\| = \sup_{\|x\|=1} \langle Tx, x\rangle. \tag{6.2}$$

Proof. Recall the Cauchy-Schwarz inequality $|[x,y]|^2 \leq [x,x][y,y]$ for a semi-definite inner product $[\cdot,\cdot]$; cf. Remark 2.42. Since T is nonnegative it follows that T is selfadjoint. Therefore $[x,y] = \langle Tx, y\rangle$ defines a semi-definite inner product on X and, hence, one has

$$|\langle Tx, y\rangle|^2 \leq \langle Tx, x\rangle \langle Ty, y\rangle, \quad x, y \in X.$$

Choose $y = Tx$, then one obtains

$$\|Tx\|^4 \leq \langle Tx, x\rangle \langle TTx, Tx\rangle \leq \|T\| \langle Tx, x\rangle \|Tx\|^2, \quad x \in X,$$

since $|\langle TTx, Tx\rangle| \leq \|T^2 x\|\|Tx\| \leq \|T\|\|Tx\|^2$. This gives (6.1).

It is clear that $\langle Tx, x\rangle \leq \|T\|\|x\|^2$, so that $\sup_{\|x\|=1}\langle Tx, x\rangle \leq \|T\|$. It follows from (6.1) that

$$\|T\|^2 = \left(\sup_{\|x\|=1} \|Tx\|\right)^2 \leq \|T\| \sup_{\|x\|=1} \langle Tx, x\rangle.$$

This gives (6.2). ☺

Corollary 6.18. Let X be a Hilbert space and let $T \in B(X)$. If $\langle Tx, x\rangle = 0$ for all $x \in X$, then $T = 0$.

Proof. If $\langle Tx, x\rangle = 0$ for all $x \in X$, then certainly $\langle Tx, x\rangle \geq 0$ and Lemma 6.17 applies. Therefore $Tx = 0$ for all $x \in X$, which implies that $T = 0$. ☺

Definition 6.19. Let X be a Hilbert space and let T and S be selfadjoint operators in $B(X)$. Then the inequality $T \leq S$ is defined by

$$\langle Tx, x\rangle \leq \langle Sx, x\rangle \quad x \in X.$$

Corollary 6.20. Let X be a Hilbert space and let T and S be nonnegative operators in $B(X)$. Then the inequality $T \leq S$ implies that $\|T\| \leq \|S\|$.

Let X be a Hilbert space and let $T \in B(X)$. If $T \geq 0$, then by Lemma 6.15 it follows that T is selfadjoint. Hence also $T^2 \geq 0$ as $\langle T^2 x, x\rangle = \langle Tx, Tx\rangle \geq 0$. Furthermore a similar reasoning gives $T^3 \geq 0$ as $\langle T^3 x, x\rangle = \langle TTx, Tx\rangle \geq 0$. It is useful to observe the general case for $T \in B(X)$ and any $n \in \mathbb{N}$:

$$T \geq 0 \quad \Rightarrow \quad T^n \geq 0. \tag{6.3}$$

Let X and Y be Hilbert spaces and recall the following notions (which were originally introduced in a Banach space setting). An operator $T \in B(X,Y)$ is called *contractive* if $\|Tx\| \leq \|x\|$, $x \in X$; cf. Example 4.8, and an operator $T \in B(X,Y)$ is called *isometric* if $\|Tx\| = \|x\|$, $x \in X$. Moreover, an operator $T \in B(X,Y)$ is called *isometrically isomorphic* or *unitary* if T is isometric and T maps X onto Y. The structure of Hilbert spaces X and Y makes it possible to give an alternative description involving the operators T^*T or TT^* which are selfadjoint in X and in Y, respectively. In fact, for any $T \in B(X,Y)$ one has:

$$\langle T^*Tx, x\rangle = \|Tx\|^2, \quad x \in X, \quad \text{and} \quad \langle TT^*y, y\rangle = \|T^*y\|^2, \quad y \in Y, \tag{6.4}$$

and therefore $T^*T \geq 0$ in X and $TT^* \geq 0$ in Y.

Lemma 6.21. Let X and Y be Hilbert spaces and let $T \in B(X,Y)$. Then
1. T is contractive $\Leftrightarrow$ $T^*T \leq I_X$;
2. T is isometric $\Leftrightarrow$ $T^*T = I_X$;
3. T is unitary $\Leftrightarrow$ $T^*T = I_X$ and $TT^* = I_Y$.

Proof. (1) and (2) One has $\|Tx\|^2 \leq \|x\|^2$, i.e., $\langle Tx, Tx\rangle \leq \langle x, x\rangle$ or

$$\langle T^*Tx, x\rangle \leq \langle x, x\rangle, \quad x \in X.$$

The isometric case is completely similar.

(3) Assume that T is unitary. The isometry of T gives $T^*T = I_X$, which also leads to $TT^*T = T$. Hence $TT^*(Tx) = Tx$ for all $x \in X$ and, since $\operatorname{ran} T = Y$, it follows that $TT^* = I_Y$. The converse is trivial. ☺

Observe that T is unitary if and only if T is invertible and $T^{-1} = T^*$. An operator $T \in B(X,Y)$ is called *partially isometric* if T is isometric on $(\ker T)^\perp$. This means that $T^*T \restriction \overline{\operatorname{ran}} T^* = I_{\overline{\operatorname{ran}} T^*}$. Since $T^*T \in B(X)$ this equality is valid if and only if T^*T and $I_{\overline{\operatorname{ran}} T^*}$ agree on the dense subset $\operatorname{ran} T^*$. Thus T is partially isometric $\Leftrightarrow T^*TT^* = T^*$.

Definition 6.22. Let X be a Hilbert space and let $T \in B(X)$. Then T is called *normal* if $T^*T = TT^*$.

Lemma 6.23. Let X be a Hilbert space and let $T \in B(X)$. Then
1. T is normal $\Leftrightarrow \|Tx\| = \|T^*x\|$ for all $x \in X$;
2. T is selfadjoint $\Rightarrow T$ is normal;
3. T is unitary $\Rightarrow T$ is normal.

Proof. (1) A straightforward computation gives

$$\|Tx\|^2 - \|T^*x\|^2 = \langle (T^*T - TT^*)x, x\rangle.$$

so that ($\Rightarrow$) is clear. For ($\Leftarrow$) note that $\langle (T^*T - TT^*)x, x\rangle = 0$ for all $x \in X$ implies that $T^*T = TT^*$; cf. Corollary 6.18.

(2) and (3) are trivial. ☺

Note that if T is normal then $\ker T = \ker T^*$. If T is normal and $\lambda \in \mathbb{C}$ then also $T - \lambda$ is normal and hence

$$\ker (T - \lambda) = \ker (T^* - \bar{\lambda}), \quad \lambda \in \mathbb{C}.$$

Lemma 6.24. Let X be a Hilbert space and let $T \in B(X)$ be normal. Assume $Tx = \lambda x$ and $Ty = \mu y$ for some $x, y \in X$. If $\lambda \neq \mu$ then $\langle x, y\rangle = 0$.

Proof. As $\ker (T - \mu) = \ker (T^* - \bar{\mu})$ it follows that

$$\lambda \langle x, y\rangle = \langle \lambda x, y\rangle = \langle Tx, y\rangle = \langle x, T^*y\rangle = \langle x, \bar{\mu} y\rangle = \mu \langle x, y\rangle.$$

Thus if $\lambda \neq \mu$ then $\langle x, y\rangle = 0$. ☺

Lemma 6.25. Let X be a Hilbert space and let $T \in B(X)$ be normal. Then $\sigma(T)$ consists of approximate eigenvalues of T.

Proof. By Proposition 5.36 one has in general $\lambda \in \rho(T)$ if and only if
1. $\operatorname{ran}(T - \lambda)$ is dense in X or, equivalently, $\ker (T^* - \bar{\lambda}) = \{0\}$;
2. $\|(T - \lambda)x\| \geq \alpha \|x\|$ for some $c > 0$.

However since T is normal one has $\ker (T - \lambda) = \ker (T^* - \bar{\lambda})$. This means that if T is normal (1) is a consequence of (2). Hence $\lambda \in \rho(T)$ if and only if $\|(T - \lambda)x\| \geq c\|x\|$ for some $c > 0$. This is equivalent to $\lambda \in \sigma(T)$ if and only if λ is an approximate eigenvalue of T. ☺

In a Hilbert space there is an important role for orthogonal projections; cf. Definition 5.20 and Lemma 5.21. Here is a characterization in terms of adjoints.

Lemma 6.26. Let X be a Hilbert space and let $P \in B(X)$ be a projection. Then P is orthogonal if and only if P is selfadjoint. In this case P is nonnegative.

Proof. Recall that $P \in B(X)$ and $P^2 = P$ imply that $\ker P = \operatorname{ran}(I-P)$ and

$$X = \operatorname{ran} P + \ker P, \quad \text{direct sum},$$

cf. Lemma 5.16. Let the projection $P \in B(X)$ be orthogonal. Then

$$\begin{aligned} \langle Px, y \rangle &= \langle Px, Py + (I-P)y \rangle = \langle Px, Py \rangle \\ &= \langle Px + (I-P)x, Py \rangle = \langle x, Py \rangle, \quad x, y \in X, \end{aligned}$$

shows that P is selfadjoint. Conversely, let the projection $P \in B(X)$ be selfadjoint. Then

$$\langle Px, (I-P)y \rangle = \langle x, P(I-P)y \rangle = 0, \quad x, y \in X,$$

shows that $\operatorname{ran} P$ and $\ker P = \operatorname{ran}(I-P)$ are orthogonal.

To see the last statement, observe that $\langle Px, x \rangle = \langle P^2 x, x \rangle = \langle Px, Px \rangle = \|Px\|^2$ for all $x \in X$. Hence P is nonnegative. ☺

Definition 6.27. Let X be a Hilbert space and let $T \in B(X)$. Let $V \subset X$ be a closed linear subspace and let P be the orthogonal projection onto V. Then $V \subset X$ is said to be *invariant* under T if

$$TV \subset V \quad \text{or, equivalently,} \quad (I-P)TP = 0.$$

In this case the restriction of T to V, which belongs to $B(V)$, is denoted by $T \restriction V$.

Lemma 6.28. Let X be a Hilbert space and let $T \in B(X)$. Assume that $V \subset X$ is a closed linear subspace, then

$$TV \subset V \quad \Leftrightarrow \quad T^* V^\perp \subset V^\perp.$$

Proof. Let P be the orthogonal projection onto V. Use that $TV \subset V$ is equivalent to $(I-P)TP = 0$. Taking adjoints it follows that $PT^*(I-P) = 0$, so that $T^* V^\perp \subset V^\perp$. The converse is now clear. ☺

Corollary 6.29. Let X be a Hilbert space and let $T \in B(X)$ be selfadjoint. Let $V \subset X$ be a closed linear subspace with orthogonal projection P. Then

$$TV \subset V \quad \Leftrightarrow \quad TV^\perp \subset V^\perp.$$

In this case, $T \restriction V$ acts as a selfadjoint operator in $B(V)$ and $T \restriction V^\perp$ acts as a selfadjoint operator in $B(V^\perp)$, while the operator T has the orthogonal sum decomposition

$$T = T \restriction V + T \restriction V^\perp.$$

Proof. The first statement follows from Lemma 6.28. The restriction of T to V acts as a selfadjoint operator in $B(V)$, since $(Tx, y) = (x, Ty)$ when $x, y \in V$. Likewise, the restriction of T to $V^\perp$ acts as a selfadjoint operator in $B(V^\perp)$. It is clear that $T = T \restriction V + T \restriction V^\perp$ is an orthogonal sum. ☺

6.4 Selfadjoint operators in Hilbert spaces

Definition 6.30. Let X be a Hilbert space and let $T \in B(X)$ be selfadjoint. Then the numbers a and b are defined by

$$a = \inf_{\|x\|=1} \langle Tx, x \rangle \quad \text{and} \quad b = \sup_{\|x\|=1} \langle Tx, x \rangle.$$

Remark 6.31. The numbers a and b are also obtained via

$$a = \inf_{\|x\|\leq 1} \langle Tx, x \rangle \quad \text{and} \quad b = \sup_{\|x\|\leq 1} \langle Tx, x \rangle.$$

It suffices to show this for the endpoint b. First observe that

$$\sup_{\|x\|=1} \langle Tx, x \rangle \leq \sup_{\|x\|\leq 1} \langle Tx, x \rangle.$$

Next observe that for a nontrivial $x_0 \in X$ with $\|x_0\| \leq 1$ one has

$$\langle Tx_0, x_0 \rangle = \|x_0\|^2 \left\langle T\frac{x_0}{\|x_0\|}, \frac{x_0}{\|x_0\|} \right\rangle \leq \left\langle T\frac{x_0}{\|x_0\|}, \frac{x_0}{\|x_0\|} \right\rangle \leq \sup_{\|x\|=1} \langle Tx, x \rangle.$$

This inequality leads to

$$\sup_{\|x\|\leq 1} \langle Tx, x \rangle \leq \sup_{\|x\|=1} \langle Tx, x \rangle.$$

□

Theorem 6.32. Let X be a Hilbert space and let $T \in B(X)$ be selfadjoint. Then $\sigma(T)$ consists of approximate eigenvalues and $\sigma(T) \subset [a,b]$, while $a, b \in \sigma(T)$. Moreover, if $Tx = \lambda x$, $x \neq 0$, and $Ty = \mu y$, $y \neq 0$, then $\lambda \neq \mu$ implies $\langle x, y \rangle = 0$.

Proof. Let $\lambda \in \sigma(T)$. Since T is selfadjoint λ belongs to the approximate point spectrum of T; cf. Lemma 6.25. Hence there exists a sequence (x_n) in X with $\|x_n\| = 1$ and $(T-\lambda)x_n \to 0$, which leads to

$$|\langle Tx_n, x_n \rangle - \lambda \langle x_n, x_n \rangle| = |\langle (T-\lambda)x_n, x_n \rangle| \leq \|(T-\lambda)x_n\|.$$

In particular one has

$$\|x_n\| = 1 \quad \text{and} \quad \langle Tx_n, x_n \rangle \to \lambda.$$

Now each $\langle Tx_n, x_n \rangle \in [a,b]$, so that $\lambda \in \mathbb{R}$ and, in fact, $a \leq \lambda \leq b$.

Note that $b - T \geq 0$. Hence one has by Lemma 6.17

$$\|(b-T)x\|^2 \leq \|b-T\| \langle (b-T)x, x \rangle, \quad x \in X.$$

As $b = \sup_{\|x\|=1} \langle Tx, x \rangle$ there is a sequence (x_n) with $\|x_n\| = 1$ and $\langle Tx_n, x_n \rangle \to b$. Therefore $(b-T)x_n \to 0$, which implies $b \in \sigma(T)$. Similarly one can show that $a \in \sigma(T)$.

The last statement is a direct consequence of Lemma 6.24. ☺

Corollary 6.33. Let X be a Hilbert space and let $T \in B(X)$. Then the following statements are equivalent:

1. T or, equivalently, T^* is invertible;
2. T^*T and TT^* are invertible;
3. there exists $\varepsilon > 0$ for which $T^*T \geq \varepsilon$ and $TT^* \geq \varepsilon$;

4. there exists $\eta > 0$ for which

$$\|Tx\| \geq \eta \|x\| \quad \text{and} \quad \|T^*x\| \geq \eta \|x\|, \quad x \in X.$$

Proof. (1) $\Rightarrow$ (2) Assume that T or, equivalently, T^* is invertible; cf. Lemma 6.9. Then also T^*T and TT^* are invertible; cf. Corollary 4.24.

(2) $\Rightarrow$ (1) Assume that T^*T and TT^* are invertible. Then

$$(T^*T)^{-1}T^*T = I, \;\; TT^*(TT^*)^{-1} = I,$$

and hence T has a left inverse $(T^*T)^{-1}T^*$ and a right inverse $T^*(TT^*)^{-1}$. Thus T is invertible.

(2) $\Leftrightarrow$ (3) Observe that T^*T and TT^* are nonnegative. Hence T^*T and TT^* are invertible if and only if the lower bounds of T^*T and TT^* are positive; see Theorem 6.32.

(3) $\Leftrightarrow$ (4) This is trivial. ☺

Corollary 6.34. Let X be a Hilbert space and let $T \in B(X)$ be selfadjoint.
1. If $a = \min_{\|x\|=1} \langle Tx, x\rangle$, then $a \in \sigma_{\mathrm{p}}(T)$.
2. If $b = \max_{\|x\|=1} \langle Tx, x\rangle$, then $b \in \sigma_{\mathrm{p}}(T)$.

Proof. It suffices to show (2) as the proof of (1) is completely similar. Due to the assumption there exists $x_0 \in X$ for which $\|x_0\| = 1$ and

$$b = \langle Tx_0, x_0\rangle \quad \text{or, equivalently,} \quad \langle (b-T)x_0, x_0\rangle = 0.$$

An application of Lemma 6.17 leads to the desired result. ☺

Theorem 6.35. Let X be a Hilbert space and let $T \in B(X)$ be selfadjoint. Then

$$\|T\| = \max\{|a|, |b|\},$$

or, equivalently,

$$\|T\| = \sup_{\|x\|=1} |\langle Tx, x\rangle|.$$

Proof. Denote $M = \sup\{|\langle Tx, x\rangle| : \|x\| = 1\}$. Then it is clear that

$$M \leq \|T\|.$$

For the converse inequality use that T is selfadjoint and observe that therefore

$$\begin{aligned} 4\mathrm{Re}\,(Tx, y) &= 2[\langle Tx, y\rangle + \overline{\langle Tx, y\rangle}] \\ &= 2[\langle Tx, y\rangle + \langle Ty, x\rangle] \\ &= \langle T(x+y), x+y\rangle - \langle T(x-y), x-y\rangle. \end{aligned}$$

Note that $|\langle T(x \pm y, x \pm y\rangle| \leq M\|x \pm y\|^2$, and hence

$$\begin{aligned} 4\mathrm{Re}\,\langle Tx, y\rangle &\leq |\langle T(x+y), x+y\rangle| + |\langle T(x-y), x-y\rangle| \\ &\leq M(\|x+y\|^2 + \|x-y\|^2) \\ &= 2M(\|x\|^2 + \|y\|^2), \end{aligned}$$

where in the last step the parallelogram law has been used. Now in this inequality take

$$y = \frac{\|x\|}{\|Tx\|} Tx, \quad \text{so that} \quad \langle Tx, y\rangle = \|Tx\|\,\|x\|, \quad \|y\| = \|x\|,$$

which leads to

$$4\|Tx\|\,\|x\| \leq 4M\|x\|^2, \quad x \in X,$$

or $\|T\| \leq M$. Hence the equality $\|T\| = M$ has been shown. ☺

A alternative way to state the above result is that for a selfadjoint operator $T \in B(X)$ one has $r_\sigma(T) = \|T\|$; see the definition in (5.10).

6.5 Monotonicity and square roots

Theorem 6.36. Let X be a Hilbert space and let $T_n \in B(X)$ be a sequence of operators satisfying:

$$0 \leq T_m \leq T_n \leq S, \quad n \geq m, \tag{6.5}$$

for some nonnegative selfadjoint operator $S \in B(X)$. Then there exists $T \in B(X)$ such that $0 \leq T_n \leq T \leq S$ and

$$T_n x \to Tx, \quad x \in X.$$

Proof. Note that by (6.5) and Lemma 6.17 one sees $\|T_n\| \leq \|S\|$, i.e., the sequence (T_n) is uniformly bounded. It also follows from Lemma 6.17 that for $n \geq m$ and $x \in X$

$$\|(T_n - T_m)x\|^2 \leq \|T_n - T_m\|\langle (T_n - T_m)x, x\rangle \leq 2\|S\|\langle (T_n - T_m)x, x\rangle. \tag{6.6}$$

Now in $\mathbb{R}$ the sequence $(\langle T_n x, x\rangle)$ is monotone and bounded above and, hence, converges in $\mathbb{R}$, so that it is Cauchy. Therefore by (6.6) the sequence $(T_n x)$ is Cauchy in the Hilbert space X, so there is a limit, say, Tx, and $T_n x \to Tx$, $x \in X$.

It is clear that T is linear. It follows from $T_n x \to Tx$, $x \in X$, that also $\|T_n x\| \to \|Tx\|$. Since

$$\|T_n x\| \leq \|T_n\|\|x\| \leq \|S\|\|x\|,$$

one obtains $\|Tx\| \leq \|S\|\|x\|$. Hence $T \in B(X)$ and $\|T\| \leq \|S\|$.

The inequalities $0 \leq T_m \leq T \leq S$ follow by taking $n \to \infty$ in (6.5). ☺

Corollary 6.37. Let X be a Hilbert space and let $P_n \in B(X)$ be a sequence of orthogonal projections satisfying:

$$0 \leq P_m \leq P_n, \quad n \geq m. \tag{6.7}$$

Then there exists an orthogonal projection $P \in B(X)$ such that $0 \leq P_n \leq P \leq I$ and

$$P_n x \to Px, \quad x \in X.$$

Theorem 6.38 (Existence and uniqueness of square root)**.** Let X be a Hilbert space and let $T \in B(X)$ be a nonnegative operator. Then there exists a unique nonnegative operator $U \in B(X)$, denoted by $U = T^{\frac{1}{2}}$, such that $U^2 = T$. This operator commutes with every bounded linear operator which commutes with T. Moreover, $\|T^{\frac{1}{2}}\| = \sqrt{\|T\|}$.

Proof. The proof of the existence and uniqueness of the root of T will be given in two steps under the assumption that T is a contraction. In a third step the general case will be considered.

Existence for contractions. Assume that $0 \leq T \leq I$ and let $S = I - T$, so that $S \in B(X)$ with $0 \leq S \leq I$. Then an operator $A \in B(X)$ with $0 \leq A \leq I$ satisfies $A^2 = T$ if and only if $D \in B(X)$ with $0 \leq D \leq I$ satisfies

$$D = \frac{1}{2}(S + D^2), \tag{6.8}$$

where A and D are connected via $D = I - A$. In order to solve (6.8), define

$$D_0 = 0 \quad \text{and} \quad D_{n+1} = \frac{1}{2}(S + D_n^2), \quad n \in \mathbb{N} \cup \{0\}. \tag{6.9}$$

Therefore one sees that $D_n \in B(X)$ and that all operators D_n are nonnegative. It follows from (6.9) that

$$I - D_{n+1} = \frac{1}{2}(I - S) + \frac{1}{2}(I - D_n^2), \quad n \in \mathbb{N} \cup \{0\},$$

so that inductively one also sees that $I - D_{n+1} \geq 0$ or $D_{n+1} \leq I$.

Now the monotonicity of the operators D_n will be shown. For this purpose it is helpful to know that in fact $D_n = q_n(S)$, where the polynomials q_n are defined inductively by

$$q_0(z) = 0 \quad \text{and} \quad q_{n+1}(z) = \frac{1}{2}(z + q_n(z)^2), \quad n \in \mathbb{N} \cup \{0\}, \tag{6.10}$$

cf. (6.9). It is clear that all coefficients of the polynomials q_n are nonnegative. It is a consequence of (6.10) that the polynomials $q_{n+1} - q_n$ satisfy

$$\begin{cases} q_1(z) - q_0(z) = \frac{1}{2}z, \\ q_{n+1}(z) - q_n(z) = \frac{1}{2}(q_n(z) + q_{n-1}(z))(q_n(z) - q_{n-1}(z)), \quad n \in \mathbb{N}. \end{cases} \tag{6.11}$$

Since the polynomials $q_n + q_{n-1}$ have all nonnegative coefficients one sees from (6.11) by induction that the polynomials $q_{n+1} - q_n$ have all nonnegative coefficients. Recall that $S \geq 0$ implies that $S^n \geq 0$, $n \in \mathbb{N}$; cf. (6.3). Thus it follows that

$$D_{n+1} - D_n = q_{n+1}(S) - q_n(S) = (q_{n+1} - q_n)(S) \geq 0, \quad n \in \mathbb{N} \cup \{0\}.$$

Therefore one has

$$0 \leq D_n \leq D_{n+1} \leq I.$$

Hence by Theorem 6.36 there exists a unique nonnegative $D \in B(X)$ with

$$0 \leq D_n \leq D \leq I \quad \text{and} \quad D_n x \to Dx, \quad x \in X.$$

Observe that

$$\begin{aligned} \langle Dx, x \rangle &= \lim_{n\to\infty} \langle D_{n+1}x, x \rangle = \lim_{n\to\infty} \frac{1}{2} \langle (S + D_n^2)x, x \rangle \\ &= \frac{1}{2}\left[\langle Sx, x \rangle + \lim_{n\to\infty} \|D_n x\|^2 \right] = \frac{1}{2} \langle (S + D^2)x, x \rangle, \quad x \in X. \end{aligned}$$

Polarization now gives (6.8), which proves the existence.

Now assume that $B \in B(X)$ commutes with T. Since $S = I - T$, $D = I - A$, and $D_n = q_n(S)$, one sees that

$$\begin{aligned} BT = TB \quad &\Rightarrow \quad BS = SB \quad \Rightarrow \quad BD_n = D_n B \\ &\underset{n\to\infty}{\Rightarrow} \quad BD = DB \quad \Rightarrow \quad BA = AB. \end{aligned}$$

Hence B also commutes with A.

Uniqueness for contractions. Still assuming that T is a nonnegative contraction, let $A \in B(X)$ be the operator just constructed, so that $A^2 = T$. Let $B \in B(X)$ be any nonnegative operator with $B^2 = T$. It follows from

$$BT = BB^2 = B^2B = TB,$$

that B commutes with T. Then, by the previous step, B also commutes with A. Now let $N = \ker T$, then it is not difficult to see that

$$N = \ker A = \ker B = \ker (A+B).$$

To verify this, use Lemma 6.17 and $(Tx,x) = \|Ax\|^2 = \|Bx\|^2$; for the last identity observe that $(A+B)x = 0$ implies $\langle Ax,x\rangle + \langle Bx,x\rangle = 0$ and so $\langle Ax,x\rangle = 0$ and $\langle Bx,x\rangle = 0$. In particular, A and B coincide on N, and it suffices to show that A and B coincide on $N^\perp$.

Note that the selfadjoint operator $A+B$ maps $N^\perp$ into $N^\perp$ and that $(A+B)N^\perp$ is dense in $N^\perp$; just remember that a selfadjoint $S \in B(X)$ maps $(\ker S)^\perp = \overline{\operatorname{ran}} S$ into itself; cf. Lemma 6.10. It follows from $A^2 = T = B^2$ and $AB = BA$ that

$$(A-B)(A+B) = A^2 + AB - BA - B^2 = 0,$$

which implies that A and B coincide on the dense set $(A+B)N^\perp$, and therefore on all of $N^\perp$, since $A, B \in B(X)$. One concludes that $A = B$.

Existence and uniqueness for nonnegative operators. Let $T \in B(X)$ be a nonnegative operator. Assume that T is not trivial. Then $Q = T/\|T\|$ is a nonnegative contraction. Let R be the unique nonnegative square root of Q, as constructed in the previous steps, so that

$$(\|T\|^{\frac{1}{2}}R)^2 = \|T\|Q = T.$$

Hence the operator $\|T\|^{\frac{1}{2}}R$ is a nonnegative root of T. Now let U be any nonnegative operator which is a square root: $U^2 = T$. Then clearly

$$\left(\frac{U}{\sqrt{\|T\|}}\right)^2 = \frac{T}{\|T\|} = Q$$

so that

$$\frac{U}{\sqrt{\|T\|}} = R \quad \text{or} \quad U = \sqrt{\|T\|}R.$$

Thus $\|T\|^{\frac{1}{2}}R$ is the unique nonnegative root of T.

Finally let C be bounded linear operator which commutes with T. Then clearly C commutes with $Q = T/\|T\|$. Hence by the above steps C also commutes with the unique nonnegative square root R of Q. But then C also commutes with unique nonnegative square root $\|T\|^{\frac{1}{2}}R$ of T.

Finally note that for the nonnegative operator $U \in B(X)$, such that $U^2 = T$ one has $\|U^2\| = \|T\|$. But U is nonnegative and hence selfadjoint so that it follows from Lemma 6.8 that $\|U^2\| = \|U^*U\| = \|U\|^2$. This leads to $\|U\| = \sqrt{\|T\|}$. ☺

Let $S, T \in B(X)$ be selfadjoint. Then, cleary, the product ST is selfadjoint if and only if S and T commute: $ST = TS$.

Corollary 6.39. Let X be a Hilbert space and let $S, T \in B(X)$ be nonnegative operators with $ST = TS$. Then
1. $ST = TS$ is nonnegative;
2. $S^{\frac{1}{2}}T^{\frac{1}{2}} = T^{\frac{1}{2}}S^{\frac{1}{2}}$;

3. the square root of $ST = TS$ is given by $S^{\frac{1}{2}}T^{\frac{1}{2}} = T^{\frac{1}{2}}S^{\frac{1}{2}}$.

Proof. (1) Since $S, T \in B(X)$ are nonnegative and $ST = TS$, it follows that $S^{\frac{1}{2}}T = TS^{\frac{1}{2}}$. Therefore

$$\langle STx, x\rangle = \langle S^{\frac{1}{2}}S^{\frac{1}{2}}Tx, x\rangle = \langle S^{\frac{1}{2}}Tx, S^{\frac{1}{2}}x\rangle = \langle TS^{\frac{1}{2}}x, S^{\frac{1}{2}}x\rangle \geq 0, \quad x \in X.$$

Thus $ST = TS$ implies that the product is nonnegative.

(2) Since $S, T \in B(X)$ are nonnegative, it follows as in the proof of (1) with one more step, that

$$ST = TS \quad \Rightarrow \quad S^{\frac{1}{2}}T = TS^{\frac{1}{2}} \quad \Rightarrow \quad S^{\frac{1}{2}}T^{\frac{1}{2}} = T^{\frac{1}{2}}S^{\frac{1}{2}}.$$

(3) The square roots $S^{\frac{1}{2}}$ and $T^{\frac{1}{2}}$ are nonnegative and $S^{\frac{1}{2}}T^{\frac{1}{2}} = T^{\frac{1}{2}}S^{\frac{1}{2}}$. Hence also $S^{\frac{1}{2}}T^{\frac{1}{2}}$ is nonnegative and thus selfadjoint. The identity

$$ST = S^{\frac{1}{2}}S^{\frac{1}{2}}T^{\frac{1}{2}}T^{\frac{1}{2}} = S^{\frac{1}{2}}T^{\frac{1}{2}}S^{\frac{1}{2}}T^{\frac{1}{2}}$$

shows that $(ST)^{\frac{1}{2}} = S^{\frac{1}{2}}T^{\frac{1}{2}}$. ☺

Definition 6.40. Let X and Y be Hilbert spaces and let $T \in B(X,Y)$. Define the *absolute value* $|T| \in B(X)$ by

$$|T| = (T^*T)^{\frac{1}{2}},$$

where the right-hand side stands for the unique nonnegative square root of $T^*T \geq 0$.

Remark 6.41. Let X and Y be Hilbert spaces and let $T \in B(X,Y)$. Then

$$\ker T = \ker T^*T = \ker |T|.$$

To see this note that the first identity follows from (6.4). However then also $\ker |T| = \ker |T|^2$, which is equal to $\ker T^*T$. By taking orthogonal complements one has, equivalently,

$$\overline{\operatorname{ran}}\, T^* = \overline{\operatorname{ran}}\, T^*T = \overline{\operatorname{ran}}\, |T|.$$

□

Theorem 6.42 (Polar decomposition)**.** Let X and Y be Hilbert spaces and let $T \in B(X,Y)$. Then there exists a partial isometry $U \in B(X,Y)$ such that $T = U|T|$. The partial isometry is unique if $\ker U = \ker T$.

Proof. Note that $|T|$ is selfadjoint and that

$$\| \, |T|x\|^2 = \langle |T|x, |T|x\rangle = \langle |T|^2x, x\rangle = \langle T^*Tx, x\rangle = \|Tx\|^2, \quad x \in X.$$

Define the operator U initially from $\operatorname{ran}|T|$ to $\operatorname{ran} T$ by

$$|T|x \mapsto Tx.$$

Note that U is a well defined isometry. It extends by continuity to an isometry, again denoted by U, from $\overline{\operatorname{ran}}\,|T|$ to $\overline{\operatorname{ran}}\,T$. Furthermore define U on $(\overline{\operatorname{ran}}\,|T|)^\perp = \ker |T| = \ker T$ by $U = 0$. Finally observe that $U|T|x = Tx$, $x \in X$, shows that $T = U|T|$. ☺

6.6 Spectral theory for compact selfadjoint operators

Proposition 6.43. Let X be a Hilbert space and let $T \in K(X)$ be selfadjoint. Then either $\|T\|$ or $-\|T\|$ is an eigenvalue of T.

Proof. It suffices to consider the case that T is non-trivial. It has been shown that $a, b \in \sigma(T)$ and that $\|T\| = \max(|a|, |b|)$. Assume for the sake of simplicity that $b = \|T\|$. There is a sequence (x_n) with $\|x_n\| = 1$ and $(b-T)x_n \to 0$. Note that

$$bx_n = (b-T)x_n + Tx_n.$$

Since T is compact there is a subsequence (x_{n_k}) of (x_n) such that Tx_{n_k} converges to, say, $y \in X$. Hence $x_{n_k} \to y/b$ and since $T \in B(X)$ one sees that $Tx_{n_k} \to Ty/b$. It follows that $y = Ty/b$, i.e., $Ty = by$. Since $\|y\| = b \neq 0$, one concludes that b is an eigenvalue of T.

The proof in the other cases is similar. ☺

Theorem 6.44. Let X be a Hilbert space and let $T \in K(X)$ be selfadjoint. Then there exists an orthonormal system of eigenvectors (e_i) with corresponding real eigenvalues (λ_i) such that for all $x \in X$:

$$Tx = \sum_i \lambda_i \langle x, e_i \rangle e_i.$$

If (λ_i) is an infinite sequence then $\lambda_i \to 0$.

Proof. Denote the operator T by T_1. Assume that T_1 is non-trivial. For simplicity of exposition assume that T_1 is nonnegative, so that $b = \|T_1\|$. By Proposition 6.43 $\lambda_1 = \|T_1\|$ is a nonzero eigenvalue of T and its multiplicity m_1 is finite; cf. Theorem 5.44. Choose a finite orthonormal basis (e_i^1) in the eigenspace $X_1 = \ker\,(T - \lambda_1)$ and observe that X_1 is invariant under T_1. Hence its orthogonal complement $X_1^\perp$ is invariant under T_1 and note that

$$x - \sum_{i=1}^{m_1} \langle x, e_i^1 \rangle e_i^1 \in X_1^\perp.$$

Let T_2 be the restriction of T to $X_1^\perp$. Assume that T_2 is not trivial. Then T_2 is a compact nonnegative selfadjoint operator in $X_1^\perp$. Therefore $\lambda_2 = \|T_2\|$ is an eigenvalue of T_2 and its multiplicity m_2 is finite, while $0 < \lambda_2 \leq \lambda_1$. Choose a finite orthonormal basis (e_i^2) in the eigenspace $X_2 = \ker\,(T_2 - \lambda_2)$ and observe that X_2 is invariant under T_2. Hence the orthogonal complement $(X_1 \oplus X_2)^\perp$ is invariant under T and note that

$$x - \left(\sum_{i=1}^{m_1} \langle x, e_i^1 \rangle e_i^1 + \sum_{i=1}^{m_2} \langle x, e_i^2 \rangle e_i^1 \right) \in (X_1 \oplus X_2)^\perp.$$

This leads to a sequence of eigenspaces $X_1, \ldots, X_{n-1}$ corresponding to eigenvalues $\lambda_1 \geq \cdots \geq \lambda_{n-1} > 0$ with multiplicities $m_1, \ldots, m_{n-1}$ such that

$$x - \sum_{j=1}^{n-1} \sum_{i=1}^{m_j} \langle x, e_i^j \rangle e_i^j \in (X_1 \oplus \cdots \oplus X_{n-1})^\perp.$$

This can be continued inductively as long as T_n as the restriction of T to

$$(X_1 \oplus \cdots \oplus X_{n-1})^\perp$$

is not trivial.

If for some $n \in \mathbb{N}$ the operator T_n is trivial, then

$$Tx = \sum_{j=1}^{n-1} \sum_{i=1}^{m_j} \lambda_j \langle x, e_i^j \rangle e_i^j,$$

which is a finite sum expansion.

If for all $n \in \mathbb{N}$ the operator T_n is nontrivial, then with

$$y_n = x - \sum_{j=1}^{n-1} \sum_{i=1}^{m_j} \langle x, e_i^j \rangle e_i^j \in (X_1 \oplus \cdots \oplus X_{n-1})^{\perp},$$

one obtains for all $n \in \mathbb{N}$

$$x = \sum_{j=1}^{n-1} \sum_{i=1}^{m_j} \langle x, e_i^j \rangle e_i^j + y_n.$$

By the orthogonality of the decompositions one sees that

$$\|x\|^2 = \sum_{j=1}^{n-1} \sum_{i=1}^{m_j} |\langle x, e_i^j \rangle|^2 + \|y_n\|^2,$$

and it follows that $\|y_n\| \leq \|x\|$. Observe that by the construction

$$\|Ty_n\| = \|T_n y_n\| \leq \|T_n\| \|y_n\| \leq |\lambda_n| \|x\|.$$

Hence it is clear that

$$\left\| Tx - \sum_{j=1}^{n-1} \sum_{i=1}^{m_j} \lambda_j \langle x, e_i^j \rangle e_i^j \right\| \leq \|Ty_n\| \leq |\lambda_n| \|x\|.$$

By Theorem 5.44 one has $\lambda_n \to 0$, which completes the proof for this case.

If T is not nonnegative, then a similar argument involving some bookkeeping leads to the expansion result. ☺

Corollary 6.45. Let X be a Hilbert space and let $T \in K(X)$ be selfadjoint. Then the orthonormal system of eigenvectors (e_i) forms an orthonormal basis for the Hilbert space $\overline{\operatorname{ran}}\, T$.

Proof. This is immediate from Theorem 3.46. ☺

Let $T \in K(X)$ be selfadjoint. Then there is the orthogonal decomposition $X = \ker T \oplus \overline{\operatorname{ran}}\, T$. Clearly $\ker T$ is invariant under T, hence also $\overline{\operatorname{ran}}\, T$ is invariant under T. Every element $Y \in \overline{\operatorname{ran}}\, T$ has the expansion

$$y = \sum_i \langle y, e_i \rangle e_i, \quad y \in X.$$

It follows from this expansion that the restriction of T to $\overline{\operatorname{ran}}\, T$ has a point spectrum which consists entirely of the nonzero eigenvalues. If there are infinitely many eigenvalues then they accumulate at 0, and 0 belongs to $\sigma(T)$ as an approximate eigenvalue.

Example 6.46. Let $X = L^2(0,1)$ and let $T \in B(X)$ be the Volterra operator defined by

$$Tf(x) = \int_0^x f(t)\, dt, \quad f \in X,$$

cf. Example 4.15 and Example 6.13. Then $T^* \in B(X)$ has the representation

$$T^* f(x) = \int_x^1 f(t)\, dt, \quad f \in X,$$

see Example 6.13. The nonnegative selfadjoint operator T^*T belongs to $K(X)$, since $T \in K(X)$. A short calculation shows that T^*T is given by

$$T^*Tf(x) = \int_0^x (1-x)f(t)\,dt + \int_x^1 (1-t)f(t)\,dt, \quad f \in X.$$

In order to find the eigenvalues of T^*T one has to solve $T^*Tf = \lambda f$ with $f \in X$, i.e.,

$$\int_0^x (1-x)f(t)\,dt + \int_x^1 (1-t)f(t)\,dt = \lambda f(x), \quad x \in [0,1], \quad f \in X.$$

Assume that $\lambda > 0$. Since f is integrable, the left-hand side defines a continuously differentiable function. Hence it follows that f is continuously differentiable. Therefore differentiation gives

$$-\int_0^x f(t)\,dt = \lambda f'(x), \quad x \in [0,1].$$

A similar reasoning leads to f' being continuously differentiable and

$$-f(x) = \lambda f''(x), \quad x \in [0,1].$$

Since $\lambda > 0$ one reads off from the equations that $f(1) = 0$ and $f'(0) = 0$. The general solution f of the differential equation $-f = \lambda f''$ is given by

$$f(x) = A\cos\frac{x}{\sqrt{\lambda}} + B\sin\frac{x}{\sqrt{\lambda}}, \quad x \in [0,1], \quad \lambda > 0,$$

so that the derivative f' is given by

$$f'(x) = -\frac{A}{\sqrt{\lambda}}\sin\frac{x}{\sqrt{\lambda}} + \frac{B}{\sqrt{\lambda}}\cos\frac{x}{\sqrt{\lambda}}, \quad x \in [0,1], \quad \lambda > 0.$$

It follows from the boundary condition $f'(0) = 0$ that $B = 0$. Hence one sees that $f(x) = A\cos\frac{x}{\sqrt{\lambda}}$. The boundary condition $f(1) = 0$ then leads to $\cos\frac{1}{\sqrt{\lambda}} = 0$, so that

$$\frac{1}{\sqrt{\lambda}} = \frac{\pi}{2} + k\pi \quad \text{or} \quad \lambda = \frac{4}{(2k+1)^2\pi^2}, \quad k \in \mathbb{N} \cup \{0\}.$$

Therefore the largest eigenvalue of T^*T is given by $4/\pi^2$, so that $\|T^*T\| = 4/\pi^2$. One concludes that $\|T\| = 2/\pi$; cf. Example 4.15.

Note that if $\lambda = 0$ it follows that $\int_0^x f(t)\,dt = 0$, $x \in [0,1]$. Since f is integrable, this implies that $\lambda = 0$ is not an eigenvalue of T^*T; therefore $\lambda = 0$ is an approximate eigenvalue of T^*T. □

Proposition 6.47 (Minimum-maximum property)**.** Let X be a Hilbert space and let $T \in K(X)$ be selfadjoint and nonnegative. Let $\lambda_1 \geq \lambda_2 \geq \cdots \geq 0$ be the eigenvalues of T. Then

$$\min_{\{V \subset X, \dim V = n-1\}} \left\{ \max_{x \perp V} \frac{\langle Tx, x\rangle}{\langle x, x\rangle} \right\} = \lambda_n. \tag{6.12}$$

Proof. Let $\lambda_1 \geq \lambda_2 \geq \cdots \geq 0$ be the eigenvalues of T, and let $e_1, e_2, \ldots$ be the corresponding orthonormal eigenvectors. Recall that by successively splitting off the (orthogonal) eigenspaces corresponding to $\lambda_1, \ldots, \lambda_{n-1}$ one obtains

$$\max\left\{ \frac{\langle Tx, x\rangle}{\langle x, x\rangle} : x \perp \operatorname{span}\{e_1, \ldots, e_{n-1}\} \right\} = \lambda_n. \tag{6.13}$$

Let $V \subset X$ be any linear subspace with $\dim V = n-1$. It will be shown that

$$\max\left\{\frac{\langle Tx,x\rangle}{\langle x,x\rangle} : x \perp V\right\} \geq \lambda_n. \tag{6.14}$$

Then the choice of the linear subspace $V = \operatorname{span}\{e_1,\dots,e_{n-1}\}$ in the left-hand side of (6.14) actually produces the right-hand side; cf. (6.13). Therefore the statement in (6.12) follows.

In order to show (6.14) let $V \subset X$ be a linear subspace with $\dim V = n-1$. Note that $\operatorname{span}\{e_1,\dots,e_n\}$ is n-dimensional, hence

$$V^{\perp} \cap \operatorname{span}\{e_1,\dots,e_n\} \neq \{0\}, \tag{6.15}$$

cf. Corollary 2.53. In general any vector of the form $x = \sum_{k=1}^{n}\langle x,e_k\rangle e_k$ leads to the inequality

$$\langle Tx,x\rangle = \sum_{k=1}^{n}\lambda_k|\langle x,e_k\rangle|^2 \geq \lambda_n\langle x,x\rangle.$$

In particular the last inequality holds for some nontrivial $x \in \operatorname{span}\{e_1,\dots,e_n\} \cap V^{\perp}$, see (6.15), which shows that

$$\sup_{x\perp V}\frac{\langle Tx,x\rangle}{\langle x,x\rangle} \geq \lambda_n. \tag{6.16}$$

Let P be the orthogonal projection from X onto $V^{\perp}$. Then it is clear that PTP is selfadjoint and compact in $V^{\perp}$. Hence one has in fact

$$\sup_{x\perp V}\frac{\langle Tx,x\rangle}{\langle x,x\rangle} = \sup_{x\perp V}\frac{\langle PTPx,x\rangle}{\langle x,x\rangle} = \max_{x\perp V}\frac{\langle PTPx,x\rangle}{\langle x,x\rangle} = \max_{x\perp V}\frac{\langle Tx,x\rangle}{\langle x,x\rangle},$$

which, together with (6.16), gives (6.14). ☺

The minimum-maximum property has a bonus: a characterization of the so-called singular numbers of any $T \in K(X,Y)$.

Definition 6.48. Let X and Y be Hilbert spaces and let $T \in K(X,Y)$. For any $n \in \mathbb{N}$ the *singular value* $s_n(T)$ is defined by

$$s_n(T) = \inf\{\|T-R\| : R \in B(X,Y),\ \operatorname{rank} R \leq n-1\}.$$

Thus $s_n(T)$ is the distance in the uniform norm of $T \in K(X,Y)$ to the operators of rank at most $n-1$ in $B(X,Y)$. It is clear that

$$s_1(T) \geq s_2(T) \geq \cdots \geq 0.$$

There is a simple geometric characterization of the singular values.

Theorem 6.49. Let X and Y be Hilbert spaces and let $T \in K(X,Y)$. Then

$$s_n(T) = \inf\{\,\|T\restriction W\| : W \subset X \text{ closed linear subspace, } \operatorname{codim} W \leq n-1\,\}.$$

Proof. Let $W \subset X$ be a closed linear subspace of X with $\operatorname{codim} W \leq n-1$. Then $X = W \oplus W^{\perp}$ with $\dim W^{\perp} \leq n-1$. Define $R \in B(X,Y)$ by

$$Rx = \begin{cases} Tx, & x \in W^{\perp},\\ 0, & x \in W,\end{cases}$$

so that $\operatorname{rank} R \le n-1$. Clearly $T - R = 0$ on $W^\perp$, so that

$$\|T\upharpoonright W\| = \|(T-R)\upharpoonright W\| = \|T-R\| \ge s_n(T),$$

where the last inequality is due to Definition 6.48. Hence

$$\inf\{\|T\upharpoonright W\| : W \subset X \text{ closed linear subspace, } \operatorname{codim} W \le n-1\} \ge s_n(T).$$

For the reverse inequality choose $\varepsilon > 0$. Then there is an operator $R \in B(X,Y)$ with $\operatorname{rank} R \le n-1$ with

$$\|T-R\| < s_n(T) + \varepsilon.$$

Observe that with $R \in B(X,Y)$ one has $X = \ker R \oplus (\ker R)^\perp$ and that R is bijective between $(\ker R)^\perp$ and $\operatorname{ran} R$, so that

$$\dim (\ker R)^\perp = \dim \operatorname{ran} R \le n-1.$$

Furthermore note that

$$\|T\upharpoonright \ker R\| = \|(T-R)\upharpoonright \ker R\| \le \|T-R\| < s_n(T) + \varepsilon.$$

Since $\ker R$ is a closed linear subspace of X with $\operatorname{codim} \ker R \le n-1$ it follows that

$$\inf\{\|T\upharpoonright W\| : W \subset X \text{ closed linear subspace, } \operatorname{codim} W \le n-1\} \le s_n(T).$$

This completes the proof of the identity. ☺

There is an important connection between the singular values of a, not necessarily selfadjoint, operator $T \in K(X,Y)$ and the eigenvalues of the nonnegative selfadjoint operator $T^*T \in B(X)$ via the minimum-maximum property.

Corollary 6.50. Let X and Y be Hilbert spaces and let $T \in K(X,Y)$. Then the singular value $s_n(T)$ is the square root of the n-th eigenvalue of T^*T.

Proof. Let $T \in K(X,Y)$, then by Theorem 6.49 one has explicitly

$$\inf_{\{W \subset X,\, \operatorname{codim} W \le n-1\}} \left\{ \sup_{x \in W} \frac{\|Tx\|^2}{\|x\|^2} \right\} = s_n(T)^2. \tag{6.17}$$

Observe that $T^*T \in K(X)$ is selfadjoint and nonnegative. Let $\lambda_1 \ge \lambda_2 \ge \cdots \ge 0$ be the eigenvalues of T^*T. Then by Theorem 6.47

$$\min_{\{V \subset X,\, \dim V \le n-1\}} \left\{ \max_{x \perp V} \frac{\langle T^*Tx, x\rangle}{\langle x, x\rangle} \right\} = \lambda_n, \tag{6.18}$$

where it is clear that the inequality may be allowed. The result follows by comparing (6.17) and (6.18). ☺

6.7 Canonical decompositions and Hilbert-Schmidt operators

Recall that $T \in B(X,Y)$ has a polar decomposition $T = U|T|$ where $|T| = (T^*T)^{\frac{1}{2}}$ is the unique nonnegative square root of the nonnegative selfadjoint operator T^*T in $B(X)$ and U is an isometric operator from $(\ker U)^\perp$ onto $\operatorname{ran} U$. If $T \in K(X,Y)$ then clearly the nonnegative operator $T^*T \in K(X)$. The singular value (number) $s_n(T)$ is the root of the positive eigenvalue λ_n of T^*T, $n \in \mathbb{N}$.

Lemma 6.51. Let X and Y be Hilbert spaces and let $T \in K(X,Y)$. Then $T^*T \in K(X)$ admits the decomposition

$$T^*Tx = \sum_{n=1}^{\infty} s_n(T)^2 \langle x, e_n \rangle e_n, \quad x \in X, \tag{6.19}$$

in terms of its eigenvectors corresponding to nonzero eigenvalues. Moreover, $|T| \in K(X)$ and

$$|T|x = \sum_{n=1}^{\infty} s_n(T) \langle x, e_n \rangle e_n, \quad x \in X. \tag{6.20}$$

Proof. Since $T^*T \in K(X)$ is selfadjoint one may apply Theorem 6.44 to obtain (6.19), where $s_n(T)^2$ are the eigenvalues of T^*T.

Denote the operator defined by the right-hand side of (6.20) by R. It clearly belongs to $B(X)$ and is nonnegative. Since $T^*T \in K(X)$ it follows that $s_n(T) \to 0$, which then implies that R is compact; cf. Example 4.49.

It is easily seen that $R^2 = T$. Hence by the uniqueness of the nonnegative square root it follows that $R = |T|$, i.e., the right-hand side of (6.20) coincides with the operator $|T|$. ☺

Theorem 6.52 (Canonical decomposition of compact operators)**.** Let X and Y be Hilbert spaces and let $T \in K(X,Y)$. Then there exists orthonormal systems (e_n) in X and (f_n) in Y, respectively, such that

$$Tx = \sum_{n=1}^{\infty} s_n(T) \langle x, e_n \rangle f_n, \quad x \in X. \tag{6.21}$$

Proof. Use the polar decomposition $T = U|T|$. Since $T \in K(X,Y)$ it follows from Lemma 6.51 that $|T| \in K(X)$. Then use the expansion in (6.20) to obtain

$$Tx = U|T|x = \sum_{n=1}^{\infty} s_n(T) \langle x, e_n \rangle Ue_n, \quad x \in X.$$

Here (e_n) stands for the orthonormal sytem of eigenvectors of T^*T and where one defines $f_n = Ue_n$. Then it is clear that

$$\langle f_i, f_j \rangle = \langle Ue_i, Ue_j \rangle = \langle e_i, e_j \rangle = \delta_{ij},$$

since U is isometric on $(\ker U)^{\perp}$. ☺

Remark 6.53. By applying (6.21) to $x = e_m$ one obtains for any $m \in \mathbb{N}$

$$Te_m = s_m(T) f_m, \quad \|Te_m\| = s_m(T). \tag{6.22}$$

From the identity $\langle Tx, y \rangle = \langle x, T^*y \rangle$ it follows immediately from (6.21) that T^* has the expansion

$$T^*y = \sum_{n=1}^{\infty} s_n(T) \langle y, f_n \rangle e_n, \quad y \in Y.$$

Furthermore one has

$$|T^*|y = \sum_{n=1}^{\infty} s_n(T) \langle y, f_n \rangle f_n, \quad y \in Y.$$

To see last identity observe

$$TT^*y = \sum_{n=1}^{\infty} s_n(T) \langle y, f_n \rangle Te_n = \sum_{n=1}^{\infty} s_n(T)^2 \langle y, f_n \rangle f_n, \quad y \in Y,$$

where (6.22) has been used. Observe that $|T^*|$ is the unique nonnegative root of TT^*. □

Lemma 6.54. Let X and Y be Hilbert spaces and let $T \in K(X,Y)$. Assume that (e_n) and (f_n) are orthonormal bases for $\overline{\operatorname{ran}}\, T^*$ and $\overline{\operatorname{ran}}\, T$, respectively. Then

$$\sum_{n=1}^{\infty} \|Te_n\|^2 = \sum_{n=1}^{\infty} \|T^* f_n\|^2.$$

Here the sums may diverge.

Proof. Since (e_n) and (f_n) are orthonormal bases for $\overline{\operatorname{ran}}\, T^*$ and $\overline{\operatorname{ran}}\, T$, respectively, one has

$$\begin{aligned}
\sum_{n=1}^{\infty} \|Te_n\|^2 &= \sum_{n=1}^{\infty} \left(\sum_{m=1}^{\infty} |\langle Te_n, f_m \rangle|^2 \right) \\
&= \sum_{n=1}^{\infty} \left(\sum_{m=1}^{\infty} |\langle e_n, T^* f_m \rangle|^2 \right) \\
&= \sum_{m=1}^{\infty} \left(\sum_{n=1}^{\infty} |\langle e_n, T^* f_m \rangle|^2 \right) = \sum_{m=1}^{\infty} \|T^* f_m\|^2.
\end{aligned}$$

Note that changing the order of the summation is allowed since all terms are nonnegative. ☺

Definition 6.55. Let X and Y be Hilbert spaces and let $T \in K(X,Y)$. Then the operator T is said to belong to the *Hilbert-Schmidt class* if for some and, hence, for all orthonormal bases (e_n) for $\overline{\operatorname{ran}}\, T^*$

$$\sum_{n=1}^{\infty} \|Te_n\|^2 < \infty.$$

Lemma 6.56. Let X and Y be Hilbert spaces and let $T \in K(X,Y)$. Then the operator T belongs to the Hilbert-Schmidt class if and only if

$$\sum_{n=1}^{\infty} s_n(T)^2 < \infty.$$

Proof. Recall that T^*T is a nonnegative selfadjoint operator in $K(X)$. Its eigenvectors form an orthonormal basis for $\overline{\operatorname{ran}}\, T^*T = \overline{\operatorname{ran}}\, T^*$, cf. Remark 6.41, and recall the expansion

$$T^*Tx = \sum_{n=1}^{\infty} s_n(T)^2 \langle x, e_n \rangle e_n, \quad x \in X,$$

which leads to $\|Te_m\|^2 = \langle T^*Te_m, e_m \rangle = s_m(T)^2$; cf. (6.22). Hence one sees the identity

$$\sum_{n=1}^{\infty} \|Te_n\|^2 = \sum_{n=1}^{\infty} s_n(T)^2,$$

where the sums may diverge. This shows the lemma. ☺

Note that T belongs to the Hilbert-Schmidt class if and only if T^* belongs to the Hilbert-Schmidt class; cf. Lemma 6.54.

Lemma 6.57. Let X, Y, and Z be Hilbert spaces and let $T \in K(X,Y)$ belong to the Hilbert-Schmidt class.

1. If $S \in B(Y,Z)$, then ST belongs to the Hilbert-Schmidt class.

2. If $S \in B(Z,X)$, then TS belongs to the Hilbert-Schmidt class.

Proof. (1) Assume that $S \in B(X,Z)$. Let (e_n) be an orthonormal basis for $\overline{\operatorname{ran}}\, T^*$. Then $\|STe_n\| \leq \|S\|\|Te_n\|$, and thus

$$\sum_{n=1}^{\infty} \|STe_n\|^2 \leq \|S\|^2 \sum_{n=1}^{\infty} \|Te_n\|^2 < \infty.$$

Now observe that

$$\overline{\operatorname{ran}}\, T^* \ominus \overline{\operatorname{ran}}\,(ST)^* = \{x \in \overline{\operatorname{ran}}\, T^* : Tx \in \ker S\}$$

Hence by deleting from (e_n) those elements for which $Te_n \in \ker S$ one obtains an orthonormal basis for $\overline{\operatorname{ran}}\,(ST)^*$. Hence ST belongs to the Hilbert-Schmidt class.

(2) It suffices to show that $(TS)^*$ belongs to the Hilbert-Schmidt class. Now $(TS)^* = S^*T^*$ and the result follows from (1). ☺

In Remark 4.58 it was shown that integral operators in $L^2(a,b)$ with a square-integrable kernel are compact. In fact such operators belong to the Hilbert-Schmidt class.

Theorem 6.58. Let $K : [c,d] \times [a,b] \to \mathbb{R}$ be square integrable, i.e.,

$$\int_c^d \int_a^b |K(x,y)|^2 \, dxdy < \infty.$$

Then the operator T, defined by

$$Tf(x) = \int_a^b K(x,y)f(y)\, dy, \quad x \in [c,d], \quad f \in L^2(a,b),$$

belongs to the Hilbert-Schmidt class from $L^2(a,b)$ to $L^2(c,d)$.

Proof. Let (e_n) be a complete orthonormal system for $L^2(a,b)$ and define

$$Te_n(x) = \int_a^b K(x,y)e_n(y)\, dy = \langle K_x, \bar{e}_n \rangle, \quad K_x(y) = K(x,y), \quad x \in [c,d].$$

Note that it is clear from the Fubini theorem that $K_x \in L^2(a,b)$ for almost all $x \in (c,d)$ and that

$$\|K_x\|^2 = \sum_{n=1}^{\infty} |\langle K_x, \bar{e}_n \rangle|^2;$$

observe that also $(\bar{e}_n)$ is a complete orthonormal system for $L^2(a,b)$. Hence it follows from

$$\|Te_n\|^2 = \int_c^d |Te_n(x)|^2 \, dx = \int_c^d |\langle K_x, \bar{e}_n \rangle|^2 \, dx,$$

and the monotone convergence theorem that

$$\begin{aligned}
\sum_{n=1}^{\infty} \|Te_n\|^2 &= \sum_{n=1}^{\infty} \int_c^d |\langle K_x, \bar{e}_n \rangle|^2 \, dx \\
&= \int_c^d \left(\sum_{n=1}^{\infty} |\langle K_x, \bar{e}_n \rangle|^2 \right) dx \\
&= \int_c^d \|K_x\|^2 \, dx \\
&= \int_c^d \left(\int_a^b |K(x,y)|^2 \, dy \right) dx < \infty.
\end{aligned}$$

☺

Chapter 7

Dual spaces, reflexive spaces, and weak convergence

The bounded linear functionals on a normed linear space can be given the structure of a normed linear space which turns out to be complete: the so-called dual space. The dual spaces of some frequently used Banach spaces will be determined. Bounded linear operators between Banach spaces have their counterparts on the level of dual spaces; the so-called conjugate linear operators. The double dual is an object of great interest: the original space has an isometric copy in the double dual space. If the isometric copy equals the double dual one speaks of a reflexive space. Reflexive spaces possess a number of very useful properties. In particular, Hilbert spaces are reflexive. Sometimes one needs a weakening of the usual convergence in normed linear spaces: this is where weak convergence and weak* convergence come into play.

7.1 Dual spaces and the Hahn-Banach theorem

Let X be a normed linear space with dual space $X' = B(X,\mathbb{K})$; cf. Definition 4.34. Recall that X' consists of all linear functionals $f : X \to \mathbb{K}$ which are bounded and whose norm is given by

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

Since $\mathbb{K}$ is complete the dual space X' is a Banach space, even if X itself is not complete; see Theorem 4.31.

Remark 7.1. Let X be a normed linear space and let $V \subset X$ be a linear subspace. If $F \in X'$, then $f = F\restriction_V$, the restriction of F to V, belongs to V', as can be seen from

$$|f(v)| = |F(v)| \leq \|F\|\,\|v\|, \quad v \in V.$$

The converse situation will be discussed in the next theorem. □

Theorem 7.2 (Hahn-Banach theorem for normed linear spaces)**.** Let X be a normed linear space and let $V \subset X$ be a linear subspace. If $f \in V'$, then there exists $F \in X'$ such that

$$F(v) = f(v), \quad v \in V, \quad \text{and} \quad \|F\| = \|f\|.$$

Proof. Since $f \in V'$ it is clear that $|f(x)| \leq \|f\|\,\|x\|$, $x \in V$. Define the map $p : X \to \mathbb{K}$ by $p(x) = \|f\|\,\|x\|$, $x \in X$. It is clear that

$$p(x+y) = \|f\|\,\|x+y\| \leq \|f\|\,\|x\| + \|f\|\,\|y\| = p(x) + p(y), \quad x, y \in X.$$

It is also clear that

$$p(\lambda x) = \|f\|\,\|\lambda x\| = \|\lambda|\,\|f\|\,\|x\| = |\lambda|\,p(x), \quad x \in X,\ \lambda \in \mathbb{K}.$$

In terms of the map p one sees that $|f(x)| \leq p(x)$, $x \in V$. Hence, by Theorem 1.45, there exists a linear functional $F : X \to \mathbb{K}$, extending f to all of X, such that

$$|F(x)| \leq p(x), \quad x \in X.$$

In other words, the extension F satisfies $|F(x)| \leq \|f\|\|x\|$, $x \in X$. This leads to $\|F\| \leq \|f\|$. However, since F is an extension of f, it is clear that $\|f\| \leq \|F\|$. Thus $\|F\| = \|f\|$. ☺

Theorem 7.3. Let X be a normed linear space, let $V \subset X$ be a linear subspace, and let $x_0 \in X$. Assume that

$$\delta = \operatorname{dist}(x_0, V) = \inf\{\|x_0 - v\| : v \in V\} > 0.$$

Then there exists $F \in X'$ such that

$$F(x_0) = \delta, \quad F \restriction V = 0, \quad \text{and} \quad \|F\| = 1.$$

Proof. Define the map $f : \operatorname{span}(x_0) + V \to \mathbb{K}$ by

$$f(\lambda x_0 + v) = \lambda\delta, \quad \lambda \in \mathbb{K}, \quad v \in V.$$

Then f is a well-defined linear map such that

$$f(x_0) = \delta, \quad f(v) = 0, v \in V.$$

Note that

$$\|f\| = \sup\left\{\frac{|\lambda\delta|}{\|\lambda x_0 + v\|} : v \in V\right\} = \sup\left\{\frac{\delta}{\|x_0 - v\|} : v \in V\right\} = \frac{\delta}{\delta} = 1.$$

Hence $f \in V'$ and $\|f\| = 1$. Now apply the Hahn-Banach theorem to extend $f \in V'$ to $F \in X'$, while preserving the norm. ☺

Corollary 7.4. Let X be a normed linear space, let $V \subset X$ be a linear subspace, and let $x_0 \in X$. Then $x_0 \in \overline{V}$ if and only if for every $f \in X'$ with $f \restriction V = 0$ also $f(x_0) = 0$.

Proof. If $f(V) = 0$ and $x_0 \in \overline{V}$, then clearly $f(x_0) = 0$.

Conversely, if $x_0 \notin \overline{V}$, then $\delta = d(x_0, V) > 0$ and by Theorem 7.3 there exists $f \in X'$ with $f \restriction V = 0$ and $f(x_0) = \delta$. ☺

Corollary 7.5. Let X be a normed linear space and let $x_0 \in X$ be nontrivial. Then there exists $F \in X'$ such that

$$F(x_0) = \|x_0\| \quad \text{and} \quad \|F\| = 1.$$

Proof. Take $V = \{0\}$ in Theorem 7.3. ☺

Corollary 7.6. Let X be a normed linear space. Then for each $x \in X$

$$\|x\| = \sup\{|f(x)| : f \in X', \|f\| = 1\}.$$

Proof. First observe that $|f(x)| \leq \|f\| \, \|x\|$, which implies

$$\sup\{|f(x)| : f \in X', \|f\| = 1\} \leq \|x\|.$$

Equality follows from the existence of an element $f \in X'$ such that $f(x) = \|x\|$ and $\|f\| = 1$. ☺

Theorem 7.7. Let X be a normed linear space. If X' is separable, then X is separable.

Proof. Since X' is separable, there is a countable set $E = \{f_n \in X' : n \in \mathbb{N}\}$ which is dense in X'; it is assumed that all f_n are nontrivial. For each $f_n \in E$ there exists $v_n \in X$ with $\|v_n\| = 1$ and

$$|f_n(v_n)| \geq \frac{1}{2}\|f_n\|.$$

Introduce $V = \overline{\text{span}}\,(v_n)$, so that V is a closed separable subspace of X. It will be shown that $V = X$, so that X is separable.

Assume that V is a proper closed linear subspace of X. Then there exists $f \in X'$ with $\|f\| = 1$ and the property

$$f(v) = 0, \quad v \in V.$$

Associated with this $f \in X'$ there is a sequence (f_{n_k}) from E such that $f_{n_k} \to f$ in X', and hence also $\|f_{n_k}\| \to \|f\| = 1$. Therefore the inequalities in

$$\frac{1}{2}\|f_{n_k}\| \leq |f_{n_k}(v_{n_k})| = |f_{n_k}(v_{n_k}) - f(v_{n_k})| \leq \|f_{n_k} - f\| \, \|v_{n_k}\| = \|f_{n_k} - f\|$$

lead to a contradiction. Hence $V = X$. ☺

By means of the Hahn-Banach theorem one can now show that in a normed linear space every finite-dimensional subspace is complemented.

Lemma 7.8. Let X be a normed linear space and let $V \subset X$ be a linear subspace. If V is finite-dimensional, then there exists a projection $P \in B(X)$ such that $V = \operatorname{ran} P$.

Proof. Let $\{e_1, \ldots, e_n\}$ be a basis for V. Let the elements $\{f_1, \ldots, f_n\}$ in X' be dual to this basis

$$f_i(e_j) = \delta_{ij}.$$

Define the linear operator $P : X \to X$ by

$$Px = \sum_{i=1}^{n} f_i(x) e_i, \quad x \in X.$$

Then it follows that $\operatorname{ran} P \subset V$ and it is clear that $Pe_j = e_j$, which shows that $\operatorname{ran} P = V$. Furthermore,

$$P^2 x = \sum_{i=1}^{n} f_i(x) Pe_i = \sum_{i=1}^{n} f_i(x) e_i = Px, \quad x \in X.$$

so that P is a projection onto V. Finally,

$$\|Px\| \leq \sum_{i=1}^{n} |f_i(x)| \|e_i\| \leq \left(\sum_{i=1}^{n} \|f_i\| \|e_i\| \right) \|x\|,$$

shows that $P \in B(X)$. ☺

7.2 The dual space of ℓ^p

In the next two sections a number of examples is presented. Recall that the dual space of a Hilbert space has already been discussed in Theorem 6.1; see also Section 7.6 of the present chapter.

Theorem 7.9. Assume that $1 \le p < \infty$ and $1/p + 1/q = 1$, so that $1 < q \le \infty$. For $a \in \ell^q$ the functional $f_a : \ell^p \to \mathbb{K}$ given by

$$f_a(x) = \sum_{i=1}^{\infty} x_i a_i, \quad x \in \ell^p, \tag{7.1}$$

belongs to $(\ell^p)'$. The map $a \mapsto f_a$ is an isometric isomorphism from ℓ^q onto $(\ell^p)'$.

Proof. For each $1 \le p < \infty$ the sequence $(x_i a_i)$ belongs to ℓ^1 due to the Hölder inequality. Hence f_a is a well-defined map from ℓ^p to $\mathbb{K}$ and it is clear that f_a is linear. Furthermore

$$\|f_a\| = \sup_{x \ne 0} \frac{|f_a(x)|}{\|x\|_p} \le \|a\|_q,$$

so that $f_a \in (\ell^p)'$.

Case $1 < p < \infty$. Let $f \in (\ell^p)'$ and define $a_i = f(e_i)$, where e_i is the i-th standard vector (which belongs to ℓ^p). Fix $n \in \mathbb{N}$ and define

$$b = \left(\frac{|a_1|^q}{a_1}, \dots, \frac{|a_n|^q}{a_n}, 0, 0, \dots \right),$$

with the understanding that whenever $a_i = 0$ the corresponding entry in b is defined as zero. Observe that

$$f(b) = f\left(\sum_{i=1}^{n} b_i e_i \right) = \sum_{i=1}^{n} b_i a_i = \sum_{i=1}^{n} |a_i|^q,$$

and that

$$|f(b)| \le \|f\| \|b\|_p = \|f\| \left(\sum_{i=1}^{n} |a_i|^q \right)^{1/p},$$

where the last equality follows from the observation that $|b_i|^p = |a_i|^q$. Combining these results one obtains

$$\left(\sum_{i=1}^{n} |a_i|^q \right)^{1/q} = \left(\sum_{i=1}^{n} |a_i|^q \right)^{1-1/p} \le \|f\|.$$

Since this is true for any $n \in \mathbb{N}$ it follows that $a \in \ell^q$ and $\|a\|_q \le \|f\|$.

Note that for this $a \in \ell^q$ the map $f_a \in (\ell^p)'$ is well-defined and that for any $n \in \mathbb{N}$

$$f\left(\sum_{i=1}^{n} x_i e_i \right) = \sum_{i=1}^{n} x_i a_i = f_a\left(\sum_{i=1}^{n} x_i e_i \right).$$

Thus f and f_a coincide on a dense subset of ℓ^p, which implies $f = f_a$. Hence it follows that

$$\|a\|_q \le \|f\| = \|f_a\| \le \|a\|_q,$$

and therefore $\|a\|_q = \|f\| = \|f_a\|$.

Case $p = 1$. Let $f \in (\ell^1)'$ and define $a_i = f(e_i)$, where e_i is the i-th standard vector (which belongs to ℓ^1). It is clear that

$$|a_i| = |f(e_i)| \le \|f\| \|e_i\| = \|f\|,$$

so that $a \in \ell^\infty$ and $\|a\|_\infty \le \|f\|$.

Note that $f_a \in (\ell^1)'$ is well-defined and that for any $n \in \mathbb{N}$

$$f\left(\sum_{i=1}^n x_i e_i\right) = \sum_{i=1}^n x_i a_i = f_a\left(\sum_{i=1}^n x_i e_i\right).$$

Thus f and f_a coincide on a dense subset of ℓ^1, which implies $f = f_a$. Hence it follows that

$$\|a\|_\infty \le \|f\| = \|f\|_a \le \|a\|_\infty,$$

and therefore $\|a\|_\infty = \|f\| = \|f_a\|$. ☺

Remark 7.10. The fact that $(\ell^p)'$ is isometrically isomorphic with ℓ^q for $1 \le p < \infty$) via (7.1) is usually denoted by $(\ell^p)' \simeq \ell^q$. For $1 < p < \infty$ each space ℓ^p and its dual space $(\ell^p)' \simeq \ell^q$ are separable. However, for $p = 1$ the space ℓ^1 is separable, but its dual $(\ell^1)' \simeq \ell^\infty$ is not separable. □

Theorem 7.11. For $a \in \ell^1$ the functional $f_a : \ell^\infty \to \mathbb{K}$ defined by

$$f_a(x) = \sum_{i=1}^\infty x_i a_i, \quad x \in \ell^\infty, \tag{7.2}$$

belongs to $(\ell^\infty)'$. The map $a \mapsto f_a$ is
1. an isometry from ℓ^1 to $(\ell^\infty)'$ which is not surjective;
2. an isometric isomorphism from ℓ^1 onto $(c_0)'$.

Proof. Let $a \in \ell^1$ and define f_a by (7.2). Then for all $x \in \ell^\infty$

$$|f_a(x)| \le \sum_{i=1}^\infty |x_i||a_i| \le \left(\sum_{i=1}^\infty |a_i|\right) \sup_{i \in \mathbb{N}} |x_i|. \tag{7.3}$$

Hence $f_a \in (\ell^\infty)'$ and $\|f_a\| \le \|a\|_1$.

(1) To see the reverse inequality, fix $n \in \mathbb{N}$ and define

$$b = \left(\frac{|a_1|}{a_1}, \dots, \frac{|a_n|}{a_n}, 0, 0, \dots\right), \tag{7.4}$$

with the understanding that whenever $a_i = 0$ the corresponding entry in b is defined as zero. Then $b \in c_0 \subset \ell^\infty$ and $\|b\|_\infty = 1$, so that

$$\sum_{i=1}^n |a_i| = \sum_{i=1}^n a_i \frac{|a_i|}{a_i} = f_a(b) \le \|f_a\|.$$

Since $n \in \mathbb{N}$ is arbitrary this shows that $\|a\|_1 \le \|f_a\|$. Assume that the isometric map $a \mapsto f_a$ from ℓ^1 is onto $(\ell^\infty)'$. Since ℓ^1 is separable, this implies that $(\ell^\infty)'$ is separable, which leads to a contradiction; cf. Theorem 7.7.

(2) For $x \in c_0$ the result in (7.3) shows that $f_a \in (c_0)'$ and $\|f_a\| \le \|a\|_1$. Now consider any $f \in (c_0)'$ and define $a_i = f(e_i)$, where e_i is the i-th standard vector (which belongs to c_0). Fix $n \in \mathbb{N}$ and define b as in (7.4). Note that $b \in c_0$ and that $\|b\|_\infty = 1$. Observe that

$$f(b) = f\left(\sum_{i=1}^n b_i e_i\right) = \sum_{i=1}^n b_i a_i = \sum_{i=1}^n |a_i|,$$

and that, since $f \in (c_0)'$,

$$|f(b)| \le \|f\|\|b\|_\infty \le \|f\|.$$

Hence
$$\sum_{i=1}^{n} |a_i| \le \|f\|.$$
Since $n \in \mathbb{N}$ is arbitrary this shows that $\|a\|_1 \le \|f\|$.

Note that $f_a \in (c_0)'$ is well-defined and that for any $n \in \mathbb{N}$
$$f\left(\sum_{i=1}^{n} x_i e_i\right) = \sum_{i=1}^{n} x_i a_i = f_a\left(\sum_{i=1}^{n} x_i e_i\right).$$
Thus f and f_a coincide on a dense subset of c_0, which implies $f = f_a$ and $\|a\|_1 = \|f_a\|$. ☺

Remark 7.12. Every bounded linear functional $f : c \to \mathbb{K}$ has the form
$$f(x) = x_0 a_0 + \sum_{i=1}^{\infty} x_i a_i, \quad x \in c, \quad x_0 = \lim_{n\to\infty} x_n,$$
where $(a_0, a_1, a_2, \dots) \in \ell^1$. Moreover
$$\|f\| = |a_0| + \sum_{i=1}^{\infty} |a_i|.$$
This follows by observing that $x - x_0 e \in c_0$, where $e = (1,1,\dots) \in \ell^\infty$. □

7.3 The dual space of $\mathcal{C}([a,b],\mathbb{C})$

Consider the Banach space $\mathcal{C}([a,b],\mathbb{C})$ provided with the sup-norm; cf. Example 2.10. In order to describe its dual space one needs the notion of the Lebesgue-Stieltjes integral or the Riemann-Stieltjes integral. A function $\alpha : [a,b] \to \mathbb{C}$ is said to be of *bounded variation* on $[a,b]$ if
$$V(\alpha) = \sup_P \sum_{i=1}^{n} |\alpha(t_i) - \alpha(t_{i-1})| < \infty,$$
where the supremum is taken over all partitions P of $[a,b]$ of the form
$$a = t_0 < t_1 < \cdots < t_n = b, \tag{7.5}$$
with $n \in \mathbb{N}$ arbitrary. The quantity $V(\alpha)$ is called the *total variation* of α. Note that $\alpha : [a,b] \to \mathbb{C}$ is of bounded variation if and only if its real and imaginary parts are of bounded variation. If $\alpha : [a,b] \to \mathbb{R}$ is monotonically nondecreasing, then α is of bounded variation. In fact, any real-valued function of bounded variation is the difference of two monotonically nondecreasing functions.

Let $f \in \mathcal{C}([a,b],\mathbb{C})$ be any continuous function and let $\alpha : [a,b] \to \mathbb{C}$. Then there exists a number I such for every $\varepsilon > 0$ there exists $\delta > 0$ so that for every partition P of the form (7.5) with $|P| = \max_{i=1,\dots,n} |t_i - t_{i-1}| < \delta$ one has
$$\left| I - \sum_{i=1}^{n} f(t_{i-1})[\alpha(t_i) - \alpha(t_{i-1})] \right| < \varepsilon. \tag{7.6}$$
The number I is the integral of f over $[a,b]$ with respect to α and is denoted by
$$I = \int_a^b f(t)\, d\alpha(t).$$

Note that if α is monotonically nondecreasing then the above integral can be developed as the Lebesgue-Stieltjes integral with the measure generated by α or, alternatively, as the Riemann-Stieltjes integral completely parallel to the Riemann integral for the case $\alpha(t) = t$. In the general case of a function of bounded variation α one may decompose α as above and piece the integral together by linearity. The resulting integral has the usual linearity properties and there is the usual inequality

$$\left| \int_a^b f(t)d\alpha(t) \right| \leq \|f\| V(\alpha), \quad f \in \mathcal{C}([a,b],\mathbb{C}), \tag{7.7}$$

where $\|f\|$ stands for the sup-norm of f on $[a,b]$.

As a direct consequence of the above definitions one observes the following. Let the function $\alpha : [a,b] \to \mathbb{R}$ be of bounded variation and define the functional $F : (\mathcal{C}([a,b],\mathbb{C})) \to \mathbb{C}$ by

$$F(f) = \int_a^b f(t)d\alpha(t), \quad f \in (\mathcal{C}([a,b],\mathbb{C})). \tag{7.8}$$

Then F is linear and due to (7.7) satisfies

$$|F(f)| \leq \|f\| V(\alpha), \quad f \in (\mathcal{C}([a,b],\mathbb{C})).$$

Hence $F \in (\mathcal{C}([a,b],\mathbb{C}))'$ and, in addition, $\|F\| \leq V(\alpha)$. Moreover, it will now be shown that every bounded linear functional on $\mathcal{C}([a,b],\mathbb{C})$ is of this form.

Theorem 7.13. Let $F \in (\mathcal{C}([a,b],\mathbb{C}))'$ be a bounded linear functional. Then there exists a function $\alpha : [a,b] \to \mathbb{R}$ of bounded variation such that (7.8) and $\|F\| = V(\alpha)$ hold.

Proof. Let $F : \mathcal{C}([a,b],\mathbb{C}) \to \mathbb{C}$ be a bounded linear functional. Note that $\mathcal{C}([a,b],\mathbb{C})$ is a closed linear subspace of the Banach space $\mathcal{B}([a,b],\mathbb{C})$ provided with the sup-norm; cf. Example 2.10. Hence by Theorem 7.2 there exists a bounded linear functional $\widetilde{F}$ on $\mathcal{B}([a,b],\mathbb{C})$ which extends F and satisfies $\|\widetilde{F}\| = \|F\|$. Note that for each $s \in [a,b]$ the characteristic function

$$\chi_s(t) = \begin{cases} 1, & \text{if } a \leq t \leq s, \\ 0 & \text{if } s < t \leq b, \end{cases} \tag{7.9}$$

belongs to $\mathcal{B}([a,b],\mathbb{C})$. Therefore, by means of the characteristic functions χ_s, the function $\alpha : [a,b] \to \mathbb{C}$ may be defined as follows:

$$\alpha(s) = \widetilde{F}(\chi_s), \quad s \in [a,b]. \tag{7.10}$$

First it will be shown that α is of bounded variation and that $V(\alpha) \leq \|\widetilde{F}\|$. In order to do this, choose an arbitrary partition P of $[a,b]$ of the form (7.5), let $\varepsilon_i \in \mathbb{C}$ be given by

$$\varepsilon_i = \frac{\overline{\alpha(t_i)} - \overline{\alpha(t_{i-1})}}{|\alpha(t_i) - \alpha(t_{i-1})|}, \quad i = 1, \dots, n, \tag{7.11}$$

with the understanding that $\varepsilon_i = 0$ when $\alpha(t_i) - \alpha(t_{i-1}) = 0$, and define the step function $h : [a,b] \to \mathbb{C}$ by

$$h(t) = \begin{cases} \varepsilon_1, & \text{if } t_0 \leq t \leq t_1, \\ \varepsilon_i, & \text{if } t_{i-1} < t \leq t_i, \ i = 2, \dots, n. \end{cases}$$

Since $|\varepsilon_i| \leq 1$, $i = 1, \dots, n$, one sees that $h \in \mathcal{B}([a,b],\mathbb{C})$ with $\|h\| \leq 1$, and one may write h in the equivalent form

$$h(t) = \sum_{i=1}^n \varepsilon_i [\chi_{t_i}(t) - \chi_{t_{i-1}}(t)],$$

see (7.9). Now apply $\widetilde{F}$ to this identity and use the definition (7.11) of ε_i:

$$\begin{aligned}\widetilde{F}(h) &= \sum_{i=1}^{n} \varepsilon_i [\widetilde{F}(\chi_{t_i}) - \widetilde{F}(\chi_{t_{i-1}})] \\ &= \sum_{i=1}^{n} \varepsilon_i [\alpha(t_i) - \alpha(t_{i-1})] \\ &= \sum_{i=1}^{n} |\alpha(t_i) - \alpha(t_{i-1})|.\end{aligned}$$

It follows from this identity that

$$\sum_{i=1}^{n} |\alpha(t_i) - \alpha(t_{i-1})| = |\widetilde{F}(h)| \le \|\widetilde{F}\| \|h\| \le \|\widetilde{F}\|.$$

Since this is true for any partition P of $[a,b]$, as in (7.5), one concludes by means of Theorem 7.2 that $V(\alpha) \le \|\widetilde{F}\| = \|F\|$.

Next it will be shown that (7.8) holds. Let $f \in \mathcal{C}([a,b],\mathbb{C})$ and let P be a partition of $[a,b]$ as in (7.5). Define the function g, depending on P, by

$$g(t) = \sum_{i=1}^{n} f(t_{i-1})[\chi_{t_i}(t) - \chi_{t_{i-1}}(t)], \tag{7.12}$$

so that $g \in \mathcal{B}([a,b],\mathbb{C})$. Apply $\widetilde{F}$ to this identity, then by (7.10) one obtains

$$\widetilde{F}(g) = \sum_{i=1}^{n} f(t_{i-1})[\alpha(t_i) - \alpha(t_{i-1})]. \tag{7.13}$$

The function g in (7.12) is a step function and it is therefore clear that

$$|g(t) - f(t)| = \begin{cases} |f(t_0) - f(t)|, & \text{if } a \le t \le t_1, \\ |f(t_{i-1}) - f(t)|, & \text{if } t_{i-1} < t \le t_i,\ i = 2, \dots, n. \end{cases} \tag{7.14}$$

If $|P| = \max_{i=1,\dots,n} |t_i - t_{i-1}| \to 0$, then it follows from (7.14) that $\|g - f\| \to 0$, since f is uniformly continuous on $[a,b]$. This in turn implies that $\widetilde{F}(g) \to \widetilde{F}(f)$ as $\widetilde{F}$ is a bounded linear functional on $\mathcal{B}([a,b],\mathbb{C})$. The continuity of the function f shows via (7.13) that also

$$\widetilde{F}(g) = \sum_{i=1}^{n} f(t_{i-1})[\alpha(t_i) - \alpha(t_{i-1})] \to \int_a^b f(t)\, d\alpha(t),$$

as $|P| \to 0$; see (7.6). Thus

$$\widetilde{F}(f) = \int_a^b f(t)\, d\alpha(t),$$

which confirms (7.8) as $\widetilde{F}$ agrees with F on $\mathcal{C}([a,b],\mathbb{C})$. The inequality $\|F\| \le V(\alpha)$ is clear from (7.8). ☺

The function α in (7.8) is not uniquely determined by F. However, it will be if a normalization is required. For instance, the conditions $\alpha(a) = 0$ and α is right continuous on (a,b) make α uniquely determined.

7.4 Conjugate operators in dual spaces

Definition 7.14. Let X and Y be normed linear spaces and let $T \in B(X,Y)$. Then the *conjugate operator* $T^\times : Y' \to X'$ is defined by

$$(T^\times f)(x) = f(Tx), \quad f \in Y', \quad x \in X.$$

Theorem 7.15. Let X and Y be normed linear spaces and let $T \in B(X,Y)$. Then $T^\times \in B(Y',X')$ and, in fact,

$$\|T^\times\| = \|T\|.$$

Proof. To see that the operator $T^\times$ is linear, choose $f,g \in Y'$ and $\lambda,\mu \in \mathbb{K}$. Then for all $x \in X$

$$\begin{aligned}(T^\times(\lambda f + \mu g))(x) &= (\lambda f + \mu g)(Tx) \\ &= \lambda f(Tx) + \mu g(Tx) = \lambda (T^\times f)(x) + \mu (T^\times g)(x),\end{aligned}$$

which implies $T^\times(\lambda f + \mu g) = \lambda T^\times f + \mu T^\times g$.

As to the boundedness of $T^\times$, observe that for all $f \in X'$ and $x \in X$ it follows from the definition that

$$|(T^\times f)(x)| = |f(Tx)| \leq \|f\|\|Tx\| \leq \|f\|\|T\|\|x\|.$$

Thus for all $f \in Y'$ one sees that $T^\times f \in X'$ and

$$\|T^\times f\| = \sup_{\|x\| \leq 1} |(T^\times f)(x)| \leq \|f\|\|T\|,$$

which leads to $\|T^\times\| = \sup_{\|f\| \leq 1} \|T^\times f\| \leq \|T\|$.

For $x_0 \in X$ there exists an element $f \in Y'$ such that $\|f\| = 1$ and $f(Tx_0) = \|Tx_0\|$; cf. Corollary 7.5. From this it follows that

$$\|Tx_0\| = f(Tx_0) = (T^\times f)(x_0) \leq \|T^\times f\|\|x_0\| \leq \|T^\times\|\|f\|\|x_0\| = \|T^\times\|\|x_0\|.$$

Since $x_0 \in X$ is arbitrary, this implies that $\|T\| \leq \|T^\times\|$. ☺

Lemma 7.16. Let X, Y, and Z be normed linear spaces.

1. Let $T,S \in B(X,Y)$ and $\lambda,\mu \in \mathbb{K}$, then $(\lambda T + \mu S)^\times = \lambda T^\times + \mu S^\times$;
2. Let $T \in B(X,Y)$ and $S \in B(Y,Z)$, then $(ST)^\times = T^\times S^\times$.

Proof. (1) For all $x \in X$ and $f \in Y'$ one has

$$\begin{aligned}((\lambda T + \mu S)^\times f)(x) &= f((\lambda T + \mu S)x) = f(\lambda Tx + \mu Sx) \\ &= f(\lambda Tx) + f(\mu Sx) = \lambda f(Tx) + \mu f(Sx) \\ &= \lambda (T^\times f)(x) + \mu (S^\times f)(x),\end{aligned}$$

so that for all $f \in Y'$

$$(\lambda T + \mu S)^\times f = \lambda T^\times f + \mu S^\times f,$$

which leads to the desired result.

(2) For all $x \in X$ and $f \in Z'$ one has

$$((ST)^\times f)(x) = f(STx) = f(S(Tx)) = (S^\times f)(Tx) = (T^\times(S^\times f))(x),$$

so that for $f \in Z'$

$$(ST)^\times f = T^\times S^\times f,$$

which leads to the desired result. ☺

Corollary 7.17. Let X and Y be normed linear spaces. If T is an isomorphism from X onto Y, then $T^\times$ is an isomorphism from Y' onto X'. Moreover, in this case

$$(T^\times)^{-1} = (T^{-1})^\times. \tag{7.15}$$

If the isomorphism T is isometric, then the isomorphism $T^\times$ is isometric.

Proof. Let T be an isomorphism from X onto Y. Then the statement in the first part is clear from Lemma 7.16.

Let T be an isometric isomorphism from X onto Y. To show the statement in the second part let $f \in Y'$. Then for all $x \in X$:

$$|(T^\times f)(x)| = |f(Tx)| \leq \|f\|\|Tx\| = \|f\|\|x\|,$$

so that $\|T^\times f\| \leq \|f\|$. To see the equality, choose $\varepsilon > 0$. Then there exists $y \in Y$ with $\|y\| = 1$ and $\|f\| - \varepsilon < |f(y)|$. Now define $x = T^{-1}y$, so that $x \in X$ and $\|x\| = 1$, and obtain

$$\|f\| - \varepsilon \leq |f(y)| = |f(Tx)| = |T^\times f(x)| \leq \|T^\times f\|.$$

Since $\varepsilon > 0$ is arbitrary it follows that $\|f\| \leq \|T^\times f\|$. One concludes therefore that $\|T^\times f\| = \|f\|$ for all $f \in Y'$. ☺

Lemma 7.18. Let X and Y be normed linear spaces and let $T \in B(X,Y)$. Let $f \in Y'$, then

$$f \in \ker T^\times \quad \Leftrightarrow \quad \operatorname{ran} T \subset \ker f.$$

Proof. According to the definition one has

$$(T^\times f)(x) = f(Tx), \quad x \in X.$$

If $f \in \ker T^\times$, then for all $x \in X$ one has $(T^\times f)(x) = 0$, which implies $f(Tx) = 0$ for all $x \in X$ or $\operatorname{ran} T \subset \ker f$. The converse goes similarly. ☺

Theorem 7.19. Let X and Y be normed linear spaces and let $T \in K(X,Y)$. Then $T^\times \in K(Y',X')$

Proof. Let (g_n) be a bounded sequence in Y', i.e., $\|g_n\| \leq c$ for some $c > 0$. It will be shown via the theorem of Arzela-Ascoli that there is subsequence (g_{n_k}) such that $(T^\times g_{n_k})$ converges in X'.

Let B be the unit ball in X. Since T is compact it follows that TB is relatively compact in Y. Hence $F = \overline{TB}$ is a compact subspace in Y. Consider F as a compact metric space and let $\mathcal{C}(F,\mathbb{K})$ be the linear space of continuous functions f from F to $\mathbb{K}$ provided with the sup norm

$$\|f\|_{\sup} = \sup_{y \in F} |f(y)|.$$

Recall that the elements $g_n \in Y'$ are bounded linear functionals from Y to $\mathbb{K}$, so that in fact these functions are continuous. Hence the restrictions

$$f_n = g_n \restriction F$$

are continuous functions from F to $\mathbb{K}$. The sequence (f_n) is bounded in the sup-norm on $\mathcal{C}(F,\mathbb{K})$:

$$\|f_n\|_{\sup} = \sup_{y \in F} |f_n(y)| = \sup_{y \in F} |g_n(y)| \leq c \sup_{y \in F} \|y\|,$$

since $\|g_n\| \leq c$. Moreover, by the same reasoning the sequence (f_n) is equicontinuous:

$$\begin{aligned}|f_n(y_1) - f_n(y_2)| &= |g_n(y_1) - g_n(y_2)| = |g_n(y_1 - y_2)| \\ &\leq \|g_n\|\|y_1 - y_2\| \leq c\|y_1 - y_2\|, \quad y_1, y_2 \in F.\end{aligned}$$

By Theorem A.9 there exists a subsequence (f_{n_k}) converging in $\mathcal{C}(F, \mathbb{K})$ provided with the sup-norm. Note that for the corresponding subsequence (g_{n_k}) one has by definition

$$\|T^\times g_{n_k} - T^\times g_{n_l}\| = \sup_{x \in B} |T^\times g_{n_k}(x) - T^\times g_{n_l}(x)| = \sup_{x \in B} |g_{n_k}(Tx) - g_{n_l}(Tx)|.$$

Now recall that $f_{n_k} = g_{n_k} \upharpoonright F$, so that

$$\|T^\times g_{n_k} - T^\times g_{n_l}\| = \sup_{x \in B} |f_{n_k}(Tx) - f_{n_l}(Tx)| = \sup_{y \in F} |f_{n_k}(y) - f_{n_l}(y)|,$$

where the last equality is valid, since TB is dense in F. Hence $(T^\times g_{n_k})$ is a Cauchy sequence in X', as the subsequence (f_{n_k}) is Cauchy in $\mathcal{C}(F, \mathbb{K})$. Since X' is a Banach space, it follows that $(T^\times g_{n_k})$ converges. ☺

7.5 The second dual space and reflexive spaces

Lemma 7.20. Let X be a normed linear space and let $x \in X$. Then the map $F_x : X' \to \mathbb{K}$ defined by

$$F_x(f) = f(x), \quad f \in X', \tag{7.16}$$

belongs to X''. In fact, $\|F_x\| = \|x\|$.

Proof. For all $f, g \in X'$ and $\lambda, \mu \in \mathbb{K}$ one has

$$F_x(\lambda f + \mu g) = (\lambda f + \mu g)(x) = \lambda f(x) + \mu g(x) = \lambda F_x(f) + \mu F_x(g),$$

which shows that the map $F_x : X' \to \mathbb{K}$ is linear. It follows from Corollary 7.6 that

$$\|F_x\| = \sup_{f \neq 0} \frac{|F_x(f)|}{\|f\|} = \sup_{f \neq 0} \frac{|f(x)|}{\|f\|} = \|x\|,$$

so that $F_x \in X''$. ☺

Definition 7.21. Let X be a normed linear space and let $x \in X$. The *natural map* $J_X : X \to X''$ is defined by

$$J_X(x) = F_x, \quad x \in X,$$

where F_x is defined in (7.16). In other words

$$J_X(x)(f) = f(x), \quad x \in X, \quad f \in X'.$$

Lemma 7.22. Let X be a normed linear space. The map J_X is a linear isometry from X into the Banach space X''.

Proof. For all $x, y \in X$ and $f \in X'$ it follows that

$$J_X(x+y)(f) = f(x+y) = f(x) + f(y) = J_X(x)(f) + J_X(y)(f),$$

so that $J_X(x+y) = J_X(x) + J_X(y)$. Similarly, it follows that $J_X(\lambda x) = \lambda J_X(x)$ for all $x \in X$ and $\lambda \in \mathbb{K}$. Hence, $J_X : X \to X''$ is linear.

It follows from Lemma 7.20 that

$$\|J_X(x)\| = \|F_x\| = \|x\|,$$

which shows that $J_X : X \to X''$ is an isometry. ☺

Remark 7.23. The normed linear space X is isometrically isomorphic to the subset $J_X(X)$ of the Banach space X''. In particular, since $\overline{J_X(X)}$ is closed in X'' and hence a Banach space, X is isometrically isomorphic to a dense subspace of a Banach space. This has been shown in a different context in Theorem 3.30. If X is not complete, then we can interpret $\overline{J_X(X)}$ as a completion of X. □

Definition 7.24. Let X be a normed linear space. Then X is called *reflexive* if $J_X(X) = X''$.

Proposition 7.25. Let X be a normed linear space. If X is reflexive, then X is complete.

Proof. The second dual X'', being the dual of X', is a Banach space. Since X is reflexive, the isometry J_X maps X isometrically onto X''. Hence X is a Banach space. ☺

Lemma 7.26. Let X be a separable normed linear space. If X is reflexive, then X' is separable.

Proof. Let X be separable. If X is reflexive, then also X'' is separable. But then Theorem 7.7 implies that X' is separable. ☺

Proposition 7.27. If X is a finite-dimensional normed linear space, then X is reflexive.

Proof. If $\dim X = n$, then $\dim X' = n$. Hence by the same reasoning $\dim X'' = n$. This implies that the linear isometry J_X maps X onto X''. ☺

Theorem 7.28. Let X be a Banach space. Then X is reflexive if and only if X' is reflexive.

Proof. ($\Rightarrow$) Assume that X is reflexive. Let $\rho \in X'''$, then $f = \rho \circ J_X$ belongs to X' and it will be shown that $J_{X'}(f) = \rho$. To verify this identity observe that every $\omega \in X''$ has the form $\omega = J_X(x)$ for some $x \in X$, since X is reflexive. Therefore

$$J_{X'}(f)(\omega) = \omega(f) = J_X(x)(f) = f(x) = (\rho \circ J_X)(x) = \rho(\omega), \quad \omega \in X'',$$

which shows the claim $J_{X'}(f) = \rho$. Hence $J_{X'}$ is surjective and thus X' is reflexive.

($\Leftarrow$) Assume that X' is reflexive. Hence for every $\rho \in X'''$ there exists $f \in X'$ such that $\rho = J_{X'}(f)$. Observe that with this representation

$$\rho(J_X(x)) = J_{X'}(f)(J_X(x)) = J_X(x)(f) = f(x), \quad x \in X. \tag{7.17}$$

Now assume that X is not reflexive. Then there exists a nontrivial $\omega \in X'' \setminus J_X(X)$. Since X is a Banach space the linear subspace $J_X(X)$ is closed in X''. Thus, according to Theorem 7.3, there exists some $\rho \in X'''$ for which

$$\rho(\omega) \neq 0 \quad \text{and} \quad \rho(J_X(x)) = 0, \quad x \in X.$$

Now write $\rho = J_{X'}(f)$ with $f \in X'$. Then according to (7.17) it follows that $f = 0$. However, $\rho(\omega) = J_{X'}(f)(\omega) = \omega(f) = 0$, which leads to a contradiction. Hence X' is reflexive. ☺

Theorem 7.29. Let X be a normed linear space and let $V \subset X$ be closed linear subspace. If X is reflexive, then V is reflexive.

Proof. Let $g \in V''$ be an arbitrary element. It will be shown that there exists an element $x \in V$ such that $g(f) = f(x)$ for all $f \in V'$.

For this purpose define $G : X' \to \mathbb{K}$ by

$$G(f) = g(f\restriction_V), \quad f \in X'.$$

Since $f\restriction_V$ belongs to V', it is clear that G is linear and that G is bounded as the composition of bounded operators:

$$|G(f)| \leq \|g\| \|f\restriction_V\| \leq \|g\| \|f\|, \quad f \in X',$$

cf. Remark 7.1. Therefore $G \in X''$. Since X is reflexive it follows that there exists $x \in X$ such that

$$G(f) = f(x), \quad f \in X'.$$

Assume that $x \notin V$. Then by Theorem 7.2 there exists $f \in X'$ such that

$$f(x) \neq 0 \quad \text{and} \quad f(v) = 0,\ v \in V.$$

Then

$$f(x) = G(f) = g(f\restriction_V) = 0,$$

which leads to a contradiction. Hence $x \in V$ and

$$g(f\restriction_V) = G(f) = f(x), \quad f \in X'.$$

In other words for all $f \in X'$ one has

$$g(f\restriction_V) = f\restriction_V(x), \quad x \in V.$$

Note that any element in V' is of the form $f\restriction_V$ for some $f \in X'$. Therefore V is reflexive. ☺

Proposition 7.30. The reflexivity or non-reflexivity of the sequence spaces ℓ^p, c_0, and c is as follows:

1. ℓ^p is reflexive for $1 < p < \infty$;
2. ℓ^1 is not reflexive;
3. c_0 is not reflexive;
4. c is not reflexive;
5. ℓ^∞ is not reflexive.

Proof. (1) Let $1 \leq p < \infty$. Denote by $T_p : \ell^q \to (\ell^p)'$ the isometric isomorphism $a \mapsto f_a$ as defined in Theorem 7.9. Then $T_p^\times : (\ell^p)'' \to (\ell^q)'$ is an isometric isomorphism. Let $J : \ell^p \to (\ell^p)''$ be the natural map. Then for $x \in \ell^p$ one has $Jx \in (\ell^p)''$, so that $T_p^\times(Jx) \in (\ell^q)'$. Moreover, for $y \in \ell^q$ one has

$$(T_p^\times(Jx))(y) = (Jx)(T_p y) = (T_p y)(x) = \sum_{i=1}^\infty y_i x_i.$$

Now observe that for $1 < p < \infty$ one also has

$$(T_q x)(y) = \sum_{i=1}^\infty y_i x_i.$$

Therefore, for $1 < p < \infty$ and all $x \in \ell^p$ one obtains:

$$T_p^\times \circ J(x) = T_q(x),$$

which shows $T_p^\times \circ J = T_q$. Since T_p and T_q are isometric isomorphisms, also $T_p^\times$ is an isometric isomorphism. This implies that J is an isometric isomorphism. Thus ℓ^p is reflexive for $1 < p < \infty$.

(2) Since $(\ell^1)' \simeq \ell^\infty$ is not separable, it follows from Lemma 7.26 that ℓ^1 is not reflexive.

(3) It follows from $c_0' \simeq \ell^1$ and $(\ell^1)' \simeq \ell^\infty$ that c_0'' is isometrically isomorphic to ℓ^∞. This implies that c_0'' is not separable. Hence c_0 is not reflexive.

(4) and (5) The non-reflexive space c_0 is a closed linear subspace of c and of ℓ^∞. Hence c and of ℓ^∞ are not reflexive by Theorem 7.29. ☺

7.6 Hilbert spaces and dual spaces

Let X be a Hilbert space and let $f \in X'$. According to Theorem 6.1 the map $\mathfrak{J}_X$ defined by $y \mapsto f_y$ is a conjugate linear isometry from X onto X'; recall that for $y \in X$ the functional $f_y \in X'$ is defined by

$$f_y(x) = \langle x, y \rangle, \quad x \in X.$$

belongs to X'. Recall that the dual space X' can be viewed as an inner product space when the following definition is used:

$$\langle f_y, f_z \rangle = \langle z, y \rangle, \quad y, z \in X,$$

and then X' is a Hilbert space. The special structure of the Hilbert space makes it possible to interpret the definition of conjugate operator as an object in the original space.

Theorem 7.31 (Adjoint operators in Hilbert spaces)**.** Let X and Y be Hilbert spaces and let $T \in B(X,Y)$. Then the adjoint operator $T^* \in B(Y,X)$ is given in terms of the conjugate operator $T^\times \in B(Y',X')$ by the formal identity

$$T^* = (\mathfrak{J}_X)^{-1} T^\times \mathfrak{J}_Y. \tag{7.18}$$

Proof. Let $w \in Y$, so that $\mathfrak{J}_Y w \in Y'$. Recall that $T^\times$ is a linear map from Y' to X', hence there is a unique $v \in X$ such that

$$T^\times[\mathfrak{J}_Y w] = \mathfrak{J}_X v \quad \text{or} \quad v = (\mathfrak{J}_X)^{-1} T^\times [\mathfrak{J}_Y w].$$

For any $x \in X$ it now follows that

$$\langle Tx, w \rangle = (\mathfrak{J}_Y w)(Tx) = [T^\times(\mathfrak{J}_Y w)](x) = [\mathfrak{J}_X v](x) = \langle x, v \rangle.$$

It follows from Theorem 6.7 that $v = T^* w$. Hence (7.18) follows. ☺

The identification of T^* with $T^\times$ in Theorem 7.31 makes it possible to obtain a number of results for the adjoint operator T^* in terms of the conjugate operator $T^\times$. Assume that X, Y, and Z are Hilbert spaces.

1. Let $T, S \in B(X,Y)$ and $\lambda, \mu \in \mathbb{K}$, then

$$\begin{aligned}(\lambda T + \mu S)^* &= (\mathfrak{J}_X)^{-1}(\lambda T^\times + \mu S^\times)\mathfrak{J}_Y \\ &= (\mathfrak{J}_X)^{-1}(\lambda T^\times)\mathfrak{J}_Y + (\mathfrak{J}_X)^{-1}(\mu S^\times)\mathfrak{J}_Y \\ &= \bar{\lambda}(\mathfrak{J}_X)^{-1}(T^\times)\mathfrak{J}_Y + \bar{\mu}(\mathfrak{J}_X)^{-1}S^\times)\mathfrak{J}_Y \\ &= \bar{\lambda} T^* + \bar{\mu} S^*,\end{aligned}$$

where it has been used that $\mathfrak{J}_X$, and hence its inverse $(\mathfrak{J}_X)^{-1}$, is anti-linear.

2. Let $T \in B(X,Y)$ and $S \in B(Y,Z)$, then

$$\begin{aligned}(ST)^* &= (\mathfrak{J}_X)^{-1}(ST)^\times \mathfrak{J}_Z = (\mathfrak{J}_X)^{-1}(T^\times S^\times)\mathfrak{J}_Z \\ &= [(\mathfrak{J}_X)^{-1}T^\times \mathfrak{J}_Z][(\mathfrak{J}_Z)^{-1}S^\times \mathfrak{J}_Z] = T^* S^*.\end{aligned}$$

3. Let $T \in B(X,Y)$ be an isomorphism. Then by (7.15) it follows that $T^* \in B(Y,X)$ is an isomorphism and

$$(T^*)^{-1} = (\mathfrak{J}_Y)^{-1}(T^\times)^{-1}\mathfrak{J}_X = (\mathfrak{J}_Y)^{-1}(T^{-1})^\times \mathfrak{J}_X = (T^{-1})^*.$$

4. Let $T \in K(X,Y)$. Then it follows from (7.18), Theorem 7.19, and Lemma 4.43 that

$$T^* \in K(Y,X).$$

Proposition 7.32. If X is a Hilbert space, then X is reflexive.

Proof. Let X be a Hilbert space and let X' be the dual space. Recall that $x \mapsto f_x$ defined by $f_x(z) = \langle z,x\rangle$ is an anti-linear isometry from X onto X'. The dual space X' has a Hilbert space structure with

$$\langle f_y, f_z\rangle = \langle z,y\rangle, \quad y,z \in X.$$

Now let $\rho \in X''$, then there exists $\varphi \in X'$ with

$$\rho(\psi) = \langle \psi, \varphi\rangle, \quad \psi \in X'.$$

Observe that $\varphi = f_x$ and $\psi = f_y$ for unique $x,y \in X$, and hence

$$\rho(\psi) = \langle f_y, f_x\rangle = \langle x,y\rangle = f_y(x) = \psi(x) = J_X(x)(\psi).$$

Therefore $\rho = J_X(x)$ and X is reflexive. ☺

7.7 Weak convergence and weak* convergence

Let X be a normed linear space. Recall the notions of convergent sequences and Cauchy sequences in X. In Chapter 6 it has been shown that for the special case that X is a Hilbert space, these notions also exist in a weak sense. In the present section it is shown how the usual convergence in a Banach space has weak equivalents which involve the dual space X'. First the notions of weak convergence and weakly bounded are considered.

Definition 7.33. Let X be a normed linear space and let (x_n) be a sequence in X. Then (x_n) *converges weakly* to $x \in X$ if $f(x_n) \to f(x)$ for all $f \in X'$. The sequence (x_n) is a *weak Cauchy sequence* on X if $(f(x_n))$ is a Cauchy sequence in $\mathbb{K}$ for all $f \in X'$.

Weak limits are uniquely determined. Note that the usual convergence of (x_n) implies weak convergence of (x_n). If $\dim X < \infty$ then weak convergence implies strong convergence with the same limit.

Definition 7.34. Let X be a normed linear space and let $E \subset X$ be a subset. Then E is called *weakly bounded* if for every $f \in X'$

$$\sup_{x\in E} |f(x)| < \infty.$$

For instance, weakly convergent sequences are weakly bounded.

Theorem 7.35. Let X be a normed linear space and let $E \subset X$ be a subset. Then E is weakly bounded if and only if E is bounded.

Proof. ($\Leftarrow$) There exists $M > 0$ such that $\|x\| \leq M$ for all $x \in E$. Hence, for all $f \in X'$ and $x \in E$ one has

$$|f(x)| \leq \|f\|\|x\| \leq M\|f\|,$$

which shows that E is weakly bounded.

($\Rightarrow$) Observe that for each $x \in E$ the operator $Jx \in (X')'$ is defined by

$$(Jx)(f) = f(x), \quad f \in X',$$

and recall that X' is a Banach space. By assumption

$$\sup_{x\in E} |(Jx)(f)| = \sup_{x\in E} |f(x)| = c_f < \infty.$$

Therefore by the uniform boundedness principle in Theorem 5.24 one obtains the uniform estimate $\sup_{x\in E} \|Jx\| < \infty$. This leads to

$$\sup_{x\in E} \|x\| = \sup_{x\in E} \|Jx\| < \infty,$$

which shows that E is bounded ☺

Lemma 7.36. Let X be a normed linear space, let $F \subset X'$ be a dense subset, and let (x_n) be a sequence in X. Then (x_n) converges weakly to $x \in X$ if and only if

1. the sequence $(\|x_n\|)$ is bounded;
2. $f(x_n) \to f(x)$ for all $f \in F$.

Proof. ($\Rightarrow$) Assume that (x_n) converges weakly to $x \in X$. Then (1) holds by Theorem 7.35 and (2) is trivial.

($\Leftarrow$) Assume that (1) and (2) hold. It will be shown that $f(x_n) \to f(x)$ for all $f \in X'$. There is $c > 0$ such that $\|x_n\| \leq c$, $\|x\| \leq c$. Let $\varepsilon > 0$, then there is $g \in F$ with

$$\|f - g\| \leq \frac{\varepsilon}{3c}.$$

Morover, since $g \in F$, there is $N \in \mathbb{N}$ such that for $n > N$

$$|g(x_n) - g(x)| < \frac{\varepsilon}{3}.$$

Hence for $n > N$

$$\begin{aligned} |f(x_n) - f(x)| &\leq |f(x_n) - g(x_n)| + |g(x_n) - g(x)| + |g(x) - f(x)| \\ &\leq \|f - g\|\|x_n\| + \frac{\varepsilon}{3} + \|f - g\|\|x\| < \varepsilon. \end{aligned}$$ ☺

Definition 7.37. Let X be a normed linear space. Then X is called *weakly complete* if every weak Cauchy sequence converges weakly in X.

Theorem 7.38. Let X be a reflexive normed linear space. Then X is weakly complete.

Proof. Let (x_n) be a sequence in X which is weakly Cauchy. Thus for any $f \in X'$ the sequence $(f(x_n))$ is Cauchy in $\mathbb{K}$, and hence it has a limit

$$\varphi(f) = \lim_{n\to\infty} f(x_n),$$

which clearly defines $\varphi : X' \to \mathbb{K}$ as a linear map.

Since (x_n) is weakly Cauchy it follows from Theorem 7.35 that the sequence (x_n) is bounded in X: there exists $M > 0$ such that $\|x_n\| \leq M$. Hence

$$|f(x_n)| \leq \|f\|\|x_n\| \leq M\|f\|,$$

which implies that

$$|\varphi(f)| \leq M\|f\|, \quad f \in X',$$

and thus $\varphi \in X''$. Since X is reflexive, there exists $x \in X$ with $Jx = \varphi$. Thus

$$f(x) = (Jx)(f) = \varphi(f) = \lim_{n\to\infty} f(x_n), \quad f \in X'.$$

Therefore (x_n) converges weakly to $x \in X$. ☺

The convergence of sequences in the dual space X' is defined in precisely the same way as above. Here is the formal definition.

Definition 7.39. Let X be a Banach space and let X' be its dual space. A sequence (f_n) in X' is said to *converge* to $f \in X'$ if $\|f_n - f\| \to 0$ where the norm is the one of X'. A sequence (f_n) in X' *converges weakly* to $f \in X'$ if for each $\varphi \in X''$ one has $\varphi(f_n) \to \varphi(f)$.

However, there is another natural way in which a sequence (f_n) in X' converges to $f \in X'$.

Definition 7.40. A sequence (f_n) in X' is said to converge in the *weak* sense* if there exists an element $f \in X'$ such that $f_n(x) \to f(x)$ for all $x \in X$.

The limit of a weak* convergent sequence is uniquely determined. Moreover, a weak* convergent sequence is bounded. Note that ordinary convergence of (f_n) implies weak convergence of (f_n); and that weak convergence of (f_n) implies weak* convergence of (f_n).

Weak* convergence has important applications in analysis. Note the following result which is based on the uniform boundedness principle.

Theorem 7.41. Let X be a Banach space and let $f_n \in X'$. If $f_n(x)$ converges in $\mathbb{K}$ for every $x \in X$, then the sequence (f_n) is uniformly bounded: $\sup_n \|f_n\| < \infty$. Moreover,

$$f(x) = \lim_{n\to\infty} f_n(x)$$

defines an operator $f \in X'$ and $\|f\| \leq \sup_n \|f_n\|$. In particular (f_n) converges in the weak* sense to $f \in X'$.

Proof. This follows as a special case from Theorem 5.27. ☺

Proposition 7.42. Let X be a Banach space, let $E \subset X$ be a dense subset, and let (f_n) be a sequence in X'. Then (f_n) is weak* convergent if and only if

1. the sequence $(\|f_n\|)$ is bounded;
2. for all $x \in E$ the sequence $(f_n(x))$ is Cauchy.

Proof. This follows from Corollary 5.28. ☺

Lemma 7.43. Let X be a separable Banach space and let (f_n) be a bounded sequence in X'. Then (f_n) has a weak* convergent subsequence.

Proof. The Banach space X is assumed to be separable. Hence there exists a subset $E = \{x_1, x_2, x_3, \ldots\}$ which is dense in X. The selection of the weak* convergent subsequence of (f_n) will be done via the elements of E.

Since (f_n) is bounded the sequence $(f_n(x_1))$ is bounded in $\mathbb{K}$ and thus there exists a subsequence (f_n^1) of f_n such that $(f_n^1(x_1))$ converges. Now consider $(f_n^1(x_2))$ which is bounded in $\mathbb{K}$. Hence there exists a subsequence (f_n^2) of (f_n^1) such that $(f_n^2(x_2))$ converges. Thus one gets the array of sequences

$$\begin{array}{cccc} f_1^1, & f_2^1, & f_3^1, & \cdots \\ f_1^2, & f_2^2, & f_3^2, & \cdots \\ f_1^3, & f_2^3, & f_3^3, & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{array}$$

where each sequence is a subsequence of the one above it, while for each $i \in \mathbb{N}$ the sequence

$$f_1^i(x_i), f_2^i(x_i), f_3^i(x_i), \ldots$$

converges. Now define the diagonal sequence $g_1, g_2, g_3, \ldots$, where

$$g_i = f_i^i.$$

It is clear that for each x_i the sequence $g_1(x_i), g_2(x_i), g_3(x_i), \ldots$ converges, as this sequence is apart from the first $i-1$ entries a subsequence of $(f_n^i(x_i))$, $n \in \mathbb{N}$. Thus the diagonal sequence is a bounded sequence which converges on each element of E. Hence by Proposition 7.42 the diagonal sequence converges in the weak* sense. ☺

Theorem 7.44. Let X be a reflexive normed linear space and let (x_n) be a bounded sequence in X. Then (x_n) has a weakly convergent subsequence.

Proof. By means of the sequence (x_n) define $V = \overline{\operatorname{span}}\{x_n : n \in \mathbb{N}\}$. Then V is a closed linear subspace which is separable. Moreover, by Theorem 7.29, V is reflexive. Therefore V' is separable by Lemma 7.26.

Since the sequence (x_n) is bounded in V, the sequence $(J_V x_n)$ is bounded in V''. By Lemma 7.43 it has a weak* convergent subsequence, again denoted by $(J_V(x_n))$, converging to $J_V(y)$ for some $y \in V \subset X$ as V is reflexive. Therefore for all $f \in X'$:

$$f(x_n) = (J_V(x_n))(f) \to (J_V(y)(f) = f(y),$$

which proves the assertion. Note that in the final conclusion it was tacitly assumed that the restriction of $f \in X'$ to V belongs to V'; see Remark 7.1. ☺

Chapter 8

Sturm-Liouville problems and compact operators

Boundary value problems from mathematical physics are often treated by Hilbert space methods. As a simple illustration a special boundary value problem for a one-dimensional Sturm-Liouville equation on a compact interval $[a,b]$ is considered. A solution of the corresponding inhomogeneous equation is written as an integral operator by means of Green's function, which is continuous on the square $[a,b] \times [a,b]$. Since the interval is compact the integral operator is compact in the Hilbert space $L^2(a,b)$, which gives the connection with Chapter 6. It is briefly indicated how a Sturm-Liouville expression gives rise to an unbounded linear operator in a Hilbert space.

8.1 Motivation

The motivation for the present topic comes from the method of separation of variables to solve partial differential equations[1]. This method results in ordinary differential equations with boundary conditions induced by the geometric configuration of the original problem.

Example 8.1. Consider the following differential equation:

$$-u'' = \lambda u, \tag{8.1}$$

on the interval $[0,\pi]$. Observe that for each $\lambda \in \mathbb{C}$ the differential equation $-u'' = \lambda u$ has two linearly independent complex-valued solutions. Furthermore, it is known that with the condition

$$u(c) = \alpha, \quad u'(c) = \beta,$$

where $0 \leq c \leq \pi$, the corresponding initial value problem has precisely one solution for any $\lambda \in \mathbb{C}$. However, the method of separation leads to so-called boundary value problems which involve both endpoints simultaneously.

As a first example, one may pose with the differential equation (8.1) the following boundary conditions:

$$u(0) = 0, \quad u(\pi) = 0. \tag{8.2}$$

The question is to obtain a non-trivial solution of this problem and it turns out that now one needs restrictions on the parameter $\lambda \in \mathbb{C}$. For instance when $\lambda = 0$ the differential equation $-u'' = 0$ has two linearly independent conditions 1 and x and it is clear that the only linear combination of these functions which satisfies the boundary conditions in (8.2) is the trivial one. In order to see

[1] See W.A. Strauss. Partial differential equations. An introduction. Wiley, 2008. Chapter 4.

what solutions there are, assume there is a nontrivial solution $u(\cdot,\lambda)\in \mathcal{C}^2([0,\pi],\mathbb{C})$ corresponding to $\lambda\in\mathbb{C}$. Then it follows from (8.2) and integration by parts that

$$\begin{aligned}\lambda\int_0^\pi |u(x,\lambda)|^2\,dx &= -\int_0^\pi u''(x,\lambda)\overline{u(x,\lambda)}\,dx\\ &= -u'(x,\lambda)\overline{u(x,\lambda)}\Big|_{x=0}^{\pi}+\int_0^\pi |u'(x,\lambda)|^2\,dx \\ &= \int_0^\pi |u'(x,\lambda)|^2\,dx.\end{aligned}\tag{8.3}$$

Since $u(\cdot,\lambda)$ is assumed to be nontrivial, it follows that $\lambda\geq 0$ and, in fact, that $\lambda>0$. So one may restrict the search to positive λ, and in this case every nontrivial solution of $-u''=\lambda u$ is of the form

$$u(x,\lambda)=c_1(\lambda)\cos\sqrt{\lambda}x+c_2(\lambda)\frac{\sin\sqrt{\lambda}x}{\sqrt{\lambda}},\quad x\in[0,\pi].$$

The boundary condition $u(0,\lambda)=0$ implies that $c_1(\lambda)=0$, while then the boundary condition $u(\pi,\lambda)=0$ leads to the condition

$$\sin\sqrt{\lambda}\pi=0.$$

Hence one obtains $\lambda=n^2$, $n\in\mathbb{N}$, with corresponding solution $\sin nx$. Thus the multiplicity of the eigenvalue $\lambda=n^2$ is one.

As a second example one may pose with the differential equation (8.1) the following boundary conditions:

$$u(0)=u(\pi),\quad u'(0)=u'(\pi).\tag{8.4}$$

Again one sees that for the existence of a nontrivial solution $u(\cdot,\lambda)$ one needs $\lambda\geq 0$; cf. (8.3). But this time there is a nontrivial solution for $\lambda=0$, namely $u(x,0)=1$, thus the multiplicity of the eigenvalue $\lambda=0$ is one. However, for $\lambda>0$ a similar reasoning leads to

$$\cos\sqrt{\lambda}\pi=1,$$

so that $\lambda=4n^2$, $n\in\mathbb{N}$, while the corresponding solution is spanned by $\cos 2nx$ and $\sin 2nx$, $n\in\mathbb{N}$. Thus the multiplicity of the eigenvalue $\lambda=4n^2$, $n\in\mathbb{N}$, is two. □

In general one encounters second order Sturm-Liouville differential expressions which are a bit more involved because also potentials will be taken into account. For the convenience of the reader a brief update about ordinary differential equations is provided. Then the connection between boundary value problems as above and selfadjoint compact operators in Hilbert spaces will be discussed; cf. Chapter 6. At the end of this section one can see that Sturm-Liouville differential expressions on an interval $[a,b]$ generate linear operators in the Hilbert space $L^2(a,b)$ which are unbounded.

8.2 Sturm-Liouville differential equations

Let the second-order differential expression L be given by

$$L=-\frac{d^2}{dx^2}+q(x)$$

on the compact interval $[a,b]$ and $q\in\mathcal{C}([a,b],\mathbb{R})$. For an arbitrary function $g\in\mathcal{C}([a,b],\mathbb{C})$ and $\lambda\in\mathbb{C}$ the inhomogeneous Sturm-Liouville equation is defined by

$$(L-\lambda)u=g.\tag{8.5}$$

One searches for solutions $u = u(\cdot,\lambda) \in \mathcal{C}^2([a,b],\mathbb{C})$. There is also occasion to consider the corresponding homogeneous equation

$$(L-\lambda)u = 0, \quad \lambda \in \mathbb{C}. \tag{8.6}$$

In order to state existence and uniqueness results for solutions of these equations the connection with first order systems of differential equations is explained.

An equivalent formulation of (8.5) in the form of a first order system is as follows

$$\begin{pmatrix} u(\cdot,\lambda) \\ u'(\cdot,\lambda) \end{pmatrix}' = \begin{pmatrix} 0 & 1 \\ q-\lambda & 0 \end{pmatrix} \begin{pmatrix} u(\cdot,\lambda) \\ u'(\cdot,\lambda) \end{pmatrix} + \begin{pmatrix} 0 \\ -g \end{pmatrix}. \tag{8.7}$$

It is known that for $c \in [a,b]$ and each pair $\alpha,\beta \in \mathbb{C}$ there is precisely one solution of (8.7) such that

$$\begin{pmatrix} u(c,\lambda) \\ u'(c,\lambda) \end{pmatrix} = \begin{pmatrix} \alpha \\ \beta \end{pmatrix}. \tag{8.8}$$

This can be seen by writing the initial value problem for the system as an integral equation of Volterra type, which can be solved by means of successive approximations[2]. This method also shows that for each $x \in [a,b]$ the map

$$\lambda \mapsto \begin{pmatrix} u(x,\lambda) \\ u'(x,\lambda) \end{pmatrix} \tag{8.9}$$

is entire in λ. Hence from the general theory one obtains the following useful theorem.

Theorem 8.2. Let $g \in \mathcal{C}([a,b],\mathbb{C})$. Then for each $c \in [a,b]$ and each pair $\alpha,\beta \in \mathbb{C}$ there is precisely one function $u(\cdot,\lambda) \in \mathcal{C}^2([a,b],\mathbb{C})$ which satisfies (8.5) and has the initial values

$$u(c,\lambda) = \alpha, \quad u'(c,\lambda) = \beta. \tag{8.10}$$

Moreover, for each $x \in [a,b]$ the functions

$$\lambda \mapsto u(x,\lambda), \quad \lambda \mapsto u'(x,\lambda) \tag{8.11}$$

are entire in $\lambda \in \mathbb{C}$.

It is often required to check if two solutions of the homogeneous equation (8.6) are linearly dependent or independent. Let the functions φ,ψ belong to $\mathcal{C}^1([a,b],\mathbb{C})$ and define their Wronskian by $W(\varphi,\psi) = \varphi\psi' - \varphi'\psi$.

Lemma 8.3. Let $\varphi(\cdot,\lambda),\psi(\cdot,\lambda) \in \mathcal{C}^2([a,b],\mathbb{C})$ be solutions of the homogeneous equation (8.6) with $\lambda \in \mathbb{C}$. Then there exists a constant $\gamma \in \mathbb{C}$ such that

$$W(\varphi(\cdot,\lambda),\psi(\cdot,\lambda)) = \gamma \tag{8.12}$$

for all $\lambda \in \mathbb{C}$. Moreover, the solutions $\varphi(\cdot,\lambda)$ and $\psi(\cdot,\lambda)$ are linearly independent if and only if $\gamma \neq 0$.

Proof. Differentiation of the left-hand side of (8.12) leads to

$$\begin{aligned} W(\varphi(\cdot,\lambda),\psi(\cdot,\lambda))' &= [\varphi(\cdot,\lambda)\psi'(\cdot,\lambda) - \varphi'(\cdot,\lambda)\psi(\cdot,\lambda)]' \\ &= \varphi(\cdot,\lambda)\psi''(\cdot,\lambda) - \varphi''(\cdot,\lambda)\psi(\cdot,\lambda) \\ &= \varphi(\cdot,\lambda)[q-\lambda]\psi(\cdot,\lambda) - [(q-\lambda)\varphi(\cdot,\lambda)]\psi(\cdot,\lambda) = 0, \end{aligned}$$

[2]See W. Walter, Ordinary differential equations. Springer, 1998. Chapter III, §13.

so that there exists a constant $\gamma \in \mathbb{C}$ for which (8.12) holds.

If $\gamma = 0$ then $W(\varphi(\cdot,\lambda),\psi(\cdot,\lambda)) = 0$ and there is non-trivial pair $c_1(\lambda), c_2(\lambda) \in \mathbb{C}$ such that

$$\begin{pmatrix} \varphi(\cdot,\lambda) & \psi(\cdot,\lambda) \\ \varphi'(\cdot,\lambda) & \psi'(\cdot,\lambda) \end{pmatrix} \begin{pmatrix} c_1(\lambda) \\ c_2(\lambda) \end{pmatrix} = 0,$$

and it follows that $\varphi(\cdot,\lambda), \psi(\cdot,\lambda)$ are linearly dependent. The converse is shown similarly. ☺

Define the solutions $u_1(\cdot,\lambda), u_2(\cdot,\lambda) \in \mathcal{C}^2([a,b],\mathbb{C})$ of the homogeneous equation (8.6) with a fixed $\lambda \in \mathbb{C}$ by the following initial vales at a:

$$\begin{cases} u_1(a,\lambda) = 1, \\ u_1'(a,\lambda) = 0, \end{cases} \qquad \begin{cases} u_2(a,\lambda) = 0, \\ u_2'(a,\lambda) = 1. \end{cases} \tag{8.13}$$

Then for each $x \in [a,b]$ the functions $\lambda \mapsto u_i(x,\lambda)$, $i = 1,2$, are entire. Moreover, $W(u_1(\cdot,\lambda),u_2(\cdot,\lambda)) = 1$ and the following result is clear.

Lemma 8.4. The functions $u_1(\cdot,\lambda)$ and $u_2(\cdot,\lambda)$ form a fundamental system of solutions of the homogeneous equation (8.6): every solution of the homogeneous equation is a linear combination of $u_1(\cdot,\lambda)$ and $u_2(\cdot,\lambda)$

It will now be shown that the general solution of (8.5) is a linear combination of the fundamental system of the homogeneous equation (8.6) and one particular solution of the inhomogeneous equation (8.5).

Lemma 8.5. Let $g \in \mathcal{C}([a,b],\mathbb{C})$ and let $u \in \mathcal{C}^2([a,b],\mathbb{C})$ be a solution of (8.5). Then there exists $c_1(\lambda), c_2(\lambda) \in \mathbb{C}$ such that

$$\begin{aligned} u(x,\lambda) &= c_1(\lambda)u_1(x,\lambda) + c_2(\lambda)u_2(x,\lambda) \\ &\quad + u_1(x,\lambda)\int_a^x u_2(t,\lambda)g(t)\,dt - u_2(x,\lambda)\int_a^x u_1(t,\lambda)g(t)\,dt. \end{aligned} \tag{8.14}$$

Proof. Define the function h by

$$h(x) = u_1(x,\lambda)\int_a^x u_2(t,\lambda)g(t)\,dt - u_2(x,\lambda)\int_a^x u_1(t,\lambda)g(t)\,dt$$

Then it is clear that $h \in \mathcal{C}^2([a,b],\mathbb{C})$ and that $(L-\lambda)h = g$. This implies that $(L-\lambda)(u-h) = 0$ and it follows that $u-h$ is a linear combination of $u_1(\cdot,\lambda)$ and $u_2(\cdot,\lambda)$; cf. Lemma 8.4. ☺

8.3 Sturm-Liouville boundary value problems

Sturm-Liouville boundary value problems are problems where solutions of the Sturm-Liouville equation (8.7) are searched which satisfy conditions at *both* endpoints a and b. For simplicity only the simultaneous conditions

$$u(a) = 0, \quad u(b) = 0, \tag{8.15}$$

will be treated here. But one could treat in a similar manner also conditions like $u'(a) = 0$, $u'(b) = 0$ or $u(a) = 0$, $u'(b) = 0$; or, even, periodic bounday conditions of the form $u(a) = u(b)$, $u'(a) = u'(b)$; see Section 8.4 for more details.

Theorem 8.6. Let $[a,b] \subset \mathbb{R}$ be a compact interval and let $q \in \mathcal{C}([a,b],\mathbb{R})$. Then the eigenvalues of the eigenvalue problem

$$(L-\lambda)u = 0, \quad u(a) = u(b) = 0, \quad \lambda \in \mathbb{C}, \tag{8.16}$$

i.e., the values of λ for which (8.16) has a nontrivial solution $u(\cdot,\lambda)$, coincide with the values of $\lambda \in \mathbb{C}$ for which $u_2(b,\lambda) = 0$. Consequently, there are at most countably many eigenvalues which do not cluster except possibly at ∞. The eigenvalues are all real and bounded below by $\min_{x\in[a,b]} q(x)$:

$$\lambda_1 < \lambda_2 < \lambda_3 < \cdots. \tag{8.17}$$

The multiplicity of each eigenvalue λ_n is one and the corresponding eigenspace is spanned by $u_2(\cdot,\lambda_n)$.

Proof. Let $u_1(\cdot,\lambda)$ and $u_2(\cdot,\lambda)$ be solutions of (8.7) fixed by the initial conditions (8.13). It is clear that $u_1(\cdot,\lambda)$ and $u_2(\cdot,\lambda)$ are linearly independent. Any nontrivial solution of (8.7) has the form $u(\cdot,\lambda) = c_1(\lambda)u_1(\cdot,\lambda) + c_2(\lambda)u_2(\cdot,\lambda)$ with not both $c_1(\lambda)$ and $c_2(\lambda)$ being trivial simultaneously. Hence if $u(a,\lambda) = 0$ it follows that $c_1(\lambda) = 0$ and, in fact, $u(\cdot,\lambda) = c_2(\lambda)u_1(\cdot,\lambda)$. Furthermore it is clear that $u(b,\lambda) = 0$ if and only if $u_2(b,\lambda) = 0$. As the function $\lambda \mapsto u_2(b,\lambda)$ is entire, its zeros $(\lambda_n) \in \mathbb{C}$ are discrete and do not cluster except possibly at ∞. Note that the multiplicity of each eigenvalue λ_n is one: each eigenspace is spanned by $u_2(\cdot,\lambda_n)$.

Let $u(\cdot,\lambda)$ be a nontrivial solution of (8.16) corresponding to $\lambda \in \mathbb{C}$. Then

$$\begin{aligned}
\lambda \int_a^b |u(x,\lambda)|^2\,dx &= \int_a^b \left[-u''(x,\lambda) + q(x)u(x,\lambda)\right]\overline{u(x,\lambda)}\,dx \\
&= \int_a^b -u''(x,\lambda)\overline{u(x,\lambda)}\,dx + \int_a^b q(x)u(x,\lambda)\overline{u(x,\lambda)}\,dx \\
&= \int_a^b |u'(x,\lambda)|^2\,dx + \int_a^b q(x)|u(x,\lambda)|^2\,dx.
\end{aligned}$$

Since q is real-valued, this shows that $\lambda \in \mathbb{R}$ and moreover that

$$\lambda \int_a^b |u(x,\lambda)|^2\,dx \geq \int_a^b q(x)|u(x,\lambda)|^2\,dx \geq \min_{x\in[a,b]} q(x) \int_a^b |u(x,\lambda)|^2\,dx,$$

so that $\lambda \geq \min_{x\in[a,b]} q(x)$. ☺

Now it will be shown that the inhomogeneous equation $(L-\lambda)u = g$ has a unique solution if λ is not an eigenvalue of (8.16). Hence one may choose $\lambda \in \mathbb{C}\setminus\mathbb{R}$ or $\lambda \in \mathbb{R}$ with $\lambda \neq \lambda_n$; for instance any $\lambda < \min_{x\in[a,b]} q(x)$ is not an eigenvalue.

Theorem 8.7. Let $[a,b] \subset \mathbb{R}$ be a compact interval and let $q \in \mathcal{C}([a,b],\mathbb{R})$. Assume that $\lambda \in \mathbb{C}$ is not an eigenvalue of (8.16), or equivalently, $u_2(b,\lambda) \neq 0$. Then the inhomogeneous problem

$$(L-\lambda)u = g, \quad u(a) = u(b) = 0,$$

where $g \in \mathcal{C}([a,b],\mathbb{C})$, has a unique solution given by

$$u(x) = \chi(x,\lambda)\int_a^x u_2(t,\lambda)g(t)\,dt + u_2(x,\lambda)\int_x^b \chi(t,\lambda)g(t)\,dt, \tag{8.18}$$

when the solution $\chi(\cdot,\lambda)$ is defined by

$$\chi(x,\lambda) = u_1(x,\lambda) + M(\lambda)\, u_2(x,\lambda), \quad M(\lambda) = -\frac{u_1(b,\lambda)}{u_2(b,\lambda)}. \tag{8.19}$$

Proof. Use Lemma 8.5 and determine $c_1(\lambda)$ and $c_2(\lambda)$ in (8.14) so that u satisfies $u(a) = u(b) = 0$. Clearly, the condition $u(a) = 0$ gives $c_1(\lambda) = 0$. Then the condition $u(b) = 0$ gives

$$c_2(\lambda) = -\frac{u_1(b,\lambda)}{u_2(b,\lambda)} \int_a^b u_2(t,\lambda) g(t)\, dt + \int_a^b u_1(t,\lambda) g(t)\, dt.$$

Putting these values of $c_1(\lambda)$ and $c_2(\lambda)$ back into (8.14) gives the desired result. ☺

The solution in (8.18) of Theorem 8.7, with $\lambda \in \mathbb{C}$ not being an eigenvalue, can be written as as an integral

$$u(x,\lambda) = \int_a^b G(x,t,\lambda) g(t)\, dt,$$

where the so called *Green's function* $G(\cdot,\cdot,\lambda)$ is defined by

$$G(x,t,\lambda) = \begin{cases} \chi(x,\lambda) u_2(t,\lambda), & a \le t \le x \le b, \\ u_2(x,\lambda) \chi(t,\lambda), & a \le x \le t \le b. \end{cases} \tag{8.20}$$

The function $G(\cdot,\cdot,\lambda)$ is continuous on the square $[a,b] \times [a,b]$.

Definition 8.8. Let $[a,b] \subset \mathbb{R}$ be a compact interval and let $q \in \mathcal{C}([a,b],\mathbb{R})$. For each $\lambda \in \mathbb{C}$ which is not an eigenvalue of (8.16), or equivalently, $u_2(b,\lambda) \neq 0$, the integral operator R_λ is defined by

$$R_\lambda g(x) = \int_a^b G(x,t,\lambda) g(t)\, dt. \tag{8.21}$$

Due to the continuity of the Green's function in (8.20) the operator R_λ is a bounded linear operator from $\mathcal{C}([a,b],\mathbb{C})$ to $\mathcal{C}([a,b],\mathbb{C})$ with the sup norm; cf. Proposition 4.54.

Corollary 8.9. Let $[a,b] \subset \mathbb{R}$ be a compact interval and let $q \in \mathcal{C}([a,b],\mathbb{R})$. Assume that $\lambda \in \mathbb{C}$ is not an eigenvalue of (8.16), or equivalently, $u_2(b,\lambda) \neq 0$. Then

$$\operatorname{ran} R_\lambda = \{u \in \mathcal{C}^2([a,b],\mathbb{C}) : u(a) = u(b)\}.$$

Proof. Let $u \in \operatorname{ran} R_\lambda$. Then $u = R_\lambda g$ for some $g \in \mathcal{C}([a,b],\mathbb{C})$. It follows from (8.18) that u belongs to $\mathcal{C}^2([a,b],\mathbb{C})$ and $u(a) = u(b) = 0$.

Conversely, assume that $u \in \mathcal{C}^2([a,b],\mathbb{C})$ and $u(a) = u(b) = 0$. With $g = (L-\lambda)u$ it follows that $g \in \mathcal{C}([a,b],\mathbb{C})$, and $u = R_\lambda g$ by Theorem 8.7. Hence $u \in \operatorname{ran} R_\lambda$. ☺

Example 8.10. Return to the Sturm-Liouville equation in Example 8.1 with the boundary value problem (8.2). A fundamental system for the homogeneous Sturm-Liouville equation $(L-\lambda)u = 0$ on $[0,\pi]$ is given by

$$u_1(x,\lambda) = \cos\sqrt{\lambda}x, \quad u_2(x,\lambda) = \frac{\sin\sqrt{\lambda}x}{\sqrt{\lambda}},$$

as (8.13) is satisfied. Now Theorem 8.7 may be applied. In particular, note that $M(\lambda)$ and $\chi(\cdot,\lambda)$ in (8.19) are given by

$$M(\lambda) = -\frac{u_1(\pi,\lambda)}{u_2(\pi,\lambda)} = -\sqrt{\lambda}\cot\sqrt{\lambda}\pi, \quad \chi(x,\lambda) = \cos\sqrt{\lambda}x + M(\lambda)\frac{\sin\sqrt{\lambda}x}{\sqrt{\lambda}}.$$

Therefore, the corresponding Green's function $G(\cdot,\cdot,\lambda)$ in (8.20) is given by

$$G(x,t,\lambda) = \begin{cases} \left(\cos\sqrt{\lambda}x + M(\lambda)\,\frac{\sin\sqrt{\lambda}x}{\sqrt{\lambda}}\right)\frac{\sin\sqrt{\lambda}t}{\sqrt{\lambda}}, & \text{if } 0 \le t \le x \le \pi, \\ \frac{\sin\sqrt{\lambda}x}{\sqrt{\lambda}}\left(\cos\sqrt{\lambda}t + M(\lambda)\,\frac{\sin\sqrt{\lambda}t}{\sqrt{\lambda}}\right), & \text{if } 0 \le x \le t \le \pi. \end{cases}$$

□

8.4 Compact operators for Sturm-Liouville boundary value problems

Assume that $\lambda \in \mathbb{C}$ is not an eigenvalue of (8.16). The integral operator R_λ in (8.21) of Definition 8.8 is a bounded linear operator from $\mathcal{C}([a,b],\mathbb{C})$ to $\mathcal{C}([a,b],\mathbb{C})$ with the sup norm. Furthermore, it is clear that the integral operator R_λ can be seen as a bounded linear operator from $\mathcal{C}([a,b],\mathbb{C})$ to $\mathcal{C}([a,b],\mathbb{C})$ with the L^2-norm. Since $\mathcal{C}([a,b],\mathbb{C})$ is dense in $L^2(a,b)$ the operator R_λ can be uniquely extended to a bounded linear operator $\dot{R}_\lambda$ from $L^2(a,b)$ to $L^2(a,b)$:

$$\dot{R}_\lambda g(x) = \int_a^b G(x,t,\lambda)g(t)\,dt, \quad g \in L^2(a,b); \tag{8.22}$$

cf. Proposition 4.54. Due to its structure the integral operator R_λ in (8.21) is a compact operator from $\mathcal{C}([a,b],\mathbb{C})$ to $\mathcal{C}([a,b],\mathbb{C})$. Hence, also the operator $\dot{R}_\lambda$ in (8.22) from $L^2(a,b)$ to $L^2(a,b)$ is compact; cf. Proposition 4.54.

Theorem 8.11. Assume that $\lambda \in \mathbb{C}$ is not an eigenvalue of (8.16). Let the operator $\dot{R}_\lambda$ from $L^2(a,b)$ to $L^2(a,b)$ be given by (8.22). Then the adjoint $(\dot{R}_\lambda)^*$ is given by

$$(\dot{R}_\lambda)^* = \dot{R}_{\bar{\lambda}}. \tag{8.23}$$

Moreover, the operator $\dot{R}_\lambda$ is compact and has the following properties:
1. $\ker \dot{R}_\lambda = \{0\}$;
2. $\ker(\dot{R}_\lambda - \zeta) = \ker(R_\lambda - \zeta)$, $\zeta \in \mathbb{C}$.

Now assume that $\lambda \in \mathbb{R}$ is not an eigenvalue of (8.16). Then the operator $\dot{R}_\lambda$ is selfadjoint.

Proof. The right-hand side of (8.22) defines a bounded linear operator in $L^2(a,b)$. Its restriction to $\mathcal{C}([a,b],\mathbb{C})$ coincides with R_λ. The identity (8.23) follows from

$$\langle (R_\lambda)^* h, g\rangle = \langle h, R_\lambda g\rangle = \langle R_{\bar{\lambda}} h, g\rangle, \quad h,g \in \mathcal{C}([a,b],\mathbb{C}),$$

where in the last equality the Fubini theorem has been used.

Note that $\ker \dot{R}_\lambda = (\operatorname{ran}(\dot{R}_\lambda)^*)^\perp = (\operatorname{ran}\dot{R}_{\bar{\lambda}})^\perp$; cf. Lemma 6.10. Therefore, since $\operatorname{ran} R_{\bar{\lambda}} \subset \operatorname{ran}\dot{R}_{\bar{\lambda}}$, it suffices to show that $\operatorname{ran} R_{\bar{\lambda}}$ is dense in $L^2(a,b)$. Recall from Lemma 8.9 that

$$\operatorname{ran} R_{\bar{\lambda}} = \{u \in \mathcal{C}^2([a,b],\mathbb{C}) : u(a) = u(b)\}.$$

The set of all such functions is dense in $L^2(a,b)$; cf. Corollary 3.53.

Let $\zeta \in \mathbb{C}$. It is clear that $\ker(R_\lambda - \zeta) \subset \ker(\dot{R}_\lambda - \zeta)$. To show the reverse inclusion let $u \in \ker(\dot{R}_\lambda - \zeta)$. This means that $u \in L^2(a,b)$ and

$$\zeta u(x) = \int_a^b G(x,t,\lambda)u(t)\,dt,$$

and, since $\zeta \neq 0$, it follows from Corollary 8.9 that $u \in \mathcal{C}([a,b],\mathbb{C})$, so that $\zeta u = R_\lambda u$.

Since the operator R_λ is compact, it automatically follows that also the operator $\dot{R}_\lambda$ is compact; cf. Proposition 4.52. ☺

It should be emphasized that the boundary conditions $u(a) = 0$, $u(b) = 0$ in Theorem 8.7 give rise to the integral operator $\dot{R}_\lambda$ in (8.21) which is selfadjoint in $L^2(a,b)$, when $\lambda \in \mathbb{R}$ is not an eigenvalue. In a similar way, selfadjoint boundary conditions give rise to selfadjoint integral operators. For instance $u'(a) = 0$, $u'(b) = 0$; $u(a) = 0$, $u'(b) = 0$; or $u(a) = u(b)$, $u'(a) = u'(b)$ are selfadjoint boundary conditions.

Proposition 8.12. Assume that $\mu \in \mathbb{C}$ is not an eigenvalue of (8.16). Then there is a one-to-one correspondence between the eigenvalues λ of (8.16) and the eigenvalues ζ of $\dot{R}_\mu$ via

$$\zeta = \frac{1}{\lambda - \mu}.$$

Moreover, the nontrivial function $u \in \mathcal{C}^2([a,b],\mathbb{C})$ satisfies (8.16) if and only if u is an eigenfunction of $\dot{R}_\mu$:

$$\dot{R}_\mu u = \frac{1}{\lambda - \mu} u.$$

Proof. Since $\mu \in \mathbb{C}$ is not an eigenvalue of (8.16), we have the following equivalences:

$$\begin{aligned} (L-\lambda)u = 0, \quad u(a) = u(b) = 0 &\Leftrightarrow (L-\mu)u = (\lambda-\mu)u, \quad u(a) = u(b) = 0 \\ &\Leftrightarrow u = (\lambda-\mu)R_\mu u \\ &\Leftrightarrow u = (\lambda-\mu)\dot{R}_\mu u. \end{aligned}$$

Hence the proposition is clear. ☺

The eigenfunctions of the boundary value problem (8.16) corresponding to the eigenvalues λ_n have been determined in Theorem 8.6. For the following result it is customary to normalize the eigenfunctions in the sense of $L^2(a,b)$. The normalized eigenfunction e_n corresponding to λ_n is given by

$$e_n(x) = \frac{u_2(x,\lambda_n)}{\sqrt{\int_a^b |u_2(x,\lambda_n)|^2\,dx}}, \quad n \in \mathbb{N}.$$

Theorem 8.13. Let $[a,b] \subset \mathbb{R}$ be a compact interval and let $q \in \mathcal{C}([a,b],\mathbb{R})$. Then the normalized eigenvectors (e_n) of (8.16) form an orthonormal basis for $L^2(a,b)$: every element $u \in L^2(a,b)$ has the expansion

$$u = \sum_{n=1}^{\infty} \langle u, e_n \rangle e_n, \quad \langle u, e_n \rangle = \int_a^b u(t)\overline{e_n(t)}\,dt, \tag{8.24}$$

in the sense of $L^2(a,b)$.

Proof. Assume that $\mu \in \mathbb{R}$ is not an eigenvalue of (8.16). By Theorem 8.11 the operator $\dot{R}_\mu$ is compact and selfadjoint. ☺

The functions $\chi(\cdot,\lambda)$ and $M(\lambda)$ in Theorem 8.7 appear in the Green's function for the boundary value problem $u(a) = u(b) = 0$. There is an interesting connection between them.

Lemma 8.14. The functions $\chi(\cdot,\lambda)$ and $M(\lambda)$ in (8.19) of Theorem 8.7 satisfy

$$\int_a^b \chi(t,\lambda)\overline{\chi(t,\mu)}\,dt = \frac{M(\lambda) - \overline{M(\mu)}}{\lambda - \overline{\mu}}, \quad \lambda,\mu \in \mathbb{C}\setminus\mathbb{R}. \tag{8.25}$$

Proof. The definition of $\chi(\cdot,\zeta) = u_1(\cdot,\zeta) + M(\zeta)u_2(\cdot,\zeta)$ in (8.19) implies the identity $L\chi(\cdot,\zeta) = \zeta\chi(\cdot,\zeta)$. Hence it follows that

$$\begin{aligned}
(\lambda - \overline{\mu}) &\int_a^b \chi(t,\lambda)\overline{\chi(t,\mu)}\,dt \\
&= \int_a^b \lambda\chi(t,\lambda)\overline{\chi(t,\mu)}\,dt - \int_a^b \chi(t,\lambda)\overline{\mu\chi(t,\mu)}\,dt \\
&= \int_a^b L\chi(t,\lambda)\overline{\chi(t,\mu)}\,dt - \int_a^b \chi(t,\lambda)\overline{L\chi(t,\mu)}\,dt \\
&= -\chi'(t,\lambda)\overline{\chi(t,\mu)}\big|_a^b + \chi(t,\lambda)\overline{\chi'(t,\mu)}\big|_a^b,
\end{aligned}$$

where the last equality has been obtained via integration by parts. Since $\chi(a,\zeta) = 1$ and $\chi'(a,\zeta) = M(\zeta)$, whenever ζ is not an eigenvalue, it is clear that

$$\chi'(a,\lambda)\overline{\chi(a,\mu)} - \chi(a,\lambda)\overline{\chi'(a,\mu)} = M(\lambda) - \overline{M(\mu)}.$$

Likewise, since $\chi(b,\zeta) = 0$, whenever ζ is not an eigenvalue, it is also clear that

$$-\chi'(b,\lambda)\overline{\chi(b,\mu)} + \chi(b,\lambda)\overline{\chi'(b,\mu)} = 0,$$

which completes the proof. ☺

The function $\lambda \mapsto M(\lambda)$ is sometimes called a Titchmarsh-Weyl function. It is holomorphic on $\mathbb{C}\setminus\mathbb{R}$, $\overline{M(\lambda)} = M(\overline{\lambda})$, and $\operatorname{Im} M(\lambda)/\operatorname{Im}\lambda \geq 0$. The poles of the meromorphic function $\lambda \mapsto M(\lambda)$ coincide with the eigenvalues of (8.16) and, in fact, this function carries all the information about the boundary value problem (8.16). Such functions appear frequently in the study of Schrödinger operators.

8.5 Unbounded operators and Sturm-Liouville boundary value problems

In the Hilbert space $L^2(a,b)$ one can associate a linear operator T with the Sturm-Liouville boundary value problem in the following way: the domain is defined by

$$\operatorname{dom} T = \{u \in \mathcal{C}^2([a,b],\mathbb{C}) : u(a) = u(b) = 0\},$$

while the action of T is then given by

$$Tu = Lu.$$

Note that $\operatorname{ran} T = \mathcal{C}([a,b],\mathbb{C}) \subset L^2(a,b)$. It is clear that the linear operator T is only densely defined. Note that the eigenvalues and eigenvectors of T are precisely those of the Sturm-Liouville problem. In particular, the operator T is unbounded. Also note that

$$\langle Tu, u\rangle \geq \left[\min_{x\in[a,b]} q(x)\right]\langle u,u\rangle, \quad u \in \operatorname{dom} T.$$

To see this, observe that by integration by parts one has for all $u \in \operatorname{dom} T$

$$\begin{aligned}
\langle Tu, u\rangle &= \int_a^b \left[-u''(x) + q(x)u(x)\right]\overline{u(x)}\,dx \\
&= \int_a^b \left[|u'(x)|^2 + q(x)|u(x)|^2\right]dx \geq \int_a^b q(x)|u(x)|^2\,dx.
\end{aligned}$$

Similarly one sees that T is symmetric in the following sense

$$\langle Tu,v\rangle = \langle u,Tv\rangle, \quad u,v \in \operatorname{dom} T. \tag{8.26}$$

The following result shows that the bounded linear operator R_λ is the "inverse" of the unbounded linear operator $T-\lambda$, when $\ker(T-\lambda)=\{0\}$. A situation like this has been met before in the context of bounded linear operators, see Section 5.5.7.

Lemma 8.15. Assume that $\lambda \in \mathbb{C}$ is not an eigenvalue of (8.16) or, equivalently, that $\ker(T-\lambda)=\{0\}$. Then

$$(T-\lambda)R_\lambda g = g, \quad g \in \mathcal{C}([a,b],\mathbb{C}), \tag{8.27}$$

and

$$R_\lambda(T-\lambda)u = u, \quad u \in \operatorname{dom} T. \tag{8.28}$$

Proof. The identity (8.27) is clear. But then it follows that

$$(T-\lambda)R_\lambda(T-\lambda)u = (T-\lambda)u, \quad u \in \operatorname{dom} T,$$

or, equivalently,

$$(T-\lambda)(R_\lambda(T-\lambda)u-u)=0.$$

By assumption $\ker(T-\lambda)=\{0\}$, so that (8.28) follows. ☺

With these facts in mind there is an interesting interpretation for the expansion in (8.24): when the function u is sufficiently smooth and satisfies the boundary conditions, then there is also a pointwise interpretation.

Corollary 8.16. Let $u \in \operatorname{dom} T$. Then

$$u(x) = \sum_{n=1}^{\infty} \langle u,e_n\rangle e_n(x), \quad x \in [a,b],$$

where the convergence is uniform on $[a,b]$.

Proof. Assume that $u \in \operatorname{dom} T$. Since $e_k \in \operatorname{dom} T$, it is clear from (8.26) that

$$\begin{aligned} T\left(\sum_{k=1}^{n} \langle u,e_k\rangle e_k\right) &= \sum_{k=1}^{n} \langle u,e_k\rangle Te_k = \sum_{k=1}^{n} \langle u,e_k\rangle \lambda_k e_k \\ &= \sum_{k=1}^{n} \langle u,\lambda_k e_k\rangle e_k = \sum_{k=1}^{n} \langle u,Te_k\rangle e_k = \sum_{k=1}^{n} \langle Tu,e_k\rangle e_k. \end{aligned}$$

Now choose $\lambda \in \mathbb{C}$ such that $\ker(T-\lambda)=\{0\}$. Since $u,e_k \in \operatorname{dom} T$ it follows from the previous identity and Lemma 8.15 that

$$\begin{aligned} u-\sum_{k=1}^{n} \langle u,e_k\rangle e_k &= R_\lambda(T-\lambda)\left(u-\sum_{k=1}^{n} \langle u,e_k\rangle e_k\right) \\ &= R_\lambda\left((T-\lambda)u-\sum_{k=1}^{n} \langle (T-\lambda)u,e_k\rangle e_k\right). \end{aligned}$$

Observe that the functions

$$u-\sum_{k=1}^{n} \langle u,e_k\rangle e_k \quad \text{and} \quad (T-\lambda)u-\sum_{k=1}^{n} \langle (T-\lambda)u,e_k\rangle e_k$$

are both continuous on $[a,b]$, and it follows that

$$\left\| u - \sum_{k=1}^{n} \langle u, e_k \rangle e_k \right\|_\infty = \left\| R_\lambda \left((T-\lambda)u - \sum_{k=1}^{n} \langle (T-\lambda)u, e_k \rangle e_k \right) \right\|_\infty .$$

However, since R_λ is an integral operator with continuous kernel $G(\cdot,\cdot,\lambda)$ it is clear that

$$\begin{aligned} &\left\| R_\lambda \left(Tu - \sum_{k=1}^{n} \langle Tu, e_k \rangle e_k \right) \right\|_\infty \\ &\qquad \le M_\lambda \sqrt{b-a} \left\| (T-\lambda)u - \sum_{k=1}^{n} \langle (T-\lambda)u, e_k \rangle e_k \right\|_{L^2(a,b)} , \end{aligned}$$

which follows from the Cauchy-Schwarz inequality. An application of Theorem 8.13, with u replaced by $(T-\lambda)u$, shows that the right-hand side converges to 0 as $n \to \infty$. This proves the assertion. ☺

Remark 8.17. The graph of the operator T is not closed. However it follows from the lemma that

$$R_\lambda = (T-\lambda)^{-1}.$$

Note that the graph of $\dot{R}_\lambda \in B(L^2(a,b))$ is in fact the closure of the graph of R_λ. Therefore $\dot{R}_\lambda$ is the closure of the graph of $(T-\lambda)^{-1}$, which leads to

$$\dot{R}_\lambda = (\dot{T}-\lambda)^{-1},$$

where $\dot{T}$ is the closure of the graph of T. For $\lambda \in \mathbb{R}$ the selfadjointness of $\dot{R}_\lambda$ is the same as the selfadjointness of $(\dot{T}-\lambda)^{-1}$. In terms of $\dot{T}$ it follows that $(\dot{T}-\lambda)^{-1}$ is selfadjoint if and only if

$$\langle \dot{T}f, g \rangle = \langle f, \dot{T}g \rangle, \quad f, g \in \operatorname{dom} \dot{T},$$

and if for some h,k one has

$$\langle \dot{T}f, h \rangle = \langle f, k \rangle, \quad f \in \operatorname{dom} \dot{T},$$

then $h \in \operatorname{dom} \dot{T}$ and $\dot{T}h = k$. It can be shown that the closure $\dot{T}$ is again a differential operator, but its domain has to be described in terms of integration theory. In fact $\operatorname{dom} \dot{T}$ is given by all $u \in \mathcal{C}^1([a,b],\mathbb{C})$ for which

$$u' \in AC([a,b]),\ -u'' + qu \in L^2(a,b),\ u(a) = u(b) = 0,$$

while the action of $\dot{T}$ is then given by

$$\dot{T}u = -u'' + qu, \quad u \in \operatorname{dom} \dot{T}.$$

Here AC stands for the absolutely continuous functions, i.e., functions of the form $\int_a^x \varphi\, dx$, $\varphi \in L^1(a,b)$, and these functions are differentiable almost everywhere with derivative φ. In particular, $\dot{T}u$ is defined only almost everywhere. □

Appendix A

Compactness in metric spaces

For the convenience of the reader the following facts about compact subsets of metric spaces are briefly reviewed.

Definition A.1. Let X be a metric space and let $V \subset X$ be a subset. Then V is called *(sequentially) compact* if every sequence in V has a subsequence which converges to an element in V. Moreover, V is called *relatively compact* if $\overline{V}$ is compact.

Lemma A.2. Let X be a metric space and let $V \subset X$. Then V is relatively compact if and only if every sequence in V has a subsequence which converges (to an element in $\overline{V}$).

Proof. Let V be relatively compact, i.e., $\overline{V}$ is compact. Hence every sequence in V (thus in $\overline{V}$), has a subsequence which converges to an element in $\overline{V}$.

Conversely, assume that every sequence in V has a subsequence which converges (to an element in $\overline{V}$). Now let (x_n) be a sequence in $\overline{V}$ and choose a sequence $y_n \in V$ with $d(x_n, y_n) < 1/n$. Then y_n has a subsequence y_{n_k} which converges to an element in $\overline{V}$. Hence the corresponding subsequence x_{n_k} converges to the same element. Thus $\overline{V}$ is compact and V is relatively compact. ☺

Definition A.3. Let (X,d) be a metric space and let $V \subset X$ be a subset. For $\varepsilon > 0$ a subset $V_\varepsilon \subset X$ is called an *ε-net* for V if for every point $x \in V$ there is $u \in V_\varepsilon$ such that $d(x,u) < \varepsilon$.

Definition A.4. Let X be a metric space and let $V \subset X$ be a subset. Then V is called *totally bounded* if for every $\varepsilon > 0$ there exists a finite ε-net for V.

Thus $V \subset X$ is totally bounded if for every $\varepsilon > 0$ the set V is contained in the union of finitely many open balls of radius ε. Clearly, total boundedness implies boundedness but the converse does not hold; cf. Theorem 2.32. A subset of a totally bounded set is totally bounded.

Lemma A.5. Let X be a metric space and let $V \subset X$ be totally bounded. Then

1. for every $\varepsilon > 0$ there exists a finite ε-net $V_\varepsilon \subset V$;
2. V is separable.

Proof. (1) Let $\varepsilon > 0$ then there exists a finite collection of open balls with radius $\varepsilon/2$ which cover V. Denote these balls by $B_1, \dots, B_n$ with centers $x_1, \dots, x_n$. Consider those balls B_i which have a nonempty intersection with V and choose $z_i \in V \cap B_i$. Then the open balls with radius ε and center z_i form an ε-net for V. To see this let $x \in V$, then there exists $B(x_i; \varepsilon/2)$ such that $x \in B(x_i; \varepsilon/2)$.

Then

$$d(x,z_i) \le d(x,x_i) + d(x_i,z_i) < \varepsilon/2 + \varepsilon/2 = \varepsilon,$$

so that $x \in B(z_i;\varepsilon)$.

(2) For each $n \in \mathbb{N}$ there is a finite $1/n$-net in V for V; see (1). The union of these nets is countable. Now let $x \in V$ and let $\varepsilon > 0$. Then choose $n \in \mathbb{N}$ with $1/n < \varepsilon$. There exists $x_n \in V$ with $d(x,x_n) < 1/n$. Hence V is separable. ☺

Lemma A.6. Let X be a metric space and let $V \subset X$. If V is relatively compact, then V is totally bounded.

Proof. Assume that V is relatively compact and that V is not totally bounded. Then there exists $\varepsilon > 0$ for which there is no ε-net for V. Now a sequence (a_n) in V without a convergent subsequence will be constructed by induction. Choose $a_1 \in V$ arbitrarily. Assume that $a_1,\dots,a_n$ have been chosen in V such that

$$d(a_i,a_j) \ge \varepsilon.$$

Then there exists $a_{n+1} \in V$ with $d(a_i,a_{n+1}) \ge \varepsilon$, since otherwise $a_1,\dots,a_n$ would be an ε-net for V. The sequence (a_n) constructed in this way has no convergent subsequence, which gives a contradiction. ☺

Due to Lemma A.6 one concludes that a relatively compact set is bounded and separable; this conclusion is applied, for instance, in Lemma 4.41.

In these notes the notion of compactness is in fact the notion of sequentially compactness. However in a metric space the notion of compactness involving open covers coincides with the notion of sequentially compact.

Theorem A.7. Let X be a metric space and let $V \subset X$. Then the following statements are equivalent:

1. V is sequentially compact;
2. V is totally bounded and complete;
3. every open cover of V has a finite subcover.

Proof. (1) $\Rightarrow$ (2) Let (a_n) be a Cauchy sequence in V. Since V is sequentially compact, (a_n) has a subsequence (a_{n_k}) which converges to some $a \in V$. Hence (a_n) itself converges, which shows that V is complete. Furthermore, it follows from Lemma A.6 that V is totally bounded.

(2) $\Rightarrow$ (3) Assume that V is totally bounded and complete. Assume that there exists an open cover of V, say $V \subset \bigcup_i O_i$, which has no finite subcover. First it is proven that there exists a sequence of open balls $B_n = B(x_n;1/2^n)$ with the following properties:

(i) $V \cap B_n$ has no finite subcover;
(ii) $B_{n-1} \cap B_n \neq \emptyset$ for $n \ge 2$.

Since V is totally bounded there is a finite number of open balls with radius $1/2$ which cover V. Hence there is at least one open ball $B_1 = B(x_1;1/2)$ such that $V \cap B_1$ has no finite subcover (otherwise V would have a finite subcover). This proves the claim for $n=1$. Assume that $B_{n-1} = B(x_{n-1};1/2^{n-1})$ is an open ball such that $V \cap B_{n-1}$ has no finite subcover. Since V is totally bounded there is a finite number of open balls, say $V_1,\dots,V_m$, with radius $1/2^n$ which cover V. There is at least one V_i such that $V \cap B_{n-1} \cap V_i \neq \emptyset$ and $V \cap B_{n-1} \cap V_i$ has no finite subcover (otherwise $V \cap B_{n-1}$ would have a finite subcover). Denote this V_i by $B_n = B(x_n;1/2^n)$. Note that $V \cap B_n$ has no finite subcover (otherwise $V \cap B_{n-1} \cap B_n$ would have a finite subcover). The claim now follows by induction.

The above construction with the open balls B_n now leads to a contradiction. First observe that with $x \in B_{n-1} \cap B_n$ it follows that

$$d(x_{n-1},x_n) \leq d(x_{n-1},x) + d(x,x_n) \leq \frac{1}{2^{n-1}} + \frac{1}{2^n} < \frac{1}{2^{n-2}}.$$

Now choose for every $n \in \mathbb{N}$ an element $a_n \in V \cap B_n$ and observe

$$d(a_{n-1},a_n) \leq d(a_{n-1},x_{n-1}) + d(x_{n-1},x_n) + d(x_n,a_n) \leq \frac{3}{2^{n-2}}.$$

This gives the following estimate for every $p \in \mathbb{N}$:

$$\begin{aligned} d(a_n,a_{n+p}) &\leq d(a_n,a_{n+1}) + \cdots + d(a_{n+p-1},a_{n+p}) \\ &\leq \frac{3}{2^{n-1}}\left(1 + \frac{1}{2} + \cdots\right) = \frac{3}{2^{n-2}}. \end{aligned}$$

Hence (a_n) is a Cauchy sequence in V and since V is assumed to be complete it follows that $a_n \to a$ with $a \in V$. Due to the open cover $V \subset \bigcup_i O_i$ there exists i_0 such that $a \in O_{i_0}$. Since O_{i_0} is open there exists $\varepsilon > 0$ such that $B(a;\varepsilon) \subset O_{i_0}$. Now choose $n \in \mathbb{N}$ such that

$$d(a,a_n) < \frac{\varepsilon}{2} \quad \text{and} \quad \frac{1}{2^n} < \frac{\varepsilon}{4}.$$

Recall that $a_n \in V \cap B_n$. Hence for all $x \in V \cap B_n$ it follows that

$$d(x,a) \leq d(x,x_n) + d(x_n,a_n) + d(a_n,a) \leq \frac{1}{2^n} + \frac{1}{2^n} + \frac{\varepsilon}{2} < \varepsilon,$$

which implies that $V \cap B_n \subset O_{i_0}$, a contradiction.

(3) $\Rightarrow$ (1) Assume that every open cover of V has a finite subcover. Let (a_n) be any sequence in V. Define the closed sets F_n by

$$F_n = \operatorname{clos}\{a_n, a_{n+1}, a_{n+2}, \ldots\}.$$

Then the claim is that

$$V \cap \bigcap_{n=1}^{\infty} F_n \neq \emptyset.$$

To show this, assume the opposite, i.e., assume that

$$V \subset X \setminus \bigcap_{n=1}^{\infty} F_n = \bigcup_{n=1}^{\infty} (X \setminus F_n).$$

By assumption there is a finite collection which covers V:

$$V \subset \bigcup_{i=1}^{k} (X \setminus F_{n_i}) = X \setminus \bigcap_{i=1}^{k} F_{n_i}.$$

Observe for $n > \max\{n_1,\ldots,n_k\}$ that $F_n \subset \bigcap_{i=1}^k F_{n_i}$ which leads to $V \cap F_n = \emptyset$, which is a contradiction.

Hence there exists an element such that $a \in V$ and such that $a \in \bigcap_{n=1}^{\infty} F_n$. As a consequence one sees that for all $n \in \mathbb{N}$:

$$a \in F_n = \operatorname{clos}\{a_n, a_{n+1}, a_{n+2}, \ldots\ldots\}.$$

Thus for each $k \in \mathbb{N}$ there exists $i_k > k$ with $d(a,a_{i_k}) < 1/k$. This shows that (a_n) has a convergent subsequence which converges to $a \in V$. ☺

Corollary A.8. Let X be a complete metric space. If $V \subset X$ is totally bounded then V is relatively compact.

Proof. Note that $\overline{V}$ is closed, so that it is complete. Observe that $\overline{V}$ is totally bounded since every ε-net for V is a 2ε-net for $\overline{V}$. ☺

An interesting question is how to characterize relatively compact or compact subsets of a normed linear space or of a Banach space. Here only a special, but useful, situation is treated. Recall that if F is a metric space with metric d, a subset $V \subset \mathcal{C}(F,\mathbb{K})$ is called *equicontinuous* if for all $\varepsilon > 0$ there exists $\delta > 0$ so that for all $x, y \in F$ and for all $f \in V$

$$d(x,y) < \delta \quad \Rightarrow \quad |f(x) - f(y)| < \varepsilon.$$

Theorem A.9 (Arzela-Ascoli)**.** Let F be a compact metric space and let V be a subset of the Banach space $\mathcal{C}(F,\mathbb{K})$. Then V is relatively compact if and only if

1. V is bounded;
2. V is equicontinuous.

Moreover, V is compact if and only if

3. V is bounded;
4. V is equicontinuous;
5. V is closed.

Proof. ($\Rightarrow$) Since V is relatively compact, V is totally bounded and, in particular, bounded. Choose $\varepsilon > 0$, then there is a finite $\varepsilon/3$-net for V made up of $f_1, \dots, f_n \in V$. Since F is compact, each f_k is uniformly continuous, so that there exists $\delta_k > 0$ so that $d(x,y) < \delta_k$ implies that $|f_k(x) - f_k(y)| < \varepsilon/3$. Define $\delta = \min(\delta_1, \dots, \delta_n)$, then $\delta > 0$. Let $f \in V$, then there is f_k with $\|f - f_k\| < \varepsilon/3$, and for $d(x,y) < \delta$ one obtains

$$|f(x) - f(y)| \leq |f(x) - f_k(x)| + |f_k(x) - f_k(y)| + |f_k(y) - f(y)| < \varepsilon.$$

Hence V is equicontinuous.

($\Leftarrow$) The space F is assumed to be compact and therefore F is separable. Hence there exists a dense and countable set $E = \{x_1, x_2, \dots\}$. Now let (f_n) be a (bounded) sequence in V. Since (f_n) is bounded there exists K such that $|f_n(x)| \leq K$ for all $n \in \mathbb{N}$ and all $x \in F$. The selection of the convergent subsequence of (f_n) will be done in terms of the elements of E.

Since the sequence $(f_n(x_1))$ is bounded in $\mathbb{K}$, the Bolzano-Weierstrass theorem implies the existence of a subsequence (f_n^1) such that $(f_n^1(x_1))$ converges. Now consider $(f_n^1(x_2))$ which is bounded in $\mathbb{K}$. Hence there is a subsequence (f_n^2) of (f_n^1) such that $(f_n^2(x_2))$ converges. Thus one gets the array of sequences

$$\begin{array}{cccc} f_1^1, & f_2^1, & f_3^1, & \dots \\ f_1^2, & f_2^2, & f_3^2, & \dots \\ f_1^3, & f_2^3, & f_3^3, & \dots \\ \vdots & \vdots & \vdots & \ddots \end{array}$$

where each sequence is a subsequence of the one above it, while for each $i \in \mathbb{N}$ the sequence

$$f_1^i(x_i), f_2^i(x_i), f_3^i(x_i), \dots$$

converges. Now define the diagonal sequence $g_1, g_2, g_3, \dots$, where

$$g_i = f_i^i.$$

It is clear that for each x_i the sequence $g_1(x_i), g_2(x_i), g_3(x_i), \dots$ converges, as this sequence is apart from the first $i-1$ entries a subsequence of $f_n^i(x_i)$, $n \in \mathbb{N}$. It remains to show that the sequence (g_i) converges in $\mathcal{C}(F,\mathbb{K})$. Since $\mathcal{C}(F,\mathbb{K})$ is complete it suffices show that (g_i) is a Cauchy sequence.

Let $\varepsilon > 0$. Since V is equicontinous there exists $\delta > 0$ such that $d(x,y) < \delta$ implies $|h(x) - h(y)| < \varepsilon/3$ for all $h \in V$. The elements in $E = \{x_1, x_2, \dots\}$ form a dense set, hence the balls $B(x_i, \delta)$, $i \in \mathbb{N}$, form an open cover of F. Since F is compact, there is a finite subcover

$$F = \bigcup_{i=1}^{r} B(x_i;\delta).$$

There exists $N \in \mathbb{N}$ such that $n,m \geq N$ implies

$$|g_n(x_i) - g_m(x_i)| < \varepsilon/3, \quad i = 1,2,\dots,r.$$

Let $x \in F$ then there is x_i for some $i = 1,2,\dots,r$, such that $d(x,x_i) < \delta$ and hence for $n,m \geq N$ one has

$$|g_n(x) - g_m(x)| \leq |g_n(x) - g_n(x_i)| + |g_n(x_i) - g_m(x_i)| + |g_m(x_i) - g_m(x)| < \varepsilon,$$

which implies that (g_n) is a Cauchy sequence in $\mathcal{C}(F,\mathbb{K})$. ☺

Appendix B

Dirichlet and Fejér kernels

What follows is an elementary treatment of rewriting the partial sums of a Fourier series and the corresponding Cesàro means. These rewritings lead to the classical Dirichlet and Fejér kernels. The Fejér kernel will play a role in Appendix C. It is briefly indicated how these kernels can be used to show the pointwise convergence of a Fourier series under additional conditions.

Let $f \in L^1(-\pi,\pi)$ and define the coefficients

$$a_n = \frac{1}{\pi}\int_{-\pi}^{\pi} f(t)\cos nt\,dt, \quad b_n = \frac{1}{\pi}\int_{-\pi}^{\pi} f(t)\sin nt\,dt. \tag{B.1}$$

These coefficients are well-defined since the functions $\cos nt$ and $\sin nt$ are bounded.

Lemma B.1. If $f \in L^1(a,b)$, then $a_n \to 0$ and $b_n \to 0$.

Proof. The assertion is clearly true for characteristic functions of compact intervals of $[-\pi,\pi]$; and thus also for any step function. Since the step functions are dense in $L^1(a,b)$ the general result follows. To see this, let $\varepsilon > 0$ and let φ be a step function such that $\int_{-\pi}^{\pi} |f(t)-\varphi(t)|\,dt < \varepsilon$, then for instance for b_n one has:

$$\begin{aligned}\left|\int_{-\pi}^{\pi} f(t)\sin nt\,dt\right| &\le \left|\int_{-\pi}^{\pi} (f(t)-\varphi(t))\sin nt\,dt\right| + \left|\int_{-\pi}^{\pi} \varphi(t)\sin nt\,dt\right| \\ &\le \int_{-\pi}^{\pi} |f(t)-\varphi(t)|\,dt + \left|\int_{-\pi}^{\pi} \varphi(t)\sin nt\,dt\right|.\end{aligned}$$

The first term in the right-hand side is less than ε and the second term can be made small by taking n large. ☺

With the coefficients in (B.1) introduce the following finite sums for $n \in \mathbb{N} \cup \{0\}$:

$$S_n f(x) = \frac{a_0}{2} + \sum_{k=1}^{n} (a_k \cos kx + b_k \sin kx), \quad x \in [-\pi,\pi]. \tag{B.2}$$

Lemma B.2. If $f \in L^1(-\pi,\pi)$, then for all $n \in \mathbb{N} \cup \{0\}$:

$$S_n f(x) = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x+t)D_n(t)\,dt,$$

where the Dirichlet kernel D_n is defined by

$$D_n(t) = \frac{\sin(n+\frac{1}{2})t}{2\sin\frac{1}{2}t}.$$

Proof. Inserting the coefficients from (B.1) into the sum (B.2) and using the periodicity gives

$$\begin{aligned} S_n f(x) &= \frac{1}{\pi}\int_{-\pi}^{\pi} f(t)\left[\frac{1}{2} + \sum_{k=1}^{n} \cos k(x-t)\right] dt \\ &= \frac{1}{\pi}\int_{-\pi}^{\pi} f(x+t)\left[\frac{1}{2} + \sum_{k=1}^{n} \cos kt\right] dt. \end{aligned}$$

The claim is now that $\frac{1}{2} + \sum_{k=1}^{n} \cos kt = D_n(t)$. This is best seen by induction. For $n = 0$ the identity is trivial and assuming the identity for $n-1$ gives

$$\frac{\sin(n-\frac{1}{2})t}{2\sin\frac{1}{2}t} + \cos nt = \frac{1}{2\sin\frac{1}{2}t}\left[\sin(n-\tfrac{1}{2})t + 2\cos nt \sin\tfrac{1}{2}t\right] = \frac{\sin(n+\frac{1}{2})t}{2\sin\frac{1}{2}t},$$

where it has been used that $\sin(\alpha \pm \beta) = \sin\alpha\cos\beta \pm \cos\alpha\sin\beta$. This completes the proof. ☺

It is easy to see that S_n has the nice property: $S_n\mathbf{1} = \mathbf{1}$. Note that $D_n(0) = n + \frac{1}{2}$ and that D_n changes sign rapidly as n grows.

In addition one may also introduce the Cesàro means C_n of S_n, $n \in \mathbb{N}$:

$$C_n f(x) = \frac{1}{n}\left[S_0 f(x) + S_1 f(x) + \cdots + S_{n-1} f(x)\right], \quad x \in [-\pi, \pi].$$

Lemma B.3. If $f \in L^1(-\pi,\pi)$, then for all $n \in \mathbb{N}$

$$C_n f(x) = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x+t) F_n(t)\, dt,$$

where the Fejér kernel F_n is defined by

$$F_n(t) = \frac{1}{2n}\left(\frac{\sin\frac{1}{2}nt}{\sin\frac{1}{2}t}\right)^2.$$

Proof. By means of Lemma B.2 one obtains for $n \in \mathbb{N}$:

$$\begin{aligned} C_n f(x) &= \frac{1}{n}\sum_{k=0}^{n-1} \frac{1}{\pi}\int_{-\pi}^{\pi} f(x+t)\frac{\sin(k+\frac{1}{2})t}{2\sin\frac{1}{2}t}\, dt \\ &= \frac{1}{2n\pi}\int_{-\pi}^{\pi} f(x+t)\left[\sum_{k=0}^{n-1} \frac{\sin(k+\frac{1}{2})t}{\sin\frac{1}{2}t}\right] dt. \end{aligned}$$

The claim is now that

$$\sum_{k=0}^{n-1} \frac{\sin(k+\frac{1}{2})t}{\sin\frac{1}{2}t} = \left(\frac{\sin\frac{1}{2}nt}{\sin\frac{1}{2}t}\right)^2,$$

or, due to $2\sin^2\alpha = 1 - \cos 2\alpha$, equivalently,

$$\sum_{k=0}^{n-1} \sin(k+\tfrac{1}{2})t = \frac{1-\cos nt}{2\sin\frac{1}{2}t}.$$

The last identity will be shown by induction. It is clear for $n=1$ and next observe that assuming the identity holds for n gives

$$\begin{aligned}
\sum_{k=1}^{n} \sin(k+\tfrac{1}{2})t &= \frac{1-\cos nt}{2\sin\frac{1}{2}t} + \sin(n+\tfrac{1}{2})t \\
&= \frac{1}{2\sin\frac{1}{2}t}\left[1-\cos nt + 2\sin(n+\tfrac{1}{2})t\sin\tfrac{1}{2}t\right] \\
&= \frac{1}{2\sin\frac{1}{2}t}\left[1-\cos nt+\cos nt-\cos(n+1)t\right] = \frac{1-\cos(n+1)t}{2\sin\frac{1}{2}t},
\end{aligned}$$

where it has been used that $2\sin\alpha\sin\beta = \cos(\alpha-\beta)-\cos(\alpha+\beta)$. This completes the proof. ☺

Note that $C_n\mathbf{1}=\mathbf{1}$ and now the function F_n is nonnegative.

An important question is what can be said about the *pointwise* convergence of the following trigonometric series

$$\frac{a_0}{2} + \sum_{k=1}^{\infty}(a_k\cos kx + b_k\sin kx), \quad -\pi \le x \le \pi, \tag{B.3}$$

when $f \in L^1(-\pi,\pi)$ and the coefficients are given by (B.1). In general the series (B.3) does not necessarily converge pointwise, even when f is continuous; see Example 5.29. However, from Lemma B.2 one obtains the following statement.

Lemma B.4. Assume that $f \in \mathcal{C}([-\pi,\pi],\mathbb{C})$ is periodic. If f is differentiable at x, then the Fourier series in (B.3) converges to $f(x)$.

Proof. Due to $S_n\mathbf{1}=\mathbf{1}$ one sees that

$$\begin{aligned}
(S_nf)(x)-f(x) &= \frac{1}{\pi}\int_{-\pi}^{\pi}[f(x+t)-f(x)]\frac{\sin(n+\frac{1}{2})t}{2\sin\frac{1}{2}t}\,dt \\
&= \frac{1}{\pi}\int_{-\pi}^{\pi} g(t)\sin(n+\tfrac{1}{2})t\,dt,
\end{aligned}$$

where the continuous function g is defined by

$$g(t) = \begin{cases} \dfrac{f(x+t)-f(x)}{t}\,\dfrac{t}{2\sin\frac{1}{2}t} & \text{if } t \neq 0, \\ f'(x) & \text{if } t = 0. \end{cases}$$

The result now follows from Lemma B.1. ☺

More generally, Lemma B.2 also implies the following practical result. A function $f:[-\pi,\pi]\to\mathbb{R}$ is said to be *piecewise smooth* if there exists a partition $-\pi = x_0 < x_1 < \cdots < x_n = \pi$, such that the function f is continuously differentiable on each subinterval $(c,d)=(x_{i-1},x_i)$, and the limits

$$f(c+) = \lim_{x\downarrow c} f(x), \quad f(d-) = \lim_{x\uparrow d} f(x),$$

and

$$f'(c+) = \lim_{x\downarrow c} f'(x), \quad f'(d-) = \lim_{x\uparrow d} f'(x),$$

exist. If the 2π-periodic function f is piecewise smooth, then the Fourier series (B.3) converges pointwise and the sum in $x \in [-\pi, \pi]$ is equal to

$$\frac{f(x+) + f(x-)}{2}.$$

Moreover in this case the series (B.3) converges uniformly on each interval where f is continuous.

A similar result is obtained via Cesàro means: the Cesàro means of a continuous function which is periodic converges uniformly to that function; see Appendix C.

Finally the following observation may sometimes be helpful. Assume that the sequences (a_n) and (b_n) are given and that they satisfy

$$\sum_{k=1}^{\infty} |a_k| < \infty, \quad \sum_{k=1}^{\infty} |b_k| < \infty. \tag{B.4}$$

Then the series in (B.3) clearly converges uniformly on $[-\pi, \pi]$ and defines a continuous 2π-periodic function $f : [-\pi, \pi] \to \mathbb{R}$. Moreover, due to the trigonometric identities

$$\int_{-\pi}^{\pi} \cos mt \, \sin nt \, dt = 0, \quad \int_{-\pi}^{\pi} \cos mt \, \cos nt \, dt = \int_{-\pi}^{\pi} \sin mt \, \sin nt \, dt = \pi \delta_{mn},$$

it is clear that a_m and b_m can be expressed in terms of f by (B.1).

Appendix C

The Weierstrass approximation theorems

The Weierstrass approximation theorem states that continuous functions on a compact interval can be uniformly approximated by polynomials. Moreover, there is a version for periodic continuous functions which are uniformly approximated by trigonometric polynomials. For the convenience of the reader these classical results are proved in detail by means of Korovkin's theorems. A more general, abstract, treatment can be found in the next section.

In what follows, (X,d) is a metric space and the space $\mathcal{C}(X,\mathbb{R})$ is provided with the norm $\|f\|_\infty = \sup_{x\in X}|f(x)|$. Recall that convergence with respect to this norm is equivalent to uniform convergence. In addition, $\mathbf{1}: X \to \mathbb{R}$ denotes the function such that $\mathbf{1}(x) = 1$ for all $x \in X$.

Definition C.1. A linear operator $T: \mathcal{C}(X,\mathbb{R}) \to \mathcal{C}(X,\mathbb{R})$ is said to be *positive*, i.e. $T \geq 0$, if $f \geq 0$ implies $Tf \geq 0$ for all $f \in \mathcal{C}(X,\mathbb{R})$.

The notion of a positive operator in the present context should not be confused with the concept of a non-negative operator on a Hilbert space. Note that a positive operator preserves inequalities, i.e., $f \leq g$ implies $Tf \leq Tg$ for all $f,g \in \mathcal{C}(X,\mathbb{R})$. In addition, a positive operator $T: \mathcal{C}(X,\mathbb{R}) \to \mathcal{C}(X,\mathbb{R})$ is bounded. To see this, note that

$$-\|f\|_\infty \mathbf{1}(x) \leq f(x) \leq \|f\|_\infty \mathbf{1}(x), \quad f \in \mathcal{C}(X,\mathbb{R}), \quad x \in X.$$

Since T preserves this inequality, it follows that

$$\|Tf\|_\infty \leq \|T\mathbf{1}\|_\infty \|f\|_\infty, \quad f \in \mathcal{C}(X,\mathbb{R}).$$

Theorem C.2 (Korovkin's first theorem)**.** Let (X,d) be a compact metric space, and define for each $x \in X$ the function

$$\psi_x(y) = d(x,y)^2, \quad y \in X.$$

Assume that $T_n: \mathcal{C}(X,\mathbb{R}) \to \mathcal{C}(X,\mathbb{R})$ is a sequence of linear operators that satisfies the following three properties:

1. $T_n \geq 0$ for each $n \in \mathbb{N}$;
2. $\|T_n\mathbf{1} - \mathbf{1}\|_\infty \to 0$;
3. $\sup_{x\in X}|T_n\psi_x(x)| \to 0$.

Then $\|T_n f - f\|_\infty \to 0$ for each $f \in \mathcal{C}(X,\mathbb{R})$.

Proof. Let $\varepsilon > 0$ be arbitrary. Since f is uniformly continuous on X, there exists $\delta > 0$ such that

$|f(x) - f(y)| < \varepsilon$ whenever $d(x,y) < \delta$. If $d(x,y) \geq \delta$, then

$$|f(x) - f(y)| \leq 2\|f\|_\infty \leq \frac{2\|f\|_\infty}{\delta^2} d(x,y)^2.$$

Hence, for all $x, y \in X$ we obtain

$$|f(x) - f(y)| \leq \frac{2\|f\|_\infty}{\delta^2} \psi_x(y) + \varepsilon \mathbf{1}(y),$$

or, equivalently,

$$-\frac{2\|f\|_\infty}{\delta^2} \psi_x(y) - \varepsilon \mathbf{1}(y) \leq f(y) - f(x)\mathbf{1}(y) \leq \frac{2\|f\|_\infty}{\delta^2} \psi_x(y) + \varepsilon \mathbf{1}(y).$$

If this is considered as an inequality of functions in the variable y where $x \in X$ is fixed, then the positivity of the operator T_n gives for each $n \in \mathbb{N}$ the following inequality:

$$|T_n f(y) - f(x) T_n \mathbf{1}(y)| \leq \frac{2\|f\|_\infty}{\delta^2} T_n \psi_x(y) + \varepsilon T_n \mathbf{1}(y).$$

In particular, for $y = x$, the triangle inequality gives

$$|T_n f(x) - f(x)| \leq |T_n f(x) - f(x) T_n \mathbf{1}(x)| + |f(x) T_n \mathbf{1}(x) - f(x)|,$$

which implies that

$$\|T_n f - f\|_\infty \leq \frac{2\|f\|_\infty}{\delta^2} \sup_{x \in X} T_n \psi_x(x) + \varepsilon + \|f\|_\infty \|T_n \mathbf{1} - \mathbf{1}\|_\infty.$$

The proof is completed by noting that the right-hand side can be made arbitrarily small for n sufficiently large. ☺

In concrete applications, the conditions (2) and (3) are replaced by conditions which are easy to verify.

Theorem C.3 (Bohman's theorem)**.** Let $T_n : \mathcal{C}([0,1],\mathbb{R}) \to \mathcal{C}([0,1],\mathbb{R})$ be a sequence of linear operators that satisfies the following two properties:

1. $T_n \geq 0$ for each $n \in \mathbb{N}$;
2. $\|T_n f - f\|_\infty \to 0$ for each $f \in \{1, x, x^2\}$.

Then $\|T_n f - f\|_\infty \to 0$ for each $f \in \mathcal{C}([0,1],\mathbb{R})$.

Proof. With the functions $f_i \in \mathcal{C}([0,1],\mathbb{R})$, $i = 0,1,2$, defined by

$$f_0(x) = 1, \quad f_1(x) = x, \quad f_2(x) = x^2,$$

it follows that

$$\psi_x(y) = d(x,y)^2 = |x - y|^2 = x^2 f_0(y) - 2x f_1(y) + f_2(y),$$

so that clearly

$$x^2 f_0(x) - 2x f_1(x) + f_2(x) = 0.$$

For fixed $x \in [0,1]$ this gives

$$T_n \psi_x(y) = x^2 T_n f_0(y) - 2x T_n f_1(y) + T_n f_2(y).$$

In particular, setting $y = x$ gives

$$T_n \psi_x(x) = x^2 (T_n f_0 - f_0)(x) - 2x (T_n f_1 - f_1)(x) + (T_n f_2 - f_2)(x).$$

Therefore,

$$\sup_{x\in[0,1]} |T_n\psi_x(x)| \le \|T_nf_0 - f_0\|_\infty + 2\|T_nf_1 - f_1\|_\infty + \|T_nf_2 - f_2\|_\infty \to 0.$$

The result now follows from Korovkin's first theorem. ☺

It is now shown how the Weierstrass approximation theorem follows from Theorem C.3. For this purpose, introduce the operators

$$B_n : \mathcal{C}([0,1],\mathbb{R}) \to \mathcal{C}([0,1],\mathbb{R}), \quad B_nf(x) = \sum_{k=0}^{n} f\left(\frac{k}{n}\right)\binom{n}{k} x^k(1-x)^{n-k},$$

which are clearly linear and positive. The function B_nf is a polynomial of degree at most n and referred to as the n-th Bernstein polynomial of $f \in \mathcal{C}([0,1],\mathbb{R})$. In particular, the Bernstein polynomials of the functions

$$f_0(x) = 1, \quad f_1(x) = x, \quad f_2(x) = x^2,$$

are given by

$$B_nf_0(x) = 1, \quad B_nf_1(x) = x, \quad B_nf_2(x) = x^2 + \frac{x-x^2}{n},$$

respectively. Clearly, $\|B_nf_i - f_i\|_\infty \to 0$ for each $i = 0,1,2$. The next Corollary now follows from Bohman's theorem.

Corollary C.4 (Weierstrass approximation theorem)**.** Let $f \in \mathcal{C}([a,b],\mathbb{R})$. For every $\varepsilon > 0$ there exists a polynomial P such that

$$\|P - f\|_\infty < \varepsilon.$$

We define the space of continuous 2π-periodic functions by

$$\mathcal{C}_{\text{per}}([-\pi,\pi],\mathbb{R}) = \{f \in \mathcal{C}([-\pi,\pi],\mathbb{R}) \,:\, f(-\pi) = f(\pi)\}.$$

The following theorem follows from Korovkin's first theorem by considering 2π-periodic functions as functions on a circle.

Theorem C.5 (Korovkin's second theorem)**.** Let the sequence of linear operators

$$T_n : \mathcal{C}_{\text{per}}([-\pi,\pi],\mathbb{R}) \to \mathcal{C}_{\text{per}}([-\pi,\pi],\mathbb{R})$$

satisfy the following two properties:

1. $T_n \ge 0$ for each $n \in \mathbb{N}$;
2. $\|T_nf - f\|_\infty \to 0$ for each $f \in \{1, \cos t, \sin t\}$.

Then $\|T_nf - f\|_\infty \to 0$ for each $f \in \mathcal{C}_{\text{per}}([-\pi,\pi],\mathbb{R})$.

Proof. The set $X = \{x = (x_1,x_2) \in \mathbb{R}^2 \,:\, x_1^2 + x_2^2 = 1\}$ provided with the metric inherited from $\mathbb{R}^2$ is a compact metric space. For any $f \in \mathcal{C}_{\text{per}}([-\pi,\pi],\mathbb{R})$ we define $\widehat{f} \in \mathcal{C}(X,\mathbb{R})$ by

$$\widehat{f}(\cos t, \sin t) = f(t).$$

Note that $\|\widehat{f}\|_\infty = \|f\|_\infty$ so that the map $f \mapsto \widehat{f}$ is an isometric isomorphism between the spaces $\mathcal{C}_{\text{per}}([-\pi,\pi],\mathbb{R})$ and $\mathcal{C}(X,\mathbb{R})$. Clearly, the operators

$$\widehat{T}_n : \mathcal{C}(X,\mathbb{R}) \to \mathcal{C}(X,\mathbb{R}), \quad \widehat{T}_n\widehat{f} = \widehat{T_nf},$$

are linear and positive.

Define the following functions in $\mathcal{C}_{\text{per}}([-\pi,\pi],\mathbb{R})$:

$$f_0(t)=1, \quad f_1(t)=\cos t, \quad f_2(t)=\sin t.$$

Clearly,

$$\|\widehat{T}_n\widehat{f}_0-\widehat{f}_0\|_\infty=\|T_nf_0-f_0\|_\infty\to 0.$$

For all $x=(\cos t,\sin t)\in X$ and $y=(\cos s,\sin s)\in X$ one obtains

$$\begin{aligned}\widehat{\psi}_x(y)&=d(x,y)^2\\&=(\cos t-\cos s)^2+(\sin t-\sin s)^2\\&=2-2\cos t\cos s-2\sin t\sin s\\&=2\widehat{f}_0(y)-2\cos t\,\widehat{f}_1(y)-2\sin t\,\widehat{f}_2(y),\end{aligned}$$

so that clearly,

$$2f_0(t)-2\cos t\,f_1(t)-2\sin t\,f_2(t)=0.$$

For fixed $x\in X$ it follows that

$$\begin{aligned}\widehat{T}_n\widehat{\psi}_x(y)&=2\widehat{T}_n\widehat{f}_0(y)-2\cos t\,\widehat{T}_n\widehat{f}_1(y)-2\sin t\,\widehat{T}_n\widehat{f}_2(y)\\&=2T_nf_0(s)-2\cos t\,T_nf_1(s)-2\sin t\,T_nf_2(s).\end{aligned}$$

In particular, setting $s=t$, so that $y=x$, gives

$$\widehat{T}_n\widehat{\psi}_x(x)=2(T_nf_0-f_0)(t)-2\cos t\,(T_nf_1-f_1)(t)-2\sin t\,(T_nf_2-f_2)(t),$$

which implies that

$$\sup_{x\in X}|\widehat{T}_n\widehat{\psi}_x(x)|\le 2(\|T_nf_0-f_0\|_\infty+\|T_nf_1-f_1\|_\infty+\|T_nf_2-f_2\|_\infty)\to 0.$$

For any $f\in\mathcal{C}_{\text{per}}([-\pi,\pi],\mathbb{R})$ Korovkin's first theorem gives

$$\|T_nf-f\|_\infty=\|\widehat{T}_n\widehat{f}-\widehat{f}\|_\infty\to 0,$$

which completes the proof. ☺

Next is a proof of the Weierstrass approximation theorem for periodic functions. Let $f\in\mathcal{C}_{\text{per}}([-\pi,\pi],\mathbb{R})$ and recall from Appendix B the Cesàro means of the Fourier partial sums:

$$C_n:\mathcal{C}_{\text{per}}([-\pi,\pi],\mathbb{R})\to\mathcal{C}_{\text{per}}([-\pi,\pi],\mathbb{R}), \quad C_nf(t)=\frac{1}{\pi}\int_{-\pi}^{\pi}f(t+s)F_n(s)\,ds,$$

where the Fejér kernel F_n is defined by

$$F_n(s)=\frac{1}{2n}\left(\frac{\sin\frac{1}{2}ns}{\sin\frac{1}{2}s}\right)^2.$$

Clearly, $C_n\ge 0$ for each $n\in\mathbb{N}$. For the functions

$$f_0(t)=1, \quad f_1(t)=\cos t, \quad f_2(t)=\sin t,$$

it follows that

$$C_nf_0(t)=1, \quad C_nf_1(t)=\frac{n-1}{n}\cos t, \quad C_nf_2(t)=\frac{n-1}{n}\sin t,$$

which implies that $\|C_nf_i-f_i\|_\infty\to 0$ for each $i=0,1,2$. The next Corollary now follows from Korovkin's second theorem.

Corollary C.6 (Weierstrass approximation theorem; trigonometric version)**.** For any function $f \in \mathcal{C}_{\text{per}}([-\pi,\pi],\mathbb{R})$ and every $\varepsilon > 0$ there exists a trigonometric polynomial P such that

$$\|P - f\|_\infty < \varepsilon.$$

Appendix D

The Stone-Weierstrass theorem

The original Weierstrass theorem can be put in a more abstract context giving the following Stone-Weierstrass theorem. This theorem is about certain subalgebras of the Banach algebra $\mathcal{C}(X,\mathbb{K})$, i.e., linear subspaces of $\mathcal{C}(X,\mathbb{K})$ which are closed under pointwise multiplication, which are dense in $\mathcal{C}(X,\mathbb{K})$[1]. The presentation in this section is independent of the presentation in the foregoing section. It all depends on a very specific classical series expansion, which will be proved in detail for the convenience of the reader.

Lemma D.1. The series expansion

$$\sqrt{1-t} = 1 - \sum_{n=1}^{\infty} a_n t^n, \tag{D.1}$$

where

$$a_n = 2^{1-2n} \frac{(2n-2)!}{n!(n-1)!}, \quad n \in \mathbb{N}, \tag{D.2}$$

holds for $t \in (-1,1)$. In fact, the convergence is uniform on $[-1,1]$ and the equality holds on $[-1,1]$.

Proof. First remember the following equality:

$$\sqrt{1-t} = \sum_{n=0}^{\infty} c_n t^n, \quad c_n = (-1)^n \binom{\frac{1}{2}}{n}, \quad t \in (-1,1). \tag{D.3}$$

Comparing (D.1) with (D.3) shows the identification

$$a_n = -c_n = (-1)^{n+1} \binom{\frac{1}{2}}{n} = 2^{1-2n} \frac{(2n-2)!}{n!(n-1)!}, \quad n \in \mathbb{N}. \tag{D.4}$$

Although (D.3) is well-known here is a quick proof of (D.3). Observe that (D.4) implies that

$$\frac{c_{n+1}}{c_n} = \frac{a_{n+1}}{a_n} = \frac{2n-1}{2(n+1)}. \tag{D.5}$$

In particular the series in (D.3) is convergent for $t \in (-1,1)$ by the test of d'Alembert. Denote the right-hand side in (D.3) by $\varphi(t)$. By termwise differentiation one obtains

$$\varphi'(t) = \sum_{n=1}^{\infty} n c_n t^{n-1}, \quad t \in (-1,1).$$

[1] For the Stone-Weierstrass theorem see the exposition by M.H. Stone, "A generalized Weierstrass approximation theorem", in R.C. Buck (ed.), Studies in Modern Analysis 1, Math. Assoc. Amer. (1962), 30–87.

A straightforward calculation involving (D.5) shows that

$$\varphi(t) = -2(1-t)\varphi'(t), \quad t \in (-1,1).$$

Integration of this differential equation leads to

$$\varphi(t) = c\sqrt{1-t}, \quad t \in (-1,1).$$

By definition $\varphi(0) = 1$ and it follows that $c = 1$. Therefore

$$\varphi(t) = \sqrt{1-t}, \quad t \in (-1,1),$$

which gives the identity in (D.3) and thus in (D.1) for $t \in (-1,1)$.

Before looking at the convergence of the right-hand side of (D.1) in the endpoints briefly review the following simple estimates. Recall that $1 - x \le e^{-x}$, $x \in \mathbb{R}$, so that

$$\begin{aligned}(-1)^n \binom{-\frac{1}{2}}{n} &= \left(1 - \frac{\frac{1}{2}}{n}\right)\left(1 - \frac{\frac{1}{2}}{n-1}\right)\cdots\left(1 - \frac{\frac{1}{2}}{2}\right)\left(1 - \frac{\frac{1}{2}}{1}\right) \\ &\le e^{-1/2(1+1/2+\cdots+1/n)} \le e^{-1/2(1+\log n)} \le \frac{1}{n^{1/2}},\end{aligned}$$

which gives $\binom{-\frac{1}{2}}{n} \to 0$. A similar reasoning shows that $\binom{\frac{1}{2}}{n} \to 0$.

To see that the series in (D.1) or, equivalently, the series in (D.3) converges for $t = 1$ it suffices to observe that

$$\sum_{k=0}^{n} c_k = \sum_{k=0}^{n} (-1)^k \binom{\frac{1}{2}}{k} = (-1)^n \binom{-\frac{1}{2}}{n} \to 0.$$

Note that the above identity can be checked via induction. To see that the series in (D.1) converges for $t = -1$ it suffices to observe that all $a_n > 0$ and that $a_{n+1} < a_n$, cf. (D.4) and (D.5), respectively. Thus the series in (D.1) for $t = -1$ is an alternating series with $a_n = (-1)^{n+1}\binom{\frac{1}{2}}{n} \to 0$. Therefore the Leibniz criterium about series with alternating terms guarantees that the series in (D.1) converges at $t = -1$. An application of Abel's theorem shows that there is uniform convergence on $[-1,1]$ in the right-hand side of (D.1) and equality on $[-1,1]$ in (D.1). Thus the assertion in the lemma holds. ☺

Theorem D.2 (Stone-Weierstrass, real case)**.** Let X be a compact metric space and let $\mathcal{A}$ be subspace of the Banach algebra $\mathcal{C}(X,\mathbb{R})$ such that:

1. $\mathcal{A}$ is an algebra over $\mathbb{R}$;
2. the constant function $\mathbf{1}$ belongs to $\mathcal{A}$;
3. $\mathcal{A}$ separates points, i.e., for all $x, y \in X$ with $x \neq y$ there is a function $g \in \mathcal{A}$ such that $g(x) \neq g(y)$.

Then $\mathcal{A}$ is dense in $\mathcal{C}(X,\mathbb{R})$.

Proof. The proof is given in a number of steps.

Step 1. Since $\mathcal{A}$ is an algebra by (1) it is clear that the closure $\overline{\mathcal{A}}$ of $\mathcal{A}$ is an algebra and that $\overline{\mathcal{A}} \subset \mathcal{C}(X,\mathbb{R})$. The aim is to show that $\overline{\mathcal{A}} = \mathcal{C}(X,\mathbb{R})$.

Step 2. First it will be shown that $h \in \overline{\mathcal{A}}$ with $h(x) \ge 0$, $x \in X$, implies that $\sqrt{h} \in \overline{\mathcal{A}}$. Without loss of generality one may assume that $0 \le h(x) \le 1$, $x \in X$. Then it is clear that

$$k(x) = 1 - h(x) \in \overline{\mathcal{A}} \quad \text{and} \quad 0 \le k(x) \le 1,$$

and therefore one has

$$\sqrt{h(x)} = \sqrt{1-k(x)} = 1 - \sum_{n=1}^{\infty} a_n k^n(x), \quad x \in X,$$

cf. Lemma D.1. Moreover, one sees that

$$\begin{aligned}\left\| \sqrt{h} - \left(1 - \sum_{n=1}^{N} a_n k^n\right) \right\| &= \sup_{x \in X} \left| \sqrt{h(x)} - \left(1 - \sum_{n=1}^{N} a_n k^n(x)\right) \right| \\ &\leq \sup_{t \in [0,1]} \left| \sqrt{1-t} - \left(1 - \sum_{n=1}^{N} a_n t^n\right) \right|,\end{aligned}$$

and the right-hand side goes to 0 as $N \to \infty$ by Lemma D.1. This shows that $\sqrt{h}$ can be approximated by elements of the form $1 - \sum_{n=1}^{N} a_n k^n \in \overline{\mathcal{A}}$, which implies that $\sqrt{h} \in \overline{\mathcal{A}}$.

Next it will be shown for any $h \in \overline{\mathcal{A}}$ one has $|h| \in \overline{\mathcal{A}}$. All one has to observe is that $h \in \overline{\mathcal{A}}$ implies that $h^2 \in \overline{\mathcal{A}}$, and consequenly $|h| = \sqrt{h^2} \in \overline{\mathcal{A}}$.

Step 3. If $h_1, \dots, h_n \in \overline{\mathcal{A}}$, then

$$\max\{h_1, \dots, h_n\} \in \overline{\mathcal{A}}, \quad \min\{h_1, \dots, h_n\} \in \overline{\mathcal{A}}.$$

To see this it suffices to prove the statements for $n = 2$ in which case

$$\max\{h_1, h_2\} = \frac{1}{2}(h_1 + h_2 + |h_1 - h_2|), \quad \min\{h_1, h_2\} = \frac{1}{2}(h_1 + h_2 - |h_1 - h_2|).$$

Due to Step 2 it follows that $\max\{h_1, h_2\}$ and $\min\{h_1, h_2\}$ belong to $\overline{\mathcal{A}}$.

Step 4. Let $f \in \mathcal{C}(X, \mathbb{R})$. Then for each pair $x, y \in X$ there exists a function $\varphi_{x,y} \in \mathcal{A} \subset \overline{\mathcal{A}}$ such that

$$\varphi_{x,y}(x) = f(x), \quad \varphi_{x,y}(y) = f(y). \tag{D.6}$$

To see this assume $x \neq y$ and let $g \in \mathcal{A}$ separate these points: $g(x) \neq g(y)$; here (3) is being used. Define the function $\varphi_{x,y}$ by

$$\varphi_{x,y}(z) = \frac{f(x) - f(y)}{g(x) - g(y)} g(z) + \frac{f(y)g(x) - f(x)g(y)}{g(x) - g(y)}, \quad z \in X.$$

The function $\varphi_{x,y}$ is well-defined, belongs to $\mathcal{A}$, and satisfies (D.6). If $x = y$ define the function $\varphi_{x,y}$ by

$$\varphi_{x,y}(z) = f(z), \quad z \in X.$$

Also in this case the function $\varphi_{x,y}$ is well-defined, belongs to $\mathcal{A}$, and satisfies (D.6).

Step 5. Let $f \in \mathcal{C}(X, \mathbb{R})$ and let $\varphi_{x,y} \in \mathcal{A}$ such that (D.6) holds. Let $\varepsilon > 0$ and fix $x \in X$. For each $y \in X$ there exists an open neighborhood $U(y)$ such that

$$\varphi_{x,y}(z) > f(z) - \varepsilon \quad \text{for all} \quad z \in U(y). \tag{D.7}$$

This is clear since $\varphi_{x,y}$ and f are continuous and $\varphi_{x,y}(y) = f(y)$; cf. Step 4. The compactness of X implies the existence of $y_1, \dots, y_k \in X$ such that

$$X = \bigcup_{j=1}^{k} U(y_j).$$

By means of $y_1, \dots, y_k \in X$ define the function $\varphi_x : X \to \mathbb{R}$ by

$$\varphi_x = \max\{\varphi_{x,y_1}, \varphi_{x,y_2}, \dots, \varphi_{x,y_n}\}.$$

Then by Step 3, Step 4, and (D.7) one sees that $\varphi_x \in \overline{\mathcal{A}}$, $\varphi_x(x) = f(x)$, and

$$\varphi_x(z) > f(z) - \varepsilon \quad \text{for all} \quad z \in X. \tag{D.8}$$

For each $x \in X$ there exists an open neighborhood $V(x)$ of x such that

$$\varphi_x(z) < f(z) + \varepsilon \quad \text{for all} \quad z \in V(x). \tag{D.9}$$

This is clear since φ_x and f are continuous and $\varphi_x(x) = f(x)$; cf. Step 4. The compactness of X implies the existence of $x_1, \dots, x_l \in X$ such that

$$X = \bigcup_{j=1}^{l} V(x_j).$$

Define the function $\varphi : X \to \mathbb{R}$ by

$$\varphi = \min\{\varphi_{x_1}, \varphi_{x_2}, \dots, \varphi_{x_l}\}.$$

Then by Step 3, (D.8), and (D.7) one sees that $\varphi \in \overline{\mathcal{A}}$, $\varphi(x) = f(x)$, and

$$f(z) - \varepsilon < \varphi(z) < f(z) + \varepsilon \quad \text{for all} \quad z \in X.$$

Thus $\|f - \varphi\| \leq \varepsilon$. Since $\overline{\mathcal{A}}$ is closed, it follows that $f \in \overline{\mathcal{A}}$. Hence $\overline{\mathcal{A}} = \mathcal{C}(X, \mathbb{R})$. ☺

Corollary D.3. Let $[a,b] \subset \mathbb{R}$ be a compact interval. Then the algebra of all real polynomials is dense in $\mathcal{C}([a,b], \mathbb{R})$.

Corollary D.4. Let X be a compact subset of $\mathbb{R}^n$. Then the algebra of all real polynomials in the coordinates $x_1, \dots, x_n$ is dense in $\mathcal{C}(X, \mathbb{R})$.

Corollary D.5. The algebra of all real trigonometric polynomials is dense in $\mathcal{C}_{\text{per}}([-\pi, \pi], \mathbb{R})$.

Proof. Clearly the periodic polynomials, to which **1** belongs, form an algebra; just recall the various trigonometric identities. The periodic polynomials separate points. To use Theorem D.2, identify $\mathcal{C}_{\text{per}}([-\pi, \pi], \mathbb{R})$ with $\mathcal{C}(\mathbb{T}, \mathbb{R})$, where $\mathbb{T} = \{(x_1, x_2) \in \mathbb{R}^2 : x_1^2 + x_2^2 = 1\}$ is the unit circle. The mapping

$$t \in [-\pi, \pi] \to (\cos t, \sin t) \in \mathbb{T}$$

gives a correspondence between the trigonometric polynomials with polynomials in (x_1, x_2) on $\mathbb{T}$, which are dense in $\mathcal{C}(\mathbb{T}, \mathbb{R})$ by Corollary D.4. ☺

Theorem D.6 (Stone-Weierstrass, complex case)**.** Let X be a compact metric space and let $\mathcal{A}$ be subspace of the Banach algebra $\mathcal{C}(X, \mathbb{C})$ such that:

1. $\mathcal{A}$ is an algebra over $\mathbb{C}$;
2. the constant function **1** belongs to $\mathcal{A}$;
3. $\mathcal{A}$ separates points, i.e, for all $x, y \in X$ there is a function $g \in \mathcal{A}$ such that $g(x) \neq g(y)$ whenever $x, y \in X$ and $x \neq y$;
4. $f \in \mathcal{A}$ implies $\bar{f} \in \mathcal{A}$.

Then $\mathcal{A}$ is dense in $\mathcal{C}(X, \mathbb{C})$.

Proof. The proof consists of reducing to the real case. Define $\mathcal{A}'$ by

$$\mathcal{A}' = \mathcal{A} \cap \mathcal{C}(X, \mathbb{R}).$$

Then $\mathcal{A}'$ is a real algebra by (1) and the constant function **1** belongs to $\mathcal{A}'$ by (2). Furthermore one sees that for $f \in \mathcal{A}$

$$\operatorname{Re} f = \frac{1}{2}(f + \bar{f}) \in \mathcal{A}' \quad \text{and} \quad \operatorname{Im} f = \frac{1}{2i}(f - \bar{f}) \in \mathcal{A}',$$

here (4) has been used. Now let $x, y \in X$ and $x \neq y$. Then by (3) there is a function $g \in \mathcal{C}(X, \mathbb{R})$ such that $g(x) \neq g(y)$ whenever $x, y \in X$ and $x \neq y$. This implies that either $(\operatorname{Re} g)(x) \neq (\operatorname{Re} g)(y)$ or $(\operatorname{Im})g(x) \neq (\operatorname{Im})g(y)$. Hence $\mathcal{A}'$ separates points of X. Thus by Theorem D.2 the algebra $\mathcal{A}'$ is dense in $\mathcal{C}(X, \mathbb{R})$.

Let $f \in \mathcal{C}(X, \mathbb{C})$ and let $\varepsilon > 0$. Write $f = f_1 + i f_2$, where $f_1, f_2 \in \mathcal{C}(X, \mathbb{R})$. By Theorem D.2 there exist $\varphi_1, \varphi_2 \in \mathcal{A}'$ with

$$\|f_1 - \varphi_1\| < \frac{\varepsilon}{2}, \quad \|f_2 - \varphi_2\| < \frac{\varepsilon}{2}.$$

Then $\varphi = \varphi_1 + i\varphi_2 \in \mathcal{A}$ and $\|f - \varphi\| < \varepsilon$. Thus $f \in \overline{\mathcal{A}}$, which shows $\overline{\mathcal{A}} = \mathcal{C}(X, \mathbb{C})$. ☺

Bibliography

[1] E.W. Cheney, *Analysis for Applied Mathematics*, Springer, 2001.

[2] E.W. Cheney, *Introduction to Approximation Theory*, McGraw Hill, 1966.

[3] P.G. Ciarlet, *Linear and Nonlinear Functional Analysis with Applications*, Society for Industrial and Applied Mathematics, 2013.

[4] J.B. Conway, *A Course in Functional Analysis*, Springer, 1990.

[5] I.C. Gohberg & S. Goldberg, *Basic Operator Theory*, Birkhäuser Verlag, 2001.

[6] W. Kaballo, *Grundkurs Funktionalanalysis*, Spektrum, Akademischer Verlag, 2011.

[7] G. Klambauer, *Real Analysis*, Dover, 2005.

[8] E. Kreyszig, *Introductory Functional Analysis with Applications*, John Wiley, 1978.

[9] P.D. Lax, *Functional Analysis*, John Wiley, 2002.

[10] R.E. Megginson, *An Introduction to Banach Space Theory*, Springer, 1998.

[11] R. Meise & D. Vogt, *Einführung in die Funktionalanalysis*, Vieweg+Teubner, 2011.

[12] A. Pinkus, "Weierstrass and Approximation Theory", Journal of Approximation Theory, 107 (2000), 1–66.

[13] M. Reed & B. Simon, *Methods of Modern Mathematical Physics I: Functional Analysis*, Academic Press, 1980.

[14] M. Reed & B. Simon, *Methods of Modern Mathematical Physics II: Fourier Analysis, Self-Adjointness*, Academic Press, 1975.

[15] W. Rudin, *Principles of Mathematical Analysis*, McGraw-Hill, 1991.

[16] W. Rudin, *Real and Complex Analysis*, McGraw-Hill, 1966.

[17] W. Rudin, *Functional Analysis*, McGraw-Hill, 1991.

[18] B.R. Rynne & M.A. Youngson, *Linear Functional Analysis*, Springer, 2008.

[19] M. Schechter, *Principles of Functional Analysis*, American Mathematical Soc., 1973.

[20] G.F. Simmons, *Introduction to Topology and Modern Analysis*, McGraw-Hill, 1963.

[21] W.A. Sutherland, *Introduction to Metric and Topological Spaces*, Oxford University Press, 2nd edition, 2009.

[22] A.E. Taylor & D.C. Lay, *Introduction to Functional Analysis*, John Wiley, 1980.

[23] J. Weidmann, *Lineare Operatoren in Hilberträumen*, Teubner, 1976.

[24] D. Werner, *Funktionalanalysis*, Springer Verlag, 2011.

[25] N. Young, *An Introduction to Hilbert Spaces*, Cambridge University Press, 1988.

[26] E. Zeidler, *Applied Functional Analysis: Applications to Mathematical Physics*, Springer Verlag, 1995.

[27] E. Zeidler, *Applied Functional Analysis: Main Principles and their Applications*, Springer Verlag, 1995.