FUNCTIONAL ANALYSIS

CHRISTIAN REMLING

Contents

These are lecture notes that have evolved over time. Center stage is given to the Spectral Theorem for (bounded, in this first part) normal operators on Hilbert spaces; this is approached through the Gelfand representation of commutative C^*-algebras. Banach space topics are ruthlessly reduced to the mere basics (some would argue, less than that); topological vector spaces aren't mentioned at all.

1. Metric and topological spaces

A *metric space* is a set on which we can measure distances. More precisely, we proceed as follows: let $X \neq \emptyset$ be a set, and let $d : X \times X \to [0, \infty)$ be a map.

Definition 1.1. (X, d) is called a *metric space* if d has the following properties, for arbitrary $x, y, z \in X$:

(1) $d(x, y) = 0 \iff x = y$
(2) $d(x, y) = d(y, x)$
(3) $d(x, y) \leq d(x, z) + d(z, y)$

Property 3 is called the *triangle inequality*. It says that a detour via z will not give a shortcut when going from x to y.

The notion of a metric space is very flexible and general, and there are many different examples. We now compile a preliminary list of metric spaces.

Example 1.1. If $X \neq \emptyset$ is an arbitrary set, then

$$d(x, y) = \begin{cases} 0 & x = y \\ 1 & x \neq y \end{cases}$$

defines a metric on X.

Exercise 1.1. Check this.

This example does not look particularly interesting, but it does satisfy the requirements from Definition 1.1.

Example 1.2. $X = \mathbb{C}$ with the metric $d(x, y) = |x - y|$ is a metric space. X can also be an arbitrary non-empty subset of $\mathbb{C}$, for example $X = \mathbb{R}$.

In fact, this works in complete generality: If (X, d) is a metric space and $Y \subset X$, then Y with the same metric is a metric space also.

Example 1.3. Let $X = \mathbb{C}^n$ or $X = \mathbb{R}^n$. For each $p \geq 1$,

$$d_p(x, y) = \left(\sum_{j=1}^{n} |x_j - y_j|^p \right)^{1/p}$$

defines a metric on X. Properties 1, 2 are clear from the definition, but if $p > 1$, then the verification of the triangle inequality is not completely straightforward here. We leave the matter at that for the time being, but will return to this example later.

An additional metric on X is given by

$$d_\infty(x, y) = \max_{j=1,\ldots,n} |x_j - y_j|$$

Exercise 1.2. (a) Show that (X, d_∞) is a metric space.
(b) Show that $\lim_{p\to\infty} d_p(x,y) = d_\infty(x,y)$ for fixed $x, y \in X$.

Example 1.4. Similar metrics can be introduced on function spaces. For example, we can take

$$X = C[a,b] = \{f : [a,b] \to \mathbb{C} : f \text{ continuous }\}$$

and define, for $1 \le p < \infty$,

$$d_p(f,g) = \left(\int_a^b |f(x) - g(x)|^p \, dx \right)^{1/p}$$

and

$$d_\infty(f,g) = \max_{a \le x \le b} |f(x) - g(x)|.$$

Again, the proof of the triangle inequality requires some care if $1 < p < \infty$. We will discuss this later.

Exercise 1.3. Prove that (X, d_∞) is a metric space.

Actually, we will see later that it is often advantageous to use the spaces

$$X_p = L^p(a,b) = \{f : [a,b] \to \mathbb{C} : f \text{ measurable}, \int_a^b |f(x)|^p \, dx < \infty\}$$

instead of X if we want to work with these metrics. We will discuss these issues in much greater detail in Section 2.

On a metric space, we can define convergence in a natural way. We just interpret "$d(x,y)$ small" as "x close to y", and we are then naturally led to make the following definition.

Definition 1.2. Let (X,d) be a metric space, and $x_n, x \in X$. We say that x_n *converges to* x (in symbols: $x_n \to x$ or $\lim x_n = x$, as usual) if $d(x_n, x) \to 0$.

Similarly, we call x_n a *Cauchy sequence* if for every $\epsilon > 0$, there exists an $N = N(\epsilon) \in \mathbb{N}$ so that $d(x_m, x_n) < \epsilon$ for all $m, n \ge N$.

We can make some quick remarks on this. First of all, if a sequence x_n is convergent, then the limit is unique because if $x_n \to x$ and $x_n \to y$, then, by the triangle inequality,

$$d(x,y) \le d(x, x_n) + d(x_n, y) \to 0,$$

so $d(x,y) = 0$ and thus $x = y$. Furthermore, a convergent sequence is a Cauchy sequence: If $x_n \to x$ and $\epsilon > 0$ is given, then we can find an $N \in \mathbb{N}$ so that $d(x_n, x) < \epsilon/2$ if $n \ge N$. But then we also have that

$$d(x_m, x_n) \le d(x_n, x) + d(x, x_m) < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon \qquad (m, n \ge N),$$

so x_n is a Cauchy sequence, as claimed.

The converse is wrong in general metric spaces. Consider for example $X = \mathbb{Q}$ with the metric $d(x, y) = |x - y|$ from Example 1.2. Pick a sequence $x_n \in \mathbb{Q}$ that converges in $\mathbb{R}$ (that is, in the traditional sense) to an irrational limit ($\sqrt{2}$, say). Then x_n is a Cauchy sequence in (X, d) because it is convergent in the bigger space $(\mathbb{R}, d)$, so, as just observed, x_n must be a Cauchy sequence in $(\mathbb{R}, d)$. But then x_n is also a Cauchy sequence in $(\mathbb{Q}, d)$ because this is actually the exact same condition (only the distances $d(x_m, x_n)$ matter, we don't need to know how big the total space is). However, x_n can not converge in $(\mathbb{Q}, d)$ because then it would have to converge to the same limit in the bigger space $(\mathbb{R}, d)$, but by construction, in this space, it converges to a limit that was not in $\mathbb{Q}$.

Please make sure you understand exactly how this example works. There's nothing mysterious about this divergent Cauchy sequence. The sequence really wants to converge, but, unfortunately, the putative limit fails to lie in the space.

Spaces where Cauchy sequences do always converge are so important that they deserve a special name.

Definition 1.3. Let X be a metric space. X is called *complete* if every Cauchy sequence converges.

The mechanism from the previous example is in fact the only possible reason why spaces can fail to be complete. Moreover, it is always possible to *complete* a given metric space by including the would-be limits of Cauchy sequences. The bigger space obtained in this way is called the *completion* of X.

We will have no need to apply this construction, so I don't want to discuss the (somewhat technical) details here. In most cases, the completion is what you think it should be; for example, the completion of $(\mathbb{Q}, d)$ is $(\mathbb{R}, d)$.

Exercise 1.4. Show that $(C[-1, 1], d_1)$ is not complete.

Suggestion: Consider the sequence

$$f_n(x) = \begin{cases} -1 & -1 \leq x < -1/n \\ nx & -1/n \leq x \leq 1/n \\ 1 & 1/n < x \leq 1 \end{cases} .$$

A more general concept is that of a topological space. By definition, a *topological space* X is a non-empty set together with a collection $\mathcal{T}$ of distinguished subsets of X (called open sets) with the following properties:

(1) $\emptyset, X \in \mathcal{T}$
(2) If $U_\alpha \in \mathcal{T}$, then also $\bigcup U_\alpha \in \mathcal{T}$.
(3) If $U_1, \ldots, U_N \in \mathcal{T}$, then $U_1 \cap \ldots \cap U_N \in \mathcal{T}$.

This structure allows us to introduce some notion of closeness also, but things are fuzzier than on a metric space. We can zoom in on points, but there is no notion of one point being closer to a given point than another point.

We call $V \subset X$ a *neighborhood* of $x \in X$ if $x \in V$ and $V \in \mathcal{T}$. (*Warning:* This is sometimes called an open neighborhood, and it is also possible to define a more general notion of not necessarily open neighborhoods. We will always work with open neighborhoods here.) We can then say that x_n converges to x if for every neighborhood V of x, there exists an $N \in \mathbb{N}$ so that $x_n \in V$ for all $n \geq N$. However, on general topological spaces, sequences are not particularly useful; for example, if $\mathcal{T} = \{\emptyset, X\}$, then (obviously, by just unwrapping the definitions) every sequence converges to every limit.

Here are some additional basic notions for topological spaces. Please make sure you're thoroughly familiar with these (the good news is that we won't need much beyond these definitions).

Definition 1.4. Let X be a topological space.
(a) $A \subset X$ is called *closed* if A^c is open.
(b) For an arbitrary subset $B \subset X$, the *closure* of $B \subset X$ is defined as

$$\overline{B} = \bigcap_{A \supset B; A \text{ closed}} A;$$

this is the smallest closed set that contains B (in particular, there always is such a set).
(c) The *interior* of $B \subset X$ is the biggest open subset of B (such a set exists). Equivalently, the complement of the interior is the closure of the complement.
(d) $K \subset X$ is called *compact* if every open cover of K contains a finite subcover.
(e) $\mathcal{B} \subset \mathcal{T}$ is called a *neighborhood base* of X if for every neighborhood V of some $x \in X$, there exists a $B \in \mathcal{B}$ with $x \in B \subset V$.
(f) Let $Y \subset X$ be an arbitrary, non-empty subset of X. Then Y becomes a topogical space with the *induced* (or *relative*) topology

$$\mathcal{T}_Y = \{U \cap Y : U \in \mathcal{T}\}.$$

(g) Let $f : X \to Y$ be a map between topological spaces. Then f is called *continuous* at $x \in X$ if for every neighborhood W of $f(x)$ there exists a neighborhood V of x so that $f(V) \subset W$. f is called continuous

if it is continuous at every point.
(h) A topological space X is called a *Hausdorff space* if for every pair of points $x, y \in X$, $x \neq y$, there exist disjoint neighborhoods V_x, V_y of x and y, respectively.

Continuity on the whole space could have been (and usually is) defined differently:

Proposition 1.5. *f is continuous (at every point $x \in X$) if and only if $f^{-1}(V)$ is open (in X) whenever V is open (in Y).*

Exercise 1.5. Do some reading in your favorite (point set) topology book to brush up your topology. (Folland, Real Analysis, Ch. 4 is also a good place to do this.)

Exercise 1.6. Prove Proposition 1.5.

Metric spaces can be equipped with a natural topology. More precisely, this topology is natural because it gives the same notion of convergence of sequences. To do this, we introduce balls

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

and use these as a neighborhood base for the topology we're after. So, by definition, $U \in \mathcal{T}$ if for every $x \in U$, there exists an $\epsilon > 0$ so that $B_\epsilon(x) \subset U$. Notice also that on $\mathbb{R}$ or $\mathbb{C}$ with the absolute value metric (see Example 1.2), this gives just the usual topology; in fact, the general definition mimics this procedure.

Theorem 1.6. *Let X be a metric space, and let $\mathcal{T}$ be as above. Then $\mathcal{T}$ is a topology on X, and $(X, \mathcal{T})$ is a Hausdorff space. Moreover, $B_r(x)$ is open and*

$$x_n \xrightarrow{d} x \quad \Longleftrightarrow \quad x_n \xrightarrow{\mathcal{T}} x.$$

Proof. Let's first check that $\mathcal{T}$ is a topology on X. Clearly, $\emptyset, X \in \mathcal{T}$. If $U_\alpha \in \mathcal{T}$ and $x \in \bigcup U_\alpha$, then $x \in U_{\alpha_0}$ for some index α_0, and since U_{α_0} is open, there exists a ball $B_r(x) \subset U_{\alpha_0}$, but then $B_r(x)$ is also contained in $\bigcup U_\alpha$.

Similarly, if $U_1, \ldots, U_N$ are open sets and $x \in \bigcap U_j$, then $x \in U_j$ for all j, so we can find N balls $B_{r_j}(x) \subset U_j$. It follows that $B_r(x) \subset \bigcap U_j$, with $r := \min r_j$.

Next, we prove that $B_r(x) \in \mathcal{T}$ for arbitrary $r > 0$, $x \in X$. Let $y \in B_r(x)$. We want to find a ball about y that is contained in the original ball. Since $y \in B_r(x)$, we have that $\epsilon := r - d(x,y) > 0$, and I now claim that $B_\epsilon(y) \subset B_r(x)$. Indeed, if $z \in B_\epsilon(y)$, then, by the triangle inequality,

$$d(z,x) \le d(z,y) + d(y,x) < \epsilon + d(y,x) = r,$$

so $z \in B_r(x)$, as desired.

The Hausdorff property also follows from this, because if $x \neq y$, then $r := d(x,y) > 0$, and $B_{r/2}(x)$, $B_{r/2}(y)$ are disjoint neighborhoods of x, y.

Exercise 1.7. It seems intuitively obvious that $B_{r/2}(x)$, $B_{r/2}(y)$ are disjoint. Please prove it formally.

Finally, we discuss convergent sequences. If $x_n \xrightarrow{d} x$ and V is a neighborhood of x, then, by the way $\mathcal{T}$ was defined, there exists $\epsilon > 0$ so that $B_\epsilon(x) \subset V$. We can find $N \in \mathbb{N}$ so that $d(x_n, x) < \epsilon$ for $n \geq N$, or, equivalently, $x_n \in B_\epsilon(x)$ for $n \geq N$. So $x_n \in V$ for large enough n. This verifies that $x_n \xrightarrow{\mathcal{T}} x$.

Conversely, if this is assumed, it is clear that we must also have that $x_n \xrightarrow{d} x$ because we can take $V = B_\epsilon(x)$ as our neighborhood of x and we know that $x_n \in V$ for all large n. □

In metrizable topological spaces (that is, topological spaces where the topology comes from a metric, in this way) we can always work with sequences. This is a big advantage over general topological spaces.

Theorem 1.7. *Let (X, d) be a metric space, and introduce a topology $\mathcal{T}$ on X as above. Then:*
(a) $A \subset X$ is closed $\iff$ If $x_n \in A$, $x \in X$, $x_n \to x$, then $x \in A$.
(b) Let $B \subset X$. Then

$$\overline{B} = \{x \in X : \text{There exists a sequence } x_n \in B, x_n \to x\}.$$

(c) $K \subset X$ is compact precisely if every sequence $x_n \in K$ has a subsequence that is convergent in K.

These statements are false in general topological spaces (where the topology does not come from a metric).

Proof. We will only prove part (a) here. If A is closed and $x \notin A$, then, since A^c is open, there exists a ball $B_r(x)$ that does not intersect A. This means that if $x_n \in A$, $x_n \to x$, then we also must have that $x \in A$.

Conversely, if the condition on sequences from A holds and $x \notin A$, then there must be an $r > 0$ so that $B_r(x) \cap A = \emptyset$ (if not, pick an x_n from $B_{1/n}(x) \cap A$ for each n; this gives a sequence $x_n \in A$, $x_n \to x$, but $x \notin A$, contradicting our assumption). This verifies that A^c is open and thus A is closed.

Exercise 1.8. Prove Theorem 1.7 (b), (c).

□

Similarly, sequences can be used to characterize continuity of maps between metric spaces. Again, this doesn't work on general topological spaces.

Theorem 1.8. *Let (X, d), (Y, e) be metric spaces, let $f : X \to Y$ be a function, and let $x \in X$. Then the following are equivalent:*
(a) f is continuous at x (with respect to the topologies induced by d, e).
(b) For every $\epsilon > 0$, there exists a $\delta > 0$ so that $e(f(x), f(t)) < \epsilon$ for all $t \in X$ with $d(x, t) < \delta$.
(c) If $x_n \to x$ in X, then $f(x_n) \to f(x)$ in Y.

Proof. If (a) holds and $\epsilon > 0$ is given, then, since $B_\epsilon(f(x))$ is a neighborhood of $f(x)$, there exists a neighborhood U of x so that $f(U) \subset B_\epsilon(f(x))$. From the way the topology on a metric space is defined, we see that U must contain a ball $B_\delta(x)$, and (b) follows.

If (b) is assumed and $\epsilon > 0$ is given, pick $\delta > 0$ according to (b) and then $N \in \mathbb{N}$ so that $d(x_n, x) < \delta$ for $n \geq N$. But then we also have that $e(f(x), f(x_n)) < \epsilon$ for all $n \geq N$, that is, we have verified that $f(x_n) \to f(x)$.

Finally, if (c) holds, we argue by contradiction to obtain (a). So assume that, contrary to our claim, we can find a neighborhood V of $f(x)$ so that for every neighborhood U of x, there exists $t \in U$ so that $f(t) \notin V$. In particular, we can then pick an $x_n \in B_{1/n}(x)$ for each n, so that $f(x_n) \notin V$. Since V is a neighborhood of $f(x)$, there exists $\epsilon > 0$ so that $B_\epsilon(f(x)) \subset V$. Summarizing, we can say that $x_n \to x$, but $e(f(x_n), f(x)) \geq \epsilon$; in particular, $f(x_n) \not\to f(x)$. This contradicts (c), and so we have to admit that (a) holds. $\square$

The following fact is often useful:

Proposition 1.9. *Let (X, d) be a metric space and $Y \subset X$. As above, write $\mathcal{T}$ for the topology generated by d on X.*

Then (Y, d) is a metric space, too (this is obvious and was also already observed above). Moreover, the topology generated by d on Y is the relative topology of Y as a subspace of $(X, \mathcal{T})$.

Exercise 1.9. Prove this (this is done by essentially chasing definitions, but it is a little awkward to write down).

We conclude this section by proving our first fundamental functional analytic theorem. We need one more topological definition: We call a set $M \subset X$ *nowhere dense* if $\overline{M}$ has empty interior. If X is a metric space, we can also say that $M \subset X$ is nowhere dense if $\overline{M}$ contains no (non-empty) open ball.

Theorem 1.10 (Baire)**.** *Let X be a complete metric space. If the sets $A_n \subset X$ are nowhere dense, then $\bigcup_{n\in\mathbb{N}} A_n \neq X$.*

Completeness is crucial here:

Exercise 1.10. Show that there are (necessarily: non-complete) metric spaces that are countable unions of nowhere dense sets.

Suggestion: $X = \mathbb{Q}$

Proof. The following proof is similar in spirit to Cantor's famous diagonal trick, which proves that $[0, 1]$ is uncountable. We will construct an element that is not in $\bigcup A_n$ by avoiding these sets step by step.

First of all, we may assume that the A_n's are closed (if not, replace A_n with $\overline{A_n}$; note that these sets are still nowhere dense).

Then, since A_1 is nowhere dense, we can find an $x_1 \in A_1^c$. In fact, A_1^c is also open, so we can even find an open ball $B_{r_1}(x_1) \subset A_1^c$, and here we may assume that $r_1 \le 2^{-1}$ (decrease r_1 if necessary).

In the next step, we pick an $x_2 \in B_{r_1/2}(x_1) \setminus A_2$. There must be such a point because A_2 is nowhere dense and thus cannot contain the ball $B_{r_1/2}(x_1)$. Moreover, we can again find $r_2 > 0$ so that

$$B_{r_2}(x_2) \cap A_2 = \emptyset, \quad B_{r_2}(x_2) \subset B_{r_1/2}(x_1), \quad r_2 \le 2^{-2}.$$

We continue in this way and construct a sequence $x_n \in X$ and radii $r_n > 0$ with the following properties:

$$B_{r_n}(x_n) \cap A_n = \emptyset, \quad B_{r_n}(x_n) \subset B_{r_{n-1}/2}(x_{n-1}), \quad r_n \le 2^{-n}$$

It follows that x_n is a Cauchy sequence. Indeed, if $m \ge n$, then x_m lies in $B_{r_n/2}(x_n)$, so

$$d(x_m, x_n) \le \frac{r_n}{2}. \tag{1.1}$$

Since X is complete, $x := \lim x_n$ exists. Moreover,

$$d(x_n, x) \le d(x_n, x_m) + d(x_m, x)$$

for arbitrary $m \in \mathbb{N}$. For $m \ge n$, (1.1) shows that $d(x_n, x_m) \le r_n/2$, so if we let $m \to \infty$, it follows that

$$d(x_n, x) \le \frac{r_n}{2}. \tag{1.2}$$

By construction, $B_{r_n}(x_n) \cap A_n = \emptyset$, so (1.2) says that $x \notin A_n$ for all n. □

Baire's Theorem can be (and often is) formulated differently. We need one more topological definition: We call a set $M \subset X$ *dense* if $\overline{M} = X$. For example, $\mathbb{Q}$ is dense in $\mathbb{R}$. Similarly, $\mathbb{Q}^c$ is also dense in $\mathbb{R}$. However, note that (of course) $\mathbb{Q} \cap \mathbb{Q}^c = \emptyset$.

Theorem 1.11 (Baire)**.** *Let X be a complete metric space. If U_n ($n \in \mathbb{N}$) are dense open sets, then $\bigcap_{n\in\mathbb{N}} U_n$ is dense.*

Exercise 1.11. Derive this from Theorem 1.10.

Suggestion: If U is dense and open, then $A = U^c$ is nowhere dense (prove this!). Now apply Theorem 1.10. This will not quite give the full claim, but you can also apply Theorem 1.10 on suitable subspaces of the original space.

An immediate consequence of this, in turn, is the following slightly stronger looking version. By definition, a G_δ *set* is a countable intersection of open sets.

Exercise 1.12. Give an example that shows that a G_δ set need not be open (but, conversely, open sets are of course G_δ sets).

Theorem 1.12 (Baire)**.** *Let X be a complete metric space. Then a countable intersection of dense G_δ sets is a dense G_δ set.*

Exercise 1.13. Derive this from the previous theorem.

Given this result, it makes sense to interpret dense G_δ sets as big sets, in a topological sense, and their complements as small sets. Theorem 1.12 then says that even a countable union of small sets will still be small. Call a property of elements of a complete metric space *generic* if it holds at least on a dense G_δ set.

Theorem 1.12 has a number of humoristic applications, which say that certain unexpected properties are in fact generic. Here are two such examples:

Example 1.5. Let $X = C[a, b]$ with metric $d(f, g) = \max |f(x) - g(x)|$ (compare Example 1.4). This is a complete metric space (we'll prove this later). It can now be shown, using Theorem 1.12, that *the generic continuous function is nowhere differentiable.*

Example 1.6. *The generic coin flip refutes the law of large numbers.*

More precisely, we proceed as follows. Let $X = \{(x_n)_{n\geq 1} : x_n = 0 \text{ or } 1\}$ and $d(x, y) = \sum_{n=1}^{\infty} 2^{-n}|x_n - y_n|$. This is a metric and X with this metric is complete, but we don't want to prove this here. In fact, this metric is a natural choice here; it generates the product topology on X.

From probability theory, we know that if the x_n are independent random variables and the coin is fair, then, with probability 1, we have that $S_n/n \to 1/2$, where $S_n = x_1 + \ldots + x_n$ is the number of heads (say) in the first n coin tosses.

The generic behavior is quite different: For a generic sequence $x \in X$,

$$\liminf_{n\to\infty} \frac{S_n}{n} = 0, \quad \limsup_{n\to\infty} \frac{S_n}{n} = 1.$$

Since these examples are for entertainment only, we will not prove these claims here.

Baire's Theorem is fundamental in functional analysis, and it will have important consequences. We will discuss these in Chapter 3.

Exercise 1.14. Consider the space $X = C[0,1]$ with the metric $d(f,g) = \max_{0\le x\le 1} |f(x) - g(x)|$ (compare Example 1.4). Define $f_n \in X$ by

$$f_n(x) = \begin{cases} 2^n x & 0 \le x \le 2^{-n} \\ 1 & 2^{-n} < x \le 1 \end{cases}.$$

Work out $d(f_n, 0)$ and $d(f_m, f_n)$, and deduce from the results of this calculation that $S = \{f \in X : d(f, 0) = 1\}$ is not compact.

Exercise 1.15. Let X, Y be topological spaces, and let $f : X \to Y$ be a continuous map. True or false (please give a proof or a counterexample):
(a) $U \subset X$ open $\Longrightarrow f(U)$ open
(b) $A \subset Y$ closed $\Longrightarrow f^{-1}(A)$ closed
(c) $K \subset X$ compact $\Longrightarrow f(K)$ compact
(d) $L \subset Y$ compact $\Longrightarrow f^{-1}(L)$ compact

Exercise 1.16. Let X be a metric space, and define, for $x \in X$ and $r > 0$,

$$\overline{B}_r(x) = \{y \in X : d(y, x) \le r\}.$$

(a) Show that $\overline{B}_r(x)$ is always closed.
(b) Show that $\overline{B_r(x)} \subset \overline{B}_r(x)$. (By definition, the first set is the closure of $B_r(x)$.)
(c) Show that it can happen that $\overline{B_r(x)} \neq \overline{B}_r(x)$.

2. Banach spaces

Let X be a complex vector space. So the elements of X ("vectors") can be added and multiplied by complex numbers ("scalars"), and these operations obey the usual algebraic rules.

Definition 2.1. A map $\|\cdot\| : X \to [0,\infty)$ is called a norm (on X) if it has the following properties for arbitrary $x, y \in X$, $c \in \mathbb{C}$:

1. $\|x\| = 0 \iff x = 0$
2. $\|cx\| = |c|\,\|x\|$
3. $\|x+y\| \le \|x\| + \|y\|$

We may interpret a given norm as assigning a length to a vector. Property (3) is again called the *triangle inequality.* It has a similar interpretation as in the case of a metric space. A vector space with a norm defined on it is called a *normed space.*

If $(X, \|\cdot\|)$ is a normed space, then $d(x,y) := \|x-y\|$ defines a metric on X.

Exercise 2.1. Prove this remark.

Therefore, all concepts and results from Chapter 1 apply to normed spaces also. In particular, a norm generates a topology on X. We repeat here some of the most basic notions: A sequence $x_n \in X$ is said to converge to $x \in X$ if $\|x_n - x\| \to 0$ (note that these norms form a sequence of *numbers*, so it's clear how to interpret this latter convergence). We call x_n a Cauchy sequence if $\|x_m - x_n\| \to 0$ as $m, n \to \infty$. The open ball of radius $r > 0$ about $x \in X$ is defined as

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

This set is indeed open in the topology mentioned above; more generally, an arbitrary set $U \subset X$ is open precisely if for every $x \in U$, there exists an $r = r(x) > 0$ so that $B_r(x) \subset U$. Finally, recall that a space is called complete if every Cauchy sequence converges. Complete normed spaces are particularly important; for easier reference, they get a special name:

Definition 2.2. A *Banach space* is a complete normed space.

The following basic properties of norms are relatively direct consequences of the definition, but they are extremely important when working on normed spaces.

Exercise 2.2. (a) Prove the *second triangle inequality:*

$$\big|\, \|x\| - \|y\| \,\big| \le \|x - y\|$$

(b) Prove that the norm is a continuous map $X \to \mathbb{R}$; put differently, if $x_n \to x$, then also $\|x_n\| \to \|x\|$.

Exercise 2.3. Prove that the vector space operations are continuous. In other words, if $x_n \to x$ and $y_n \to y$ (and $c \in \mathbb{C}$), then also $x_n + y_n \to x + y$ and $cx_n \to cx$.

Let's now collect some examples of Banach spaces. It turns out that most of the examples for metric spaces that we considered in Chapter 1 actually have a natural vector space structure and the metric comes from a norm.

Example 2.1. The simplest vector spaces are the finite-dimensional spaces. Every n-dimensional (complex) vector space is isomorphic to $\mathbb{C}^n$, so it will suffice to consider $X = \mathbb{C}^n$. We would like to define norms on this space, and we can in fact turn to Example 1.3 for inspiration. For $x = (x_1, \ldots, x_n) \in X$, let

$$\|x\|_p = \left(\sum_{j=1}^{n} |x_j|^p \right)^{1/p}, \tag{2.1}$$

for $1 \le p < \infty$, and

$$\|x\|_\infty = \max_{j=1,\ldots,n} |x_j|. \tag{2.2}$$

I claim that this defines a family of norms (one for each p, $1 \le p \le \infty$), but we will not prove this in this setting. Rather, we will right away prove a more general statement in Example 2.2 below. (Only the triangle inequality for $1 < p < \infty$ needs serious proof; everything else is fairly easy to check here anyway.)

Example 2.2. We now consider infinite-dimensional versions of the Banach spaces from the previous example. Instead of finite-dimensional vectors $(x_1, \ldots, x_n)$, we now want to work with infinite sequences $x = (x_1, x_2, \ldots)$, and we want to continue to use (2.1), (2.2), or at least something similar. We first of all introduce the maximal spaces on which these formulae seem to make sense. Let

$$\ell^p = \left\{ x = (x_n)_{n \ge 1} : \sum_{n=1}^{\infty} |x_n|^p < \infty \right\}$$

(for $1 \le p < \infty$) and

$$\ell^\infty = \left\{ x = (x_n)_{n \ge 1} : \sup_{n \ge 1} |x_n| < \infty \right\}.$$

Then, as expected, for $x \in \ell^p$, define

$$\|x\|_p = \left(\sum_{n=1}^{\infty} |x_n|^p \right)^{1/p} \qquad (p < \infty),$$
$$\|x\|_\infty = \sup_{n \geq 1} |x_n|.$$

Proposition 2.3. *ℓ^p is a Banach space for $1 \leq p \leq \infty$.*

Here, the algebraic operations on ℓ^p are defined in the obvious way: we perform them componentwise; for example, $x + y$ is the sequence whose nth element is $x_n + y_n$.

Proof. We will explicitly prove this only for $1 < p < \infty$; the cases $p = 1$, $p = \infty$ are easier and can be handled by direct arguments. First of all, we must check that ℓ^p is a vector space. Clearly, if $x \in \ell^p$ and $c \in \mathbb{C}$, then also $cx \in \ell^p$. Moreover, if $x, y \in \ell^p$, then, since $|x_n + y_n|^p \leq (2|x_n|)^p + (2|y_n|)^p$, we also have that $x + y \in \ell^p$. So addition and multiplication by scalars can be defined on all of ℓ^p, and it is clear that the required algebraic laws hold because all calculations are performed on individual components, so we just inherit the usual rules from $\mathbb{C}$.

Next, we verify that $\| \cdot \|_p$ is a norm on ℓ^p. Properties (1), (2) from Definition 2.1 are obvious. The proof of the triangle inequality will depend on the following very important inequality:

Theorem 2.4 (Hölder's inequality)**.** *Let $x \in \ell^p$, $y \in \ell^q$, where p, q satisfy*

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

($1/0 := \infty$, $1/\infty := 0$ in this context). Then $xy \in \ell^1$ and

$$\|xy\|_1 \leq \|x\|_p \|y\|_q.$$

Proof of Theorem 2.4. Again, we focus on the case $1 < p < \infty$; if $p = 1$ or $p = \infty$, an uncomplicated direct argument is available.

The function $\ln x$ is concave, that is, the graph lies above line segments connecting any two of its points (formally, this follows from the fact that $(\ln x)'' = -1/x^2 < 0$). In other words, if $a, b > 0$ and $0 \leq \alpha \leq 1$, then

$$\alpha \ln a + (1 - \alpha) \ln b \leq \ln \left(\alpha a + (1 - \alpha) b \right).$$

We apply the exponential function on both sides and obtain that $a^\alpha b^{1-\alpha} \leq \alpha a + (1 - \alpha) b$. If we introduce the new variables c, d by writing $a = c^p$,

$b = d^q$, with $1/p = \alpha$ (so $1/q = 1 - \alpha$), then this becomes

$$cd \leq \frac{c^p}{p} + \frac{d^q}{q}. \tag{2.3}$$

This holds for all $c, d \geq 0$ (the original argument is valid only if $c, d > 0$, but of course (2.3) is trivially true if $c = 0$ or $d = 0$). In particular, we can use (2.3) with $c = |x_n|/\|x\|_p$, $d = |y_n|/\|y\|_q$ (at least if $\|x\|_p, \|y\|_q \neq 0$, but if that fails, then the claim is trivial anyway) and then sum over $n \geq 1$. This shows that

$$\sum_{n=1}^{\infty} \frac{|x_n y_n|}{\|x\|_p \|y\|_q} \leq \sum_{n=1}^{\infty} \frac{|x_n|^p}{p\|x\|_p^p} + \sum_{n=1}^{\infty} \frac{|y_n|^q}{q\|y\|_q^q} = \frac{1}{p} + \frac{1}{q} = 1,$$

so $xy \in \ell^1$, as claimed, and we obtain Hölder's inequality. □

We are now in a position to establish the triangle inequality on ℓ^p:

Theorem 2.5 (Minkowski's inequality = triangle inequality on ℓ^p)**.** *Let $x, y \in \ell^p$. Then $x + y \in \ell^p$ and*

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

Proof of Theorem 2.5. Again, we will discuss explicitly only the case $1 < p < \infty$. We already know that $x + y \in \ell^p$. Hölder's inequality with the given p (and thus $q = p/(p-1)$) shows that

$$\begin{aligned} \|x + y\|_p^p = \sum |x_n + y_n|^p &= \sum |x_n + y_n|\, |x_n + y_n|^{p-1} \\ &\leq \sum |x_n|\, |x_n + y_n|^{p-1} + \sum |y_n|\, |x_n + y_n|^{p-1} \\ &\leq (\|x\|_p + \|y\|_p)\, \|x + y\|_p^{p-1}. \end{aligned}$$

If $x+y \neq 0$, we can divide by $\|x+y\|_p^{p-1}$ to obtain the desired inequality, and if $x + y = 0$, then the claim is trivial. □

It remains to show that ℓ^p is complete. So let $x^{(n)} \in \ell^p$ be a Cauchy sequence (since the elements of ℓ^p are themselves sequences, we really have a sequence whose members are sequences; we use a *super*script to label the elements of the Cauchy sequence from $X = \ell^p$ to avoid confusion with the index labeling the components of a fixed element of ℓ^p). Clearly,

$$\left| x_j^{(m)} - x_j^{(m)} \right|^p \leq \|x^{(m)} - x^{(n)}\|_p^p$$

for each fixed $j \geq 1$, so $\left(x_j^{(n)} \right)_{n \geq 1}$ is a Cauchy sequence of complex numbers. Now $\mathbb{C}$ is complete, so these sequences have limits in $\mathbb{C}$. Define

$$x_j = \lim_{n \to \infty} x_j^{(n)}.$$

I claim that $x = (x_j) \in \ell^p$ and $x^{(n)} \to x$ in the norm of ℓ^p. To verify that $x \in \ell^p$, we observe that for arbitrary $N \in \mathbb{N}$,

$$\sum_{j=1}^{N} |x_j|^p = \lim_{n\to\infty} \sum_{j=1}^{N} \left| x_j^{(n)} \right|^p \le \limsup_{n\to\infty} \|x^{(n)}\|^p.$$

Exercise 2.4. Let $x_n \in X$ be Cauchy sequence in a normed space X. Prove that x_n is bounded in the following sense: There exists $C > 0$ so that $\|x_n\| \le C$ for all $n \ge 1$.

Exercise 2.4 now shows that

$$\sum_{j=1}^{N} |x_j|^p \le C$$

for some fixed, N independent constant C, so $x \in \ell^p$, as required.

It remains to show that $\|x^{(n)} - x\|_p \to 0$. Let $\epsilon > 0$ be given and pick $N_0 \in \mathbb{N}$ so large that $\|x^{(n)} - x^{(m)}\| < \epsilon$ if $m, n \ge N_0$ (this is possible because $x^{(n)}$ is a Cauchy sequence). Then, for fixed $N \in \mathbb{N}$, we have that

$$\sum_{j=1}^{N} \left| x_j^{(n)} - x_j \right|^p = \lim_{m\to\infty} \sum_{j=1}^{N} \left| x_j^{(n)} - x_j^{(m)} \right|^p \le \epsilon$$

if $n \ge N_0$. Since $N \in \mathbb{N}$ was arbitrary, it also follows that $\|x^{(n)} - x\|_p^p \le \epsilon$ for $n \ge N_0$. $\square$

Similar spaces can be defined for arbitrary index sets I instead of $\mathbb{N}$. For example, by definition, the elements of $\ell^p(I)$ are complex valued functions $x : I \to \mathbb{C}$ with

$$\sum_{j\in I} |x_j|^p < \infty. \tag{2.4}$$

If I is uncountable, this sum needs interpretation. We can do this by hand, as follows: (2.4) means that $x_j \ne 0$ only for countably many $j \in I$, and the corresponding sum is finite. Alternatively, and more elegantly, we can also use the counting measure on I and interpret the sum as an integral.

If we want to emphasize the fact that we're using $\mathbb{N}$ as the index set, we can also denote the spaces discussed above by $\ell^p(\mathbb{N})$. When no confusion has to be feared, we will usually prefer the shorter notation ℓ^p. We can also consider finite index sets $I = \{1, 2, \dots, n\}$. We then obtain that $\ell^p(\{1, 2, \dots, n\}) = \mathbb{C}^n$ as a set, and the norms on these spaces are the ones that were already introduced in Example 2.1 above.

Example 2.3. Two more spaces of sequences are in common use. In both cases, the index set is usually $\mathbb{N}$ (or sometimes $\mathbb{Z}$). Put

$$c = \left\{ x : \lim_{n\to\infty} x_n \text{ exists} \right\},$$
$$c_0 = \left\{ x : \lim_{n\to\infty} x_n = 0 \right\}.$$

It is clear that $c_0 \subset c \subset \ell^\infty$. In fact, more is true: the smaller spaces are (algebraic linear) subspaces of the bigger spaces. On c and c_0, we also use the norm $\|\cdot\|_\infty$ (as on the big space ℓ^∞).

Proposition 2.6. *c and c_0 are Banach spaces.*

Proof. We can make use of the observation made above, that $c_0 \subset c \subset \ell^\infty$ and then refer to the following fact:

Proposition 2.7. *Let $(X, \|\cdot\|)$ be a Banach space, and let $Y \subset X$. Then $(Y, \|\cdot\|)$ is a Banach space if and only if Y is a closed (linear) subspace of X.*

Exercise 2.5. Prove Proposition 2.7. Recall that on metric (and thus also normed and Banach) spaces, you can use sequences to characterize topological notions. So a subset is closed precisely if all limits of convergent sequences from the set lie in the set again.

So we only need to show that c and c_0 are closed in ℓ^∞.

Exercise 2.6. Complete the proof of Proposition 2.6 along these lines.

□

Example 2.4. *Function spaces* provide another very important class of Banach spaces. The discussion is in large parts analogous to our treatment of sequence spaces (Examples 2.2, 2.3); sometimes, sequence spaces are somewhat more convenient to deal with and, as we will see in a moment, they can actually be interpreted as function spaces of a particular type.

Let $(X, \mathcal{M}, \mu)$ be a measure space (with a positive measure μ). The discussion is most conveniently done in this completely general setting, but if you prefer a more concrete example, you could think of $X = \mathbb{R}^n$ with Lebesgue measure, as what is probably the most important special case.

Recalling what we did above, it now seems natural to introduce (for $1 \le p < \infty$)

$$\mathcal{L}^p(X, \mu) = \left\{ f : X \to \mathbb{C} : f \text{ measurable}, \int_X |f(x)|^p \, d\mu(x) < \infty \right\}.$$

Note that this set also depends on the σ-algebra $\mathcal{M}$, but this dependence is not made explicit in the notation. We would then like to define

$$\|f\|_p = \left(\int_X |f|^p \, d\mu \right)^{1/p} .$$

This, however, does not give a norm in general because $\|f\|_p = 0$ precisely if $f = 0$ almost everywhere, so usually there will be functions of zero "norm" that are not identically equal to zero. Fortunately, there is an easy fix for this problem: we simply identify functions that agree almost everywhere. More formally, we introduce an equivalence relation on $\mathcal{L}^p$, as follows:

$$f \sim g \iff f(x) = g(x) \quad \text{for } \mu\text{-almost every } x \in X$$

We then let L^p be the set of equivalence classes:

$$L^p(X, \mu) = \{(f) : f \in \mathcal{L}^p(X, \mu)\} ,$$

where $(f) = \{g \in \mathcal{L}^p : g \sim f\}$. We obtain a vector space structure on L^p in the obvious way; for example, $(f) + (g) := (f + g)$ (it needs to be checked here that the equivalence class on the right-hand side is independent of the choice of representatives f, g, but this is obvious from the definitions). Moreover, we can put

$$\|(f)\|_p := \|f\|_p;$$

again, it doesn't matter which function from (f) we take on the right-hand side, so this is well defined.

In the same spirit ("ignore what happens on null sets"), we define

$$\mathcal{L}^\infty(X, \mu) = \{f : X \to \mathbb{C} : f \text{ essentially bounded}\}.$$

A function f is called *essentially bounded* if there is a null set $N \subset X$ so that $|f(x)| \le C$ for $x \in X \setminus N$. Such a C is called an *essential bound.* If f is essentially bounded, its *essential supremum* is defined as the best essential bound:

$$\begin{aligned} \text{ess sup } |f(x)| &= \inf_{N:\mu(N)=0} \sup_{x \in X \setminus N} |f(x)| \\ &= \inf\{C \ge 0 : \mu(\{x \in X : |f(x)| > C\}) = 0\} \end{aligned}$$

Exercise 2.7. (a) Prove that both formulae give the same result.
(b) Prove that ess sup $|f|$ is itself an essential bound: $|f| \le$ ess sup $|f|$ almost everywhere.

Finally, we again let

$$L^\infty = \{(f) : f \in \mathcal{L}^\infty\} ,$$

and we put

$$\|(f)\|_\infty = \text{ess sup } |f(x)|.$$

Strictly speaking, the elements of the spaces L^p are not functions, but equivalence classes of functions. Sometimes, it is important to keep this distinction in mind; for example, it doesn't make sense to talk about $f(0)$ for an $(f) \in L^1(\mathbb{R}, m)$, say, because $m(\{0\}) = 0$, so we can change f at $x = 0$ without leaving the equivalence class (f). However, for most purposes, no damage is done if, for convenience and as a figure of speech, we simply refer to the elements of L^p as "functions" anyway (as in "let f be a function from L^1", rather than the pedantic and clumsy "let F be an element of L^1 and pick a function $f \in \mathcal{L}^1$ that represents the equivalence class F"). This convention is in universal use (it is similar to, say, "right lane must exit").

Proposition 2.8. *$L^p(X, \mu)$ is a Banach space for $1 \le p \le \infty$.*

We will not give the complete proof of this because the discussion is reasonably close to our previous treatment of ℓ^p. Again, the two main issues are the triangle inequality and completeness. The proof of the triangle inequality follows the pattern of the above proof very closely. To establish completeness, we (unsurprisingly) need facts from the theory of the Lebesgue integral, so this gives us a good opportunity to review some of these tools. We will give this proof only for $p = 1$ ($1 < p < \infty$ is similar, and $p = \infty$ can again be handled by a rather direct argument).

So let $f_n \in L^1$ be a Cauchy sequence. Pick a subsequence $n_k \to \infty$ so that $\|f_{n_{k+1}} - f_{n_k}\| < 2^{-k}$.

Exercise 2.8. Prove that n_k's with these properties can indeed be found.

Let

$$S_j(x) = \sum_{k=1}^{j} \left| f_{n_{k+1}}(x) - f_{n_k}(x) \right|.$$

Then S_j is measurable, non-negative, and $S_{j+1} \ge S_j$. So, if we let $S(x) = \lim_{j\to\infty} S_j(x) \in [0, \infty]$, then the Monotone Convergence Theorem shows that

$$\int_X S\, d\mu = \lim_{j\to\infty} \int_X S_j\, d\mu = \lim_{j\to\infty} \sum_{k=1}^{j} \int_X \left| f_{n_{k+1}} - f_{n_k} \right| d\mu$$

$$= \lim_{j\to\infty} \sum_{k=1}^{j} \left\| f_{n_{k+1}} - f_{n_k} \right\| < \sum_{k=1}^{\infty} 2^{-k} = 1.$$

In particular, $S \in L^1$, and this implies that $S < \infty$ almost everywhere.

The same conclusion can be obtained from Fatou's Lemma; let us do this too, to get some additional practice:

$$\int_X S\,d\mu = \int_X \lim_{j\to\infty} S_j\,d\mu = \int_X \liminf_{j\to\infty} S_j\,d\mu \le \liminf_{j\to\infty} \int_X S_j\,d\mu$$

We can conclude the argument as in the preceding paragraph, and we again see that $\int S < 1$, so $S < \infty$ almost everywhere.

For almost every $x \in X$, we can define

$$f(x) := f_{n_1}(x) + \sum_{k=1}^{\infty} \left(f_{n_{k+1}}(x) - f_{n_k}(x) \right);$$

indeed, we just verified that this series actually converges *absolutely* for almost every $x \in X$. Moreover, the sum is telescoping, so in fact

$$f(x) = \lim_{j\to\infty} f_{n_j}(x)$$

for a.e. x. Also,

$$\left| f(x) - f_{n_j}(x) \right| \le \sum_{k=j}^{\infty} \left| f_{n_{k+1}}(x) - f_{n_k}(x) \right|.$$

Since this latter sum is dominated by $S \in L^1$, this shows, first of all, that $|f - f_{n_j}| \in L^1$ and thus also $f \in L^1$ (because $|f| \le |f_{n_j}| + |f - f_{n_j}|$). Moreover, the functions $|f - f_{n_j}|$ satisfy the hypotheses of Dominated Convergence, so we obtain that

$$\lim_{j\to\infty} \int_X \left| f - f_{n_j} \right| \, d\mu = 0.$$

To summarize: given the Cauchy sequence $f_n \in L^1$, we have constructed a function $f \in L^1$, and $\|f_{n_j} - f\| \to 0$. This is almost what we set out to prove. For the final step, we can refer to the following general fact.

Exercise 2.9. Let x_n be a Cauchy sequence from a metric space Y. Suppose that $x_{n_j} \to x$ for some subsequence (and some $x \in Y$). Prove that then in fact $x_n \to x$.

We also saw in this proof that $f_{n_j} \to f$ pointwise almost everywhere. This is an extremely useful fact, so it's a good idea to state it again (for general p).

Corollary 2.9. *If $\|f_n - f\|_p \to 0$, then there exists a subsequence f_{n_j} that converges to f pointwise almost everywhere.*

Exercise 2.10. Give a (short) direct argument for the case $p = \infty$. Show that in this case, it is not necessary to pass to a subsequence.

If I is an arbitrary set (the case $I = \mathbb{N}$ is of particular interest here), $\mathcal{M} = \mathcal{P}(I)$ and μ is the counting measure on I (so $\mu(A)$ equals the number of elements of A), then $L^p(I, \mu)$ is the space $\ell^p(I)$ that was discussed earlier, in Example 2.2. Note that on this measure space, the only null set is the empty set, so there's no difference between $\mathcal{L}^p$ and L^p here.

Example 2.5. Our final example can perhaps be viewed as a mere variant of L^∞, but this space will become very important for us later on. We start out with a compact Hausdorff space K. A popular choice would be $K = [a, b]$, with the usual topology, but the general case will also be needed. We now consider

$$C(K) = \{f : K \to \mathbb{C} : f \text{ continuous }\},$$

with the norm

$$\|f\| = \|f\|_\infty = \max_{x \in K} |f(x)|.$$

The maximum exists because $|f|(K)$, being a continuous image of a compact space, is a compact subset of $\mathbb{R}$. As anticipated, we then have the following:

Proposition 2.10. *$\|\cdot\|_\infty$ is a norm on $C(K)$, and $C(K)$ with this norm is a Banach space.*

The proof is very similar to the corresponding discussion of L^∞; I don't want to discuss it in detail here. In fact, if there is a measure on K that gives positive weight to all non-empty open sets (such as Lebesgue measure on $[a, b]$), then $C(K)$ can be thought of as a subspace of L^∞.

Exercise 2.11. Can you imagine why we want the measure to give positive weight to open sets?

Hint: Note that the elements of $C(K)$ are genuine functions, while the elements of $L^\infty(K, \mu)$ were defined as equivalence classes of functions, so if we want to think of $C(K)$ as a subset of L^∞, we need a way to identify continuous functions with equivalence classes.

Exercise 2.12. Prove that $C(K)$ is complete.

In the sequel, we will be interested mainly in linear maps between Banach spaces (and not so much in the spaces themselves). More generally, let X, Y be normed spaces. Recall that a map $A : X \to Y$ is called *linear* if $A(x_1 + x_2) = Ax_1 + Ax_2$ and $A(cx) = cAx$. In functional analysis, we usually refer to linear maps as (linear) *operators*. The *null* *space* (or *kernel*) and the *range* (or *image*) of an operator A are defined as follows:

$$N(A) = \{x \in X : Ax = 0\},$$
$$R(A) = \{Ax : x \in X\}$$

Theorem 2.11. *Let $A : X \to Y$ be a linear operator. Then the following are equivalent:*
(a) A is continuous (everywhere);
(b) A is continuous at $x = 0$;
(c) A is bounded: *There exists a constant $C \ge 0$ so that $\|Ax\| \le C\|x\|$ for all $x \in X$.*

Proof. $(a) \implies (b)$: This is trivial.

$(b) \implies (c)$: Suppose that A was not bounded. Then we can find, for every $n \in \mathbb{N}$, a vector $x_n \in X$ so that $\|Ax_n\| > n\|x_n\|$. Let $y_n = (1/(n\|x_n\|))x_n$. Then $\|y_n\| = 1/n$, so $y_n \to 0$, but $\|Ay_n\| > 1$, so Ay_n can not go to the zero vector, contradicting (b).

$(c) \implies (a)$: Suppose that $x_n \to x$. We want to show that then also $Ax_n \to Ax$, and indeed this follows immediately from the linearity and boundedness of A:

$$\|Ax_n - Ax\| = \|A(x_n - x)\| \le C\|x_n - x\| \to 0$$

□

Given two normed spaces X, Y, we introduce the space $B(X, Y)$ of bounded (or continuous) linear operators from X to Y. The special case $X = Y$ is of particular interest; in this case, we usually write $B(X)$ instead of $B(X, X)$.

$B(X, Y)$ becomes a vector space if addition and multiplication by scalars are defined in the obvious way (for example, $(A+B)x := Ax + Bx$). We can go further and also introduce a norm on $B(X, Y)$, as follows:

$$\|A\| = \sup_{x \ne 0} \frac{\|Ax\|}{\|x\|}$$

Since A is assumed to be bounded here, the supremum will be finite. We call $\|A\|$ the *operator norm* of A (that this is a norm will be seen in Theorem 2.12 below).

There are a number of ways to compute $\|A\|$.

Exercise 2.13. Prove the following formulae for $\|A\|$ (for $A \in B(X,Y)$):

$$\begin{aligned} \|A\| &= \inf\{C \geq 0 : \|Ax\| \leq C\|x\| \text{ for all } x \in X\} \\ &= \min\{C \geq 0 : \|Ax\| \leq C\|x\| \text{ for all } x \in X\} \\ \|A\| &= \sup_{\|x\|=1} \|Ax\| \end{aligned}$$

In particular, we have that $\|Ax\| \leq \|A\| \, \|x\|$, and $\|A\|$ is the smallest constant for which this inequality holds.

Exercise 2.14. However, it is not necessarily true that $\|A\| = \max_{\|x\|=1} \|Ax\|$. Provide an example of such an operator A.

Suggestion: $X = Y = c_0$ (or ℓ^1 if you prefer, this also works very smoothly), and define $(Ax)_n = a_n x_n$, where a_n is a suitably chosen bounded sequence.

Theorem 2.12. *(a) $B(X,Y)$ with the operator norm is a normed space.*
(b) If Y is a Banach space, then $B(X,Y)$ (with the operator norm) is a Banach space.

The special case $Y = \mathbb{C}$ (recall that this is a Banach space if we use the absolute value as the norm) is particularly important. We use the alternative notation $X^* = B(X, \mathbb{C})$, and we call the elements of X^* (continuous, linear) *functionals.* X^* itself is called the *dual space* (or just the dual) of X.

This must not be confused with the dual space from linear algebra, which is defined as the set of all linear maps from the original vector space back to its base field (considered as a vector space also). This is of limited use in functional analysis. The (topological) dual X^* consists only of continuous maps; it is usually much smaller than the algebraic dual described above.

Proof. (a) We observed earlier that $B(X,Y)$ is a vector space, so we need to check that the operator norm satisfies the properties from Definition 2.1. First of all, we will have $\|A\| = 0$ precisely if $Ax = 0$ for all $x \in X$, that is, precisely if A is the zero map or, put differently, $A = 0$ in $B(X,Y)$. Next, if $c \in \mathbb{C}$ and $A \in B(X,Y)$, then

$$\|cA\| = \sup_{\|x\|=1} \|cAx\| = \sup_{\|x\|=1} |c|\|Ax\| = |c|\|A\|.$$

A similar calculation establishes the third property from Definition 2.1:

$$\|A + B\| = \sup_{\|x\|=1} \|(A + B)x\| \leq \sup_{\|x\|=1} (\|Ax\| + \|Bx\|) \leq \|A\| + \|B\|$$

(b) Let A_n be a Cauchy sequence from $B(X,Y)$. We must show that A_n converges. Observe that for fixed x, $A_n x$ will be a Cauchy sequence in Y. Indeed,

$$\|A_m x - A_n x\| \le \|A_m - A_n\| \|x\|$$

can be made arbitrarily small by taking both m and n large enough. Since Y is now assumed to be complete, the limits $Ax := \lim_{n\to\infty} A_n x$ exist, and we can define a map A on X in this way. We first check that A is linear:

$$\begin{aligned} A(x_1 + x_2) &= \lim_{n\to\infty} A_n(x_1 + x_2) = \lim_{n\to\infty} (A_n x_1 + A_n x_2) \\ &= \lim_{n\to\infty} A_n x_1 + \lim_{n\to\infty} A_n x_1 = Ax_1 + Ax_2, \end{aligned}$$

and a similar (if anything, this is easier) argument shows that $A(cx) = cAx$.

A is also bounded because

$$\|Ax\| = \|\lim A_n x\| = \lim \|A_n x\| \le (\sup \|A_n\|) \|x\|;$$

the supremum is finite because $|\|A_m\| - \|A_n\|| \le \|A_m - A_n\|$, so $\|A_n\|$ forms a Cauchy sequence of real numbers and thus is convergent and, in particular, bounded. Notice also that we used the continuity of the norm for the second equality (see Exercise 2.2(b)).

Summing up: we have constructed a map A and confirmed that in fact $A \in B(X,Y)$. The final step will be to show that $A_n \to A$, with respect to the operator norm in $B(X,Y)$. Let $x \in X$, $\|x\| = 1$. Then, by the continuity of the norm again,

$$\|(A - A_n)x\| = \lim_{m\to\infty} \|(A_m - A_n)x\| \le \limsup_{m\to\infty} \|A_m - A_n\|.$$

Since x was arbitrary, it also follows that

$$\|A - A_n\| \le \limsup_{m\to\infty} \|A_m - A_n\|.$$

Since A_n is a Cauchy sequence, the $\limsup$ can be made arbitrarily small by taking n large enough. □

Perhaps somewhat surprisingly, there are discontinuous linear maps if the first space, X, is infinite-dimensional. We can then even take $Y = \mathbb{C}$. An abstract construction can be done as follows: Let $\{e_\alpha\}$ be an algebraic basis of X (that is, every $x \in X$ can be written in a unique way as a linear combination of (finitely many) e_α's). For arbitrary complex numbers c_α, there exists a linear map $A : X \to \mathbb{C}$ with $Ae_\alpha = c_\alpha \|e_\alpha\|$.

Exercise 2.15. This problem reviews the linear algebra fact needed here. Let V, W be vector spaces (over $\mathbb{C}$, say), and let $\{e_\alpha\}$ be a basis of V. Show that for every collection of vectors $w_\alpha \in W$, there exists a unique linear map $A : V \to W$ so that $Ae_\alpha = w_\alpha$ for all α.

Since $\|Ae_\alpha\|/\|e_\alpha\| = |c_\alpha|$, we see that A can not be bounded if $\sup_\alpha |c_\alpha| = \infty$.

On the other hand, if $\dim X < \infty$, then linear operators $A : X \to Y$ are always bounded. We will see this in a moment; before we do this, we introduce a new concept and prove a related result.

Definition 2.13. Two norms on a common space X are called *equivalent* if they generate the same topology.

This can be restated in a less abstract way:

Proposition 2.14. *The norms $\|\cdot\|_1$, $\|\cdot\|_2$ are equivalent if and only if there are constants $C_1, C_2 > 0$ so that*

$$C_1\|x\|_1 \le \|x\|_2 \le C_2\|x\|_1 \qquad \text{for all } x \in X. \tag{2.5}$$

Proof. Consider the identity as a map from $(X, \|\cdot\|_1)$ to $(X, \|\cdot\|_2)$. Clearly, this is bijective, and, by Theorem 2.11 this map and its inverse are continuous precisely if (2.5) holds. Put differently, (2.5) is equivalent to the identity map being a homeomorphism (a bijective continuous map with continuous inverse), and this holds if and only if $(X, \|\cdot\|_1)$ and $(X, \|\cdot\|_2)$ have the same topology. □

Exercise 2.16. (a) Let $\|\cdot\|_1$, $\|\cdot\|_2$ be equivalent norms on X. Show that then $(X, \|\cdot\|_1)$ and $(X, \|\cdot\|_2)$ are either both complete or both not complete.

(b) Construct a metric d on $\mathbb{R}$ that produces the usual topology, but $(\mathbb{R}, d)$ is not complete. (Since $(\mathbb{R}, |\cdot|)$ has the same topology and is complete, this shows that the analog of (a) for metric spaces is false.)

Theorem 2.15. *Let X be a (complex) vector space with* $\dim X < \infty$. *Then all norms on X are equivalent.*

In particular, by combining Example 2.1 with Exercise 2.16, we see that finite-dimensional normed spaces are automatically complete and thus Banach spaces.

Proof. By fixing a basis on X, we may assume that $X = \mathbb{C}^n$. We will show that every norm on $\mathbb{C}^n$ is equivalent to $\|\cdot\|_1$. We will do this by verifying (2.5). So let $\|\cdot\|$ be a norm. Then, first of all,

$$\|x\| = \left\|\sum_{j=1}^n x_j e_j\right\| \le \sum_{j=1}^n |x_j|\,\|e_j\| \le \left(\max_{j=1,\dots,n} \|e_j\|\right)\|x\|_1. \tag{2.6}$$

To obtain the other inequality, consider again the identity as a map from $(\mathbb{C}^n, \|\cdot\|_1)$ to $(\mathbb{C}^n, \|\cdot\|)$. As we have just seen in (2.6), this map is bounded, thus continuous. Since a norm always defines a continuous map, we obtain that the composite map from $(\mathbb{C}^n, \|\cdot\|_1)$ to $\mathbb{R}$, $x \mapsto \|x\|$ is also continuous. Now $\{x \in \mathbb{C}^n : \|x\|_1 = 1\}$ is a compact subset of $\mathbb{C}^n$, with respect to the topology generated by $\|\cdot\|_1$ (which is just the usual topology on $\mathbb{C}^n$). Therefore, the image under our map, which is given by $\{\|x\| : \|x\|_1 = 1\}$ is a compact subset of $\mathbb{R}$, and it doesn't contain zero, so

$$\inf_{\|x\|_1=1} \|x\| = \min_{\|x\|_1=1} \|x\| =: c > 0,$$

and the homogeneity of norms now implies that $\|x\| \geq c\|x\|_1$ for all $x \in \mathbb{C}^n$, as required. □

Corollary 2.16. *Suppose that* $\dim X < \infty$*, and let* $A : X \to Y$ *be a linear operator. Then* A *is bounded.*

Proof. By Theorem 2.15, it suffices to discuss the case $X = \mathbb{C}^n$, equipped with the norm $\|\cdot\|_1$. As above, we estimate

$$\|Ax\| = \left\| A\left(\sum_{j=1}^n x_j e_j\right)\right\| \leq \sum_{j=1}^n |x_j|\, \|Ae_j\| \leq \left(\max_{j=1,\ldots,n} \|Ae_j\|\right) \|x\|_1.$$

□

We conclude this chapter by discussing sums and quotients of Banach spaces. Let $X_1, \ldots, X_n$ be Banach spaces. We form their direct sum (as vector spaces). More precisely, we introduce

$$X = \{(x_1, \ldots, x_n) : x_j \in X_j\};$$

this becomes a vector space in the obvious way: the algebraic operations are defined componentwise. Of course, we want more: We want to introduce a norm on X that makes X a Banach space, too. This can be done in several ways; for example, the following works.

Theorem 2.17. $\|x\| = \sum_{j=1}^n \|x_j\|_j$ *defines a norm on* X*, and with this norm,* X *is a Banach space.*

Exercise 2.17. Prove Theorem 2.17.

We will denote this new Banach space by $X = \bigoplus_{j=1}^n X_j$.

Moving on to quotients now, we consider a Banach space X and a closed subspace $M \subset X$.

Exercise 2.18. (a) In general, subspaces need not be closed. Give an example of a dense subspace $M \subset \ell^1$, $M \neq \ell^1$ (in other words, we want $\overline{M} = \ell^1$, $M \neq \ell^1$; in particular, such an M is definitely not closed).

(b) What about open subspaces of a normed space?

Exercise 2.19. However, show that *finite-dimensional* subspaces of a normed space are always closed.

Suggestion: Use Theorem 2.15.

As a vector space, we define the quotient X/M as the set of equivalence classes (x), $x \in X$, where $x, y \in X$ are equivalent if $x-y \in M$. So $(x) = x+M = \{x+m : m \in M\}$, and to obtain a vector space structure on X/M, we do all calculations with representatives. In other words, $(x) + (y) := (x+y)$, $c(x) := (cx)$, and this is well defined, because the right-hand sides are independent of the choice of representatives x, y.

Theorem 2.18. $\|(x)\| := \inf_{y \in (x)} \|y\|$ *defines a norm on* X*, and* X/M *with this norm is a Banach space.*

Proof. First of all, we must check the conditions from Definition 2.1. We have that $\|(x)\| = 0$ precisely if there are $m_n \in M$ so that $\|x - m_n\| \to 0$. This holds if and only if $x \in \overline{M}$, but M is assumed to be closed, so $\|(x)\| = 0$ if and only if $x \in M$, that is, if and only if x represents the zero vector from X/M (equivalently, $(x) = (0)$).

If $c \in \mathbb{C}$, $c \neq 0$, then

$$\begin{aligned} \|c(x)\| = \|(cx)\| &= \inf_{m \in M} \|cx - m\| = \inf_{m \in M} \|cx - cm\| \\ &= |c| \inf_{m \in M} \|x - m\| = |c| \, \|(x)\|. \end{aligned}$$

If $c = 0$, then this identity ($\|0(x)\| = 0\|(x)\|$) is also true and in fact trivial.

The triangle inequality follows from a similar calculation:

$$\begin{aligned} \|(x) + (y)\| = \|(x+y)\| &= \inf_{m \in M} \|x + y - m\| = \inf_{m,n \in M} \|x + y - m - n\| \\ &\le \inf_{m,n \in M} (\|x - m\| + \|y - n\|) = \|(x)\| + \|(y)\| \end{aligned}$$

Finally, we show that X/M is complete. Let (x_n) be a Cauchy sequence. Pass again to a subsequence, so that $\|(x_{n_{j+1}}) - (x_{n_j})\| < 2^{-j}$ (see Exercise 2.8). Since the quotient norm was defined as the infimum of the norms of the representatives, we can now also (inductively) find representatives (we may assume that these are the x_n's themselves) so that $\|x_{n_{j+1}} - x_{n_j}\| < 2^{-j}$. Since $\sum 2^{-j} < \infty$, it follows that x_{n_j} is a Cauchy sequence in X, so $x = \lim_{j \to \infty} x_{n_j}$ exists. But then we also have that

$$\|(x) - (x_{n_j})\| \le \|x - x_{n_j}\| \to 0,$$

so a subsequence of the original Cauchy sequence (x_n) converges, and this forces the whole sequence to converge; see Exercise 2.9. $\square$

Exercise 2.20. Let X be a normed space, and define

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

Show that $\overline{B}_r(x) = \overline{B_r(x)}$, where the right-hand side is the closure of the (open) ball $B_r(x)$. (Compare Exercise 1.16, which discussed the analogous problem on metric spaces.)

Exercise 2.21. Call a subset B of a Banach space X *bounded* if there exists $C \ge 0$ so that $\|x\| \le C$ for all $x \in B$.
(a) Show that if $K \subset X$ is compact, then K is closed and bounded.
(b) Consider $X = \ell^\infty$, $B = \overline{B}_1(0) = \{x \in \ell^\infty : \|x\| \le 1\}$. Show that B is closed and bounded, but not compact (in fact, the closed unit ball of an infinite-dimensional Banach space is never compact).

Exercise 2.22. If x_n are elements of a normed space X, we define, as usual, the *series* $\sum_{n=1}^\infty x_n$ as the limit as $N \to \infty$ of the partial sums $S_N = \sum_{n=1}^N x_n$, if this limit exists (of course, this limit needs to be taken with respect to the norm, so $S = \sum_{n=1}^\infty x_j$ means that $\|S - S_N\| \to 0$). Otherwise, the series is said to be divergent. Call a series *absolutely convergent* if $\sum_{n=1}^\infty \|x_n\| < \infty$.

Prove that a normed space is complete if and only if every absolutely convergent series converges.

Exercise 2.23. Find the operator norm of the identity map $(x \mapsto x)$ as an operator
(a) from $(\mathbb{C}^n, \|\cdot\|_1)$ to $(\mathbb{C}^n, \|\cdot\|_2)$;
(b) from $(\mathbb{C}^n, \|\cdot\|_2)$ to $(\mathbb{C}^n, \|\cdot\|_1)$.

Exercise 2.24. Find the operator norms of the following operators on $\ell^2(\mathbb{Z})$. In particular, prove that these operators are bounded.

$$(Ax)_n = x_{n+1} + x_{n-1}, \qquad (Bx)_n = \frac{n^2}{n^2+1}\, x_n$$

Exercise 2.25. Let X, Y, Z be Banach spaces, and let $S \in B(X, Y)$, $T \in B(Y, Z)$. Show that the composition TS lies in $B(X, Z)$ and $\|TS\| \le \|T\|\,\|S\|$. Show also that strict inequality is possible here. Give an example; as always, it's sound strategy to try to keep this as simple as possible. Here, finite-dimensional spaces X, Y, Z should suffice.

Exercise 2.26. Let X, Y be Banach spaces and let M be a dense subspace of X (there is nothing unusual about that on infinite-dimensional spaces; compare Exercise 2.18). Prove the following: Every $A_0 \in B(M, Y)$ has a unique continuous extension to X. Moreover, if we call this extension A, then $A \in B(X, Y)$ (by construction, A is continuous, so we're now claiming that A is also linear), and $\|A\| = \|A_0\|$.

Exercise 2.27. (a) Let $A \in B(X, Y)$. Prove that $N(A)$ is a closed subspace of X.
(b) Now assume that F is a linear functional on X, that is, a linear map $F : X \to \mathbb{C}$. Show that F is continuous if $N(F)$ is closed (so, for linear functionals, continuity is equivalent to $N(F)$ being closed).

Suggestion: Suppose F is not continuous, so that we can find $x_n \in X$ with $\|x_n\| = 1$ and $|F(x_n)| \geq n$, say. Also, fix another vector $z \notin N(F)$ (what if $N(F) = X$?). Use these data to construct a sequence $y_n \in N(F)$ that converges to a vector not from $N(F)$. (If this doesn't seem helpful, don't give up just yet, but try something else; the proof is quite short.)

3. Consequences of Baire's Theorem

In this chapter, we discuss four fundamental functional analytic theorems that are direct descendants of Baire's Theorem (Theorem 1.10). All four results have a somewhat paradoxical character; the assumptions look too weak to give the desired conclusions, but somehow we get these anyway.

Theorem 3.1 (Uniform boundedness principle)**.** *Let X be a Banach space and let Y be a normed space. Assume that $\mathcal{F} \subset B(X,Y)$ is a family of bounded linear operators that is bounded* pointwise *in the following sense: For each $x \in X$, there exists $C_x \ge 0$ so that $\|Ax\| \le C_x$ for all $A \in \mathcal{F}$. Then $\mathcal{F}$ is* uniformly *bounded, that is,* $\sup_{A\in\mathcal{F}} \|A\| < \infty$.

Proof. Let $M_n = \{x \in X : \|Ax\| \le n \text{ for all } A \in \mathcal{F}\}$. Then M_n is a closed subset X. Indeed, we can write

$$M_n = \bigcap_{A\in\mathcal{F}} \{x \in X : \|Ax\| \le n\},$$

and these sets are closed because they are the inverse images under A of the closed ball $\overline{B}_n(0)$. Moreover, the assumption that $\mathcal{F}$ is pointwise bounded says that $\bigcup_{n\in\mathbb{N}} M_n = X$. Therefore, by Baire's Theorem, at least one of the M_n's is not nowhere dense. Fix such an n, and let $B_r(x_0)$ be an open ball contained in M_n. In other words, we now know that if $\|y - x_0\| < r$, then $\|Ay\| \le n$ for all $A \in \mathcal{F}$. In particular, if $x \in X$ is arbitrary with $\|x\| = 1$, then $y = x_0 + (r/2)x$ is such a vector and thus

$$\begin{aligned} \|Ax\| &= \frac{2}{r} \|A(y - x_0)\| \le \frac{2}{r} (\|Ay\| + \|Ax_0\|) \\ &\le \frac{2}{r} (n + C_{x_0}) \equiv D. \end{aligned}$$

The constant D is independent of x, so we also obtain that $\|A\| \le D$, and since D is also independent of $A \in \mathcal{F}$, this is what we claimed. $\square$

Theorem 3.2 (The open mapping theorem)**.** *Let X, Y be Banach spaces, and assume that $A \in B(X,Y)$ is surjective (that is, $R(A) = Y$). Then A is an open map: if $U \subset X$ is open, then $A(U)$ is also open (in Y).*

The condition defining an open map is of course similar to the corresponding property of continuous maps (see Proposition 1.5), but it goes in the other direction. In particular, that means that the inverse of an open map, if it exists, is continuous. Therefore, the open mapping theorem has the following consequence:

Corollary 3.3. *Let X, Y be Banach spaces, and assume that $A \in B(X,Y)$ is bijective. Then $A^{-1} \in B(Y,X)$.*

Exercise 3.1. Prove the following linear algebra fact: The inverse of an invertible linear map is linear.

Proof. By Exercise 3.1, A^{-1} is linear. By the open mapping theorem and the subsequent remarks, A^{-1} is continuous. □

Proof of Theorem 3.2. Let $U \subset X$ be an open set, and let $y \in A(U)$, so $y = Ax$ for some $x \in X$ (perhaps there are several such x, but then we just pick one of these). We want to show that there exists $r > 0$ so that $B_r(y) \subset A(U)$. Since $y \in A(U)$ was arbitrary, this will prove that $A(U)$ is open.

We know that $B_\epsilon(x) \subset U$ for some $\epsilon > 0$, so it actually suffices to discuss the case where $U = B_\epsilon(x)$. In fact, this can be further reduced: it is enough to consider $x, y = 0$, and it then suffices to show that for *some* $R > 0$, the set $A(B_R(0))$ contains a ball $B_r(0)$ for some $r > 0$. Indeed, if this holds, then, using the linearity of A, we will also obtain that

$$A(B_\epsilon(x)) = Ax + \frac{\epsilon}{R} A(B_R(0)) \supset Ax + \frac{\epsilon}{R} B_r(0) = B_{\epsilon r/R}(Ax) = B_{\epsilon r/R}(y),$$

and this is exactly what we originally wanted to show.

Since A is surjective, we have that

$$Y = \bigcup_{n\in\mathbb{N}} A(B_n(0)) = \bigcup_{n\in N} \overline{A(B_n(0))}.$$

By Baire's Theorem, one of the closed sets in the second union has to contain an open ball, say $B_r(v) \subset \overline{A(B_n(0))}$. In other words, $B_r(0) \subset \overline{A(B_n(0))} - v$. Now again $v = Au$ for some $u \in X$, so

$$B_r(0) \subset \overline{A(B_n(0))} - Au = \overline{A(B_n(-u))}, \tag{3.1}$$

and if we take $N \geq n + \|u\|$, then $B_N(0) \supset B_n(-u)$, so

$$B_r(0) \subset \overline{A(B_N(0))}. \tag{3.2}$$

Except for the closure, this is what we wanted to show.

Exercise 3.2. In (3.1), we used the following fact: If $M \subset X$ and $x \in X$, then $\overline{M} + x = \overline{M + x}$. Prove this and also the analogous property that $\overline{cM} = c\overline{M}$ $(c \in \mathbb{C})$.

We will now finish the proof by showing that $\overline{A(B_N(0))} \subset A(B_{2N}(0))$. So let $y \in \overline{A(B_N)}$ (since all balls will be centered at 0, we will use this simplified notation).

We can find an $x_1 \in B_N$ with $\|y - Ax_1\| < r/2$. Since, by (3.2) and Exercise 3.2,

$$B_{r/2} = \frac{1}{2} B_r \subset \frac{1}{2} \overline{A(B_N)} = \overline{A(B_{N/2})},$$

we then also have that $y - Ax_1 \in \overline{A(B_{N/2})}$. Thus there exists an $x_2 \in B_{N/2}$ with $\|y - Ax_1 - Ax_2\| < 2^{-2} r$. We continue in this way and obtain a sequence x_n with the following properties:

$$x_n \in B_{2^{-n+1}N}, \qquad \left\| y - \sum_{j=1}^{n} Ax_j \right\| < 2^{-n} r \tag{3.3}$$

This shows, first of all, that the series $\sum_{n=1}^{\infty} x_n$ is absolutely convergent. Indeed, $\sum_{n=1}^{\infty} \|x_n\| < 2N \sum_{n=1}^{\infty} 2^{-n} = 2N < \infty$. By Exercise 2.22, $x := \sum_{n=1}^{\infty} x_n$ exists. Moreover, by the calculation just carried out, $\|x\| \le \sum \|x_n\| < 2N$, so $x \in B_{2N}$. Since A is continuous, we obtain that $Ax = \lim_{n\to\infty} \sum_{j=1}^{n} Ax_j$, and the second property from (3.3) now shows that $Ax = y$. In other words, $y \in A(B_{2N})$, as desired. □

The *graph* of an operator $A : X \to Y$ is defined as the set $\mathcal{G}(A) = \{(x, Ax) : x \in X\}$. We can think of $\mathcal{G}(A)$ as a subset of the Banach space $X \oplus Y$ that was introduced in Chapter 2; see especially Theorem 2.17.

Exercise 3.3. Show that $\mathcal{G}(A)$ is a (linear) subspace of $X \oplus Y$ if A is a linear operator.

Definition 3.4. Let X, Y be Banach spaces. A linear operator $A : X \to Y$ is called *closed* if $\mathcal{G}(A)$ is closed in $X \oplus Y$.

If we recall how the norm on $X \oplus Y$ was defined, we see that $(x_n, y_n) \to (x, y)$ in $X \oplus Y$ precisely if $x_n \to x$ and $y_n \to y$. Therefore, using sequences, we can characterize closed operators as follows: $A : X \to Y$ is closed precisely if the following holds: If $x_n \to x$ and $Ax_n \to y$, then $y = Ax$.

On the other hand, A is continuous precisely if $x_n \to x$ implies that $Ax_n \to y$ and $y = Ax$ (formulated in a slightly roundabout way here to facilitate the comparison). This looks clearly stronger than the condition from above: what was part of the hypothesis has become part of the conclusion. In particular, continuous operators are always closed. When viewed against this background, the following result is quite stunning.

Theorem 3.5 (The closed graph theorem)**.** *Let X, Y be Banach spaces and assume that $A : X \to Y$ is linear and closed. Then $A \in B(X, Y)$.*

Proof. We introduce the projections $P_1 : X \oplus Y \to X$, $P_2 : X \oplus Y \to Y$, $P_1(x, y) = x$, $P_2(x, y) = y$. It is clear that P_1, P_2 are linear and continuous. By hypothesis and Exercise 3.3, $\mathcal{G}(A)$ is a closed linear subspace of $X \oplus Y$. By Proposition 2.7, it is therefore a Banach space itself (with the same norm as $X \oplus Y$). Now P_1, restricted to $\mathcal{G}(A)$ is a bijection onto X. Corollary 3.3 shows that the inverse $P_1^{-1} : X \to \mathcal{G}(A)$, $P_1^{-1}x = (x, Ax)$ is continuous. It follows that $A = P_2 P_1^{-1}$ is a composition of continuous maps and thus continuous itself. $\square$

Exercise 3.4. Let X, Y be Banach spaces and $A_n, A \in B(X, Y)$. We say that A_n converges (to A) *strongly* if $A_n x \to Ax$ for all $x \in X$. In this case, we write $A_n \xrightarrow{s} A$. Prove that this has the following properties:
(a) $\|A_n - A\| \to 0 \Longrightarrow A_n \xrightarrow{s} A$;
(b) The converse does not hold;
(c) If $A_n \xrightarrow{s} A$, then $\sup_n \|A_n\| < \infty$ (*Hint:* use the uniform boundedness principle).

Exercise 3.5. Suppose that for some measure space (X, μ) and exponents p, q, we have that $L^p(X, \mu) \subset L^q(X, \mu)$. Show that then there exists a constant $C > 0$ so that $\|f\|_q \le C\|f\|_p$ for all $f \in L^p(X, \mu)$.

Suggested strategy: If $L^p \subset L^q$, we can define the inclusion map $I : L^p \to L^q$, $If = f$. Use Corollary 2.9 to show that this map is closed, and then apply the closed graph theorem.

4. Dual spaces and weak topologies

Recall that if X is a Banach space, we write X^* for its dual. This was defined as the space of all continuous (or bounded) linear functionals $F : X \to \mathbb{C}$. We know from the special case $Y = \mathbb{C}$ of Theorem 2.12 that X^* itself is a Banach space, too, if we use the operator norm

$$\|F\| = \sup_{\|x\|=1} |F(x)| \qquad (F \in X^*).$$

The following fundamental result makes sure that there is a large supply of bounded linear functionals on every normed space.

Theorem 4.1 (Hahn-Banach)**.** *Let X be a normed space and let M be a subspace of X. Suppose that $F : M \to \mathbb{C}$ is a linear map satisfying $|F(x)| \le C\|x\|$ ($x \in M$). Then there exists a linear extension $G : X \to \mathbb{C}$ of F so that $|G(x)| \le C\|x\|$ for all $x \in X$.*

In other words, a bounded linear functional on a subspace can always be extended to the whole space *without increasing the norm.* This latter property is the point here; it is easy, at least in principle, to linearly extend a given functional. (Sketch: Fix a basis of M as a vector space, extend to a basis of the whole space and assign arbitrary values on these new basis vectors.)

Proof. We first prove a real version of theorem. So, for the time being, let X be a *real* vector space, and assume that $F : M \to \mathbb{R}$ is a bounded linear functional on a subspace.

Roughly speaking, the extension will be done one step at a time. So our first goal is to show that F can be extended to a one-dimensional extension of M in such a way that the operator norm is preserved. We are assuming that $|F(x)| \le C\|x\|$; in fact, since $C\|x\|$ defines a new norm on X (if $C > 0$), we can assume that $C = 1$ here.

Now let $x_1 \in X$, $x_1 \notin M$. We want to define a linear extension F_1 of F on the (slightly, by one dimension) bigger space

$$M_1 = \{x + cx_1 : x \in M, c \in \mathbb{R}\}.$$

Such a linear extension is completely determined by the value $f = F_1(x_1)$ (and, conversely, every $f \in \mathbb{R}$ will define an extension). Since we also want an extension that still satisfies $|F_1(y)| \le \|y\|$ ($y \in M_1$), we're looking for an $f \in \mathbb{R}$ so that

$$-\|x + cx_1\| \le F(x) + cf \le \|x + cx_1\| \tag{4.1}$$

for all $c \in \mathbb{R}$, $x \in M$. By assumption, we already have this for $c = 0$, and by discussing the cases $c > 0$ and $c < 0$ separately, we see that

(4.1) is equivalent to

$$-\left\|\frac{x}{c}+x_1\right\| - F\left(\frac{x}{c}\right) \le f \le \left\|\frac{x}{c}+x_1\right\| - F\left(\frac{x}{c}\right)$$

for all $c \neq 0$, $x \in M$. In other words, there will be an extension F_1 with the desired properties if (and only if, but of course that is not our concern here)

$$\|z + x_1\| - F(z) \geq -\|y + x_1\| - F(y)$$

for arbitrary $y, z \in M$. This is indeed the case, because

$$F(z) - F(y) = F(z-y) \le \|z - y\| \le \|z + x_1\| + \|x_1 + y\|.$$

We now use Zorn's Lemma to obtain a norm preserving extension to all of X (this part of the proof can be safely skipped if you're not familiar with this type of argument). We consider the set of all linear extensions G of F that satisfy $|G(x)| \le \|x\|$ on the subspace on which they are defined. This set can be partially ordered by declaring $G \prec G'$ if G' is an extension of G. Now if $\{G_\alpha\}$ is a totally ordered subset (any two G_α's can be compared) and if we denote the domain of G_α by M_α, then $G : \bigcup M_\alpha \to \mathbb{R}$, $G(x) = G_\alpha(x)$ defines an extension of all the G_α's, that is, $G \succ G_\alpha$ for all α. Note that there are no consistency problems in the definition of G because if there is more than one possible choice for α for any given x, then the corresponding G_α's must give the same value on x because one of them is an extension of the other.

We have verified the hypotheses of Zorn's Lemma. The conclusion is that there is a G that is maximal in the sense that if $H \succ G$, then $H = G$. This G must be defined on the whole space X because otherwise the procedure described above would give an extension H to a strictly bigger space. We have proved the real version of the Hahn-Banach Theorem.

The original, complex version can be derived from this by some elementary, but ingenious trickery, as follows: First of all, we can think of X and M as real vector spaces also (we just refuse to multiply by non-real numbers and otherwise keep the algebraic structure intact). Moreover, $L_0(x) = \text{Re } F(x)$ defines an $\mathbb{R}$-linear functional $L_0 : M \to \mathbb{R}$. By the real version of the theorem, there exists an $\mathbb{R}$-linear extension $L : X \to \mathbb{R}$, $|L(x)| \le \|x\|$.

I now claim that the following definition will work: $G(x) = L(x) - iL(ix)$ Indeed, it is easy to check that $G(x+y) = G(x) + G(y)$, and if

$c = a + ib \in \mathbb{C}$, then

$$\begin{aligned}
G(cx) &= L(ax + ibx) - iL(iax - bx) \\
&= aL(x) + bL(ix) - iaL(ix) + ibL(x) \\
&= (a + ib)L(x) + (b - ia)L(ix) = c(L(x) - iL(ix)) = cG(x).
\end{aligned}$$

So G is $\mathbb{C}$-linear. It is also an extension of F because if $x \in M$, then $L(x) = L_0(x) = \text{Re } F(x)$ and thus

$$\begin{aligned}
G(x) &= \text{Re } F(x) - i \text{ Re } F(ix) = \text{Re } F(x) - i \text{ Re}(iF(x)) \\
&= \text{Re } F(x) + i \text{ Im } F(x) = F(x).
\end{aligned}$$

Finally, if we write $G(x) = |G(x)|e^{i\varphi(x)}$, we see that

$$\begin{aligned}
|G(x)| &= G(x)e^{-i\varphi(x)} = G(e^{-i\varphi(x)}x) = \text{Re } G(e^{-i\varphi(x)}x) \\
&= L(e^{-i\varphi(x)}x) \leq \|e^{-i\varphi(x)}x\| = \|x\|.
\end{aligned}$$

□

Here are some immediate important consequences of the Hahn-Banach Theorem. They confirm that much can be learned about a Banach spaces by studying its dual. For example, part (b) says that norms can be computed by testing functionals on the given vector x.

Corollary 4.2. *Let X, Y be normed spaces.*
(a) X^* separates the points X, that is, if $x, y \in X$, $x \neq y$, then there exists an $F \in X^*$ with $F(x) \neq F(y)$.
(b) For all $x \in X$, we have that

$$\|x\| = \sup\{|F(x)| : F \in X^*, \|F\| = 1\}.$$

(c) If $T \in B(X, Y)$, then

$$\|T\| = \sup\{|F(Tx)| : x \in X, F \in Y^*, \|F\| = \|x\| = 1\}.$$

Proof of (b). If $F \in X^*$, $\|F\| = 1$, then $|F(x)| \leq \|x\|$. This implies that $\sup |F(x)| \leq \|x\|$. On the other hand, $F_0(cx) = c\|x\|$ defines a linear functional on the one-dimensional subspace $L(x)$ that satisfies $|F_0(y)| \leq \|y\|$ for all $y = cx \in L(x)$ (in fact, we have equality here). By the Hahn-Banach Theorem, there exists an extension $F \in X^*$, $\|F\| = 1$ of F_0; by construction, $|F(x)| = \|x\|$, so $\sup |F(x)| \geq \|x\|$ and the proof is complete. In fact, this argument has also shown that the supremum is attained; it is a maximum. □

Exercise 4.1. Prove parts (a) and (c) of Corollary 4.2.

Let X be a Banach space. Since X^* is a Banach space, too, we can form *its* dual $X^{**} = (X^*)^*$. We call X^{**} the *bidual* or *second dual* of X. We can identify the original space X with a closed subspace of X^{**} in a natural way, as follows: Define a map $j : X \to X^{**}$, $j(x)(F) = F(x)$ $(x \in X,\ F \in X^*)$. In other words, vectors $x \in X$ act in a natural way on functionals $F \in X^*$: we just evaluate F on x.

Proposition 4.3. *We have that $j(x) \in X^{**}$, and the map j is a (linear) isometry. In particular, $j(X) \subset X^{**}$ is a closed subspace of X^{**}.*

An operator $I : X \to Y$ is called an *isometry* if $\|Ix\| = \|x\|$ for all $x \in X$.

Exercise 4.2. (a) Show that an isometry I is always injective, that is, $N(I) = \{0\}$.
(b) Show that $S : \ell^1 \to \ell^1$, $Sx = (0, x_1, x_2, \ldots)$ is an isometry that is not onto, that is $R(S) \neq \ell^1$.

Proof. We will only check that j is an isometry and that $j(X)$ is a closed subspace.

Exercise 4.3. Prove the remaining statements from Proposition 4.3. More specifically, prove that $j(x)$ is a linear, bounded functional on X^* for every $x \in X$, and prove that the map $x \mapsto j(x)$ is itself linear.

By the definition of the operator norm and Corollary 4.2(b), we have that

$$\begin{aligned} \|j(x)\| &= \sup\{|j(x)(F)| : F \in X^*, \|F\| = 1\} \\ &= \sup\{|F(x)| : F \in X^*, \|F\| = 1\} = \|x\|, \end{aligned}$$

so j indeed is an isometry. Clearly, $j(X)$ is a subspace (being the image of a linear map). If $y_n \in j(X)$, that is, $y_n = j(x_n)$, and $y_n \to y$, then also $x_n \to x$ for some $x \in X$, because y_n is a Cauchy sequence, and since j preserves norms, so is x_n. Since j is continuous, it follows that $j(x) = \lim j(x_n) = y$, so $y \in j(X)$. $\square$

A linear isometry preserves all structures on a Banach space (the vector space structure and the norm), and thus provides an identification of its domain with its image. Using j and Proposition 4.3, we can therefore think of X as a closed subspace of X^{**}. If, in this sense, $X = X^{**}$, we call X *reflexive.* This really means that $j(X) = X^{**}$. In particular, note that for X to be reflexive, it is not enough to have X isometrically isomorphic to X^{**}; rather, we need this isometric isomorphism to be specifically j.

We now use dual spaces to introduce new topologies on Banach spaces. If $\mathcal{T}_1$, $\mathcal{T}_2$ are two topologies on a common space X, we say that $\mathcal{T}_1$ is *weaker* than $\mathcal{T}_2$ (or $\mathcal{T}_2$ is *stronger* than T_1) if $\mathcal{T}_1 \subset T_2$. In topology, *coarse* and *fine* mean the same thing as weak and strong, respectively, but it would be uncommon to use these alternative terms in functional analysis.

Given a set X and a topological space $(Y, \mathcal{T})$ and a family $\mathcal{F}$ of maps $F : X \to Y$, there exists a weakest topology on X that makes all $F \in \mathcal{F}$ continuous. Let us try to give a description of this topology (and, in fact, we also need to show that such a topology exists). We will denote it by $\mathcal{T}_w$.

Clearly, we must have $F^{-1}(U) \in \mathcal{T}_w$ for all $F \in \mathcal{F}$, $U \in \mathcal{T}$. Conversely, any topology that contains these sets will make the F's continuous. So we could stop here and say that $\mathcal{T}_w$ is the topology generated by these sets. (Given any collection of sets, there always is a weakest topology containing these sets.) However, we would like to be somewhat more explicit. It is clear that finite intersections of sets of this type have to be in $\mathcal{T}_w$, too; in other words,

$$\{x \in X : F_1(x) \in U_1, \ldots, F_n(x) \in U_n\} \tag{4.2}$$

belongs to $\mathcal{T}_w$ for arbitrary choices of $n \in \mathbb{N}$, $F_j \in \mathcal{F}$, $U_j \in \mathcal{T}$. If these sets are open, then arbitrary unions of such sets need to belong to $\mathcal{T}_w$, and, fortunately, the process stops here: we don't get additional sets if we now take finite intersections again. So the claim is that

$$\mathcal{T}_w = \{ \text{ arbitrary unions of sets of the type (3.3) } \}. \tag{4.3}$$

We must show that $\mathcal{T}_w$ is a topology; by its construction, any other topology that makes all $F \in \mathcal{F}$ continuous must then be stronger than $\mathcal{T}_w$. This verification is quite straightforward, but a little tedious to write down, so I'll make this an exercise:

Exercise 4.4. Prove that (4.2), (4.3) define a topology.

We now apply this process to a Banach space X, with $Y = \mathbb{C}$ and $\mathcal{F} = X^*$. Of course, we already have a topology on X (the norm topology); this new topology will be different, unless X is finite-dimensional. Here's the formal definition:

Definition 4.4. Let X be a Banach space. The *weak topology* on X is defined as the weak topology $\mathcal{T}_w$ generated by X^*.

If we denote the norm topology by $\mathcal{T}$, then, since all $F \in X^*$ are continuous if we use $\mathcal{T}$ (by definition of X^*!), we see that $\mathcal{T}_w \subset \mathcal{T}$; in other words, the weak topology is weaker than the norm topology. By the discussion above, (4.3) gives a description of $\mathcal{T}_w$. A slightly more convenient variant of this can be obtained by making use of the vector space structure. First of all, the sets

$$U(F_1, \ldots, F_n; \epsilon_1, \ldots, \epsilon_n) = \{x \in X : |F_j(x)| < \epsilon_j \ (j = 1, \ldots, n)\} \tag{4.4}$$

are in $\mathcal{T}_w$ for arbitrary $n \in \mathbb{N}$, $F_j \in X^*$, $\epsilon_j > 0$. In fact, they are of the form (4.2), with $U_j = \{z : |z| < \epsilon_j\}$. I now claim that $V \in \mathcal{T}_w$ if and only if for every $x \in V$, there exists a set $U = U(F_j; \epsilon_j)$ of this form with $x + U \subset V$.

Exercise 4.5. Prove this claim.

We can rephrase this as follows: The U's form a neighborhood base at 0 (that is, any neighborhood of $x = 0$ contains some U) and the neighborhoods of an arbitrary $x \in X$ are precisely the translates $x+W$ of the neighborhoods W of 0.

We'll make two more observations on the weak topology and then leave the matter at that. First of all, $\mathcal{T}_w$ is a Hausdorff topology: If $x, y \in X$, $x \neq y$, then there exist $V, W \in \mathcal{T}_w$ with $x \in V$, $y \in W$, $V \cap W = \emptyset$. To prove this, we use the fact that X^* separates the points of X; see Corollary 4.2(a). So there is an $F \in X^*$ with $F(x) \neq F(y)$. We can now take $V = x + U(F; \epsilon)$, $W = y + U(F; \epsilon)$ with a sufficiently small $\epsilon > 0$.

Exercise 4.6. Provide the details for this last step. You can (and should) make use of the description of $\mathcal{T}_w$ established above, in Exercise 4.5.

Finally, if x_n is a sequence from X, then $x_n \to x$ in $\mathcal{T}_w$ (this is usually written as $x_n \xrightarrow{w} x$; x_n goes to x weakly) if and only if $F(x_n) \to F(x)$ for all $F \in X^*$.

Exercise 4.7. Prove this. Again, by the results of Exercise 4.5, $x_n \xrightarrow{w} x$ precisely if for every U of the form (4.4), we eventually have that $x_n - x \in U$.

This gives a characterization of convergent sequences and thus some idea of what the topology does. However, it can happen that $\mathcal{T}_w$ is not metrizable and then the topological notions (closed sets, compactness, continuity etc.) can *not* be characterized using sequences.

Definition 4.5. Let X be a Banach space. The *weak-$*$ topology* $\mathcal{T}_{w^*}$ on X^* is defined as the weak topology generated by X, viewed as a subset of X^{**}. Put differently, $\mathcal{T}_{w^*}$ is the weakest topology that turns all point evaluations $j(x) : X^* \to \mathbb{C}$, $F \mapsto F(x)$ $(x \in X)$ into continuous functions on X^*.

We have an analogous description of $\mathcal{T}_{w^*}$. The sets

$$U(x_1, \ldots, x_n; \epsilon_1, \ldots, \epsilon_n) = \{F \in X^* : |F(x_j)| < \epsilon_j\}$$

are open, and $V \subset X^*$ is open in the weak-$*$ topology if and only for every $F \in V$, there exists such a U so that $F + U \subset V$.

Exercise 4.8. Prove that $\mathcal{T}_{w^*}$ is a Hausdorff topology. *Hint:* If $F \neq G$, then $F(x) \neq G(x)$ for some $x \in X$. Now you can build disjoint neighborhoods of F, G as above, using this x; see also Exercise 4.6.

Exercise 4.9. Let X be a Banach space, and let $F_n, F \in X^*$. Show that $F_n \to F$ in the weak-$*$ topology if and only if $F_n(x) \to F(x)$ for all $x \in X$.

Since X^* is a Banach space, we can also define a weak topology on X^*. This is the topology generated by X^{**}. The weak-$*$ topology is generated by X, which in general is a smaller set of maps, so the weak-$*$ topology is weaker than the weak topology. It can only be defined on a dual space. If X is reflexive, then there's no difference between the weak and weak-$*$ topologies.

Despite its clumsy and artificial looking definition, the weak-$*$ topology is actually an extremely useful tool. All the credit for this is due to the following fundamental result.

Theorem 4.6 (Banach-Alaoglu)**.** *Let X be a Banach space. Then the closed unit ball $\overline{B}_1(0) = \{F \in X^* : \|F\| \leq 1\}$ is compact in the weak-$*$ topology.*

Proof. This will follow from *Tychonoff's Theorem:* The product of compact topological spaces is compact in the product topology. To get this argument started, we look at the Cartesian product set

$$K = \prod_{x \in X} \{z \in \mathbb{C} : |z| \leq \|x\|\}.$$

As a set, this is defined as the set of maps $F : X \to \mathbb{C}$ with $|F(x)| \leq \|x\|$. The individual factors $\{|z| \leq \|x\|\}$ come with a natural topology, and we endow K with the product topology, which, by definition, is the weak topology generated by the projections $p_x : K \to \{|z| \leq \|x\|\}$, $p_x(F) = F(x)$ (equivalently, you can also produce it from cylinder sets, if this is more familiar to you). By Tychonoff's Theorem, K is compact.

Now $\overline{B}_1(0) \subset K$; more precisely, $\overline{B}_1(0)$ consists of those maps $F \in K$ that are also linear. I now claim that the topology induced by K on $\overline{B}_1(0)$ is the same as the induced topology coming from the weak-$*$ topology on $X^* \supset \overline{B}_1(0)$. This should not come as a surprise because both the product topology and $\mathcal{T}_{w^*}$ are weak topologies generated by the point evaluations $x \mapsto F(x)$. Writing it down is a slightly unpleasant task that is best delegated to the reader.

Exercise 4.10. Show that K and $(X^*, \mathcal{T}_{w^*})$ indeed induce the same topology on $\overline{B}_1(0)$. Come to think of it, we perhaps really want to prove the following abstract fact: Let $(Z, \mathcal{T})$ be a topological space, $\mathcal{F}$ a family of maps $F : X \to Z$ and let $Y \subset X$. Then we can form the weak topology $\mathcal{T}_w$ on X; this induces a relative topology on Y. Alternatively, we can restrict the maps $F \in \mathcal{F}$ to Y and let the restrictions generate a weak topology on Y. Prove that both methods lead to the same topology. As usual, this is mainly a matter of unwrapping definitions. You could use the description (4.2), (4.3) of the weak topologies and look at what happens when these induce relative topologies.

Exercise 4.11. (a) Let Y be a compact topological space and let $A \subset Y$ be closed. Prove that then A is compact, too.
(b) Let Y be a topological space. Show that a subset $B \subset Y$ is compact if and only if B with the relative topology is a compact topological space. (Sometimes compactness is defined in this way; recall that we defined compact sets by using covers by open sets $U \subset Y$. It is now in fact almost immediate that we get the same condition from both variants, but this fact will be needed here, so I thought I'd point it out.)

With these preparations out of the way, it now suffices to show that $\overline{B}_1(0)$ is closed in K. So let $F \in K \setminus \overline{B}_1(0)$. We want to find a neighborhood of F that is contained in $K \setminus \overline{B}_1(0)$ (note that we cannot use sequences here because there is no guarantee that our topologies are metrizable). Since $F \notin \overline{B}_1(0)$, F is not linear and thus there are $c, d \in \mathbb{C}$, $x, y \in X$ so that

$$\epsilon \equiv |F(cx + dy) - cF(x) - dF(y)| > 0.$$

But then

$$V = \Big\{ G \in K : |G(cx + dy) - F(cx + dy)| < \frac{\epsilon}{3}, \\ |c|\,|G(x) - F(x)| < \frac{\epsilon}{3}, |d|\,|G(y) - F(y)| < \frac{\epsilon}{3} \Big\}$$

is an open set in K with $F \in V$, and if $G \in V$, then still

$$|G(cx + dy) - cG(x) - dG(y)| > 0,$$

so V does not contain any linear maps and thus is the neighborhood of F we wanted to find. $\square$

We have already seen how the fact that $\mathcal{T}_{w^*}$ need not be metrizable makes this topology a bit awkward to deal with. The following result provides some relief. We call a metric space X *separable* if X has a countable dense subset (that is, there exist $x_n \in X$ so that if $x \in X$ and $\epsilon > 0$ are arbitrary, then $d(x, x_n) < \epsilon$ for some $n \in \mathbb{N}$).

Exercise 4.12. (a) Show that ℓ^p (with index set $\mathbb{N}$, as usual) is separable for $1 \le p < \infty$. You can use the result of Exercise 4.13 below if you want.
(b) Show that ℓ^∞ is not separable. *Suggestion:* Consider all $x \in \ell^\infty$ that only take the values 0 and 1. How big is this set? What can you say about $\|x - x'\|$ for two such sequences?

Exercise 4.13. Let X be a Banach space. Show that X will be separable if there is a countable *total* subset, that is, if there is a countable set $M \subset X$ so that the (finite) linear combinations of elements from M are dense in X (in other words, if $x \in X$ and $\epsilon > 0$, we must be able to find $m_j \in M$ and coefficients $c_j \in \mathbb{C}$ so that $\left\|\sum_{j=1}^N c_j m_j - x\right\| < \epsilon$.)

Theorem 4.7. *If X is a separable Banach space, then the weak-$*$ topology on $\overline{B}_1(0) \subset X^*$ (more precisely: the relative topology induced by $\mathcal{T}_{w^*}$) is metrizable.*

We don't want to prove this in detail, but the basic idea is quite easy to state. The formula

$$d(F,G) = \sum_{n=1}^{\infty} 2^{-n} \frac{|F(x_n) - G(x_n)|}{1 + |F(x_n) - G(x_n)|}$$

(say), where $\{x_n\}$ is a dense subset of X, defines a metric that generates the desired topology.

Corollary 4.8. *Let X be a separable Banach space. If $F_n \in X^*$, $\|F_n\| \le 1$, then there exist $F \in X^*$, $\|F\| \le 1$ and a subsequence $n_j \to \infty$ so that $F_{n_j}(x) \to F(x)$ for all $x \in X$.*

Proof. This follows by just putting things together. By the Banach-Alaoglu Theorem, $\overline{B}_1(0)$ is compact in the weak-$*$ topology. By Theorem 4.7, this can be thought of as a metric space. By Theorem 1.7(c), compactness is therefore equivalent to sequences having convergent subsequences. By using Exercise 4.9, we now obtain the claim. □

To make good use of the results of this chapter, we need to know what the dual space of a given space is. We now investigate this question for the Banach spaces from our list that was compiled in Chapter 2.

Example 4.1. If $X = \mathbb{C}^n$ with some norm, then Corollary 2.16 implies that all linear functionals on X are bounded, so in this case X^* coincides with the algebraic dual space. From linear algebra we know that as a vector space, X^* can be identified with $\mathbb{C}^n$ again; more precisely, $y \in \mathbb{C}^n$ can be identified with the functional $x \mapsto \sum_{j=1}^n y_j x_j$. It also follows from this that X is reflexive. The norm on X^* depends on the norm on X; Example 4.3 below will throw some additional light on this.

Exercise 4.14. Show that the weak topology on $X = \mathbb{C}^n$ coincides with the norm topology. *Suggestion:* It essentially suffices to check that open balls $B_r(0)$ (say) are in $\mathcal{T}_w$. Show this and then use the definition of the norm topology $\mathcal{T}$ to show that $\mathcal{T} \subset \mathcal{T}_w$. Since always $\mathcal{T}_w \subset \mathcal{T}$, this will finish the proof.

This Exercise says that we really don't get anything new from the theory of this chapter on finite-dimensional spaces; recall also in this context that $\mathcal{T}_w = \mathcal{T}_{w^*}$ on $\mathbb{C}^n$, thought of as the dual space X^* of $X = \mathbb{C}^n$, because X is reflexive.

Example 4.2. Let K be a compact Hausdorff space and consider the Banach space $C(K)$. Then $C(K)^* = \mathcal{M}(K)$, where $\mathcal{M}(K)$ is defined as the space of all complex, regular Borel measures on K. Here, we call a (complex) measure μ (inner and outer) *regular* if its total variation $\nu = |\mu|$ is regular in the sense that

$$\nu(B) = \sup_{L \subset B: L \text{ compact}} \nu(L) = \inf_{U \supset B: U \text{ open}} \nu(U)$$

for all Borel sets $B \subset K$. $\mathcal{M}(K)$ becomes a vector space if the vector space operations are introduced in the obvious way as $(\mu+\nu)(B) = \mu(B) + \nu(B)$, $(c\mu)(B) = c\mu(B)$. In fact, $\mathcal{M}(K)$, equipped with the norm

$$\|\mu\| = |\mu|(K), \tag{4.5}$$

is a Banach space. This is perhaps most elegantly deduced from the main assertion of this Example, namely the fact that $C(K)^*$ can be identified with $\mathcal{M}(K)$, and, as we will see in a moment, the operator norm on $\mathcal{M}(K) = C(K)^*$ turns out to be exactly (4.5). More precisely, the claim is that every $\mu \in \mathcal{M}(K)$ generates a functional $F_\mu \in C(K)^*$ via

$$F_\mu(f) = \int_K f(x)\, d\mu(x), \tag{4.6}$$

and we also claim that the corresponding map $\mathcal{M}(K) \to C(K)^*$, $\mu \mapsto F_\mu$ is an isomorphism between Banach spaces (in other words, a bijective, linear isometry).

The *Riesz Representation Theorem* does the lion's share of the work here; it implies that $\mu \mapsto F_\mu$ is a bijection from $\mathcal{M}(K)$ onto $C(K)^*$; see for example, Folland, Real Analysis, Corollary 7.18. It is also clear that this map is linear.

Exercise 4.15. Suppose we introduce a norm on $\mathcal{M}(K)$ by just declaring $\|\mu\| = \|F_\mu\|$ (operator norm), that is, we just move the norm on $C(K)^*$ over to $\mathcal{M}(K)$. Show that this leads to (4.5); put differently, show that the operator norm of F_μ from (4.6) equals $|\mu|(K)$.

This identification of the dual of $C(K)$ (which is basically a way of stating (one version of) the Riesz Representation Theorem) is an extremely important fundamental result; we have every right to be very excited about this.

One pleasing consequence is the fact that $\mathcal{M}(K)$, being a dual space, can be equipped with a weak-$*$ topology. This, in turn, has implications of the following type:

Exercise 4.16. Show that $C[a,b]$ (so $K = [a,b]$, with the usual topology) is separable. *Suggestion:* Deduce this from the Weierstraß approximation theorem.

Exercise 4.17. Let μ_n be a sequence of complex Borel measures on $[a,b]$ with $|\mu_n|([a,b]) \le 1$ (in particular, these could be arbitrary positive measures with $\mu_n([a,b]) \le 1$). Show that there exists another Borel measure μ on $[a,b]$ with $|\mu|([a,b]) \le 1$ and a subsequence n_j so that

$$\lim_{j\to\infty} \int_{[a,b]} f(x)\, d\mu_{n_j}(x) = \int_{[a,b]} f(x)\, d\mu(x)$$

for all $f \in C[a,b]$.

Hint: This in fact follows quickly from Corollary 4.8 and Exercise 4.16.

Example 4.3. We now move on to ℓ^p spaces. We first claim that if $1 \le p < \infty$, then $(\ell^p)^* = \ell^q$, where $1/p + 1/q = 1$. More precisely, the claim really is that every $y \in \ell^q$ generates a functional $F_y \in (\ell^p)^*$, as follows:

$$F_y(x) = \sum_{j=1}^{\infty} y_j x_j \tag{4.7}$$

Moreover, the corresponding map $y \mapsto F_y$ is an isomorphism between the Banach spaces ℓ^q and $(\ell^p)^*$.

Let us now prove these assertions. First of all, Hölder's inequality shows that the series from (4.7) converges, and in fact $|F_y(x)| \le \|y\|_q \|x\|_p$. Since (4.7) is also obviously linear in x, this shows that $F_y \in (\ell^p)^*$ and $\|F_y\| \le \|y\|_q$. We will now explicitly discuss only the case $1 < p < \infty$. If $p = 1$ (and thus $q = \infty$), the same basic strategy works and actually the technical details get easier, but some slight adjustments are necessary.

To compute $\|F_y\|$, we set

$$x_n = \begin{cases} \frac{|y_n|^q}{y_n} & n \le N, y_n \ne 0 \\ 0 & \text{else} \end{cases}, \tag{4.8}$$

with $N \in \mathbb{N}$. It is then clear that $x \in \ell^p$,

$$\|x\|_p^p = \sum_{n=1}^N |y_n|^{(q-1)p} = \sum_{n=1}^N |y_n|^q,$$

and thus

$$F_y(x) = \sum_{n=1}^N |y_n|^q = \left(\sum_{n=1}^N |y_n|^q \right)^{1/q} \|x\|_p.$$

Thus $\|F_y\| \ge \left(\sum_{n=1}^N |y_n|^q \right)^{1/q}$, and this holds for arbitrary $N \in \mathbb{N}$, so it follows that $\|F_y\| \ge \|y\|_q$, so $\|F_y\| = \|y\|_q$. This says that the identification $y \mapsto F_y$ is isometric, and it is obviously linear (in y!), so it remains to show that it is surjective, that is, every $F \in (\ell^p)^*$ equals some F_y for suitable $y \in \ell^q$. To prove this, fix $F \in (\ell^p)^*$. It is clear from (4.7) that only $y_n = F(e_n)$ can work, so define a sequence y in this way (here, $e_n(j) = 1$ if $j = n$ and $e_n(j) = 0$ otherwise). Consider again the $x \in \ell^p$ from (4.8). Then we have that

$$F(x) = F\left(\sum_{n=1}^N \frac{|y_n|^q}{y_n} e_n \right) = \sum_{n=1}^N |y_n|^q.$$

As above, $\|x\|_p = \left(\sum_{n=1}^N |y_n|^q \right)^{1/p}$, so

$$\sum_{n=1}^N |y_n|^q \le \|F\| \left(\sum_{n=1}^N |y_n|^q \right)^{1/p}$$

or $\left(\sum_{n=1}^N |y_n|^q \right)^{1/q} \le \|F\|$. Again, N is arbitrary here, so $y \in \ell^q$. By construction of y, we have that $F(z) = F_y(z)$ if z is a finite linear combination of e_n's. These vectors z, however, are dense in ℓ^p, so it follows from the continuity of both F and F_y that $F(w) = F_y(w)$ for *all* $w \in \ell^p$, that is, $F = F_y$.

As a consequence, ℓ^p is reflexive for $1 < p < \infty$, basically because $(\ell^p)^{**} = (\ell^q)^* = \ell^p$, by applying the above result on the dual of ℓ^r twice.

Exercise 4.18. Give a careful version of this argument, where the identification $j : X \to X^{**}$ from the definition of reflexivity is taken seriously.

We can't be sure about ℓ^1 and ℓ^∞ at this point because we don't know yet what $(\ell^\infty)^*$ is. It will turn out that these spaces are not reflexive.

Example 4.4. Similar discussions let us identify the duals of c_0 and c. We claim that $c_0^* = \ell^1 = \ell^1(\mathbb{N})$ and $c^* = \ell^1(\mathbb{N}_0)$; as usual, we really mean that there are Banach space isomorphisms (linear, bijective isometries) $y \mapsto F_y$ that provide identifications between these spaces, and in the case at hand, these are given by:

$$F_y(x) = \sum_{j=1}^\infty y_j x_j \quad (y \in \ell^1(\mathbb{N}), x \in c_0),$$

$$F_y(x) = y_0 \cdot \left(\lim_{n\to\infty} x_n\right) + \sum_{j=1}^\infty y_j \left(x_j - \lim_{n\to\infty} x_n\right) \quad (y \in \ell^1(\mathbb{N}_0), x \in c)$$

Since this discussion is reasonably close to Example 4.3, I don't want to do it here. The above representations of the dual spaces as $\ell^1(\mathbb{N})$ and $\ell^1(\mathbb{N}_0)$ seem natural, but note that these are of course isometrically isomorphic as Banach spaces: $(y_n)_{n\ge 1} \mapsto (y_{n+1})_{n\ge 0}$ is an isometry from $\ell^1(\mathbb{N})$ onto $\ell^1(\mathbb{N}_0)$. Roughly speaking, the one additional dimension of $\ell^1(\mathbb{N}_0)$ doesn't alter the Banach space structure. (At the same time, $\ell^1(\mathbb{N})$ can of course also be identified with the codimension 1 subspace $\{(x_n)_{n\ge 0} : x_0 = 0\}$ of $\ell^1(\mathbb{N}_0)$; there is nothing unusual about this in infinite-dimensional situations: the whole space can be isomorphic to a proper subspace.)

Example 4.5. We now discuss $(\ell^\infty)^*$. We will obtain an explicit looking description of this dual space, too, but, actually, this result will not be very useful. This is so because the objects that we will obtain are not particularly well-behaved and there is no well developed machinery that would recommend them for further use.

It will turn out that $(\ell^\infty)^* = \mathcal{M}_{fa}(\mathbb{N})$, the space of bounded, finitely additive set functions on $\mathbb{N}$. More precisely, the elements of $\mathcal{M}_{fa}(\mathbb{N})$ are set functions $\mu : \mathcal{P}(\mathbb{N}) \to \mathbb{C}$ that satisfy $\sup_{M\subset\mathbb{N}} |\mu(M)| < \infty$ and if $M_1, M_2 \subset \mathbb{N}$ are disjoint, then $\mu(M_1 \cup M_2) = \mu(M_1) + \mu(M_2)$. Note that the complex measures on the σ-algebra $\mathcal{P}(\mathbb{N})$ are precisely those $\mu \in \mathcal{M}_{fa}(\mathbb{N})$ that are σ-additive rather than just finitely additive.

Finitely additive bounded set functions will act on vectors $x \in \ell^\infty$ by an integration of sorts. We discuss this new integral first (as far as I can see, this integral does not play any major role in analysis except in this particular context). Let $x \in \ell^\infty$. We subdivide the disk $\{z \in \mathbb{C} : |z| \le \|x\|\}$ into squares (say) Q_j, and we fix a number $z_j \in Q_j$ from each square. Let $M_j = \{n \in \mathbb{N} : x_n \in Q_j\}$ be the inverse image under x_n. It's quite easy to check that for $\mu \in \mathcal{M}_{fa}(\mathbb{N})$, the sums $\sum z_j \mu(M_j)$ will approach a limit as the subdivision gets finer and finer, and this limit is independent of the choice of the Q_j and z_j. We call this limit the *Radon integral* of x_n with respect to μ, and we denote it by

$$R - \int_{\mathbb{N}} x_n \, d\mu(n).$$

Next, we show how to associate a set function $\mu \in \mathcal{M}_{fa}(\mathbb{N})$ with a given functional $F \in (\ell^\infty)^*$. Define $\mu(M) = F(\chi_M)$ $(M \subset \mathbb{N})$. Then $|\mu(M)| \leq \|F\| \, \|\chi_M\| \leq \|F\|$, so μ is a bounded set function. Also, if $M_1 \cap M_2 = \emptyset$, then

$$\begin{aligned} \mu(M_1 \cup M_2) &= F(\chi_{M_1 \cup M_2}) = F(\chi_{M_1} + \chi_{M_2}) \\ &= F(\chi_{M_1}) + F(\chi_{M_2}) = \mu(M_1) + \mu(M_2). \end{aligned}$$

Thus $\mu \in \mathcal{M}_{fa}(\mathbb{N})$. Moreover, if squares Q_j and points $z_j \in Q_j$ are chosen as above and if we again set $M_j = \{n \in \mathbb{N} : x_n \in Q_j\}$, then $\left\| x - \sum z_j \chi_{M_j} \right\|$ is bounded by the maximal diameter of the Q_j's, so this will go to zero if we again consider a sequence of subdivisions becoming arbitrarily fine. It follows that

$$F(x) = \lim F\left(\sum z_j \chi_{M_j} \right) = \lim \sum z_j \mu(M_j) = R - \int_{\mathbb{N}} x_n \, d\mu(n).$$

Conclusion: Every $F \in (\ell^\infty)^*$ can be represented as a Radon integral. Conversely, one can show that every $\mu \in \mathcal{M}_{fa}(\mathbb{N})$ generates a functional F_μ on ℓ^∞ by Radon integration:

$$F_\mu(x) = R - \int_{\mathbb{N}} x_n \, d\mu(n)$$

(The boundedness of F_μ requires some work; Exercise 4.19 below should help to clarify things.)

We obtain a bijection $\mathcal{M}_{fa}(\mathbb{N}) \to (\ell^\infty)^*$, $\mu \mapsto F_\mu$, and, as in the previous examples, it's now a relatively easy matter to check that this actually sets up an isometric isomorphism between Banach spaces if we endow $\mathcal{M}_{fa}(\mathbb{N})$ with the natural vectors space structure ($(\mu+\nu)(M) := \mu(M) + \nu(N)$ etc.) and the norm

$$\|\mu\| = \sup_{\|x\|=1} \left| R - \int_{\mathbb{N}} x_n \, d\mu(n) \right|.$$

I'll leave the details of these final steps to the reader.

Exercise 4.19. Show that $\|\mu\| = \sup \sum |\mu(M_j)|$, where the supremum is over all partitions of $\mathbb{N}$ into finitely many sets $M_1, \ldots, M_N$. Moreover,

$$\sum_{n=1}^{\infty} |\mu(\{n\})| \le \|\mu\|;$$

can you also show that strict inequality is possible? (Exercise 4.21 might be helpful here.)

Exercise 4.20. Show that $\ell^1 \subset \mathcal{M}_{fa}(\mathbb{N})$ in the sense that if $y \in \ell^1$, then $\mu(M) = \sum_{n \in M} y_n$ defines a bounded, finitely additive set function. Show that in fact these μ's are exactly the (complex) measures on $(\mathbb{N}, \mathcal{P}(\mathbb{N}))$.

Remark: Since X can be identified with a subspace of X^{**} for any Banach space X and since $\ell^\infty = (\ell^1)^*$, we knew right away that $\ell^1 \subset (\ell^\infty)^*$, provided this is suitably interpreted.

Exercise 4.21. The fact that $\ell^1 \subsetneqq (\ell^\infty)^*$ can also be seen more directly, without giving a description of $(\ell^\infty)^*$, as follows:
(a) Show that every $y \in \ell^1$ generates a functional $F_y \in (\ell^\infty)^*$ by letting

$$F_y(x) = \sum_{n=1}^{\infty} y_n x_n \qquad (x \in \ell^\infty). \tag{4.9}$$

(b) Show that not every $F \in (\ell^\infty)^*$ is of this form, by using the Hahn-Banach Theorem. More specifically, choose a subspace $Y \subset \ell^\infty$ and define a bounded functional F_0 on Y in such a way that no extension F of F_0 can be of the form (4.9). (This is an uncomplicated argument if done efficiently; it all depends on a smart choice of Y and F_0.)

Example 4.6. I'll quickly report on the spaces $L^p(X, \mu)$ here. The situation is similar to the discussion above; see Examples 4.3, 4.5. If $1 \le p < \infty$, then $(L^p)^* = L^q$, where $1/p + 1/q = 1$. This holds in complete generality for $1 < p < \infty$, but if $p = 1$, then we need the additional hypothesis that μ is σ-finite (which means that X can be written as a countable union of sets of finite measure). Again, this is an abbreviated way of stating the result; it really involves an identification of Banach spaces: the function $f \in L^q$ is identified with the functional $F_f \in (L^p)^*$ defined by $F_f(g) = \int_X f g \, d\mu$.

$(L^\infty)^*$ is again a complicated space that can be described as a space of finitely additive set functions, but this description is only moderately useful. In particular, except in special cases, $(L^\infty)^*$ is (much) bigger than L^1. In fact, for example $L^1(\mathbb{R}, m)$ is not the dual space of *any* Banach space: there is no Banach space X for which X^* is isometrically isomorphic to $L^1(\mathbb{R}, m)$!

Exercise 4.22. What is wrong with the following sketch of a "proof" that $(\ell^\infty)^* = \ell^1$:

Follow the strategy from Example 4.3. Obviously, if $y \in \ell^1$, then $F_y \in (\ell^\infty)^*$, if F_y is defined as in (4.7). Conversely, given an $F \in (\ell^\infty)^*$, let $y_n = F(e_n)$. Define $x \in \ell^\infty$ by

$$x_n = \begin{cases} \frac{|y_n|}{y_n} & n \le N, y_n \ne 0 \\ 0 & \text{otherwise} \end{cases} .$$

Then $\|x\|_\infty \le 1$, so $F(x) = \sum_{n=1}^N |y_n| \le \|F\|$, and it follows that $y \in \ell^1$. By construction, $F = F_y$.

Exercise 4.23. Let X be a Banach space. Show that every weakly convergent sequence is bounded: If $x_n, x \in X$, $F(x_n) \to F(x)$ for all $F \in X^*$, then $\sup \|x_n\| < \infty$.

Hint: Think of the x_n as elements of the bidual $X^{**} \supset X$ and apply the uniform boundedness principle.

Exercise 4.24. (a) Show that $e_n \xrightarrow{w} 0$ in ℓ^2.
(b) Construct a sequence f_n with similar properties in $C[0,1]$: we want that $\|f_n\| = 1$, $f_n \xrightarrow{w} 0$.

5. Hilbert spaces

Definition 5.1. Let H be a (complex) vector space. A *scalar product* (or *inner product*) is a map $\langle \cdot, \cdot \rangle : H \times H \to \mathbb{C}$ with the following properties:
(1) $\langle x, x\rangle \ge 0$ and $\langle x, x\rangle = 0 \iff x = 0$;
(2) $\langle x, y\rangle = \overline{\langle y, x\rangle}$;
(3) $\langle x, y+z\rangle = \langle x, y\rangle + \langle x, z\rangle$;
(4) $\langle x, cy\rangle = c\langle x, y\rangle$.

(3), (4) say that a scalar product is linear in the second argument, and by combining this with (2), we see that it is *antilinear* in the first argument, that is $\langle x+y, z\rangle = \langle x, z\rangle + \langle y, z\rangle$ as usual, but $\langle cx, y\rangle = \overline{c}\,\langle x, y\rangle$.

Example 5.1. It is easy to check that

$$\langle x, y\rangle = \sum_{n=1}^{\infty} \overline{x_n} y_n$$

defines a scalar product on $H = \ell^2$. Indeed, the series converges by Hölder's inequality with $p = q = 2$, and once we know that, it is clear that (1)–(4) from above hold.

In fact, this works for arbitrary index sets I: there is a similarly defined scalar product on $\ell^2(I)$. I mention this fact here because we will actually make brief use of this space later in this chapter.

Similarly,

$$\langle f, g\rangle = \int_X \overline{f(x)} g(x)\, d\mu(x)$$

defines a scalar product on $L^2(X, \mu)$.

Theorem 5.2. *Let H be a space with a scalar product. Then:*
(a) $\|x\| := \sqrt{\langle x, x\rangle}$ defines a norm on H;
(b) The Cauchy-Schwarz inequality *holds:*

$$|\langle x, y\rangle| \le \|x\|\,\|y\|;$$

(c) We have equality in (b) if and only if x, y are linearly dependent.

Proof. We first discuss parts (b) and (c). Let $x, y \in H$. Then, by property (1) from Definition 5.1, we have that

$$0 \le \langle cx + y, cx + y\rangle = |c|^2\|x\|^2 + \|y\|^2 + c\langle y, x\rangle + \overline{c}\langle x, y\rangle, \tag{5.1}$$

for arbitrary $c \in \mathbb{C}$. If $x \ne 0$, we can take $c = -\langle x, y\rangle/\|x\|^2$ here (note that (1) makes sure that $\|x\| = \sqrt{\langle x, x\rangle} > 0$, even though we don't know yet that this really is a norm). Then (5.1) says that

$$0 \leq \|y\|^2 - \frac{|\langle x, y\rangle|^2}{\|x\|^2},$$

and this implies the Cauchy-Schwarz inequality. Moreover, we can get equality in (5.1) only if $cx + y = 0$, so x, y are linearly dependent in this case. Conversely, if $y = cx$ or $x = cy$, then it is easy to check that we do get equality in (b).

We can now prove (a). Property (1) from Definition 5.1 immediately implies condition (1) from Definition 2.1. Moreover, $\|cx\| = \sqrt{\langle cx, cx\rangle} = \sqrt{|c|^2\langle x, x\rangle} = |c|\,\|x\|$, and the triangle inequality follows from the Cauchy-Schwarz inequality, as follows:

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

□

Notice that we recover the usual norms on ℓ^2 and L^2, respectively, if we use the scalar products introduced in Example 5.1 here. It now seems natural to ask if every norm is of the form $\|x\| = \sqrt{\langle x, x\rangle}$ for some scalar product $\langle \cdot, \cdot\rangle$. This question admits a neat, satisfactory answer (although it must be admitted that this result doesn't seem to have meaningful applications):

Exercise 5.1. Let H be a vector space with a scalar product, and introduce a norm $\|\cdot\|$ on H as in Theorem 5.2(a). Then $\|\cdot\|$ satisfies the *parallelogram identity:*

$$\|x + y\|^2 + \|x - y\|^2 = 2\|x\|^2 + 2\|y\|^2 \tag{5.2}$$

One can now show that (5.2) is also a sufficient condition for a norm to come from a scalar product (the *Jordan-von Neumann Theorem*). This converse is much harder to prove; we don't want to discuss it in detail here. However, I will mention how to get this proof started. The perhaps somewhat surprising fact is that the norm already completely determines its scalar product (assuming now that the norm comes from a scalar product). In fact, we can be totally explicit, as Proposition 5.3 below will show. A slightly more general version is often useful; to state this, we need an additional definition: A *sesquilinear form* is a map $s : H \times H \to \mathbb{C}$ that is linear in the second argument and antilinear in the first ("sesquilinear" = one and a half linear):

$$s(x, cy + dz) = cs(x, y) + ds(x, z)$$
$$s(cx + dy, z) = \overline{c}s(x, z) + \overline{d}s(y, z)$$

A scalar product has these properties, but this new notion is more general.

Proposition 5.3 (The polarization identity). *Let s be a sesquilinear form, and let $q(x) = s(x, x)$. Then*

$$s(x, y) = \frac{1}{4}\left[q(x + y) - q(x - y) + iq(x - iy) - iq(x + iy)\right].$$

Exercise 5.2. Prove Proposition 5.3, by a direct calculation.

This is an extremely useful tool and has many applications. The polarization identity suggest the principle "it is often enough to know what happens on the diagonal."

In the context of the Jordan-von Neumann Theorem, it implies that the scalar product can be recovered from its norm, as already mentioned above. This is in fact immediate now because if $s(x, y) = \langle x, y\rangle$, then $q(x) = \|x\|^2$, so the polarization identity gives $\langle x, y\rangle$ in terms of the norms of four other vectors.

Exercise 5.3. Use the result from Exercise 5.1 to prove that the norms $\|\cdot\|_p$ on ℓ^p are not generated by a scalar product for $p \neq 2$.

Given a scalar product on a space H, we always equip H with the norm from Theorem 5.2(a) also. In particular, all constructions and results on normed spaces can be used in this setting, and we have a topology on H. The following observation generalizes the result from Exercise 2.2(b) in this setting:

Corollary 5.4. *The scalar product is continuous: if $x_n \to x$, $y_n \to y$, then also $\langle x_n, y_n\rangle \to \langle x, y\rangle$.*

Exercise 5.4. Deduce this from the Cauchy-Schwarz inequality.

As usual, complete spaces are particularly important, so they again get a special name:

Definition 5.5. A complete space with scalar product is called a *Hilbert space.*

Or we could say a Hilbert space is a Banach space whose norm comes from a scalar product. By Example 5.1, ℓ^2 and L^2 are Hilbert spaces (we of course know that these are Banach spaces, so there's nothing new to check here). On the other hand, Exercise 5.3 says that ℓ^p cannot be given a Hilbert space structure (that leaves the norm intact) if $p \neq 2$. Hilbert spaces are very special Banach spaces. Roughly speaking, the scalar product allows us to introduce angles between vectors, and this additional structure makes things much more pleasant. There is no such notion on a general Banach space.

In particular, a scalar product leads to a natural notion of orthogonality, and this can be used to great effect. In the sequel, H will always assumed to be a Hilbert space. We say that $x, y \in H$ are *orthogonal* if $\langle x, y\rangle = 0$. In this case, we also write $x \perp y$. If $M \subset H$ is an arbitrary subset of H, we define its *orthogonal complement* by

$$M^{\perp} = \{x \in H : \langle x, m\rangle = 0 \text{ for all } m \in M\}.$$

Exercise 5.5. Prove the following formula, which is reminiscent of the Pythagorean theorem: If $x \perp y$, then

$$\|x+y\|^2 = \|x\|^2 + \|y\|^2. \tag{5.3}$$

Theorem 5.6. *(a) $M^{\perp}$ is a closed subspace of H.*
(b) $M^{\perp} = L(M)^{\perp} = \overline{L(M)}^{\perp}$

Here, $L(M)$ denotes the linear span of M, that is, $L(M)$ is the smallest subspace containing M. A more explicit description is also possible: $L(M) = \{\sum_{j=1}^{n} c_j m_j : c_j \in \mathbb{C}, m_j \in M, n \in \mathbb{N}\}$

Proof. (a) To show that $M^{\perp}$ is a subspace, let $x, y \in M^{\perp}$. Then, for arbitrary $m \in M$, $\langle x+y, m\rangle = \langle x, m\rangle + \langle y, m\rangle = 0$, so $x + y \in M^{\perp}$ also. A similar argument works for multiples of vectors from $M^{\perp}$.

If $x_n \in M^{\perp}$, $x \in H$, $x_n \to x$ and $m \in M$ is again arbitrary, then, by the continuity of the scalar product (Corollary 5.4),

$$\langle x, m\rangle = \lim_{n\to\infty} \langle x_n, m\rangle = 0,$$

so $x \in M^{\perp}$ also and $M^{\perp}$ turns out to be closed, as claimed.

(b) From the definition of $A^{\perp}$, it is clear that $A^{\perp} \supset B^{\perp}$ if $A \subset B$. Since obviously $M \subset L(M) \subset \overline{L(M)}$, we obtain that $\overline{L(M)}^{\perp} \subset L(M)^{\perp} \subset M^{\perp}$. On the other hand, if $x \in M^{\perp}$, then $\langle x, m\rangle = 0$ for all $m \in M$. Since the scalar product is linear in the second argument, this implies that $\langle x, y\rangle = 0$ for all $y \in L(M)$. Since the scalar product is also continuous, it now follows that in fact $\langle x, z\rangle = 0$ for all $z \in \overline{L(M)}$, that is, $x \in \overline{L(M)}^{\perp}$. □

Exercise 5.6. (a) Show that the closure of a subspace is a subspace again. (This shows that $\overline{L(M)}$ can be described as the smallest closed subspace containing M.)
(b) Show that $L(\overline{M}) \subset \overline{L(M)}$.
(c) Show that it can happen that $L(\overline{M}) \neq \overline{L(M)}$.
Suggestion: Consider $M = \{e_n : n \geq 1\} \subset \ell^2$.

Theorem 5.7. *Let $M \subset H$ be a closed subspace of H, and let $x \in H$. Then there exists a unique best approximation to x in M, that is, there exists a unique $y \in M$ so that*

$$\|x - y\| = \inf_{m \in M} \|x - m\|.$$

Proof. Write $d = \inf_{m \in M} \|x - m\|$ and pick a sequence $y_n \in M$ with $\|x - y_n\| \to d$. The parallelogram identity (5.2) implies that

$$\begin{aligned}\|y_m - y_n\|^2 &= \|(y_m - x) - (y_n - x)\|^2 \\ &= 2\|y_m - x\|^2 + 2\|y_n - x\|^2 - \|y_m + y_n - 2x\|^2 \\ &= 2\|y_m - x\|^2 + 2\|y_n - x\|^2 - 4\|(1/2)(y_m + y_n) - x\|^2.\end{aligned}$$

Now if $m, n \to \infty$, then the first two terms in this final expression both converge to $2d^2$, by the choice of y_n. Since $(1/2)(y_m + y_n) \in M$, we have that $\|(1/2)(y_m + y_n) - x\| \geq d$ for all m, n. It follows $\|y_m - y_n\| \to 0$ as $m, n \to \infty$, so y_n is a Cauchy sequence. Let $y = \lim_{n\to\infty} y_n$. Since M is closed, $y \in M$, and by the continuity of the norm, $\|x - y\| = \lim \|x - y_n\| = d$, so y is a best approximation.

To prove the uniqueness of y, assume that $y' \in M$ also satisfies $\|x - y'\| = d$. Then, by the above calculation, with y_m, y_n replaced by y, y', we have that

$$\begin{aligned}\|y - y'\|^2 &= 2\|y - x\|^2 + 2\|y' - x\|^2 - 4\|(1/2)(y + y') - x\|^2 \\ &= 4d^2 - 4\|(1/2)(y + y') - x\|^2.\end{aligned}$$

Again, since $(1/2)(y + y') \in M$, this last norm is $\geq d$, so the whole expression is ≤ 0 and we must have that $y = y'$, as desired. □

These best approximations can be used to project orthogonally onto closed subspaces of a Hilbert space. More precisely, we have the following:

Theorem 5.8. *Let $M \subset H$ be a closed subspace. Then every $x \in H$ has a unique representation of the form $x = y + z$, with $y \in M$, $z \in M^\perp$.*

Proof. Use Theorem 5.7 to define $y \in M$ as the best approximation to x from M, that is, $\|x - y\| \leq \|x - m\|$ for all $m \in M$. Let $z = x - y$. We want to show that $z \in M^\perp$. If $w \in M$, $w \neq 0$, and $c \in \mathbb{C}$, then

$$\|z\|^2 \leq \|x - (y + cw)\|^2 = \|z - cw\|^2 = \|z\|^2 + |c|^2 \, \|w\|^2 - 2 \text{ Re } c\langle z, w\rangle.$$

In particular, with $c = \frac{\langle w, z\rangle}{\|w\|^2}$, this shows that $|\langle w, z\rangle|^2 \leq 0$, so $\langle w, z\rangle = 0$, and since this holds for every $w \in M$, we see that $z \in M^\perp$, as desired.

To show that the decomposition from Theorem 5.8 is unique, suppose that $x = y + z = y' + z'$, with $y, y' \in M$, $z, z' \in M^\perp$. Then $y - y' = z' - z \in M \cap M^\perp = \{0\}$, so $y = y'$ and $z = z'$. □

Corollary 5.9. *For an arbitrary subset $A \subset H$, we have that $A^{\perp\perp} = \overline{L(A)}$.*

Proof. From the definition of $(\ldots)^{\perp}$, we see that $B \subset B^{\perp\perp}$, so Theorem 5.6(b) implies that $\overline{L(A)} \subset A^{\perp\perp}$.

On the other hand, if $x \in A^{\perp\perp}$, we can use Theorem 5.8 to write $x = y + z$ with $y \in \overline{L(A)}$, $z \in \overline{L(A)}^{\perp} = A^{\perp}$. The last equality again follows from Theorem 5.6(b). As just observed, we then also have that $y \in A^{\perp\perp}$ and thus $z = x - y \in A^{\perp} \cap A^{\perp\perp} = \{0\}$, so $x = y \in \overline{L(A)}$. □

We now introduce a linear operator that produces the decomposition from Theorem 5.8. Let $M \subset H$ be a closed subspace. We then define $P_M : H \to H$, $P_M x = y$, where $y \in M$ is as in Theorem 5.8; P_M is called the *(orthogonal) projection* onto M.

Proposition 5.10. $P_M \in B(H)$, $P_M^2 = P_M$, *and if* $M \neq \{0\}$, *then* $\|P_M\| = 1$.

Proof. We will only compute the operator norm of P_M here.

Exercise 5.7. Prove that P_M is linear and $P_M^2 = P_M$.

Write $x = P_M x + z$. Then $P_M x \in M$, $z \in M^{\perp}$, so, by the Pythagorean formula (5.3),

$$\|x\|^2 = \|P_M x\|^2 + \|z\|^2 \geq \|P_M x\|^2.$$

Thus $P_M \in B(H)$ and $\|P_M\| \leq 1$. On the other hand, if $x \in M$, then $P_M x = x$, so $\|P_M\| = 1$ if $M \neq \{0\}$. □

We saw in Chapter 4 that $(\ell^2)^* = \ell^2$, $(L^2)^* = L^2$. This is no coincidence.

Theorem 5.11 (Riesz Representation Theorem)**.** *Every $F \in H^*$ has the form $F(x) = \langle y, x\rangle$, for some $y = y_F \in H$. Moreover, $\|F\| = \|y_F\|$.*

We can say more: conversely, every $y \in H$ generates a bounded, linear functional $F = F_y$ via $F_y(x) = \langle y, x\rangle$. So we can define a map $I : H \to H^*$, $y \mapsto F_y$. This map is injective (why?), and, by the Riesz representation theorem, I is also surjective and isometric, so we obtain an identification of H with H^*. We need to be a little careful here, because I is *antilinear*, that is, $F_{y+z} = F_y + F_z$, as usual, but $F_{cy} = \overline{c} F_y$.

Exercise 5.8. Deduce from this that Hilbert spaces are reflexive. If we ignore the identification maps and just pretend that $H = H^*$ and proceed formally, then this becomes obvious: $H^{**} = (H^*)^* = H^* = H$. Please give a careful argument. Recall that you really need to show that $j(H) = H^{**}$, where j was defined in Chapter 4. (This is surprisingly awkward to write down; perhaps you want to use the fact that $F : X \to \mathbb{C}$ is antilinear precisely if $\overline{F}$ is linear.)

Exercise 5.9. Let X be a (complex) vector space and let $F : X \to \mathbb{C}$ be a linear functional, $F \neq 0$.
(a) Show that codim $N(F) = 1$, that is, show that there exists a one-dimensional subspace $M \subset X$, $M \cap N(F) = \{0\}$, $M + N(F) = X$. (This is an immediate consequence of linear algebra facts, but you can also do it by hand.)
(b) Let F, G be linear functionals with $N(F) = N(G)$. Then $F = cG$ for some $c \in \mathbb{C}$, $c \neq 0$.

Proof of Theorem 5.11. This is surprisingly easy; Exercise 5.9 provides the motivation for the following argument and also explains why this procedure (take an arbitrary vector from $N(F)^{\perp}$) works.

If $F = 0$, we can of course just take $y = 0$. If $F \neq 0$, then $N(F) \neq H$, and $N(F)$ is a closed subspace because F is continuous. Therefore, Theorem 5.8 shows that $N(F)^{\perp} \neq \{0\}$. Pick a vector $z \in N(F)^{\perp}$, $z \neq 0$. Then, for arbitrary $x \in H$, $F(z)x - F(x)z \in N(F)$, so

$$0 = \langle z, F(z)x - F(x)z \rangle = F(z)\langle z, x \rangle - F(x)\|z\|^2.$$

Rearranging, we obtain that $F(x) = \langle y, x \rangle$, with $y = \frac{\overline{F(z)}}{\|z\|^2} z$.

Since $|\langle y, x \rangle| \leq \|y\| \, \|x\|$, we have that $\|F\| \leq \|y\|$. On the other hand, $F(y) = \|y\|^2$, so $\|F\| = \|y\|$. □

Exercise 5.10. Corollary 4.2(b), when combined with the Riesz Representation Theorem, implies that

$$\|x\| = \sup_{\|y\|=1} |\langle y, x \rangle| \, .$$

Give a quick direct proof of this fact.

Exercise 5.11. We don't need the Hahn-Banach Theorem on Hilbert spaces because the Riesz Representation Theorem gives a much more explicit description of the dual space. Show that it in fact implies the following stronger version of Hahn-Banach: If $F_0 : H_0 \to \mathbb{C}$ is a bounded linear functional on a subspace H_0, then there exists a *unique* bounded linear extension $F : H \to \mathbb{C}$ with $\|F\| = \|F_0\|$.

Remark: If you want to avoid using the Hahn-Banach Theorem here, you could as a first step extend F_0 to $\overline{H_0}$, by using Exercise 2.26.

The Riesz Representation Theorem also shows that on a Hilbert space, $x_n \xrightarrow{w} x$ if and only if $\langle y, x_n \rangle \to \langle y, x \rangle$; compare Exercise 4.7.

Exercise 5.12. Assume that $x_n \xrightarrow{w} x$.
(a) Show that $\|x\| \le \liminf_{n\to\infty} \|x_n\|$.
(b) Show that it can happen that $\|x\| < \liminf \|x_n\|$.
(c) On the other hand, if $\|x_n\| \to \|x\|$, then $\|x_n - x\| \to 0$ (prove this).

As our final topic in this chapter, we discuss orthonormal bases in Hilbert spaces.

Definition 5.12. A subset $\{x_\alpha : \alpha \in I\}$ is called an *orthogonal system* if $\langle x_\alpha, x_\beta \rangle = 0$ for all $\alpha \ne \beta$. If, in addition, all x_α are normalized (so $\langle x_\alpha, x_\beta \rangle = \delta_{\alpha\beta}$), we call $\{x_\alpha\}$ an *orthonormal system* (ONS). A maximal ONS is called an *orthonormal basis* (ONB).

Theorem 5.13. *Every Hilbert space has an ONB. Moreover, any ONS can be extended to an ONB.*

This looks very plausible (if an ONS isn't maximal yet, just keep adding vectors). The formal proof depends on Zorn's Lemma; we don't want to do it here.

Theorem 5.14. *Let $\{x_\alpha\}$ be an ONB. Then, for every $y \in H$, we have that*

$$y = \sum_{\alpha\in I} \langle x_\alpha, y \rangle x_\alpha,$$

$$\|y\|^2 = \sum_{\alpha\in I} |\langle x_\alpha, y \rangle|^2 \qquad \textit{(Parseval's identity).}$$

If, conversely, $c_\alpha \in \mathbb{C}$, $\sum_{\alpha\in I} |c_\alpha|^2 < \infty$, then the series $\sum_{\alpha\in I} c_\alpha x_\alpha$ converges to an element $y \in H$.

To make this precise, we need to define sums over arbitrary index sets. We encountered this problem before, in Chapter 2, when defining the space $\ell^2(I)$, and we will use the same procedure here: $\sum_{\alpha\in I} w_\alpha = z$ means that $w_\alpha \ne 0$ for at most countably many $\alpha \in I$ and if $\{\alpha_n\}$ is an arbitrary enumeration of these α's, then $\lim_{N\to\infty} \sum_{n=1}^{N} w_{\alpha_n} = z$. In this definition, we can have $w_\alpha, z \in H$ or $\in \mathbb{C}$. In this latter case, we can also again use counting measure on I to obtain a more elegant formulation.

Theorem 5.14 can now be rephrased in a more abstract way. Consider the map

$$U : H \to \ell^2(I), \qquad (Uy)_\alpha = \langle x_\alpha, y \rangle.$$

Theorem 5.14 says that this is well defined, bijective, and isometric. Moreover, U is also obviously linear. So, summing up, we have a bijection $U \in B(H, \ell^2)$ that also preserves the scalar product: $\langle Ux, Uy \rangle = \langle x, y \rangle$.

Exercise 5.13. Prove this last statement. *Hint:* Polarization!

Such maps are called *unitary;* they preserve the complete Hilbert space structure. In other words, we can now say that Theorem 5.14 shows that $H \cong \ell^2(I)$ for an arbitrary Hilbert space H; more specifically, I can be taken as the index set of an ONB. So we have a one-size-fits-all model space, namely $\ell^2(I)$; there is no such universal model for Banach spaces.

There is a version of Theorem 5.14 for ONS; actually, we will prove the two results together.

Theorem 5.15. *Let $\{x_\alpha\}$ be an ONS. Then, for every $y \in H$, we have that*

$$P_{\overline{L(x_\alpha)}}\, y = \sum_{\alpha \in I} \langle x_\alpha, y\rangle x_\alpha,$$
$$\|y\|^2 \geq \sum_{\alpha \in I} |\langle x_\alpha, y\rangle|^2 \qquad \textit{(Bessel's inequality).}$$

Proof of Theorems 5.14, 5.15. We start by establishing Bessel's inequality for finite ONS $\{x_1, \ldots, x_N\}$. Let $y \in H$ and write

$$y = \sum_{n=1}^N \langle x_n, y\rangle x_n + \left(y - \sum_{n=1}^N \langle x_n, y\rangle x_n \right).$$

A calculation shows that the two terms on the right-hand side are orthogonal, so

$$\begin{aligned} \|y\|^2 &= \left\| \sum_{n=1}^N \langle x_n, y\rangle x_n \right\|^2 + \left\| y - \sum_{n=1}^N \langle x_n, y\rangle x_n \right\|^2 \\ &= \sum_{n=1}^N |\langle x_n, y\rangle|^2 + \left\| y - \sum_{n=1}^N \langle x_n, y\rangle x_n \right\|^2 \geq \sum_{n=1}^N |\langle x_n, y\rangle|^2. \end{aligned}$$

This is Bessel's inequality for finite ONS. It now follows that the sets $\{\alpha \in I : |\langle x_\alpha, y\rangle| \geq 1/n\}$ are finite, so $\{\alpha : \langle x_\alpha, y\rangle \neq 0\}$ is countable. Let $\{\alpha_1, \alpha_2, \ldots\}$ be an enumeration. Then, by Bessel's inequality (we're still referring to the version for finite ONS), $\lim_{N\to\infty} \sum_{n=1}^N |\langle x_{\alpha_n}, y\rangle|^2$ exists, and, since we have absolute convergence here, the limit is independent of the enumeration. If we recall how $\sum_{\alpha\in I} \ldots$ was defined, we see that we have proved the general version of Bessel's inequality.

As the next step, define $y_n = \sum_{j=1}^n \langle x_{\alpha_j}, y\rangle x_{\alpha_j}$. If $n \geq m$ (say), then

$$\|y_m - y_n\|^2 = \left\| \sum_{j=m+1}^n \langle x_{\alpha_j}, y\rangle x_{\alpha_j} \right\|^2 = \sum_{j=m+1}^n \left| \langle x_{\alpha_j}, y\rangle \right|^2.$$

This shows that y_n is a Cauchy sequence. Let $y' = \lim y_n = \sum_{j=1}^{\infty} \langle x_{\alpha_j}, y\rangle x_{\alpha_j}$. By the continuity of the scalar product,

$$\langle x_{\alpha_k}, y - y'\rangle = \langle x_{\alpha_k}, y\rangle - \sum_{j=1}^{\infty} \langle x_{\alpha_j}, y\rangle \delta_{jk} = 0$$

for all $k \in \mathbb{N}$, and if $\alpha \in I \setminus \{\alpha_j\}$, then we also obtain that

$$\langle x_\alpha, y - y'\rangle = -\langle x_\alpha, y'\rangle = -\sum_{j=1}^{\infty} \langle x_{\alpha_j}, y\rangle \langle x_\alpha, x_{\alpha_j}\rangle = 0.$$

So $y - y' \in \{x_\alpha\}^\perp = \overline{L(x_\alpha)}^\perp$, and, by its construction, $y' \in \overline{L(x_\alpha)}$. Thus $y' = P_{\overline{L(x_\alpha)}}\, y$, as claimed in Theorem 5.15. It now also follows that $\sum_{\alpha \in I} \langle x_\alpha, y\rangle x_\alpha$ exists because we always obtain the same limit $y' = P_{\overline{L(x_\alpha)}}\, y$, no matter how the α_j are arranged.

To obtain Theorem 5.14, we observe that if $\{x_\alpha\}$ is an ONB, then $\overline{L(x_\alpha)} = H$. Indeed, if this were not true, then the closed subspace $\overline{L(x_\alpha)}$ would have to have a non-zero orthogonal complement, by Theorem 5.8, and we could pass to a bigger ONS by adding a normalized vector from this orthogonal complement. So $\overline{L(x_\alpha)} = H$ if $\{x_\alpha\}$ is an ONB, but then also $y' = y$, and Parseval's identity now follows from the continuity of the norm:

$$\|y\|^2 = \lim_{N \to \infty} \left\| \sum_{i=1}^{N} \langle x_{\alpha_i}, y\rangle x_{\alpha_i} \right\|^2 = \lim_{N \to \infty} \sum_{i=1}^{N} |\langle x_{\alpha_i}, y\rangle|^2 = \sum_{\alpha \in I} |\langle x_\alpha, y\rangle|^2$$

Finally, similar arguments show that $\sum c_\alpha x_\alpha$ exists for $c \in \ell^2(I)$ (consider the partial sums and check that these form a Cauchy sequence). $\square$

We can try to summarize this as follows: once an ONB is fixed, we may use the coefficients with respect to this ONB to manipulate vectors; in particular, there is an easy formula (Parseval's identity) that will give the norm in terms of these coefficients. The situation is quite similar to linear algebra: coefficients with respect to a fixed basis is all we need to know about vectors. Note, however, that ONB's are not bases in the sense of linear algebra: we use *infinite* linear combinations (properly defined as limits) to represent vectors. *Algebraic* bases on infinite-dimensional Hilbert spaces exist, too, but they are almost entirely useless (for example, they can never be countable).

Exercise 5.14. Show that $\{e_n : n \in \mathbb{N}\}$ is an ONB of $\ell^2 = \ell^2(\mathbb{N})$.

Exercise 5.15. Show that $\{e^{inx} : n \in \mathbb{Z}\}$ is an ONB of $L^2\left((-\pi,\pi); \frac{dx}{2\pi}\right)$.

Suggestion: You should not try to prove the maximality directly, but rather refer to suitable facts from Analysis (continuous functions are dense in L^2, and they can be uniformly approximated, on compact sets, by trigonometric polynomials).

Exercise 5.16. (a) For $f \in L^2(-\pi,\pi)$, define the nth *Fourier coefficient* as

$$\widehat{f}_n = \int_{-\pi}^{\pi} f(x)e^{-inx}\, dx,$$

and use the result from Exercise 5.15 to establish the following formula, which is also known as *Parseval's identity:*

$$\sum_{n=-\infty}^{\infty} \left|\widehat{f}_n\right|^2 = 2\pi \int_{-\pi}^{\pi} |f(x)|^2\, dx$$

(b) Prove that

$$\sum_{n=1}^{\infty} \frac{1}{n^2} = \frac{\pi^2}{6}.$$

Suggestion: Part (a) with $f(x) = x$.

Exercise 5.17. The *Rademacher functions* $R_0(x) = 1$,

$$R_n(x) = \begin{cases} 1 & x \in \bigcup_{k=0}^{2^{n-1}-1} [k2^{1-n}, (2k+1)2^{-n}) \\ -1 & \text{else} \end{cases}$$

form an ONS, but not an ONB in $L^2(0,1)$. (Please plot the first few functions to get your bearings here.)

Exercise 5.18. (a) Let $U : H_1 \to H_2$ be a unitary map between Hilbert spaces, and let $\{x_\alpha\}$ be an ONB of H_1. Show that $\{Ux_\alpha\}$ is an ONB of H_2.
(b) Conversely, let $U : H_1 \to H_2$ be a linear map that maps an ONB to an ONB again. Show that U is unitary.

Exercise 5.19. Show that a Hilbert space is separable precisely if it has a countable ONB.

6. Operators in Hilbert spaces

Let H be a Hilbert space. In this chapter, we are interested in basic properties of operators $T \in B(H)$ on this Hilbert space. First of all, we would like to define an adjoint operator T^*, and its defining property should be given by $\langle T^*y, x\rangle = \langle y, Tx\rangle$. It is not completely clear, however, that this indeed defines a new operator T^*. To make this idea precise, we proceed as follows: Fix $y \in H$ and consider the map $H \to \mathbb{C}$, $x \mapsto \langle y, Tx\rangle$. It is clear that this is a linear map, and

$$|\langle y, Tx\rangle| \le \|y\|\,\|Tx\| \le \|y\|\,\|T\|\,\|x\|,$$

so the map is also bounded. By the Riesz Representation Theorem, there exists a unique vector $z = z_y \in H$ so that $\langle y, Tx\rangle = \langle z_y, x\rangle$ for all $x \in H$. We can now define a map $T^* : H \to H$, $T^*y = z_y$. By construction, we then indeed have that $\langle T^*y, x\rangle = \langle y, Tx\rangle$ for all $x, y \in H$; conversely, this condition uniquely determines T^*y for all $y \in H$. We call T^* the *adjoint operator* (of T).

Theorem 6.1. *Let $S, T \in B(H)$, $c \in \mathbb{C}$. Then:*
(a) $T^* \in B(H)$;
(b) $(S+T)^* = S^* + T^*$, $(cT)^* = \overline{c}T^*$;
(c) $(ST)^* = T^*S^*$;
(d) $T^{**} = T$;
(e) If T is invertible, then T^* is also invertible and $(T^*)^{-1} = (T^{-1})^*$;
(f) $\|T\| = \|T^*\|$, $\|TT^*\| = \|T^*T\| = \|T\|^2$ (the C^* property*)*

Here, we call $T \in B(H)$ *invertible* (it would be more precise to say: invertible in $B(H)$) if there exists an $S \in B(H)$ so that $ST = TS = 1$. In this case, S with these properties is unique and we call it the inverse of T and write $S = T^{-1}$. Notice that this version of invertibility requires more than just injectivity of T as a map: we also require the inverse map to be continuous and defined everywhere on H (and linear, but this is automatic). So we can also say that $T \in B(H)$ is invertible (in this sense) precisely if T is bijective on H and has a continuous inverse. Actually, Corollary 3.3 shows that this continuity is automatic also, so $T \in B(H)$ is invertible precisely if T is a bijective map.

Exercise 6.1. (a) Show that it is not enough to have just one of the equations $ST = 1$, $TS = 1$: Construct two non-invertible maps $S, T \in B(H)$ (on some Hilbert space H; $H = \ell^2$ seems a good choice) that nevertheless satisfy $ST = 1$.
(b) However, if H is finite-dimensional and $ST = 1$, then both S and T will be invertible. Prove this.

Proof. (a) The (anti-)linearity of the scalar product implies that T^* is linear; for example, $\langle cT^*y, x\rangle = \langle T^*y, \overline{c}x\rangle = \langle y, T(\overline{c}x)\rangle = \langle cy, Tx\rangle$ for all $x \in H$, so $T^*(cy) = cT^*y$. Furthermore,

$$\sup_{\|y\|=1} \|T^*y\| = \sup_{\|x\|=\|y\|=1} |\langle T^*y, x\rangle| = \sup_{\|x\|=\|y\|=1} |\langle y, Tx\rangle| = \|T\|,$$

so $T^* \in B(H)$ and $\|T^*\| = \|T\|$.

Parts (b), (c) follow directly from the definition of the adjoint operator. For example, to verify (c), we just observe that $\langle y, STx\rangle = \langle S^*y, Tx\rangle = \langle T^*S^*y, x\rangle$, so $(ST)^*y = T^*S^*y$.

(d) We have that $\langle y, T^*x\rangle = \overline{\langle T^*x, y\rangle} = \overline{\langle x, Ty\rangle} = \langle Ty, x\rangle$, thus $T^{**}y = Ty$.

(e) Obviously, $1^* = 1$. So if we take adjoints in $TT^{-1} = T^{-1}T = 1$ and use (c), we obtain that $(T^{-1})^*T^* = T^*(T^{-1})^* = 1$, and since $(T^{-1})^* \in B(H)$, this says that T^* is invertible and $(T^*)^{-1} = (T^{-1})^*$.

(f) We already saw in the proof of part (a) that $\|T^*\| = \|T\|$. It is then also clear that $\|T^*T\| \le \|T^*\|\|T\| = \|T\|^2$. On the other hand,

$$\|T^*T\| = \sup_{\|x\|=\|y\|=1} |\langle y, T^*Tx\rangle| \ge \sup_{\|x\|=1} |\langle x, T^*Tx\rangle| = \sup_{\|x\|=1} \|Tx\|^2 = \|T\|^2,$$

so $\|T^*T\| = \|T\|^2$. If applied to T^* in place of T, this also gives that $\|TT^*\| = \|T^{**}T^*\| = \|T^*\|^2 = \|T\|^2$. $\square$

Theorem 6.2. *Let $T \in B(H)$. Then $N(T^*) = R(T)^\perp$.*

Proof. We have $x \in N(T^*)$ precisely if $\langle T^*x, y\rangle = 0$ for all $y \in H$, and this happens if and only if $\langle x, Ty\rangle = 0$ ($y \in H$). This, in turn, holds if and only if $x \in R(T)^\perp$. $\square$

We will be especially interested in Hilbert space operators with additional properties.

Definition 6.3. Let $T \in B(H)$. We call T *self-adjoint* if $T = T^*$, *unitary* if $TT^* = T^*T = 1$ and *normal* if $TT^* = T^*T$.

So self-adjoint and unitary operators are also normal. We introduced unitary operators earlier, in Chapter 5, in the more general setting of operators between two Hilbert spaces; recall that we originally defined these as maps that preserve the complete Hilbert space structure (that is, the algebraic structure and the scalar product). Theorem 6.4(b) below will make it clear that the new definition is equivalent to the old one (for maps on one space). Also, notice that U is unitary precisely if U is invertible (in $B(H)$, as above) and $U^{-1} = U^*$.

Here are some additional reformulations:

Theorem 6.4. *Let $U \in B(H)$. Then the following statements are equivalent:*
(a) U is unitary;
(b) U is bijective and $\langle Ux, Uy\rangle = \langle x, y\rangle$ for all $x, y \in H$;
(c) U is surjective and isometric (that is, $\|Ux\| = \|x\|$ for all $x \in H$).

Exercise 6.2. Prove Theorem 6.4. *Suggestion:* Use polarization to derive (b) from (c).

We now take a second look at (orthogonal) projections. Recall that, by definition, the projection on M (where $M \subset H$ is a closed subspace) is the operator that sends $x \in H$ to $y \in M$, where y is the part from M in the (unique) decomposition $x = y + z$, $y \in M$, $z \in M^\perp$. If $P = P_M$ is such a projection, then it has the following properties: $P^2 = P$ (see Proposition 5.10), $R(P) = M$, $N(P) = M^\perp$

Exercise 6.3. Prove these latter two properties. Also, show that $Px = x$ if and only if $x \in M = R(P)$.

Theorem 6.5. *Let $P \in B(H)$. Then the following are equivalent:*
(a) P is a projection;
(b) $1 - P$ is a projection;
(c) $P^2 = P$ and $R(P) = N(P)^\perp$;
(d) $P^2 = P$ and P is self-adjoint;
(e) $P^2 = P$ and P is normal.

Proof. (a) $\Longrightarrow$ (b): It is clear from the definition of P_M and the fact that $M^{\perp\perp} = M$ that $1 - P$ is the projection onto $M^\perp$ if P is the projection onto M.

(b) $\Longrightarrow$ (a): This is the same statement, applied to $1 - P$ in place of P.

(a) $\Longrightarrow$ (c): This was already observed above, see Exercise 6.3.

(c) $\Longrightarrow$ (a): If $y \in R(P)$, so $y = Pu$ for some $u \in H$, then we obtain that $Py = P^2u = Pu = y$. On the other hand, if $z \in R(P)^\perp = N(P)$ (we make use of the fact that $N(P)$ is a closed subspace, because P is continuous), then $Pz = 0$. Now let $x \in H$ be arbitrary and use Theorem 5.8 to decompose $x = y + z$, $y \in R(P)$, $z \in R(P)^\perp$. Note that $R(P)$ is a closed subspace because it is the orthogonal complement of $N(P)$ by assumption. By our earlier observations, $Px = Py + Pz = y$, so indeed P is the projection on $R(P)$.

(a) $\Longrightarrow$ (d): Again, we already know that $P^2 = P$. Moreover, for arbitrary $x, y \in H$, we have that

$$\langle Px, Py\rangle = \langle x, Py\rangle = \langle Px, y\rangle, \tag{6.1}$$

because, for example, $x = Px + (1-P)x$, but $(1-P)x \perp Py$. The second equality in (6.1) says that $P^* = P$, as desired.

(d) $\Longrightarrow$ (e) is trivial.

(e) $\Longrightarrow$ (c): Since P is normal, we have that

$$\|Px\|^2 = \langle Px, Px\rangle = \langle P^*Px, x\rangle = \langle PP^*x, x\rangle = \|P^*x\|^2.$$

In particular, this implies that $N(P) = N(P^*)$, and Theorem 6.2 then shows that $N(P) = R(P)^{\perp}$. We could finish the proof by passing to the orthogonal complements here if we also knew that $R(P)$ is closed. We will establish this by showing that $R(P) = N(1-P)$ (which is closed, being the null space of a continuous operator). Clearly, if $x \in R(P)$, then $x = Py$ for some $y \in H$ and thus $(1-P)x = P^2y - Py = 0$, so $x \in N(1-P)$. Conversely, if $x \in N(1-P)$, then $x = Px \in R(P)$. $\square$

For later use, we also note the following technical property of projections:

Proposition 6.6. *Let P, Q be projections. Then PQ is a projection if and only if $PQ = QP$. In this case, $R(PQ) = R(P) \cap R(Q)$.*

Proof. If PQ is a projection, then it satisfies condition (d) from Theorem 6.5, so $PQ = (PQ)^* = Q^*P^* = QP$. Conversely, if we assume that $PQ = QP$, then the same calculation shows that PQ is self-adjoint. Since we also have that $(PQ)^2 = PQPQ = PPQQ = P^2Q^2 = PQ$, it now follows from Theorem 6.5 that PQ is a projection.

To find its range, we observe that $R(PQ) \subset R(P)$, but also $R(PQ) = R(QP) \subset R(Q)$, so $R(PQ) \subset R(P) \cap R(Q)$. On the other hand, if $x \in R(P) \cap R(Q)$, then $Px = Qx = x$, so $PQx = x$ and thus $x \in R(PQ)$. $\square$

On the finite-dimensional Hilbert space $H = \mathbb{C}^n$, every operator $T \in B(\mathbb{C}^n)$ (equivalently, every matrix $T \in \mathbb{C}^{n\times n}$) can be brought to a relatively simple canonical form (the Jordan normal form) by a change of basis. In fact, usually operators are diagonalizable.

Exercise 6.4. Can you establish the following precise version: The set of diagonalizable matrices is a dense open subset of $\mathbb{C}^{n\times n}$, where we use the topology generated by the operator norm. (In fact, by Theorem 2.15, any other norm will give the same topology.)

The situation on infinite-dimensional Hilbert spaces is much more complicated. We cannot hope for a normal form theory for *general* Hilbert space operators. In fact, the following much more modest question is a famous long-standing open problem:

Does every $T \in B(H)$ have a non-trivial invariant subspace?
(the invariant subspace problem)

Here, a closed subspace $M \subset H$ is called invariant if $TM \subset M$; the trivial invariant subspaces are $\{0\}$ and H.

Exercise 6.5. (a) Show that every $T \in \mathbb{C}^{n\times n} = B(\mathbb{C}^n)$ has a non-trivial invariant subspace.
(b) Show that $\overline{L(\{T^n x : n \geq 1\})}$ is an invariant subspace (possibly trivial) for every $x \in H$.
(c) Deduce from (b) that every $T \in B(H)$ on a *non-separable* Hilbert space H has a non-trivial invariant subspace.

Of course, we wouldn't really gain very much even from a positive answer to the invariant subspace problem; this would just make sure that every operator has some smaller part that could be considered separately. The fact that the invariant subspace problem is universally recognized as an exceedingly hard problem makes any attempt at a general structure theory for Hilbert space operators completely hopeless.

We will therefore focus on normal operators, which form an especially important subclass of Hilbert space operators. Here, we will be able to develop a powerful theory. The fundamental result here is the *spectral theorem;* we will prove this in Chapter 10, after a few detours. It is also useful to recall from linear algebra that a normal *matrix* $T \in \mathbb{C}^{n\times n}$ can be diagonalized; in fact, this is done by changing from the original basis to a new ONB, consisting of the eigenvectors of T.

Generally speaking, the eigenvalues and eigenvectors of a matrix take center stage in the analysis in the finite-dimensional setting, so it seems a good idea to try to generalize these notions. We do this as follows (actually, we only generalize the concept of an eigenvalue here):

Definition 6.7. For $T \in B(H)$, define

$$\rho(T) = \{z \in \mathbb{C} : T - z \text{ is invertible in } B(H)\},$$
$$\sigma(T) = \mathbb{C} \setminus \rho(T).$$

We call $\rho(T)$ the *resolvent set* of T and $\sigma(T)$ the *spectrum* of T.

Exercise 6.6. Show that $\sigma(T)$ is exactly the set of eigenvalues of T if $T \in B(\mathbb{C}^n)$ is a matrix.

This confirms that we may hope to have made a good definition, but perhaps the more obvious try would actually have gone as follows: Call $z \in \mathbb{C}$ an *eigenvalue* of $T \in B(H)$ if there exists an $x \in H$, $x \neq 0$, so that $Hx = zx$, and introduce $\sigma_p(T)$ as the set of eigenvalues of T; we also call $\sigma_p(T)$ the *point spectrum* of T.

However, this doesn't work very well in the infinite-dimensional setting:

Exercise 6.7. Consider the operator $S \in \ell^2(\mathbb{Z})$, $(Sx)_n = x_{n+1}$ (S as in shift), and prove the following facts about S:
(a) S is unitary;
(b) $\sigma_p(S) = \emptyset$.

We can also obtain an example of a self-adjoint operator with no eigenvalues from this, by letting $T = S + S^*$. Then $T = T^*$ (obvious), and again $\sigma_p(T) = \emptyset$ (not obvious, and in fact you will probably need to use a few facts about difference equations to prove this; this part of the problem is optional).

Exercise 6.8. Show that $\sigma_p \subset \sigma$.

Exercise 6.9. Here's another self-adjoint operator with no eigenvalues; compare Exercise 6.7. Define $T : L^2(0,1) \to L^2(0,1)$ by $(Tf)(x) = xf(x)$.
(a) Show that $T \in B(L^2(0,1))$ and $T = T^*$, and compute $\|T\|$.
(b) Show that $\sigma_p(T) = \emptyset$. Can you also show that $\sigma(T) = [0,1]$?

Exercise 6.10. Let $s(x,y)$ be a sesquilinear form that is bounded in the sense that

$$M \equiv \sup_{\|x\|=\|y\|=1} |s(x,y)| < \infty.$$

Show that there is a unique operator $T \in B(H)$ so that $s(x,y) = \langle x, Ty\rangle$. Show also that $\|T\| = M$.

Hint: Apply the Riesz Representation Theorem to the map $x \mapsto \overline{s(x,y)}$, for fixed but arbitrary $y \in H$.

Exercise 6.11. Let $T : H \to H$ be a linear operator, and assume that $\langle Tx, y\rangle = \langle x, Ty\rangle$ for all $x, y \in H$. Show that T is bounded (the *Hellinger-Toeplitz Theorem*).
Suggestion: Show that T is closed and apply the closed graph theorem.

7. Banach algebras

Definition 7.1. A is called a *Banach algebra* (with unit) if: (1) A is a Banach space;
(2) There is a multiplication $A \times A \to A$ that has the following properties:

$$(xy)z = x(yz), \quad (x+y)z = xz + yz, \quad x(y+z) = xy + xz,$$
$$c(xy) = (cx)y = x(cy)$$

for all $x, y, z \in A$, $c \in \mathbb{C}$. Moreover, there is a *unit element* e: $ex = xe = x$ for all $x \in A$;
(3) $\|e\| = 1$;
(4) $\|xy\| \le \|x\| \, \|y\|$ for all $x, y \in A$.

So a Banach algebra is an algebra (a vector space with multiplication, satisfying the usual algebraic rules) and also a Banach space, and these two structures are compatible (see conditions (3), (4)).

At the end of the last chapter, we decided to try to analyze normal operators on a Hilbert space H. Banach algebras will prove useful here, because of the following:

Example 7.1. If X is a Banach space, then $A = B(X)$ is a Banach algebra, with the composition of operators as multiplication and the operator norm. Indeed, we know from Theorem 2.12(b) that A is a Banach space, and composition of operators has the properties from (2) of Definition 7.1. The identity operator 1 is the unit element; of course $\|1\| = \sup_{\|x\|=1} \|x\| = 1$, as required, and (4) was discussed in Exercise 2.25.

Of course, there are more examples:

Example 7.2. $A = \mathbb{C}$ with the usual multiplication and the absolute value as norm is a Banach algebra.

Example 7.3. $A = C(K)$ with the pointwise multiplication $(fg)(x) = f(x)g(x)$ is a Banach algebra. Most properties are obvious. The unit element is the function $e(x) \equiv 1$; clearly, this has norm 1, as required. To verify (4), notice that

$$\|fg\| = \max_{x\in K} |f(x)g(x)| \le \max_{x\in K} |f(x)| \max_{x\in K} |g(x)| = \|f\| \, \|g\|.$$

Example 7.4. Similarly, $A = L^\infty$ and $A = \ell^\infty$ with the pointwise multiplication are Banach algebras.

Notice that the last three examples are in fact *commutative* Banach algebras, that is, $xy = yx$ for all $x, y \in A$. On the other hand, $B(X)$ is not commutative if $\dim X > 1$.

Example 7.5. $A = L^1(\mathbb{R})$ with the convolution product

$$(fg)(x) = (f * g)(x) = \int_{-\infty}^{\infty} f(t)g(x-t)\, dt$$

satisfies most of the properties from Definition 7.1, but does not have a unit element, so this would provide an example of a Banach algebra without a unit.

On the other hand, the discrete analog $A = \ell^1(\mathbb{Z})$ with the convolution product

$$(xy)_n = \sum_{j=-\infty}^{\infty} x_j y_{n-j}$$

is a Banach algebra with unit. Both L^1 and ℓ^1 are commutative.

Exercise 7.1. Prove the claims about the unit elements: Show that there is no function $f \in L^1(\mathbb{R})$ so that $f * g = g * f = g$ for all $g \in L^1(\mathbb{R})$. Also, find the unit element e of $\ell^1(\mathbb{Z})$.

We now start to develop the general theory of Banach algebras.

Theorem 7.2. *Multiplication is continuous in Banach algebras: If $x_n \to x$, $y_n \to y$, then $x_n y_n \to xy$.*

Proof.

$$\begin{aligned}\|x_n y_n - xy\| &\le \|(x_n - x)y_n\| + \|x(y_n - y)\| \\ &\le \|x_n - x\|\,\|y_n\| + \|x\|\,\|y_n - y\| \to 0\end{aligned}$$

□

We call $x \in A$ *invertible* if there exists $y \in A$ so that $xy = yx = e$. Note that on the Banach algebra $B(H)$, this reproduces the definition of invertibility in $B(H)$ that was given earlier, in Chapter 6. Returning to the general situation, we observe that if $x \in A$ is invertible, then y with these properties is unique. We write $y = x^{-1}$ and call x^{-1} the *inverse* of x. We denote the set of invertible elements by $G(A)$. Here, the choice of symbol is motivated by the fact that $G(A)$ is a group, with multiplication as the group operation. Indeed, we have that $xy \in G(A)$ and $x^{-1} \in G(A)$ if $x, y \in G(A)$; this can be verified by just writing down the inverses: $(xy)^{-1} = y^{-1}x^{-1}$, $(x^{-1})^{-1} = x$. Moreover, $e \in G(A)$ ($e^{-1} = e$), and of course multiplication is associative.

If A, B are algebras, then a map $\phi : A \to B$ is called a *homomorphism* if it preserves the algebraic structure. More precisely, we demand that ϕ is linear (as a map between vector spaces) and $\phi(xy) = \phi(x)\phi(y)$ for all $x, y \in A$. If $B = \mathbb{C}$ here and $\phi \not\equiv 0$, we call ϕ a *complex homomorphism.*

Proposition 7.3. *Let ϕ be a complex homomorphism. Then $\phi(e) = 1$ and $\phi(x) \neq 0$ for all $x \in G(A)$.*

Proof. Since $\phi \not\equiv 0$, there is a $y \in A$ with $\phi(y) \neq 0$. Since $\phi(y) = \phi(ey) = \phi(e)\phi(y)$, it follows that $\phi(e) = 1$. If $x \in G(A)$, then $\phi(x)\phi(x^{-1}) = \phi(e) = 1$, so $\phi(x) \neq 0$. □

Exercise 7.2. Let A be an algebra, with unit e. True or false:
(a) $fx = x$ for all $x \in A \Longrightarrow f = e$;
(b) $0x = 0$ for all $x \in A$;
(c) $xy = 0 \Longrightarrow x = 0$ or $y = 0$;
(d) $xy = zx = e \Longrightarrow x \in G(A)$ and $y = z = x^{-1}$;
(e) $xy, yx \in G(A) \Longrightarrow x, y \in G(A)$;
(f) $xy = e \Longrightarrow x \in G(A)$ or $y \in G(A)$;
(g) If B is another algebra with unit e' and $\phi : A \to B$ is a homomorphism, $\phi \not\equiv 0$, then $\phi(e) = e'$.

Theorem 7.4. *Let A be a Banach algebra. If $x \in A$, $\|x\| < 1$, then $e - x \in G(A)$ and*

$$(e-x)^{-1} = \sum_{n=0}^{\infty} x^n. \tag{7.1}$$

Moreover, if ϕ is a complex homomorphism, then $|\phi(x)| < 1$.

Here, we define $x^n = xx\cdots x$ as the n-fold product of x with itself, and $x^0 := e$. The series from (7.1) is then defined, as usual, as the norm limit of the partial sums (existence of this limit is part of the statement, of course). It generalizes the geometric series to the Banach algebra setting and is called the *Neumann series.*

Proof. Property (4) from Definition 7.1 implies that $\|x^n\| \leq \|x\|^n$. Since $\|x\| < 1$, we now see that $\sum \|x^n\|$ converges. It follows that the Neumann series converges, too (see Exercise 2.22). By the continuity of the multiplication in A,

$$\begin{aligned}(e-x)\sum_{n=0}^{\infty} x^n &= (e-x)\lim_{N\to\infty}\sum_{n=0}^{N} x^n = \lim_{N\to\infty}(e-x)\sum_{n=0}^{N} x^n \\ &= \lim_{N\to\infty}\left(\sum_{n=0}^{N} x^n - \sum_{n=0}^{N} x^{n+1}\right) = \lim_{N\to\infty}\left(e - x^{N+1}\right) = e.\end{aligned}$$

A similar calculation shows that $\left(\sum_{n=0}^{\infty} x^n\right)(e-x) = e$, so indeed $e - x \in G(A)$ and the inverse is given by (7.1).

If $c \in \mathbb{C}$, $|c| \geq 1$, then, by what has just been shown, $e - (1/c)x \in G(A)$, so $\phi(e - (1/c)x) = 1 - (1/c)\phi(x) \neq 0$ by Proposition 7.3, that is, $\phi(x) \neq c$. □

Corollary 7.5. *(a) $G(A)$ is open. More precisely, if $x \in G(A)$ and $\|h\| < \frac{1}{\|x^{-1}\|}$, then $x + h \in G(A)$ also.*
(b) If ϕ is a complex homomorphism, then $\phi \in A^*$ and $\|\phi\| = 1$.

Proof. (a) Write $x + h = x(e + x^{-1}h)$. Since $\|x^{-1}h\| \leq \|x^{-1}\| \, \|h\| < 1$, Theorem 7.4 shows that $e + x^{-1}h \in G(A)$. Since also $x \in G(A)$ and $G(A)$ is a group, it follows that $x + h \in G(A)$, too.

(b) The last part of Theorem 7.4 says that ϕ is bounded and $\|\phi\| \leq 1$. Since $\phi(e) = 1$ and $\|e\| = 1$, it follows that $\|\phi\| = 1$. □

Exercise 7.3. We can also run a more quantitative version of the argument from (a) to obtain the following: Inversion in Banach algebras is a continuous operation. More precisely, if $x \in G(A)$ and $\epsilon > 0$, then there exists $\delta > 0$ so that $y \in G(A)$ and $\|y^{-1} - x^{-1}\| < \epsilon$ if $\|y - x\| < \delta$. Prove this.

We now introduce the Banach algebra version of Definition 6.7.

Definition 7.6. Let $x \in A$. Then we define

$$\begin{aligned} \rho(x) &= \{z \in \mathbb{C} : x - ze \in G(A)\}, \\ \sigma(x) &= \mathbb{C} \setminus \rho(x), \\ r(x) &= \sup\{|z| : z \in \sigma(x)\}. \end{aligned}$$

We call $\rho(x)$ the *resolvent set*, $\sigma(x)$ the *spectrum*, and $r(x)$ the *spectral radius* of x. Also, $(x - ze)^{-1}$, which is defined for $z \in \rho(x)$, is called the *resolvent* of x.

Theorem 7.7. *(a) $\rho(x)$ is an open subset of $\mathbb{C}$.*
(b) The resolvent $R(z) = (x - ze)^{-1}$ admits power series representations about every point $z_0 \in \rho(x)$. More specifically, if $z_0 \in \rho(x)$, then there exists $r > 0$ so that $\{z : |z - z_0| < r\} \subset \rho(x)$ and

$$(x - ze)^{-1} = \sum_{n=0}^{\infty} (x - z_0 e)^{-n-1} (z - z_0)^n$$

for all z with $|z - z_0| < r$.

Here we define y^{-n}, for $n \geq 0$ and invertible y, as $y^{-n} = (y^{-1})^n$. More succinctly, we can say that the resolvent $R(z)$ is a *holomorphic function* (which takes values in a Banach algebra) on $\rho(x)$; we then simply define this notion by the property from Theorem 7.7(b).

Proof. (a) This is an immediate consequence of Corollary 7.5 because $\|x - ze - (x - z_0)e\| = |z - z_0|$.

(b) As in (a) and the proof of Corollary 7.5(a), we see that $B_r(z_0) \subset \rho(x)$ if we take $r = 1/\|(x-z_0e)^{-1}\|$. Moreover, we can use the Neumann series to expand $R(z)$, as follows:

$$\begin{aligned}(x - ze)^{-1} &= \left[(e - (z - z_0)(x - z_0e)^{-1})(x - z_0e)\right]^{-1}\\ &= (x - z_0e)^{-1}\left[e - (z - z_0)(x - z_0e)^{-1}\right]^{-1}\\ &= (x - z_0e)^{-1}\sum_{n=0}^{\infty}(x - z_0e)^{-n}(z - z_0)^n\\ &= \sum_{n=0}^{\infty}(x - z_0e)^{-n-1}(z - z_0)^n\end{aligned}$$

We have used the continuity of the multiplication in the last step. □

Theorem 7.8. *(a) $\sigma(x)$ is a compact, non-empty subset of $\mathbb{C}$.*
(b) $r(x) = \inf_{n\in\mathbb{N}} \|x^n\|^{1/n} = \lim_{n\to\infty} \|x^n\|^{1/n}$

The existence of the limit in part (b) is part of the statement. Note also that $\|x^n\| \le \|x\|^n$, by using property (4) from Definition 7.1 repeatedly, so we always have that $r(x) \le \|x\|$. Strict inequality is possible here.

The inconspicuous *spectral radius formula* from part (b) has a rather remarkable property: $r(x)$ is a purely algebraic quantity (to work out $r(x)$, find the biggest $|z|$ for which $x-ze$ does not have a multiplicative inverse), but nevertheless $r(x)$ is closely related to the norm on A via the spectral radius formula.

Proof. (a) We know from Theorem 7.7(a) that $\sigma(x) = \mathbb{C}\setminus\rho(x)$ is closed. Moreover, if $|z| > \|x\|$, then $x - ze = (-z)(e - (1/z)x) \in G(A)$ by Theorem 7.4, so $\sigma(x)$ is also bounded and thus a compact subset of $\mathbb{C}$.

We also obtain the representation

$$(x - ze)^{-1} = -\sum_{n=0}^{\infty} z^{-n-1}x^n \tag{7.2}$$

from Theorem 7.4; this is valid for $|z| > \|x\|$. Suppose now that we had $\sigma(x) = \emptyset$. For an arbitrary $F \in A^*$, we can introduce the function $g : \rho(x) \to \mathbb{C}$, $g(z) = F((x - ze)^{-1})$. Since we are assuming that $\sigma(x) = \emptyset$, this function is really defined on all of $\mathbb{C}$. Moreover, by using Theorem 7.7(b) and the continuity of F, we see that g has convergent power series representations about every point and thus is holomorphic (in the traditional sense). If $|z| \geq 2\|x\|$, then (7.2) yields

$$|g(z)| = \left| F\left(\sum_{n=0}^{\infty} z^{-n-1} x^n \right) \right| \leq \|F\| \sum_{n=0}^{\infty} |z|^{-n-1} \|x\|^n$$
$$\leq \frac{\|F\|}{|z|} \sum_{n=0}^{\infty} 2^{-n} = \frac{2\|F\|}{|z|}.$$

So g is a bounded entire function. By Liouville's Theorem, g must be constant. Since $g(z) \to 0$ as $|z| \to \infty$, this constant must be zero. This, however, is not possible, because $F(y) = 0$ for all $F \in A^*$ would imply that $y = 0$, by Corollary 4.2(b), but clearly the inverse $(x - ze)^{-1}$ can not be the zero element of A. The assumption that $\sigma(x) = \emptyset$ must be dropped.

(b) Let $n \in \mathbb{N}$ and let $z \in \mathbb{C}$ be such that $z^n \in \rho(x^n)$. We can write

$$x^n - z^n e = (x - ze)(z^{n-1}e + z^{n-2}x + \cdots + x^{n-1}),$$

and now multiplication from the right by $(x^n - z^n e)^{-1}$ shows that $x - ze$ has a right inverse. A similar calculation provides a left inverse also, so it follows that $z \in \rho(x)$ (we are using Exercise 7.2(d) here!). Put differently, $z^n \in \sigma(x^n)$ if $z \in \sigma(x)$. The proof of part (a) has shown that $|z| \leq \|y\|$ for all $z \in \sigma(y)$, so we now obtain that $|z^n| \leq \|x^n\|$ for all $z \in \sigma(x)$. Since the spectral radius $r(x)$ was defined as the maximum of the spectrum (we cautiously worked with the supremum in the original definition, but we now know that $\sigma(x)$ is a compact set), this says that $r(x) \leq \inf \|x^n\|^{1/n}$.

Next, consider again the function $g(z) = F((x - ze)^{-1})$, with $F \in A^*$. This is holomorphic on $\rho(x) \supset \{z \in \mathbb{C} : |z| > r(x)\}$. Furthermore, for $|z| > \|x\|$, we have the power series expansion (in z^{-1})

$$g(z) = -\sum_{n=0}^{\infty} F(x^n) \left(z^{-1}\right)^{n+1}.$$

This shows that g is holomorphic near $z = \infty$; more precisely, if we let $\zeta = 1/z$ and $h(\zeta) = g(1/\zeta)$, then h has a convergent power series expansion, $h(\zeta) = -\sum_{n=0}^{\infty} F(x^n)\zeta^{n+1}$, which is valid for small $|\zeta|$. Moreover, by our earlier remarks, h also has a holomorphic extension to the disk $\{\zeta : |\zeta| < 1/r(x)\}$ (the extension is provided by the original definition of g). A power series converges on the biggest disk to which the function can be holomorphically extended; thus the radius of convergence of the series $\sum F(x^n)\zeta^{n+1}$ is at least $1/r(x)$. In particular, if $0 < a < 1/r(x)$, then

$$F(x^n)a^n = F(a^n x^n) \to 0 \qquad (n \to \infty).$$

Since this is true for arbitrary $F \in A^*$, we have in fact shown that $a^n x^n \xrightarrow{w} 0$. Weakly convergent sequences are bounded (Exercise 4.23), so there exists $C = C(a) > 0$ so that $\|a^n x^n\| \le C$ ($n \in \mathbb{N}$). Hence

$$\|x^n\|^{1/n} \le \frac{1}{a} C^{1/n} \to \frac{1}{a},$$

and here $a < 1/r(x)$ was arbitrary and we can take the limit on any subsequence, so $r(x) \ge \limsup_{n\to\infty} \|x^n\|^{1/n}$. On the other hand, we have already proved that

$$r(x) \le \inf_{n\in\mathbb{N}} \|x^n\|^{1/n} \le \liminf_{n\to\infty} \|x^n\|^{1/n},$$

so we now obtain the full claim. □

You should now work out some spectra in concrete examples. The first example is particularly important for us, so I'll state this as a Proposition:

Proposition 7.9. *Consider the Banach algebra $A = C(K)$. Then, for $f \in C(K)$, we have that $\sigma(f) = f(K)$, where $f(K) = \{f(x) : x \in K\}$. Moreover, $r(f) = \|f\|$ for all $f \in C(K)$.*

Exercise 7.4. Prove Proposition 7.9.

Exercise 7.5. (a) Show that on $A = \ell^\infty$, we have that

$$\sigma(x) = \overline{\{x_n : n \in \mathbb{N}\}}.$$

Also, show that again $r(x) = \|x\|$ for all $x \in \ell^\infty$.
(b) Show that on $A = L^\infty(X, \mu)$, we have that

$$\sigma(f) = \{z \in \mathbb{C} : \mu(\{x \in X : |f(x) - z| < \epsilon\}) > 0 \text{ for all } \epsilon > 0\}.$$

(This set is also called the *essential range* of f; roughly speaking, it is the range of f, but we ignore what happens on null sets, in keeping with the usual philosophy. Also, it is again true that $r(f) = \|f\|$.)

Exercise 7.6. Show that on $A = B(\mathbb{C}^n)$, the spectrum $\sigma(T)$ of a matrix $T \in B(\mathbb{C}^n) = \mathbb{C}^{n\times n}$ is the set of eigenvalues of T (this was discussed earlier, in Chapter 6). Now find a matrix $T \in \mathbb{C}^{2\times 2}$ for which $r(T) < \|T\|$.

The fact that spectra are always non-empty has the following consequence:

Theorem 7.10 (Gelfand-Mazur)**.** *If A is a Banach algebra with $G(A) = A \setminus \{0\}$, then $A \cong \mathbb{C}$.*

More specifically, the claim is that there is an identification map between A and $\mathbb{C}$ (thought of as a Banach algebra, with the usual multiplication and the absolute value as the norm) that preserves the complete Banach algebra structure: There is a map $\varphi : A \to \mathbb{C}$ that is bijective (= preserves sets), a homomorphism (= preserves the algebraic structure), and an isometry (= preserves the norm).

Proof. By Theorem 7.8(a), we can pick a number $z(x) \in \sigma(x)$ for each $x \in A$. So $x - z(x)e \notin G(A)$, but the only non-invertible element of A is the zero vector, so we must have that $x = z(x)e$ (and we also learn that in fact $\sigma(x) = \{z(x)\}$). The map $\varphi : A \to \mathbb{C}$, $\varphi(x) = z(x)$ has the desired properties. □

In the last part of this chapter, we discuss the problem of how the spectrum of an element changes when we pass to a smaller Banach algebra. Let B be a Banach algebra, and let $A \subset B$ be a subalgebra. By this we mean that A with the structure inherited from B is a Banach algebra itself. We also insist that $e \in A$. Note that this latter requirement could be dropped, and in fact that would perhaps be the more common version of the definition of a subalgebra. The following Exercise discusses the difference between the two versions:

Exercise 7.7. Let B be a Banach algebra, and let $C \subset B$ be a subset that also is a Banach algebra *with unit element* with the structure (algebraic operations, norm) inherited from B. Give a (simple) example of such a situation where $e \notin C$.
Remark: This is very straightforward. Just make sure you don't get confused. C is required to have a unit (call it f, say), but what exactly is f required to do? Exercise 7.2(g) might also provide some inspiration.

If we now fix an element $x \in A$ of the smaller algebra, we can consider its spectrum with respect to both algebras. From the definition, it is clear that $\sigma_A(x) \supset \sigma_B(x)$: everything that is invertible in A remains invertible in B, but we may lose invertibility when going from B to A simply because the required inverse may no longer be part of the algebra.

Furthermore, Theorem 7.8(b) shows that $r_A(x) = r_B(x)$. More can be said about the relation between $\sigma_A(x)$ and $\sigma_B(x)$, but this requires some work. This material will be needed later, but is of a technical character and can be given a light reading at this point.

We need the notion of connected components in a topological space; actually, we only need this for the space $X = \mathbb{C}$. Recall that we call a topological space X *connected* if the only decomposition of X into two disjoint open sets is the trivial one: if $X = U \cup V$, $U \cap V = \emptyset$

and U, V are open, then $U = X$ or $V = X$. A subset $A \subset X$ is called *connected* if A with the relative topology is a connected topological space. A *connected component* is a maximal connected set. These connected components always exist and in fact every point lies in a unique connected component, and the whole space can be written as the disjoint union of its connected components.

For a detailed reading of this final section, the following topological warm-up should be helpful. You can either try to solve this directly or do some reading.

Exercise 7.8. (a) Prove these facts. More specifically, show that if $x \in X$, then there exists a unique set C_x so that $x \in C_x$, C_x is connected, and if also $x \in D$, D connected, then $D \subset C_x$. Also, show that if $x, y \in X$, then either $C_x \cap C_y = \emptyset$ or $C_x = C_y$.
(b) Call $A \subset X$ *arcwise connected* if any two points can be joined by a continuous curve: If $x, y \in A$, then there exists a continuous map $\varphi : [0,1] \to A$ with $\varphi(0) = x$, $\varphi(1) = y$. Show that an arcwise connected set is connected.
(c) Show that if $U \subset \mathbb{C}$ is open, then all connected components of U are open subsets of $\mathbb{C}$.

We are heading towards the following general result:

Theorem 7.11. *We have a representation of the following type:*

$$\sigma_A(x) = \sigma_B(x) \cup C,$$

where C is a (necessarily disjoint) union of connected components of $\rho_B(x)$ ($C = \emptyset$ is possible, of course).

This has the following consequences (whose relevance is more obvious):

Corollary 7.12. *(a) If $\rho_B(x)$ is connected, then $\sigma_A(x) = \sigma_B(x)$. In particular, this conclusion holds if $\sigma_B(x) \subset \mathbb{R}$.*
(b) If $\overset{\circ}{\sigma}_A(x) = \emptyset$, then $\sigma_A(x) = \sigma_B(x)$.

Here, $\overset{\circ}{C}$ denotes the interior of C, defined as the largest open subset of C.

To prove the Corollary (given the Theorem), note that the hypothesis that $\rho_B(x)$ is connected means that the only connected component of this set is $\rho_B(x)$ itself, but we cannot have $\sigma_A(x) = \sigma_B(x) \cup \rho_B(x)$ because $\rho_B(x)$ is unbounded (being the complement of the compact set $\sigma_B(x)$), and $\sigma_A(x)$ needs to be compact. If $\sigma_B(x)$ is a (compact!) subset of $\mathbb{R}$, then clearly its complement $\rho_B(x)$ is arcwise connected, thus connected. Compare Exercise 7.8(b).

Part (b) follows from the fact that the connected components of the open set $\rho_B(x)$ are open (Exercise 7.8(c)), so if we had $C \neq \emptyset$, then automatically $\sigma_A(x)$ would have non-empty interior.

To prove Theorem 7.11, we need the following topological fact.

Lemma 7.13. *Let $U, V \subset X$ be open sets and assume that $U \subset V$, $(\overline{U} \setminus U) \cap V = \emptyset$. Then $U = \bigcup V_\alpha$, where the V_α are connected components of V (but not necessarily all of these, of course).*

Proof. We must show that if W is a connected component of V with $W \cap U \neq \emptyset$, then $W \subset U$ (assuming this, we can then indeed write U as the union of those components of V that intersect U). So let W be such a component. From the assumption of the Lemma, we have that $W \cap (\overline{U} \setminus U) = \emptyset$. Therefore,

$$W = (W \cap U) \cup (W \cap \overline{U}^c).$$

This is a decomposition of W into two disjoint relatively (!) open subsets. Since W is connected by assumption, one of these must be all of W, and since $W \cap U \neq \emptyset$, the first set is this set, that is, $W \cap U = W$ or $W \subset U$. □

We are now ready for the

Proof of Theorem 7.11. We will verify the hypotheses of Lemma 7.13 for $U = \rho_A(x)$, $V = \rho_B(x)$. The Lemma will then show that $\rho_A(x) = \bigcup_{\alpha \in I_0} V_\alpha$, where the V_α are connected components of $\rho_B(x)$. Also, $\rho_B(x) = \bigcup_{\alpha \in I} V_\alpha$, and $I_0 \subset I$, so we indeed obtain that

$$\sigma_A(x) = \mathbb{C} \setminus \rho_A(x) = \sigma_B(x) \cup \bigcup_{\alpha \in I \setminus I_0} V_\alpha.$$

Clearly, $\rho_A(x) \subset \rho_B(x)$, so we must check that $(\overline{\rho_A(x)} \setminus \rho_A(x)) \cap \rho_B(x) = \emptyset$. Let $z \in \overline{\rho_A(x)} \setminus \rho_A(x)$. Then there are $z_n \in \rho_A(x)$, $z_n \to z$. I now claim that

$$\left\| (x - z_n e)^{-1} \right\| \to \infty \qquad (n \to \infty). \tag{7.3}$$

Suppose this were wrong. Then $|z - z_n| \, \|(x - z_n e)^{-1}\| < 1$ for some (large) n, and hence

$$(x - z_n e)^{-1}(x - ze) = e - (z - z_n)(x - z_n e)^{-1}$$

would be in $G(A)$ by Theorem 7.4, but then also $x - ze \in G(A)$, and this contradicts $z \notin \rho_A(x)$. Thus (7.3) holds. Now (7.3) also prevents $x - ze$ from being invertible in B, because inversion is a continuous operation in Banach algebras (Exercise 7.3). More precisely, if we had $x - ze \in G(B)$, then, since $x - z_n e \to x - ze$, it would follow that

$(x - z_n e)^{-1} \to (x - ze)^{-1}$, but this convergence is ruled out by (7.3). So $x - ze \notin G(B)$, or, put differently, $z \notin \rho_B(x)$. $\square$

Exercise 7.9. Show that $r(xy) = r(yx)$. *Hint:* Use the formula $(xy)^n = x(yx)^{n-1}y$.

Exercise 7.10. Prove that $\sigma(xy)$ and $\sigma(yx)$ can at most differ by the point 0. (In particular, this again implies the result from Exercise 7.9, but of course the direct proof suggested there was much easier.)
Suggested strategy: This essentially amounts to showing that $e - xy$ is invertible if and only if $e - yx$ is invertible. So assume that $e - xy \in G(A)$. Assume also that $\|x\|, \|y\| < 1$ and write $(e-xy)^{-1}$, $(e-yx)^{-1}$ as Neumann series. Use the formula from the previous problem to obtain one inverse in terms of the other. Then show that this formula actually works in complete generality, without the assumptions on x, y.

8. Commutative Banach algebras

In this chapter, we analyze *commutative* Banach algebras in greater detail. So we always assume that $xy = yx$ for all $x, y \in A$ here.

Definition 8.1. Let A be a (commutative) Banach algebra. A subset $I \subset A$ is called an *ideal* if I is a (linear) subspace and $xy \in I$ whenever $x \in I$, $y \in A$. An ideal $I \neq A$ is called *maximal* if the only ideals $J \supset I$ are $J = I$ and $J = A$.

Ideals are important for several reasons. First of all, we can take quotients with respect to ideals, and we again obtain a Banach algebra.

Theorem 8.2. *Let $I \neq A$ be a closed ideal. Then A/I is a Banach algebra.*

This needs some clarification. The quotient A/I consists of the equivalence classes $(x) = x + I = \{x + y : y \in I\}$, and we define the algebraic operations on A/I by working with representatives; the fact that I is an ideal makes sure that everything is well defined (independent of the choice of representative). Since I is in particular a closed subspace, we also have the quotient norm available, and we know from Theorem 2.18 that A/I is a Banach space with this norm. Recall that this norm was defined as

$$\|(x)\| = \inf_{y \in I} \|x + y\|.$$

Proof. From the above remarks, we already know that A/I is a Banach space and a commutative algebra with unit (e). We need to discuss conditions (3), (4) from Definition 7.1. To prove (4), let $x_1, x_2 \in A$, and let $\epsilon > 0$. We can then find $y_1, y_2 \in I$ so that $\|x_j + y_j\| < \|(x_j)\| + \epsilon$. It follows that

$$\begin{aligned} \|(x_1)(x_2)\| = \|(x_1 x_2)\| &\le \|[x_1 + y_1][x_2 + y_2]\| \\ &\le \|x_1 + y_1\| \, \|x_2 + y_2\| \le (\|(x_1)\| + \epsilon)(\|(x_2)\| + \epsilon). \end{aligned}$$

Since $\epsilon > 0$ is arbitrary here, we have that $\|(x_1)(x_2)\| \le \|(x_1)\| \, \|(x_2)\|$, as required.

Next, notice that $\|(e)\| \le \|e\| = 1$. On the other hand, for all $x \in A$, we have that $\|(x)\| = \|(x)(e)\| \le \|(x)\| \, \|(e)\|$, so $\|(e)\| \ge 1$. □

Theorem 8.3. *(a) If $I \neq A$ is an ideal, then $I \cap G(A) = \emptyset$.*
(b) The closure of an ideal is an ideal.
(c) Every maximal ideal is closed.
(d) Every ideal $I \neq A$ is contained in some maximal ideal $J \supset I$.

Proof. (a) If $x \in I \cap G(A)$, then $y = x(x^{-1}y) \in I$ for all $y \in A$, so $I = A$.

(b) The closure of a subspace is a subspace, and if $x \in \overline{I}$, $y \in A$, then there are $x_n \in I$, $x_n \to x$. Thus $x_n y \in I$ and $x_n y \to xy$ by the continuity of the multiplication, so $xy \in \overline{I}$, as required.

(c) Let I be a maximal ideal. Then, by (b), $\overline{I}$ is another ideal that contains I. Since $I \cap G(A) = \emptyset$, by (a), and since $G(A)$ is open, $\overline{I}$ still doesn't intersect $G(A)$. In particular, $\overline{I} \neq A$, so $\overline{I} = I$ because I was maximal.

(d) This follows in the usual way from Zorn's Lemma. Also as usual, we don't want to discuss the details of this argument here. □

Definition 8.4. The *spectrum* or *maximal ideal space* Δ of a commutative Banach algebra A is defined as

$$\Delta = \{\phi : A \to \mathbb{C} : \phi \text{ complex homomorphism}\}.$$

The term *maximal ideal space* is justified by parts (a) and (b) of the following result, which set up a one-to-one correspondence between complex homomorphisms and maximal ideals.

Theorem 8.5. *(a) If I is a maximal ideal, then there exists a unique $\phi \in \Delta$ with $N(\phi) = I$.*
(b) Conversely, if $\phi \in \Delta$, then $N(\phi)$ is a maximal ideal.
(c) $x \in G(A) \iff \phi(x) \neq 0$ for all $\phi \in \Delta$.
(d) $x \in G(A) \iff x$ does not belong to any ideal $I \neq A$.
(e) $z \in \sigma(x) \iff \phi(x) = z$ for some $\phi \in \Delta$.

Proof. (a) A maximal ideal is closed by Theorem 8.3(c), so the quotient A/I is a Banach algebra by Theorem 8.2. Let $x \in A$, $x \notin I$, and put $J = \{ax + y : a \in A, y \in I\}$. It's easy to check that J is an ideal, and $J \supset I$, because we can take $a = 0$. Moreover, $x = ex + 0 \in J$, but $x \notin I$, so, since I is maximal, we must have that $J = A$. In particular, $e \in J$, so there are $a \in A$, $y \in I$ so that $ax + y = e$. Thus $(a)(x) = (e)$ in A/I. Since $x \in A$ was an arbitrary vector with $x \notin I$, we have shown that every $(x) \in A/I$, $(x) \neq 0$ is invertible. By the Gelfand-Mazur Theorem, $A/I \cong \mathbb{C}$. More precisely, there exists an isometric homomorphism $f : A/I \to \mathbb{C}$. The map $A \to A/I$, $x \mapsto (x)$ also is a homomorphism (the algebraic structure on A/I is *defined* in such a way that this would be true), so the composition $\phi(x) := f((x))$ is another homomorphism: $\phi \in \Delta$. Since f is injective, its kernel consists of exactly those $x \in A$ that are sent to zero by the first homomorphism, that is, $N(\phi) = I$.

It remains to establish uniqueness. If $N(\phi) = N(\psi)$, then $x - \psi(x)e \in N(\phi)$ for all $x \in A$, so $0 = \phi(x) - \psi(x)$.

(b) Homomorphisms are continuous, so $N(\phi)$ is a closed linear subspace. If $x \in N(\phi)$, $y \in A$, then $\phi(xy) = \phi(x)\phi(y) = 0$, so $xy \in N(\phi)$ also, and $N(\phi)$ is an ideal. Since $\phi : A \to \mathbb{C}$ is a linear map to the one-dimensional space $\mathbb{C}$, we have that codim $N(\phi) = 1$, so $N(\phi)$ is already maximal as a subspace (the only strictly bigger subspace is A).

(c) $\Longrightarrow$: This was proved earlier, in Proposition 7.3.

$\Longleftarrow$: Suppose that $x \notin G(A)$. Then $I_0 = \{ax : a \in A\}$ is an ideal with $I_0 \neq A$ (because $e \notin I_0$). By Theorem 8.3(d), there exists a maximal ideal $I \supset I_0$. By part (a), there is a $\phi \in \Delta$ with $N(\phi) = I$. In particular, $\phi(x) = 0$.

(d) This follows immediately from what we have shown already, plus Theorem 8.3(d) again.

(e) We have $z \in \sigma(x)$ if and only if $x - ze \notin G(A)$, and by part (c), this holds if and only if $\phi(x - ze) = \phi(x) - z = 0$ for some $\phi \in \Delta$. $\square$

In particular, this says that a commutative Banach algebra always admits complex homomorphisms, that is, we always have $\Delta \neq \emptyset$. Indeed, notice that Theorem 8.3(d) with $I = \{0\}$ shows that there are maximal ideals, so we obtain the claim from Theorem 8.5(a). Alternatively, we could use Theorem 8.5(e) together with the fact that spectra are always non-empty (Theorem 7.8(a)). The situation can be quite different on non-commutative algebras:

Exercise 8.1. Consider the algebra $\mathbb{C}^{2\times 2} = B(\mathbb{C}^2)$ of 2×2-matrices (this becomes a Banach algebra if we fix an arbitrary norm on $\mathbb{C}^2$ and use the corresponding operator norm; however, as this is a purely algebraic exercise, the norm plays no role here). Show that there are no complex homomorphisms $\phi \not\equiv 0$ on this algebra.

Here is a rather spectacular application of the ideas developed in Theorem 8.5:

Example 8.1. Consider the Banach algebra of absolutely convergent trigonometric series:

$$A = \left\{ f(e^{ix}) = \sum_{n=-\infty}^{\infty} a_n e^{inx} : a \in \ell^1(\mathbb{Z}) \right\}$$

We have written $f(e^{ix})$ rather than $f(x)$ because it will be convenient to think of f as a function on the unit circle $S = \{z \in \mathbb{C} : |z| = 1\} = \{e^{ix} : x \in \mathbb{R}\}$. Notice that the series converges uniformly, so $A \subset C(S)$.

Exercise 8.2. Show that if $f \equiv 0$, then $a_n = 0$ for all $n \in \mathbb{Z}$.
Suggestion: Recall that $\{e^{inx}\}$ is an ONB of $L^2((-\pi, \pi), dx/(2\pi))$. Use this fact to derive a formula that recovers the a_n's from f.

The algebraic operations on A are defined pointwise; for example, $(f+g)(z) := f(z) + g(z)$. It is not entirely clear that the product of two functions from A will be in A again, but this issue will be addressed later.

Consider the map $\varphi : \ell^1 \to A$, $\varphi(a) = \sum a_n e^{inx}$. It is clear that φ is linear and surjective. Moreover, Exercise 8.2 makes sure that φ is injective. Therefore, we can define a norm on A by $\|\varphi(a)\| = \|a\|_1$. This makes A isometrically isomorphic to $\ell^1(\mathbb{Z})$ as a Banach space. We claim that these spaces are actually isometrically isomorphic as Banach algebras, where we endow ℓ^1 with the convolution product, as in Example 7.5:

$$(a * b)_n = \sum_{j=-\infty}^{\infty} a_j b_{n-j}$$

Exercise 8.3. Show that φ is a homomorphism. Since we already know that φ is linear, you must show that $\varphi(a * b) = \varphi(a)\varphi(b)$.

In particular, this does confirm that $fg \in A$ if $f, g \in A$ (the sequence corresponding to fg is $a * b$ if a and b correspond to f and g, respectively). Since $\ell^1(\mathbb{Z})$ is a Banach algebra, A is a Banach algebra also, or perhaps it would be more appropriate to say that A is another realization of the same Banach algebra.

Proposition 8.6. *Every $\phi \in \Delta$ on this Banach algebra is an evaluation: There exists a $z \in S$ so that $\phi(f) = f(z)$. Conversely, this formula defines a complex homomorphism for every $z = e^{it} \in S$.*

Exercise 8.4. Prove Proposition 8.6, by using the following strategy: Let $\phi \in \Delta$. What can you say about $|\phi(e^{ix})|$ and $|\phi(e^{-ix})|$? Conclude that $|\phi(e^{ix})| = 1$, say $\phi(e^{ix}) = e^{it}$. Now use the continuity of ϕ to prove that for an arbitrary $f \in A$, we have that $\phi(f) = f(e^{it})$.

The converse is much easier, of course.

This material leads to an amazingly elegant proof of the following result:

Theorem 8.7 (Wiener). *Consider an absolutely convergent trigonometric series: $f(e^{ix}) = \sum_{n=-\infty}^{\infty} a_n e^{inx}$, $a \in \ell^1(\mathbb{Z})$. Suppose that $f(z) \neq 0$ for all $z \in S$. Then $1/f$ also has an absolutely convergent trigonometric expansion: There exists $b \in \ell^1(\mathbb{Z})$ so that $1/f(e^{ix}) = \sum_{n=-\infty}^{\infty} b_n e^{inx}$.*

This result is interesting because it is usually very hard to tell whether the expansion coefficients ("Fourier coefficients") of a given function lie in ℓ^1.

Proof. By Proposition 8.6, the hypothesis says that $\phi(f) \neq 0$ for all $\phi \in \Delta$. By Theorem 8.5(c), $f \in G(A)$. Clearly, the inverse is given by the function $1/f$. □

We now come to the most important topic of this chapter. With each $x \in A$, we can associate a function $\widehat{x} : \Delta \to \mathbb{C}$, $\widehat{x}(\phi) = \phi(x)$. We have encountered this type of construction before (see Proposition 4.3); it will work especially well in this new context. We call $\widehat{x}$ the *Gelfand transform* of x. The *Gelfand topology* on $\Delta \subset A^*$ is defined as the relative topology that is induced by the weak-$*$ topology on A^*. By Exercise 4.10, this is also the weak topology that is generated by the maps $\{\widehat{x} : A \to \mathbb{C} : x \in A\}$. We also write $\widehat{A}$ for this collection of maps.

Here are the fundamental properties of the Gelfand transform.

Theorem 8.8. *(a) Δ with the Gelfand topology is a compact Hausdorff space.*
(b) $\widehat{A} \subset C(\Delta)$ and the Gelfand transform $\widehat{\ }: A \to C(\Delta)$ is a homomorphism between Banach algebras.
(c) $\sigma(x) = \widehat{x}(\Delta) = \{\widehat{x}(\phi) : \phi \in \Delta\}$; in particular, $\|\widehat{x}\|_\infty = r(x) \leq \|x\|$.

Note that we use the term *Gelfand transform* for the function $\widehat{x} \in C(\Delta)$, but also for the homomorphism $\widehat{\ }: A \to C(\Delta)$ that sends x to $\widehat{x}$. Recall from Proposition 7.9 that in the Banach algebra $C(\Delta)$, $\sigma(\widehat{x}) = \widehat{x}(\Delta)$, so part (c) of the Theorem really says that the Gelfand transform preserves spectra: $\sigma(\widehat{x}) = \sigma(x)$. It also preserves the algebraic structure (by part (b)) and is continuous (by part (c) again).

Proof. (a) This is very similar to the proof of the Banach-Alaoglu Theorem, so we will just provide a sketch. From that result, we know that $\Delta \subset \overline{B}_1(0) = \{F \in A^* : \|F\| \leq 1\}$ is a subset of the compact Hausdorff space $\overline{B}_1(0)$, and so it again suffices to show that Δ is closed in the weak-$*$ topology. A procedure very similar to the one used in the original proof works again: If $\psi \in \overline{B}_1(0) \setminus \Delta$, then either $\psi \equiv 0$ or there exist $x, y \in A$ so that $\epsilon := |\psi(xy) - \psi(x)\psi(y)| > 0$. Let us indicate how to finish the proof in the second case: Let

$$U = \Big\{\phi \in \overline{B}_1(0) : |\phi(xy) - \psi(xy)| < \frac{\epsilon}{3}, |\psi(x)|\,|\phi(y) - \psi(y)| < \frac{\epsilon}{3},$$
$$|\phi(y)| < |\psi(y)| + 1, (|\psi(y)| + 1)\,|\phi(x) - \psi(x)| < \frac{\epsilon}{3}\Big\}.$$

Then U is an open set in the weak-$*$ topology that contains ψ. Moreover, if $\phi \in U$, then

$$|\phi(xy) - \phi(x)\phi(y)| \geq |\psi(xy) - \psi(x)\psi(y)| - |\phi(xy) - \psi(xy)| - \\ |\phi(y)|\,|\phi(x) - \psi(x)| - |\psi(x)|\,|\phi(y) - \psi(y)| > \epsilon - \frac{\epsilon}{3} - \frac{\epsilon}{3} - \frac{\epsilon}{3} = 0,$$

so $\phi \notin \Delta$ either and indeed $\Delta \cap U = \emptyset$. We have shown that $\overline{B}_1(0) \setminus \Delta$ is open, as claimed.

(b) It is clear that $\widehat{A} \subset C(\Delta)$, from the second description of the Gelfand topology as the weakest topology that makes all maps $\widehat{x} \in \widehat{A}$ continuous. To prove that $\widehat{\ }: A \to C(\Delta)$ is a homomorphism of algebras, we compute

$$(xy)\widehat{\ }(\phi) = \phi(xy) = \phi(x)\phi(y) = \widehat{x}(\phi)\widehat{y}(\phi) = (\widehat{x}\widehat{y})\,(\phi);$$

in other words, $(xy)\widehat{\ } = \widehat{x}\widehat{y}$. Similar arguments show that $\widehat{\ }$ is also linear.

(c) This is an immediate consequence of Theorem 8.5(e). $\square$

Let us summarize this one more time and also explore the limitations of the Gelfand transform. The maximal ideal space Δ with the Gelfand topology is a compact Hausdorff space, and the Gelfand transform provides a map from the original (commutative!) Banach algebra A to $C(\Delta)$ that

- preserves the algebraic structure: it is a homomorphism;
- preserves spectra: $\sigma(\widehat{x}) = \sigma(x)$;
- is continuous: $\|\widehat{x}\| \leq \|x\|$.

However, in general, it

- does not preserve the norm: it need not be isometric; in fact, it can have a non-trivial null space;
- need not be surjective; worse still, its range $\widehat{A}$ need not be a closed subspace of $C(\Delta)$.

Another remarkable feature of the Gelfand transform is the fact that it is a purely algebraic construction: it is independent of the norm that is being used on A. Indeed, all we need to do is construct the complex homomorphisms on A and then evaluate these on x to find $\widehat{x}$. We also let the $\widehat{x}$ generate a weak topology on Δ, but again, if formulated this way, this procedure does not involve the norm on A.

We *are* using the fact that there is *some* norm on A, though, for example to make sure that Δ is a compact space in the Gelfand topology. However, the Gelfand transform does not change if we switch to a different norm on A (in many situations, there will be only one norm that makes A a Banach algebra).

The following examples illustrate the last two properties from the above list.

Example 8.2. Let A be the set of matrices of the form $T = \left(\begin{smallmatrix} a & b \\ 0 & a \end{smallmatrix}\right)$. This is a commutative Banach algebra if we use matrix multiplication and an arbitrary operator norm on A; in fact, A is a (commutative) subalgebra of $\mathbb{C}^{2\times 2} = B(\mathbb{C}^2)$.

Exercise 8.5. Find all complex homomorphisms. Then show that there are $T \in A$, $T \neq 0$ with $\phi(T) = 0$ for all $\phi \in \Delta$. In other words, $\widehat{T} = 0$, so the Gelfand transform on A is not injective.
Remark: To get this started, you could use the fact that homomorphisms are in particular linear functionals, and we know what these are on a finite-dimensional vector space.

Example 8.3. We consider again the Banach algebra of absolutely convergent trigonometric series from Example 8.1. We saw in Proposition 8.6 that as a set, Δ may be identified with the unit circle $S = \{z : |z| = 1\}$. To extract this identification from Proposition 8.6, notice also that if $z, z' \in S$, $z \neq z'$, then there will be an $f \in A$ with $f(z) \neq f(z')$. Actually, there will be a trigonometric polynomial (that is, $a_n = 0$ for all large $|n|$) with this property. So if $z \neq z'$, then the corresponding homomorphisms are also distinct.

With this identification of Δ with S, the Gelfand transform $\widehat{f}$ of an $f \in A$ is the function that sends $z \in S$ to $\phi_z(f) = f(z)$; in other words, $\widehat{f}$ is just f itself. The Gelfand topology on S is the weakest topology that makes all $\widehat{f}$ continuous. Clearly, these functions are continuous if we use the usual topology on S. Moreover, S with both topologies is a compact Hausdorff space. Now the following simple but very important Lemma shows that the Gelfand topology is just the usual topology on S.

Lemma 8.9. *Let $\mathcal{T}_1 \subset \mathcal{T}_2$ be topologies on a common space X. If X is a compact Hausdorff space with respect to both topologies, then $\mathcal{T}_1 = \mathcal{T}_2$.*

Proof. We use the fact that on a compact Hausdorff space, a subset is compact if and only if it is closed. Now let $U \in \mathcal{T}_2$. Then U^c is closed in $\mathcal{T}_2$, thus compact. But then U^c is also compact with respect to $\mathcal{T}_1$, because $\mathcal{T}_1$ is a weaker topology (there are fewer open covers to consider). Thus U^c is $\mathcal{T}_1$-closed, so $U \in \mathcal{T}_1$. $\square$

$\widehat{A} = A$ is dense in $C(S) = C(\Delta)$ because, by (a suitable version of) the Weierstraß approximation theorem, every continuous function on

S can be uniformly (that is, with respect to $\|\cdot\|_\infty$) approximated by trigonometric polynomials, and these manifestly are in A. However, $A \neq C(S)$. This is a well known fact from the theory of Fourier series. The following Exercise outlines an argument of a functional analytic flavor.

Exercise 8.6. Suppose that we had $\widehat{A} = C(S)$. First of all, use Corollary 3.3 to show that then

$$\|a\|_1 \le C\|f\|_\infty \tag{8.1}$$

for all $a \in \ell^1$ and $f(x) = \sum a_n e^{inx}$, for some $C > 0$.

However, (8.1) can be refuted by considering approximations f_N to the (discontinuous!) function $f(e^{ix}) = \chi_{(0,\pi)}(x)$. More precisely, proceed as follows: Notice that if $f = \sum a_n e^{inx}$ with $a \in \ell^1$, then the series also converges in $L^2(-\pi,\pi)$. Recall that $\{e^{inx}\}$ is an ONS (in fact, an ONB) in $L^2((-\pi,\pi), dx/(2\pi))$, so it follows that

$$a_n = \langle e^{inx}, f\rangle = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(e^{ix})e^{-inx}\, dx$$

for all $f \in C(S)$. Use this to approximately compute the $a_n^{(N)}$ for functions $f_N \in C(S)$ that satisfy $0 \le f_N \le 1$, $f_N(e^{ix}) = 1$ for $0 < x < \pi$ and $f_N(e^{ix}) = 0$ for $-\pi + 1/N < x < -1/N$. Show that $\|a^{(N)}\|_1$ can be made arbitrarily large by taking N large enough. Since $\|f_N\|_\infty = 1$, this contradicts (8.1).

Exercise 8.7. (a) Show that c with pointwise multiplication is a Banach algebra.
(b) Show that $\ell^1 \subset c$ is an ideal.
(c) Show that there is a *unique* maximal ideal $I \supset \ell^1$. Find I and also the unique $\phi \in \Delta$ with $N(\phi) = I$.

Exercise 8.8. Consider the Banach algebra ℓ^∞. Show that

$$I_n = \{x \in \ell^\infty : x_n = 0\}$$

is a maximal ideal for every $n \in \mathbb{N}$. Find the corresponding homomorphisms $\phi_n \in \Delta$ with $N(\phi_n) = I_n$. Finally, show that there must be additional complex homomorphisms (*Suggestion:* Find another ideal J that is not contained in any I_n.)

Exercise 8.9. Let A be a commutative Banach algebra. Show that the spectral radius satisfies

$$r(xy) \le r(x)r(y), \qquad r(x+y) \le r(x) + r(y)$$

for all $x, y \in A$.

Exercise 8.10. Show that the inequalities from Exercise 8.9 can fail on non-commutative Banach algebras. More specifically, show that they fail on $A = \mathbb{C}^{2\times 2}$.

Remark: Recall that on this Banach algebra, the spectrum of a matrix is the set of its eigenvalues, so $r(T)$ is the absolute value of the biggest eigenvalue of T.

9. C^*-ALGEBRAS

We are especially interested in the Banach algebra $B(H)$, and here we have an additional structure that we have not taken into account so far: we can form adjoints T^* of operators $T \in B(H)$. We now discuss such an operation in the abstract setting.

Unless stated otherwise, the algebras in this chapter are not assumed to be commutative.

Definition 9.1. Let A be a Banach algebra. A map $* : A \to A$ is called an *involution* if it has the following properties:

$$(x+y)^* = x^* + y^*, \quad (cx)^* = \overline{c}x^*, \quad (xy)^* = y^*x^*, \quad x^{**} = x$$

for all $x, y \in A$, $c \in \mathbb{C}$.

We call $x \in A$ *self-adjoint (normal)* if $x = x^*$ $(xx^* = x^*x)$.

Example 9.1. Parts (a)–(d) of Theorem 6.1 show that the motivating example "adjoint operator on $B(H)$" indeed is an involution on $B(H)$ in the sense of Definition 9.1.

Example 9.2. $f^*(x) := \overline{f(x)}$ defines an involution on $C(K)$ and also on $L^\infty(X,\mu)$. Similarly, $(x^*)_n := \overline{x_n}$ defines an involution on ℓ^∞.

Theorem 9.2. *Let A be a Banach algebra with involution, and let $x \in A$. Then:*
(a) $x + x^*$, $-i(x - x^*)$, xx^* are self-adjoint;
(b) x has a unique representation of the form $x = u + iv$ with u, v self-adjoint;
(c) $e = e^*$;
(d) If $x \in G(A)$, then also $x^* \in G(A)$ and $(x^*)^{-1} = (x^{-1})^*$;
(e) $z \in \sigma(x) \iff \overline{z} \in \sigma(x^*)$.

Proof. (a) can be checked by direct calculation; for example, $(x+x^*)^* = x^* + x^{**} = x^* + x$.

(b) We can write

$$x = \frac{1}{2}(x + x^*) + i\,\frac{-i}{2}(x - x^*),$$

and by part (a), this is a representation of the desired form. To prove uniqueness, assume that $x = u+iv = u'+iv'$, with self-adjoint elements u, u', v, v'. Then both $w := u - u'$ and $iw = i(u - u') = v - v'$ are self-adjoint, too, so $iw = (iw)^* = -iw$ and hence $w = 0$.

(c) $e^* = ee^*$, and this is self-adjoint by part (a). So $e^* = e^{**} = e$, and thus e itself is self-adjoint, too.

(d) Let $x \in G(A)$. Then we can take adjoints in $xx^{-1} = x^{-1}x = e$; by part (c), $e^* = e$, so we obtain that

$$\left(x^{-1}\right)^* x^* = x^* \left(x^{-1}\right)^* = e,$$

and this indeed says that $x^* \in G(A)$ and $(x^*)^{-1} = (x^{-1})^*$.

(e) If $z \notin \sigma(x)$, then $x - ze \in G(A)$, so $(x - ze)^* = x^* - \overline{z}e \in G(A)$ by part (d), that is, $\overline{z} \notin \sigma(x^*)$. We have established "$\Longleftarrow$", and the converse is the same statement, applied to x^* in place of x. $\square$

The involution on $B(H)$ has an additional property that does not follow from the conditions of Definition 9.1: we always have that $\|TT^*\| = \|T^*T\| = \|T\|^2$; see Theorem 6.1(f). This innocuous looking identity is so powerful and has so many interesting consequences that it deserves a special name:

Definition 9.3. Let A be a Banach algebra with involution. A is called a *C^*-algebra* if $\|xx^*\| = \|x\|^2$ for all $x \in A$ (the *C^*-property*).

From this, we automatically get analogs of the other properties from Theorem 6.1(f) also; in other words, these could have been included in the definition.

Proposition 9.4. *Let A be a C^*-algebra. Then $\|x\| = \|x^*\|$ and $\|x^*x\| = \|x\|^2$ for all $x \in A$.*

Exercise 9.1. Prove Proposition 9.4.

Example 9.3. $B(H)$, $C(K)$, $L^\infty(X, \mu)$, and ℓ^∞ with the involutions introduced above are C^*-algebras. For $B(H)$ (which again was the motivating example) this of course follows from Theorem 6.1(f), and on the other algebras, we obtain the C^*-property from an easy direct argument. For example, if $f \in C(K)$, then

$$\|ff^*\| = \max_{x\in K} |f(x)\overline{f(x)}| = \max_{x\in K} |f(x)|^2 = \left(\max_{x\in K} |f(x)|\right)^2 = \|f\|^2.$$

Example 9.4. This really is a non-example. Consider again the Banach algebra

$$A = \left\{ f(e^{ix}) = \sum_{n=-\infty}^{\infty} a_n e^{inx} : a \in \ell^1(\mathbb{Z}) \right\}$$

of absolutely convergent trigonometric series. Recall that we multiply functions from A pointwise (equivalently, we take the convolution product of the corresponding sequences from ℓ^1), and we use the norm $\|f\| = \|a\|_1$.

It is not very difficult to verify that $f^*(z) := \overline{f(z)}$ again defines an involution on A. The algebraic properties from Definition 9.1 are in fact obvious, we just need to make sure that $f^* \in A$ again, but this is easy: if $f = \sum a_n e^{inx}$, then $f^* = \sum b_n e^{inx}$, with $b_n = \overline{a_{-n}}$ (or we can rephrase and say that this last formula defines an involution on $\ell^1(\mathbb{Z})$).

Exercise 9.2. Show that this involution does *not* have the C^*-property, that is, A is not a C^*-algebra.

We can now formulate and prove the central result of this chapter.

Theorem 9.5 (Gelfand-Naimark)**.** *Let A be a commutative C^*-algebra. Then the Gelfand transform $\widehat{\ }: A \to C(\Delta)$ is an isometric $*$-isomorphism between the C^*-algebras A and $C(\Delta)$.*

We call a map $\varphi : A \to B$ between C^*-algebras an isometric $*$-isomorphism if φ is bijective, a homomorphism, an isometry, and preserves the involution: $\varphi(x^*) = (\varphi(x))^*$. In other words, such a map preserves the complete C^*-algebra structure (set, algebraic structure, norm, involution).

It now becomes clear that the Gelfand-Naimark Theorem is a very powerful structural result; it says that $C(K)$ provides a universal model for arbitrary commutative C^*-algebras. Every commutative C^*-algebra can be identified with $C(K)$; in fact, we can be more specific: K can be taken to be the maximal ideal space Δ with the Gelfand topology, and then the Gelfand transform provides an identification map.

Note also that the Gelfand transform on C^*-algebras has much better properties than on general Banach algebras; see again our discussion at the end of Chapter 8.

For the proof, we will need the following result.

Theorem 9.6 (Stone-Weierstraß)**.** *Let K be a compact Hausdorff space, and suppose that $A \subset C(K)$ has the following properties:*
(a) A is a subalgebra (possibly without unit);
(b) If $f \in A$, then $\overline{f} \in A$;
(c) A separates the points of K: if $x, y \in K$, $x \neq y$, then there is an $f \in A$ with $f(x) \neq f(y)$;
(d) For every $x \in K$, there exists an $f \in A$ with $f(x) \neq 0$.
Then $\overline{A} = C(K)$.

This closure is taken with respect to the norm topology. So we could slightly rephrase the statement as follows: if $g \in C(K)$ and $\epsilon > 0$ are given, then we can find an $f \in A$ so that $\|f - g\|_\infty < \epsilon$.

This result is a far-reaching generalization of the classical Weierstraß approximation theorem, which says that every continuous function on a compact interval $[a, b]$ can be uniformly approximated by polynomials. To obtain this as a special case of Theorem 9.6, just put $K = [a, b]$ and check that

$$A = \left\{ p(x) = \sum_{n=0}^{N} a_n x^n : a_n \in \mathbb{C}, N \in \mathbb{N}_0 \right\}$$

satisfies hypotheses (a)–(d). We don't want to prove the Stone-Weierstraß Theorem here; a proof can be found in most topology books. Or see Folland, Real Analysis, Theorem 4.51. We are now ready for the

Proof of the Gelfand-Naimark Theorem. We first claim that $\phi(u) \in \mathbb{R}$ for all $\phi \in \Delta$ if $u \in A$ is self-adjoint. To see this, write $\phi(u) = c + id$, with $c, d \in \mathbb{R}$, and put $x = u + ite$, with $t \in \mathbb{R}$. Then $\phi(x) = c + i(d + t)$ and $xx^* = u^2 + t^2 e$, so

$$c^2 + (d + t)^2 = |\phi(x)|^2 \le \|x\|^2 = \|xx^*\| \le \|u^2\| + t^2.$$

It follows that $2dt \le C$, with $C := \|u^2\| - d^2 - c^2$, and this holds for arbitrary $t \in \mathbb{R}$. Clearly, this is only possible if $d = 0$, so $\phi(u) = c \in \mathbb{R}$, as asserted.

It now follows that the Gelfand transform preserves the involution: if $x \in A$, then we can write $x = u + iv$ with u, v self-adjoint, and it follows that

$$\phi(x^*) = \phi(u - iv) = \phi(u) - i\phi(v) = \overline{\phi(u) + i\phi(v)} = \overline{\phi(x)}.$$

Recall that the involution on $C(\Delta)$ was defined as the pointwise complex conjugate, so, since $\phi \in \Delta$ is arbitrary here, this calculation indeed says that $\widehat{x^*} = \overline{\widehat{x}} = (\widehat{x})^*$.

We also learn from this that $\widehat{A} \subset C(\Delta)$ satisfies assumption (b) from the Stone-Weierstraß Theorem. It is straightforward to establish the other conditions, too; for example, to verify (c), just note that if $\phi, \psi \in \Delta$, $\phi \ne \psi$, then $\phi(x) \ne \psi(x)$ for some $x \in A$, so $\widehat{x}(\phi) \ne \widehat{x}(\psi)$. So Theorem 9.6 shows that $\overline{\widehat{A}} = C(\Delta)$.

As the next step, we want to show that the Gelfand transform is isometric. Let $x \in A$, and put $y = xx^*$. Then y is self-adjoint, and therefore the C^*-property gives that $\|y^2\| = \|y\|^2$, $\|y^4\| = \|y^2 y^2\| = \|y^2\|^2 = \|y\|^4$, and so forth. The general formula is $\|y^n\| = \|y\|^n$, if $n = 2^k$ is a power of 2. Now we can compute the spectral radius by using the formula from Theorem 7.8(b) along this subsequence. It follows that $r(y) = \lim_{n\to\infty} \|y^n\|^{1/n} = \|y\|$. Since $\|\widehat{y}\| = r(y)$ by Theorem 8.8(c), this shows that $\|\widehat{y}\| = \|y\|$ for y of the form $y = xx^*$. We can now use the C^*-property on both algebras $C(\Delta)$ and A to conclude that also $\|\widehat{x}\| = \|x\|$ for arbitrary $x \in A$.

So the Gelfand transform is an isometry, and this implies that this map is injective (obvious, because only the zero vector can get mapped to zero) and its range $\widehat{A}$ is a *closed* subspace of $C(\Delta)$ (not completely obvious, but we have encountered this argument before; see the proof of Proposition 4.3). We proved earlier that $\overline{\widehat{A}} = C(\Delta)$, so it now follows that $\widehat{A} = C(\Delta)$. We have established all the properties of the Gelfand transform that were stated in Theorem 9.5. □

We now discuss in detail the Gelfand transform for the three commutative C^*-algebras $C(K)$, c, $L^\infty(0,1)$.

Example 9.5. Let K be a compact Hausdorff space and consider the C^*-algebra $A = C(K)$. We know from the Gelfand-Naimark Theorem that $C(K) \cong C(\Delta)$, but we would like to explicitly identify Δ and the Gelfand transforms of functions $f \in C(K)$.

We will need the following tool:

Lemma 9.7 (Urysohn)**.** *Let K be a compact Hausdorff space. If A, B are disjoint closed subsets of K, then there exists $f \in C(K)$ with $0 \le f \le 1$, $f = 0$ on A and $f = 1$ on B.*

See, for example, Folland, Real Analysis, Lemma 4.15 (plus Proposition 4.25) for a proof.

It is clear that the point evaluations $\phi_x(f) = f(x)$ are complex homomorphisms for all $x \in K$. So we obtain a map $\Psi : K \to \Delta$, $\Psi(x) = \phi_x$. Urysohn's Lemma shows that Ψ is injective: if $x, y \in K$, $x \ne y$, then there exists $f \in C(K)$ with $f(x) \ne f(y)$ (just take $A = \{x\}$, $B = \{y\}$ in Lemma 9.7). So $\phi_x(f) \ne \phi_y(f)$ and thus $\phi_x \ne \phi_y$.

I now claim that Ψ is also surjective. If this were wrong, then there would be a $\phi \in \Delta$, $\phi \notin \{\phi_x : x \in K\}$. Let $I = N(\phi)$, $I_x = N(\phi_x) = \{f \in C(K) : f(x) = 0\}$ be the corresponding maximal ideals. By assumption and (the uniqueness part of) Theorem 8.5(a), $I \ne I_x$ for all $x \in K$. Since I is also maximal, this implies that I is not contained in any I_x. So for every $x \in K$, there exists an $f_x \in I$ with $f_x(x) \ne 0$. Since the f_x are continuous, we can find neighborhoods U_x of x so that $f_x(y) \ne 0$ for all $y \in U_x$. By compactness, K is covered by finitely many of these, say $K = \bigcup_{j=1}^N U_{x_j}$. Now let $g = \sum_{j=1}^N f_{x_j}\overline{f_{x_j}}$. Then $g \in I$ and $g > 0$ on K (because on U_{x_j}, the jth summand is definitely positive), so g is invertible in $C(K)$ (with inverse $1/g$). This is a contradiction because the ideal $I \ne C(K)$ cannot contain invertible elements; see Theorem 8.3(a).

We conclude that $\Delta = \{\phi_x : x \in K\}$. This identifies Δ as a set with K. Moreover, $\widehat{f}(\phi_x) = \phi_x(f) = f(x)$, so if we use this identification, then the Gelfand transform of a function $f \in C(K)$ is just f itself.

We now want to show that the identification map Ψ is a homeomorphism, so in fact Δ (with the Gelfand topology) can be identified with K *as a topological space.* We introduce some notation: write $\mathcal{T}_G$ for the Gelfand topology on Δ, and let $\mathcal{T}_K$ be the given topology on K, but moved over to Δ. More precisely, $\mathcal{T}_K = \{\Psi(U) : U \subset K \text{ open }\}$. Since Ψ is a bijection, it preserves the set operations and thus $\mathcal{T}_K$ indeed is a topology.

Notice that every $\widehat{f} : \Delta \to \mathbb{C}$ is continuous if we use the topology $\mathcal{T}_K$ on Δ. This is almost a tautology because $\mathcal{T}_K$ is essentially the original topology and $\widehat{f}$ is essentially f, and these were continuous functions to start with. For a more formal verification, notice that $\widehat{f} = f \circ \Psi^{-1}$, so if $V \subset \mathbb{C}$ is open, then $\widehat{f}^{-1}(V) = \Psi(f^{-1}(V))$, which is in $\mathcal{T}_K$.

So $\mathcal{T}_K$ is a topology that makes all $\widehat{f}$ continuous. This implies that $\mathcal{T}_G \subset \mathcal{T}_K$, because $\mathcal{T}_G$ can be defined as the weakest such topology. Moreover, Δ is a compact Hausdorff space with respect to both topologies. This follows from Theorem 8.8(a) (for $\mathcal{T}_G$) and the fact that by construction of $\mathcal{T}_K$, $(\Delta, \mathcal{T}_K)$ is homeomorphic to K. Lemma 8.9 now shows that $\mathcal{T}_G = \mathcal{T}_K$. We summarize:

Theorem 9.8. *Let K be a compact Hausdorff space. Then the maximal ideal space Δ of the C^*-algebra $C(K)$ is homeomorphic to K. A homeomorphism between these spaces is given by $\Psi : K \to \Delta$, $\Psi(x) = \phi_x$, $\phi_x(f) = f(x)$. Moreover, if Δ is identified in this way with K, then the Gelfand transform of a function $f \in C(K)$ is just f itself.*

At least with hindsight, this does not come as a big surprise. The Gelfand transform gives a representation of a commutative C^*-algebra A as continuous functions on a compact Hausdorff space (namely, Δ), but if the algebra is already given in this form, there is no work left to be done, and indeed the Gelfand transform does not do anything (except change names) on $C(K)$. From that point of view, Theorem 9.8 seems somewhat disappointing, but we can in fact draw interesting conclusions:

Theorem 9.9. *Let K and L be compact Hausdorff spaces. Then K is homeomorphic to L if and only if the algebras $C(K)$ and $C(L)$ are (algebraically!) isomorphic.*

In this case, $C(K)$ and $C(L)$ are in fact isometrically $*$-isomorphic as C^*-algebras.

Here, we say that A and B are algebraically isomorphic if there exists a bijective homomorphism (in other words, an isomorphism) $\varphi : A \to B$. We do *not* require φ to be isometric or preserve the conjugation.

Proof. Suppose that $C(K)$ and $C(L)$ are isomorphic as algebras. By Theorem 9.8, $K \cong \Delta_K$, $L \cong \Delta_L$, but the construction of Δ and its Gelfand topology only uses the algebraic structure (we already discussed this feature of the Gelfand transform in Chapter 8), so $\Delta_K \cong \Delta_L$. Or, to spell this out somewhat more explicitly, if $\varphi : C(K) \to C(L)$ is an algebraic isomorphism, then $\phi_L \mapsto \phi_K = \phi_L \circ \varphi$ defines a homeomorphism from Δ_L onto Δ_K.

Exercise 9.3. Prove the converse statement. Actually, prove right away the stronger version that $C(K)$ and $C(L)$ are isometrically $*$-isomorphic if $K \cong L$. Also, if the above sketch doesn't convince you, try to write this down in greater detail. More specifically, give a more detailed argument that shows that the map defined at the end of the proof indeed is a homeomorphism.

□

Example 9.6. Our next example is $A = c$. This is a C^*-algebra with the conjugation $(x^*)_n = \overline{x_n}$; in fact, c is a subalgebra of the C^*-algebra ℓ^∞. We want to discuss its Gelfand representation $c \cong C(\Delta)$. We start out by finding Δ. I claim that we can identify Δ with $\mathbb{N}_\infty \equiv \mathbb{N} \cup \{\infty\}$ (this is just $\mathbb{N}$ with an additional point, which we choose to call "∞"). More precisely, $n \in \mathbb{N}$ corresponds to the complex homomorphism $\phi_n(x) = x_n$, and $\phi_\infty(x) = \lim_{n\to\infty} x_n$. It's easy to check that these ϕ's are indeed complex homomorphisms. Moreover, these are in fact *all* homomorphisms. This could be seen as in Example 9.5, but we can also just recall that the dual space c^* can be identified with $\ell^1(\mathbb{N}_\infty)$: we associate with $y \in \ell^1(\mathbb{N}_\infty)$ the functional

$$F_y(x) = \sum_{n=1}^{\infty} y_n x_n + y_\infty \cdot \lim_{n\to\infty} x_n.$$

See Example 4.4; we called the additional point 0 there (rather than ∞), but that of course is irrelevant.

Exercise 9.4. Show that F_y is a homomorphism precisely if $y = e_n$ or $y = e_\infty$.

With this identification of Δ with $\mathbb{N}_\infty$, the Gelfand transform of an $x \in c$ becomes the function $\widehat{x}(n) = \phi_n(x) = x_n$, $\widehat{x}(\infty) = \lim x_n$. So $\widehat{x}$ is just the sequence x_n itself, with the limit added as the value at the additional point ∞.

Now what is the Gelfand topology on $\mathbb{N}_\infty$? First of all, all subsets of $\mathbb{N}$ are open. To see this, just note that (with $e = (1, 1, 1, \ldots)$)

$$\{m\} = \{n \in \mathbb{N}_\infty : |\widehat{e}(n) - \widehat{e_m}(n)| < 1\} = \widehat{e - e_m}^{-1}(\{z : |z| < 1\}),$$

so this is indeed an open set for all $m \in \mathbb{N}$. Similarly, the sets $\{n \in \mathbb{N} : n \geq k\} \cup \{\infty\}$ are open for all $k \in \mathbb{N}$ because they are again inverse images of open sets $U \subset \mathbb{C}$ under suitable functions $\widehat{x}$. For example, we can take $U = \{|z| < 1\}$ (again) and $x_n = 1$ for $n < k$ and $x_n = 0$ for $n \geq k$.

By combining these observations, we see that a subset $U \subset \mathbb{N}_\infty$ is open in the Gelfand topology if:

- $\infty \notin U$ or
- $U \supset \{n : n \geq k\} \cup \{\infty\}$ for some $k \in \mathbb{N}$

This actually gives a complete list of the open sets. We can prove this remark as follows: First of all, the collection of sets U described above clearly defines a topology on $\mathbb{N}_\infty$. It now suffices to show that every $\widehat{x} : \mathbb{N}_\infty \to \mathbb{C}$ is continuous with respect to this topology, because the Gelfand topology was defined as the weakest topology with this property. Continuity of $\widehat{x}$ at $n \in \mathbb{N}$ is obvious because $\{n\}$ is a neighborhood of n. To check continuity at ∞, let $\epsilon > 0$ be given. Since $\widehat{x}(\infty) = \lim_{n\to\infty} \widehat{x}(n)$, there exists $k \in \mathbb{N}$ so that

$$|\widehat{x}(n) - \widehat{x}(\infty)| < \epsilon \qquad \text{for } n \geq k.$$

Since $U = \{n : n \geq k\} \cup \{\infty\}$ is a neighborhood of ∞, this verifies that $\widehat{x}$ is continuous at ∞ also.

This topology $\mathcal{T}_G$ is a familiar object: the space $(\mathbb{N}_\infty, \mathcal{T}_G)$ is called the 1-*point compactification* of $\mathbb{N}$; please refer to a topology book for further information. Here, the compactness of $(\mathbb{N}_\infty, \mathcal{T}_G)$ also follows from Theorem 8.8(a). In the case at hand, $\mathcal{T}_G$ also has the following characterization:

Exercise 9.5. Show that $\mathcal{T}_G$ is the only topology on $\mathbb{N}_\infty$ that induces the given topology on $\mathbb{N}$ (all sets open) and makes $\mathbb{N}_\infty$ a compact space.

We summarize:

Theorem 9.10. *The maximal ideal space Δ of c is homeomorphic to the* 1*-point compactification $\mathbb{N}_\infty$ of $\mathbb{N}$. The Gelfand transform of an $x \in c$ is just the original sequence, supplemented by its limit: $\widehat{x}(n) = x_n$, $\widehat{x}(\infty) = \lim x_n$.*

Example 9.7. In the previous two examples, the final results could have been guessed at the very beginning: it was not very hard to realize the given C^*-algebra as continuous functions on a compact Hausdorff space.

Matters are very different for $A = L^\infty(0,1)$, which is our final example. Neither Δ as a set nor its Gelfand topology are directly accessible, but we will obtain useful information anyway. It will turn out that the topological space $(\Delta, \mathcal{T}_G)$ has rather exotic properties.

We introduce a measure on Δ as follows: Consider the functional $C(\Delta) \to \mathbb{C}$, $\widehat{f} \mapsto \int_0^1 f(x)\, dx$. This is well defined because every continuous function on Δ is the Gelfand transform of a unique element of $L^\infty(0,1)$, by the Gelfand-Naimark Theorem. Moreover, the functional is also linear and positive: if $\widehat{f} \ge 0$, then $f \ge 0$ almost everywhere, because the Gelfand transform preserves spectra, and on $C(\Delta)$ and L^∞, these are given by the range and essential range of the function, respectively (see Proposition 7.9 and Exercise 7.5(b)). Therefore, $\int_0^1 f\, dx \ge 0$ if $\widehat{f} \ge 0$. The Riesz Representation Theorem now shows that there is a unique regular positive Borel measure $\mu \in \mathcal{M}(\Delta)$ so that

$$\int_0^1 f(x)\, dx = \int_\Delta \widehat{f}(\phi)\, d\mu(\phi)$$

for all $f \in L^\infty(0,1)$. See Folland, Real Analysis, Theorem 7.2 (and Proposition 7.5 for the regularity). We can think of μ as Lebesgue measure on $(0,1)$, moved over to Δ. Notice also that $\widehat{1} = 1$, so $\mu(\Delta) = \int_0^1 dx = 1$.

We will now use μ as our main tool to establish the following properties of Δ and the Gelfand topology. Taken together, these are rather strange.

Theorem 9.11. *(a) If $V \subset \Delta$, $V \ne \emptyset$ is open, then $\mu(V) > 0$.*
(b) If $g : \Delta \to \mathbb{C}$ is a bounded, (Borel) measurable function, then there exists an $\widehat{f} \in C(\Delta)$ so that $g = \widehat{f}$ μ-almost everywhere.
(c) If $V \subset \Delta$ is open, then $\overline{V}$ is also open.
(d) If $E \subset \Delta$ is a Borel set, then $\mu(\mathring{E}) = \mu(E) = \mu(\overline{E})$.
(e) Δ does not have isolated points, that is, $\{\phi\}$ is not open for any $\phi \in \Delta$.
(f) Δ does not have non-trivial convergent sequences: If $\phi_n, \phi \in \Delta$, $\phi_n \to \phi$, then $\phi_n = \phi$ for all large n.

Some comments are in order. Parts (a) and (b) imply that $L^\infty(\Delta, \mu) = C(\Delta)$: every bounded measurable function has exactly one continuous representative.

The property stated in part (c) is sometimes referred to by saying that Δ is *extremally disconnected.* Part (c) in particular implies that Δ is *totally disconnected:* the only connected subsets of Δ are the single points.

Exercise 9.6. Prove this fact. In fact, please prove the corresponding general statement: If X is a topological Hausdorff space in which the closure of every open set is open and $M \subset X$ has more than one point, then there are disjoint open sets U, V that both intersect M with $M \subset U \cup V$.

So far, none of this is particularly outlandish; indeed, discrete topological spaces such as $\mathbb{N}$ or finite collections of points (all subsets are open) have all these properties. However, part (e) says that Δ is decidedly not of this type. We must give up all attempts at visualizing Δ and admit that Δ is such a complicated space that no easy intuition will do justice to it. Note also that some of the above properties (for example, (b), (c), and (d)) seem to suggest that Δ might have many open subsets, but we also know that Δ is compact, and that works in the other direction.

Proof. (a) Let $V \subset \Delta$ be a non-empty open set. Pick $\phi \in V$. By Urysohn's Lemma, there exists $\widehat{f} \in C(\Delta)$ with $0 \leq \widehat{f} \leq 1$, $\widehat{f}(\phi) = 1$, and $\widehat{f} = 0$ on V^c. Again, since the Gelfand transform preserves spectra, we then also have that $f \geq 0$, but f is not equal to zero (Lebesgue) almost everywhere. Thus

$$0 < \int_0^1 f(x)\, dx = \int_\Delta \widehat{f}(\phi)\, d\mu(\phi) = \int_V \widehat{f}(\phi)\, d\mu(\phi),$$

and we can conclude that $\mu(V) > 0$.

(b) Let $g : \Delta \to \mathbb{C}$ be a Borel function with $|g(\phi)| \leq M$. We now use the fact that continuous functions are dense in L^p spaces ($p < \infty$) if (like here) the underlying measure is a regular Borel measure on a compact space. See Folland, Real Analysis, Proposition 7.9 for a slightly more general version of this result.

In particular, we can find $\widehat{f}_n \in C(\Delta)$ so that $\|\widehat{f}_n - g\|_2 \to 0$. In fact, we may assume that $|\widehat{f}_n| \leq M$ also.

Exercise 9.7. Prove this remark. *Suggestion:* If $|\widehat{f}| > M$ at certain points, we could just redefine $\widehat{f}$ on this set so that the new function is bounded by M, and we would in fact obtain a better approximation to g. However, we also need to make sure that the new function is still continuous. Use Urysohn's Lemma to give a careful version of this argument.

By the basic properties of the Gelfand transform, we now obtain that

$$\begin{aligned}\int_0^1 |f_m(x) - f_n(x)|^2\, dx &= \int_0^1 (\overline{f}_m(x) - \overline{f}_n(x))(f_m(x) - f_n(x))\, dx\\ &= \int_\Delta \left((\overline{f}_m - \overline{f}_n)(f_m - f_n)\right)\widehat{}\, d\mu\\ &= \int_\Delta \left|\widehat{f}_m(x) - \widehat{f}_n(x)\right|^2\, d\mu(x) \to 0 \quad (m,n\to\infty).\end{aligned}$$

So $f := \lim_{n\to\infty} f_n$ exists in $L^2(0,1)$. On a suitable subsequence, we can obtain $f(x)$ as a pointwise limit. This shows that $|f| \le M$ almost everywhere, so $f \in L^\infty(0,1)$. By the same calculation as above, we now see that

$$\int_\Delta \left|\widehat{f}_n(x) - \widehat{f}(x)\right|^2\, d\mu(x) = \int_0^1 |f_n(x) - f(x)|^2\, dx \to 0,$$

that is, $\widehat{f}_n \to \widehat{f}$ in $L^2(\Delta,\mu)$. On the other hand, $\widehat{f}_n \to g$ in this space by construction of the $\widehat{f}_n$, so $g = \widehat{f}$ in $L^2(\Delta,\mu)$, that is, almost everywhere with respect to μ, and $\widehat{f} \in C(\Delta)$, as desired.

(c) $g = \chi_{V^c}$ is a bounded Borel function because the only preimages that can occur here are $\emptyset$, V, V^c, Δ. By part (b), there exists $\widehat{f} \in C(\Delta)$ so that $g = \widehat{f}$ μ-almost everywhere. Now $\widehat{f}^{-1}(\mathbb{C}\setminus\{0,1\})$ is an open set of μ measure zero. By part (a), the set is actually empty, and thus $\widehat{f}$ only takes the values 0 and 1. This argument also shows that the sets $V \cap \widehat{f}^{-1}(\mathbb{C}\setminus\{0\})$ and $\overline{V}^c \cap \widehat{f}^{-1}(\mathbb{C}\setminus\{1\})$ are empty. Put differently, we have that $\widehat{f} = 0$ on V and $\widehat{f} = 1$ on $\overline{V}^c$. Therefore,

$$V \subset \widehat{f}^{-1}(\{0\}) \subset \overline{V}.$$

Now $\widehat{f}^{-1}(\{0\})$ is also closed (it is the preimage of a closed set), and since $\overline{V}$ is the smallest closed set that contains V, we must have that $\widehat{f}^{-1}(\{0\}) = \overline{V}$. We can also obtain this set as $\widehat{f}^{-1}(\mathbb{C}\setminus\{1\})$, which is open, so indeed $\overline{V}$ is an open set.

(d) First of all, let $V \subset \Delta$ be open. Consider again the function $g = \chi_{V^c}$ and its continuous representative $\widehat{f}$ from the proof of part (c). We saw above that $\widehat{f} = 0$ exactly on $\overline{V}$. On the other hand, $g = 0$ on V, and since $g = \widehat{f}$ almost everywhere, this implies that $\mu(V) = \mu(\overline{V})$. By passing to the complements, we also obtain from this that $\mu(\mathring{A}) = \mu(A)$ if $A \subset \Delta$ is closed.

If $E \subset \Delta$ is an arbitrary Borel set and $\epsilon > 0$ is given, we can use the regularity of μ to find a compact set $K \subset E$ and an open set $V \supset E$

so that $\mu(V) < \mu(K) + \epsilon$. It then follows that

$$\mu(\overline{E}) \le \mu(\overline{V}) = \mu(V) < \mu(K) + \epsilon = \mu(\overset{\circ}{K}) + \epsilon < \mu(\overset{\circ}{E}) + \epsilon.$$

Now $\epsilon > 0$ was arbitrary, so $\mu(\overline{E}) \le \mu(\overset{\circ}{E})$. Since clearly $\mu(\overset{\circ}{E}) \le \mu(E) \le \mu(\overline{E})$, we obtain the claim.

(e) Suppose that $\{\phi_0\}$ were an open set. Since points in Hausdorff spaces are always closed, the function $\chi_{\{\phi_0\}}$ would then be continuous and thus be equal to $\widehat{f}$ for some $f \in L^\infty(0,1)$. We can now again use the fact that the Gelfand transform preserves spectra to deduce that f itself is the characteristic function of some measurable set $M \subset (0,1)$, $|M| > 0$: $f = \chi_M$ (this follows because the essential range of f has to be $\{0,1\}$). Pick a subset $M' \subset M$ so that both M' and $M \setminus M'$ have positive Lebesgue measure.

Exercise 9.8. Prove the existence of such a set M'. Does a corresponding result hold on arbitrary measure spaces (do positive measure sets always have subsets of strictly smaller positive measure)?

Let $g = \chi_{M'}$. Then clearly $fg = g$, so $\widehat{f}\widehat{g} = \widehat{g}$. Since $\widehat{f}(\phi) = 0$ for $\phi \ne \phi_0$, this says that $\widehat{g} = c\widehat{f}$ for some $c \in \mathbb{C}$. On the other hand, it is not true that $g = cf$ almost everywhere, so we have reached a contradiction. We have to admit that $\{\phi_0\}$ is not open.

(f) Let $\phi_n \to \phi$ be a convergent sequence, and assume that ϕ_n is not eventually constant. By passing to a subsequence, we may then in fact assume that $\phi_n \ne \phi$ for all $n \in \mathbb{N}$. Pick disjoint neighborhoods U_1 and V_1 of ϕ_1 and ϕ, respectively. Since $\phi_n \to \phi$, we can find an index n_2 so that $\phi_{n_2} \in V_1$. Now pick disjoint neighborhoods U_2' and V_2' of ϕ_{n_2} and ϕ, respectively, and put $U_2 = U_2' \cap V_1$, $V_2 = V_2' \cap V_1$. These are still (possibly smaller) neighborhoods of the same points.

We can continue this procedure. We obtain pairwise disjoint neighborhoods $U_1, U_2, U_3, \ldots$ of the members of the subsequence $\phi_1, \phi_{n_2}, \phi_{n_3}, \ldots$. Since all the U_j's are in particular open, the formula

$$g(\phi) = \begin{cases} 1 & \phi \in \bigcup_{j\in\mathbb{N}} U_{2j-1} \\ -1 & \phi \in \bigcup_{j\in\mathbb{N}} U_{2j} \\ 0 & \text{otherwise} \end{cases}$$

defines a (bounded) Borel function g. By part (b), $g = \widehat{f}$ almost everywhere for some $\widehat{f} \in C(\Delta)$. We observe that we also must have that $\widehat{f}(\phi_{n_{2j-1}}) = 1$, $\widehat{f}(\phi_{n_{2j}}) = -1$, because if $\widehat{f}$ took a different value at one of these points, then $\widehat{f}$ and g would differ on an open set, and this has positive measure by (a).

Exercise 9.9. Let $f : X \to Y$ be a continuous function between topological spaces. Show that f is also *sequentially continuous,* that is, if $x_n \to x$, then $f(x_n) \to f(x)$.

From this Exercise, we obtain that $\widehat{f}(\phi_n) \to \widehat{f}(\phi)$, but clearly this is not possible if these values alternate between 1 and -1. $\square$

We now return to the general theory of C^*-algebras.

Theorem 9.12. *Suppose that A is a commutative C^*-algebra that is generated by one element $x \in A$. Then $\Delta \cong \sigma(x)$.*

If A is a (not necessarily commutative) C^*-algebra and $C \subset A$, then we define the *C^*-algebra generated by C* to be the smallest C^*-subalgebra $B \subset A$ that contains C. It is very important to recall here that we are using the convention that subalgebras always contain the original unit $e \in A$. The following Exercise clarifies basic aspects of this definition:

Exercise 9.10. (a) Show that there always exists such a C^*-algebra $B \subset A$ by defining B to be the intersection of all C^*-algebras B' with $e \in B'$ and $C \subset B' \subset A$.
(b) Prove that B has the following somewhat more explicit alternative description:

$$B = \overline{\{p(b_1, \ldots, b_M, b_1^*, \ldots, b_N^*) : p \text{ polynomial }, b_j \in C\}}$$

More precisely, the p's are polynomials in *non-commuting* variables; these are, as usual, linear combinations of products of powers of the variables, but the order of the variables matters, and we need to work with all possible arrangements.

Back to the case under consideration: The hypothesis of Theorem 9.12 means that the only C^*-algebra $B \subset A$ with $e, x \in B$ is $B = A$. Equivalently, the polynomials $p(x, x^*) = \sum_{j,k=0}^{N} c_{jk} x^j (x^*)^k$ are dense in A; notice also that we don't need to insist on non-commuting variables in p here because A is commutative.

The conclusion of the Theorem states that Δ and $\sigma(x)$ (with the relative topology coming from $\mathbb{C}$) are homeomorphic.

Proof of Theorem 9.12. The Gelfand transform of x provides the homeomorphism we are looking for: $\widehat{x} : \Delta \to \sigma(x)$ is continuous and onto. If $\widehat{x}(\phi_1) = \widehat{x}(\phi_2)$ or, equivalently, $\phi_1(x) = \phi_2(x)$, then also

$$\phi_1(x^*) = \overline{\phi_1(x)} = \overline{\phi_2(x)} = \phi_2(x^*),$$

and thus $\phi_1(p) = \phi_2(p)$ for all polynomials in x, x^*. Since these are dense in A by assumption and ϕ_1, ϕ_2 are continuous, we conclude that $\phi_1(y) = \phi_2(y)$ for all $y \in A$. So $\widehat{x}$ is also injective.

Summing up: $\widehat{x} : \Delta \to \sigma(x)$ is a continuous bijection between compact Hausdorff spaces. In this situation, the inverse is automatically continuous also, so we have our homeomorphism. To prove this last remark, we can argue as in Lemma 8.9 (or we could in fact use this result itself): Suppose $A \subset \Delta$ is closed. Then A is compact, so $\widehat{x}(A) \subset \sigma(x)$ is compact, thus closed. We have shown that the inverse image of a closed set under $\widehat{x}^{-1}$ is closed, which is one of the characterizations of continuity. □

Exercise 9.11. (a) Let $B \subset A$ be the C^*-algebra that is generated by $C \subset A$. Show that B is commutative if and only if

$$xy = yx, \quad xy^* = y^*x$$

for all $x, y \in C$.
(b) Show that the C^*-algebra generated by x is commutative if and only if x is normal.

Theorem 9.12 in particular shows that $A \cong C(\sigma(x))$ if the commutative C^*-algebra is generated by a single element. We can be a little more specific here:

Theorem 9.13. *Suppose that the commutative C^*-algebra A is generated by the single element $x \in A$. Then there exists a unique isometric $*$-isomorphism $\Psi : C(\sigma(x)) \to A$ with $\Psi(\mathrm{id}) = x$.*

Here, id refers to the function $\mathrm{id}(z) = z$ ("identity").

Proof. Uniqueness is clear because x generates the algebra, so Ψ^{-1} is determined as soon as we know $\Psi^{-1}(x)$. To prove existence, we can simply define Ψ^{-1} as the Gelfand transform, where we also identify Δ with $\sigma(x)$, as in Theorem 9.12. More precisely, let $\Psi^{-1}(y) = \widehat{y} \circ \widehat{x}^{-1}$. □

Exercise 9.12. If you have doubts about this definition of Ψ^{-1}, the following should be helpful: Let $\varphi : K \to L$ be a homeomorphism between compact Hausdorff spaces. Show that then $\Phi : C(L) \to C(K)$, $\Phi(f) = f \circ \varphi$ is an isometric $*$-isomorphism between C^*-algebras. ("Change of variables on K preserves the C^*-algebra structure of $C(K)$.")

We will use Theorem 9.13 to define $f(x) := \Psi(f)$, for $f \in C(\sigma(x))$ and $x \in A$ as above. We interpret $f(x) \in A$ as "f, applied to x", as is already suggested by the notation. There is some logic to this terminology; indeed, if we move things over to the realization $C(\sigma(x))$

of A, then f is applied to the variable (which corresponds to x) in a very literal sense.

So we can talk about continuous functions of elements of C^*-algebras, at least in certain situations. We have just made our first acquaintance with the *functional calculus.*

It may appear that the previous results are rather limited in scope because we specifically seem to need commutative C^*-algebras that are generated by a single element. That, however, is not the case because we can often use these tools on smaller subalgebras of a given C^*-algebra. Here are some illustrations of this technique.

Definition 9.14. Let A be a C^*-algebra. An element $x \in A$ is called *positive* (notation: $x \geq 0$) if $x = x^*$ and $\sigma(x) \subset [0, \infty)$.

Theorem 9.15. *Let A be a (not necessarily commutative) C^*-algebra.*
(a) If $x = x^*$, then $\sigma(x) \subset \mathbb{R}$.
(b) If x is normal, then $r(x) = \|x\|$.
(c) If $x, y \geq 0$, then $x + y \geq 0$.
(d) $xx^* \geq 0$ for all $x \in A$.

Proof. (a) Consider the C^*-algebra $B \subset A$ that is generated by x. Since x is normal (even self-adjoint), B is commutative by Exercise 9.11(b). So the Gelfand theory applies to B. In particular, $\sigma_B(x) = \{\phi(x) : \phi \in \Delta_B\}$, and this is a subset of $\mathbb{R}$, because $\overline{\phi(x)} = \phi(x^*) = \phi(x)$. Since $\sigma_A(x) \subset \sigma_B(x)$, this gives the claim.

(b) Consider again the commutative C^*-algebra $B \subset A$ that is generated by x. By the Gelfand theory (on B), $r_B(x) = \|x\|$, but, as observed earlier, in Chapter 7, the spectral radius formula shows that $r_A(x) = r_B(x)$.

(c) We will make use of the following simple transformation property of spectra, which follows directly from the definition:

Exercise 9.13. Show that if $c, d \in \mathbb{C}$, $x \in A$, then $\sigma(cx+de) = c\sigma(x)+d$; this second set is of course defined as the collection of numbers $cz + d$, with $z \in \sigma(x)$.

By hypothesis, $\sigma(x) \subset [0, \|x\|]$. By the Exercise, $\sigma(x - \|x\|e) \subset [-\|x\|, 0]$ also, and now (b) implies that $\big\|x - \|x\|e\big\| \leq \|x\|$. Similarly, $\big\|y - \|y\|e\big\| \leq \|y\|$. Thus

$$\big\|x + y - (\|x\| + \|y\|)e\big\| \leq \|x\| + \|y\|,$$

and now a final application of the Exercise yields

$$\sigma(x + y) \subset [0, 2(\|x\| + \|y\|)].$$

(d) Obviously, $y = xx^*$ is self-adjoint. We will again consider the commutative C^*-algebra $B \subset A$ that is generated by y. We know that $B \cong C(\Delta_B)$. The function $|\widehat{y}|-\widehat{y}$ is continuous, so there exists $z \in B$ so that $\widehat{z} = |\widehat{y}|-\widehat{y}$. Since $\widehat{z}$ is also real valued, this function is a self-adjoint element of $C(\Delta_B)$, so we also have that $z = z^*$. Let $w = zx$ and write $w = u + iv$, with u, v self-adjoint. Then $ww^* = zxx^*z = zyz = z^2y$; in the last step, we used the fact that y and z both lie in the commutative algebra B. On the other hand,

$$ww^* = (u + iv)(u - iv) = u^2 + v^2 + i(vu - uv),$$
$$w^*w = (u - iv)(u + iv) = u^2 + v^2 + i(uv - vu),$$

so $w^*w = 2u^2+2v^2-ww^* = 2u^2+2v^2-z^2y$. We now claim that $u^2, v^2 \geq 0$. Since u, v are self-adjoint, this can again be seen by investigating the ranges of the Gelfand transforms on suitable commutative subalgebras, as in the proof of part (a). Moreover, we also have that

$$-\widehat{z}^2\widehat{y} = -(|\widehat{y}| - \widehat{y})^2\widehat{y} = 2\widehat{y}^2(|\widehat{y}| - \widehat{y}) \geq 0, \tag{9.1}$$

so $-z^2y \geq 0$. By part (c), $w^*w \geq 0$. Now Exercise 7.10 implies that $ww^* \geq 0$, and by Corollary 7.12(a), this also holds in the subalgebra B. But, as computed earlier, $ww^* = z^2y$, so by combining this with (9.1), we conclude that $\widehat{z}^2\widehat{y} \equiv 0$, so at all points of Δ_B, either $\widehat{y} = 0$ or $\widehat{z} = 0$. In both cases, $\widehat{y} \geq 0$, so we obtain that $\sigma_A(y) \subset \sigma_B(y) \subset [0, \infty)$, as claimed. □

Here's a very important and pleasing consequence of this material:

Theorem 9.16. *Let B be a C^*-algebra and let $A \subset B$ be a C^*-subalgebra. Then $\sigma_A(x) = \sigma_B(x)$ for all $x \in A$.*

Proof. It is clear that $\sigma_A(x) \supset \sigma_B(x)$ (see also our discussion in Chapter 7), so it suffices to show that if $y \in A\cap G(B)$, then also $y \in G(A)$. Now if $y \in A\cap G(B)$, then $y^* \in A\cap G(B)$ and thus also $yy^* \in A\cap G(B)$. In particular, $0 \notin \sigma_B(yy^*)$. Theorem 9.15(d) now shows that $\sigma_B(yy^*) \subset (0, \infty)$. By Corollary 7.12(a), $\sigma_A(yy^*) = \sigma_B(yy^*)$. Hence $0 \notin \sigma_A(yy^*)$, so $(yy^*)^{-1} \in A$, and thus also $y^{-1} = y^*(yy^*)^{-1} \in A$. □

We conclude this chapter with a short digression. Suppose that $xu = ux$. Does this imply that also $x^*u = ux^*$? For arbitrary u, this can only be true if x is normal (take $u = x$). This condition is indeed sufficient, and in fact we can prove a more general result along these lines.

Theorem 9.17. *Let A be a C^*-algebra and let $x, y, u \in A$. Suppose that x, y are normal and $xu = uy$. Then we also have that $x^*u = uy^*$.*

Proof. We need some preparation. For $w \in A$, define $e^w := \sum_{n=0}^{\infty} \frac{1}{n!} w^n$. This series converges absolutely, and just as for the ordinary exponential function, one shows that $e^{v+w} = e^v e^w = e^w e^v$ if $vw = wv$. Involution is a continuous operation (it is in fact isometric), and this implies that $(e^w)^* = e^{w^*}$. When applied to $w = t - t^*$ (where $t \in A$ is arbitrary), these formulae show that

$$e^w \left(e^w\right)^* = e^w e^{w^*} = e^w e^{-w} = e^{w-w} = 1;$$

here we denote the unit element of A by 1 (rather than e, as usual), to avoid confusion with the base of the exponential function. It follows that $1 = \|e^w \left(e^w\right)^*\| = \|e^w\|^2$ or

$$\|e^{t-t^*}\| = 1 \quad \text{for all } t \in A. \tag{9.2}$$

The assumption that $xu = uy$ can be used repeatedly, and we also obtain that $x^n u = uy^n$ for all $n \geq 0$. Multiplication is continuous, so this implies that $e^x u = ue^y$ or $u = e^{-x} u e^y$. We now multiply this identity by e^{x^*} and e^{-y^*} (from the left and right, respectively). Since x, y are normal, this gives

$$e^{x^*} u e^{-y^*} = e^{x^* - x} u e^{y - y^*},$$

and now (9.2) shows that $\|e^{x^*} u e^{-y^*}\| \leq \|u\|$. This whole argument can be repeated with x, y replaced by $\overline{z}x, \overline{z}y$, with $z \in \mathbb{C}$, so it is also true that $\|f(z)\| \leq \|u\|$, where $f(z) = e^{zx^*} u e^{-zy^*}$. For every $F \in A^*$, the new function $g(z) = F(f(z))$ is an entire function; the analyticity follows from the series representations of the exponential functions. Since g is also bounded ($|g(z)| \leq \|F\| \, \|u\|$), this function is constant by Liouville's theorem. Since this is true for every $F \in A^*$, f itself has to be constant:

$$f(z) = e^{zx^*} u e^{-zy^*} = u = f(0),$$

or $e^{zx^*} u = ue^{zy^*}$ for all $z \in \mathbb{C}$. We obtain the claim by comparing the first order terms in the series expansions of both sides (more formally, subtract u, divide by z and let $z \to 0$). □

Exercise 9.14. Let A be a commutative algebra with unit. True or false:
(a) There exist at most one norm and one involution on A so that A becomes a C^*-algebra.
(b) There exist a norm and an involution on A so that A becomes a C^*-algebra.

Exercise 9.15. Let A be a C^*-algebra and let x, y be normal elements of A that commute: $xy = yx$. Show that

$$\sigma(x+y) \subset \sigma(x) + \sigma(y) := \{w + z : w \in \sigma(x), z \in \sigma(y)\},$$
$$\sigma(xy) \subset \sigma(x)\sigma(y) := \{wz : w \in \sigma(x), z \in \sigma(y)\}.$$

Also show that both inclusions can fail if x, y don't commute.
Suggestion: Consider suitable 2×2-matrices for the counterexamples.

Exercise 9.16. Let A be a C^*-algebra and let $x \in A$ be normal. Then we can define $f(x) \in A$, for $f \in C(\sigma(x))$, as follows: Consider the commutative C^*-algebra $B \subset A$ that is generated by x, and then use Theorem 9.16 and the original definition of $f(x) \in B$, which was based on Theorem 9.13.

Prove the *spectral mapping theorem:* $\sigma(f(x)) = f(\sigma(x))$.
Hint: This follows very quickly from Theorem 9.16 and the fact that the map $f \mapsto f(x)$ sets up an isometric $*$-isomorphism between $C(\sigma(x))$ and B. Just make sure you don't get confused.

Exercise 9.17. Consider the following subalgebra of $\mathbb{C}^{2\times 2} = B(\mathbb{C}^2)$:

$$A = \left\{ y = \begin{pmatrix} a & b \\ b & a \end{pmatrix} : a, b \in \mathbb{C} \right\}$$

(a) Show that A is a commutative C^*-algebra (with the structure inherited from $B(\mathbb{C}^2)$; in particular, $\left(\begin{smallmatrix} a & b \\ b & a \end{smallmatrix}\right)^* = \left(\begin{smallmatrix} \overline{a} & \overline{b} \\ \overline{b} & \overline{a} \end{smallmatrix}\right)$). *Remark:* Most of this is already clear because we know that the bigger algebra $B(\mathbb{C}^2)$ is a C^*-algebra.
(b) Show that A is generated by $x = \left(\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right)$.
(c) Show that $\Delta = \{\phi_1, \phi_2\}$, where $\phi_1(y) = a + b$, $\phi_2(y) = a - b$.
(d) Find $\sigma(x)$ and confirm the (here: obvious) fact that $\Delta \cong \sigma(x)$, as asserted by Theorem 9.12.
(e) Find $f(x) \in A$, for the functions $f(z) = |z|$ and $f(z) = 1/2(|z| + z)$.

10. The Spectral Theorem

The big moment has arrived, and we are now ready to prove several versions of the spectral theorem for normal operators in Hilbert spaces. Throughout this chapter, it should be helpful to compare our results with the more familiar special case when the Hilbert space is finite-dimensional. In this setting, the spectral theorem says that every normal matrix $T \in \mathbb{C}^{n\times n}$ can be diagonalized by a unitary transformation. This can be rephrased as follows: There are numbers $z_j \in \mathbb{C}$ (the eigenvalues) and orthogonal projections $P_j \in B(\mathbb{C}^n)$ so that $T = \sum_{j=1}^m z_j P_j$. The subspaces $R(P_j)$ are orthogonal to each other. From this representation of T, it is then also clear that P_j is the projection onto the eigenspace belonging to z_j.

In fact, we have already proved one version of the (general) spectral theorem: The Gelfand theory of the commutative C^*-algebra $A \subset B(H)$ that is generated by a normal operator $T \in B(H)$ provides a *functional calculus:* We can define $f(T)$, for $f \in C(\sigma(T))$ in such a way that the map $C(\sigma(T)) \to A$, $f \mapsto f(T)$ is an isometric $*$-isomorphism between C^*-algebras, and this is the spectral theorem in one of its many disguises! See Theorem 9.13 and the discussion that follows. As a warm-up, let us use this material to give a quick proof of the result about normal matrices $T \in \mathbb{C}^{n\times n}$ that was stated above.

Consider the C^*-algebra $A \subset \mathbb{C}^{n\times n}$ that is generated by T. Since T is normal, A is commutative. By Theorem 9.13, $A \cong C(\sigma(T)) = C(\{z_1, \ldots, z_m\})$, where $z_1, \ldots, z_m$ are the eigenvalues of T. We also use the fact that by Theorem 9.16, $\sigma_A(T) = \sigma_{B(H)}(T)$.

All subsets of the discrete space $\{z_1, \ldots, z_m\}$ are open, and thus all functions $f : \{z_1, \ldots, z_m\} \to \mathbb{C}$ are continuous. We will make use of the functional calculus notation: $f(T) \in A$ will denote the operator that corresponds to the function f under the isometric $*$-isomorphism that sends the identity function $\mathrm{id}(z) = z$ to $T \in A$. Write $f_j = \chi_{\{z_j\}}$ and let $P_j = f_j(T)$. Since $\overline{f_j} = f_j$ and $f_j^2 = f_j$, we also have that $P_j^* = P_j$ and $P_j^2 = P_j$, so each P_j is an orthogonal projection by Theorem 6.5. Furthermore, $f_j f_k = 0$ if $j \neq k$, so $P_j P_k = 0$, and thus

$$\langle P_j x, P_k y\rangle = \langle x, P_j P_k y\rangle = 0$$

for all $x, y \in H$ if $j \neq k$. This says that $R(P_j) \perp R(P_k)$ for $j \neq k$. Also, $P_1 + \ldots + P_m = 1$ because we have the same identity for the f_j's. It follows that $\bigoplus_{j=1}^m R(P_j) = H = \mathbb{C}^n$. Finally, since $\mathrm{id} = \sum_{j=1}^m z_j f_j$, we obtain the representation $T = \sum_{j=1}^m z_j P_j$, as asserted.

On infinite-dimensional Hilbert spaces, we have a continuous analog of this representation: every normal $T \in B(H)$ can be written as

$T = \int z\, dP(z)$. We first need to address the question of how such an integral can be meaningfully defined. We will also switch to the more common symbol E (rather than P) for these "measures" (if that's what they are).

Definition 10.1. Let $\mathcal{M}$ be a σ-algebra on a set Ω, and let H be a Hilbert space. A *resolution of the identity* (or *spectral resolution*) on $(\Omega, \mathcal{M})$ is a map $E : \mathcal{M} \to B(H)$ with the following properties:
(1) Every $E(\omega)$ ($\omega \in \mathcal{M}$) is a projection;
(2) $E(\emptyset) = 0$, $E(\Omega) = 1$;
(3) $E(\omega_1 \cap \omega_2) = E(\omega_1)E(\omega_2) \quad (\omega_1, \omega_2 \in \mathcal{M})$;
(4) For all $x, y \in H$, the set function $\mu_{x,y}(\omega) = \langle x, E(\omega)y\rangle$ is a complex measure on $(\Omega, \mathcal{M})$. If Ω is a locally compact Hausdorff space and $\mathcal{M} = \mathcal{B}$ is the Borel σ-algebra, then we also demand that every $\mu_{x,y}$ is a regular (Borel) measure.

We can think of E as a projection valued measure (of sorts) on $(\Omega, \mathcal{M})$: the "measure" $E(\omega)$ of a set $\omega \in \mathcal{M}$ is a projection. The $E(\omega)$ are also called *spectral projections.*

Let's start out with some quick observations. For every $x \in H$, we have that

$$\mu_{x,x}(\omega) = \langle x, E(\omega)x\rangle = \langle x, E(\omega)^2 x\rangle = \langle E(\omega)x, E(\omega)x\rangle = \|E(\omega)x\|^2,$$

so $\mu_{x,x}$ is a finite positive measure with $\mu_{x,x}(\Omega) = \|x\|^2$. Property (3) implies that any two spectral projections $E(\omega_1)$, $E(\omega_2)$ commute. Moreover, if $\omega_1 \subset \omega_2$, then $R(E(\omega_1)) \subset R(E(\omega_2))$. If $\omega_1 \cap \omega_2 = \emptyset$, then $R(E(\omega_1)) \perp R(E(\omega_1))$, as the following calculation shows:

$$\langle E(\omega_1)x, E(\omega_2)y\rangle = \langle x, E(\omega_1)E(\omega_2)y\rangle = \langle x, E(\omega_1 \cap \omega_2)y\rangle = 0$$

for arbitrary $x, y \in H$.

E is *finitely additive:* If $\omega_1, \ldots, \omega_n \in \mathcal{M}$ are disjoint sets, then $E\left(\bigcup_{j=1}^n \omega_j\right) = \sum_{j=1}^n E(\omega_j)$. To prove this, notice that (4) implies that

$$\begin{aligned}\langle x, E\left(\bigcup_{j=1}^n \omega_j\right) y\rangle &= \mu_{x,y}\left(\bigcup_{j=1}^n \omega_j\right) = \sum_{j=1}^n \mu_{x,y}(\omega_j)\\ &= \sum_{j=1}^n \langle x, E(\omega_j)y\rangle = \langle x, \sum_{j=1}^n E(\omega_j)y\rangle\end{aligned}$$

for arbitrary $x, y \in H$, and this gives the claim.

Is E also σ-additive (as it ought to be, if we are serious about interpreting E as a new sort of measure)? In other words, if $\omega_n \in \mathcal{M}$ are disjoint sets, does it follow that

$$E\left(\bigcup_{n\in\mathbb{N}}\omega_n\right)=\sum_{n=1}^{\infty}E(\omega_n). \tag{10.1}$$

The answer to this question depends on how one defines the right-hand side of (10.1). We observe that if $E(\omega_n)\neq 0$ for infinitely many n, then this series can never be convergent in operator norm. Indeed, $\|E(\omega_n)\|=1$ if $E(\omega_n)\neq 0$, and thus the partial sums do not form a Cauchy sequence. However, (10.1) will hold if we are satisfied with strong operator convergence: We say that $T_n\in B(H)$ *converges strongly* to $T\in B(H)$ (notation: $T_n\xrightarrow{s}T$) if $T_nx\to Tx$ for all $x\in H$.

To prove that (10.1) holds in this interpretation, fix $x\in H$ and use the fact that the $E(\omega_n)x$ form an orthogonal system (because the ranges of the projections are orthogonal subspaces for disjoint sets). We normalize the non-zero vectors: let $y_n=\frac{E(\omega_n)x}{\|E(\omega_n)x\|}$ if $E(\omega_n)x\neq 0$. Then the y_n form an ONS, and thus, by Theorem 5.15, the series $\sum\langle y_n,x\rangle y_n=\sum E(\omega_n)x$ converges. Now if $y\in H$ is arbitrary, then the continuity of the scalar product and the fact that $\mu_{x,y}$ is a complex measure give that

$$\langle y,\sum_{n=1}^{\infty}E(\omega_n)x\rangle=\sum_{n=1}^{\infty}\langle y,E(\omega_n)x\rangle=\langle y,E\left(\bigcup_{n\in\mathbb{N}}\omega_n\right)x\rangle.$$

Since this holds for every $y\in H$, it follows that $\sum_{n=1}^{\infty}E(\omega_n)x=E\left(\bigcup_{n\in\mathbb{N}}\omega_n\right)x$, and this is (10.1), with the series interpreted as a strong operator limit.

Definition 10.2. A set $N\in\mathcal{M}$ with $E(N)=0$ is called an *E-null set.* We define $L^{\infty}(\Omega,E)$ as the set of equivalence classes of measurable, essentially bounded functions $f:\Omega\to\mathbb{C}$. Here, $f\sim g$ if f and g agree off an E-null set. Also, as usual, we say that f is essentially bounded if $|f(x)|\leq M$ $(x\in\Omega\setminus N)$ for some $M\geq 0$ and some E-null set $N\subset\Omega$.

Exercise 10.1. Prove that a countable union of E-null sets is an E-null set.

Recall that for an arbitrary positive measure μ on X, the space $L^{\infty}(X,\mu)$ only depends on what the μ-null sets are and not on the specific choice of the measure μ. For this reason and because of Exercise 10.1, we can also, and if fact without any difficulties, introduce L^{∞} spaces that are based on resolutions of the identity. These spaces have the same basic properties: $L^{\infty}(\Omega,E)$ with the essential supremum of $|f|$ as the norm and the involution $f^*(x)=\overline{f(x)}$ is a commutative

C^*-algebra. The spectrum of a function $f \in L^\infty(\Omega, E)$ is its essential range.

Exercise 10.2. Write down precise definitions of the essential supremum and the essential range of a function $f \in L^\infty(\Omega, E)$.

We would like to define an integral $\int_\Omega f(t)\, dE(t)$ for $f \in L^\infty(\Omega, E)$. This integral should be an operator from $B(H)$, and it also seems reasonable to demand that

$$\langle x, \left(\int_\Omega f(t)\, dE(t) \right) y \rangle = \int_\Omega f(t)\, d\mu_{x,y}(t)$$

for all $x, y \in H$. It is clear that this condition already suffices to uniquely determine $\int_\Omega f(t)\, dE(t)$, should such an operator indeed exist. As for existence, we have the following result; it will actually turn out that the integral with respect to a resolution of the identity has many other desirable properties, too.

Theorem 10.3. *Let E be a resolution of the identity. Then there exists a unique map $\Psi : L^\infty(\Omega, E) \to A$ onto a C^*-subalgebra $A \subset B(H)$, so that*

$$\langle x, \Psi(f) y \rangle = \int_\Omega f(t)\, d\mu_{x,y}(t) \tag{10.2}$$

for all $f \in L^\infty(\Omega, E)$, $x, y \in H$. Moreover, Ψ is an isometric $*$-isomorphism from $L^\infty(\Omega, E)$ onto A, and

$$\|\Psi(f)x\|^2 = \int_\Omega |f(t)|^2\, d\mu_{x,x}(t). \tag{10.3}$$

So we can (and will) define $\int_\Omega f(t)\, dE(t) := \Psi(f)$. Let us list the properties of the integral that are guaranteed by Theorem 10.3 one more time, using this new notation:

$$\int (f+g)\, dE = \int f\, dE + \int g\, dE, \quad \int (cf)\, dE = c \int f\, dE,$$
$$\int fg\, dE = \int f\, dE \int g\, dE,$$
$$\left(\int f\, dE \right)^* = \int \overline{f}\, dE, \quad \left\| \int f\, dE \right\| = \|f\|_\infty$$

The multiplicativity of the integral (see the second line) may seem a bit strange at first, but it becomes plausible again if we recall that all $E(\omega)$ are projections.

Exercise 10.3. Show that $\sum_{j=1}^m f_j P_j \sum_{j=1}^m g_j P_j = \sum_{j=1}^m f_j g_j P_j$ if the P_j are projections with orthogonal ranges, as at the beginning of this chapter, and $f_j, g_j \in \mathbb{C}$.

Proof. This is not a particularly short proof, but it follows a standard pattern. First of all, we certainly know how we want to define $\int f\,dE$ for *simple functions* $f \in L^\infty(\Omega, E)$, that is, functions of the form $f = \sum_{j=1}^n c_j \chi_{\omega_j}$ with $c_j \in \mathbb{C}$ and $\omega_j \in \mathcal{M}$. For such an f, put $\Psi(f) = \sum_{j=1}^n c_j E(\omega_j)$. For $x, y \in H$, we then obtain that

$$\langle x, \Psi(f)y \rangle = \sum_{j=1}^n c_j \langle x, E(\omega_j)y \rangle = \sum_{j=1}^n c_j \mu_{x,y}(\omega_j) = \int_\Omega f(t)\, d\mu_{x,y}(t).$$

This is (10.2) for simple functions f, and this identity also confirms that $\Psi(f)$ was indeed well defined ($\Psi(f)$ is determined by the *function* f, and it is independent of the particular representation of f that was chosen to form $\Psi(f)$).

We also have that $\Psi(f)^* = \sum \overline{c_j} E(\omega_j) = \Psi(\overline{f})$, and if $g = \sum_{k=1}^m d_k \chi_{\omega'_k}$ is a second simple function, then

$$\Psi(f)\Psi(g) = \sum_{j,k} c_j d_k E(\omega_j) E(\omega'_k) = \sum_{j,k} c_j d_k E(\omega_j \cap \omega'_k) = \Psi(fg).$$

For the last equality, we use the fact that fg is another simple function, with representation $fg = \sum_{j,k} c_j d_k \chi_{\omega_j \cap \omega'_k}$. Similar arguments show that Ψ is linear (on simple functions). Finally, (10.3) (for simple functions) follows from the identity $\Psi(f)^*\Psi(f) = \Psi(\overline{f})\Psi(f) = \Psi(|f|^2)$:

$$\|\Psi(f)x\|^2 = \langle x, \Psi(f)^*\Psi(f)x \rangle = \langle x, \Psi(|f|^2)x \rangle = \int_\Omega |f(t)|^2\, d\mu_{x,x}(t)$$

This also implies that $\|\Psi(f)x\|^2 \le \|f\|_\infty^2 \|x\|^2$, so $\|\Psi(f)\| \le \|f\|$. On the other hand, the sets ω_j in the representation $f = \sum c_j \chi_{\omega_j}$ can be taken to be disjoint (just take $\omega_j = f^{-1}(\{c_j\})$). Now if $E(\omega_j) \ne 0$, then there exists $x \in R(E(\omega_j))$, $x \ne 0$. Clearly, $\Psi(f)x = c_j x$, and since $\|f\|_\infty = \max_{j: E(\omega_j) \ne 0} |c_j|$, we now see that $\|\Psi(f)\| = \|f\|$. So Ψ is isometric (on simple functions).

We now want to extend these results to arbitrary functions $f \in L^\infty(\Omega, E)$ by using an approximation procedure.

Exercise 10.4. Let $f \in L^\infty(\Omega, E)$. Show that there exists a sequence of *simple* functions $f_n \in L^\infty(\Omega, E)$ so that $\|f_n - f\| \to 0$.

Let $f \in L^\infty(\Omega, E)$ and pick an approximating sequence f_n of simple functions, as in Exercise 10.4. Notice that $\Psi(f_n)$ converges in $B(H)$:

indeed,

$$\|\Psi(f_m) - \Psi(f_n)\| = \|\Psi(f_m - f_n)\| = \|f_m - f_n\|,$$

so this is a Cauchy sequence. The same argument shows that the limit is independent of the specific choice of the approximating sequence, so we can define $\Psi(f) := \lim \Psi(f_n)$. The continuity of the scalar product gives

$$\langle x, \Psi(f)y\rangle = \lim_{n\to\infty} \langle x, \Psi(f_n)y\rangle = \lim_{n\to\infty} \int_\Omega f_n(t)\, d\mu_{x,y}(t).$$

Every E-null set is a $|\mu_{x,y}|$-null set, so f_n converges $\mu_{x,y}$-almost everywhere to f. Moreover, $|f_n| \le \|f_n\|_\infty \le C$ off an E-null set, so again $\mu_{x,y}$-almost everywhere. The constant function C lies in $L^1(\Omega, d|\mu_{x,y}|)$ because $|\mu_{x,y}|$ is a finite measure. We have just verified the hypotheses of the Dominated Convergence Theorem. It follows that $\lim_{n\to\infty} \int_\Omega f_n\, d\mu_{x,y} = \int_\Omega f\, d\mu$, and we obtain (10.2) (for arbitrary $f \in L^\infty(\Omega, E)$).

Exercise 10.5. Establish (10.3) in a similar way.

The remaining properties follow easily by passing to limits. For example, if $f, g \in L^\infty$, pick approximating simple functions f_n, g_n and use the continuity of the multiplication to deduce that

$$\begin{aligned}\Psi(f)\Psi(g) &= \lim \Psi(f_n) \lim \Psi(g_n) = \lim \Psi(f_n)\Psi(g_n)\\ &= \lim \Psi(f_n g_n) = \Psi(fg).\end{aligned}$$

In the last step, we use the fact that $f_n g_n$ is a sequence of simple functions that converges to fg in the norm of $L^\infty(\Omega, E)$.

Exercise 10.6. Prove at least two more properties of Ψ (Ψ linear, isometric, $\Psi(f)^* = \Psi(\overline{f})$) in this way.

Finally, since Ψ is an isometry, its image $A = \Psi(L^\infty(\Omega, E))$ is closed (compare the proof of Proposition 4.3), and it is also a subalgebra that is closed under the involution $*$ because Ψ is a $*$-homomorphism. $\square$

We now have the tools to prove the next version of the Spectral Theorem (the first version being the existence of a functional calculus for normal operators). We actually obtain a more abstract version for a whole algebra of operators from our machinery; we discuss this first and then specialize to a single operator later on, in Theorem 10.5.

Theorem 10.4. *Suppose $A \subset B(H)$ is a commutative C^*-subalgebra of $B(H)$. Let Δ be its maximal ideal space.*

(a) There exists a unique resolution of the identity on the Borel sets of Δ (with its Gelfand topology) so that

$$T = \int_{\Delta} \widehat{T}(t)\, dE(t) \tag{10.4}$$

for all $T \in A$.

Moreover, E has the following additional properties:
(b) $B = \{\int_{\Delta} f(t)\, dE(t) : f \in L^{\infty}(\Omega, E)\}$ is a commutative C^*-algebra satisfying $A \subset B \subset B(H)$.
(c) The finite linear combinations of the $E(\omega)$, $\omega \in \mathcal{M}$ are dense in B.
(d) If $\omega \subset \Delta$ is a non-empty open set, then $E(\omega) \neq 0$.

Proof. By the Gelfand-Naimark Theorem, $A \cong C(\Delta)$. We will now use the Riesz Representation Theorem: $C(\Delta)^* = \mathcal{M}(\Delta)$, the space of regular complex Borel measures on Δ. See Example 4.2. The uniqueness of E follows immediately from this: If E satisfies (10.4), then $\int_{\Delta} \widehat{T}(t)\, d\mu_{x,y}(t) = \langle x, Ty\rangle$, and every continuous function on Δ is of the form $\widehat{T}$ for some $T \in A$, so the functionals (on $C(\Delta)$) associated with the measures $\mu_{x,y}$ and thus also the measures themselves are already determined by (10.4). Since $x, y \in H$ are arbitrary here, E itself is determined by (10.4).

To prove existence of E, we fix $x, y \in H$ and consider the map $C(\Delta) \to \mathbb{C}$, $\widehat{T} \mapsto \langle x, Ty\rangle$. Since the inverse of the Gelfand transform, $\widehat{T} \mapsto T$, is linear, this map is linear, too, and also bounded, as we see from

$$|\langle x, Ty\rangle| \le \|x\|\,\|Ty\| \le \|x\|\,\|T\|\,\|y\| = \|x\|\,\|y\|\,\|\widehat{T}\|_{\infty}.$$

By the Riesz Representation Theorem, there is a regular complex Borel measure on Δ (call it $\mu_{x,y}$) so that

$$\langle x, Ty\rangle = \int_{\Delta} \widehat{T}(t)\, d\mu_{x,y}(t) \tag{10.5}$$

for all $T \in A$. Our goal is to construct a resolution of the identity E for which $\langle x, E(\omega)y\rangle = \mu_{x,y}(\omega)$. That will finish the proof of part (a).

As a function of x, y, $\langle x, Ty\rangle$ is sesquilinear. From this, it follows that that $(x, y) \mapsto \mu_{x,y}$ is sesquilinear, too. This means that $\mu_{x+y,z} = \mu_{x,z} + \mu_{y,z}$, $\mu_{cx,y} = \overline{c}\mu_{x,y}$, and $\mu_{x,y}$ is linear in y.

Exercise 10.7. Prove this claim.

If now $f : \Delta \to \mathbb{C}$ is a bounded measurable function, then $(x, y) \mapsto \int_{\Delta} f(t)\, d\mu_{x,y}(t)$ defines another sesquilinear form. In fact, this form is bounded in the sense that

$$\left| \int_\Delta f(t)\, d\mu_{x,y}(t) \right| \le \left(\sup_{t\in\Delta} |f(t)| \right) |\mu_{x,y}|(\Delta) \le \left(\sup_{t\in\Delta} |f(t)| \right) \|x\|\, \|y\|.$$

By Exercise 6.10, there is a unique operator $\Phi(f) \in B(H)$, so that

$$\langle x, \Phi(f)y\rangle = \int_\Delta f(t)\, d\mu_{x,y}(t)$$

for all $x, y \in H$. If $f \in C(\Delta)$ here, then a comparison with (10.5) shows that $\Phi(f) = T$, where $T \in A$ is the unique operator with $\widehat{T} = f$. Now

$$\int_\Delta \widehat{T}\, d\mu_{x,y} = \langle x, Ty\rangle = \overline{\langle y, T^*x\rangle} = \overline{\int_\Delta \widehat{T^*}\, d\mu_{y,x}} = \int_\Delta \widehat{T}\, d\overline{\mu_{y,x}},$$

and this holds for all functions $\widehat{T} \in C(\Delta)$, so we conclude that $\mu_{x,y} = \overline{\mu_{y,x}}$, where, as expected, the measure $\overline{\nu}$ is defined by $\overline{\nu}(\omega) = \overline{\nu(\omega)}$. But then we can use this for integrals of arbitrary bounded Borel functions f:

$$\langle x, \Phi(\overline{f})y\rangle = \int_\Delta \overline{f}\, d\mu_{x,y} = \overline{\int_\Delta f\, d\mu_{y,x}} = \overline{\langle y, \Phi(f)x\rangle} = \langle \Phi(f)x, y\rangle,$$

so $\Phi(f)^* = \Phi(\overline{f})$. Next, for $S, T \in A$, we have that

$$\int_\Delta \widehat{S}\widehat{T}\, d\mu_{x,y} = \int_\Delta (ST)\widehat{\ }\, d\mu_{x,y} = \langle x, STy\rangle = \int_\Delta \widehat{S}\, d\mu_{x,Ty},$$

so $\widehat{T}\, d\mu_{x,y} = d\mu_{x,Ty}$. Again, we can apply this to integrals of arbitrary bounded Borel functions f: $\int f\widehat{T}\, d\mu_{x,y} = \int f\, d\mu_{x,Ty}$, and this implies that

$$\int_\Delta f\widehat{T}\, d\mu_{x,y} = \langle x, \Phi(f)Ty\rangle = \langle \Phi(f)^*x, Ty\rangle = \int_\Delta \widehat{T}\, d\mu_{\Phi(f)^*x,y}.$$

Since $\widehat{T} \in C(\Delta)$ is arbitrary here, this says that $f\, d\mu_{x,y} = d\mu_{\Phi(f)^*x,y}$, so $\int fg\, d\mu_{x,y} = \int g\, d\mu_{\Phi(f)^*x,y}$ for all bounded Borel functions g. Now $\int fg\, d\mu_{x,y} = \langle x, \Phi(fg)y\rangle$ and

$$\int_\Delta g\, d\mu_{\Phi(f)^*x,y} = \langle \Phi(f)^*x, \Phi(g)y\rangle = \langle x, \Phi(f)\Phi(g)y\rangle,$$

so we finally obtain the desired conclusion that $\Phi(fg) = \Phi(f)\Phi(g)$.

We can now define $E(\omega) = \Phi(\chi_\omega)$. I claim that E is a resolution of the identity. Clearly, by construction,

$$\langle x, E(\omega)y\rangle = \langle x, \Phi(\chi_\omega)y\rangle = \int_\Delta \chi_\omega\, d\mu_{x,y} = \mu_{x,y}(\omega),$$

as required. This also verifies (4) from Definition 10.1. It remains to check the conditions (1)–(3).

Notice that $E(\omega)^* = \Phi(\chi_\omega)^* = \Phi(\overline{\chi_\omega}) = \Phi(\chi_\omega) = E(\omega)$, so $E(\omega)$ is self-adjoint. Similarly, $E(\omega)^2 = \Phi(\chi_\omega)^2 = \Phi(\chi_\omega^2) = \Phi(\chi_\omega) = E(\omega)$. By Theorem 6.5, $E(\omega)$ is a projection, so (1) holds. A similar computation lets us verify (3). Finally, moving on to (2), it is clear that $E(\emptyset) = \Phi(0) = 0$, and $E(\Delta) = \Phi(1)$. Now the constant function 1 is continuous, so, as observed above, $\Phi(1)$ is the operator whose Gelfand transform is identically equal to one, but this is the identity operator $1 \in A \subset B(H)$ (the multiplicative unit of A and $B(H)$). So $E(1) = 1$, as desired.

(b) We know from Theorem 10.3 that B is a C^*-subalgebra of $B(H)$, and since continuous functions are in $L^\infty(\Delta, E)$, we clearly have that $B \supset A$.

(c) This is immediate from the way the integral $\int f\, dE$ was constructed, in the proof of Theorem 10.3.

(d) Let $\omega \subset \Delta$ be a non-empty open set. Pick $t_0 \in \omega$ and use Urysohn's Lemma to find a continuous function f with $f(t_0) = 1$, $f = 0$ on ω^c. Then $f = \widehat{T}$ for some $T \in A$, and if we had $E(\omega) = 0$, then $T = \int_\Delta \widehat{T}\, dE = 0$, but this is impossible because $\widehat{T} = f$ is not the zero function. $\square$

We now specialize to (algebras generated by) a single normal operator.

Theorem 10.5 (The Spectral Theorem for normal operators)**.** *Let $T \in B(H)$ be a normal operator. Then there exists a unique resolution of the identity E on the Borel sets of $\sigma(T)$ so that*

$$T = \int_{\sigma(T)} z\, dE(z). \tag{10.6}$$

Proof. Consider, as usual, the commutative C^*-algebra $A \subset B(H)$ that is generated by T. Existence of E now follows from Theorem 10.4(a) because we can make the following identifications: By Theorems 9.12 and 9.13, Δ_A is homeomorphic to $\sigma(T)$, and $A \cong C(\sigma(T))$. Here we may interpret $\sigma(T)$ as $\sigma_{B(H)}(T)$ because $\sigma_A(T)$ is the same set by Theorem 9.16. From a formal point of view, perhaps the most satisfactory argument runs as follows: Reexamine the proof of Theorem 10.4 to confirm that we obtain the representation $T = \int_K f\, dE$ as soon as we have an isometric $*$-isomorphism between A and $C(K)$ that sends T to f (it is not essential that this isomorphism is specifically the Gelfand transform). In the case at hand, $A \cong C(\sigma(T))$, by Theorem 9.13, and the corresponding isomorphism sends T to $\mathrm{id}(z) = z$, so we obtain (10.6). For later use, we also record that, by the same argument, $f(T) = \int_{\sigma(T)} f(z)\, dE(z)$ for all $f \in C(\sigma(T))$, where $f(T) \in A$ is defined as in Chapter 9; see especially the discussion following Theorem 9.13.

Let us now prove uniqueness of E. By Theorem 10.3, if (10.6) holds, then also $p(T,T^*) = \int_{\sigma(T)} p(z,\overline{z})\, dE(z)$ for all polynomials p in two variables. When viewed as functions of z only, this set

$$\{f : \sigma(T) \to \mathbb{C} : f(z) = p(z,\overline{z}),\ p \text{ polynomial in two variables}\}$$

satisfies the hypotheses of the Stone-Weierstraß Theorem. Therefore, if $f \in C(\sigma(T))$ is arbitrary, there are polynomials p_n so that $\|f(z) - p_n(z,\overline{z})\|_\infty \to 0$. Alternatively, this conclusion can also be obtained from the fact that T generates A, so $\{p(T,T^*)\}$ is dense in A, and we can then move things over to $C(\sigma(T))$.

The Dominated Convergence Theorem shows that

$$\int_{\sigma(T)} f(z)\, d\mu_{x,y}(z) = \lim_{n\to\infty} \int_{\sigma(T)} p_n(z,\overline{z})\, d\mu_{x,y}(z) = \lim_{n\to\infty} \langle x, p_n(T,T^*)y\rangle.$$

So the measures $\mu_{x,y}$ and thus also E itself are uniquely determined. □

This proof has also established the following fact, which we state again because it will prove useful in the sequel:

Proposition 10.6. *If E is the spectral resolution of $T \in B(H)$, as in the Spectral Theorem, then $E(U \cap \sigma(T)) \neq 0$ for all open sets $U \subset \mathbb{C}$ with $U \cap \sigma(T) \neq \emptyset$.*

This follows from Theorem 10.4(d) and our identification of Δ_A with $\sigma(T)$.

We introduce some new notation. It will occasionally be convenient to write $d\langle x, E(z)y\rangle$ for the measure $d\mu_{x,y}(z)$. Similarly, $d\langle x, E(z)x\rangle$ and $d\|E(z)x\|^2$ both refer to the measure $d\mu_{x,x}(z)$. This notation is reasonable because $\langle x, E(\omega)x\rangle = \|E(\omega)x\|^2$.

We can now also extend the *functional calculus* from Chapter 9. More precisely, for a normal $T \in B(H)$ and $f \in L^\infty(\sigma(T),E)$, where E is the resolution of the identity of T, as in the Spectral Theorem, let

$$f(T) := \int_{\sigma(T)} f(z)\, dE(z). \tag{10.7}$$

As observed above, in the proof of Theorem 10.5, this is consistent with our earlier definition of $f(T)$ for $f \in C(\sigma(T))$ from Chapter 9.

By Theorem 10.3, the functional calculus $f \mapsto f(T)$ is an isometric $*$-isomorphism between $L^\infty(\sigma(T),E)$ and a subalgebra of $B(H)$. Note also that if $p(z)$ is a polynomial, $p(z) = \sum_{j=0}^n c_j z^j$, then $p(T)$ could have been defined directly as $p(T) = \sum_{j=0}^n c_j T^j$, and the functional calculus gives the same result. A similar remark applies to functions of the form $p(z, \overline{z})$.

We state the basic properties of the functional calculus one more time:

$$(cf + dg)(T) = cf(T) + dg(T), \quad (fg)(T) = f(T)g(T) = g(T)f(T)$$

$$f(T)^* = \overline{f}(T), \quad \|f(T)\| = \|f\|_\infty, \quad \|f(T)x\|^2 = \int_{\sigma(T)} |f(z)|^2 \, d\|E(z)x\|^2$$

Moreover, if f is continuous, then we have the *spectral mapping theorem:* $\sigma(f(T)) = f(\sigma(T))$. This was discussed in Exercise 9.16.

We want to prove still another version of the Spectral Theorem. This last version will be an analog of the statement: a normal matrix can be diagonalized by a unitary transformation. We will needs sums of Hilbert spaces to formulate this result, so we discuss this topic first. If $H_1, \ldots, H_n$ are Hilbert spaces, then we can construct a new Hilbert space $H = \bigoplus_{j=1}^n H_j$, as follows: As a vector space, H is the sum of the vector spaces H_j, and if $x, y \in H$, say $x = (x_1, \ldots, x_n)$, $y = (y_1, \ldots, y_n)$, then we define $\langle x, y \rangle = \sum_{j=1}^n \langle x_j, y_j \rangle_{H_j}$.

Exercise 10.8. Verify that this defines a scalar product on H and that H is complete with respect to the corresponding norm.

Note that each H_j can be naturally identified with a closed subspace of H, by sending $x_j \in H_j$ to $x = (0, \ldots, 0, x_j, 0, \ldots 0)$. In fact, the H_j, viewed in this way as subspaces of H, are pairwise orthogonal. Conversely, if H is a Hilbert space and the H_j are orthogonal subspaces of H, then $\bigoplus H_j$ can be naturally identified with a subspace of H (by mapping (x_j) to $\sum x_j$).

An analogous construction works for infinitely many summands H_α, $\alpha \in I$. We now define H to be the set of vectors $x = (x_\alpha)_{\alpha \in I}$ ($x_\alpha \in H_\alpha$) that satisfy $\sum_{\alpha \in I} \|x_\alpha\|^2 < \infty$. If I is uncountable, then, as usual, this means that $x_\alpha \neq 0$ for only countably many α and the corresponding series is required to converge. We can again define $\langle x, y \rangle = \sum_{\alpha \in I} \langle x_\alpha, y_\alpha \rangle$; the convergence of this series follows from the definition on H and the Cauchy-Schwarz inequality for both the individual scalar products and then also the sum over $\alpha \in I$.

Exercise 10.9. Again, prove that this defines a scalar product and that H is a Hilbert space.

Theorem 10.7 (Spectral representation of normal operators). *Let $T \in B(H)$ be a normal operator. Then there exists a collection $\{\rho_\alpha : \alpha \in I\}$ of finite positive Borel measures on $\sigma(T)$ and a unitary map $U : H \to$*

$\bigoplus_{\alpha \in I} L^2(\sigma(T), d\rho_\alpha)$ *so that*

$$UTU^{-1} = M_z, \qquad (M_z f)_\alpha(z) = z f_\alpha(z).$$

The minimal cardinality of such a set I is called the *spectral multiplicity* of T; if H is separable (as almost all Hilbert spaces that occur in practice are), then I can always taken to be a countable set (say $I = \mathbb{N}$). Sometimes, a finite I will suffice or even an I consisting of just one element, so that T would then be unitarily equivalent to a multiplication operator by the variable on a sinle $L^2(\rho)$ space.

Exercise 10.10. Let $T \in \mathbb{C}^{n\times n}$ be a normal matrix, with eigenvalues $\sigma(T) = \{z_1, \ldots, z_m\}$. Prove the existence of a spectral representation directly, by providing the details in the following sketch: Choose the ρ_α as counting measures on (subsets of) $\sigma(T)$, and to define U, send a vector $x \in \mathbb{C}^n$ to its expansion coefficients with respect to an ONB consisting of eigenvectors of T.

Exercise 10.11. Use the discussion of the previous Exercise to show that for a normal $T \in \mathbb{C}^{n\times n}$, the spectral multiplicity (as defined above) is the maximal degeneracy of an eigenvalue, or, put differently, it is equal to $\max_{z\in\sigma(T)} \dim N(T - z)$.

The measures ρ_α from Theorem 10.7 are called *spectral measures.* They are *not* uniquely determined by the operator T; Exercise 10.17 below will shed some additional light on this issue.

Proof. For $x \in H$, $x \neq 0$, let

$$H_x = \overline{\{f(T)x : f \in C(\sigma(T))\}}.$$

We also define an operator $U_x : H_x \to L^2(\sigma(T), d\mu_{x,x})$, as follows: For $f \in C(\sigma(T))$, put $U_x^{(0)} f(T)x = f$. Then

$$\|U_x^{(0)} f(T)x\|^2 = \int_{\sigma(T)} |f(z)|^2 \, d\mu_{x,x}(z) = \|f(T)x\|^2,$$

by Theorem 10.3. By Exercise 2.26, the operator $U_x^{(0)} : \{f(T)x\} \to L^2(\mu_{x,x})$ has a unique continuous extension to H_x (call it U_x). Since the norm is continuous, U_x will also be isometric. In particular, $R(U_x)$ is closed, but clearly $R(U_x)$ also contains every continuous function on $\sigma(T)$, and these are dense in $L^2(\sigma(T), d\mu_{x,x})$, so $R(U_x) = L^2(\sigma(T), d\mu_{x,x})$. Summing up: U_x is a unitary map (a linear bijective isometry) from H_x onto $L^2(\sigma(T), d\mu_{x,x})$.

Now let $f \in C(\sigma(T))$ and write $zf(z) = g(z)$. Then

$$U_x T U_x^{-1} f = U_x T f(T)x = U_x g(T)x = g = M_z f,$$

where M_z denotes the operator of multiplication by z (here: in $L^2(\sigma(T), d\mu_{x,x})$). Since these functions f are dense in $L^2(\sigma(T), d\mu_{x,x})$ and both operators $U_x T U_x^{-1}$ and M_z are continuous, it follows that $U_x T U_x^{-1} = M_z$.

We now consider those collections of such spaces $\{H_x : x \in I\}$ for which the individual spaces are orthogonal: $H_x \perp H_y$ if $x, y \in I$, $x \neq y$. One can now use Zorn's Lemma to show that there is such collection of H_x spaces so that $\bigoplus_{x\in I} H_x = H$. As always, we don't want to discuss the details of this argument. The crucial fact is this: If $\bigoplus_{x\in I} H_x \neq H$, then there is another space H_y that is orthogonal to all H_x ($x \in I$). This can be proved as follows: Just pick an arbitrary $y \in \left(\bigoplus_{x\in I} H_x\right)^{\perp}$, $y \neq 0$. Then $\langle y, g(T)x\rangle = 0$ for all $x \in I$ and continuous functions g. But then it also follows that for all continuous f

$$\langle f(T)y, g(T)x\rangle = \langle y, \overline{f}(T)g(T)x\rangle = 0,$$

because $\overline{f}g$ is another continuous function. So $f(T)y \perp H_x$ and thus $H_y \perp H_x$ by the continuity of the scalar product.

We can now define the unitary map U as $U = \bigoplus_{x\in I} U_x$, where I is chosen so that $\bigoplus_{x\in I} H_x = H$, as discussed in the preceding paragraph. More precisely, by this we mean the following:

$$U : H \to \bigoplus_{x\in I} L^2(\sigma(T), d\mu_{x,x}),$$

and if $y = \sum_{x\in I} y_x$ is the unique decomposition of $y \in H$ into components $y_x \in H_x$, then we put $(Uy)_x = U_x y_x$. This map has the desired properties.

Exercise 10.12. Check this in greater detail.

□

We have now discussed three versions of the Spectral Theorem. We originally obtained the functional calculus for normal operators from the theory of C^*-algebras, especially the Gelfand-Naimark Theorem. This was then used to derive the existence of a spectral resolution E and a spectral representation. Conversely, spectral resolutions can be used to construct (in fact: an extended version of) the functional calculus, and it is also easy to recover E, starting from a spectral representation $UTU^{-1} = M_z$ (we sometimes write this as $T \cong M_z$). We summarize symbolically:

$$\boxed{\text{functional calculus} \quad\Longleftrightarrow\quad T = \int_{\sigma(T)} z\, dE(z) \quad\Longleftrightarrow\quad T \cong M_z}$$

Every version has its merits, and it's good to have all three statements available. Note, however, that the original functional calculus (obtained from the theory of C^*-algebras) becomes superfluous now because we obtain more powerful versions from the other statements (this was already pointed out above).

The spectrum of T will not always be known, and so it is sometimes more convenient to have statements that do not explicitly involve $\sigma(T)$. This is very easy to do: Given E, we can also get a spectral resolution on the Borel sets of $\mathbb{C}$ by simply declaring $E(\mathbb{C}\setminus\sigma(T))=0$. Similarly, in a spectral representation, we can think of the ρ_α as measures on $\mathbb{C}$ (with $\rho_\alpha(\mathbb{C}\setminus\sigma(T))=0$).

In this case, we can recover the spectrum from the measures ρ_α. We discuss the case of one space $L^2(\mathbb{C},d\rho)$ and leave the discussion of the effect of the orthogonal sum to an exercise. Given a Borel measure ρ on $\mathbb{C}$, we define its *topological support* as the smallest closed set A that supports ρ in the sense that $\rho(A^c)=0$. We denote it by $A=\text{top supp}\,\rho$.

Exercise 10.13. Prove that such a set exists. *Suggestion:* It is tempting to try to define $(\text{top supp}\,\rho)^c=\bigcup U$, where the union is over all open sets $U\subset\mathbb{C}$ with $\rho(U)=0$. This works, but note that the union will be uncountable, which could be a minor nuisance because we want to show that it has ρ measure zero.

Proposition 10.8. *If $T=M_z$ on $L^2(\mathbb{C},d\rho)$, then $\sigma(T)=\text{top supp}\,\rho$.*

Proof. Abbreviate $S=\text{top supp}\,\rho$. We must show that M_z-w is invertible in $B(L^2)$ precisely if $w\notin S$. Now if $w\notin S$, then $|w-z|\ge\epsilon>0$ for ρ-almost every $z\in\mathbb{C}$ (by definition of S), and this implies that $M_{(z-w)^{-1}}$ is a bounded linear operator. Obviously, it is the inverse of M_z-w.

Conversely, if $w\in S$, then $\rho(B_n)>0$ for all $n\in\mathbb{N}$, where $B_n=\{z\in\mathbb{C}:|z-w|<1/n\}$. Again, this follows from the definition of S. This means that $\|\chi_{B_n}\|>0$ in $L^2(\mathbb{C},d\rho)$. Let $f_n=\chi_{B_n}/\|\chi_{B_n}\|$, so $\|f_n\|=1$. Then $\|(M_z-w)f_n\|<1/n$, and this shows that (M_z-w) is not invertible: if it were, then it would follow that

$$1=\|f_n\|=\|(M_z-w)^{-1}(M_z-w)f_n\|\le C\|(M_z-w)f_n\|<\frac{C}{n},$$

a statement that seems hard to believe. □

As for the orthogonal sum, we have the following result:

Proposition 10.9. *Let H_α be Hilbert spaces, and let $T_\alpha \in B(H_\alpha)$, with* $\sup_{\alpha\in I} \|T_\alpha\| < \infty$. *Write $H = \bigoplus_{\alpha \in I} H_\alpha$ and define $T : H \to H$ as follows: $(Tx)_\alpha = T_\alpha x_\alpha$ (if $x = (x_\alpha)_{\alpha\in I}$). Then $T \in B(H)$ and*

$$\sigma(T) = \overline{\bigcup_{\alpha\in I} \sigma(T_\alpha)}.$$

It is customary to write this operator as $T = \bigoplus_{\alpha\in I} T_\alpha$, and actually we already briefly mentioned this notation in the proof of Theorem 10.7. If I is finite, then no closure is necessary in the statement of Proposition 10.9.

The situation of Theorem 10.7 is as discussed in the Proposition, with $T_\alpha = M_z$ for all α. So we can now say that the spectrum of M_z on $\bigoplus L^2(\mathbb{C}, d\rho_\alpha)$ is the closure of the union of the topological supports of the ρ_α.

Exercise 10.14. Prove Proposition 10.9.

The following basic facts are very useful when dealing with spectral representations. They provide further insight into the functional calculus and also a very convenient way of performing these operations once a spectral representation has been found.

Proposition 10.10. *Let $f : \mathbb{C} \to \mathbb{C}$ be a bounded Borel function. Then:*
(a) $f(M_z) = M_{f(z)}$;
(b) Let $U : H_1 \to H_2$ be a unitary map and let $T \in B(H_1)$ be a normal operator. Then

$$f(UTU^{-1}) = Uf(T)U^{-1}.$$

Sketch of proof. We argue as in the second part of the proof of Theorem 10.5. First of all, the assertions hold for functions of the type $f(z) = p(z, \overline{z})$, with a polynomial p, because for such functions we have an alternative direct description of $f(T)$, which lets us verify (a), (b) directly. Again, by the Stone-Weierstraß Theorem, these functions are dense in $C(K)$ for compact subsets $K \subset \mathbb{C}$. Since $f_n(T) \to f(T)$ in $B(H)$ if $\|f_n - f\|_\infty \to 0$, this gives the claim for continuous functions. Now $\|(f(T) - g(T))x\|^2 = \int |f - g|^2\, d\mu_{x,x}$ and continuous functions are dense in L^2 spaces. From this, we obtain the statements for arbitrary bounded Borel functions. □

Exercise 10.15. Give a detailed proof by filling in the details.

If T is of the form M_z on $L^2(\mathbb{C}, d\rho)$, as in a spectral representation (where we assume, for simplicity, that there is just one L^2 space), what is the spectral resolution E of this operator? In general, we can recover E from T as $E(A) = \chi_A(T)$, so Proposition 10.10 shows that $E(A) = M_{\chi_A}$ if $A \subset \mathbb{C}$ is a Borel set.

Exercise 10.16. Verify directly that this defines a resolution of the identity on the Borel sets of $\mathbb{C}$ (and the Hilbert space $L^2(\mathbb{C}, d\rho)$).

We observed earlier that the spectral measures ρ_α are (in fact: highly) non-unique. The following Exercise helps to clarify the situation. We call two operators $T_j \in B(H_j)$ *unitarily equivalent* if $T_2 = UT_1U^{-1}$ for some unitary map $U : H_1 \to H_2$. So, if we use this terminology, then Theorem 10.7 says that every normal operator is unitarily equivalent to the operator of multiplication by the variable in a sum of spaces $L^2(\mathbb{C}, \rho_\alpha)$.

Exercise 10.17. Consider the multiplication operators $T_1 = M_z^{(\mu)}$ and $T_2 = M_z^{(\nu)}$ on $L^2(\mu)$ and $L^2(\nu)$, respectively, where μ, ν are finite Borel measures on $\mathbb{C}$. Show that T_1, T_2 are unitarily equivalent if and only if μ and ν are equivalent measures (that is, they have the same null sets). *Suggestion:* For one direction, use the fact that μ and ν are equivalent if and only if $d\mu = f\, d\nu$, with $f \in L^1(\nu)$ and $f > 0$ almost everywhere with respect to μ (or ν).

Example 10.1. Let us now discuss the operator $(Tx)_n = x_{n+1}$ on $\ell^2(\mathbb{Z})$. By Exercise 6.7(a), T is unitary, so the Spectral Theorem applies. It is easiest to start out with a spectral representation because this can be guessed directly. Consider the operator

$$F : L^2(S, dx/(2\pi)) \to \ell^2(\mathbb{Z}), \quad (Ff)_n = \frac{1}{2\pi}\int_0^{2\pi} f(e^{ix})e^{inx}\, dx$$

(F as in *Fourier transform*). Here, $S = \{z \in \mathbb{C} : |z| = 1\}$ denotes again the unit circle; when convenient, we also use $x \in [0, 2\pi)$ to parametrize S by writing $z = e^{ix}$. Note that $(Ff)_n = \langle e_n, f\rangle$, with $e_n(z) = z^{-n}$. Since these functions form an ONB (compare Exercise 5.15), Theorem 5.14 shows that F is unitary.

Observe that the function $g(z) = zf(z)$ has Fourier coefficients $(Fg)_n = (Ff)_{n+1}$. In other words, $F^{-1}TF = M_z$, and this is a spectral representation, with $U = F^{-1}$. The spectral measure $dx/(2\pi)$ has the unit circle as its topological support, so $\sigma(T) = S$. Since only one L^2 space is necessary here, the operator T has spectral multiplicity one.

What is the spectral resolution of T? We already identified this spectral resolution on $L^2(S, dx/(2\pi))$, the space from the spectral representation, and we can now map things back to the original Hilbert space $\ell^2(\mathbb{Z})$ by using Proposition 10.10. More specifically,

$$E(A) = \chi_A(T) = \chi_A(FM_zF^{-1}) = F\chi_A(M_z)F^{-1} = FM_{\chi_A(z)}F^{-1}.$$

We can rewrite this if we recall that $(F^{-1}y)(z) = \sum y_n z^{-n}$, so $(M_{\chi_A} F^{-1}y)(z) = \sum y_n \chi_A(z) z^{-n}$ (both series converge in $L^2(S)$), and thus

$$(E(A)y)_n = \sum_{m=-\infty}^{\infty} \widehat{\chi_A}(m-n) y_m,$$

where $\widehat{\chi_A}(k) = 1/(2\pi) \int_0^{2\pi} \chi_A(e^{ix}) e^{ikx}\, dx$. Formally, this follows immediately from the preceding formulae, and for a rigorous argument, we use the fact that $(Ff)_n$ may be interpreted as a scalar product, the continuity of the scalar product and the L^2 convergence of the series that are involved here.

We now prove some general statements that illustrate how the Spectral Theorem helps to analyze normal operators.

Theorem 10.11. *Let $T \in B(H)$ be normal. Then:*
(a) T is self-adjoint $\iff \sigma(T) \subset \mathbb{R}$;
(b) T is unitary $\iff \sigma(T) \subset S = \{z \in \mathbb{C} : |z| = 1\}$.

The assumption that T is normal is needed here: if, for example, $T = \left(\begin{smallmatrix} 0 & 1 \\ 0 & 0 \end{smallmatrix}\right) \in B(\mathbb{C}^2)$, then $\sigma(T) = \{0\} \subset \mathbb{R}$, but T is not self-adjoint.

Proof. (a) $\Longrightarrow$: This was established earlier, in Theorem 9.15(a).
$\Longleftarrow$: By the Spectral Theorem and functional calculus,

$$T^* = \int_{\sigma(T)} \overline{z}\, dE(z) = \int_{\sigma(T)} z\, dE(z) = T.$$

(b) $\Longleftarrow$: This follows as in (a) from

$$TT^* = T^*T = \int_{\sigma(T)} z\overline{z}\, dE(z) = \int_{\sigma(T)} dE(z) = 1.$$

$\Longrightarrow$: If $z \in \sigma(T)$, then $E(B_{1/n}(z)) \neq 0$ for all $n \in \mathbb{N}$ by Proposition 10.6, so we can pick $x_n \in R(E(B_{1/n}(z)))$, $\|x_n\| = 1$. Then

$$\begin{aligned} \mu_{x_n,x_n}((B_{1/n}(z))^c) &= \langle x_n, E((B_{1/n}(z))^c) x_n \rangle \\ &= \langle x_n, E((B_{1/n}(z))^c) E(B_{1/n}(z)) x_n \rangle = 0, \end{aligned}$$

so it follows that
(10.8)
$$\big| \|Tx_n\| - |z|\, \|x_n\| \big|^2 \le \|(T-z)x_n\|^2 = \int_{\sigma(T)} |t-z|^2\, d\mu_{x_n,x_n}(t) \le \frac{1}{n^2}.$$

Since $\|Ty\| = \|y\|$ for all $y \in H$ for a unitary operator, this shows that $|z| = 1$, as claimed. ∎

Theorem 10.12. *If $T \in B(H)$ is normal, then*
$$\|T\| = \sup_{\|x\|=1} |\langle x, Tx\rangle| .$$

Proof. Clearly, $|\langle x, Tx\rangle| \le \|T\|\,\|x\|^2$, so the sup is $\le \|T\|$. On the other hand, we know from Theorem 9.15(b) that $\|T\| = r(T)$, so there exists a $z \in \sigma(T)$ with $|z| = \|T\|$. As in the previous proof, if $\epsilon > 0$ is given, then $E(B_\epsilon(z)) \ne 0$, so we can find an $x \in R(E(B_\epsilon(z)))$, $\|x\| = 1$. Then
$$|\langle x, Tx\rangle - z| = |\langle x, (T-z)x\rangle| = \left| \int (t-z)\, d\|E(t)x\|^2 \right| < \epsilon$$
because (again, as in the previous proof) $\mu_{x,x}((B_\epsilon(z))^c) = 0$ (and $\mu_{x,x}(\mathbb{C}) = \|x\|^2 = 1$). So $\sup |\langle x, Tx\rangle| \ge \|T\| - \epsilon$, and $\epsilon > 0$ is arbitrary here. $\square$

Theorem 10.13. *Let $T \in B(H)$. Then $T \ge 0$ (in the C^*-algebra $B(H)$; see Definition 9.14) if and only if $\langle x, Tx\rangle \ge 0$ for all $x \in H$.*

Proof. If $T \ge 0$, then T is self-adjoint and $\sigma(T) \subset [0,\infty)$, so the Spectral Theorem shows that
$$\langle x, Tx\rangle = \int_{[0,\infty)} t\, d\mu_{x,x}(t) \ge 0$$
for all $x \in H$.

Conversely, if this condition holds, then in particular $\langle x, Tx\rangle \in \mathbb{R}$ for all $x \in H$, so $\langle x, T^*x\rangle = \langle Tx, x\rangle = \overline{\langle x, Tx\rangle} = \langle x, Tx\rangle$. Polarization now shows that $\langle x, T^*y\rangle = \langle x, Ty\rangle$ for all $x, y \in H$, that is, $T = T^*$ and T is self-adjoint.

Now if $t > 0$, then
$$t\|x\|^2 = \langle x, tx\rangle \le \langle x, (T+t)x\rangle \le \|x\|\,\|(T+t)x\|,$$
so it follows that
$$\|(T+t)x\| \ge t\|x\|. \tag{10.9}$$
This shows, first of all, that $N(T+t) = \{0\}$. Moreover, we also see from (10.9) that $R(T+t)$ is closed: if $y_n \in R(T+t)$, say $y_n = (T+t)x_n$ and $y_n \to y \in H$, then (10.9) shows that x_n is a Cauchy sequence, so $x_n \to x$ for some $x \in H$ and thus $y = (T+t)x \in R(T+t)$ also, by the continuity of $T+t$. Finally, we observe that $R(T+t)^\perp = N((T+t)^*) = N(T+t) = \{0\}$ (by Theorem 6.2). Putting things together, we see that $R(T+t) = H$, so $T+t$ is bijective and thus $-t \notin \sigma(T)$. This holds for every $t > 0$, so, since T is self-adjoint, $\sigma(T) \subset [0,\infty)$ and $T \ge 0$. $\square$

Theorem 10.14. *Let $T \in B(H)$, $T \ge 0$. Then there exists a unique $S \in B(H)$, $S \ge 0$ so that $S^2 = T$.*

Proof. Existence is very easy: By the Spectral Theorem, $T = \int_{[0,\infty)} t\,dE(t)$. The operator $S = \int_{[0,\infty)} t^{1/2}\,dE(t)$ has the desired properties (here, $t^{1/2}$ of course denotes the positive square root).

Uniqueness isn't hard either, but more technical, and we just sketch this part: If S_0 is another operator with $S_0 \ge 0$, $S_0^2 = T$, write $S_0 = \int_{[0,\infty)} s\,dE_0(s)$, so $T = \int_{[0,\infty)} s^2\,dE_0(s)$. Now we can run a "substitution" $s^2 = t$ (of sorts) and rewrite this as $T = \int_{[0,\infty)} t\,d\widetilde{E}_0(t)$, where $\widetilde{E}_0(M) = E_0(\{s^2 : s \in M\})$ (this part would need a more serious discussion if a full proof is desired). By the uniqueness of the spectral resolution E (see Theorem 10.5), $\widetilde{E}_0 = E$, and this will imply that $S_0 = S$. □

Exercise 10.18. Let $T \in \mathbb{C}^{n\times n}$ be a normal matrix with n distinct, nonzero eigenvalues. Show that there are precisely 2^n normal (!) matrices $S \in \mathbb{C}^{n\times n}$ with $S^2 = T$.

Exercise 10.19. Recall that $\sigma_p(T)$ was defined as the set of eigenvalues of T; equivalently, $z \in \sigma_p(T)$ precisely if $N(T - z) \ne \{0\}$. Show that if $T \in B(H)$ is normal, then $z \in \sigma_p(T)$ if and only if $E(\{z\}) \ne 0$ (here, as usual, E denotes the spectral resolution of T).

Exercise 10.20. Let $T \in B(H)$ be normal. Show that $z \in \sigma(T)$ if and only if there exists a sequence $x_n \in H$, $\|x_n\| = 1$, so that $(T - z)x_n \to 0$.

Exercise 10.21. Suppose that $T \in B(H)$ is both unitary and self-adjoint. Show that T is of the form $T = 2P - 1$, for some orthogonal projection P. Show also that, conversely, every such operator T is unitary and self-adjoint.
Suggestion: Use the Spectral Theorem and Theorem 10.11 for the first part.

Exercise 10.22. Let $T \in B(H)$. Recall that a closed subspace $M \subset H$ is called *invariant* if $TM \subset M$, that is, if $Tx \in M$ for all $x \in M$. Call M a *reducing subspace* if both M and $M^\perp$ are invariant. Show that if T is normal with spectral resolution E, then $R(E(B))$ is a reducing subspace for every Borel set $B \subset \mathbb{C}$.
Hint: $E(B) = \chi_B(T)$; now use the functional calculus.

MATH DEPARTMENT, UNIVERSITY OF OKLAHOMA, NORMAN, OK 73019
E-mail address: cremling@math.ou.edu
URL: www.math.ou.edu/~cremling