Chapter 2

Functional analytic tools

The goal of this chapter is to introduce important tools from mathematical analysis and from functional analysis which are frequently used for the solution of identification and reconstruction problems, and build a basis for the advanced study of data assimilation problems.

The functional analytic language has many strengths and advantages, not only for the mathematical specialist, but also for scientists from various applications. It captures the main features of the problems and brings arguments into a form which applies to many diverse phenomena.

In this chapter, we are making use of parts of lectures by our colleague and teacher Rainer Kress, see [1], but put things into our framework to prepare the upcoming chapters. The following material has been extensively used in lectures at the University of Hokkaido, Japan, at Inha University, Korea, at Reading University, UK, and at the University of Göttingen, Germany.

Today a broad range of functional analysis books are available. But we believe that some classical books are worth reading, for example Reed and Simon [2], Bachman and Narici [3] or the very good book by Harro Heuser [4] (in German).

2.1 Normed spaces, elementary topology and compactness

2.1.1 Norms, convergence and the equivalence of norms

We first collect the main definitions of a distance in linear spaces. The natural idea of a distance between points in a space is usually captured by the mathematical term *metric*. The term *norm* combines these ideas with the *linear structure* of a vector space. Here we move directly into the linear structure.

Definition 2.1.1 (Norm, normed space). *Let X be a real or complex linear space (i.e. vector space). A function $\|\cdot\| : X \to \mathbb{R}$ with the properties*

$$\text{(N1)} \quad \|\varphi\| \geqslant 0 \qquad (\textit{positivity}) \tag{2.1.1}$$

$$\text{(N2)} \quad \|\varphi\| = 0 \;\; \textit{if and only if} \;\; \varphi = 0 \qquad (\textit{definiteness}) \tag{2.1.2}$$

doi:10.1088/978-0-7503-1218-9ch2

$$\text{(N3)} \quad \|\alpha\varphi\| = |\alpha| \, \|\varphi\| \qquad (\textit{homogeneity}) \tag{2.1.3}$$

$$\text{(N4)} \quad \|\varphi + \psi\| \leqslant \|\varphi\| + \|\psi\| \qquad (\textit{triangle inequality}) \tag{2.1.4}$$

for all $\varphi, \psi \in X$ *and* $\alpha \in \mathbb{R}$ *or* $\mathbb{C}$ *is called a norm on X. A linear space X equipped with a norm is called a normed space. We remark that the dimension of normed space X is defined in terms of the dimension of linear space.*

First, consider some examples of normed spaces.

Example 2.1.2 (*n*-dimensional real numbers). *The space* $\mathbb{R}^m$ *equipped with the* Euclidean norm

$$\|x\|_2 := \left(\sum_{j=1}^{n} |x_j|^2 \right)^{\frac{1}{2}}, \quad x \in \mathbb{R}^m \tag{2.1.5}$$

is a normed space which can be seen by checking the axioms (N1) to (N4). The only difficult part is the triangle inequality, which can be shown based on the Schwarz inequality.

We can also use the norm

$$\|x\|_\infty := \max_{j=1}^{n} |x_j|, \quad x \in \mathbb{R}^m \tag{2.1.6}$$

on $\mathbb{R}^m$. *This is called the* maximum norm. *Here the establishment of the axioms is straightforward.*

Further, we can define the one norm

$$\|x\|_1 := \sum_{j=1}^{n} |x_j|, \quad x \in \mathbb{R}^m. \tag{2.1.7}$$

It is indeed a norm according to our definition. The one norm is also called the Manhattan metric, *since it fits the travel time with roads as in parts of the city of New York.*

We can now transfer the basic terms from calculus into the setting of normed spaces. Most of the well-known concepts are the same.

Definition 2.1.3 (Convergence, $\epsilon - N$ criterion). *We say that a sequence* $(\varphi_n) \subset X$ *in a normed space X* converges *to an element* $\varphi \in X$, *if for every* $\epsilon > 0$ *there is an integer* $N \in \mathbb{N}$ *depending on* ϵ *such that for any* $n \geq N$

$$\|\varphi_n - \varphi\| \leqslant \epsilon. \tag{2.1.8}$$

Then φ *is called the* limit *of the sequence* (φ_n) *and we write*

$$\varphi_n \to \varphi, \quad n \to \infty \qquad \text{or} \qquad \lim_{n\to\infty} \varphi_n = \varphi. \tag{2.1.9}$$

Example 2.1.4 (Function spaces). *We consider the space* $C([a, b])$ *of continuous functions on the interval* $[a, b] \subset \mathbb{R}$. *The space equipped with the* maximum norm

$$\|\varphi\|_\infty := \max_{x\in[a,b]} |\varphi(x)| \tag{2.1.10}$$

is a normed space.

A different norm on $C([a, b])$ *can be defined by the* mean square norm

$$\|\varphi\|_2 := \left(\int_a^b |\varphi(x)|^2 \right)^{\frac{1}{2}}. \tag{2.1.11}$$

Convergence in the maximum norm is usually called uniform convergence. *Convergence in the mean square norm is referred to as* mean square convergence. *These two types of convergence are not equivalent. For example, the sequence*

$$\varphi_n(x) := \frac{1}{(1+x)^n}, \quad x \in [0, 1] \tag{2.1.12}$$

in $C([0, 1])$ *is convergent towards* $\varphi(x) \equiv 0$ *in the mean square norm, but it is not convergent at all in the maximum norm, since at* $x = 0$ *the functions all take the value* $\varphi_n(0) = 1$.

Definition 2.1.5 (Continuity of mappings). *A mapping* A *from* $U \subset X$ *with a normed space* X *into a normed space* Y *is called continuous at* $\varphi \in U$, *if*

$$\lim_{n\to\infty} A\varphi_n = A\varphi \tag{2.1.13}$$

for every sequence $(\varphi_n) \subset U$ *with* $\varphi_n \to \varphi$, $n \to \infty$. *The mapping* $A : U \to Y$ *is called continuous, if it is continuous at all* $\varphi \in U$.

Definition 2.1.6 (Equivalence of norms). *We call two norms* $\|\cdot\|_1, \|\cdot\|_2$ *on a linear space* X equivalent, *if each sequence* $(\varphi_n) \subset X$ *which converges with respect to* $\|\cdot\|_1$ *also converges with respect to* $\|\cdot\|_2$ *and vice versa.*

Theorem 2.1.7 (Estimates for equivalent norms). *Two norms* $\|\cdot\|_1, \|\cdot\|_2$ *are equivalent if and only if there exist constants* $C, c > 0$ *such that*

$$c\|\varphi\|_1 \leqslant \|\varphi\|_2 \leqslant C\|\varphi\|_1 \tag{2.1.14}$$

is satisfied for all $\varphi \in X$. *Also, the limits with respect to either norm coincide.*

Proof. First, if the estimates are satisfied, they immediately imply that convergence of a sequence with respect to $\|\cdot\|_1$ implies convergence of the sequence with respect to $\|\cdot\|_2$ and vice versa.

To show the other direction of the equivalence statement assume that there is no constant $C > 0$ such that the second part of the estimate (2.1.14) is satisfied. Then there is a sequence (φ_n) with $\|\varphi_n\|_1 = 1$ and $\|\varphi_n\|_2 \geqslant n^2$, $n \in \mathbb{N}$. We define

$$\psi_n := \frac{1}{n}\varphi_n, \quad n \in \mathbb{N}. \tag{2.1.15}$$

Then $\psi_n \to 0$, $n \to \infty$ with respect to $\|\cdot\|_1$, but

$$\|\psi_n\|_2 = \frac{1}{n}\|\varphi_n\|_2 \geqslant n \to \infty, \quad n \to \infty, \tag{2.1.16}$$

i.e. the sequence (ψ_n) is not convergent towards 0 with respect to $\|\cdot\|_2$. This shows that for equivalent norms the second part of the estimate must be satisfied. The arguments for the first part with the constant c work in the same way. □

We close this subsection with some observations on finite-dimensional normed spaces. Consider a space

$$X = \operatorname{span}\left\{u_1, \ldots, u_n\right\} = \left\{\alpha_1 u_1 + \cdots + \alpha_n u_n : \alpha_j \in \mathbb{R} \text{ (or } \mathbb{C})\right\} \tag{2.1.17}$$

with linearly independent elements $u_1, \ldots, u_n \in X$. Then every element $\varphi \in X$ can be expressed as a unique *linear combination*

$$\varphi = \sum_{j=1}^{n} \alpha_j u_j. \tag{2.1.18}$$

In this case we can define a norm in X by

$$\|\varphi\|_\infty := \max_{j=1}^{n} \left|\alpha_j\right|. \tag{2.1.19}$$

We obtain the estimate

$$\begin{aligned} \|\varphi\| &= \left\| \sum_{j=1}^{n} \alpha_j u_j \right\| \\ &\leqslant \sum_{j=1}^{n} \left\|\alpha_j u_j\right\| \\ &= \sum_{j=1}^{n} \left|\alpha_j\right| \left\|u_j\right\| \\ &\leqslant \max_{j=1}^{n} \left|\alpha_j\right| \cdot \sum_{j=1}^{n} \left\|u_j\right\| \\ &= C\|\varphi\|_\infty \end{aligned} \tag{2.1.20}$$

with

$$C := \sum_{j=1}^{n} \left\|u_j\right\|. \tag{2.1.21}$$

We will also obtain an equivalence in the other direction, stated fully as follows.

Theorem 2.1.8. *On a finite-dimensional linear space all norms are equivalent.*

Proof. For the norm $\|\cdot\|$ of X and the norm $\|\cdot\|_\infty$ defined by (2.1.19) we have already shown one part of the estimate (2.1.14). The other direction is proven as follows. Suppose that there is no $c > 0$ which satisfies $c\|\varphi\|_\infty \leqslant \|\varphi\|$ for all $\varphi \in X$. Then there exists a sequence $(\varphi_n) \subset X$ with $\|\varphi_n\| = 1$ and $\|\varphi\|_\infty \geqslant n$. Define

$$\psi_n := \frac{\varphi_n}{\|\varphi_n\|_\infty}, \quad n \in \mathbb{N} \tag{2.1.22}$$

and consider the representation in terms of the basis elements

$$\psi_n = \sum_{j=1}^{n} \alpha_{n,j} u_j, \quad n \in \mathbb{N}. \tag{2.1.23}$$

The coefficients $\alpha_{n,j}$ are bounded, since we have $\|\psi_n\|_\infty = 1$. Using the theorem of Bolzano–Weierstrass we have convergent subsequences of the sequence of coefficients such that $\alpha_{n(k),j} \to \alpha_{0,j}$, $k \to \infty$ for each $j = 1, \ldots, n$. We define

$$\psi_0 := \sum_{j=1}^{n} \alpha_{0,j} u_j \tag{2.1.24}$$

and have the convergence

$$\psi_{n(k)} \to \psi_0, \quad k \to \infty \tag{2.1.25}$$

with respect to $\|\cdot\|_\infty$. By $\|\cdot\| \leqslant C\|\cdot\|_\infty$ we obtain $\psi_{n(k)} \to \psi_0$ with respect to $\|\cdot\|$. But on the other hand we have $\|\psi_n\| = 1/\|\varphi\|_\infty \to 0$, $n \to \infty$. Since the limit is unique this yields $\psi_0 = 0$ and thus $\|\psi_0\|_\infty = 0$. But this contradicts $\|\psi_n\|_\infty = 1$ for all $n \in \mathbb{N}$. Thus also the other part of (2.1.14) must be satisfied for the norms $\|\cdot\|$ and $\|\cdot\|_\infty$.

Finally, given two norms $\|\cdot\|_1$ and $\|\cdot\|_2$ on X we first use the equivalence of $\|\cdot\|_1$ to $\|\cdot\|_\infty$ and then the equivalence of $\|\cdot\|_\infty$ to $\|\cdot\|_2$ to derive the statement of the theorem. $\square$

2.1.2 Open and closed sets, Cauchy sequences and completeness

For working with normed spaces, we need to transfer all the facts about open and closed sets, sequences and completeness in the Euclidean space to this setting. Let X be a normed space and $\varphi \in X$. We call the set

$$B(\varphi, \rho) := \left\{\psi \in X : \|\psi - \varphi\| < \rho\right\} \tag{2.1.26}$$

the *open ball* of radius ρ and center φ in X. The set

$$B[\varphi, \rho] := \left\{\psi \in X : \|\psi - \varphi\| \leqslant \rho\right\} \tag{2.1.27}$$

is called the *closed ball*.

Definition 2.1.9. *We note the following basic definitions.*

- *A subset U of a normed space X is called open, if for each element $\varphi \in U$ there is $\epsilon > 0$ such that the ball $B(\varphi, \epsilon) \subset U$.*
- *A subset U is called closed, if it contains all limits of convergent subsequences of U.*
- *The closure $\bar{U}$ of a set is the set of all limits of convergent sequences of U.*
- *A set U is called dense in another set V in a normed space X, if $V \subset \bar{U}$.*
- *A set U is called* bounded, *if there is a constant $C > 0$ such that*

$$\|\varphi\| \leqslant C \tag{2.1.28}$$

for all $\varphi \in U$.

We consider some examples.

Example 2.1.10 (Ball in different norms). *Consider the unit ball in different norms on $\mathbb{R}^2$. The ball in the Euclidean norm is what we usually call a* disc. *The ball in the infinity norm is in fact a square. The ball in the one-norm is a diamond. Other norms lead to different shapes.*

Example 2.1.11 (Weierstrass approximation theorem). *Consider the space of continuous functions $C([a, b])$. The* Weierstrass approximation theorem *states that every continuous function φ on $[a, b]$ can be approximated up to any precision by a polynomial*

$$p_n(x) = \sum_{j=0}^{n} a_j x^j \tag{2.1.29}$$

with appropriate coefficients a_j and degree n. In the functional analytic language this means that the space Π of polynomials is dense *in the set $C([a, b])$.*

Note that the space

$$\Pi_n := \left\{ p_n(x) = \sum_{j=0}^{n} a_j x^j \ : \ a_j \in \mathbb{R} \ (or\ \mathbb{C}) \right\} \tag{2.1.30}$$

for fixed $n \in \mathbb{N}$ is not *dense in $C([a, b])$.*

Next, we transfer further elementary definitions from calculus into the environment of normed spaces.

Definition 2.1.12 (Cauchy sequence). *A sequence (φ_n) in a normed space X is called a* Cauchy sequence, *if for every $\epsilon > 0$ there is $N \in \mathbb{N}$ such that*

$$\|\varphi_n - \varphi_m\| \leqslant \epsilon, \quad \forall n, m \geqslant N. \tag{2.1.31}$$

Every convergent sequence is a Cauchy sequence. However, if we have a Cauchy sequence then the limits do not need to be an element of the space. The term Cauchy sequence has been developed to capture convergence without talking about the limit. This is very advantageous, however, now the question is what we can say about the limit.

Example 2.1.13. *Consider the sequence of functions*

$$f_n(x) := \begin{cases} \dfrac{1}{(1+x)^n}, & 0 < x \leqslant 1 \\ 1, & -1 \leqslant x \leqslant 0. \end{cases} \tag{2.1.32}$$

It is a sequence in $C([-1, 1])$ *and also in* $L^2([-1, 1])$. *In the mean square norm the sequence converges towards the function*

$$f(x) := \begin{cases} 0, & 0 < x \leqslant 1 \\ 1, & -1 \leqslant x \leqslant 0. \end{cases} \tag{2.1.33}$$

Clearly, this function is not continuous and, thus, it is not an element of $C([-1, 1])$. *Thus, the sequence does not converge in the normed space* $(C([-1, 1]), \|\cdot\|_2)$.

To describe sets which contain the limits of their Cauchy sequences, we proceed as follows.

Definition 2.1.14. *A subset U of a normed space X is called* complete, *if every Cauchy sequence* (φ_n) *in U converges towards an element* φ *of U. We call a complete normed space X a* Banach space.

2.1.3 Compact and relatively compact sets

We first provide two equivalent definitions of compact sets. One comes from general topology, the other from metric spaces. In our framework both are equivalent. However, the range of the topological definition is far greater, if you want to proceed further into pure mathematics.

Definition 2.1.15. *A subset U of a normed space X is called* compact *if every open covering of U contains a finite subcovering. In more detail, for every family* V_j, $j \in J$ *with some index set J (which is in general infinite, it can be countable or non-countable) of open sets with*

$$U \subset \bigcup_{j \in J} V_j \tag{2.1.34}$$

there is a finite subfamily $V_{j(k)}$, $j(k) \in J$, $k = 1, \ldots, n$ *with*

$$U \subset \bigcup_{k=1}^{n} V_{j(k)}. \tag{2.1.35}$$

The set U is called totally bounded, *if for each* $\epsilon > 0$ *there exists a finite number of elements* $\varphi_1, \ldots, \varphi_n$ *in U such that*

$$U \subset \bigcup_{j=1}^{n} B(\varphi_j, \epsilon), \tag{2.1.36}$$

i.e. each element $\varphi \in U$ *has a distance less than* ϵ *from at least one of the elements* $\varphi_1, \ldots, \varphi_n$.

To obtain some experience with this definition consider examples for non-compact sets.

Example 2.1.16 (Unbounded set). *The set* $U := R^+$ *in* $\mathbb{R}$ *is not compact. Consider a covering with bounded open sets*

$$V_j := (j - 2, j + 2), \quad j = 1, 2, 3, \dots \tag{2.1.37}$$

There is not a finite subcovering which would cover the whole unbounded positive real axis U.

Example 2.1.17 (Not-complete set). *The set* $U := (0,\ 1]$ *in* $\mathbb{R}$ *is not compact. Consider a covering with open sets*

$$V_j := \left(\frac{1}{j+1}, \frac{1}{j - \frac{1}{2}} \right), \quad j = 1, 2, 3, \dots \tag{2.1.38}$$

There is not a finite subcovering which would cover the whole interval, since any finite subcovering has a maximal index N *and then the points smaller than* $x = \frac{1}{N+1}$ *are not contained in the finite subcovering.*

Definition 2.1.18. *A subset* U *in a normed space* X *is called* sequentially compact *if every sequence in* U *has a convergent subsequence in* U.

Remark. It is clear from their definitions that any finite sets are compact and sequentially compact.

Lemma 2.1.19. *A sequentially compact set* U *is totally bounded.*

Proof. Assume that the set U is not totally bounded. Then there is some number $\epsilon > 0$ for which no finite number N of balls $B(\varphi_j, \epsilon)$ with $\varphi_j \in U$, $j = 1, \dots, N$ covers U. For every finite number $\varphi_1, \dots, \varphi_n$ of elements we can find an element $\varphi_{n+1} \in U$ such that $\|\varphi_{n+1} - \varphi_j\| \geqslant \epsilon$ for all $j = 1, \dots, n$. This leads to a sequence (φ_n) such that

$$\|\varphi_n - \varphi_m\| \geqslant \epsilon, \quad n \neq m. \tag{2.1.39}$$

This sequence cannot contain a convergent subsequence, thus it is not sequentially compact. As a consequence we conclude that sequentially compact sequences are totally bounded. □

Lemma 2.1.20. *For each totally bounded set* U *there is a dense sequence* $(\varphi_n)_{n \in \mathbb{N}}$ *of elements.*

Proof. For $n \in \mathbb{N}$ we choose $\epsilon = 1/n$ and collect the finitely many elements $\varphi_1, \dots, \varphi_N$ for which $B(\varphi_j, \epsilon)$, $j = 1, \dots, N$ covers U. Putting these together into a sequence for $n = 1, 2, \dots$ we obtain a dense sequence in U. □

Lemma 2.1.21. *Consider a sequentially compact set* U *and an open covering* V_j, $j \in J$ *of* U. *Then there are* $\epsilon > 0$ *such that for any* $\varphi \in U$ *the ball* $B(\varphi, \epsilon)$ *is contained in one of the domains* V_j, $j \in J$.

Proof. If there is no $\epsilon > 0$, then for $\epsilon = 1/n$, $n \in \mathbb{N}$, there is an element φ_n for which $B(\varphi_n, 1/n)$ is not contained in one of the V_j, $j \in J$. The sequence of these φ_n has a convergent subsequence $(\varphi_{n(k)})_{k\in\mathbb{N}}$ (since U is sequentially compact). Its limit φ is in U and thus in one of the V_j, $j \in J$. Since V_j is open, $\varphi_{n(k)}$ tends to φ and $\epsilon = 1/n$ tends to 0, there is $K \in \mathbb{N}$ such that $B(\varphi_{n(k)}, 1/n(k))$ is in V_j for $k \geqslant K$. But this is a contradiction to our construction above and thus we obtain the statement of the lemma. □

We are now prepared for the following basic equivalence result.

Theorem 2.1.22 (Sequence criterion). *A subset U of a normed space is compact if and only if it is sequentially compact, i.e. if every sequence $(\varphi_n) \subset U$ of elements of U contains a convergent subsequence to an element of U.*

Proof. First, we assume that U is not sequentially compact. In this case there is a sequence (φ_n) in U which does not contain a convergent subsequence in U. Since every $\varphi \in U$ is not a limit point of a subsequence of (φ_n), there is a radius $\rho(\varphi)$ such that the ball $B(\varphi, \rho(\varphi))$ only contains finitely many elements of (φ_n). The set of all these balls clearly is a covering of U. If it contained a finite subcovering, then one of the finitely many balls would necessarily contain an infinite number of elements, thus there is no finite subcovering of this covering and thus U is not compact. This shows that if U is compact it is also sequentially compact.

Next, we show the other direction of the equivalence statement. Assume that U is sequentially compact. Then U is also totally bounded. Consider some open covering V_j, $j \in J$, of U. We need to show that there is a finite subcovering. According to lemma 2.1.21 we know that there is $\epsilon > 0$ such that $B(\varphi, \epsilon)$ is contained in one of the V_j for each $\varphi \in U$. However, since U is totally bounded, there is a finite number of elements $\varphi_1, \ldots, \varphi_n$ such that $B(\varphi_j, \epsilon)$ cover U. But they are all contained in one of the V_j, thus a finite number of the sets V_j will cover U. □

The following theorem of Bolzano and Weierstrass is one of the well-known results of analysis, which we recall here for further use and for illustration.

Theorem 2.1.23 (Bolzano–Weierstrass). *A bounded sequence in $\mathbb{R}^n$ has a convergent subsequence.*

Proof. Let us have a brief look at the proof in $\mathbb{R}$. The n-dimensional version of the proof works analogously.

Let $(z_n)_{n\in\mathbb{N}}$ be a bounded sequence in $\mathbb{R}$. We construct a convergent subsequence via interval partition. Since the sequence is bounded there is $a_1, b_1 \in \mathbb{R}$ such that $a_1 \leqslant z_n \leqslant b_1$ for all $n \in \mathbb{N}$. Then in at least one of the intervals

$$\left[a_1, \frac{a_1 + b_1}{2}\right], \qquad \left(\frac{a_1 + b_1}{2}, b_1\right]$$

there are infinitely many numbers of the sequence. We call this interval $[a_2, b_2]$ with appropriately chosen numbers a_1 and b_2. Then we apply the same argument to this new interval. Again we obtain an interval $[a_3, b_3]$ which contains infinitely many elements of the sequence. We repeat this to obtain two sequences (a_n) and (b_n). The sequence (a_n) is monotonously increasing, the sequence (b_n) is monotonously decreasing. For the distance between the interval boundaries we obtain

$$\|b_n - a_n\| \leqslant \frac{b - a}{2^{n-1}}, \quad n \in \mathbb{N}. \tag{2.1.40}$$

Thus both sequences are convergent towards a point $a = b$. □

Definition 2.1.24. *A subset U of a normed space is called relatively compact if its closure is compact.*

The $\mathbb{R}^m$ is a very important prototype of a normed space, which is studied extensively in calculus modules and courses on linear algebra. Next, we will employ the fact that it is the basic model for any *finite-dimensional* space.

Theorem 2.1.25. *Let U be a bounded and finite-dimensional subset of a normed space, i.e. there is $C > 0$ such that $\|\varphi\| \leqslant C$ for all $\varphi \in U$ and there are linearly independent elements $u_1, \ldots, u_n$ with $n \in \mathbb{N}$*

$$V := \operatorname{span}\{u_1, \ldots, u_n\} = \left\{\alpha_1 u_1 + \cdots + \alpha_n u_n : \alpha_j \in \mathbb{C} \text{ or } \mathbb{R}\right\} \tag{2.1.41}$$

such that $U \subset V$. Then the set U is relatively compact.

Proof. We map $\mathbb{R}^m$ (or $\mathbb{C}^n$) into V via

$$\Psi : \alpha \mapsto \alpha_1 u_1 + \cdots + \alpha_n u_n, \quad \text{for} \quad \alpha = \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{pmatrix} \in \mathbb{R}^m. \tag{2.1.42}$$

This mapping is bijective (one-to-one) and there are constants C, $c > 0$ such that we have the norm estimates

$$c\|\alpha\| \leqslant \|\Psi(\alpha)\| \leqslant C\|\alpha\|, \tag{2.1.43}$$

where $\|\cdot\|$ denotes the Euclidean norm in $\mathbb{R}^m$. Now, via (2.1.43) convergence in V and convergence in $\mathbb{R}^m$ are equivalent. The bounded set U is mapped into a bounded set $W := \Psi^{-1}(U)$ in $\mathbb{R}^m$. We apply the theorem of Bolzano–Weierstrass to conclude that W is relatively compact and again by (2.1.43) this also yields for U, which completes the proof. □

2.2 Hilbert spaces, orthogonal systems and Fourier expansion

2.2.1 Scalar products and orthonormal systems

So far we have used a distance concept which takes into account the linear structure of a space. However, from $\mathbb{R}^m$ we are used to geometry, in particular angles and orthogonality. Can this be transferred and exploited in a more general setting? The answer is yes. The concept of a Hilbert space is the appropriate setting to use orthogonality in a wider framework. The success of this concept has been overwhelming. The whole of modern physics with its quantum processes has been built onto the concept of the Hilbert space.

Definition 2.2.1 (Scalar product). *Let X be a real or complex linear space (i.e. vector space). A function* $\langle\cdot,\cdot\rangle$: $X \times X \to \mathbb{C}$ *with the properties*

$$\text{(S1)}\quad \langle\varphi, \varphi\rangle \geqslant 0 \qquad (\textit{positivity}) \tag{2.2.1}$$

$$\text{(S2)}\quad \langle\varphi, \varphi\rangle = 0 \text{ if and only if } \varphi = 0 \qquad (\textit{definiteness}) \tag{2.2.2}$$

$$\text{(S3)}\quad \langle\varphi, \psi\rangle = \overline{\langle\psi, \varphi\rangle} \qquad (\textit{symmetry}) \tag{2.2.3}$$

$$\text{(S4)}\quad \langle\alpha\varphi + \beta\psi, \xi\rangle = \alpha\langle\varphi, \xi\rangle + \beta\langle\varphi, \xi\rangle \qquad (\textit{linearity}) \tag{2.2.4}$$

for all $\varphi, \psi, \xi \in X$ *and* $\alpha \in \mathbb{R}$ *or* $\mathbb{C}$ *is called a* scalar product *on X. A linear space X equipped with a scalar product is called a* pre-Hilbert space. *Combining (S3) and (S4) we have*

$$\langle\xi, \alpha\varphi + \beta\psi\rangle = \overline{\alpha}\langle\xi, \varphi\rangle + \overline{\beta}\langle\xi, \psi\rangle. \tag{2.2.5}$$

We call this kind of property anti-linear *not only for what we have here with respect to the second variable of the scalar product.*

A basic inequality for pre-Hilbert spaces is given as follows.

Theorem 2.2.2. *For a pre-Hilbert space X we have the* Cauchy–Schwarz inequality

$$|\langle\varphi, \psi\rangle|^2 \leqslant \langle\varphi, \varphi\rangle\langle\psi, \psi\rangle \tag{2.2.6}$$

for $\varphi, \psi \in X$. *Here the equality holds if and only if* φ *and* ψ *are linearly dependent.*

Proof. For $\varphi = 0$ the inequality is clearly satisfied. For $\varphi \neq 0$ we define

$$\alpha = -\langle\varphi, \varphi\rangle^{-1/2}\overline{\langle\varphi, \psi\rangle}, \qquad \beta = \langle\varphi, \varphi\rangle^{1/2}. \tag{2.2.7}$$

Then we derive

$$0 \leqslant \langle\alpha\varphi + \beta\psi, \alpha\varphi + \beta\psi\rangle \tag{2.2.8}$$

$$= |\alpha|^2\langle\varphi, \varphi\rangle + 2\,\mathrm{Re}\left(\alpha\overline{\beta}\langle\varphi, \psi\rangle\right) + |\beta|^2\langle\psi, \psi\rangle \tag{2.2.9}$$

$$= \langle\varphi, \varphi\rangle\langle\psi, \psi\rangle - |\langle\varphi, \psi\rangle|^2. \tag{2.2.10}$$

This proves the inequality. We have equality if and only if $\alpha\varphi + \beta\psi = 0$, i.e. if φ and ψ are linearly dependent.

A scalar product $\langle\cdot,\cdot\rangle$ defines a norm via

$$\|\varphi\| := \langle\varphi, \varphi\rangle^{1/2}, \quad \varphi \in X. \tag{2.2.11}$$

Please check the norm axioms! Thus, a pre-Hilbert space is a normed space. If a pre-Hilbert space is complete, we call it a *Hilbert space*.

We can now introduce elements of geometry into an abstract space using the scalar product.

Definition 2.2.3. *We call two elements φ, ψ of a pre-Hilbert space X* orthogonal, *if*

$$\langle \varphi, \psi \rangle = 0. \tag{2.2.12}$$

In this case we write $\varphi \perp \psi$. Two subsets V, W of a Hilbert space are called orthogonal if for each pair of elements $\varphi \in V$ and $\psi \in W$ the equation (2.2.12) is satisfied. We write $V \perp W$.

We know orthogonality in $\mathbb{R}^m$. Here, an appropriate scalar product is given by the *Euclidean scalar product*

$$x \cdot y = \sum_{j=1}^{n} x_j y_j, \quad x, y \in \mathbb{R}^m. \tag{2.2.13}$$

For the special case $n = 2, 3$ we know that two vectors x, y are orthogonal if and only if the angle between these vectors is $90°$ or $\pi/2$. Moreover, using the series definition of $\cos(\theta)$ and $\sin(\theta)$ we can define the angle θ ($0 \leqslant \theta \leqslant 180°$) between vectors x, y via $\|x\|_2 \|y\|_2 \cos(\theta) = x \cdot y$.

In $\mathbb{R}^m$ we know that $e_1, \ldots, e_n \in \mathbb{R}^m$ where

$$e_1 = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \quad e_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \ldots, \quad e_n = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix} \tag{2.2.14}$$

is a basis of orthogonal vectors with respect to the Euclidean scalar product (2.2.13). We have also learned in linear algebra that there are many sets of vectors which span the whole space and are orthogonal to each other. This leads to the definition.

Definition 2.2.4. *A set U of elements $\varphi \in X$ in a pre-Hilbert space X is called an* orthogonal system, *if*

$$\langle \varphi, \psi \rangle = 0, \qquad \varphi \neq \psi \in U. \tag{2.2.15}$$

If in addition we have $\|\varphi\| = 1$ for all $\varphi \in U$, then the set is called an orthonormal system.

Such orthonormal systems are of importance in many mathematical disciplines, for example in approximation theory, interpolation theory and numerical mathematics, but also in very new mathematical areas such as wavelet theory and inverse problems. We will extensively explore orthonormal systems for the study of inversion problems.

We complete this subsection with some further basic definition.

Definition 2.2.5. *Given some set U in a pre-Hilbert space we call*

$$U^{\perp} := \left\{ \varphi \in X : \varphi \perp U \right\} \tag{2.2.16}$$

the orthogonal complement *of U.*

Example 2.2.6. *For example, the orthogonal complement of the set* $U = \{(0, 0, 1)^T\}$ *in* $\mathbb{R}^3$ *is the x–y plane* $\{(s, t, 0) : s, t \in \mathbb{R}\}$. *Given some vector* $y \in \mathbb{R}^3$, $y \neq 0$, *determine its orthogonal complement using the vector product*

$$x \times y = \begin{pmatrix} x_2 y_3 - x_3 y_2 \\ -(x_1 y_3 - x_3 y_1) \\ x_1 y_2 - x_2 y_3 \end{pmatrix} \tag{2.2.17}$$

defined for arbitrary $x, y \in \mathbb{R}^3$. *Note that we have*

$$x \cdot (x \times y) = 0, \qquad y \cdot (x \times y) = 0 \tag{2.2.18}$$

for vectors $x, y \in \mathbb{R}^3$ *with* $x \neq y$, $x \neq 0$, $y \neq 0$. *Thus we choose any vector* $x \in \mathbb{R}^3$ *with* $x \neq y$ *and* $x \neq 0$ *and define vectors*

$$z_1 := x \times y, \qquad z_2 := z_1 \times y. \tag{2.2.19}$$

Then $y \cdot z_1 = 0$ *and* $y \cdot z_2 = y \cdot ((x \times y) \times y) = 0$ *and thus*

$$V := \operatorname{span}\{z_1, z_2\} = \{\alpha z_1 + \beta z_2 : \alpha, \beta \in \mathbb{R}\} \tag{2.2.20}$$

satisfies

$$V = \{y\}^{\perp}. \tag{2.2.21}$$

2.2.2 Best approximations and Fourier expansion

Here we collect some facts about best approximations in pre-Hilbert spaces. These are of interest by themselves, but can also be seen as a basis for the Fourier expansion. We would like to remark that the Fourier theory in Hilbert spaces is much easier than the classical Fourier theory.

Further, note that most solutions of inverse problems can be understood as some type of *best approximation*, for example three-dimensional or four-dimensional variational data assimilation in chapter 5, but also Tikhonov regularization which we introduce in section 3.1.4. Fourier theory is a core ingredient to understand and analyze regularization techniques.

Definition 2.2.7 (Best approximation). *Consider a set* U *in a normed space* X *and* $\varphi \in X$. *Then* $\psi \in U$ *is called the* best approximation *to the element* φ *with respect to* U *if*

$$\|\varphi - \psi\| = \inf_{\xi \in U} \|\varphi - \xi\| \tag{2.2.22}$$

is satisfied. The best approximation is an element with the smallest distance

$$d(\varphi, U) := \inf_{\xi \in U} \|\varphi - \xi\| \tag{2.2.23}$$

of the set U *to the element* φ.

As basic questions of approximation we need to investigate existence and uniqueness of such 'best' approximations. We will provide some results for particular settings for further use.

Theorem 2.2.8. *If U is a finite-dimensional subspace of a normed space X, then for every element $\varphi \in X$ there exists a best approximation with respect to U.*

Proof. We remark that the finite-dimensional subspace U is complete (this is a consequence of the theorem of Bolzano–Weierstrass) and every bounded sequence in U has a convergent subsequence in U. We now construct a minimizing sequence (φ_n) in U, i.e. a sequence with

$$\|\varphi_n - \varphi\| \to d(U, \varphi), \quad n \to \infty. \tag{2.2.24}$$

It is clearly bounded, since $\|\varphi_n\| \leqslant \|\varphi_n - \varphi\| + \|\varphi\|$. Now we choose a convergent subsequence, which tends to an element $\psi \in U$. Passing to the limit for the subsequence $\varphi_{n(k)}$ of φ_n in (2.2.24) we obtain $\|\psi - \varphi\| = d(\varphi, U)$, i.e. ψ is best approximation to φ. □

In normed spaces best approximations do not need to be unique.

Example 2.2.9. *Consider the space $\mathbb{R}^2$ equipped with the maximum norm $\|\cdot\|_\infty$. In fact take $U = \{\lambda e_2 : \lambda \in \mathbb{R}\}$ and $\varphi = (1, 0)$. Then, any point in $\Lambda = \{(0, \mu) : \mu \in \mathbb{R}, |\mu| \leqslant 1\}$ is a best approximation to φ with respect to the maximum norm.*

Next, we employ the scalar product to formulate criteria for best approximations with respect to linear subspaces of pre-Hilbert spaces. In this case we can also answer the uniqueness question.

Theorem 2.2.10. *Let X be a pre-Hilbert space with inner product $\langle\cdot,\cdot\rangle$ and $\varphi \in X$.*

(i) *Let U be a convex subset of X. That is, U satisfies the condition*

$$\alpha\varphi_1 + (1 - \alpha)\varphi_2 \in U, \quad \alpha \in [0, 1] \tag{2.2.25}$$

for any $\varphi_1, \varphi_2 \in U$. Then, $\psi \in U$ is the best approximation to φ if and only if

$$\mathrm{Re}\langle \varphi - \psi, u - \psi\rangle \leqslant 0, \quad u \in U. \tag{2.2.26}$$

(ii) *If U is a linear subspace of X. Then, $\psi \in U$ is the best approximation to φ if and only if*

$$\langle \varphi - \psi, \xi\rangle = 0, \quad \xi \in U, \tag{2.2.27}$$

i.e. if $\varphi - \psi \perp U$.

For both (i) and (ii) there is at most one best approximation.

Proof. Since (ii) easily follows from (i), we only prove (i). Take any $u \in U$ and fix it. Consider the function

$$\begin{aligned} f(\alpha) &= \|\varphi - ((1 - \alpha)\psi + \alpha u)\|^2 \\ &= \|\varphi - \psi\|^2 - 2\alpha \mathrm{Re}\langle \varphi - \psi, u - \psi\rangle + \alpha^2\|u - \psi\|^2 \end{aligned} \tag{2.2.28}$$

for $\alpha \in [0, 1]$. First, assume (2.2.26). Then we have

$$f(1) \geqslant \|\varphi - \psi\|^2 + \|u - \psi\|^2 \geqslant \|\varphi - \psi\|^2. \tag{2.2.29}$$

Hence ψ is the best approximation to φ.

Next we prove the converse. We prove this by a contradictory argument. Let ψ is the best approximation to φ and suppose (2.2.26) does not hold. Then, there exists $u \in U$ such that

$$\mathrm{Re}\langle \varphi - \psi, u - \psi \rangle > 0. \tag{2.2.30}$$

Then by (2.2.28),

$$f(0) > f(\alpha), \qquad 0 < \alpha \ll 1. \tag{2.2.31}$$

Hence, there exist $\psi' = (1 - \alpha)\psi + \alpha u = \psi + \alpha(u - \psi) \in U$ with some $\alpha \in (0, 1)$ such that $\|\varphi - \psi\| > \|\varphi - \psi'\|$, which contradicts the fact that ψ is the best approximation to φ.

Finally, assume that there are two best approximations $\psi_1, \psi_2 \in U$ to φ. Then with $\xi := \psi_2 - \psi_1 \in U$ we obtain

$$0 \leqslant \|\psi_2 - \psi_1\|^2 = \langle \varphi - \psi_1, \psi_2 - \psi_1 \rangle + \langle \varphi - \psi_2, \psi_1 - \psi_2 \rangle \leqslant 0 \tag{2.2.32}$$

thus $\psi_1 = \psi_2$. □

Next, we note some further results about the existence of a unique best approximation preparing our main theorem about the Fourier expansion.

Theorem 2.2.11. *Consider a complete linear subspace U of a pre-Hilbert space X. Then for every element $\varphi \in X$ there exists a unique best approximation to φ with respect to U. Hence, X has the decomposition $X = U \oplus U^{\perp}$, that is, each $\varphi \in X$ can be uniquely written in the form $\varphi = \psi + \chi$ with $\psi \in U$, $\chi \in U^{\perp}$. The mapping which maps each φ to ψ in the above decomposition of φ is called the* orthogonal projection *in X onto U.*

Proof. The idea is to take a minimizing sequence such that

$$\|\varphi - \varphi_n\|^2 \leqslant d(\varphi, U)^2 + \frac{1}{n}, \quad n \in \mathbb{N}. \tag{2.2.33}$$

For $n, m \in \mathbb{N}$ we estimate

$$\begin{aligned}
&\left\|(\varphi - \varphi_n) + (\varphi - \varphi_m)\right\|^2 + \left\|\varphi_n - \varphi_m\right\|^2 \\
&= \left\|(\varphi - \varphi_n) + (\varphi - \varphi_m)\right\|^2 + \left\|\varphi - \varphi_m - (\varphi - \varphi_n)\right\|^2 \\
&= \left(\text{calculations via } \|v\|^2 = \langle v, v\rangle\right) \\
&= 2\|\varphi - \varphi_n\|^2 + 2\|\varphi - \varphi_m\|^2 \\
&\leqslant 4d(\varphi, U)^2 + \frac{2}{n} + \frac{2}{m}, \quad n, m \in \mathbb{N}.
\end{aligned} \tag{2.2.34}$$

Since $\|\varphi - 1/2(\varphi_n + \varphi_m)\| \geqslant d(\varphi, U)$ we now obtain

$$\|\varphi_n - \varphi_m\|^2 \leqslant 4d(\varphi, U)^2 + \frac{2}{n} + \frac{2}{m} - 4\left\|\varphi - \frac{1}{2}(\varphi_n + \varphi_m)\right\|^2 \leqslant \frac{2}{n} + \frac{2}{m}$$

for $n, m \in \mathbb{N}$, thus (φ_n) is a Cauchy sequence. Since U is complete, there is a limit ψ of this Cauchy sequence for which from (2.2.33) we obtain (2.2.22), i.e. ψ is the best approximation to φ with respect to U.

The uniqueness is a consequence of the previous theorem 2.2.10. □

The usual term series from analysis can be easily transferred to a normed space setting. Consider a sequence (φ_n) in a normed space X. Then we can define the *partial sums*

$$S_N := \sum_{n=1}^{N} \varphi_n, \quad N \in \mathbb{N}, \tag{2.2.35}$$

which is just a finite sum of elements in X. For each $N \in \mathbb{N}$ the partial sum S_N is an element of X. Thus, we obtain a sequence $(S_N)_{N\in\mathbb{N}}$ in X. If this sequence is convergent towards an element $S \in X$ for $N \to \infty$, then we say that the *infinite series* (in short 'series')

$$\sum_{n=1}^{\infty} \varphi_n \tag{2.2.36}$$

is convergent and has value S. We write $S = \sum_{n=1}^{\infty} \varphi_n$.

We now arrive at the culmination point of this chapter, deriving some quite general theorem about the representation of elements with respect to an orthogonal system.

Theorem 2.2.12 (Fourier representation). *We consider a pre-Hilbert space X with an orthonormal system $\{\varphi_n : n \in \mathbb{N}\}$. Then the following properties are equivalent:*

(1) *The set* span $\{\varphi_n : n \in \mathbb{N}\}$ *is dense in X. Recall that* span$\{\}$ *is the set of all linear combinations of a* finite *subset of* $\{\varphi_n : n \in \mathbb{N}\}$.

(2) *Each element $\varphi \in X$ can be expanded in a* Fourier series

$$\varphi = \sum_{n=1}^{\infty} \langle \varphi, \varphi_n \rangle \varphi_n. \tag{2.2.37}$$

(3) *For each $\varphi \in X$ there is the* Parseval identity

$$\|\varphi\|^2 = \sum_{n=1}^{\infty} |\langle \varphi, \varphi_n \rangle|^2. \tag{2.2.38}$$

Usually the polarization of this is called the Parseval identity and (2.2.38) *itself is called the Plancherel theorem. If an orthonormal system satisfies these properties it is called* complete.

Proof. First we show that (1) implies (2). According to theorems 2.2.10 and 2.2.11 the partial sum

$$\psi_N = \sum_{n=1}^{N} \langle \varphi, \varphi_n \rangle \varphi_n \tag{2.2.39}$$

is the best approximation to φ with respect to the finite-dimensional linear space $\operatorname{span}\{\varphi_1, \ldots, \varphi_N\}$. Since according to (1) the span is dense in X, the best approximation ψ_N will converge towards φ for $N \to \infty$, which establishes the convergence of (2.2.37).

To show the implication from (2) to (3) we consider

$$\langle \varphi, \psi_N \rangle = \sum_{n=1}^{N} \overline{\langle \varphi, \varphi_n \rangle} \langle \varphi, \varphi_n \rangle = \sum_{n=1}^{N} |\langle \varphi, \varphi_n \rangle|^2 \tag{2.2.40}$$

for the previous ψ_N. Passing to the limit $N \to \infty$ we obtain (2.2.38).

To pass from (3) to (1) we calculate

$$\left\| \varphi - \sum_{n=1}^{N} \langle \varphi, \varphi_n \rangle \varphi_n \right\|^2 = \|\varphi\|^2 - \sum_{n=1}^{N} |\langle \varphi, \varphi_n \rangle|^2. \tag{2.2.41}$$

The right-hand side tends to zero and the left-hand side is the approximation of φ by elements of $\operatorname{span}\{\varphi_n : n \in \mathbb{N}\}$. This yields (1).

With the three implications (1) ⇒ (2), (2) ⇒ (3) and (3) ⇒ (1) we have proven the equivalence of all three statements. □

Remark 2.2.13. *The proof of the previous theorem provides the* Bessel inequality

$$\sum_{n=1}^{\infty} |\langle \varphi, \varphi_n \rangle|^2 \leqslant \|\varphi\|^2 \tag{2.2.42}$$

for any $\varphi \in X$ when $\{\varphi_n : n \in \mathbb{N}\}$ is an orthonormal system in X.

Before dealing with convergence of the Fourier series of a function $\varphi \in L^2((0, 2\pi))$ in the mean square sense, we prepare the following lemma which is necessary for its necessity.

Lemma 2.2.14. *$C^\infty([a, b])$ is dense in $L^2((a, b))$. In particular $C([a, b])$ is dense in $L^2((a, b))$.*

Proof. We use the method of mollification for the proof. The details are as follows. Let $f \in L^2((a, b))$. Then for $\delta(0 < \delta < 1)$ we define $f_\delta \in L^2((\alpha, \beta))$ with $\alpha = (a + b)/2 - (b - a)/(2\delta)$, $\beta = (a + b)/2 + (b - a)/(2\delta)$ by

$$f_\delta(x) = f\left(\delta x + \frac{a + b}{2}(1 - \delta)\right), \qquad x \in (\alpha, \beta). \tag{2.2.43}$$

Then it is easy to see by the Lebesgue dominated convergence theorem that $f_\delta \to f\,(\delta \to 1)$ in $L^2((a, b))$. We extend f_δ to the whole $\mathbb{R}$ by extending it 0 outside (α, β) and denote its extension by $\tilde{f}_\delta$, i.e.

$$\tilde{f}_\delta(x) = \begin{cases} f_\delta(x) & \text{if } x \in (\alpha, \beta) \\ 0 & \text{if otherwise.} \end{cases} \tag{2.2.44}$$

Now we mollify $\tilde{f}_\delta$ by a mollifier $\chi_\varepsilon \in C_0^\infty(\mathbb{R})$ given by $\chi_\varepsilon(x) = \varepsilon^{-1}\chi(\varepsilon^{-1}x)$ with a $\chi \in C_0^\infty(\mathbb{R})$ with the properties $0 \leqslant \chi(x) \leqslant 1$ for any $x \in \mathbb{R}$, $\chi(x) = 0$ if $|x| \geqslant 1$ and $\int_{\mathbb{R}} \chi(x)\mathrm{d}x = 1$. The mollification $\tilde{f}_{\delta,\varepsilon}$ of $\tilde{f}_\delta$ by this mollifier is defined by their convolution, i.e.

$$\tilde{f}_{\delta,\varepsilon}(x) = \left(\tilde{f}_\delta * \chi_\varepsilon\right)(x) = \int_{\mathbb{R}} \tilde{f}_\delta(x-y)\chi(y)\,\mathrm{d}y. \tag{2.2.45}$$

By using $\sup_{|y|\leqslant\varepsilon} \int_{\mathbb{R}} |\tilde{f}_\delta(x-y) - \tilde{f}_\delta(x)|^2\,\mathrm{d}x \to 0(\varepsilon \to)$ because $\tilde{f}_\delta \in L^2(\mathbb{R})$, it can be easily seen that

$$\tilde{f}_{\delta,\varepsilon} \to \tilde{f}_\delta(\varepsilon \to 0) \text{ in } L^2(\mathbb{R}). \tag{2.2.46}$$

Then the proof will be finished by just noting that $\tilde{f}_\delta = f_\delta$ in (a, b) and $\tilde{f}_{\delta,\varepsilon} \in C_0^\infty(\mathbb{R})$. □

Example 2.2.15. *As an example consider the set $L^2([0, 2\pi])$ of square integrable functions on the interval $[0, 2\pi]$. Equipped with the scalar product*

$$\langle \varphi, \psi \rangle := \int_0^{2\pi} \varphi(y)\overline{\psi(y)}\,\mathrm{d}y \tag{2.2.47}$$

this space is a pre-Hilbert space. The norm induced by this scalar product is the mean square norm which we have discussed above.

For $n = 1, 2, 3, \ldots$ consider the functions

$$\varphi_{2n}(x) := \frac{1}{\sqrt{\pi}}\sin(nx), \qquad x \in [0, 2\pi], \quad n = 1, 2, 3, \ldots \tag{2.2.48}$$

$$\varphi_{2n+1}(x) := \frac{1}{\sqrt{\pi}}\cos(nx), \qquad x \in [0, 2\pi], \quad n = 0, 1, 2, \ldots. \tag{2.2.49}$$

We note the orthogonality relations

$$\int_0^{2\pi} \cos(nx)\cos(mx)\,\mathrm{d}x = \begin{cases} 0, & n \neq m \\ \pi, & n = m \end{cases} \tag{2.2.50}$$

$$\int_0^{2\pi} \cos(nx)\sin(mx)\,\mathrm{d}x = 0 \tag{2.2.51}$$

$$\int_0^{2\pi} \sin(nx)\sin(mx)\,\mathrm{d}x = \begin{cases} 0, & n \neq m \\ \pi, & n = m \end{cases} \tag{2.2.52}$$

for all $n, m \in \mathbb{N}_0 = \mathbb{N} \cup \{0\}$. Thus, the set

$$\mathcal{B} := \left\{\varphi_n : n \in \mathbb{N}\right\} \tag{2.2.53}$$

with φ_n defined by (2.2.48), (2.2.49) is an orthonormal system in $L^2([0, 2\pi])$.

From the approximation theorem for trigonometric polynomials we obtain that the set of trigonometric polynomials

$$t_N(x) := a_0 + \sum_{n=1}^{N} (a_n \cos(nx) + b_n \sin(nx)), \quad x \in [0, 2\pi] \tag{2.2.54}$$

where $N \in \mathbb{N}$ *and* $a_j, b_j \in \mathbb{R}$ *(or* $\mathbb{C}$*) is dense in the space* $C([0, 2\pi])$ *by the Weierstrass approximation theorem. By lemma 2.2.14, the set of trigonometric polynomials is also dense in* $X = L^2([0, 2\pi])$ *equipped with the scalar product (2.2.47).*

This shows that property (1) of theorem 2.2.12 is satisfied. As a consequence, all equivalent properties hold and we can represent any function in $L^2([0, 2\pi])$ *in terms of its* Fourier series

$$\varphi(x) = a_0 + \sum_{n=1}^{\infty} (a_n \cos(nx) + b_n \sin(nx)), \quad x \in [0, 2\pi], \tag{2.2.55}$$

where the convergence is valid in the mean square sense.

2.3 Bounded operators, Neumann series and compactness

2.3.1 Bounded and linear operators

Linear mappings in $\mathbb{R}^m$ are well known from elementary courses on linear algebra. Analogously, a mapping $A : X \to Y$ from a linear space X into some linear space Y is called *linear*, if

$$A(\alpha\varphi + \beta\psi) = \alpha A\varphi + \beta A\psi \tag{2.3.1}$$

for all $\varphi, \psi \in X$ and $\alpha, \beta \in \mathbb{R}$ (or $\mathbb{C}$). Linear operators have quite nice properties. First, we note:

Theorem 2.3.1. *A linear operator is continuous in the whole space if it is continuous at one element (for example in* $\varphi = 0$*).*

Proof. If A is continuous everywhere it is clearly continuous in a particular element $\varphi_0 \in X$. If now A is continuous in $\varphi_0 \in X$, i.e. for every sequence $(\varphi_n) \subset X$ with $\varphi_n \to \varphi_0$ for $n \to \infty$, we have $A\varphi_n \to A\varphi_0$, $n \to \infty$. Consider an arbitrary element $\varphi \in X$ and a sequence $(\tilde{\varphi}_n)$ with $\tilde{\varphi}_n \to \varphi$ for $n \to \infty$. We define $\varphi_n := \tilde{\varphi}_n + \varphi_0 - \varphi$, which tends towards φ_0 for $n \to \infty$. Now we conclude

$$A\tilde{\varphi}_n = A\varphi_n + A(\varphi - \varphi_0) \to A\varphi_0 + A(\varphi - \varphi_0) = A\varphi \tag{2.3.2}$$

for $n \to \infty$, which completes the proof.

Definition 2.3.2. *A linear operator* $A : X \to Y$ *with normed spaces* X, Y *is called* bounded, *if there is a constant* $C > 0$ *such that*

$$\|A\varphi\| \leqslant C\|\varphi\|, \quad \varphi \in X. \tag{2.3.3}$$

We call such a constant C *a* bound *for the operator* A*.*

We can restrict ourselves in the definition of the bound to a set of non-zero elements $\varphi \in X$ with $\|\varphi\| \leqslant 1$ or even with $\|\varphi\| = 1$. This is due to the fact that

$$\|A\varphi\| = \left\| A\left(\frac{\varphi}{\|\varphi\|} \right) \right\| \cdot \|\varphi\| \tag{2.3.4}$$

for all non-zero $\varphi \in X$. Thus, if (2.3.3) is satisfied for all φ with $\|\varphi\| \leqslant 1$, then via (2.3.4) we estimate

$$\|A\varphi\| \leqslant C \left\| \frac{\varphi}{\|\varphi\|} \right\| \cdot \|\varphi\| = C\|\varphi\| \tag{2.3.5}$$

for arbitrary non-zero $\varphi \in X$.

Definition 2.3.3. *An operator is bounded if and only if*

$$\|A\| := \sup_{\|\varphi\|=1} \|A\varphi\| = \sup_{\|\varphi\|\leqslant 1} \|A\varphi\| < \infty. \tag{2.3.6}$$

In this case we call $\|A\|$ *the* operator norm *of* A.

Example 2.3.4. *Examples for bounded operators on* $\mathbb{R}^m$ *are given by matrices. Note that in* $\mathbb{R}^m$ *every linear operator is bounded.*

Example 2.3.5. *Consider the integral operator*

$$(A\varphi)(x) := \int_a^b k(x, y)\varphi(y)\,\mathrm{d}y, \quad x \in [a, b] \tag{2.3.7}$$

with some continuous kernel $k : [a, b] \times [a, b] \to \mathbb{C}$. *Then we can estimate*

$$\begin{aligned} |A\varphi(x)| &= \left| \int_a^b k(x, y)\varphi(y)\mathrm{d}y \right| \\ &\leqslant \int_a^b |k(x, y)| \cdot |\varphi(y)|\,\mathrm{d}y \\ &\leqslant \int_a^b |k(x, y)| \cdot \sup_{y\in[a,b]} |\varphi(y)|\,\mathrm{d}y \\ &= C\|\varphi\|_\infty \end{aligned} \tag{2.3.8}$$

with

$$C := \sup_{x\in[a,b]} \int_a^b |k(x, y)|\,\mathrm{d}y. \tag{2.3.9}$$

This proves that the operator $A : C([a, b]) \to C([a, b])$ *is bounded where* $C([a, b])$ *is equipped with the maximum norm.*

We can add operators $A_1, A_2 : X \to Y$ by pointwise summation

$$(A_1 + A_2)\varphi := A_1\varphi + A_2\varphi, \quad \varphi \in X \tag{2.3.10}$$

and multiply an operator by a real or complex number α via

$$(\alpha A)\varphi := \alpha A\varphi. \tag{2.3.11}$$

Thus, the set of all linear operators is itself a *linear space*. Also, each linear combination of *bounded* linear operators is again a bounded operator. We call this space $BL(X, Y)$, the *space of bounded linear operators* from the normed space X into the normed space Y.

Theorem 2.3.6. *The linear space $BL(X, Y)$ of bounded linear operators for normed spaces X, Y is a normed space with the norm (2.3.6). If Y is a Banach space (i.e. complete), then $BL(X, Y)$ is also a Banach space.*

Proof. It is easy to see that $BL(X, Y)$ is a normed space. Assume that (A_n) is a Cauchy sequence of operators $X \to Y$. Since

$$\|A_n\| \leqslant \|A_n - A_1\| + \|A_1\| \leqslant C, \quad n \in \mathbb{N}, \tag{2.3.12}$$

the sequence A_n is bounded. Then also $(A_n\varphi)$ is a Cauchy sequence in Y for each point $\varphi \in X$. Since Y is a Banach space, $(A_n\varphi)$ converges towards an element $\psi \in Y$. We define the operator A via

$$A\varphi := \psi = \lim_{m\to\infty} A_m\varphi. \tag{2.3.13}$$

Then it is easy to see that A is linear. From

$$\|\psi\| = \lim_{m\to\infty} \|A_m\varphi\| \leqslant C\|\varphi\| \tag{2.3.14}$$

we obtain that A is a bounded operator. We have

$$\begin{aligned}
\|A - A_n\| &= \sup_{\|\varphi\|\leqslant 1} \|A_n\varphi - A\varphi\| \\
&\leqslant \sup_{\|\varphi\|\leqslant 1} \limsup_{m\to\infty} \|A_n\varphi - A_m\varphi\| \\
&\leqslant \limsup_{m\to\infty} \|A_n - A_m\| \\
&\leqslant \epsilon
\end{aligned} \tag{2.3.15}$$

for all n sufficiently large. Thus A is the operator limit of the sequence (A_n) in $BL(X, Y)$. $\square$

Take note of the two different concepts of convergence for operators. First, there is *pointwise convergence*, when

$$A_n\varphi \to A\varphi, \quad n \to \infty \tag{2.3.16}$$

for each fixed point $\varphi \in X$. Second, there is *norm convergence* if

$$\|A_n - A\| \to 0, \quad n \to \infty. \tag{2.3.17}$$

Example 2.3.7. *A standard example for operators which are pointwise convergent, but not norm convergent, are* interpolation operators. *For $a < b \in \mathbb{R}$ consider the grid points*

$$x_k^{(N)} := a + \frac{b-a}{N}k, \quad k = 0, \ldots, N \tag{2.3.18}$$

and piecewise linear functions $p_k^{(N)}$ in $C([a, b])$ with

$$p_k^{(N)}\left(x_j^{(N)}\right) = \begin{cases} 1, & k = j \\ 0, & \text{otherwise}. \end{cases} \tag{2.3.19}$$

Such a set of functions is called a Lagrange basis *for the interpolation problem to find a piecewise linear function on $C([a, b])$ which is equal to a given function φ in $x_j^{(N)}$, $j = 1, \ldots, N$. A solution to this interpolation problem is given by the interpolation operator*

$$(P_N\varphi)(x) := \sum_{k=1}^{N} p_k^{(N)}(x)\varphi(x_k), \quad x \in [a, b]. \tag{2.3.20}$$

The infinity norm of P_N can be estimated by

$$\begin{aligned} |(P_N\varphi)(x)| &\leqslant \|\varphi\|_\infty \left| \sum_{k=0}^{N} p_k^{(N)}(x) \right| \\ &= \|\varphi\|_\infty \cdot 1. \end{aligned} \tag{2.3.21}$$

For every fixed element $\varphi \in C([a, b])$ we have the convergence

$$\|P_N\varphi - \varphi\|_\infty = \sup_{x\in[a,b]} \left| \sum_{k=1}^{N} p_k^{(N)}(x)\varphi(x_k) - \varphi(x) \right| \to 0 \tag{2.3.22}$$

for $N \to \infty$ due to the continuity of the function φ on the closed interval $[a, b]$. Thus the interpolation operator converges pointwise towards the identity operator. However, the norm difference between P_N and I satisfies

$$\|P_N - I\|_\infty = \sup_{x\in[a,b], \|\varphi\|_\infty \leqslant 1} \left| \sum_{k=1}^{N} p_k^{(N)}(x)\varphi(x_k) - \varphi(x) \right| = 2 \tag{2.3.23}$$

for all $N \in \mathbb{N}$. Thus P_N does not converge towards I.

Example 2.3.8. *As a second example consider integral operators on the real axis*

$$(A_n\varphi)(x) := \int_{\mathbb{R}} \underbrace{\frac{1}{1+|x|^2} \cdot \frac{1}{1+|n-y|^3}}_{=:k_n(x,y)} \varphi(y)\,\mathrm{d}y, \quad x \in \mathbb{R} \tag{2.3.24}$$

which we study as an operator from the set $L^2(\mathbb{R})$ into itself. The kernel is visualized via `MATLAB` *or Scilab in figure 2.1.*

Figure 2.1. The figure shows the kernel of (2.3.24) in dependence on y for $x = 0$ for $n = 2, 5, 8$. The kernel moves along the $\mathbb{R}$-axis.

Let us investigate the operator sequence for a fixed L^2-function φ in $\mathbb{R}$. We remark that given $\epsilon > 0$ we can find $A > 0$ such that

$$\int_{|x|\geqslant a} |\varphi(y)|^2 \, \mathrm{d}y \leqslant \epsilon \tag{2.3.25}$$

for all $a > A$. Further, for fixed b and given ϵ we can find N such that

$$|k_n(x, \ y)| \leqslant \epsilon, \qquad |x|, |y| \leqslant b, \quad n \geqslant N. \tag{2.3.26}$$

Clearly the kernel $k(x, y)$ is bounded by 1 for all $x, y \in \mathbb{R}$.

We estimate using the Cauchy–Schwarz inequality in $L^2(\mathbb{R})$

$$\begin{aligned} \int_{\mathbb{R}} \left| \int_{\mathbb{R}} k_n(x, y)\varphi(y) \, \mathrm{d}y \right|^2 \mathrm{d}x &\leqslant \int_{\mathbb{R}} \left(\int_{\mathbb{R}} |k_n(x, y)|^2 \, \mathrm{d}y \int_{\mathbb{R}} |\varphi(y)|^2 \, \mathrm{d}y \right) \mathrm{d}x \\ &= \|\varphi\|_2^2 \cdot \int_{\mathbb{R}} \int_{\mathbb{R}} |k_n(x, y)|^2 \, \mathrm{d}y \ \mathrm{d}x. \end{aligned} \tag{2.3.27}$$

Since the kernel $k_n(x, y)$ is square integrable over $\mathbb{R} \times \mathbb{R}$, this shows the boundedness of the operators A_n from $L^2(\mathbb{R})$ into $L^2(\mathbb{R})$. Further, it can be seen that the sequence is not norm convergent, but it converges pointwise towards zero for functions in $L^2(\mathbb{R})$.

Bounded linear operators have the following extremely useful property.

Theorem 2.3.9. *A linear operator is continuous if and only if it is bounded.*

Proof. First, assume that an operator $A : X \to Y$ is bounded and consider a sequence (φ_n) in X with $\varphi_n \to 0$, $n \to \infty$. Then from $\|A\varphi_n\| \leqslant C\|\varphi_n\|$ it follows that $A\varphi_n \to 0$, $n \to \infty$. Thus A is continuous in 0 and therefore continuous everywhere according to theorem 2.3.1.

Now let us assume that A is continuous, but not bounded. This means there is not a constant C with $\|A\varphi\| \leqslant C\|\varphi\|$ for all $\varphi \in X$. Thus for every constant $C = n$ we can find φ_n such that the estimate is violated, i.e. such that $\|A\varphi_n\| > n\|\varphi_n\|$. We define $\psi_n := \varphi_n/\|A\varphi_n\|$. Then by construction we have $\|\psi_n\| \to 0$, $n \to \infty$. Since A is continuous we have $A\psi_n \to 0$. But this is a contradiction to $\|A\psi_n\| = 1$ for all $n \in \mathbb{N}$. Thus A must be bounded. $\square$

We close the subsection with some remarks about the *space of square summable sequences*

$$\ell^2 := \left\{ a = (a_j)_{j\in\mathbb{N}} : \text{each } a_j \in \mathbb{R}, \ \sum_{j=1}^{\infty} \left|a_j\right|^2 < \infty \right\}. \tag{2.3.28}$$

Equipped with the scalar product

$$\langle a, b \rangle := \sum_{j=1}^{\infty} a_j \overline{b_j} \tag{2.3.29}$$

it is a Hilbert space. The canonical Fourier basis of ℓ^2 is given by the elements $e^{(j)} \in \ell^2$ defined by

$$e_k^{(j)} := \begin{cases} 1 & j = k \\ 0 & \text{otherwise.} \end{cases} \tag{2.3.30}$$

In ℓ^2 the Cauchy–Schwarz inequality

$$\left| \sum_{j=1}^{\infty} a_j \overline{b_j} \right| \leqslant \left(\sum_{j=1}^{\infty} |a_j|^2 \right)^{\frac{1}{2}} \left(\sum_{j=1}^{\infty} |b_j|^2 \right)^{\frac{1}{2}} \tag{2.3.31}$$

is proven analogously to theorem 2.2.2.

Consider a linear operator $L : \ell^2 \to \mathbb{C}$. We define $l = (l_j)_{j\in\mathbb{N}}$ with $l_j := L(e^{(j)}) \in \mathbb{C}$. Then we obtain

$$L(a) = L\left(\sum_{j=1}^{\infty} a_j e^{(j)} \right) = \sum_{j=1}^{\infty} a_j L(e^{(j)}) = \sum_{j=1}^{\infty} l_j a_j = \langle a, \bar{l} \rangle. \tag{2.3.32}$$

On the other hand, every sequence $l = (l_j)_{j\in\mathbb{N}}$ with $l_j \in \mathbb{R}$ defines a linear operator L on ℓ^2 by

$$L(a) := \sum_{j=1}^{\infty} l_j a_j = \langle a, \bar{l} \rangle. \tag{2.3.33}$$

Theorem 2.3.10. *A linear operator L on ℓ^2 with corresponding sequence $l = (l_j)_{j\in\mathbb{N}}$ is bounded if and only if $l \in \ell^2$.*

Proof. If $l = (l_j)_j \in \mathbb{N}$ is bounded in ℓ^2, then by (2.3.31) applied to (2.3.32) the linear operator L is bounded. On the other side, if L is bounded, then there is a constant C such that $|L(a)| < C$ for all $a \in \ell^2$ with $\|a\|_{\ell^2} < 1$. Assume that $l \notin \ell^2$, i.e.

$$\sum_{j=1}^{\infty} |l_j|^2 = \infty. \tag{2.3.34}$$

We define

$$\rho_n := \sum_{j=1}^{n} |l_j|^2 \tag{2.3.35}$$

and set

$$a^{(n)} = \left(a_j^{(n)} \right)_{j\in\mathbb{N}} \text{ with } a_j^{(n)} := \frac{1}{\sqrt{\rho_n}} \cdot \begin{cases} l_j, & j \leqslant n \\ 0, & \text{otherwise.} \end{cases} \tag{2.3.36}$$

Then,

$$\|a^{(n)}\|^2 = \frac{1}{\rho_n}\sum_{j=1}^{n}\left|l_j\right|^2 = 1 \tag{2.3.37}$$

and further

$$L(a^{(n)}) = \sum_{j=1}^{\infty}\bar{l}_j\, a_j^{(n)} = \sum_{j=1}^{n}\bar{l}_j l_j = \rho_n \to \infty \tag{2.3.38}$$

for $n \to \infty$, in contradiction to $|L(a^{(n)})| \leqslant C$. This means that for bounded L the corresponding sequence l needs to be in ℓ^2 and the proof is complete. □

2.3.2 The solution of equations of the second kind and the Neumann series

In the introduction we discussed integral equations of the form

$$(I - A)\varphi = f.$$

Integral equations of this form are called *integral equations of the second kind*. We will develop two different solution approaches to such integral equations, the first for a bounded linear operator A for which the norm is sufficiently small. The second approach consists of Riesz theory for the case when A is compact, which will be defined in the next subsection.

First, recall the meaning of convergence of a series in a normed space as described in (2.2.36). Clearly, this also applies to operator series, i.e. to sums

$$\sum_{k=1}^{\infty} A_k \tag{2.3.39}$$

where $A_k : X \to Y$, $k \in \mathbb{N}$, are bounded linear operators. Here, we will restrict our attention to the case where $A : X \to X$ with a Banach space X, i.e. the operator maps a Banach into itself. In this case we can define *powers* of the operator A recursively via

$$A^k := A(A^{k-1}), \quad k = 1, 2, 3, \ldots \quad A^0 = I. \tag{2.3.40}$$

A simple example is given by the multiple application of a square matrix in $\mathbb{R}^{n\times n}$.

Example 2.3.11. *Consider an integral operator*

$$(A\varphi)(x) := \int_a^b k(x, y)\varphi(y)\,\mathrm{d}y, \quad x \in [a, b] \tag{2.3.41}$$

with continuous kernel $k : [a, b] \times [a, b] \to \mathbb{R}$. *Such an integral operator is a bounded linear operator in* $C([a, b])$.

In analogy to the geometric series we proceed as follows. First we observe that if $q := \|A\| < 1$, then

$$\|A^k\| \leqslant \|A\|^k = q^k \to 0, \quad k \to \infty. \tag{2.3.42}$$

We define the partial sum

$$S_n := \sum_{k=0}^{n} A^k. \tag{2.3.43}$$

For $n > m$, $m, n \in \mathbb{N}$ we estimate

$$\|S_n - S_m\| = \left\| \sum_{k=m+1}^{n} A^k \right\| \leqslant \sum_{k=m+1}^{n} q^k. \tag{2.3.44}$$

We use

$$(1 - q)\sum_{k=0}^{n} q^k = 1 - q^{n+1} \tag{2.3.45}$$

to calculate

$$\sum_{k=0}^{n} q^k = \frac{1 - q^{n+1}}{1 - q} \tag{2.3.46}$$

and

$$\sum_{k=m+1}^{n} q^k = \frac{q^{m+2} - q^{n+1}}{1 - q}. \tag{2.3.47}$$

For $q < 1$ we have $q^l \to 0$, for $l \to \infty$, thus S_n is a Cauchy sequence in $L(X, X)$. Since $L(X, X)$ is a Banach space, we obtain convergence of (S_n) towards some element $S \in L(X, X)$, i.e. a mapping from X into X. Imitating (2.3.45) for the operators A we obtain

$$(I - A)S_n = (I - A)\sum_{k=0}^{n} A^k = 1 - A^{n+1}. \tag{2.3.48}$$

By norm estimates we have $A^{n+1} \to 0, n \to \infty$, thus S_n converges towards the inverse $(I - A)^{-1}$ of $I - A$. We collect all results and estimates in the following theorem.

Theorem 2.3.12. *Consider a bounded linear operator $A : X \to X$ on a Banach space X. If the norm of A satisfies $\|A\| < 1$, then the operator $I - A$ has a bounded inverse on X which is given by the* Neumann series

$$(I - A)^{-1} = \sum_{k=0}^{\infty} A^k. \tag{2.3.49}$$

It satisfies the estimates

$$\|(I - A)^{-1}\| = \frac{1}{1 - \|A\|}. \tag{2.3.50}$$

In this case for each $f \in X$ *the integral equation of the second kind*

$$(I - A)\varphi = f \tag{2.3.51}$$

does have a unique solution $\varphi \in X$ *which in the norm depends continuously on the right-hand side* f.

The Neumann series can be rewritten in a nice form which is well known in numerical mathematics. We transform the series solution

$$\varphi_n := \sum_{k=0}^{n} A^k f \tag{2.3.52}$$

into the form

$$\varphi_n = \sum_{k=1}^{n} A^k f + f = A \sum_{k=1}^{n} A^{k-1} f + f = A\varphi_{n-1} + f. \tag{2.3.53}$$

This leads to:

Theorem 2.3.13. *Consider a bounded linear operator* $A : X \to X$ *on a Banach space* X. *If the norm of* A *satisfies* $\|A\| < 1$, *then the sequence* φ_n *of* successive approximations

$$\varphi_{n+1} := A\varphi_n + f, \quad n = 0, 1, 2, \ldots \tag{2.3.54}$$

with starting value $\varphi_0 = 0$ *converges to the unique solution* φ *of* $(I - A)\varphi = f$.

The Neumann series and successive approximations are very beautiful tools which are used often in all areas of mathematics. However, they have drawbacks. The theory applied only for $\|A\| < 1$. This condition, however, is violated often for the integral equations which arise from applications. This was a basic and important problem for mathematics until around 1900, when Fredholm, Riesz and Hilbert made significant progress in the treatment of such integral equations. We will study these and for this purpose next investigate *compact* operators.

2.3.3 Compact operators and integral operators

We first start with a general definition of compact operators.

Definition 2.3.14. *Let* $A : X \to Y$ *be a linear operator between normed spaces* X *and* Y. *Then* A *is called* compact *if it maps each bounded set in* X *into a relatively compact set in* Y.

This is a topological definition. Next, we provide an equivalent definition which uses sequences and which will be used again and again in the following arguments. Note that the topological definition has more potential to pass into weak topologies and deeper investigations, but for the beginning and in the framework of normed spaces using sequences is often easier.

Theorem 2.3.15. *A linear operator $A : X \to Y$ between normed spaces X and Y is compact if and only if for each bounded sequence $(\varphi_n) \subset X$ the image sequence $(A\varphi_n)$ contains a convergent subsequence in Y.*

Remark. Note that the image sequence $(A\varphi_n)$ is converging in Y, i.e. the limit element or limit point is an element of Y. If the limit is in a larger space, the completion of Y for example, then the operator is not compact.

Proof. One direction of the equivalence is very quick. If A is compact and (φ_n) is a bounded sequence, then $V := \{A\varphi_n : n \in \mathbb{N}\}$ is relatively compact, i.e. by the series arguments for compact sets its closure $\bar{V}$ is compact. Thus there is a convergent subsequence which converges towards an element of $\bar{V} \subset Y$.

The other direction is slightly more involved. Assume that there is a bounded set $U \subset X$ such that $V = \{A\varphi : \varphi \in U\}$ is not relatively compact, i.e. $\bar{V}$ is not compact. Then there is a sequence in $\bar{V}$ which does not have a convergent subsequence and by approximation arguments there is a sequence (ψ_n) in V which does not have a convergent subsequence. Thus there is a sequence (φ_n) with $\psi_n = A\varphi_n$ in the bounded set U for which $A\varphi_n$ does not have a convergent subsequence. This means that A is not sequentially compact. As a consequence sequentially compact operators must be compact and the proof is complete. □

We next collect basic properties of compact operators.

Theorem 2.3.16. *For compact linear operators we have that:*

(a) *Compact linear operators are bounded.*

(b) *If A, B are compact linear operators and $\alpha, \beta \in \mathbb{R}$ or $\mathbb{C}$, then $\alpha A + \beta B$ is a compact linear operator.*

(c) *Let X, Y, Z be normed spaces, $A : X \to Y$ and $B : Y \to Z$ bounded linear operators. If either A or B is compact, then the product $B \circ A : X \to Z$ is compact.*

Proof.

(a) Relatively compact sets are bounded, thus a compact operator maps the bounded set $B[0, 1]$ into a set bounded by a constant C, which yields $\|A\varphi\| \leqslant C$ for all $\|\varphi\| \leqslant 1$, therefore A is bounded.

(b) This can be easily seen by the definition of compact operator.

(c) Let us consider the case where A is bounded and B is compact. Consider a bounded sequence (φ_n) in X. Then the sequence (ψ_n) with $\psi_n := A\varphi_n$ is bounded in Y. Compactness of B yields the existence of a convergent subsequence of $(B\psi_n)$ in Z. Thus, for every bounded sequence (φ_n) in X the sequence $\chi_n := B(A(\varphi))$ in Z has a convergent subsequence and thus $B \circ A$ is a compact operator.

The case where A is compact and B is bounded works analogously. □

So far we have studied compact operators with series arguments. We now come to a quite important and far reaching result, which will help us to better understand the role of compact operators. They are basically the closure of the set of finite-dimensional operators.

Theorem 2.3.17. *Let X be a normed space and Y a Banach space. If a sequence $A_n : X \to Y$ of compact linear operators is norm convergent towards an operator $A : X \to Y$, then A is compact.*

Remark. Recall that norm convergence means $\|A_n - A\| \to 0$ for $n \to \infty$. Pointwise convergence is not sufficient to obtain this result!

Proof. The proof uses some important standard tools. Let (φ_k) be a bounded sequence in X with $\|\varphi_k\| \leqslant C$ for $k \in \mathbb{N}$. We want to show that $(A\varphi_k)$ has a convergent subsequence. To this end we use the standard *diagonalization procedure*. First, we choose a subsequence $(\varphi_{k(l)})$ of (φ_k) such that $(A_1\varphi_{k(1,l)})$ converges. From this subsequence we choose a subsequence $(\varphi_{k(2,l)})$ again such that $(A_2\varphi_{k(l)})$ converges. We repeat this procedure for all $j = 3, 4, 5, \ldots$. We obtain a grid of sequences

$$\begin{pmatrix} \varphi_{k(1,1)} & \varphi_{k(1,2)} & \varphi_{k(1,3)} & \varphi_{k(1,4)} & \cdots \\ \varphi_{k(2,1)} & \varphi_{k(2,2)} & \varphi_{k(2,3)} & \varphi_{k(2,4)} & \cdots \\ \varphi_{k(3,1)} & \varphi_{k(3,2)} & \varphi_{k(3,3)} & \varphi_{k(3,4)} & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \end{pmatrix} \tag{2.3.55}$$

where each row is a subsequence of the upper rows and for every $n \geqslant j$ the sequence $(A_j\varphi_{k(n,l)})$ arising from an application of A_j to the sequence in row n converges. We now choose the diagonal sequence

$$\psi_j := \varphi_{k(j,j)}, \quad j = 1, 2, 3, \ldots \tag{2.3.56}$$

Since $(\psi_j)_{j\geqslant n}$ is a subsequence of row n we know that for arbitrary fixed $n \in \mathbb{N}$ the sequence $(A_n\psi_j)_{j\in\mathbb{N}}$ converges for $j \to \infty$.

Next, we show that $(A\psi_j)$ is a Cauchy sequence. To this end choose $\epsilon > 0$. First, we find N_1 such that $\|A - A_n\| \leqslant \epsilon/3C$ for $n \geqslant N_1$. Second, since $(A_{N_1}\psi_j)$ converges there is $N_2 \in \mathbb{N}$ such that

$$\left\| A_{N_1}\psi_j - A_{N_1}\psi_l \right\| \leqslant \frac{\epsilon}{3}, \quad j, l \geqslant N_2. \tag{2.3.57}$$

Now, we estimate via the triangle inequality

$$\begin{aligned} \left\| A\psi_j - A\psi_l \right\| &\leqslant \left\| A\psi_j - A_{N_1}\psi_j \right\| + \left\| A_{N_1}\psi_j - A_{N_1}\psi_l \right\| + \left\| A_{N_1}\psi_l - A\psi_l \right\| \\ &\leqslant \frac{\epsilon}{3C} \cdot C + \frac{\epsilon}{3} + \frac{\epsilon}{3C} \cdot C \\ &= \epsilon \end{aligned} \tag{2.3.58}$$

for all $j, l \geqslant N_2 = N_2(\epsilon)$. Hence $(A\psi_j)$ is a Cauchy sequence and thus convergent in the Banach space Y. This shows that A is compact.

To complete our claim from above we need to study finite-dimensional operators.

Theorem 2.3.18. *A bounded operator $A : X \to Y$ with finite-dimensional range $A(X)$ is compact.*

Proof. Consider a bounded set $U \in X$. It is mapped into a bounded set $A(U)$ in the finite-dimensional range $A(X)$, which is norm isomorphic to $\mathbb{R}^m$. But according to the Bolzano–Weierstrass theorem 2.1.23 in the form of theorem 2.1.25 this set $A(U)$ is relatively compact and thus A is compact. □

We want to complete this subsection with a result about the identity operator, which is important for many linear inverse problems. However, we need a well-known lemma as preparation.

Lemma 2.3.19 (Riesz). *Let X be a normed space and $U \underset{\neq}{\subset} X$ a closed subspace. Then for any $\alpha \in (0, 1)$ there exists an element $\psi \in X$ such that*

$$\|\psi\| = 1, \quad \|\psi - \varphi\| \geqslant \alpha, \quad \varphi \in U. \tag{2.3.59}$$

Proof. There is an element $f \in X$ with $f \notin U$. Since U is closed we know that

$$\beta := \inf_{\varphi \in U} \|f - \varphi\| > 0. \tag{2.3.60}$$

Choose an element $g \in U$ which is close to the minimal distance, in particular with

$$\beta \leqslant \|f - g\| \leqslant \frac{\beta}{\alpha}. \tag{2.3.61}$$

Then define a unit vector ψ by

$$\psi := \frac{f - g}{\|f - g\|}. \tag{2.3.62}$$

For all $\varphi \in U$ we have the estimate

$$\begin{aligned} \|\psi - \varphi\| &= \frac{1}{\|f - g\|} \left\| f - \underbrace{g - \|f - g\|\varphi}_{\in U} \right\| \\ &\geqslant \frac{\beta}{\|f - g\|} \geqslant \alpha \end{aligned} \tag{2.3.63}$$

which completes the proof. □

Theorem 2.3.20. *The identity operator $I : X \to X$ is compact if and only if X has finite dimension.*

Proof. If X has finite dimension, then clearly I is compact, since every bounded set is mapped onto itself and it is relatively compact according to theorem 2.1.25. Next, consider the case where X is not finite-dimensional. We will construct a bounded sequence which does not have a convergent subsequence, which then shows that the identity operator cannot be compact. We start with an arbitrary $\varphi_1 \in X$ with $\|\varphi_1\| = 1$. Then $U_1 := \text{span}\{\varphi_1\}$ is a closed linear subspace. According to the Riesz lemma 2.3.19 there is an element $\varphi_2 \in X$ with $\|\varphi_2\| = 1$ and

$$\|\varphi_2 - \varphi_1\| \geqslant \frac{1}{2}. \tag{2.3.64}$$

Now, we define $U_2 := \text{span}\{\varphi_1, \varphi_2\}$. Again, we find $\varphi_3 \in X$ with $\|\varphi_3\| = 1$ and

$$\|\varphi_3 - \varphi_k\| \geqslant \frac{1}{2}, \quad k = 1, 2. \tag{2.3.65}$$

We proceed in the same way to obtain a sequence (φ_n) for which

$$\|\varphi_n\| = 1 \ (n \in \mathbb{N}), \quad \|\varphi_n - \varphi_m\| \geqslant \frac{1}{2}, \quad n \neq m. \tag{2.3.66}$$

This sequence cannot have a convergent subsequence, and the proof is complete. □

We finish this subsection with examples.

Example 2.3.21. *The integral operator (2.3.41) with a continuous kernel is a compact operator on $C([a, b])$. This can be obtained via theorems 2.3.17 and 2.3.18 as follows. First, we construct a polynomial approximation*

$$p_n(x, y) := \sum_{j,k=0}^{n} a_{j,k} x^j y^k, \quad x, y \in [a, b] \tag{2.3.67}$$

to the kernel function $k(x, y)$. According to the Weierstrass approximation theorem given ϵ there is n such that

$$\sup_{x,y \in [a,b]} \left| k(x, y) - p_n(x, y) \right| \leqslant \epsilon \tag{2.3.68}$$

for n sufficiently large. We define the approximation operator

$$(A_n \varphi)(x) := \int_a^b p_n(x, y) \varphi(y) \, \mathrm{d}y, \quad x \in [a, b]. \tag{2.3.69}$$

The range of the operator A_n is the finite-dimensional polynomial space Π_n, since integration with respect to y leads to a polynomial in x of degree n. We have the norm estimate

$$\|A_n - A\|_\infty \leqslant \sup_{x \in [a,b]} \int_a^b \left| k(x, y) - p_n(x, y) \right| \mathrm{d}y, \tag{2.3.70}$$

which can be made arbitrarily small for sufficiently large $n \in \mathbb{N}$. Thus we have approximated the operator A in the norm by a finite-dimensional and thus compact operator A_n. Hence A is compact.

A compact operator $A : X \to Y$ on an infinite-dimensional space X cannot have a bounded inverse, since otherwise $A^{-1} \circ A = I$ would be compact, which is not the case for X of infinite dimension. This is an important conclusion. Together with the previous example it shows that the integral equation of the first kind

$$\int_a^b k(x, y)\varphi(y)\,\mathrm{d}y = f(x), \quad x \in [a, b] \tag{2.3.71}$$

with continuous kernel $k(x, y)$ cannot be boundedly invertible! Here we discover the phenomenon of *ill-posedness*, which we will study in detail later, see section 3.1.1.

2.3.4 The solution of equations of the second kind and Riesz theory

We now study the solution of the integral equation

$$(I - A)\varphi = f \tag{2.3.72}$$

of the second kind with a compact operator A on a normed space. For more than 150 years there was no solution theory for this type of equation, which naturally appear in many applications arising from fluid dynamics, acoustic or electromagnetic waves, potential theory and, since 1920, in quantum mechanics. For most equations from nature the Neumann series results were not applicable. The important breakthrough was made by Fredholm in 1900. Here, we will use the related results of Riesz, which are valid in general normed spaces.

We will survey the three famous theorems of Riesz and an important conclusion. We will not present all proofs here but instead refer to the literature [1]. We define

$$L := I - A \tag{2.3.73}$$

where as usual I denotes the identity operator. Recall the *nullspace*

$$N(L) = \{\varphi \in X : L\varphi = 0\} \tag{2.3.74}$$

and its range

$$L(X) = \{L\varphi : \varphi \in X\}. \tag{2.3.75}$$

Theorem 2.3.22 (First Riesz theorem). *The nullspace of the operator L is a finite-dimensional subspace of X.*

Theorem 2.3.23 (Second Riesz theorem). *The range $L(X)$ of the operator L is a closed linear subspace.*

Theorem 2.3.24 (Third Riesz theorem). *There exists a uniquely determined integer $r \in \mathbb{N}_0$, such that*

$$\begin{aligned} &\text{(i) } \{0\} = N(L^0) \not\subset N(L^1) \not\subset \ldots \not\subset N(L^r) = N(L^{r+1}) = \ldots \\ &\text{(ii) } X = L^0(X) \not\subset L^1(X) \not\subset \ldots \not\subset L^r(X) = L^{r+1}(X) = \ldots.. \end{aligned} \tag{2.3.76}$$

The integer r is called the Riesz number *of the operator A and it is important to know that either (i) or (ii) can determine r. Furthermore, we have*

$$X = N(L^r) \oplus L^r(X). \tag{2.3.77}$$

The Riesz theory has an important application to the solution of (2.3.72).

Theorem 2.3.25 (Riesz). *Let X be a normed space and $A : X \to X$ be a compact linear operator. If $I - A$ is injective, then the inverse $(I - A)^{-1} : X \to X$ exists and is bounded.*

Proof. Let $L = I - A$. By the assumption the operator satisfies $N(L) = \{0\}$. Thus the Riesz number is $r = 0$. From the third Riesz theorem we conclude that $L(X) = X$, i.e. the operator is surjective. The inverse operator $(I - A)^{-1}$ exists. To show that L^{-1} is bounded we assume that this is not the case. Then there is a sequence (f_n) with $\|f_n\| = 1$ such that $(L^{-1}f_n)$ is unbounded. We define

$$\varphi_n := L^{-1}f_n, \qquad g_n := \frac{f_n}{\|\varphi_n\|}, \qquad \psi_n := \frac{\varphi_n}{\|\varphi_n\|} \tag{2.3.78}$$

for $n \in \mathbb{N}$. We conclude that $g_n \to 0$, $n \to \infty$ and $\|\psi_n\| = 1$. A is compact, hence we can choose a subsequence $(\psi_{n(k)})$ such that $(A\psi_{n(k)})$ converges towards $\psi \in X$. By construction we have

$$\psi_{n(k)} - A\psi_{n(k)} = g_{n(k)}, \quad k \in \mathbb{N}, \tag{2.3.79}$$

thus we obtain $\psi_{n(k)} \to \psi$, $k \to \infty$. However, then $\psi \in N(L)$, i.e. $\psi = 0$. But this contradicts $\|\psi_n\| = 1$. Thus L^{-1} must be bounded. □

With the Riesz theory we are able to solve integral equations which arise in many practical problems in the theory of waves. Acoustics, electromagnetics and elasticity lead to a variety of integral equations of the form (2.3.72).

2.4 Adjoint operators, eigenvalues and singular values

The concept of adjoint operators and adjoint problems is of far reaching importance for direct and inverse problems. Studying adjoint relationships between inversion algorithms will be an important part of chapters 12 and 15. The *tangent linear adjoint* solution is the key ingredient for state-of-the-art data assimilation algorithms, see section 5.3.1.

2.4.1 Riesz representation theorem and adjoint operators

In this first subsection we consider a linear bounded operator $A : X \to Y$ from a Hilbert space X into a Hilbert space Y. Recall that bounded linear functionals F are linear mappings $X \to \mathbb{R}$ or $X \to \mathbb{C}$

$$F(\alpha\varphi + \beta\eta) = \alpha F(\varphi) + \beta F(\eta) \tag{2.4.1}$$

for $\varphi, \eta \in X$, $\alpha, \beta \in \mathbb{R}$ or $\mathbb{C}$ with

$$\sup_{\|\varphi\| \leqslant 1} |F(\varphi)| < \infty. \tag{2.4.2}$$

Consider a bounded linear mapping $G : \mathbb{R} \to \mathbb{R}$. What do these mappings look like? We have $G(\alpha) = \alpha G(1) = G(1) \cdot \alpha$ for all $\alpha \in \mathbb{R}$. This means that we can write the mapping as a multiplication with the real number $G(1)$. In two dimensions $G : \mathbb{R}^2 \to \mathbb{R}$ we obtain

$$\begin{aligned} G(x) &= G(x_1 e_1 + x_2 e_2) = G(e_1)x_1 + G(e_2)x_2 \\ &= \begin{pmatrix} G(e_1) \\ G(e_2) \end{pmatrix} \cdot \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = g \cdot x \end{aligned} \tag{2.4.3}$$

with a vector $g \in \mathbb{R}^2$ for every $x = (x_1, x_2) \in \mathbb{R}^2$. This result can be obtained in a very general form in a Hilbert space.

Theorem 2.4.1 (Riesz representation). *Consider a Hilbert space X. Then for each bounded linear function $F : X \to \mathbb{C}$ there is a uniquely determined element $f \in X$ such that*

$$F(\varphi) = \langle \varphi, f \rangle, \quad \varphi \in X, \quad \|F\| = \|f\|. \tag{2.4.4}$$

Proof. To show the uniqueness assume that there are two elements $f_1, f_2 \in X$ with $\langle \varphi, f_1 \rangle = \langle \varphi, f_2 \rangle$ for all $\varphi \in X$. This yields $\langle \varphi, f_1 - f_2 \rangle = 0$ for all $\varphi \in X$, in particular for $\varphi = f_1 - f_2$, and thus $\|f_1 - f_2\| = \langle f_1 - f_2, f_1 - f_2 \rangle = 0$, from which we obtain $f_1 = f_2$.

To show the existence of such a function we apply the best approximation theorems from section 2.2.2. If $F = 0$, i.e. $F(\varphi) = 0$ for all $\varphi \in X$, then $f = 0$ is the right element. If $F \neq 0$ there is at least one $w \in X$ for which $F(w) \neq 0$. Further, we remark that the nullspace

$$N(F) = \left\{ \varphi \in X : F(\varphi) = 0 \right\} \tag{2.4.5}$$

is a closed subspace of X by the continuity of F. Thus according to theorems 2.2.10 and 2.2.11 there is a best approximation v to w in $N(F)$, which satisfies

$$w - v \perp N(F) \Leftrightarrow \langle w - v, \psi \rangle = 0, \quad \psi \in N(F). \tag{2.4.6}$$

Now with $g := w - v$ we observe that $\psi := F(g)\varphi - F(\varphi)g \in N(F)$ for arbitrary elements $\varphi \in X$, thus

$$\langle F(g)\varphi - F(\varphi)g, g \rangle = \overline{\langle g, F(g)\varphi - F(\varphi)g \rangle} = \overline{\langle w - v, \psi \rangle} = 0. \tag{2.4.7}$$

Now using the linearity of the scalar product in the first and the anti-linearity in the second component this yields

$$F(\varphi) = \left\langle \varphi, \frac{\overline{F(g)}g}{\|g\|^2} \right\rangle, \quad \varphi \in X, \tag{2.4.8}$$

such that $f := \overline{F(g)}g/\|g\|^2$ is the function with the desired properties. □

Combining this with theorem 2.2.11 we immediately have the following.

Corollary 2.4.2. *Let M be a subspace in a Hilbert space X. Then M is dense in X if and only if we have the following condition for any bounded linear function $F : X \to \mathbb{C}$:*

$$F(\varphi) = 0, \quad \varphi \in M \Longrightarrow F = 0. \tag{2.4.9}$$

We are now able to define adjoint operators. For matrices in $\mathbb{R}^m$ you know the transpose A^T of a matrix A already, defined as the matrix where the first row becomes the first column, the second row the second column and so forth. For vectors $x, y \in \mathbb{R}^m$ and a matrix $A \in \mathbb{R}^{n\times n}$ we have

$$x \cdot Ay = \sum_{j=1}^{n} x_j\left(\sum_{k=1}^{n} a_{jk}y_k\right) = \sum_{k=1}^{n}\left(\sum_{j=1}^{n} a_{jk}x_j\right) y_k = (A^T x) \cdot y. \tag{2.4.10}$$

This can also be extended into a Hilbert space setting. To make the distinction between the scalar products in the Hilbert spaces X and Y we sometimes use the notation $\langle\cdot,\cdot\rangle_X$ and $\langle\cdot,\cdot\rangle_Y$. Let $A \in BL(X, Y)$. Then, for each fixed element $\psi \in Y$ the mapping

$$\varphi \mapsto \langle A\varphi, \psi\rangle_Y \in \mathbb{C} \tag{2.4.11}$$

defines a bounded linear functional on X. Thus by the Riesz theorem 2.4.1 there is $f \in X$ such that $\langle A\varphi, \psi\rangle = \langle\varphi, f\rangle$ for all $\varphi \in X$. This defines a mapping $A^*\psi := f$ from Y into X. This operator is uniquely determined. We have prepared the following definition.

Definition 2.4.3. *For an operator $A \in BL(X, Y)$ between two Hilbert spaces X and Y the* adjoint *operator is uniquely defined via*

$$\langle A\varphi, \psi\rangle_Y = \langle\varphi, A^*\psi\rangle_X, \quad \varphi \in X, \quad \psi \in Y. \tag{2.4.12}$$

Adjoint operators are very important for the understanding of a variety of phenomena in the natural sciences. First, we remark that

$$(A^*)^* = A, \tag{2.4.13}$$

which is a consequence of

$$\langle\psi, (A^*)^*\varphi\rangle_Y = \langle A^*\psi, \varphi\rangle_X = \overline{\langle\varphi, A^*\psi\rangle_X} = \overline{\langle A\varphi, \psi\rangle_Y} = \langle\psi, A\varphi\rangle_Y$$

for all $\varphi \in X$, $\psi \in Y$. Further, we calculate

$$\|A^*\psi\|^2 = \langle A^*\psi, A^*\psi\rangle = \langle AA^*\psi, \psi\rangle \leqslant \|A\|\|A^*\psi\|\|\psi\| \tag{2.4.14}$$

for all $\psi \in Y$ via the Cauchy–Schwarz inequality theorem 2.2.2. Division by $\|A^*\psi\|$ yields the boundedness of the adjoint operator A^* with $\|A^*\| \leqslant \|A\|$. Via (2.4.13) we can exchange the roles of A and A^*, i.e. we have $\|A\| \leqslant \|A^*\|$. We summarize the results in the following theorem.

Theorem 2.4.4. *The norm of the adjoint operator A^* is equal to the norm of its dual A, i.e. we have*

$$\|A^*\| = \|A\|. \tag{2.4.15}$$

Example 2.4.5. *As an example we calculate the adjoint of an integral operator*

$$(A\varphi)(x) := \int_a^b k(x, y)\varphi(y)\,\mathrm{d}s(y), \quad x \in [c, d], \tag{2.4.16}$$

with continuous kernel k in the Hilbert spaces $X = L^2([a, b])$ *and* $Y = L^2([c, d])$ *with* $a < b$ *and* $c < d$ *in* $\mathbb{R}$. *Using Fubini's theorem with functions* $\varphi \in L^2([a, b])$ *and* $\psi \in L^2([c, d])$ *we derive*

$$\begin{aligned}
&\langle A\varphi, \psi\rangle \\
&= \int_c^d \left(\int_a^b k(x, y)\varphi(y)\,\mathrm{d}s(y)\right)\overline{\psi(x)}\,\mathrm{d}s(x) \\
&= \int_a^b \varphi(y)\left(\int_c^d k(x, y)\overline{\psi(x)}\,\mathrm{d}s(x)\right)\mathrm{d}s(y) \\
&= \int_a^b \varphi(y)\overline{\left(\int_c^d \overline{k(x, y)}\psi(x)\,\mathrm{d}s(x)\right)}\mathrm{d}s(y) \\
&= \langle \varphi, A^*\psi\rangle
\end{aligned} \tag{2.4.17}$$

with

$$(A^*\psi)(y) := \int_c^d \overline{k(x, y)}\psi(x)\,\mathrm{d}s(x), \quad y \in [a, b]. \tag{2.4.18}$$

This operator satisfies the equation (2.4.12) and thus is the uniquely determined adjoint operator. Thus for calculation of the adjoint we need to exchange the role of the kernel variables x and y and need to take the complex conjugate of the kernel.

Finally, we state the following without proof.

Theorem 2.4.6. *If* $A : X \to Y$ *is a compact linear operator between Hilbert spaces X and Y, then the adjoint operator* $A^* : Y \to X$ *is also compact.*

For the special case of the integral operator (2.4.16) this is an immediate consequence of example 2.3.21. For the general case, the theorem can be easily shown by using $AA^* : Y \to Y$ is compact and $\|A^*y\|_X^2 = \langle AA^*y, y\rangle_Y$ for $y \in Y$.

2.4.2 Weak compactness of Hilbert spaces

We have seen in the previous subsection that the Riesz representation theorem enables us to define the adjoint A^* of an operator $A \in BL(X, Y)$ for Hilbert spaces X and Y. Another interesting by-product of this theorem is the weak compactness of Hilbert space which means that every bounded sequence in a Hilbert space X has a subsequence which converges weakly in X. We first give the definition of weak convergence in X.

Definition 2.4.7. *Let X be a Hilbert space with inner product $\langle\cdot,\cdot\rangle$.*

(i) *A sequence (x_n) in X is weakly convergent to $x \in X$ if for any $y \in X$, $\langle x_n, y\rangle \to \langle x, y\rangle$ as $n \to \infty$. This will be denoted by $x_n \rightharpoonup x$, $n \to \infty$.*

(ii) *A sequence (x_n) in X is called weakly bounded if for any $y \in X$, the sequence $(\langle x_n, y\rangle)$ is bounded in $\mathbb{C}$.*

Theorem 2.4.8. *A sequence (x_n) in a Hilbert space X is weakly bounded if and only if (x_n) is bounded in X.*

Proof. Since it is clear that the boundedness of a sequence in X implies the weak boundedness of the sequence in X, we will prove the converse of this. Let a sequence (x_n) be weakly bounded in X. Then, we claim that there exist $x \in X, r > 0$ and $K > 0$ such that

$$|\langle x_n, y\rangle| \leqslant K, \quad n \in N, \quad y \in B(x, r). \tag{2.4.19}$$

In fact if we assume that this is not true, then using the continuity of the inner product, there exist sequences (n_j) in $\mathbb{N}$, (y_j) in X and (r_j) in $\mathbb{R}$ such that

(n_j) is monotone increasing and for any $2 \leqslant j \in \mathbb{N}$,

$$0 < r_j < 1/j, \quad B\Big[y_{j+1}, r_{j+1}\Big] \subset B\Big(y_j, r_j\Big), \quad \Big|\Big\langle x_{n_j}, y\Big\rangle\Big| > j \quad \text{for } y \in B\Big(y_j, r_j\Big).$$

Clearly $\|y_n - y_m\| < 1/m$ for $n > m$. Hence (y_n) is a Cauchy sequence in X. By the completeness of Hilbert space, there exists a unique limit $y \in X$ to this sequence. Since for any fixed $j \in \mathbb{N}$, $y_k \in B[y_j, r_j]$ for all $k > j$. By letting $k \to \infty$, y belongs to all $B[y_j, r_j]$. Hence, $|\langle x_{n_k}, y\rangle| > j$ for any $k > j$. This cannot happen because (x_n) is weakly bounded. Thus we have the claim.

By this claim, we have for any z ($\|z\| < 1$),

$$|\langle x_n, z\rangle| = \frac{1}{r}|\langle x_n, x + rz\rangle - \langle x_n, x\rangle| \leqslant \frac{2K}{r}. \tag{2.4.20}$$

Here we can even have this for any z ($\|z\| \leqslant 1$) by the continuity of the inner product. Hence, $\|x_n\| = \sup_{\|z\|\leqslant 1}|\langle x_n, z\rangle| \leqslant (2K)/r$. □

In order to prove the weak compactness of Hilbert spaces, we prepare a lemma which is very easy to prove.

Lemma 2.4.9. *Let X be a Hilbert space with inner product $\langle\cdot,\cdot\rangle$ and it has a dense subset D. Then, if $x \in X$ and a sequence (x_n) in X satisfies*

$$\langle x_n, y\rangle \to \langle x, y\rangle, \quad n \to \infty \ \text{ for all } y \in D, \tag{2.4.21}$$

then $x_n \rightharpoonup x$, $n \to \infty$, i.e. it converges weakly in X.

Now we state and prove the weak compactness of Hilbert spaces as follows.

Theorem 2.4.10 (Weak compactness of Hilbert spaces). *Let X be a Hilbert space. Then for any bounded sequence (x_n) in X, there exists a weakly convergent subsequence $(x_{n(k)})$ of (x_n).*

Proof. Let $D = \{y_n : n \in N\}$, $\langle\cdot,\cdot\rangle$ be the inner product of X and $\|x_n\| \leqslant C$ with some constant $C > 0$ for any $n \in \mathbb{N}$. Then, using the boundedness of (x_n) and the diagonal argument, there exists a subsequence $(x_{n(k)})$ of (x_n) such that for any $y \in \operatorname{span} D$, the limit $\lim_{k\to\infty}\langle x_{n(k)}, y\rangle =: f(y)$ exists. It is easy to see that f is anti-linear and satisfies $\|f\| \leqslant C$. Further, f can be extended to $\overline{\operatorname{span} D}$ preserving all its properties.

Now let $P : X \to \overline{\operatorname{span} D}$ be a orthogonal projection. Then, we can further extend f to the whole X without destroying its properties by considering the mapping $X \ni x \to f(Px)$. We will still use the same notation f for the extension of f. By the Riesz representation theorem, there exists a unique $x \in X$ such that $f(y) = \langle x, y\rangle$ for any $x \in X$. This gives

$$\left\langle x_{n(k)}, y\right\rangle \to \langle x, y\rangle, \quad k \to \infty, \quad y \in X. \tag{2.4.22}$$

□

2.4.3 Eigenvalues, spectrum and the spectral radius of an operator

In linear algebra we study linear mappings in $\mathbb{R}^m$, which can be expressed as matrices $A \in \mathbb{R}^{n\times n}$. A very simple subclass of matrices are diagonal matrices

$$\boldsymbol{D} = \operatorname{diag}\{d_1, \ldots, d_n\} = \begin{pmatrix} d_1 & 0 & \ldots & 0 \\ 0 & d_2 & 0 & \vdots \\ 0 & 0 & \ddots & 0 \\ 0 & \ldots & \ldots & d_n. \end{pmatrix} \tag{2.4.23}$$

These matrices are very easy for studying mappings, since on the unit basis vectors $e_1, \ldots, e_n$ the matrix $\boldsymbol{D}$ is just a multiplication operator

$$\boldsymbol{D} \circ (\alpha_j e_j) = \left(d_j \alpha_j\right) e_j, \quad j = 1, \ldots, n. \tag{2.4.24}$$

If we could find a basis $B = \{\varphi_1, \varphi_2, \varphi_3, ...\}$ in a space X such that for a matrix A we have

$$A\varphi_j = \lambda_j \varphi_j, \tag{2.4.25}$$

with real or complex numbers λ_j, $j = 1, \ldots, n$, then the application of the operator A would be reduced to simple multiplications on the basis functions.

The resolution of a matrix equation

$$A\varphi = f \tag{2.4.26}$$

with an invertible matrix A in such a basis is very simple. Assume that

$$f = \sum_{j=1}^{n} \beta_j \varphi_j. \tag{2.4.27}$$

Then the solution of (2.4.26) is given by

$$\varphi = \sum_{j=1}^{n} \frac{\beta_j}{\lambda_j} \varphi_j, \tag{2.4.28}$$

since

$$A\varphi = \sum_{j=1}^{n} \frac{\beta_j}{\lambda_j} A\varphi_j = \sum_{j=1}^{n} \frac{\beta_j}{\lambda_j} \lambda\varphi_j = \sum_{j=1}^{n} \beta_j \varphi_j = f. \tag{2.4.29}$$

In inverse problems we often find such matrices where λ_j becomes small for large j. In this case inversion is equivalent to the multiplication with large numbers $1/\lambda_j$. Small errors in the coefficients β_j will lead to large errors in the coefficients of the reconstructed quantity φ. This phenomenon is called *instability* of the inverse problem. In numerical mathematics we speak of *ill-conditioned* matrices.

We now collect adequate terms and definitions to treat the general situation in a Banach and Hilbert space setting.

Definition 2.4.11. *Consider a bounded linear operator A on a normed space X. We call $\lambda \in \mathbb{C}$ an* eigenvalue *of A, if there is an element $\varphi \in X$, $\varphi \neq 0$, such that*

$$A\varphi = \lambda\varphi. \tag{2.4.30}$$

In this case φ is called the eigenvector.

If λ is an eigenvalue of A, then the operator $\lambda I - A$ is not injective and thus cannot be invertible. In general, the operator might still not be invertible even if λ is not an eigenvalue of A. This motivates the following definition.

Definition 2.4.12. *If $\lambda I - A$ is boundedly invertible, then we call λ a* regular value *of A. The set $\rho(A) \subset \mathbb{C}$ of all regular values of A is called the* resolvent set, *its complement $\sigma(A) := \mathbb{C}\backslash\rho(A)$ is called the* spectrum *of A. The operator*

$$R(\lambda, A) := (\lambda I - A)^{-1} \tag{2.4.31}$$

is called the resolvent. *The* spectral radius $r(A)$ is the supremum

$$r(A) := \sup_{\lambda\in\sigma(A)} |\lambda|. \tag{2.4.32}$$

Note that since the set of eigenvalues is a subset of the spectrum, all eigenvalues lie in the ball with radius $r(A)$ around the origin in the complex plane $\mathbb{C}$.

For the spectrum of a compact operator we have the following result, which will be an important part of the basis of our study of inverse problems.

Theorem 2.4.13 (Spectrum of a compact operator). *Let $A : X \to X$ be a compact linear operator in an infinite-dimensional space X. Then $\lambda = 0$ is an element of the spectrum $\sigma(A)$ and $\sigma(A)\backslash\{0\}$ consists of at most a countable set of eigenvalues with no point of accumulation except, possibly, $\lambda = 0$.*

Proof. First, consider the case $\lambda = 0$. Then $\lambda = 0$ being regular is by definition equivalent to A being boundedly invertible and thus $I = A^{-1}A$ being compact. But since X is infinite-dimensional, the identity operator cannot be compact and thus λ cannot be regular.

For $\lambda \neq 0$ we apply the Riesz theory. Then in the case where the null space $N(\lambda I - A) = \{0\}$ the operator is injective and thus boundedly invertible. As a consequence λ is regular. Otherwise, we have $N(\lambda I - A) \neq \{0\}$, i.e. λ is an eigenvalue. Thus every point λ in $\mathbb{C}\backslash\{0\}$ is either a regular point or an eigenvalue.

Finally, we need to show that the set of eigenvalues $\sigma(A)\backslash\{0\}$ has no accumulation point other than zero. We do this by constructing sequences of subspaces spanned by the eigenvalues. Assume that we have a sequence (λ_n) of distinct eigenvalues and $\lambda \in \mathbb{C}\backslash\{0\}$ such that $\lambda_n \to \lambda$, $n \to \infty$ which implies $|\lambda_n| \geqslant R > 0$ $(n \in \mathbb{N})$. For each λ_n choose an associated eigenvector φ_n and define the subspaces

$$U_n := \operatorname{span}\{\varphi_1, \ldots, \varphi_n\}, \quad n \in \mathbb{N}. \tag{2.4.33}$$

It is not hard to see that eigenvectors of distinct eigenvalues are linearly independent. Thus $U_n \subset U_{n+1}$, $U_n \neq U_{n+1}$, $n = 1, 2, 3, \ldots$. Let $\psi_1 = \|\varphi_1\|^{-1}\varphi_1$. Then, by the Riesz lemma 2.3.19 we can choose a sequence $\psi_n \in U_n$ with $\|\psi_n\| = 1$ and

$$d(\psi_n, U_{n-1}) := \inf_{\psi \in U_{n-1}} \|\psi_n - \psi\| \geqslant \frac{1}{2}, \quad n \geqslant 2. \tag{2.4.34}$$

We can represent ψ_n as a linear combination of the basis elements, i.e.

$$\psi_n = \sum_{j=1}^{n} \alpha_{nj}\varphi_j \quad \text{with} \quad \alpha_{nj} \in \mathbb{C}\ (1 \leqslant j \leqslant n). \tag{2.4.35}$$

For $m < n$ we derive

$$A\psi_n - A\psi_m = \sum_{j=1}^{n} \lambda_j \alpha_{nj}\varphi_j - \sum_{j=1}^{m} \lambda_j \alpha_{mj}\varphi_j = \lambda_n(\psi_n - \psi) \tag{2.4.36}$$

with $\psi \in \operatorname{span}\{\varphi_1, \ldots, \varphi_{n-1}\} = U_{n-1}$. This yields

$$\|A\psi_n - A\psi_m\| \geqslant \frac{|\lambda_n|}{2} \geqslant \frac{R}{2}. \tag{2.4.37}$$

Thus the sequence $(A\psi_n)$ does not contain a convergent subsequence. However, (ψ_n) is a bounded sequence and A was assumed to be compact, such that we obtain a contradiction and our assumption was wrong. This completes the proof of the theorem. $\square$

2.4.4 Spectral theorem for compact self-adjoint operators

We are step-by-step approaching more general results about the representation of operators as multiplication operators in special basis systems. As a preparation we start with the following theorem, which will be an important tool both for the further theory as well as for inversion of linear operator equations in the framework of inverse problems.

Theorem 2.4.14. *For a bounded linear operator $A: X \to Y$ between Hilbert spaces X, Y we have*

$$A(X)^{\perp} = N(A^*), \quad N(A^*)^{\perp} = \overline{A(X)}, \tag{2.4.38}$$

where $N(A^*)$ and $A(X)$ are the null-space of A^* and range of A, respectively.

Proof. We first establish a sequence of equivalences as follows. We have

$$\begin{aligned} g \in A(X)^{\perp} &\Leftrightarrow \langle A\varphi, g\rangle = 0, \quad \varphi \in X \\ &\Leftrightarrow \langle \varphi, A^* g\rangle = 0, \quad \varphi \in X \\ &\Leftrightarrow A^* g = 0 \\ &\Leftrightarrow g \in N(A^*), \end{aligned} \tag{2.4.39}$$

which proves the first equality of spaces. To prove the second part denote by P the orthogonal projection in Y onto the closed subspace $\overline{A(X)}$. For function $\varphi \in (A(X)^{\perp})^{\perp}$, by using $A(X)^{\perp} = \overline{A(X)}^{\perp}$, we have $P\varphi - \varphi \in \overline{A(X)}^{\perp}$. Here, observe that for any $\psi \in \overline{A(X)}^{\perp}$,

$$\langle P\varphi - \varphi, \psi\rangle = \langle P\varphi, \psi\rangle - \langle \varphi, \psi\rangle = \langle \varphi, P\psi\rangle = 0.$$

which implies $\varphi \in \overline{A(X)}$. Hence, $(A(X)^{\perp})^{\perp} \subset \overline{A(X)}$. On the other hand, we have $(A(X)^{\perp})^{\perp} \subset \overline{A(X)}$ due to the fact that $U^{\perp} = \bar{U}^{\perp}$ for any subspace U of Y. Hence, we have obtained

$$(A(X)^{\perp})^{\perp} = \overline{A(X)}, \tag{2.4.40}$$

which together with the first result yields the second equality.

If we know the adjoint operator A^* and can evaluate its nullspace $N(A^*)$, via the previous theorem we also obtain the image space $\overline{A(X)}$. Exchanging the role of A and A^*, we can also use it to analyze the null-space of some operator A.

Definition 2.4.15. *Consider an operator $A: X \to X$ mapping a Hilbert space X into itself. It is called* self-adjoint *if $A = A^*$, i.e. if*

$$\langle A\varphi, \psi\rangle = \langle \varphi, A\psi\rangle, \quad \varphi, \psi \in X. \tag{2.4.41}$$

We first establish some basic equality between the norm and the quadratic form $\langle A\varphi, \varphi\rangle$ on X. In its proof we make use of the *parallelogram equality*

$$\|\varphi + \psi\|^2 + \|\varphi - \psi\|^2 = 2\left(\|\varphi\|^2 + \|\psi\|^2\right) \tag{2.4.42}$$

for all $\varphi, \psi \in X$, which is derived by elementary application of the linearity of the scalar product.

Theorem 2.4.16. *For a bounded linear self-adjoint operator A we have*

$$q := \sup_{\|\varphi\|=1} |\langle A\varphi, \varphi\rangle| = \|A\|. \tag{2.4.43}$$

Proof. Using the Cauchy–Schwarz inequality for $\varphi \in X$ with $\|\varphi\| = 1$ we have for any $\varphi, \psi \in X$

$$|\langle A\varphi, \varphi\rangle| \leqslant \|A\varphi\| \, \|\varphi\| \leqslant \|A\|, \tag{2.4.44}$$

from which we obtain $q \leqslant \|A\|$. The other direction is slightly tricky. First convince yourself by elementary calculation that we have

$$\begin{aligned} &\langle A(\varphi + \psi), \varphi + \psi\rangle - \langle A(\varphi - \psi, \varphi - \psi\rangle \\ &= 2\{\langle A\varphi, \psi\rangle + \langle A\psi, \varphi\rangle\} \\ &= 4 \operatorname{Re}\langle A\varphi, \psi\rangle. \end{aligned} \tag{2.4.45}$$

This yields the estimate

$$4\operatorname{Re}\langle A\varphi, \psi\rangle \leqslant q \cdot \left\{\|\varphi + \psi\|^2 + \|\varphi - \psi\|^2\right\} \leqslant 2q \cdot \left\{\|\varphi\|^2 + \|\psi\|^2\right\}, \tag{2.4.46}$$

where we have used (2.4.42). Finally, for φ with $\|\varphi\| = 1$ and $A\varphi \neq 0$, we define $\psi := A\varphi/\|A\varphi\|$ and calculate

$$\|A\varphi\| = \operatorname{Re} \frac{\langle A\varphi, A\varphi\rangle}{\|A\varphi\|} = \operatorname{Re}\langle A\varphi, \psi\rangle \leqslant q\frac{\|\varphi\|^2 + \|\psi\|^2}{2} = q. \tag{2.4.47}$$

Hence, $\|A\| \leqslant q$, and the proof is complete.

Theorem 2.4.17. *The eigenvalues of self-adjoint operators are real and eigenvectors to different eigenvalues are orthogonal to each other.*

Proof. First we remark that for an eigenvalue λ with eigenvector φ we have

$$\lambda\langle\varphi, \varphi\rangle = \langle A\varphi, \varphi\rangle = \langle\varphi, A\varphi\rangle = \bar{\lambda}\langle\varphi, \varphi\rangle, \tag{2.4.48}$$

hence $\lambda \in \mathbb{R}$. Let $\lambda \neq \mu$ be two eigenvalues of A with eigenvectors φ and ψ, respectively. Then we have

$$(\lambda - \mu)\langle\varphi, \psi\rangle = \langle A\varphi, \psi\rangle - \langle\varphi, A\psi\rangle = 0, \tag{2.4.49}$$

which yields $\varphi \perp \psi$ and the proof is complete.

Theorem 2.4.18. *Consider a bounded linear self-adjoint operator $A : X \to X$. Then its spectral radius is equal to its norm, i.e.*

$$r(A) = \|A\|. \tag{2.4.50}$$

If A is compact, then there is at least one eigenvalue λ with $|\lambda| = \|A\|$.

Proof. Since the theorem clearly holds for the case $A = 0$, we can assume $A \neq 0$. First, consider a complex number $\lambda \in \mathbb{C}$ with $|\lambda| > \|A\|$. Then we can study the operator $I - B$ with $B := \frac{1}{\lambda}A$. The norm of B satisfies

$$\|B\| = \frac{1}{|\lambda|}\|A\| < 1. \tag{2.4.51}$$

Now, by the Neumann series theorem 2.3.12 the operator $I - B$ is boundedly invertible, hence the same is true for $\lambda I - A = \lambda(I - B)$. Thus λ is a regular value, which yields $r(A) \leqslant \|A\|$.

By theorem 2.4.16 there is a sequence (φ_n) in X with $\|\varphi_n\| = 1$ such that

$$|\langle A\varphi_n, \varphi_n\rangle| \to \|A\|, \quad n \to \infty. \tag{2.4.52}$$

Since the ball with radius $\|A\|$ in $\mathbb{C}$ is compact, there is a λ with $|\lambda| = \|A\|$ such that

$$\langle A\varphi_n, \varphi_n\rangle \to \lambda, \quad n \to \infty. \tag{2.4.53}$$

By multiplication with a real number of modulus 1 to A we may achieve the situation where λ is positive. We now calculate

$$\begin{aligned} 0 \leqslant \|A\varphi_n - \lambda\varphi_n\|^2 &= \|A\varphi_n\|^2 - 2\lambda\langle A\varphi_n, \varphi_n\rangle + \lambda^2\|\varphi_n\|^2 \\ &\leqslant \|A\|^2 - 2\lambda\langle A\varphi_n, \varphi_n\rangle + \lambda^2 \\ &= 2\lambda(\lambda - \langle A\varphi_n, \varphi_n\rangle) \to 0, \quad n \to \infty. \end{aligned} \tag{2.4.54}$$

Therefore

$$(\lambda I - A)\varphi_n \to 0, \quad n \to \infty. \tag{2.4.55}$$

But in this case λ can not be a regular value of A, because

$$1 = \|\varphi_n\| = \|(\lambda I - A)^{-1}(\lambda I - A)\varphi_n\| \to 0, \quad n \to \infty \tag{2.4.56}$$

would generate a contradiction. This shows that $r(A) \geqslant |\lambda| = \|A\|$ and thus (2.4.50).

If A is compact, there exists a subsequence $(A\varphi_{n(k)})$ converging to some $\psi \in X$. Hence, by (2.4.55),

$$\lambda\varphi_{n(k)} \to \psi\ (k \to \infty), \text{ i.e. } \varphi_{n(k)} \to \lambda^{-1}\psi =: \varphi\ (k \to \infty). \tag{2.4.57}$$

Here, due to $\|\varphi_{n(k)}\| = 1\,(k \in \mathbb{N})$ we have $\varphi \neq 0$. Then, by using (2.4.55) once again, we have $A\varphi = \lambda\varphi$. This ends the proof. □

Example 2.4.19. *Consider a two point boundary value problem given by*

$$u''(x) + \kappa^2 u(x) = \varphi(x), \quad u(0) = u(1) = 0, \tag{2.4.58}$$

with $\varphi \in C([0, 1])$ *on* $\Omega := (0, 1)$, *where* $\kappa/(2\pi) > 0$ *is the frequency of vibration with small displacement which can be a vibration of a string with unit mass and cramped at its both ends* $x = 0$, $x = 1$ *under some small external force* $\varphi(x)$. *If* $\kappa/(2\pi)$ *is not a resonant frequency that is* $\kappa/(2\pi) \notin \mathbb{Z}_+ = \mathbb{N} \cup \{0\}$, *then this boundary value problem has a unique solution* $u \in C^2((0, 1) \cap C([0, 1])$ *and it is given by*

$$u(x) = (A\varphi)(x) = \int_\Omega G(x, \xi)\varphi(\xi)\,\mathrm{d}\xi, \ \ x \in [0, 1], \tag{2.4.59}$$

where $G(x, \xi)$ *is the so-called* Green's function *of* (2.4.58) *and it is given by*

$$G(x, \xi) = \begin{cases} \dfrac{-\sin\kappa\cos(\kappa x) + \cos\kappa\sin(\kappa x)}{\kappa\sin\kappa}\sin(\kappa\xi) & \text{if } 0 \leqslant \xi < x \leqslant 1 \\ \dfrac{\cos\kappa\sin(\kappa\xi) - \sin\kappa\cos(\kappa\xi)}{\kappa\sin\kappa}\sin(\kappa x) & \text{if } 0 \leqslant x < \xi \leqslant 0 \end{cases} \tag{2.4.60}$$

It is easy to see that $G(x, \xi) \in C^0([0, 1] \times [0, 1])$ *and* $G(x, \xi) = G(\xi, x)$ *for all* $x, \xi \in [0, 1]$, *which implies that the operator* A *defined by* (2.4.59) *is a compact self-adjoint operator on* $L^2((a, b))$.

In theorems 2.4.13 and 2.4.17 some spectral properties of compact self-adjoint operators have been given. Based on these properties, we are now prepared for the *spectral theorem for self-adjoint compact operators*. This is a big mile-stone in the study of operators and operator equations.

Theorem 2.4.20 (Spectral theorem). *Consider a non-zero compact self-adjoint operator* A *on a Hilbert space* X. *We order the sequence of non-zero eigenvalues* (λ_n) *according to its size such that*

$$|\lambda_1| \geqslant |\lambda_2| \geqslant |\lambda_3| \geqslant \ldots \tag{2.4.61}$$

with no point of accumulation except, possibly, $\lambda = 0$. *Denote the orthogonal projection operator onto the eigenspace* $N(\lambda_n I - A)$ *by* P_n *and let* $Q : X \to N(A)$ *be the orthogonal projection onto the null-space of* A. *Then we have the orthogonal decomposition*

$$X = \bigoplus_{n=1}^{\infty} N(\lambda_n I - A) \oplus N(A) \tag{2.4.62}$$

in the sense that each element $\varphi \in X$ *can be written as*

$$\varphi = \sum_{n=1}^{\infty} P_n\varphi + Q\varphi. \tag{2.4.63}$$

The operator A *can be represented as a multiplication operator by* λ_n *on the eigenspace* $N(\lambda_n I - A)$, *i.e. we have*

$$A = \sum_{n=1}^{\infty} \lambda_n P_n. \tag{2.4.64}$$

If there are only finitely many eigenvalues, then the series degenerate into finite sums.

Proof. We only consider the case where there are infinitely many eigenvalues, because the proof is easier for the other case. We first remark that the orthogonal projections P_n are self-adjoint and bounded with $\|P_n\| = 1$. Further, we have that $P_n P_k = 0$ for $n \neq k$ as a direct consequence of the orthogonality of the eigenspaces. Since their image space is the space $N(\lambda_n I - A)$ and by Riesz theory these spaces have finite dimension, the operators P_n are compact.

Consider the difference between A and the partial sum of (2.4.64), i.e. the operators

$$A_m := A - \sum_{n=1}^{m} \lambda_n P_n, \quad m \in \mathbb{N}. \tag{2.4.65}$$

We want to show that $A_m \to 0$, $m \to \infty$ and we carry this out by studying the eigenvalues of A_m. We show that the non-zero eigenvalues of A_m are given by $\lambda_{m+1}, \lambda_{m+2}, \ldots,$ Let $\lambda \neq 0$ be an eigenvalue of A_m with eigenvector φ. Then for $1 \leqslant n \leqslant m$ we have

$$\lambda P_n \varphi = P_n A_m \varphi = P_n(A\varphi - \lambda_n \varphi).$$

We use this to derive

$$\lambda^2 \|P_n \varphi\|^2 = \langle A\varphi - \lambda_n \varphi, P_n(A\varphi - \lambda_n \varphi)\rangle = \langle \varphi, (A - \lambda_n) P_n (A\varphi - \lambda_n \varphi)\rangle = 0$$

since $P_n = P_n^2$ is self-adjoint and it maps onto the nullspace $N(\lambda_n I - A)$. This yields $P_n \varphi = 0$ and hence $A\varphi = \lambda\varphi$. Next, assume that $\varphi \in N(\lambda_n I - A)$. Then for $n > m$ we obtain $A_m \varphi = \lambda_n \varphi$ and for $n \leqslant m$ we have $A_m \varphi = 0$. Thus, the eigenvalues of A_m which are different from zero are λ_{m+1}, λ_{m+2}, etc, i.e. the construction of A_m removes step-by-step the eigenvalues λ_1, λ_2, ... from A. But now theorems 2.4.13 and 2.4.18 show that

$$\|A_m\| = |\lambda_{m+1}| \to 0, \quad m \to \infty. \tag{2.4.66}$$

This shows the norm convergence of the series (2.4.64). To prove the orthogonal representation (2.4.62), (2.4.63) we remark that

$$\left\langle \varphi, \sum_{n=1}^{m} P_n \varphi \right\rangle = \underbrace{\left\langle \varphi - \sum_{n=1}^{m} P_n \varphi, \sum_{n=1}^{m} P_n \varphi \right\rangle}_{=0} + \left\langle \sum_{n=1}^{m} P_n \varphi, \sum_{n=1}^{m} P_n \varphi \right\rangle = \sum_{n=1}^{m} \|P_n \varphi\|^2. \tag{2.4.67}$$

We now calculate

$$\left\| \varphi - \sum_{n=1}^{m} P_n \varphi \right\|^2 = \|\varphi\|^2 - \sum_{n=1}^{m} \|P_n \varphi\|^2, \tag{2.4.68}$$

which shows that the sum

$$\sum_{n=1}^{\infty} P_n \varphi \tag{2.4.69}$$

is norm convergent in X. An application of the bounded operator A now yields

$$A\left(\varphi - \sum_{n=1}^{\infty} P_n \varphi\right) = A\varphi - \sum_{n=1}^{\infty} \lambda_n P_n \varphi = 0, \tag{2.4.70}$$

hence $\varphi - \sum_{n=1}^{\infty} P_n \varphi \in N(A)$ and we have shown (2.4.63). This completes our proof. □

2.4.5 Singular value decomposition

The spectral theorem as presented above is satisfied only for a very particular class of operators. Integral operators belong to this class only if their kernel is symmetric and real. However, we need to study more general equations and the spectral decomposition for general compact linear operators in Hilbert spaces is known as *singular value decomposition*.

Definition 2.4.21. *Consider a compact linear operator $A : X \to Y$ with Hilbert spaces X, Y and its adjoint A^*. Then the square roots of the non-zero eigenvalues of the self-adjoint compact operator $A^*A : X \to X$ are called the* singular values *of the operator A.*

We are now prepared for the following important result.

Theorem 2.4.22. *We order the sequence (μ_n) of the non-zero singular values of the non-zero compact operator $A : X \to Y$ such that $\mu_1 \geqslant \mu_2 \geqslant \mu_3 \geqslant \ldots$, where the values are repeated according to their multiplicity, i.e. according to the dimension $\dim N(\mu_n^2 I - A^*A)$ of the nullspace of $\mu_n^2 I - A^*A$. By $Q : X \to N(A)$ we denote the orthogonal projection onto the null-space $N(A)$ of A. Then there exist orthonormal sequences $(\varphi_n) \subset X$ and $(g_n) \subset Y$ such that*

$$A\varphi_n = \mu_n g_n, \qquad A^* g_n = \mu_n \varphi_n \tag{2.4.71}$$

for $n \in \mathbb{N}$. For each $\varphi \in X$ we have the singular value decomposition

$$\varphi = \sum_{n=1}^{\infty} \langle \varphi, \varphi_n \rangle \varphi_n + Q\varphi. \tag{2.4.72}$$

The operator A is represented by

$$A\varphi = \sum_{n=1}^{\infty} \mu_n \langle \varphi, \varphi_n \rangle g_n. \tag{2.4.73}$$

Each system (μ_n, φ_n, g_n) with these properties is called a singular system *of A. If the operator A is injective, then the orthonormal system $\{\varphi_n : n \in \mathbb{N}\}$ is complete in X.*

Proof. Let (φ_n) be the orthonormal sequence of eigenvectors of the compact self-adjoint operator A^*A with eigenvalues μ_n^2. We define

$$g_n := \frac{1}{\mu_n} A\varphi_n. \tag{2.4.74}$$

Then (2.4.71) is an immediate consequence of $A^*A\varphi = \mu_n^2 \varphi$. The representation (2.4.63) can be written as

$$\varphi = \sum_{n=1}^{\infty} \langle \varphi, \varphi_n \rangle \varphi_n + Q\varphi, \tag{2.4.75}$$

where Q is the orthogonal projection onto $N(A^*A)$. Let $\psi \in N(A^*A)$. Then $\langle A\psi, A\psi \rangle = \langle \psi, A^*A\psi \rangle = 0$, therefore $A\psi = 0$ and $\psi \in N(A)$. This implies $N(A^*A) = N(A)$ and thus Q is the orthogonal projection onto $N(A)$. We have shown (2.4.72). Finally, the representation of A given in (2.4.73) is obtained by an application of A to (2.4.72). □

We now write the solution of the equation

$$A\varphi = f \tag{2.4.76}$$

in spectral form and to understand the ill-posedness of the problem in terms of multiplication operators.

Theorem 2.4.23 (Picard). *Let $A : X \to Y$ be a compact linear operator between Hilbert spaces X, Y with a singular system (μ_n, φ_n, g_n). The equation (2.4.76) has a solution if and only if $f \in N(A^*)^\perp$ and*

$$\sum_{n=1}^{\infty} \frac{1}{\mu_n^2} |\langle f, g_n \rangle|^2 < \infty. \tag{2.4.77}$$

In this case the solution φ of (2.4.76) is given by

$$\varphi = \sum_{n=1}^{\infty} \frac{1}{\mu_n} \langle f, g_n \rangle \varphi_n. \tag{2.4.78}$$

Proof. The condition $f \in N(A^*)^\perp$ is a consequence of theorem 2.4.14. For a solution φ of (2.4.76) we calculate

$$\mu_n \langle \varphi, \varphi_n \rangle = \langle \varphi, A^* g_n \rangle = \langle A\varphi, g_n \rangle = \langle f, g_n \rangle, \tag{2.4.79}$$

which shows (2.4.78). Further, we obtain the estimate

$$\sum_{n=1}^{\infty} \frac{1}{\mu_n^2} |\langle f, g_n \rangle|^2 = \sum_{n=1}^{\infty} |\langle \varphi, \varphi_n \rangle|^2 \leqslant \|\varphi\|^2, \tag{2.4.80}$$

thus if the equation has a solution the sum (2.4.77) is bounded and the first direction of the theorem is proven.

To show the other direction we assume that $f \in N(A^*)^\perp$ and (2.4.77) is satisfied. Then the boundedness of (2.4.77) shows the convergence of the sum (2.4.78). We apply A to (2.4.78) to calculate

$$\begin{aligned} A\varphi &= \sum_{n=1}^{\infty} \frac{1}{\mu_n} \langle f, g_n \rangle A\varphi_n \\ &= \sum_{n=1}^{\infty} \langle f, g_n \rangle g_n \\ &= f, \end{aligned} \tag{2.4.81}$$

where we used $f \in N(A^*)^\perp$ and (2.4.72) for the singular system (μ_n, g_n, φ_n) of A^* for the last step. □

2.5 Lax–Milgram and weak solutions to boundary value problems

In this subsection we will introduce a theorem called the *Lax–Milgram theorem*. It is very important for solving linear elliptic partial differential equations by variational methods. We first provide the definitions of a *sesquilinear form* and define *X-coercivity*.

Definition 2.5.1. *Let X be a linear space over $\mathbb{C}$ and $a(,) : X \times X \to \mathbb{C}\,(u, v) \to \mathbb{C}$ be a mapping.*

(i) *The map $a(,)$ is called a* sesquilinear form *on V if it is linear with respect to the first variable and anti-linear with respect to the second variable. That is for any $u, u_1, u_2, v, v_1, v_2 \in X$ and $\alpha_1, \alpha_2, \beta_1, \beta_2 \in \mathbb{C}$, we have:*

$$
\begin{aligned}
a(\alpha_1 u_1 + \alpha_2 u_2, v) &= \alpha_1 a(u_1, v) + \alpha_2 a(u_2, v), \\
a\big(u, \beta_1 v_1 + \beta_2 v_2\big) &= \overline{\beta_1} a(u, v_1) + \overline{\beta_2} a(u, v_2).
\end{aligned}
\tag{2.5.1}
$$

(ii) *If X is a normed space, the sesquilinear form $a(,)$ on X is called continuous and X-coercive if it satisfies*

$$
|a(u, v)| \leqslant C\|u\|\,\|v\|, \quad u, v \in X \tag{2.5.2}
$$

for some constant $C > 0$ independent of u, v and

$$
|a(v, v)| \geqslant c\|v\|^2, \quad v \in X \tag{2.5.3}
$$

for some another constant $c > 0$ independent of v, respectively.

The Riesz representation theorem 2.4.1 implies the following lemma as a direct consequence.

Lemma 2.5.2. *Let X be a Hilbert space with inner product $\langle\cdot,\cdot\rangle_X$.*

(i) *If $a(,)$ is a continuous sesquilinear form on X, then there exist unique elements A and $B \in BL(X)$ such that*

$$
a(u, v) = \langle Au, v\rangle_X = \langle u, Bv\rangle_X, \quad u, v \in X. \tag{2.5.4}
$$

(ii) *Denote by $\tilde{X}'$ the set of all anti-linear continuous functions on X. Then, there exists an isometry $J \in BL(\tilde{X}', X)$ such that*

$$
f(v) = \langle J(f), v\rangle_X, \quad f \in \tilde{X}', \quad v \in X. \tag{2.5.5}
$$

Based on this lemma, we have

Theorem 2.5.3. *Let X and $a(,)$ be a Hilbert space over $\mathbb{C}$ and X-coercive continuous sesquilinear form on X, respectively. Then for any $f \in \tilde{X}'$, there exists a unique $u = u(f) \in X$ such that*

$$
a(u, v) = f(v), \quad v \in X, \quad \|u(f)\|_X \leqslant c^{-1}\|f\|_{\tilde{X}'},
$$

where $c > 0$ is the constant in (2.5.3).

Proof. By the X-coercivity of $a(,)$, A and B in lemma 2.5.2 are injective and

$$\|Au\|_X \geqslant c\|u\|_X, \quad u \in X, \tag{2.5.6}$$

where $c > 0$ is the constant in (2.5.3). Then, this estimate implies that $A(X)$ is closed subspace in X. Hence, by $A^* = B$,

$$A(X)^\perp = N(B) = \{0\},$$

which implies $A(X) = X$. Therefore, A is an isomorphism on X and from (2.5.6) the operator norm $\|A^{-1}\|$ satisfies

$$\|A^{-1}\| \leqslant c^{-1}. \tag{2.5.7}$$

Now

$$a(u, v) = f(v), \quad v \in X$$

is equivalent to

$$\langle Au, v\rangle_X = \langle J(f), v\rangle_X, \quad v \in X.$$

Therefore, u can be uniquely given by $u = A^{-1}J(f)$ and it satisfies the estimate $\|u\| \leqslant c^{-1}\|f\|_{\tilde{X}'}$. □

Remark 2.5.4. *Let X and $a(,)$ be a Hilbert space over $\mathbb{C}$ and a X-coercive continuous sesquilinear form on X, respectively. Then, it is easy to see that there exists a unique $\mathcal{A} \in BL(X, X')$ with $X' := BL(X, \mathbb{C})$ such that*

$$a(u, v) = \mathcal{A}u(v), \quad v \in X. \tag{2.5.8}$$

Hence for given $f \in \tilde{X}'$, $a(u, v) = f(v)$ for any $v \in X$ is equivalent to $\mathcal{A}u = f$. For solving a boundary value problem for an elliptic equation with some homogeneous boundary condition by the variational method, $\mathcal{A}$ expresses the elliptic operator with the homogeneous boundary condition.

We finish this section with a simple application of the Lax–Milgram result.

Corollary 2.5.5. *Let X be a Hilbert space and $A : X \to X$ a bounded linear operator with*

$$\langle A\varphi, \varphi\rangle \geqslant c\|\varphi\|^2, \quad \varphi \in X. \tag{2.5.9}$$

Then A is invertible with

$$\|A^{-1}\| \leqslant c^{-1}. \tag{2.5.10}$$

Proof. Clearly, by (2.2.1) a sesquilinear form is defined by

$$a(\varphi, \psi) := \langle A\varphi, \psi\rangle, \quad \varphi, \psi \in X.$$

By (2.5.9) it is coercive. For $b \in X$ an element of X' is given by $f(\psi) := \langle b, \psi\rangle$ and we have $\|f\|_{X'} = \|b\|_X$. Then, by theorem 2.5.3 there is $\varphi \in X$ such that

$$\langle A\varphi, \psi\rangle = \langle b, \psi\rangle, \quad \psi \in X,$$

which is equivalent to

$$A\varphi = b,$$

i.e. the operator A is invertible on X, and we have the norm estimate (2.5.10) by (2.5.7). □

2.6 The Fréchet derivative and calculus in normed spaces

The goal here is to collect basic definitions and notations around the Fréchet derivative in $\mathbb{R}^m$ and in normed spaces.

Let X, Y and Z denote arbitrary normed spaces and U and V are open subsets of X and Y, respectively. Recall that by $BL(X, Y)$ we denote the normed space of bounded linear operators from X to Y. For a mapping $A \in BL(X, Y)$ we write Ah for $A(h)$ indicating the linear dependence on $h \in X$. Also, recall the definition of $o(\|h\|)$ denoting a function or operator A which satisfies

$$\frac{\|A(h)\|}{\|h\|} \to 0, \quad \|h\| \to 0. \tag{2.6.1}$$

Definition 2.6.1 (Fréchet differential). *Consider a mapping $F : U \to Y$ and an element $u \in U$. We say that F is (Fréchet) differentiable at $u \in U$, if there exists $F' \in BL(X, Y)$ and a mapping $R : U \to Y$ with $R(h) = o(\|h\|)$ such that*

$$F(u + h) = F(u) + F'h + R(h). \tag{2.6.2}$$

In this case the linear mapping F' is determined uniquely and is called the Fréchet differential of F at u and is sometimes denoted by $F'(u)$. If F is differentiable at all $u \in U$, we say that F is (Fréchet) differentiable in U.

Examples. We next study several examples to gain experience with derivatives. The following examples also prepare the differentiability proofs for shape monitoring and shape reconstruction problems.

(a) For functions $f : \mathbb{R} \to \mathbb{R}$, clearly the Fréchet derivative coincides with the classical derivative. This is also true for functions $F : \mathbb{R}^m \to \mathbb{R}^m$. In this case, $F = (F_1, F_2, \ldots, F_m)^T$ a column vector and T denotes the transpose. Also its Fréchet derivative is given by the matrix

$$F = \begin{pmatrix} \frac{\partial F_1}{\partial x_1} & \cdots & \frac{\partial F_1}{\partial x_n} \\ \vdots & & \vdots \\ \frac{\partial F_m}{\partial x_1} & \cdots & \frac{\partial F_m}{\partial x_n} \end{pmatrix}. \tag{2.6.3}$$

We also use the notation $\nabla F := F$.

(b) If F is linear, then $F' = F$. This can be seen immediately from the definition of linearity

$$F(u + h) = F(u) + F(h), \tag{2.6.4}$$

which leads to $F'(h) = F(h)$ and $R \equiv 0$ in (2.6.2).

(d) Consider $F = (F_1, F_2, \ldots, F_m)^T$ again with $X = \mathbb{R}^m$ and $Y = \mathbb{R}^m$. Assume that each $F_j\,(1 \leqslant j \leqslant m)$ has a bounded continuous derivative in a neighborhood U of a point $u \in X$ and that F' depends continuously on x in a neighborhood U of a point u, i.e. $F \in BC^1(U, Y)$. Then, with $u = (u_1, \ldots, u_n)^T$ and $h = (h_1, \ldots, h_n)^T$ the *fundamental theorem of calculus* yields

$$\begin{aligned} F(u + h) &= F(u) + \nabla Fh + o\big(\|h\|\big) \\ &= F(u) + \sum_{j=1}^{n} \frac{\partial F}{\partial u_j} h_j + o\big(\|h\|\big). \end{aligned} \tag{2.6.5}$$

(e) Let Z_1 be the space of Hölder-continuously differentiable functions $C^{1,\alpha}(\partial D)$ on ∂D, where D is a domain in $\mathbb{R}^m$ with boundary of class $C^{1,\alpha}$, $\alpha > 0$. We define

$$U := \left\{ r \in Z_1 : \|r\|_{C^{1,\alpha}(\partial D)} < c \right\} \tag{2.6.6}$$

with some constant c. The derivative of the mapping

$$F_1 : U \to \mathbb{R}, \quad F_1(r) := r(x), \quad r \in U \tag{2.6.7}$$

for $x \in \partial D$ fixed is given by

$$F'_1 h = h(x), \quad h \in Z. \tag{2.6.8}$$

For the Fréchet derivative we have the *product rule* and the *chain rule* as follows. Consider $A : U \to Y$ and $B : U \to Y$. Then, we have

$$\begin{aligned} A(u + h) \cdot B(u + h) &= (A(u) + A'h + R(h)) \cdot (B(u) + B'h + S(h)) \\ &= A(u) \cdot B(u) + (A(u) \cdot B'h + A'h \cdot B(u)) \\ &\quad + (A(u) + A'h + R(h)) \cdot S(h) + R(h) \cdot (B'h + S(h)), \end{aligned} \tag{2.6.9}$$

where $\cdot$ can be any multiplicative operation in Y for example multiplication in Y by assuming Y is a Banach algebra. This leads to the *product rule*

$$(A \cdot B)' = A' \cdot B + A \cdot B'. \tag{2.6.10}$$

Assume that $A : U \to Y$ and $B : A(U) \to Z$. Then, for the composition $B \circ A$ of the two mapping we obtain

$$\begin{aligned} (B \circ A)(u + h) &= B(A(u + h)) = B(A(u) + A'h + o(\|h\|)) \\ &= B(A(u)) + B'(A'h + o(\|h\|)) + o(A'h + o(\|h\|) \\ &= B(A(u)) + B' \circ A'h + o(\|h\|), \end{aligned} \tag{2.6.11}$$

where we used $B'o(\|h\|) = o(\|h\|)$, $o(A'h + o(\|h\|) = o(\|h\|)$ and $o(\|h\|) + o(\|h\|) = o(\|h\|)$ in the sense of the definition of the *little o*, leading to the *chain rule*

$$(B \circ A)' = B' \circ A'. \tag{2.6.12}$$

We can build higher Fréchet derivatives, where here the conceptional work is a little bit more difficult than for the first order case. Note that the Fréchet derivative depends on both the point u where it is calculated as well as the argument h, i.e. F' maps $U \times X$ into Y, where the mapping is linear in the second argument.

We denote the set of all mappings

$$A : \underbrace{X \times \cdots \times X}_{n \text{ times}} \to Y, \quad (x_1, \ldots, x_n) \mapsto A(x_1, \ldots, x_n), \tag{2.6.13}$$

which are linear in every argument $x_1, \ldots, x_n$, as multi-linear mappings of order n, which are bounded in the sense that

$$\sup_{0 \neq x_j \ \|x_j\| \leqslant 1, j=1,\ldots,n} \frac{|A(x_1, \ldots, x_n)|}{\|x_1\| \cdots \|x_n\|} < \infty. \tag{2.6.14}$$

We denote this set by $BL^{(n)}(X, Y)$. Note that $BL^{(1)}(\mathbb{R}^m, \mathbb{R}^m)$ coincides with the set of $m \times n$-matrices, the set $BL^{(2)}(X, Y)$ with the set of *quadratic forms* on X with values in Y.

The first Fréchet derivative is a mapping

$$F' : U \to BL^{(1)}(X, Y), \quad u \mapsto F'(u) \tag{2.6.15}$$

via

$$F'(u, x) = F'(u)x \in Y, \quad x \in X. \tag{2.6.16}$$

The Fréchet derivative F'' as a derivative of F' is a mapping

$$F'' : U \to BL^{(2)}(X, Y), \quad u \mapsto F''(u) \tag{2.6.17}$$

via

$$F''(u, x, y) = F''(u)(x, y) \in Y, \quad x, y \in X. \tag{2.6.18}$$

The nth Fréchet derivative $F^{(n)}$ is the derivative of the $(n-1)$th derivative and thus

$$F^{(n)} : U \to BL^{(n)}(X, Y). \tag{2.6.19}$$

Often, the notation

$$F^{(n)} = \frac{\mathrm{d}^n F}{\mathrm{d}u^n}$$

is used for the nth Fréchet derivative with respect to u. We also employ the abbreviation

$$F^{(n)}(u, x) := F^{(n)}(u)(x, \ldots, x) \tag{2.6.20}$$

when the same argument $x_j = x$ for $j = 1, \ldots, n$ is used.

Now consider the situation where we have a mapping $A : U \to Y$ where $U \subset X$ and Y is the space of operators A on Z. Then, under reasonable conditions, the differentiability of A implies the differentiability of the inverse operator A^{-1} in dependence on $u \in U$.

Theorem 2.6.2. *Let X be a normed space, $U \subset X$ be an open set and Y be a Banach algebra with neutral element e. Here the Banach algebra Y is a Banach space with multiplication y_1y_2 defined for any $y_1, y_2 \in Y$ such that $\|y_1y_2\| \leqslant \|y_1\|\|y_2\|$ and e plays role of the unit of this multiplication. Let $A = A(u) \in Y$ be Fréchet differentiable in $u_0 \in U$. We further assume that $A^{-1} = A^{-1}(u)$ exists in Y for all $u \in U$ near u_0 and the inverse depends continuously on u near u_0. Then $A^{-1}(u)$ is Fréchet differentiable at u_0 with Fréchet derivative*

$$(A^{-1})' = -A^{-1}A'A^{-1}. \tag{2.6.21}$$

$$\text{i.e. } \big((A^{-1})'(u_0)\big)x = -A'(u_0)\big(A'(u_0)x)\big)A^{-1}(u_0), \quad x \in X. \tag{2.6.22}$$

Proof. We define

$$z(h) := A^{-1}(u_0 + h) - A^{-1}(u_0) + A^{-1}(u_0)A'(u_0, h)A^{-1}(u_0).$$

We need to show that $z(h) = o(\|h\|)$ to prove the statement of the theorem. We multiply by $A(u_0)$ from the right and from the left and use the continuity of the inverse $A^{-1}(u)$ to derive

$$\begin{aligned} A(u_0)z(h)A(u_0) &= A(u_0)A^{-1}(u_0 + h)A(u_0) - A(u_0) + A'(u_0, h) \\ &= (A(u_0 + h) - A(u_0))A^{-1}(u_0 + h)(A(u_0 + h) - A(u_0)) \\ &\quad - (A(u_0 + h) - A(u_0) - A'(u_0, h)) \\ &= o(\|h\|), \end{aligned} \tag{2.6.23}$$

which completes the proof. □

Usually, differentiability is used to derive high-order estimates for a function in a neighborhood of some point.

Lemma 2.6.3. *Let $f : U \to \mathbb{C}$ be two times continuously Fréchet differentiable. Then, for $h \in X$ sufficiently small and $u \in U$, we have*

$$f(u + h) = f(u) + f'(u)h + f_1(u, h) \tag{2.6.24}$$

with

$$f_1(u, h) = \int_0^1 (1 - t)f''(u + th, h)\,\mathrm{d}t. \tag{2.6.25}$$

Proof. We define a function $g : [0, 1] \to \mathbb{C}$ by $g(t) := f(u + th)$, $t \in [0, 1]$. By the chain rule the derivative of g with respect to t is given by $g'(t) = f'(u + th, h)$ and the second derivative is given by $g''(t) = f''(u + th, h, h)$, for which we use the short notation $f''(u + th, h)$. Now, applying the fundamental theorem of calculus two times we have

$$
\begin{aligned}
g(1) - g(0) &= \int_0^1 g'(t)\,\mathrm{d}t \\
&= \int_0^1 \left(g'(0) + \int_0^t g''(s)\,\mathrm{d}s\right)\mathrm{d}t \\
&= g'(0) + \int_0^1 \int_0^t g''(s)\,\mathrm{d}s\,\mathrm{d}t \\
&= g'(0) + \int_0^1 (1-\rho)g''(\rho)\,\mathrm{d}\rho,
\end{aligned}
$$

which yields (2.6.24) with (2.6.25).

Remark. The same result as in (2.6.25) holds if we replace $\mathbb{C}$ by some Banach space Y, where one needs to define integrals with values in a Banach space.

The Picard–Lindelöf method to solve the Cauchy problem for first order ordinary differential equations is well known. It transforms the problem to an integral equation and looks for a fixed point. This method can be considered as an application of the Banach fixed-point theorem which we will introduce next. To begin with we define some terminologies.

Definition 2.6.4. *Let X be a Banach space with norm $\|\cdot\|$ and $F: X \to X$ be a mapping. F is called a contraction mapping on X if it satisfies*

$$\|F(x) - F(y)\| \leqslant c\|x - y\|, \quad x, y \in X \tag{2.6.26}$$

for some constant $0 < c < 1$.

Then the Banach fixed-point theorem is given as follows.

Theorem 2.6.5 (Banach fixed-point theorem). *Let F be a contraction mapping on a Banach space X. Then, F has a unique fixed point $x_0 \in X$ that is $F(x_0) = x_0$. Further for any $x \in X$, the sequence (x_n) defined by*

$$x_1 = x, \quad x_{n+1} = F(x_{n-1}), \quad n \in \mathbb{N} \tag{2.6.27}$$

converges to x_0.

Proof. To see the uniqueness of a fixed point, let x, x' be fixed points of F. Then by (2.6.26)

$$\|x - x'\| = \|F(x) - F(x')\| \leqslant c\|x - x'\|. \tag{2.6.28}$$

Since $0 < c < 1$, we must have $x = x'$.

Next we show that the sequence (x_n) converges in X which implies (x_n) converges in X to the unique fixed point $x_0 \in X$ by (2.6.27) and the continuity of F on X due to (2.6.26). To see the convergence of sequence (x_n) it is enough to show for any $n, m \in \mathbb{N}$ with $n > m$

$$\|x_n - x_m\| \leqslant \frac{c^n}{1-c}\|x_1 - x_0\|. \tag{2.6.29}$$

Observe that for any $n \in \mathbb{N}$

$$\|x_{n+1} - x_n\| = \|F(x_n) - F(x_{n-1})\| \leqslant c\|x_n - x_{n-1}\|. \tag{2.6.30}$$

Hence $\|x_{n+1} - x_n\| \leqslant c^{n-1}\|x_2 - x_1\|$ for any $n \in \mathbb{N}$. This implies

$$\|x_n - x_m\| \leqslant \sum_{j=m}^{n-1} \left\|x_{j+1} - x_j\right\| \leqslant \|x_2 - x_1\| \sum_{j=m}^{n-1} c^j < \frac{c^m}{1-c}\|x_2 - x_1\|. \tag{2.6.31}$$

Thus (x_n) is a Cauchy sequence which converges towards some $x_* \in X$. Clearly, by choosing $n = m + 1$ we have $\|F(x_m) - x_m\| \to 0$ for $m \to \infty$, such that by continuity $F(x_*) = x_*$, i.e. x_* is a fixed point and by uniqueness we have $x_* = x_0$.

Bibliography

[1] Kress R 1999 *Linear Integral Equations* (*Applied Mathematical Sciences* vol 82) 2nd edn (New York: Springer)

[2] Reed M and Simon B 1980 *Functional Analysis* (New York: Academic)

[3] Bachman G and Narici L 2000 *Functional Analysis* (Mineola, NY: Dover)

[4] Heuser H 1986 *Funktionalanalysis* (Teubner: Leipzig)