Introduction to Functional Analysis

Michael Müger

10.12.2020

Abstract

These are notes for the course Inleiding in de Functionaalanalyse, Autumn 2020/21 (14×90 min.). They are also recommended as background for my courses on Operator Algebras.

Contents

Part I: Basics 4

1 Rough introduction 4

2 Topological groups, fields, vector spaces 6

3 Metrizable and normed vector spaces 8
   3.1 Metrizable TVS . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 8
   3.2 Normed spaces . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 9
   3.3 Brief look at locally convex and Fréchet spaces . . . . . . . . . . . . . . . . . . 12

4 Normed and Banach space basics 14
   4.1 Why we care about completeness. Closed subspaces . . . . . . . . . . . . . . . 14
   4.2 Linear maps: bounded ⇔ continuous . . . . . . . . . . . . . . . . . . . . . . . . . 16
   4.3 Spaces of bounded linear maps. First glimpse of Banach algebras . . . . . . . . 18

5 The sequence spaces and their dual spaces 20
   5.1 Basics. 1 ≤ p ≤ ∞: Hölder and Minkowski inequalities . . . . . . . . . . . . . . 20
   5.2 ∗ Aside: The translation-invariant metric d_p for 0 < p < 1 . . . . . . . . . . . . 22
   5.3 c_00 and c_0. Completeness of ℓ^p(S,F) and c_0(S,F) . . . . . . . . . . . . . . 23
   5.4 Separability of ℓ^p(S,F) and c_0(S,F) . . . . . . . . . . . . . . . . . . . . . . . . . 25
   5.5 Dual spaces of ℓ^p(S,F), 1 ≤ p < ∞, and c_0(S,F) . . . . . . . . . . . . . . . . 26
   5.6 Outlook on general L^p-spaces . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 28

6 Basics of Hilbert spaces 29
   6.1 Inner products. Cauchy-Schwarz inequality . . . . . . . . . . . . . . . . . . . . . 29
   6.2 The parallelogram and polarization identities . . . . . . . . . . . . . . . . . . . . 32
   6.3 Orthogonality, subspaces, orthogonal projections . . . . . . . . . . . . . . . . . . 33
      6.3.1 Basic Hilbert space geometry . . . . . . . . . . . . . . . . . . . . . . . . . . . 33
      6.3.2 Closed subspaces, orthogonal complement, and orthogonal projections . . . 34
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>6.4</td>
<td>The dual space $H^*$ of a Hilbert space</td>
<td>36</td>
</tr>
<tr>
<td>6.5</td>
<td>Orthonormal sets. Bases</td>
<td>36</td>
</tr>
<tr>
<td>6.6</td>
<td>Tensor products of Hilbert spaces</td>
<td>41</td>
</tr>
<tr>
<td>7</td>
<td>Quotient spaces and complemented subspaces</td>
<td>42</td>
</tr>
<tr>
<td>7.1</td>
<td>Quotient spaces of Hilbert spaces</td>
<td>42</td>
</tr>
<tr>
<td>7.2</td>
<td>Quotient spaces of Banach spaces</td>
<td>42</td>
</tr>
<tr>
<td>7.3</td>
<td>Complemented subspaces</td>
<td>44</td>
</tr>
<tr>
<td>8</td>
<td>Hahn-Banach theorem and its applications</td>
<td>45</td>
</tr>
<tr>
<td>8.1</td>
<td>First version of Hahn-Banach over $\mathbb{R}$</td>
<td>46</td>
</tr>
<tr>
<td>8.2</td>
<td>Hahn-Banach theorem for (semi)normed spaces</td>
<td>47</td>
</tr>
<tr>
<td>8.3</td>
<td>First applications</td>
<td>48</td>
</tr>
<tr>
<td>8.4</td>
<td>Reflexivity of Banach spaces</td>
<td>49</td>
</tr>
<tr>
<td>9</td>
<td>Uniform boundedness theorem: Two versions</td>
<td>50</td>
</tr>
<tr>
<td>9.1</td>
<td>The weak version, using only countable choice</td>
<td>50</td>
</tr>
<tr>
<td>9.2</td>
<td>Applications: Banach-Steinhaus, Hellinger-Toeplitz</td>
<td>51</td>
</tr>
<tr>
<td>9.3</td>
<td>The strong version, using Baire's theorem</td>
<td>53</td>
</tr>
<tr>
<td>9.4</td>
<td>Application: A dense set of continuous functions with divergent Fourier series</td>
<td>53</td>
</tr>
<tr>
<td>10</td>
<td>Open mappings and closed graphs</td>
<td>55</td>
</tr>
<tr>
<td>10.1</td>
<td>The Open Mapping Theorem</td>
<td>55</td>
</tr>
<tr>
<td>10.2</td>
<td>The Bounded Inverse Theorem</td>
<td>56</td>
</tr>
<tr>
<td>10.2.1</td>
<td>The Closed Graph Theorem</td>
<td>57</td>
</tr>
<tr>
<td>10.3</td>
<td>Boundedness below. Invertibility</td>
<td>59</td>
</tr>
<tr>
<td>Part II: Spectral theory of operators and algebras</td>
<td>60</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>Spectrum of bounded operators and of (elements of) Banach algebras</td>
<td>60</td>
</tr>
<tr>
<td>11.1</td>
<td>The spectra of $A \in B(E)$</td>
<td>60</td>
</tr>
<tr>
<td>11.2</td>
<td>The spectrum in a unital Banach algebra</td>
<td>62</td>
</tr>
<tr>
<td>11.3</td>
<td>Examples of spectra of operators</td>
<td>68</td>
</tr>
<tr>
<td>11.4</td>
<td>Characters. Spectrum of a Banach algebra</td>
<td>68</td>
</tr>
<tr>
<td>12</td>
<td>Transpose and adjoint of bounded operators. $C^*$-algebras</td>
<td>70</td>
</tr>
<tr>
<td>12.1</td>
<td>The transpose of a bounded Banach space operator</td>
<td>70</td>
</tr>
<tr>
<td>12.2</td>
<td>The adjoint of a bounded Hilbert space operator</td>
<td>71</td>
</tr>
<tr>
<td>12.3</td>
<td>Involutions. Definition of $C^*$-algebras</td>
<td>74</td>
</tr>
<tr>
<td>12.4</td>
<td>Spectrum of elements of a $C^*$-algebra</td>
<td>75</td>
</tr>
<tr>
<td>13</td>
<td>Functional calculus in Banach and $C^*$-algebras</td>
<td>77</td>
</tr>
<tr>
<td>13.1</td>
<td>Some functional calculus in Banach algebras</td>
<td>77</td>
</tr>
<tr>
<td>13.2</td>
<td>Continuous functional calculus for self-adjoint elements in a $C^*$-algebra</td>
<td>79</td>
</tr>
<tr>
<td>13.3</td>
<td>Continuous functional calculus for normal elements in a $C^*$-algebra</td>
<td>82</td>
</tr>
</tbody>
</table>
### 14 More on Hilbert space operators

14.1 Normal operators .................................................. 84
14.2 Self-adjoint operators ................................................ 86
14.3 Positive operators. Polar decomposition .......................... 87

### 15 Compact operators

15.1 Compact Banach space operators .................................. 88
15.2 Fredholm alternative. The spectrum of compact operators .... 93
15.3 Spectral theorems for compact Hilbert space operators ....... 95
15.4 * Hilbert-Schmidt operators ....................................... 97

### 16 Spectral theorems for normal Hilbert space operators

16.1 Spectral theorem: Multiplication operator version ............... 100
16.2 Borel functional calculus for normal operators .................. 103
16.3 Normal operators vs. projection-valued measures ............... 106

### 17 Weak and weak *-topologies. Alaoglu’s theorem

17.1 The weak topology of a Banach space ............................ 108
17.2 The weak-* topology on the dual space of a Banach space ..... 109

### 18 The Gelfand homomorphism for commutative Banach and C*-algebras

18.1 The topology of \( \Omega(A) \). The Gelfand homomorphism .... 111
18.2 Application: Absolutely convergent Fourier series ........... 113
18.3 C*-algebras. Continuous functional calculus revisited ....... 114

### A Some topics from topology and measure theory

A.1 Unordered sums ..................................................... 116
A.2 Nets ................................................................. 118
A.3 The Stone-Čech compactification .................................. 119
A.4 Reminder of the choice axioms and Zorn’s lemma ............... 120
A.5 Baire’s theorem ................................................... 121
A.6 Tietze’s extension theorem ....................................... 122
A.7 The Stone-Weierstrass theorem .................................... 122
A.7.1 Weierstrass’ theorem ........................................... 122
A.7.2 The Stone-Weierstrass theorem in the simplest case .......... 124
A.7.3 Generalizations .................................................. 125
A.8 Totally bounded sets in metric spaces ............................ 126
A.9 The Arzelà-Ascoli theorem ....................................... 126
A.10 Some notions from measure and integration theory ............. 128

### B Supplements for the curious. NOT part of the course

B.1 Functional analysis over fields other than \( \mathbb{R} \) and \( \mathbb{C} \)? .................................................. 129
B.2 The dual space of \( \ell^\infty(S, \mathcal{F}) \) .................................. 130
B.3 \( c_0(N, \mathcal{F}) \subseteq \ell^\infty(N, \mathcal{F}) \) is not complemented ........ 133
B.4 Banach spaces with no or multiple pre-duals ....................... 134
B.5 Normability. Separation theorems. Goldstine’s theorem ....... 135
B.5.1 Minkowski functionals. Criteria for normability and local convexity ................................. 135
B.5.2 Hahn-Banach separation theorem. Goldstine’s theorem .... 137
B.6 Strictly convex and uniformly convex Banach spaces .......... 138
1 Rough introduction

We will begin with a quick delineation of what we will discuss – and what not!

- “Classical analysis” is concerned with ‘analysis in finitely many dimensions’. ‘Functional analysis’ is the generalization or extension of classical analysis to infinitely many dimensions. Before one can try to make sense of this, one should make the first sentence more precise. Since the creation of general topology, one can talk about convergence and continuity in very general terms. As far as I see it, this is not analysis, even if infinite sums (=series) are studied. Analysis proper starts as soon as one talks about differentiation and/or integration. Differentiation has to do with approximating functions locally by linear ones, and for this one needs the spaces considered to be vector spaces (at least locally). This is the reason why most of classical analysis considers functions between the vector spaces $\mathbb{R}^n$ and $\mathbb{C}^n$ (or subsets of them). (In a second step, one can then generalize to spaces that look like $\mathbb{R}^n$ only locally by introducing topological and smooth manifolds and their generalizations, but the underlying model of $\mathbb{R}^n$ remains important.) On the other hand, integration, at least in the sense of the modern theory, can be studied much more generally, i.e. on arbitrary sets equipped with a measure (defined on some $\sigma$-algebra). Such a set can be very far from being a vector space or manifold, for example by being totally disconnected.

- In view of the above, it is not surprising that functional analysis is concerned with (possibly) infinite dimensional vector spaces and continuous maps between them. (Again, one can then generalize to spaces that look like a vector space only locally, but this would be considered infinite dimensional geometry, not functional analysis.) In addition to the vector space structure one needs a topology, which naturally leads to topological vector spaces, which I will define soon.

- The importance of topologies is not specific to infinite dimensions. The point rather is that $\mathbb{R}^n, \mathbb{C}^n$ have unique topologies making them topological vector spaces. This is no more true in infinite dimensions!

- Actually, ‘functional analysis’ most often studies only linear maps between topological vector spaces so that this domain of study should be called ‘linear functional analysis’, but this is done only rarely, e.g. [66]. Allowing non-linear maps leads to non-linear functional analysis. This course will discuss only linear functional analysis. Thorough mastery of the latter is needed anyway before one can think about non-linear FA or infinite dimensional geometry. For the simplest result of non-linear functional analysis, see Section B.9. For more, you could have a look at, e.g., [78, 13, 52]. There even is a five volume treatise [84]!
The restriction to linear maps means that the notion of differentiation becomes pointless, the derivative of \( f(x) = Ax + b \) being just \( A \) everywhere. But there are many non-trivial connections between linear FA and integration (and measure) theory. For example, every measure space \( (X, \mathcal{A}, \mu) \) gives rise to a family of topological vector spaces \( L^p(X, \mathcal{A}, \mu) \), \( p \in (0, \infty] \), and integration provides linear functionals. Proper appreciation of these matters requires some knowledge of measure and integration theory, cf. e.g. [10, 67]. I will not suppose that you have followed a course on this subject (but if you haven’t, I strongly that you do so on the next occasion or, at least, read the appendix in MacCluer’s book [41].). Yet, one can get a reasonably good idea by focusing on sequence spaces, for which no measure theory is required, see Section 5.

One should probably consider linear functional analysis as an infinite dimensional and topological version of linear algebra rather than as a branch of analysis! This might lead one to suspect linear FA to be slightly boring, but this would be wrong for many reasons:

- Functional analysis (linear or not) leads to very interesting (and arbitrarily challenging) technical questions (most of which reduce to very easy ones in finite dimensions).
- Linear FA is essential for non-linear FA, like variational calculus, and the theory of differential equations – not only linear ones!
- Quantum theory [37] is a linear theory and cannot be done properly without functional analysis, despite the fact that many physicists think so! Conversely, many developments in FA were directly motivated by quantum theory.

The above could give the impression that functional analysis arose from the wish of generalizing analysis to infinitely many dimensions. This may have played a role for some of its creators, but its beginnings (and much of what is being done now) were mostly motivated by finite dimensional “classical”¹ analysis: If \( U \subset \mathbb{R}^n \), the set of functions (possibly continuous, differentiable, etc.) from \( U \) to \( \mathbb{R}^m \) is a vector space as soon as we put \( (cf + dg)(x) = cf(x) + dg(x) \). Unless \( U \) is a finite set, this vector space will be infinite dimensional. Now one can consider certain operations on such vector spaces, like differentiation \( C^\infty(U) \rightarrow C^\infty(U), f \mapsto f' \) or integration \( f \mapsto \int_U f \). This sort of considerations provided the initial motivation for the development functional analysis, and indeed FA now is a very important tool for the study of ordinary and partial differential equations on finite dimensional spaces. See e.g. [6, 19]. The relevance of FA is even more obvious if one studies differential equations in infinitely many dimensions. In fact, it is often useful to study a partial differential equation (like heat or wave equation) by singling out one of the variables (typically ‘time’) and studying the equation as an ordinary differential equation in an infinite dimensional space of functions. FA is also essential for variational calculus (which in a sense is just a branch of differential calculus in infinitely many dimensions).

In view of the above, FA studies abstract topological vector spaces as well as ‘concrete’ spaces, whose elements are functions. In order to obtain a proper understanding of FA, one needs some familiarity with both aspects.

Before we delve into technicalities, some further general remarks:

- The history of functional analysis is quite interesting, cf. e.g. the article [4], [55, Chapter 4] and the books [14, 41]. But clearly it makes little sense to study it before one has some technical knowledge of FA. It is surprisingly intertwined with the development of linear

---

¹Ultimately, I find it quite futile to try and draw a neat line between “classical” and “modern” or functional analysis, in particular since many problems in the former require methods from the latter for their proper treatment.
algebra. One would think that (finite dimensional) vector spaces, linear maps etc. were
defined much earlier than, e.g., Banach spaces, but this is not true. In fact, Banach’s\(^2\)
book \[3\], based on his 1920 PhD thesis, is one of the first references containing the modern
definition of a vector space. Some mathematicians, mostly Italian ones, like Peano,
Pincherle and Volterra, essentially had the modern definition already in the last decades
of the 19th century, but they had no impact since the usefulness of an abstract/axiomatic
approach was not yet widely appreciated. Cf. \[36, \text{Chapter 5}\] or \[15, 45\].

Here I limit myself to mentioning that the basics of functional analysis (Hilbert and Banach
spaces and bounded linear maps between them) were developed in the period 1900-1930.
Nevertheless, many important developments (locally convex spaces, distributions, operator
algebras) took place in 1930-1960. After that, functional analysis has split up into many
very specialized subfields that interact quite little with each other. The very interesting
article \[79\] ends with the conclusion that ‘functional analysis’ has ceased to exist as a
coherent field of study!

- The study of functional analysis requires a solid background in general topology. It may
well be that you’ll have to refresh and likely also extend yours. In Appendix A I have
collected brief accounts of the topics that – sadly – you are most likely not to have en-
countered before. All of them are contained in \[63\] (written by a functional analyst!), but
my favorite (I’m admittedly biased) reference is \[46\]. You should have seen Weierstrass’
theorem, but those of Tietze and Arzelà-Ascoli tend to vanish in the (pedagogical, not
factual) gap between general topology and functional analysis.

- The main reference for this course has been \[41\] for a number of years, but I am unenthui-
siastic about a number of aspects of it, which is why I wrote these notes. If you find them
too advanced, you might want to have a look at \[53, 66, 68\]. On the other hand, if you
want more, \[54\] is a good place to start, followed by \[58, 38, 11, 62\]. (The MasterMath
course currently uses \[11\].)

One word about notation (without guarantee of always sticking to it): General vector spaces,
but also normed spaces, are denoted \(V, W, \ldots\), normed spaces also as \(E, F, \ldots\). Vectors in such
spaces are \(e, f, \ldots, x, y, \ldots\). Linear maps are always denoted \(A, B, \ldots\), except linear functionals
\(V \to F\), which are \(\varphi, \psi\). Algebras are usually denoted \(\mathcal{A}, \mathcal{B}, \ldots\) and their elements \(a, b, \ldots\). (For
\(\mathcal{A} = B(E)\) this leads to inconsistency, but I cannot bring myself to using capital letters for
abstract algebra elements.)

\textit{Acknowledgment.} I thank Bram Balkema, Victor Hissink Muller, Tim Peters and Niels Vooijs
for many corrections.

\section{Topological groups, fields, vector spaces}

As said in the Introduction, functional analysis (even most of the non-linear version) is concerned
with vector spaces, allowing infinite dimensional ones. Large parts of linear algebra of course
work equally well for finite and infinite dimensional spaces. One aspect where problems arise in
infinite dimensions is the description of linear maps by matrices, for example since multiplication
of infinite matrices involves infinite summations, which require the introduction of topologies.
(Actually, in some restricted contexts infinite matrices still are quite useful.)

\(^2\)Stefan Banach (1892-1945). Polish mathematician and pioneer of functional analysis. Also known for B. algebras,
B.’s contraction principle, the B.-Tarski paradox and the Hahn-B. and B.-Steinhaus theorems, etc.
We begin with the following

2.1 Definition A topological group is a group \((G, \cdot, 1)\) equipped with a topology \(\tau\) such that the group operations \(G \times G \to G, (g, h) \mapsto gh\) and \(G \to G, g \mapsto g^{-1}\) are continuous (where \(G \times G\) is given the product topology). (For abelian groups, one often denotes the binary operation by + instead of \(\cdot\).)

2.2 Example 1. If \((G, \cdot, 1)\) is any group then it becomes a topological group by putting \(\tau = \tau_{\text{disc}}\), the discrete topology on \(G\).

2.2 Example 2. The group \((\mathbb{R}, +, 0)\), where \(\mathbb{R}\) is equipped with its standard topology, is easily seen to be a topological group.

3. If \(n \in \mathbb{N}\) and \(F \in \{\mathbb{R}, \mathbb{C}\}\) then the set \(GL(n, F) = \{A \in M_{n \times n}(F) \mid \det(A) \neq 0\}\) of invertible \(n \times n\) matrices is a group w.r.t. matrix product and inversion and in fact a topological group when equipped with the subspace topology induced from \(M_{n \times n}(F) \cong F^{n^2}\).

2.3 Remark Topological groups – or rather matrix groups as in 3. above – are the subject of my 3rd year course on course Continuous Matrix Groups, taught again next spring. They are an important (and prototypical) case of Lie groups. The latter are a subject at Master level that is very much worthy of study!

2.4 Definition A topological field is a field \((F, +, 0, \cdot, 1)\) equipped with a topology on \(F\) such that \((F, +, 0)\) and \((F \setminus \{0\}, \cdot, 1)\) are topological groups. (Equivalently, all field operations are continuous.)

It is very easy to check that \(\mathbb{R}\) and \(\mathbb{C}\) are topological fields when equipped with their standard topologies. (So is \(\mathbb{Q}\) with the topology induced from \(\mathbb{Q}\).)

2.5 Exercise Prove the above claims.

2.6 Definition Let \(F\) be a topological field. Then a topological vector space (TVS) over \(F\) is an \(F\)-vector space equipped with a topology \(\tau_V\) (to be distinguished, obviously, from the topology \(\tau_F\) on \(F\)) such that the maps \(V \times V \to V, (x, y) \mapsto x + y\) and \(F \times V \to V, (c, x) \mapsto cx\) are continuous.

(These conditions imply that \(V \to V, x \mapsto -x\) is continuous, so that \((V, +, 0)\) is a topological group, but not conversely.)

Again it is very easy to check that \(\mathbb{R}^n\) and \(\mathbb{C}^n\) are topological vector spaces over the topological fields \(\mathbb{R}, \mathbb{C}\), respectively, when they are equipped with the euclidean topologies (=product topologies on \(F \times \cdots \times F\)).

In this course, the only topological fields considered are \(\mathbb{R}\) and \(\mathbb{C}\). When a result holds for either of the two, I will write \(\mathbb{F}\). But note that one can consider topological vector spaces over other topological fields, like the \(p\)-adic ones \(\mathbb{Q}_p\) [23]. (But the resulting \(p\)-adic functional analysis is quite different in some respects from the ‘usual’ one, cf. the comments in Section B.1 and the literature, e.g. [60, 56].)

2.7 Exercise Let \(F\) be a topological field and \(V\) an \(F\)-vector space. Is it true that \(V\), equipped with the discrete topology, is a topological vector space over \(F\)? Prove or give a counterexample.

Now we can define:
2.8 Definition Functional analysis (ordinary, as opposed to $p$-adic) is concerned with topological vector spaces over $\mathbb{R}$ or $\mathbb{C}$ and continuous maps between them. Linear functional analysis considers only linear maps.

As it turns out, the above notion of topological vector spaces is a bit too general to build a satisfactory and useful theory upon it. Just as in topology it is often (but by no means always!) sufficient to work with metric spaces, for most purposes it is usually sufficient to consider certain subclasses of topological vector spaces. The following diagram illustrates some of these classes and their relationships:

\[
\text{topological vector sp.} \supset \text{metrized/F-sp.} \\
\quad \cup \\
\text{locally convex sp.} \supset \text{Fréchet sp.} \supset \text{normed/Banach sp.} \supset (\text{pre})\text{Hilbert sp.}
\]

(Note that F-spaces, Fréchet, Banach and Hilbert spaces are assumed complete but one also has the non-complete versions. There is no special name for Fréchet spaces with completeness dropped other than metrizable locally convex spaces. In the other cases, one speaks of metrized, normed and pre-Hilbert spaces.)

The most useful of these classes are those in the bottom row. In fact, locally convex (vector) spaces are general enough for almost all applications. They are thoroughly discussed in the MasterMath course on functional analysis, while we will only briefly touch upon them. Most of the time, we will be discussing Banach and Hilbert spaces. There is much to be said for studying them in some depth before turning to locally convex spaces (or more general) spaces. (Some books on functional analysis, like [62], begin with general topological vector spaces and then turn to some special classes, but for a first encounter this does not seem appropriate. This said, I don’t see the point of beginning with proofs of many results on Hilbert spaces that literally generalize to Banach spaces.)

3 Metrizable and normed vector spaces

3.1 Metrizable TVS

I assume that you remember the notion of a metric on a set $X$: A map $d : X \times X \to [0, \infty)$ satisfying $d(x, y) = d(y, x)$ and $d(x, z) \leq d(x, y) + d(y, z)$ for all $x, y, z \in X$ and $d(x, y) = 0 \iff x = y$. Every metric $d$ on $X$ defines a topology $\tau_d$ on $X$, the smallest topology $\tau$ containing all open balls $B(x, r) = \{y \in X \mid d(x, y) < r\}$. (The open balls then form a base, not just a subbase, for $\tau$.) A topology $\tau$ on $X$ is called metrizable if there exists a metric $d$ on $X$ (not necessarily unique) such that $\tau = \tau_d$. Metrizable topologies automatically have many nice properties, like e.g. normality and, a fortiori, the Hausdorff property.

3.1 Remark Dropping the requirement that $d(x, y) = 0 \Rightarrow x = y$ (but keeping $d(x, x) = 0 \forall x$) one arrives at the notion of a pseudo-metric. Also a pseudometric $d$ defines a topology $\tau_d$ as above, and it is easy to check that $\tau_d$ is Hausdorff if and only if $d(x, y) = 0$ implies $x = y$, thus $d$ is a metric. (If $x \neq y$ but $d(x, y) = 0$ then $x$ is contained in any open ball $B(y, r)$ and vice versa so that there cannot be disjoint open neighborhoods for $x, y$.)

---


4 David Hilbert (1862-1943). Eminent German mathematician who worked on many different subjects. Considered the strongest and most influential mathematician in the decades around 1900, only Poincaré coming close.
The notion of a metric as defined above is not too well adapted to the concept of a vector space since it does not take the vector space structure into account. This motivates:

3.2 Definition 1. Let $V$ be a vector space and $d$ a (pseudo)metric on it. Then $d$ is called translation invariant if one of the following equivalent conditions holds:

$$d(x, y) = d(x - z, y - z) \quad \forall x, y, z \in V$$

$$d(x, y) = d(x - y, 0) \quad \forall x, y \in V.$$  

2. A topological vector space $(V, \tau)$ is called metrizable if there exists a translation-invariant metric $d$ on $V$ such that $\tau = \tau_d$. An $F$-space is a TVS that is metrizable by a translation-invariant and complete metric.

3.3 Remark 1. Every $\mathbb{R}$-vector space equipped with the indiscrete topology is a TVS. (Check this!) Since the indiscrete topology (on a space with more than one point) is not metrizable, this gives an example of a TVS that is not metrizable.

2. There is a nice necessary and sufficient condition for metrizability of a TVS: It must be Hausdorff and the zero element must have a countable base of open neighborhoods, cf. [62, Theorem 1.24]. We will see that every normed topological space is metrizable, but that the converse is not true.

3.4 Lemma Let $V$ be an $F$-vector space and $d$ a translation-invariant metric on it. Then equipped with the topology $\tau_d$, the abelian group $(V, +, 0)$ is a topological group. But $(V, \tau_d)$ may fail to be a TVS.

Proof. Using the translation invariance of $d$ and the triangle inequality we have

$$d(x + y, x' + y') = d(x + y - x' - y', x' + y' - y - x') = d(x - x', y' - y) \leq d(x - x', 0) + d(0, y - y') = d(x, x') + d(y', y).$$

This shows that if $(x', y')$ converges to $(x, y)$ in $V \times V$ then $x' + y'$ converges to $x + y$, so that $V \times V \rightarrow V, (x, y) \mapsto x + y$ is jointly continuous. And $d(-x, -y) = d(-x + y, 0) = d(0, -x + y) = d(x, y)$ gives continuity of the map $x \mapsto -x$. Thus $(V, +, 0)$ is a topological group (clearly abelian).

Now let $V$ be an $\mathbb{R}$-vector space of dimension $\geq 1$ and $d$ the discrete metric: $d(x, x) = 0$ $\forall x$ and $d(x, y) = 1$ for $x \neq y$. Then $\tau_d$ is the discrete topology. Then continuity of addition is obvious (or apply the above). Since $(V, \tau_d)$ is discrete and $\mathbb{R}$ connected (with the usual topology), every continuous map $F \rightarrow V$ must be constant. (Can you prove this?) But for any $0 \neq x \in V$, the map $F \rightarrow V, c \mapsto cx$ is not constant (since $0x = 0 \neq x = 1x$) and therefore not continuous. Thus also $F \times V \rightarrow V, (c, x) \mapsto cx$ certainly is not jointly continuous.

Metrizable TVS are better behaved than general TVS, but can still be quite pathological. (See Section 5.2 for examples.) Topologies coming from (semi)norms are much better.

3.2 Normed spaces

3.5 Definition Let $V$ be a vector space over $F \in \{\mathbb{R}, \mathbb{C}\}$. A seminorm on $V$ is a map $V \rightarrow [0, \infty), x \mapsto \|x\|$ such that

$$\|x + y\| \leq \|x\| + \|y\| \quad \forall x, y \in V.$$  

$$\|cx\| = |c|\|x\| \quad \forall c \in F, x \in V.$$  

(Subadditivity)

(This implies $\|0\| = 0$ and $\|-x\| = \|x\|$.)
(Note that by the above, \( \|x\| = \infty \) is not allowed!) A norm is a seminorm satisfying also \( \|x\| = 0 \Rightarrow x = 0 \).

A normed \( \mathbb{F} \)-vector space is an \( \mathbb{F} \)-vector space equipped with a norm.

3.6 Example 0. Clearly \( \mathbb{F} \in \{ \mathbb{R}, \mathbb{C} \} \) is a vector space over itself and \( \|c\| := |c| \) defines a norm, making \( \mathbb{F} \) a complete normed \( \mathbb{F} \)-vector space.

1. Let \( X \) be a compact topological space and \( V = C(X, \mathbb{F}) \). Clearly, \( V \) is an \( \mathbb{F} \)-vector space. Now \( \|f\| = \sup_{x \in X} |f(x)| \) is a norm on \( V \). You probably know that the normed space \( (V, \| \cdot \|) \) is complete.

If \( X \) is non-compact then \( \|f\| \) can be infinite, but replacing \( C(X, \mathbb{F}) \) by

\[
C_b(X, \mathbb{F}) = \{ f \in C(X, \mathbb{F}) \mid \|f\| < \infty \},
\]

\( \| \cdot \| \) again is a norm with which \( C_b(X, \mathbb{F}) \) is complete.

2. Let \( n \in \mathbb{N} \) and \( V = \mathbb{C}^n \). For \( x \in V \) and \( 1 \leq p < \infty \) (NB: \( p \) does not stand for prime!), define

\[
\|x\|_\infty = \max_{i=1,...,n} |x_i|, \quad \|x\|_p = \left( \sum_{i=1}^n |x_i|^p \right)^{1/p}.
\]

(Note that all these \( \| \cdot \|_p \) including \( p = \infty \) coincide if \( n = 1 \).) It is quite obvious that for each \( p \in [1, \infty] \) we have \( \|x\|_p = 0 \iff x = 0 \) and \( \|cx\|_p = |c| \|x\|_p \). For \( p = 1 \) and \( p = \infty \) also the subadditivity is trivial to check using only \( |c + d| \leq |c| + |d| \). Subadditivity also holds for \( 1 < p < \infty \), but is harder to prove. You have probably seen the proof for \( p = 2 \), which relies on the Cauchy-Schwarz inequality. The proof for \( 1 < p < 2 \) and \( 2 < p < \infty \) is similar, using the inequality of Hölder instead. We will return to this and also prove that \( \mathbb{R}^n, \mathbb{C}^n \) is complete w.r.t. any of the norms \( \| \cdot \|_p, p \in [1, \infty] \).

3. The above examples are easily generalized to infinite dimensions. Let \( S \) be any set. For a function \( f : S \to \mathbb{F} \) and \( 1 \leq p < \infty \) define

\[
\|f\|_\infty = \sup_{s \in S} |f(s)|, \quad \|f\|_p = \left( \sum_{s \in S} |f(s)|^p \right)^{1/p}
\]

with the understanding that \( (+\infty)^{1/p} = +\infty \). For the definition of infinite sums like \( \sum_{s \in S} f(s) \) see Appendix A.1. Now let

\[
\ell^p(S, \mathbb{F}) = \{ f : S \to \mathbb{F} \mid \|f\|_p < \infty \}.
\]

Now one can prove that \( \| \cdot \|_p \) is a complete norm on \( (\ell^p(S, \mathbb{F}), \| \cdot \|_p) \) for each \( p \in [1, \infty] \). We will do this in Section 5.

4. Let \( (X, \mathcal{A}, \mu) \) be a measure space, \( f : X \to \mathbb{F} \) measurable and \( 1 \leq p < \infty \). We define

\[
\|f\|_p = \left( \int_X |f|^p d\mu \right)^{1/p}, \quad \|f\|_\infty = \inf\{ M > 0 \mid \mu(\{ x \in X \mid |f(x)| > M \}) = 0 \}
\]

and

\[
\mathcal{L}^p(X, \mathcal{A}, \mu; \mathbb{F}) = \{ f : X \to \mathbb{F} \text{ measurable} \mid \|f\|_p < \infty \}.
\]
Then \( \|f\|_p = (\int_X |f|^p d\mu)^{1/p} \) is a seminorm on \( \mathcal{L}^p(X, \mathcal{A}, \mu; \mathbb{F}) \) for all \( 1 \leq p < \infty \).

However, in general \( \| \cdot \|_p \) it is not a norm since \( \|f\|_p \) vanishes whenever \( f \) is zero almost everywhere, i.e. \( \mu(f^{-1}(\mathbb{C}\setminus\{0\})) = 0 \), which may well happen even if \( f \) is not identically zero. In order to obtain a normed space one considers the quotient space \( \mathcal{L}^p(X, \mathcal{A}, \mu)/\{f \in \mathcal{L}^p \mid \|f\|_p = 0 \} \). Going into the details would require too much measure theory. See [41, Appendices A.1-A.3] for a crash course or [10, 67] for the full story.

There is an instructive special case: If \( S \) is a set, \( \mathcal{A} = P(S) \) and \( \mu(A) = \#A \) (the counting measure) then for every \( f : S \to \mathbb{F} \) we have \( \int_S f(s) d\mu(s) = \sum_{s \in S} f(s) \), where the integral, like the (unordered) sum, exists if and only if \( \sum_{s \in S} |f(s)| < \infty \). Thus \( \mathcal{L}^p(S, \mathcal{A}, \mu; \mathbb{F}) = \ell^p(S, \mathbb{F}) \).

3.7 Lemma Let \((V, \| \cdot \|)\) be a normed \( \mathbb{F}\)-vector space, and define \( d(x, y) = \|x - y\| \). Then

(i) \( d \) is a translation-invariant metric on \( V \).

(ii) \((V, \tau_d)\) is a topological vector space.

Proof. (i) That norms give rise to metrics should be known from topology: \( d(x, y) \geq 0 \) follows from \( \|x\| \geq 0 \), and \( d(x, y) = 0 \iff x = y \) follows from the norm axiom \( \|x\| = 0 \iff x = 0 \).

Furthermore, \( d(y, x) = \|y - x\| = \| - (x - y)\| = \|x - y\| = d(x, y) \), where we used \( \|-x\| = \|x\| \), a special case of the second seminorm axiom. Finally,

\[
d(x, z) = \|x - z\| = \|x - y + y - z\| \leq \|x - y\| + \|y - z\| = d(x, y) + d(y, z).
\]

Translation invariance is obvious: \( d(x, y) := \|x - y\| = d(x - y, 0) \).

(ii) By (i), \( d \) is a translation invariant metric, thus \(+ : V \times V \to V\) is continuous by Lemma 3.4. For continuity of the scalar action \( \mathbb{F} \times V \to V \), let \( x, x' \in V, c, c' \in \mathbb{F} \). Then

\[
\|cx - c'x'\| = \|cx - cx' + cx' - c'x'\| \leq \|c(x - x')\| + \|(c - c')x'\| = |c| \|x - x'\| + |c - c'| \|x'\|.
\]

Or, in terms of the metric: \( d(cx, c'x') \leq |c|d(x, x') + |c - c'|\|x'\| \). Thus as \((c', x')\) converges to \((c, x)\) in \( \mathbb{F} \times V \), we have convergence of \( c'x' \) to \( cx \), thus joint continuity of \( \mathbb{F} \times V \to V \), \((c, x) \mapsto cx\).

3.8 Definition A topological vector space \((V, \tau)\) is called normable if there exists a norm \( \| \cdot \| \) on \( V \) such that \( \tau = \tau_d \) with \( d(x, y) = \|x - y\| \).

3.9 Remark One can prove, see the supplementary Appendix B.5.1, that a topological vector space \((V, \tau)\) is normable if and only if there is an open \( U \subseteq V \) such that \( 0 \in U \) and \( U \) is convex, cf. Definition 6.21, and ‘bounded’. The latter property means that for every open neighborhood \( V \) of \( 0 \) there exists an \( s > 0 \) such that \( U \subseteq sV \). (In a normed space it is easy to see that \( B(0, r) \) has all these properties for each \( r > 0 \).)

3.10 Definition A normed space \((V, \| \cdot \|)\) is called

- complete, or Banach space, if the metric space \((V, d)\) with \( d(x, y) = \|x - y\| \) is complete.
- separable if the metric space \((V, d)\) is separable (i.e. has a countable dense subset).

\(^5\)In view of these facts, which we cannot prove without going deeper into measure and integration theory, I don’t find it unreasonable to ask that you understand the much simpler unordered summation.

\(^6\)Countable’ always means ‘at most countable’, otherwise we’ll say ‘countably infinite’. 

11
Just as complete metric spaces are ‘better behaved’ (in the sense of allowing stronger theo-
rems) than general metric spaces, Banach spaces are ‘better’ than normed spaces. Separability
is an annoying restriction that we will try to avoid as much as possible. (An opposite approach
restricts to separable spaces from the beginning in order to make do with weak versions of
the axiom of choice.)

3.11 Remark We have defined the notion of completeness for normed spaces. But complete-
ness can actually be defined for arbitrary topological vector spaces: A sequence \( \{x_n\}_{i \in I} \)
in a TVS \( V \) is called a Cauchy sequence if for every open set \( U \subseteq V \) containing 0 there is a \( n_0 \) such
that \( n, m \geq n_0 \Rightarrow x_n - x_m \in U \). Cauchy nets are defined analogously. (For the definition of
nets see e.g. [46].) Now a TVS \( V \) is called complete if every Cauchy net in \( V \) is convergent.

It is easy to see that a normed space \((V, \|\cdot\|)\) is complete in the sense of normed spaces if
and only if the TVS \((V, \tau_d)\) with \( d(x, y) = \|x - y\| \) is complete in the above sense.  

While metrizability of a topological space \( X \) is a property, a choice of metric on \( X \) is an
extra piece of structure. Similarly, a choice of a norm on a vector space is an extra piece of
structure. If \( \|\cdot\|_1, \|\cdot\|_2 \) are different norms on \( V \) then \((V, \|\cdot\|_1), (V, \|\cdot\|_2)\) are different as normed
spaces even if the norms give rise to the same topology!

3.12 Definition Let \( V \) be an \( F \)-vector space. Two norms \( \|\cdot\|_1, \|\cdot\|_2 \) on \( V \) are called equivalent
if \( \tau_{d_1} = \tau_{d_2} \), where \( d_i(x, y) = \|x - y\|_i \).

We will soon prove (quite easily) the following:

3.13 Proposition Two norms \( \|\cdot\|_1, \|\cdot\|_2 \) on an \( F \)-vector space \( V \) are equivalent if and only if
there are \( 0 < c' \leq c \) such that \( c'\|x\|_1 \leq \|x\|_2 \leq c\|x\|_1 \) for all \( x \in V \).

The following deeper result will be proven later:

3.14 Theorem If \( V \) is a vector space that is complete w.r.t. each of the norms \( \|\cdot\|_1, \|\cdot\|_2 \) and
\( \|\cdot\|_2 \leq c\|\cdot\|_1 \) for some \( c > 0 \) then also \( \|\cdot\|_1 \leq c'\|\cdot\|_2 \) for some \( c' > 0 \), thus the two norms are
equivalent.

3.3 Brief look at locally convex and Fréchet spaces

We have seen that every norm on a vector space gives rise to a translation-invariant metric and
a TVS structure. Analogously, if \( \|\cdot\| \) is seminorm on \( V \), but not a norm, then \( d(x, y) = \|x - y\| \)
defines only a pseudo-metric, and \( \tau_d \) is not Hausdorff (if \( x \neq 0 \) is such that \( \|x\| = 0 \) then there
are no disjoint open sets \( U, V \) containing \( x, 0 \), respectively).

3.15 Definition If \( V \) is an \( F \)-vector space and \( \mathcal{F} \) is a family of seminorms on \( V \) then the
topology \( \tau_{\mathcal{F}} \) is the smallest topology on \( V \) containing the balls

\[ B_{\|\cdot\|}(x, r) = \{y \in V \mid \|x - y\| < r\} \]

for all \( x \in V, r > 0, \|\cdot\| \in \mathcal{F} \).

More explicitly, \( \tau_{\mathcal{F}} \) consists of all unions of finite intersections of such balls, i.e. the latter
form a subbase for \( \tau_{\mathcal{F}} \). Now a sequence or net \( \{x_i\} \) in \( V \) converges to \( z \in V \) if and only if
\( \|x_i - z\| \) converges to zero for each \( \|\cdot\| \in \mathcal{F} \).
3.16 Definition We say that $F$ is separating if for any non-zero $x \in V$ there is a $\| \cdot \| \in F$ such that $\|x\| \neq 0$.

The property of being separating is important since one usually is only interested in Hausdorff topological vector spaces and the following holds:

3.17 Lemma The topology $\tau_F$ induced by a family $F$ of seminorms on $V$ is Hausdorff if and only if $F$ is separating.

Proof. $\Rightarrow$ Assume $F$ is not separating. Then there is $0 \neq x \in V$ such that $\|x\| = 0$ for all $\| \cdot \| \in F$. Then by definition of $\tau_F$, every open set containing 0 also contains $x$ and vice versa, so that $\tau_F$ is not Hausdorff.

$\Leftarrow$ Assume $x \neq y$. By assumption there is a $\| \cdot \| \in F$ such that $c = \|x - y\| > 0$. Let $U = B_{\| \cdot \|}(x, c/2), V = B_{\| \cdot \|}(y, c/2)$. Then $U, V$ are open sets containing $x, y$, respectively, and existence of $z \in U \cap V$ would imply the contradiction $d(x, y) \leq d(x, z) + d(z, y) < r/2 + r/2 = r = d(x, y)$. Thus $U \cap V = \emptyset$, so that $\tau_F$ is Hausdorff. ■

If $V$ is an $F$-vector space and $F$ is a family of seminorms on $V$, one can prove that $V$ is a topological vector space when equipped with the topology $\tau_F$. The proof is not much more complicated than for the case of one (semi)norm considered above.

3.18 Definition A topological vector space $(V, \tau)$ over $\mathbb{F}$ is called locally convex if there exists a separating family $F$ of seminorms on $V$ such that $\tau = \tau_F$.

3.19 Remark 1. Locally convex spaces were introduced in 1935 by John von Neumann (1903-1957), to whom also von Neumann algebras, the theory of unbounded operators, the spectral theorem and countless other discoveries in pure and applied mathematics are due.

2. For an equivalent, more geometric way of defining local convexity of a TVS see the supplementary Section B.5.1, and for more on locally convex spaces see, e.g., [38, 11, 62].

If the separating family $F$ has just one element, we are back at the notion of a normed, possibly Banach, space. If $F$ is finite, i.e. $F = \{\| \cdot \|_1, \ldots, \| \cdot \|_n\}$, then $\| \cdot \| = \sum_{i=1}^n \| \cdot \|_i$ is a seminorm, and it is a norm if and only if $F$ is separating. Thus the case of finite $F$ again gives a normed space. Thus $F$ must be infinite in order for interesting things to happen.

If $F$ is infinite, we cannot just put $\|x\| = \sum_{\| \cdot \| \in F} \|x\|$, since the r.h.s. has no reason to converge for all $x \in V$. But if the family $F$ of seminorms on $V$ is countable, we can label the elements of $F$ as $\| \cdot \|_n$, $n \in \mathbb{N}$ and define

$$d(x, y) = \sum_{n=1}^{\infty} 2^{-n} \min(1, \|x - y\|_n).$$

Now each term $\min(1, \|x - y\|_n)$ is a translation-invariant pseudometric on $V$ that is bounded by 1, and the sum converges to a translation-invariant metric on $V$. With just a bit more work one shows that $\tau_F = \tau_d$, thus $(V, \tau_F)$ is metrizable. (Note that we could not have defined $\|x\| = \sum_{i=1}^{\infty} 2^{-i} \|x\|_i$ since this again may fail to converge, thus need not be a norm.) If such a space is complete, it is called a Fréchet space. Clearly, every Fréchet space is an $F$-space.

Here is an example of a Fréchet space: For $f \in C^\infty(\mathbb{R}, \mathbb{C})$ and $n, m \in \mathbb{N}_0$, define

$$\|f\|_{n,m} = \sup_{x \in \mathbb{R}} |x|^n |f^{(m)}(x)|,$$
where \( f^{(m)} \) is the \( m \)-th derivative of \( f \). These are seminorms. Now 
\[
\mathcal{S} = \{ f \in C^\infty(\mathbb{R}, \mathbb{C}) \mid \| f \|_{n,m} < \infty \ \forall n, m \in \mathbb{N}_0 \}
\]
is a Fréchet space when equipped with the topology defined by the countable family \( \mathcal{F} = \{ \| \cdot \|_{n,m} \mid n, m \in \mathbb{N}_0 \} \). Its elements are called Schwartz functions. They are infinitely differentiable functions that, together with all their derivatives, vanish faster as \(|x| \to \infty\) than \(|x|^{-n}\) for any \( n \). (This definition is easily generalized to functions of several variables.) Note that the seminorm \( \| \cdot \|_{0,0} \) alone already separates the elements of \( \mathcal{S} \), thus is a norm, but having the other seminorms around gives rise to a finer topology, one that is not normable.

4 Normed and Banach space basics

4.1 Why we care about completeness. Closed subspaces

As you (should) know from topology, completeness of a metric space is convenient since leads to results that are not necessarily true without it, like Cantor’s intersection theorem and the contraction principle (or Banach’s fixed point theorem). The same holds for normed spaces. Here is one main reason:

4.1 Definition Let \((V, \| \cdot \|)\) be a normed space and \( \{x_n\}_{n \in \mathbb{N}} \) a sequence. The series \( \sum_{n=1}^\infty x_n \) is said to be absolutely convergent if \( \sum_{n=1}^\infty \|x_n\| < \infty \) and to converge to \( s \in V \) if the sequence \( S_n = \sum_{k=1}^n x_k \) of partial sums converges to \( s \).

4.2 Proposition Let \((V, \| \cdot \|)\) be a normed \( \mathbb{F} \)-vector space. Then the following are equivalent:

(i) \((V, \| \cdot \|)\) is complete, thus a Banach space.

(ii) Every absolutely convergent series \( \sum_{n=1}^\infty x_n \) in \( V \) converges.

Under these assumptions, the sum satisfies \( \| \sum_n x_n \| \leq \sum_n \|x_n\| \).

Proof. \( \Rightarrow \) Assume \( V \) to be complete and \( \sum_n x_n \) to be absolutely convergent. Let \( S_n = \sum_{k=1}^n x_k \) and \( T_n = \sum_{k=1}^n \|x_k\| \). For all \( n > m \) we have (by subadditivity of the norm)
\[
\|S_n - S_m\| = \| \sum_{k=m+1}^n x_k \| \leq \sum_{k=m+1}^n \|x_k\| = T_n - T_m.
\]

Since the sequence \( \{T_n\} \) is convergent by assumption, thus Cauchy, the above implies that \( \{S_n\} \) is Cauchy, thus convergent by completeness of \( V \). The subadditivity of the norm gives \( \| \sum_{k=1}^n x_k\| \leq \sum_{k=1}^n \|x_k\| \) for all \( n \), and since the limit \( n \to \infty \) of both sides exists, we have the inequality.

\( \Leftarrow \) Assume that every absolutely convergent series in \( V \) converges, and let \( \{y_n\}_{n \in \mathbb{N}} \) be a Cauchy sequence in \( V \). We can find (why?) a subsequence \( \{z_k\}_{k \in \mathbb{N}} = \{y_{n_k}\} \) such that \( \|z_k - z_{k-1}\| \leq 2^{-k} \ \forall k \geq 2 \). Now put \( z_0 = 0 \) and define \( x_k = z_k - z_{k-1} \) for \( k \geq 1 \). Now
\[
\sum_{k=1}^\infty \|x_k\| = \sum_{k=1}^\infty \|z_k - z_{k-1}\| \leq \|z_1\| + \sum_{k=2}^\infty 2^{-k} < \infty.
\]

\(^7\)Laurent Schwartz (1915-2002). French mathematician who invented ‘distributions’, an important notion in functional analysis.

\(^8\)Unfortunately, some authors write: “\( \sum_n \|x_n\| < \infty \), thus \( \sum_n x_n \) converges” without indicating that something needs to be proven here.
Thus $\sum_{k=1}^{\infty} x_k$ is absolutely convergent, and therefore convergent by the hypothesis. To wit, $\lim_{n \to \infty} S_n$ exists, where $S_n = \sum_{k=1}^{n} x_k = \sum_{k=1}^{n} (z_k - z_{k-1}) = z_n$. Thus $z = \lim_{k \to \infty} z_k = \lim_{k \to \infty} y_{n_k}$ exists. Now the sequence $\{y_n\}$ is Cauchy and has a convergent subsequence $\{y_{n_k}\}$. This implies (why?) that the whole sequence $\{y_n\}$ converges to the limit of the subsequence. ■

We will see various other reasons for the importance of completeness. In the next section, it will be used to prove that every finite dimensional linear subspace of a normed space is automatically closed. This is not at all true for infinite dimensional subspaces. For example, let $V = \ell^1(\mathbb{N}) = \{f : \mathbb{N} \to \mathbb{R} \mid \sum_{n=1}^{\infty} |f(n)| < \infty\}$. Now $W = \{f : \mathbb{N} \to \mathbb{R} \mid \#\{n \in \mathbb{N} \mid f(n) \neq 0\} < \infty\} \subset V$ is an infinite dimensional linear proper subspace, but not closed: One easily checks that $\overline{W} = V$.

Closedness and completeness of subsets of a metric space are related. We recall from topology (if you haven’t seen this, prove it!):

4.3 Lemma Let $(X, d)$ be a metric space and $Y \subseteq X$. Then (instead of $d|_Y$ I just write $d$)

(i) If $(X, d)$ is complete and $Y \subseteq X$ is closed (w.r.t. $\tau_d$, of course) then $(Y, d)$ is complete.
(ii) If $(Y, d)$ is complete then $Y \subseteq X$ is closed (whether or not $(X, d)$ is complete).

The above should be compared with the fact that a closed subset of a compact space is compact and that a compact subset of a Hausdorff space is closed. In the above, completeness works as a weak substitute of compactness, an interpretation that is reinforced by the fact that every compact metric space is complete.

This specializes immediately to normed spaces:

4.4 Lemma Let $(V, \| \cdot \|)$ be a normed space and $W \subseteq V$ a linear subspace. Then

(i) If $V$ is complete (=Banach) and $W \subseteq V$ is closed then $W$ is Banach.
(ii) If $W$ is complete then $W \subseteq V$ is closed (whether or not $V$ is complete).

4.5 Definition A linear map $A : V \to W$ of normed spaces is called an isometry if $\|Ax\| = \|x\|$ for all $x \in V$.

Note that an isometry is automatically injective: If $Ax = Ay$ then $\|x - y\| = \|A(x - y)\| = \|Ax - Ay\| = 0$. Furthermore, if $A : V \to W$ is a surjective isometry then it is invertible, and its inverse is also an isometry. Then $A$ is an isometric isomorphism of normed spaces.

4.6 Corollary Let $V$ be a complete normed space and $W$ a normed space. If $A : V \to W$ is an isometry then the linear subspace $AV \subseteq W$ is closed.

Proof. The map $A : V \to AV \subseteq W$ is an isometric bijection and therefore an isometric isomorphism of normed spaces. Thus $(AV, \| \cdot \|)$ is complete, thus closed in $W$ by Lemma 4.4. ■

I suppose as known that that every metric space can be completed, i.e. embedded isometrically into a complete metric space (unique up to isometry) as a dense subspace.

4.7 Exercise Prove that the completion of a normed space again is a vector space, thus a Banach space.

---

9If $f : X \to Y$ is a function and $Z \subseteq X$, we write either $f|_Z$ or $f \upharpoonright Z$ for the restriction of $f$ to $Z$. Usually the first form is typographically more convenient.
Later we will see an alternative construction of the completion.

4.8 Exercise Let \((V_1, \| \cdot \|_1), (V_2, \| \cdot \|_2)\) be normed spaces.

(i) Prove that \(\|(x_1, x_2)\|_m = \|x_1\|_1 + \|x_2\|_2\) and \(\|(x_1, x_2)\|_m = \max(\|x_1\|_1, \|x_2\|_2)\) are equivalent norms on \(V_1 \oplus V_2\).

(ii) Prove that \((V_1 \oplus V_2, \| \cdot \|_{s/m})\) is complete if and only if \((V_1, \| \cdot \|_1), (V_2, \| \cdot \|_2)\) both are complete.

Proof. (i) It is immediate that \(\|(x_1, x_2)\|_m \leq \|(x_1, x_2)\|_s \leq 2\|(x_1, x_2)\|_m\).

(ii) Assume that \(X_1, X_2\) are complete. If \(\{(x_i, y_i)\} \subseteq X_1 \oplus X_2\) is Cauchy w.r.t. \(\| \cdot \|_m\) then \(\{x_i\}\) and \(\{y_i\}\) are Cauchy in \(X, Y\), thus convergent. I.e. there are \(X \in X_1, y \in X_2\) such that \(\|x_i - x\| \to 0\), \(\|y_i - y\| \to 0\). Now \(\|(x_i, y_i) - (x, y)\| \to 0\) for both sum and maximum norm. Assume \(X_1 \oplus X_2\) is complete for the sum or maximum norm. By their equivalence, it then is complete for the other. And if \(\{x_i\}\) is Cauchy in \(X_1\) then \(\{(x_i, 0)\} \subseteq X_1 \oplus X_2\) is Cauchy, thus convergent to some \((x, 0)\). Now it is clear that \(x_i \to x\).

4.9 Exercise (i) Let \(\{(V_i, \| \cdot \|_i)\}_{i \in I}\) be a family of normed spaces, where \(I\) is any set. Put

\[\bigoplus_{i \in I} V_i = \{\{x_i\}_{i \in I} \mid \sum_{i \in I} \|f(i)\|_i < \infty\}.\]

(technically, this is a subset of \(\prod_{i \in I} V_i = \{f : I \to \bigcup_{j \in I} V_j \mid f(i) \in V_i \forall i \in I\}\).) Prove that this is a linear space and \(\|f\| = \sum_{i \in I} \|f(i)\|_i\) a norm on it.

(ii) Prove that \((\bigoplus_{i \in I} V_i, \| \cdot \|)\) is complete if all the \(V_i\) are complete. Hint: The proof is an adaptation of the one for \(\ell^1(S)\) given in Section 5.3.

If \(V_i = F\) for all \(i \in I\) with the usual norm, we have \(\bigoplus_{i \in I} V_i \cong \ell^1(I, F)\).

4.2 Linear maps: bounded \(\iff\) continuous

If \(E, F\) are vector spaces over \(F\), a map \(A : E \to F\) is called linear if \(A(x + y) = Ax + Ay\) for all \(x, y \in E\) and \(A(cx) = cAx\) for all \(x \in E, c \in F\). Note that, as in linear algebra, we write \(Ax\) instead of \(A(x)\).

NB: A map of the form \(x \mapsto Ax + b\), where \(A : E \to F\) is linear and \(b \in F\), is not called a linear map, but an affine one!

4.10 Definition Let \(E, F\) be normed spaces and \(A : E \to F\) a linear map. Then the norm \(\|A\| \in [0, \infty]\) is defined by

\[\|A\| = \sup_{0 \neq e \in E} \frac{\|Ae\|}{\|e\|} = \sup_{\|e\| = 1} \|Ae\|.\]

(The equality of the second and third expression is due to linearity of \(A\).) If \(\|A\| < \infty\) then \(A\) is called bounded.

\(^{10}\)It should be clear that writing \(\sup_{e \in E, \|e\| \leq 1} \|Ae\|\) instead would not change the result.
4.11 REMARK 1. ‘Linear operator’ is a synonym for linear map, but linear maps $A : E \to F$ are called linear functionals.

2. While unbounded linear maps exist, cf. Exercise 4.15 below, in the unbounded case it often is too restrictive to require them to be defined everywhere. See also Remark 9.10. □

4.12 EXERCISE If $E, G, H$ are normed spaces and $S : E \to G$, $T : G \to H$ are linear maps, prove that $\|T \circ S\| \leq \|S\| \|T\|$.

4.13 LEMMA Let $E, F$ be normed spaces and $A : E \to F$ a linear map. Then the following are equivalent:

(i) $A$ is bounded.

(ii) $A$ is continuous (w.r.t. the norm topologies).

(iii) $A$ is continuous at 0 $\in E$.

Proof. (i)$\Rightarrow$(ii) For $x, y \in E$ we have $\|Ax - Ay\| = \|A(x - y)\| \leq \|A\| \|x - y\|$. Thus $d(Ax, Ay) \leq \|A\| d(x, y)$, and with $\|A\| < \infty$ we have (uniform) continuity of $E$.

(ii)$\Rightarrow$(iii) This is obvious.

(iii)$\Rightarrow$(i) $B^F(0, 1) \subseteq F$ is an open neighborhood of 0 $\in F$. Since $A$ is continuous at 0, there is an open neighborhood $U \subseteq E$ of 0 $\in E$ such that $A(U) \subseteq B^F(0, 1)$. Since the balls form bases of the topologies, there is $C > 0$ such that $B^E(0, C) \subseteq U$, thus $A(B^E(0, C)) \subseteq B^F(0, 1)$.

By linearity of $A$ and the properties of the norm, this is equivalent to $A(B^E(0, 1)) \subseteq B^F(0, D)$, where $D = 1/C$. If $0 \neq x \in E$ then

$$Ax = 2\|x\| A \left( \frac{x}{2\|x\|} \right),$$

thus $\|Ax\| \leq 2\|x\||A(x/2\|x\|)| < 2\|x\|D$, and $A$ is bounded. ■

4.14 EXERCISE Let $V, W$ be normed spaces, where $V$ is finite dimensional. Prove that every linear map $V \to W$ is bounded.

4.15 EXERCISE 1. Give an example of an unbounded linear map $A : V \to W$ between normed spaces that is defined on all of $V$.

2. Bonus: Same as 1., but with $V, W$ Banach.

For linear functionals, i.e. linear maps from an $F$-vector space to $F$, there is another characterization of continuity:

4.16 EXERCISE Let $(V, \| \cdot \|)$ be a normed $F$-vector space and $\varphi : V \to F$ a linear functional. Prove that $\varphi$ is continuous if and only if $\ker \varphi = \varphi^{-1}(0) \subseteq V$ is closed.

Hint: For $\Leftarrow$, pick a ball $B(x, r) \subseteq V \setminus \ker \varphi$ and prove that $\varphi(B(0, r))$ is bounded.

4.17 LEMMA Let $V$ be a normed space, $W \subseteq V$ a dense linear subspace, $Z$ a Banach space and $A : W \to Z$ a bounded linear map. Then there is a unique linear map $\hat{A} : V \to Z$ with $\hat{A}|W = A$ and $\|\hat{A}\| = \|A\|$. If $A$ is an isometry, so is $\hat{A}$.

Proof. Let $x \in V$. Then there is a sequence $\{w_n\}$ in $W$ such that $\|w_n - x\| \to 0$. Then $\{w_n\} \subseteq W$ is a Cauchy sequence, and so is $\{Aw_n\} \subseteq Z$ by boundedness of $A$. The latter converges since $Z$ is complete. If $\{w'_n\}$ is another sequence converging to $x$ then $\|A(w_n - w'_n)\| \to 0$, so that $\lim Aw'_n = \lim Aw_n$. It thus is well-defined to put $\hat{A}x = \lim_{n \to \infty} Aw_n$. We omit the
easy proof of linearity of $\hat{A}$. If $x \in W$ then we can put $w_n = x \forall n$, obtaining $\hat{A}w_n = Ax$, thus $\hat{A}|_W = A$. Finally, $\|\hat{A}x\| = \lim \|Aw_n\| \leq \|A\|\|x\|$. Thus $\|\hat{A}\| \leq \|A\|$, and the converse inequality is obvious. If $A$ is an isometry then $\|\hat{A}x\| = \lim_n \|Aw_n\| = \lim_n \|w_n\| = \|x\|$, so that $\hat{A}$ is an isometry. ■

We recall the following from topology: If $X$ is a set and $\tau_1, \tau_2$ are topologies on $X$ then $\tau_1 = \tau_2$ holds if and only if $\text{id}_X : (X, \tau_1) \to (X, \tau_2)$ is a homeomorphism, i.e. continuous with continuous inverse.

**Proof of Proposition 3.13.** Equivalence of $\| \cdot \|_1, \| \cdot \|_2$ means that the two norms give rise to the same topology. By the above, this is equivalent to the maps $\text{id}_V : (V, \| \cdot \|_1) \to (V, \| \cdot \|_2)$ and $\text{id}_V : (V, \| \cdot \|_2) \to (V, \| \cdot \|_1)$ being continuous. Now by Lemma 4.13, this is equivalent to the existence of $C, C'$ such that $\|x\|_1 \leq C\|x\|_2$ and $\|x\|_2 \leq C'\|x\|_1$ holding for all $x \in V$. ■

4.18 **Exercise** Prove: If $V$ is a vector space and $\| \cdot \|_1, \| \cdot \|_2$ are equivalent norms on $V$ then completeness of $V$ w.r.t. $\| \cdot \|_1$ is equivalent to completeness of $V$ w.r.t. $\| \cdot \|_2$.

4.19 **Proposition** On a finite dimensional vector space, all norms are equivalent.

**Proof.** Let $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$. Let $B = \{e_1, \ldots, e_d\}$ be a basis for $V$, and define the Euclidean norm $\| \cdot \|_2$ of $x = \sum_i c_i e_i$ by $\|x\|_2 = (\sum_i |c_i|^2)^{1/2}$. Since equivalence of norms is an equivalence relation, it clearly is sufficient to show that any norm $\| \cdot \|$ is equivalent to $\| \cdot \|_2$. Using $|c_i| \leq \|x\|_2 \forall i$ and the properties of any norm, we have

$$\|x\| = \left\| \sum_{i=1}^d c_i e_i \right\| \leq \sum_{i=1}^d |c_i| \|e_i\| \leq \left( \sum_{i=1}^d \|e_i\|^2 \right)^{1/2} \|x\|_2. \quad (4.1)$$

This implies that $x \mapsto \|x\|$ is continuous w.r.t. the topology on $V$ defined by $\| \cdot \|_2$. The sphere $S = \{x \in \mathbb{F}^n \mid \|x\|_2 = 1\}$ is closed and bounded, thus compact, which implies that there is $z \in S$ such that $\lambda := \inf_{x \in S} \|x\| = \|z\|$. Since $z \in S$ implies $z \neq 0$ and $\| \cdot \|$ is a norm, we have $\lambda = \|z\| > 0$. Now, for $x \neq 0$ we have $\frac{x}{\|x\|_2} \in S$, and thus

$$\|x\| = \|x\|_2 \left\| \frac{x}{\|x\|_2} \right\| \geq \|x\|_2 \lambda. \quad (4.2)$$

Combining (4.1, 4.2), we have $c_1 \|x\|_2 \leq \|x\| \leq c_2 \|x\|_2$ with $0 < c_1 = \inf_{x \in S} \|x\| \leq \sum_i \|e_i\| = c_2$. (Note that $e_i \in S \forall i$, so that $c_2 \geq dc_1$, showing again that $V$ must be finite dimensional.) ■

4.20 **Remark** In fact, one can prove the somewhat stronger result that on a finite dimensional vector space there is precisely one topology making it a TVS.

4.21 **Exercise** Prove that every finite dimensional normed space over $\mathbb{R}$ or $\mathbb{C}$ is complete.

4.22 **Exercise** Prove that every finite dimensional subspace of a normed space is closed.

### 4.3 Spaces of bounded linear maps. First glimpse of Banach algebras

4.23 **Definition** Let $E, F$ be normed $\mathbb{F}$-vector spaces. The set of bounded linear maps from $E$ to $F$ is denoted $B(E, F)$. Instead of $B(E, E)$ and $B(E, \mathbb{F})$ one also writes $B(E)$ and $E^*$, respectively. $E^*$ is called the dual space of $E$. 18
Clearly $B(E,F)$ should not be confused with notation $B(x,r)$ for open balls!

4.24 Proposition Let $E$, $F$ be normed spaces.

(i) $B(E,F)$ is a vector space and $B(E,F) \to [0,\infty)$, $A \mapsto \|A\|$ is a norm in the sense of Definition 3.5.

(ii) If $F$ is complete (=Banach) then so is $B(E,F)$. $E^*$ is always Banach.

Proof. (i) If $T : E \to F$ is a linear map, it is clear that $\|\alpha T\| = |\alpha|\|T\|$ and that $\|T\| = 0$ if and only if $T = 0$. If $S,T \in B(E,F)$ and $x \in E$ then $\|(S+T)x\| \leq \|Sx\| + \|Tx\| \leq (\|S\| + \|T\|)\|x\|$, so that $\|S + T\| \leq \|S\| + \|T\|$. This implies that $B(E,F)$ is a vector space.

(ii) Assume $F$ is complete, and let $\{T_n\} \subseteq B(E,F)$ be a Cauchy sequence. Then there is $n_0$ such that $m,n \geq n_0 \Rightarrow \|T_m - T_n\| < 1$, in particular $T_m \in B(T_{n_0},1)$ for all $n \geq n_0$. Thus with $M = \max(\|T_1\|,\ldots,\|T_{n_0-1}\|,\|T_{n_0}\| + 1)$ we have $\|T_n\| \leq M$ for all $n$. If now $x \in E$ then $\|(T_n - T_m)x\| \leq \|T_n - T_m\|\|x\|$, so that $\{T_n x\}$ is a Cauchy sequence in $F$ and therefore convergent by completeness of $F$. Now define $T : E \to F$ by $Tx = \lim_{n \to \infty} T_n x$. It is straightforward to check that $T$ is linear. Finally, since $\|T_n x\| \leq M\|x\|$ for all $n$, we have $\|Tx\| = \lim_{n \to \infty} \|T_n x\|$, so that $T \in B(E,F)$.

4.25 Exercise Let $(V_1, \| \cdot \|_1), (V_2, \| \cdot \|_2)$ be normed spaces. Prove $(V_1 \oplus V_2, \| \cdot \|_s)^* \cong (V_1^* \oplus V_2^*, \| \cdot \|_s)$ and $(V_1 \oplus V_2, \| \cdot \|_m)^* \cong (V_1^* \oplus V_2^*, \| \cdot \|_s)$.

If $E$ is a normed $\mathbb{F}$-vector space, the same holds for $B(E) = B(E,E)$, and by Exercise 4.12, we have $\|S \circ T\| \leq \|S\|\|T\|$ for all $S,T \in B(E)$. This motivates the following definition:

4.26 Definition If $\mathbb{F}$ is a field, an $\mathbb{F}$-algebra is an $\mathbb{F}$-vector space $A$ together with an associative bilinear operation $A \times A \to A$, the ‘multiplication’.

Examples: $A = M_{n \times n}(\mathbb{F})$ with matrix product as multiplication, $A = C(X,\mathbb{F})$ with pointwise product of functions.

4.27 Definition A normed $\mathbb{F}$-algebra is a $\mathbb{F}$-algebra $A$ equipped with a norm $\| \cdot \|$ such that $\|ab\| \leq \|a\|\|b\|$ for all $a,b \in A$ (submultiplicativity). A Banach algebra is a normed algebra that is complete (as a normed space). An algebra $A$ is called unital if it has a unit $1 \neq 0$. (In fact, if $A \neq \{0\}$ then $1 = 0$ would imply the contradiction $\|a\| = \|1a\| \leq \|1\||a| = 0$ $\forall a \in A$.)

4.28 Remark 1. If $A$ is a normed algebra then for all $a,a',b,b' \in A$ we have $\|ab - a'b'\| = \|ab - ab' + ab' - a'b'\| \leq \|a\|\|b-b'\| + \|a-a'\|\|b'\|$. This proves that the multiplication map $\cdot : A \times A \to A$ is jointly continuous.

2. If $A$ is a normed algebra with unit $1$ then $1 = 1^2$, thus $\|1\| = \|1^2\| \leq \|1\|^2$. With $\|1\| \neq 0$ this implies $1 \leq \|1\|$. Some authors require all unital normed algebras to satisfy $\|1\| = 1$, but we don’t. Of course this does hold for $B(E)$.

By the above, $B(E)$ is a normed algebra for every normed space $E$, and by Proposition 4.24(ii), $B(E)$ is a Banach algebra whenever $E$ is a Banach space. There is another standard class of examples:

4.29 Example Let $X$ be a compact Hausdorff space and $A = C(X,\mathbb{F})$. We already know that $A$, equipped with the norm $\|f\| = \sup_{x \in X} |f(x)|$ is a Banach space. The pointwise product $(fg)(x) = f(x)g(x)$ of functions is bilinear, associative and clearly satisfies $\|fg\| \leq \|f\|\|g\|$. This makes $(A, \cdot, \| \cdot \|)$ a Banach algebra.
We will have more to say about Banach algebras later in the course.

Before we go on developing general theory, it seems instructive to study in some detail an important class of spaces, the $\ell^p(S)$ spaces, where everything can be done very explicitly, in particular the dual spaces can be determined.

5 The sequence spaces and their dual spaces

In this section we will consider in some detail the sequence spaces $\ell^p(S,\mathbb{F})$ for $0 < p \leq \infty$. These spaces are worth studying for several reasons:

- They provide a first encounter with the more general Lebesgue spaces $L^p(X,\mathcal{A},\mu)$ without the measure and integration theoretic baggage needed for the latter.
- They can be studied quite completely and have their dual spaces identified.
- We will see that every Hilbert space is isomorphic to $\ell^2(S,\mathbb{F})$ for some $S$.

5.1 Basics. $1 \leq p \leq \infty$: Hölder and Minkowski inequalities

5.1 Definition ($\ell^p$-Spaces) If $\mathbb{F} \in \{\mathbb{R},\mathbb{C}\}$, $0 < p < \infty$, $S$ is a set and $f : S \rightarrow \mathbb{F}$, define

$$\|f\|_\infty = \sup_{s \in S} |f(s)| \in [0, \infty], \quad \|f\|_p = \left(\sum_{s \in S} |f(s)|^p\right)^{1/p} \in [0, \infty],$$

where $\infty^{1/p} = \infty$ and we use the notion of unordered sums, cf. Appendix A.1. Now for all $p \in (0, \infty]$ put

$$\ell^p(S,\mathbb{F}) := \{f : S \rightarrow \mathbb{F} \mid \|f\|_p < \infty\}.$$

5.2 Lemma For all $p \in (0, \infty]$ and $f : S \rightarrow \mathbb{F}$ we have:

(i) $\|f\|_p = 0$ if and only if $f = 0$.

(ii) For all $c \in \mathbb{F}$ we have $\|cf\|_p = |c|\|f\|_p$ (with the understanding that $0 \cdot \infty = 0$).

(iii) If $S$ is finite then $\ell^p(S,\mathbb{F}) = \{f : S \rightarrow \mathbb{F}\} = \mathbb{F}^S$. If $\#S = 1$ then all the $\|\cdot\|_p$ coincide.

Proof. Trivial.  

5.3 Lemma (i) $(\ell^p(S,\mathbb{F}),\|\cdot\|_p)$ are normed vector spaces for $p = 1$ and $p = \infty$.

(ii) If $f \in \ell^1(S,\mathbb{F})$ and $g \in \ell^\infty(S,\mathbb{F})$ then

$$\left|\sum_{s \in S} f(s)g(s)\right| \leq \|fg\|_1 \leq \|f\|_1\|g\|_\infty.$$

Proof. $\ell^p(S,\mathbb{F})$ obviously is stable under scalar multiplication. And

$$\|f + g\|_\infty = \sup_s |f(s) + g(s)| \leq \sup_s |f(s)| + \sup_s |g(s)| = \|f\|_\infty + \|g\|_\infty,$$

$$\|f + g\|_1 = \sum_s |f(s) + g(s)| \leq \sum_s (|f(s)| + |g(s)|) = \|f\|_1 + \|g\|_1.$$

Thus for $p \in \{1,\infty\}$ and $f, g \in \ell^p(S,\mathbb{F})$ we have $f + g \in \ell^p(S,\mathbb{F})$, so that $\ell^p(S,\mathbb{F})$ is an $\mathbb{F}$-vector space and $\|\cdot\|_p$ a norm on it. For (ii) we only need the triviality $|f(s)g(s)| \leq \|g\|_\infty|f(s)|$ and
the inequality \( |\sum_{s \in S} f(s)| \leq \sum_{s \in S} |f(s)| \).

For \(1 < p < \infty\) define \( q \in (1, \infty)\) by \( \frac{1}{p} + \frac{1}{q} = 1\). (This is equivalent to \( pq = p + q\), which often is useful.) Whenever \( p, q \) appear together they are supposed to be a conjugate or dual pair in this sense. We extend this in a natural way by declaring \((1, \infty)\) and \((\infty, 1)\) to be conjugate pairs.

5.4 Proposition Let \(1 < p < \infty\) and \( q\) conjugate to \( p\), i.e. \( \frac{1}{p} + \frac{1}{q} = 1\). Then

(i) For all \( f, g : S \to \mathbb{R}\) we have \( \|fg\|_1 \leq \|f\|_p \|g\|_q\). (Inequality of Hölder\(^{11}\) (1889))

(ii) For all \( f, g : S \to \mathbb{R}\) we have \( \|f + g\|_p \leq \|f\|_p + \|g\|_p\). (Inequality of Minkowski\(^{12}\) (1896))

Proof. (i) We may assume \( \|f\|_p, \|g\|_q \) to be finite. The exponential function \( \mathbb{R} \to \mathbb{R}, x \mapsto e^x\) is convex\(^{13}\), to that with of \( \frac{1}{p} + \frac{1}{q} = 1 \) we have

\[
e^{a/p} e^{b/q} = \exp \left( \frac{a}{p} + \frac{b}{q} \right) \leq \frac{e^a}{p} + \frac{e^b}{q} \quad \forall a, b \in \mathbb{R}.
\]

(5.1)

(The validity also for \( u = 0 \) or \( v = 0 \) is obvious.)

Putting \( u = |f(s)|, v = |g(s)|\) in (5.1), we have \( |f(s)g(s)| \leq p^{-1} |f(s)|^p + q^{-1} |g(s)|^q\), so that summing over \( s \) gives \( \|fg\|_1 \leq p^{-1} \|f\|_p^p + q^{-1} \|g\|_q^q\). If \( \|f\|_p = \|g\|_q = 1 \) then this reduces to \( \|fg\|_1 \leq \frac{1}{p} + \frac{1}{q} = 1\). If now \( f, g \) are non-zero but otherwise arbitrary, put \( f' = \frac{f}{\|f\|_p}, g' = \frac{g}{\|g\|_q}\). Now \( \|f'_p\|_p = 1 = \|g'_q\|_q\), so that the above implies \( \|f'g'_p\|_1 \leq 1\). Now inserting the definitions of \( f', g'\) gives \( \|fg\|_1 \leq \|f\|_p \|g\|_q\).

Of course the inequality is trivially true if \( f \) or \( g \) vanishes.

(ii) Again we assume that \( \|f\|_p, \|g\|_p \) are finite, thus \( f, g \in \ell^p(S, \mathbb{F})\). Then with the useful inequality

\[
|a + b|^p \leq (|a| + |b|)^p \leq 2 \max(|a|, |b|)^p = 2^p \max(|a|^p, |b|^p) \leq 2^p(|a|^p + |b|^p)
\]

we have

\[
\sum_s |f(s) + g(s)|^p \leq 2^p \sum_s |f(s)|^p + |g(s)|^p = 2^p (\|f\|_p^p + \|g\|_p^p) < \infty.
\]

so that \( \|f + g\|_p < \infty \) and \( f + g \in \ell^p(S, \mathbb{C})\), and \( \ell^p(S, \mathbb{F})\) is a vector space.

If \( h \in \ell^p\) then using \( pq = p + q\) (where \( q \) is conjugate to \( p\)) we find \( \sum_s |h(s)|^{(p-1)q} = \sum_s |h(s)|^p < \infty\), so that \( s \mapsto |h(s)|^{p-1} \) is in \( \ell^q\) with \( \|h\|_{p^{-1}} = \|h\|_p^{p/q}\). Now

\[
\|f + g\|_p = \sum_s |f(s) + g(s)|^p = |f(s) + g(s)| |f(s) + g(s)|^{p-1}
\]

\[
\leq \sum_s (|f(s)| + |g(s)|) |f(s) + g(s)|^{p-1}
\]

\[
\leq (\|f\|_p + \|g\|_p) \|f + g\|_p \leq (\|f\|_p + \|g\|_p) \|f + g\|_p^{p/q},
\]

\(^{11}\)Otto Hölder (1859-1937). German mathematician. Important contributions to analysis and algebra.

\(^{12}\)Hermann Minkowski (1864-1909). German mathematician. Contributions to number theory, relativity and other fields.

\(^{13}\)\( f : [a, b] \to \mathbb{R} \) is convex if \( f(tx + (1-t)y) \leq tf(x) + (1-t)f(y) \) for all \( x, y \in [a, b] \) and \( t \in [0, 1] \). See, e.g., [22, Vol. 1, Section 7.2].
where the second ≤ comes from Hölder’s inequality applied to $|f| \in \ell^p$ and $|f + g|^{p-1} \in \ell^q$ and also to $|g| \in \ell^p$ and $|f + g|^{p-1} \in \ell^q$. If $\|f + g\|_p \neq 0$, we can divide by $\|f + g\|^{p/q}$ and with $p - p/q = p(1 - 1/q) = p_1^p = 1$ we obtain

$$\|f + g\|_p = \|f + g\|^{p-p/q} \leq \|f\|_p + \|g\|_p.$$  

Since this clearly also holds if $\|f + g\|_p = 0$, we are done.  

For $p = q = 2$, the inequality of Hölder is known as the Cauchy-Schwarz inequality. We will also call the trivial inequalities of Lemma 5.3 for $\{p, q\} = \{1, \infty\}$ Hölder and Minkowski inequalities. Now the analogue of Lemma 5.3 for $1 < p < \infty$ is clear:

5.5 **Corollary** Let $1 < p < \infty$. Then

(i) $(\ell^p(S, \mathbb{F}), \| \cdot \|_p)$ is a normed vector space.

(ii) If $q$ is conjugate to $p$ and $f \in \ell^p(S, \mathbb{F})$ and $g \in \ell^q(S, \mathbb{F})$ then

$$\left| \sum_{s \in S} f(s)g(s) \right| \leq \|fg\|_1 \leq \|f\|_p\|g\|_q.$$  

5.2 **Aside:** The translation-invariant metric $d_p$ for $0 < p < 1$

For $s \in S$, let $\delta_s : S \rightarrow \mathbb{F}$ be the function defined by $\delta_s(t) = \delta_{s,t}$ (which is 1 for $s = t$ and zero otherwise).

5.6 **Proposition** If $0 < p < 1$ and $\# S \geq 2$ then

(i) $\| \cdot \|_p$ violates subadditivity, thus is not a norm.

(ii) Nevertheless, $\ell^p(S, \mathbb{F})$ is a vector space.

(iii) Restricted to $\ell^p(S, \mathbb{F})$,

$$d_p(f, g) = \sum_{s \in S} |f(s) - g(s)|^p$$

defines a translation-invariant metric.

(iv) $\ell^p(S, \mathbb{F})$ a topological vector space when given the metric topology $\tau_{d_p}$.

**Proof.** (i) Pick $s, t \in S$, $s \neq t$ and put $f = \delta_s, g = \delta_t$. Now $\|f\|_p = \|g\|_p = 1$ and

$$2 < 2^{1/p} = \|f + g\|_p \leq \|f\|_p + \|g\|_p = 2$$

since $1/p > 1$. Thus $\| \cdot \|_p$ is not subadditive and therefore not a norm.

(ii) It is clear that $f \in \ell^p(S, \mathbb{F})$ implies $cf \in \ell^p(S, \mathbb{F})$ for all $c \in \mathbb{F}$. For $a, b \geq 0$ we have $(a + b)^p \leq (2 \max(a, b))^p \leq 2^p(a^p + b^p)$, whence the inequality

$$\|f + g\|_p = \sum_{s \in S} |f(s) + g(s)|^p \leq \sum_{s \in S} (|f(s)| + |g(s)|)^p \leq 2^p \sum_{s \in S} (|f(s)|^p + |g(s)|^p) = 2^p(\|f\|_p^p + \|g\|_p^p),$$

which still implies that $f + g \in \ell^p(S, \mathbb{F})$ for all $f, g \in \ell^p(S, \mathbb{F})$.  

22
(iii) That \( d_p(f, g) < \infty \) for all \( f, g \in \ell^p(S, \mathbb{F}) \) follows from \( \ell^p \) being a vector space. Translation invariance of \( d_p \) and the axioms \( d_p(f, g) = d_p(g, f) \) and \( d_p(f, g) = 0 \iff f = g \) are all evident from the definition. We claim that

\[
0 < p < 1, \ a, b \geq 0 \implies (a + b)^p \leq a^p + b^p.
\]

Believing this for a minute, we have

\[
d_p(f, h) = d_p(f - h, 0) = \sum_s |f(s) - h(s)|^p \leq \sum_s (|f(s) - g(s)| + |g(s) - h(s)|)^p
\]

\[
\leq \sum_s (|f(s) - g(s)|^p + |g(s) - h(s)|^p)
\]

\[
= d_p(f, g) + d_p(g, h),
\]

as wanted, where first used the triangle inequality and then the claim.

Turning to our claim \((a + b) \leq a^p + b^p\), it is clear that this holds if \( a = 0 \). For \( a = 1 \) it reduces to \((1 + b)^p \leq 1 + b^p \forall b \geq 0\). For \( b = 0 \) this is true, and for all \( b > 0 \) it follows from the fact that

\[
\frac{d}{db} \left( (1 + b)^p - (1 + b)^p\right) = p(b^{p-1} - (b + 1)^{p-1}) > 0
\]

due to \( p - 1 < 0 \). If now \( a > 0 \) then

\[
(a + b)^p = a^p(1 + (b/a))^p \leq a^p(1 + (b/a))^p = a^p + b^p,
\]

and we are done.

(iv) In view of \( d_p(f + g, 0) \leq d_p(f, 0) + d_p(g, 0) \), it is clear that the addition operation \( \ell^p \times \ell^p \to \ell^p \) is jointly continuous at \((0, 0)\), thus everywhere. It remains to show that scalar action \( \mathbb{F} \times \ell^p \to \ell^p \) is jointly continuous. By distributivity it suffices to do this at \((0, 0)\). Now

\[
d_p(cf, 0) = \sum_s |cf(s)|^p = |c|^p \sum_s |f(s)|^p = |c|^p d_p(f, 0),
\]

and this goes to zero as \((c, f)\) goes to zero in \( \mathbb{F} \times \ell^p \).

The above does not amount to a proof that the topological vector spaces \((\ell^p(S), d_p)\) with \( 0 < p < 1 \) are not normable, but this can be done, cf. Section B.5.1. (In fact they are not even locally convex.) This leads to strange behavior; for example the dual space \( \ell^p(S)^* \) is unexpected, cf. [46]. This strangeness is even more pronounced for the continuous versions \( L^p(X, \mathcal{A}, \mu) \): For \( X = [0, 1] \) equipped with Lebesgue measure, one has \( L^p(X, \mathcal{A}, \mu)^* = \{0\} \), which cannot happen for a non-zero Banach (or locally convex) space due to the Hahn-Banach theorem.

### 5.3 \( c_{00} \) and \( c_0 \). Completeness of \( \ell^p(S, \mathbb{F}) \) and \( c_0(S, \mathbb{F}) \)

In what follows we put

\[
d_p(f, g) = \left\{ \begin{array}{ll}
\|f - g\|_\infty = \sup_s |f(s) - g(s)| & \text{if } p = \infty \\
\|f - g\|_p = (\sum_s |f(s) - g(s)|^p)^{1/p} & \text{if } 1 \leq p < \infty \\
\sum_s |f(s) - g(s)|^p & \text{if } 0 < p < 1
\end{array} \right.
\]

which is a metric in all cases. For a function \( f : S \to \mathbb{F} \) we define \( \text{supp} f = \{s \in S \mid f(s) \neq 0\} \).
5.7 Definition For a set $S$ and $F \in \{ \mathbb{R}, \mathbb{C} \}$ we define

\[
\begin{align*}
    c_{00}(S, F) &= \{ f : S \to F \mid \#(\text{supp } f) < \infty \}, \\
c_0(S, F) &= \{ f : S \to F \mid \varepsilon > 0 \Rightarrow \#\{ s \in S \mid |f(s)| \geq \varepsilon \} < \infty \}.
\end{align*}
\]

(The elements of $c_0$ are the functions that ‘tend to zero at infinity’.)

5.8 Lemma If $0 < p \leq q < \infty$, we have

(i) $c_{00}(S, F) \subseteq \ell^p(S, F) \subseteq \ell^q(S, F) \subseteq c_0(S, F) \subseteq \ell^\infty(S, F),$

(ii) $\|f\|_q \leq \min(1, \|f\|_p^{p/q})$, thus $\|f\|_p \to 0 \Rightarrow \|f\|_q \to 0$ so that all inclusion maps $\ell^p(S) \hookrightarrow \ell^q(s)$ for $p \leq q$ are continuous.

Proof. (i) If $f \in c_{00}(S, F)$ then clearly $\|f\|_p < \infty$ for all $p \in (0, \infty]$. And $f \in c_0(S, F)$ implies boundedness of $f$. This gives the first and last inclusion.

If $f \in \ell^p(S, F)$ with $p \in (0, \infty)$ then finiteness of $\sum_{s \in S} |f(s)|^p$ implies that $\{ s \in S \mid |f(s)| \geq \varepsilon \}$ is finite for each $\varepsilon > 0$, thus $f \in c_0(S, F)$. In particular $F = \{ s \in S \mid |f(s)| \geq 1 \}$ is finite. If now $0 < p < q < \infty$ then

\[
\|f\|_q^q - \sum_{s \in F} |f(s)|^q = \sum_{s \in S \setminus F} |f(s)|^q = \sum_{s \in S \setminus F} |f(s)|^{p \cdot \frac{q}{p}} \leq \sum_{s \in S \setminus F} |f(s)|^p \leq \|f\|_p^p < \infty,
\]

since $q/p > 1$ and $|f(s)| < 1$, thus $|f(s)|^{q/p} \leq |f(s)|$, for all $s \in S \setminus F$. With the finiteness of $\sum_{s \in F} |f(s)|^q$ this implies $\sum_{s \in S} |f(s)|^q < \infty$, thus $f \in \ell^q(S, F)$.

(ii) It suffices to observe that $\|f\|_p < 1$ implies that the set $F$ in part (i) of the proof is empty, so that (5.3) reduces to $\|f\|_q^q \leq \|f\|_p^p$, thus $\|f\|_q \leq \|f\|_p^{p/q}$.

5.9 Exercise Let $S$ be an infinite set and $0 < p < q < \infty$. Prove that all inclusions in Lemma 5.8(i) are strict.

5.10 Lemma Let $p \in (0, \infty]$ and $d_p(x, y) = \|x - y\|_p$. Then $(\ell^p(S, F), d_p)$ is complete for every set $S$ and $F \in \{ \mathbb{R}, \mathbb{C} \}$.

Proof. Let $\{f_n\} \subseteq \ell^p(S, F)$ be a Cauchy sequence w.r.t. $d_p$, thus also w.r.t. $\| \cdot \|_p$. Then $|f_n(s) - f_m(s)| \leq \|f_n - f_m\|_p$, so that $\{f_n(s)\}$ is a Cauchy sequence in $F$, thus convergent for each $s \in S$. Defining $g(s) = \lim_n f_n(s)$, it remains to prove $g \in \ell^p(S, F)$ and $d_p(f_n, g) \to 0$.

For $p = \infty$ and $\varepsilon > 0$ we can find $n_0$ such that $n, m \geq n_0$ implies $\|f_n - f_m\|_\infty < \varepsilon$, which readily gives $\|f_m\|_\infty \leq \|f_n\|_\infty + \varepsilon$ for all $m \geq n_0$. Thus also $\|g\|_\infty \leq \|f_n\|_\infty + \varepsilon < \infty$. Taking $m \to \infty$ in $\sup_{s} |f_n(s) - f_m(s)| < \varepsilon$ gives $\lim_n f_n(s) - g(s) |\leq \varepsilon$, whence $\|f_n - g\|_\infty \to 0$.

For $0 < p < \infty$ we give a uniform argument. Since $\{f_n\}$ is Cauchy w.r.t. $d_p$, for $\varepsilon > 0$ we can find $n_0$ such that $n, m \geq n_0$ implies $d_p(f_n, f_m) < \varepsilon$. In particular $d_p(f_m, f_{n_0}) < \varepsilon$ for all $m \geq n_0$, thus also $d_p(g, f_{n_0}) < \varepsilon$, thus $g \in \ell^p(S, F)$. Applying the dominated convergence theorem (in the simple case of an infinite sum rather than a general integral, cf. Proposition A.3) to take $m \to \infty$ in $d_p(f_n, f_m) < \varepsilon$ gives $d(f_n, g) \leq \varepsilon$, whence $d(f_n, g) \to 0$.

5.11 Lemma (i) We have

\[
\frac{\|c_{00}(S, F)\|_p}{c_0(S, F)} = \begin{cases} 
\ell^p(S, F) & \text{if } 0 < p < \infty \\
c_0(S, F) & \text{if } p = \infty
\end{cases}
\]

(ii) $(c_0(S, F), \| \cdot \|_\infty)$ is complete.
5.4 Separability of $\ell^p(S, F)$. Then $\sum_{s \in S} |f(s)|^p = \|f\|_p^p$ implies that for each $\varepsilon > 0$ there is a finite $F \subseteq S$ such that $\|f\|_p - \sum_{s \in F} |f(s)|^p < \varepsilon$. Putting $g(s) = f(s)\chi_F(s)$, we have $g \in c_0(S, F)$ and $\|f - g\|_p^p = \sum_{s \in S\setminus F} |f(s)|^p < \varepsilon$. Since $\varepsilon > 0$ is arbitrary, $c_0 \in \ell^p$ is dense.

If $f \in c_0(S, F)$ and $\varepsilon > 0$ then $F = \{s \in S \mid |f(s)| \geq \varepsilon\}$ is finite. Now $g = f\chi_F$ is in $c_0(S, F)$ and $\|f - g\|_\infty < \varepsilon$, proving $f \in \overline{c_0(S, F)}|\|\infty$. And $f \in \overline{c_0(S, F)}|\|\infty$ means that for each $\varepsilon > 0$ there is a $g \in c_0(S, F)$ with $\|f - g\|_\infty < \varepsilon$. But this means $|f(s)| < \varepsilon$ for all $s \in S \setminus F$, where $F = \text{supp}(g)$ is finite. Thus $f \in c_0(S, F)$.

(ii) Being the closure of $c_0(S, F)$ in $\ell^\infty(S, F)$, $c_0(S, F)$ is closed, thus complete by completeness of $\ell^\infty(S, F)$, cf. Lemmas 5.10 and 4.4.

5.12 Remark Note that $\ell^\infty(S, F)$ is a commutative algebra under pointwise multiplication of functions. (In fact, $\ell^\infty(S, F) = C_b(S, F)$ if we equip $S$ with the discrete topology.) And $c_0(S, F) \subseteq \ell^\infty(S, F)$ is an ideal.

While the finitely supported functions are not dense in $\ell^\infty(S, F)$ (for infinite $S$), the finite-image functions are:

5.13 Lemma The set $\{f : S \to F \mid \#f(S) < \infty\}$ of functions assuming only finitely many values, equivalently, the set of finite linear combinations $\sum_{k=1}^K c_k \chi_{A_k}$ of characteristic functions, is dense in $\ell^\infty(S, F)$.

Proof. We prove this for $F = \mathbb{R}$, from which the case $F = \mathbb{C}$ is easily deduced. Let $f \in \ell^\infty(S, F)$ and $\varepsilon > 0$. For $k \in \mathbb{Z}$ define $A_k = f^{-1}(|k\varepsilon, (k + 1)\varepsilon))$. Define $K = \lceil \|f\|_\infty \rceil + 1$ and $g = \varepsilon \sum_{|k| \leq K} k \chi_{A_k}$. Then $g$ has finite image and $\|f - g\|_\infty < \varepsilon$.

5.4 Separability of $\ell^p(S, F)$ and $c_0(S, F)$

5.14 Proposition Let $p \in (0, \infty)$. The metric space $(\ell^p(S, F), d_p)$, where $d_p(f, g) = \|f - g\|_p$, is separable ($\iff$ second countable) if and only if the set $S$ is countable.

Proof. We prove this for $F = \mathbb{R}$, from which the claim for $F = \mathbb{C}$ is easily deduced. For $f : S \to \mathbb{R}$, let $\text{supp}(f) := \{s \in S \mid f(s) \neq 0\} \subseteq S$ be the support of $f$. Now, if $S$ is countable, then $Y = \{g : S \to \mathbb{Q} \mid \#(\text{supp}(g)) < \infty\} \subseteq \ell^p(S, \mathbb{R})$ is countable, and we claim that $\overline{Y} = \ell^p(S, \mathbb{R})$. To prove this, let $f \in \ell^p(S, \mathbb{R})$ and $\varepsilon > 0$. Since $\|f\|_p^p = \sum_{s \in S} |f(s)|^p < \infty$, there is a finite subset $T \subseteq S$ such that $\sum_{s \in S \setminus T} |f(s)|^p < \varepsilon/2$. On the other hand, since $\mathbb{Q} \# T \subseteq \mathbb{R} \# T$ is dense, we can choose $g : T \to \mathbb{Q}$ such that $\sum_{t \in T} |f(t) - g(t)|^p < \varepsilon/2$. Defining $g$ to be zero on $S \setminus T$, we have $g \in Y$ and

$$\|f - g\|^p = \sum_{t \in T} |f(t) - g(t)|^p + \sum_{s \in S \setminus T} |f(s)|^p < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon,$$

and since $\varepsilon > 0$ was arbitrary, $Y \subseteq \ell^p(S, \mathbb{R})$ is dense.

For the converse, assume that $S$ is uncountable. By Proposition A.2(iii), $\text{supp}(f)$ is countable for every $f \in \ell^p(S, \mathbb{R})$. Thus if $Y \subseteq \ell^p(S, \mathbb{R})$ is countable then $T = \bigcup_{f \in Y} \text{supp}(f) \subseteq S$ is a countable union of countable sets and therefore countable. Thus all functions $f \in Y$ vanish on $S \setminus T \neq \emptyset$, and the same holds for $f \in \overline{Y}$ since the coordinate maps $f \mapsto f(s)$ are continuous in view of $|f(s)| \leq \|f\|$. Thus $Y$ cannot be dense.

5.15 Exercise With $d_\infty(f, g) = \|f - g\|_\infty$ prove
(i) The space \((\ell^\infty(S, \mathbb{F}), d_\infty)\) is separable if and only if \(S\) is finite.
(ii) The space \((c_0(S, \mathbb{F}), d_\infty)\) is separable if and only if \(S\) is countable.

Hint: For (i), consider \(\{0, 1\}^S \subseteq \ell^\infty(S)\).

5.5 Dual spaces of \(\ell^p(S, \mathbb{F}), 1 \leq p < \infty, \) and \(c_0(S, \mathbb{F})\)

If \((V, \| \cdot \|)\) is a normed vector space over \(\mathbb{F}\) and \(\varphi : V \to \mathbb{F}\) is a linear functional, Definition 4.10 specializes to

\[
\|\varphi\| = \sup_{0 \neq x \in V} \frac{|\varphi(x)|}{\|x\|} = \sup_{x \in V, \|x\| \leq 1} |\varphi(x)|.
\]

Recall that the dual space \(V^* = \{ \varphi : V \to \mathbb{F} \text{ linear } | \|\varphi\| < \infty \}\) is a Banach space with norm \(\|\varphi\|\). The aim of this section is to concretely identify \(\ell^p(S, \mathbb{F})^*\) for \(1 \leq p < \infty\) and \(c_0(S, \mathbb{F})^*\).

(We will have something to say about \(\ell^\infty(S, \mathbb{F})^*\), but the complete story would lead us too far.)

For the purpose of the following proof, it will be useful to define \(\text{sgn} : \mathbb{C} \to \mathbb{C}\) by \(\text{sgn}(0) = 0\) and \(\text{sgn}(z) = z/|z|\) otherwise. Then \(z = \text{sgn}(z)|z|\) and \(|z| = \text{sgn}(z)|z|\) for all \(z \in \mathbb{C}\).

5.16 Theorem

(i) Let \(p \in [1, \infty]\) with conjugate value \(q\). Then for each \(g \in \ell^p(S, \mathbb{F})\) the map \(\varphi_g : \ell^q(S, \mathbb{F}) \to \mathbb{F}, f \mapsto \sum_{s \in S} f(s)g(s)\) satisfies \(\|\varphi_g\| \leq \|g\|_q\), thus \(\varphi_g \in \ell^p(S, \mathbb{F})^*\). And the map \(\iota : \ell^q(S, \mathbb{F}) \to \ell^p(S, \mathbb{F})^*, g \mapsto \varphi_g\), called the canonical map, is linear with \(\|\iota\| \leq 1\).

(ii) For all \(1 \leq p \leq \infty\) the canonical map \(\ell^q(S, \mathbb{F}) \to \ell^p(S, \mathbb{F})^*\) is isometric.

(iii) If \(1 \leq p < \infty\), the canonical map \(\ell^q(S, \mathbb{F}) \to \ell^p(S, \mathbb{F})^*\) is surjective, thus \(\ell^p(S, \mathbb{F})^* \cong \ell^q(S, \mathbb{F})\).

(iv) The canonical map \(\ell^1(S, \mathbb{F}) \to c_0(S, \mathbb{F})^*\) is an isometric bijection, thus \(c_0(S, \mathbb{F})^* \cong \ell^1(S, \mathbb{F})\).

(v) If \(S\) is finite, the canonical map \(\ell^1(S, \mathbb{F}) \to \ell^\infty(S, \mathbb{F})^*\) is surjective. If \(S\) is infinite, its image is a proper closed subspace of \(\ell^\infty(S, \mathbb{F})^*\).

Proof. (i) For all \(p \in [1, \infty]\) and conjugate \(q\) we have

\[
\left| \sum_s f(s)g(s) \right| \leq \sum_{s \in S} |f(s)g(s)| \leq \|f\|_p\|g\|_q < \infty \quad \forall f \in \ell^p, g \in \ell^q
\]

by Hölder’s inequality. In either case, the absolute convergence for all \(f, g\) implies that \((f, g) \mapsto \sum_s f(s)g(s)\) is bilinear.

(ii) If \(\|g\|_\infty \neq 0\) and \(\varepsilon > 0\) there is an \(s \in S\) with \(|g(s)| > \|g\|_\infty - \varepsilon\). If \(f = \delta_s : t \mapsto \delta_{s,t}\), we have \(|\varphi_g(f)| = |g(s)| > \|g\|_\infty - \varepsilon\). Since \(\|f\|_1 = 1\), this proves \(\|\varphi_g\| > \|g\|_\infty - \varepsilon\). Since \(\varepsilon > 0\) was arbitrary, we have \(\|\varphi_g\| \geq \|g\|_\infty\).

If \(\|g\|_1 \neq 0\), define \(f(s) = \text{sgn}(g(s))\). Then \(\|f\|_\infty = 1\) and \(\sum_s f(s)g(s) = \sum_s |g(s)| = \|g\|_1\). This proves \(\|\varphi_g\| \geq \|g\|_1\).

If \(1 < p, q < \infty\) and \(\|g\|_q \neq 0\), define \(f(s) = \text{sgn}(g(s))|g(s)|^{q-1}\). Then

\[
\sum_s f(s)g(s) = \sum_s |g(s)|^q = \|g\|_q^q,
\]

\[
\|f\|_p^p = \sum_s |f(s)|^p = \sum_{s, g(s) \neq 0} |g(s)|^{(q-1)p} = \sum_s |g(s)|^q = \|g\|_q^q,
\]

where we used \(p + q = pq\), whence \((q-1)p = q\). The above gives

\[
\|\varphi_g\| \geq \frac{\|\sum_s f(s)g(s)\|}{\|f\|_p} = \|g\|_q^q = \|g\|_{q'}^q = \|g\|_q^{(1-1/p)} = \|g\|_q.
\]
We thus have proven $\|\varphi_g\| \geq \|g\|_q$ in all cases and since the opposite inequality is known from (i), $g \mapsto \varphi_g$ is isometric.

(iii) Let $0 \neq \varphi \in \ell^1(S,F)^*$. Define $g : S \to F$ by $g(s) = \varphi(\delta_s)$. With $\|\varphi\|_1 = 1$, we have $\|g(s)\| = |\varphi(\delta_s)| \leq \|\varphi\|$ for all $s \in S$, thus $\|g\|_\infty \leq \|\varphi\|$. If $f \in \ell^1(S,F)$ and $F \subseteq S$ is finite, we have $\varphi(f) = \varphi(f_\delta) = \sum_{s \in F} f(s)\delta_s$.

In the limit $F \nearrow S$ this becomes $\varphi(f) = \sum_{s \in S} f(s)g(s)$ since $fg \in \ell^1$, thus the r.h.s. is absolutely convergent, and $\|f(1 - \chi_F)\|_1 = 0$ and $\varphi$ is $\|p\|_1$-continuous. This proves $\varphi = \varphi_g$ with $g \in \ell_\infty(S,F)$.

Now let $1 < p, q < \infty$, and let $0 \neq \varphi \in \ell^p(S,F)^*$. Since $\ell^1(S,F) \subseteq \ell^p(S,F)$ by Lemma 5.8, we can restrict $\varphi$ to $\ell^1(S,F)^*$, and the preceding argument gives a $g \in \ell_\infty(S,F)$ such that $\varphi(f) = \sum_{s \in S} f(s)g(s)$ for all $f \in \ell^1(S,F)$. The arguments in the proof of (ii) also show that for $1 < p, q < \infty$ and any function $g : S \to F$ we have

$$\|g\|_q = \sup \left\{ \left| \sum_{s \in S} f(s)g(s) \right| : f \in c_00(S,F), \|f\|_p \leq 1 \right\}.$$  

Using this and $\varphi(f) = \sum_{s \in S} f(s)g(s)$ for all $f \in c_00(S,F)$ we have

$$\|g\|_q = \sup \{ |\varphi(f)| : f \in c_00(S,F), \|f\|_p \leq 1\} = \|\varphi\| < \infty.$$  

Now $\varphi(f) = \sum_{s \in S} f(s)g(s) = \varphi_g(f)$ for all $f \in \ell^p(S,F)$ follows as before from $fg \in \ell^1$ and $\|f(1 - \chi_F)\|_p \to 0$ as $F \nearrow S$ and the $\|p\|_p$-continuity of $\varphi$.

(iv) Let $0 \neq g \in \ell^1(S,F)$. Then $\varphi_g \in \ell_\infty(S,F)^*$, which we can restrict to $c_0(S,F)$. For finite $F \subseteq S$ define $f_F = f\chi_F$ with $f(s) = \sgn(g(s))$. Then $f_F \in c_00(S,F)$ with $\|f_F\|_\infty = 1$ (provided $F \cap \text{supp} \, g \neq 0$) and $\varphi(f_F) = \sum_{s \in F} \sgn(g(s))$. Thus $\|\varphi\| \geq \sum_{s \in F} |g(s)|$ for all finite $F$ intersecting $\text{supp} \, g$, and this implies $\|\varphi\| \geq \|g\|_1$. The opposite being known, we have proven that $\ell^1(S,F) \to c_0(S,F)^*$ is isometric.

To prove surjectivity, let $0 \neq \varphi \in c_0(S,F)^*$ and define $g : S \to F, s \mapsto \varphi(\delta_s)$. If now $f \in c_0(S,F)$ and $F \subseteq S$ is finite, we have $f\chi_F = \sum_{s \in F} f(s)\delta_s$, thus $\varphi(f\chi_F) = \sum_{s \in F} f(s)g(s)$. In particular with $f(s) = \sgn(g(s))$ we have $\varphi(f\chi_F) = \sum_{s \in F} f(s)g(s) = \sum_{s \in F} |g(s)|$. Again we have $\|f\chi_F\|_\infty \leq \|f\|_1 = 1$, thus $|\varphi(f\chi_F)| \leq \|\varphi\|$, and combining these observations gives $\|g\|_1 \leq \|\varphi\| < \infty$, thus $g \in \ell^1(S,F)$. As $F \nearrow S$, we have $\|f(1 - \chi_F)\|_\infty = \|f\chi_S\|_\infty \to 0$ since $f \in c_0$, thus with $\|\cdot\|_\infty$-continuity of $\varphi$.

$$\varphi(f) = \lim_{F \nearrow S} \varphi(f\chi_F) = \lim_{F \nearrow S} \sum_{s \in F} f(s)g(s) = \sum_{s \in S} f(s)g(s) = \varphi_g(f),$$  

where we again used $fg \in \ell^1$. Thus $\varphi = \varphi_g$, so that $\ell^1(S,F) \to c_0(S,F)^*$ is an isometric bijection.

(v) It is clear that $\iota : \ell^1(S,F) \to \ell_\infty(S,F)^*$ is surjective if $S$ is finite. Closedness of the image of $\iota$ always follows from the completeness of $\ell^1(S,F)$ and the fact that $\iota$ is an isometry, cf. Corollary 4.6. The failure of surjectivity is deeper than the results of this section so far, so that it is illuminating to give two proofs.

First proof: If $S$ is infinite, the closed subspace $c_0(S,F) \subseteq \ell_\infty(S,F)$ is proper since $1 \in \ell_\infty(S,F) \setminus c_0(S,F)$. Thus the quotient space $Z = \ell_\infty(S,F)/c_0(S,F)$ is non-trivial. In Section 7 we will show that $Z$ is a Banach space, thus admits non-zero bounded linear maps $\psi : Z \to F$ by the Hahn-Banach theorem (Section 8), and that the quotient map $p : \ell_\infty(S,F) \to Z$ is bounded. Thus $\varphi = \psi \circ p$ is a non-zero bounded linear functional on $\ell_\infty(S,F)$ that vanishes on the closed subspace $c_0(S,F)$.

By (iv), the canonical map $\ell^1(S,F) \to c_0(S,F)^*$ is isometric, thus $\varphi_g$ with $g \in \ell^1(S,F)$ vanishes identically on $c_0(S,F)$ if and only if $g = 0$. Thus $\varphi \neq \varphi_g$ for all $g \in \ell^1(S,F)$.

Second proof: (This proof uses no unproven results from functional but the Stone-Cech compactification from general topology. Cf. Appendix A.3 and [46].) Since $S$ is discrete,
\( \ell^\infty(S, F) = C_b(S, F) \cong C(\beta S, F) \), where \( \beta S \) is the Stone-Čech compactification of \( S \). The isomorphism is given by the unique continuous extension \( C_b(S, F) \to C(\beta X, F), f \mapsto \hat{f} \) with the restriction map \( C(\beta S, \mathbb{R}) \to C_b(S, \mathbb{R}) \) as inverse. Since \( S \) is discrete and infinite, thus non-compact, \( \beta S \neq S \). If \( f \in c_0(S, F) \) then \( \hat{f}(x) = 0 \) for every \( x \in \beta S \setminus S \). (Proof: Let \( x \in \beta S \setminus S \). Since \( \beta X = \beta X \), we can find a net \( \{ x_i \} \) in \( X \) such that \( x_i \to x \). Since \( x \notin X \), \( x_i \) leaves every finite subset of \( X \). Now \( f \in c_0(S) \) and continuity of \( \hat{f} \) imply \( \hat{f}(x) = \lim \hat{f}(x_i) = \lim f(x_i) = 0. \) Thus for such an \( x \), the evaluation map \( \psi_x : C(\beta S, F) \to F, f \mapsto \hat{f}(x) \) gives rise to a non-zero bounded linear functional (in fact character) \( \varphi(f) = \hat{f}(x) \) on \( C_b(S, F) = \ell^\infty(S, F) \) that vanishes on \( c_0(S, F) \). Now we conclude as in the first proof that \( \varphi \neq \varphi_g \) for all \( g \in \ell^1(S, F) \).

5.17 REMARK 1. The two proofs of non-surjectivity of the canonical map \( \ell^1(S, F) \to \ell^\infty(S, F)^* \) for infinite \( S \) given above are both very non-constructive: The first used the Hahn-Banach theorem, which we will prove using Zorn’s lemma, equivalent to AC. The second used the Stone-Čech compactification \( \beta S \) whose usual construction relies on Tychonov’s theorem, which is equivalent to the axiom of choice. (But here we only need the restriction of Tychonov’s theorem to Hausdorff spaces, which is equivalent to the ‘ultrafilter lemma’, which is strictly weaker than AC. Also Hahn-Banach can be proven using only the ultrafilter lemma. See [46].)

2. In fact, the two proofs are essentially the same. The second proof implicitly uses the fact that \( \ell^\infty(S) \) and \( c_0(S) \) are algebras, so that we can consider characters instead of all linear functionals. Now, the characters on \( \ell^\infty(S) = C_b(S) = C(\beta S) \) correspond bijectively to the points of \( \beta S \), and those that vanish on \( c_0(S) \) correspond to \( \beta S \setminus S \). The first construction is more functional analytic, involving the Banach space quotient \( C_b(S)/c_0(S) \) and general functionals instead of characters.

3. The dual space of \( \ell^\infty(S, F) \) can be determined quite explicitly, but it is not a space of functions on \( S \) as are the spaces \( c_0(S, F)^* \) and \( \ell^p(S, F)^* \) for \( p < \infty \). It is the space \( ba(S, F) \) of ‘finitely additive \( F \)-valued measures on \( S \)’. Going into this would lead us too far, but see the supplementary Section B.2.

4. There are set theoretic frameworks without AC (but with \( DC_w \)) in which \( \ell^\infty(\mathbb{N})^* \cong \ell^1(\mathbb{N}) \), see [69, §23.10]. (In this situation, all finitely additive measures, cf. Section B.2, on \( \mathbb{N} \) are countably additive!)

5. For all \( p \in (0, 1) \), the dual space \( \ell^p(S, F)^* \) equals \( \{ \varphi_g : g \in \ell^\infty(S, F) \} = \ell^1(S, F)^* \). See [46, Appendix F.6]. Thus there is no \( p \)-dependence despite the fact that the \( \ell^p(S, F) \) are mutually non-isomorphic!

5.6 Outlook on general \( L^p \)-spaces

For an arbitrary measure space \( (X, \mathcal{A}, \mu) \) one can define normed spaces \( L^p(X, \mathcal{A}, \mu; F) \) in a broadly analogous fashion. (We will usually omit the \( F \).) Since integration on measure spaces is not among the formal prerequisites of these notes, we only sketch the basic facts referring to, e.g., [10, 67] for details. If \( f : X \to F \) is a measurable function and \( 0 < p < \infty \), then \( \| f \|_p = (\int |f(x)|^p d\mu(x))^{1/p} \in [0, \infty] \). If \( p = \infty \), put\(^{14}\)

\[
\| f \|_\infty = \text{ess sup}_x |f| = \inf \{ \lambda > 0 \mid \mu(\{ x \in X \mid |f(x)| > \lambda \}) = 0 \}.
\]

Now \( L^p(X, \mu) = \{ f : X \to \mathbb{F} \text{ measurable } \mid \| f \|_p < \infty \} \) is an \( \mathbb{F} \)-vector space for all \( p \in (0, \infty] \).

For \( 1 \leq p \leq \infty \), the proofs of the inequalities of Hölder and Minkowski extend to the present setting without any difficulties, so that the \( \| \cdot \|_p \) are seminorms on \( L^p(X, \mathcal{A}, \mu) \). But the latter

\(^{14}\)Warning: [10] defines \( \| \cdot \|_\infty \) using locally null sets instead of null sets, which is very non-standard.
fails to be a norm whenever there exists \( \emptyset \neq Y \in \mathcal{A} \) with \( \mu(Y) = 0 \) since then \( \|\chi_Y\|_p = 0 \). For this reason we define \( L^p(X,\mathcal{A},\mu) = L^p(X,\mathcal{A},\mu)/\{f \mid \|f\|_p = 0\} \). Now it is straightforward to prove that \( L^p(X,\mu) = L^p(X,\mu)/\sim \) is a normed space, and in fact complete. The proof now uses Proposition 4.2. If \( S \) is a set and \( \mu \) is the counting measure, we have \( \ell^\mu(S,\mathcal{F}) = L^\mu(S,P(S),\mu,\mathbb{F}) = L^\mu(S,P(S),\mu,\mathbb{F}) \).

A measurable function is called simple if it assumes only finitely many values. Equivalently it is of the form \( f(x) = \sum_{k=1}^K c_k \chi_{A_k}(x) \), where \( A_1, \ldots, A_k \) are measurable sets. Now one proves that the simple functions are dense in \( L^p \) for all \( p \in [1,\infty] \). If \( X \) is locally compact and \( \mu \) is nice enough, the set \( C_c(X,\mathbb{F}) \) of compactly supported continuous functions is dense in \( L^p(X,\mathcal{A},\mu;\mathbb{F}) \) for \( 1 \leq p < \infty \), while its closure in \( L^\infty \) is \( C_0(X,\mathbb{F}) \).

The inclusion \( \ell^p \subseteq \ell^q \) for \( p \leq q \) (Lemma 5.8) is false for general measure spaces! In fact, if \( \mu(X) < \infty \) then one has the reverse inclusion \( p \leq q \Rightarrow L^q(X,\mathcal{A},\mu) \subseteq L^p(X,\mathcal{A},\mu) \), while for general measure spaces there is no inclusion relation between the \( L^p \) with different \( p \).

If \( 1 < p, q < \infty \) are conjugate, the canonical map \( \varphi : L^q(X,\mathcal{A},\mu) \to L^p(X,\mathcal{A},\mu)^* \) is an isometric bijection for all measure spaces. That \( \varphi \) is an isometry is proven just as for the spaces \( \ell^p \): Hölder’s inequality gives \( \|\varphi_g\| \leq \|g\|_q \), and equality is proven as in Theorem 5.16(ii) by showing \( |\varphi_g(f)| \geq \|f\|_p \|g\|_1 \), where the \( f \in L^p \) are the same as before. However, the measure space \( X = \{x\}, \mathcal{A} = P(X) = \{\emptyset,X\} \) and \( : \emptyset \mapsto 0, X \mapsto +\infty \), for which \( L^1(X,\mathcal{A},\mu,\mathbb{F}) \cong \{0\} \) and \( \mathbb{F} \cong L^\infty(X,\mathcal{A},\mu,\mathbb{F}) \neq L^1(X,\mathcal{A},\mu,\mathbb{F})^* \), shows that isometry of \( L^\infty(X,\mathcal{A},\mu) \to (L^1(X,\mathcal{A},\mu))^* \) is not automatic. It is not hard to show that it holds if and only if \( (X,\mathcal{A},\mu) \) is semifinite, i.e.

\[
\mu(Y) = \sup\{\mu(Z) \mid Z \in \mathcal{A}, Z \subseteq Y, \mu(Z) < \infty\} \quad \forall Y \in \mathcal{A}.
\]

If \( 1 < p < \infty \), one still has surjectivity of \( L^p \to (L^q)^* \) for all measure spaces \( (X,\mathcal{A},\mu) \), but the standard proof is outside our scope since it requires the Radon-Nikodym theorem. (For a more functional-analytic proof see Section B.6.) In order for \( L^\infty \to (L^1)^* \) to be an isometric bijection, the measure space must be ‘localizable’, cf. [67]. This condition subsumes semifiniteness and is implied by \( \sigma \)-finiteness, to which case many books limit themselves.

Since we relegated the dual spaces \( \ell^\infty(S,\mathcal{F})^* \) to an appendix, we only remark that also in general \( L^\infty(X,\mathcal{A},\mu)^* \) is a space of finitely additive measures with fairly similar proofs, see [16]. For \( 0 < p < 1 \), the dual spaces \( (L^p)^* \) behave even stranger than \( (\ell^p)^* \). For example, \( L^p([0,1],\lambda;\mathbb{F})^* = \{0\} \).

6 Basics of Hilbert spaces

6.1 Inner products. Cauchy-Schwarz inequality

We have seen that every bounded linear functional \( \varphi \) on \( \ell^\mu(S,\mathcal{F}) \), where \( 1 \leq p < \infty \) is of the form \( \varphi_g : f \mapsto \sum_{s \in S} f(s)g(s) \) for a certain unique \( g \in \ell^q(S,\mathcal{F}) \). Here the conjugate exponent \( q \in (1,\infty] \) is determined by \( \frac{1}{p} + \frac{1}{q} = 1 \). Clearly we have \( p = q \) if and only if \( p = 2 \). In this space we have self-duality: \( \ell^2(S,\mathcal{F})^* \cong \ell^2(S,\mathcal{F}) \). The map

\[
\ell^2(S,\mathcal{F}) \times \ell^2(S,\mathcal{F}) \to \mathbb{F}, \ (f,g) \mapsto \sum_{s \in S} f(s)g(s)
\]

is bilinear and symmetric. Furthermore, it satisfies \( |\sum_{s \in S} f(s)g(s)| \leq \|f\|_2 \|g\|_2 \). Defining \( \overline{g}(s) = \overline{g(s)} \), we have \( \|\overline{g}\| = \|g\| \), so that also \( |\sum_{s \in S} f(s)g(s)| \leq \|f\|_2 \|g\|_2 \), which is the Cauchy-Schwarz inequality (in its incarnation for \( \ell^2(S,\mathbb{C}) \)).

For the development of a general, abstract theory it is better to adopt a slightly different definition:
6.1 Definition Let $V$ be an $\mathbb{F}$-vector space. An inner product on $V$ is a map $V \times V \to \mathbb{F}, (x, y) \mapsto \langle x, y \rangle$ such that

- The map $x \mapsto \langle x, y \rangle$ is linear for each choice of $y \in V$.
- $\langle y, x \rangle = \overline{\langle x, y \rangle} \forall x, y \in V$.
- $\langle x, x \rangle \geq 0 \forall x$, and $\langle x, x \rangle = 0 \Rightarrow x = 0$.

6.2 Remark 1. Many authors write $(x, y)$ instead of $\langle x, y \rangle$, but this leads to confusion with the notation for ordered pairs. We will use pointed brackets throughout.

2. If $\mathbb{F} = \mathbb{R}$, the complex conjugation has no effect and can be omitted. Then $\langle \cdot, \cdot \rangle$ is symmetric.

3. Combining the first two axioms one finds that the map $y \mapsto \langle x, y \rangle$ is anti-linear for each choice of $x$. This means $\langle x, cy + c'y \rangle = \overline{\langle x, y \rangle} + \overline{C} \langle x, y' \rangle$ for all $y, y' \in V$ and $c, c' \in \mathbb{F}$. Of course this reduces to linearity if $\mathbb{F} = \mathbb{R}$. A map $V \times V \to \mathbb{C}$ that is linear in one variable and anti-linear in the other is called sesquilinear.

4. A large minority of authors, mostly (mathematical) physicists, defines inner products to be linear in the second and anti-linear in the first argument. We follow the majority use like [41].

5. The first two axioms together already imply $\langle x, x \rangle \in \mathbb{R}$ for all $x$, but not the positivity assumption.

6. If $\langle x, y \rangle = 0$ for all $y \in H$ then $x = 0$. To see this, it suffices to take $y = x$. 

6.3 Example 1. If $V = \mathbb{C}^n$ then $\langle x, y \rangle = \sum_{k=1}^{n} x_k \overline{y_k}$ is an inner product and the corresponding norm (see below) is $\| \cdot \|_2$, which is complete.

2. Let $S$ be any set and $V = \ell^2(S, \mathbb{C})$. Then $\langle f, g \rangle = \sum_{s \in S} f(s) \overline{g(s)}$ converges for all $f, g \in V$ by Hölder’s inequality and is easily seen to be an inner product. Of course, 1. is a special case of 2.

3. If $(X, \mathcal{A}, \mu)$ is any measure space then $\langle f, g \rangle = \int_X f(x) \overline{g(x)} \, d\mu(x)$ is an inner product on $L^2(X, \mathcal{A}, \mu; \mathbb{F})$ turning it into a Hilbert space. (Here we allow ourselves a standard sloppiness: The elements of $L^p$ are not functions, but equivalence classes of functions. The inner product of two such classes is defined by picking arbitrary representers.)

4. Let $V = M_{n \times n}(\mathbb{C})$. For $a, b \in V$, define $\langle a, b \rangle = \text{Tr}(b^*a) = \sum_{i,j=1}^{n} a_{ij} \overline{b_{ij}}$, where $(b^*)_{ij} = \overline{b_{ji}}$. That this is an inner product turning $V$ into a Hilbert space follows from 1. upon the identification $M_{n \times n}(\mathbb{C}) \cong \mathbb{C}^{n^2}$.

In view of $\langle x, x \rangle \geq 0$ for all $x$, and we agree that $\langle x, x \rangle^{1/2}$ always is the positive root.

6.4 Lemma (Abstract Cauchy-Schwarz inequality) \footnote{Augustin-Louis Cauchy (1789-1857). French mathematician with many important contributions to analysis. Karl Hermann Armandus Schwarz (1843-1921). German mathematician, mostly active in complex analysis. Some authors (mostly Russian ones) include Viktor Yakovlevich Bunyakovski (1804-1889). Russian mathematician.} If $\langle \cdot, \cdot \rangle$ is an inner product on $V$ then

$$\forall x, y \in V.$$

This even holds if one drops the assumption that $\langle x, x \rangle = 0 \Rightarrow x = 0.$

Equality holds in (6.1) if and only if one of the vectors is zero or $x = cy$ for some $c \in \mathbb{F}$.

6.5 Exercise Prove Lemma 6.4 along the following lines:
1. Prove it for \( y = 0 \), so that we may assume \( y \neq 0 \) from now on.
2. Define \( x_1 = \|y\|^{-2}\langle x, y \rangle y \) and \( x_2 = x - x_1 \) and prove \( \langle x_1, x_2 \rangle = 0 \).
3. Use 2. to prove \( \|x\|^2 = \|x_1\|^2 + \|x_2\|^2 \geq \|x_1\|^2 \).
4. Deduce Cauchy-Schwarz from \( \|x_1\|^2 \leq \|x\|^2 \).
5. Prove the claim about equality.

The above proof is the easiest to memorize (at least in outline) and reconstruct, but there are many others, e.g.:

6.6 Exercise (i) For \( x, y \in V \), define \( P(t) = \|x + ty\|^2 \) and show this defines a quadratic polynomial in \( t \in \mathbb{C} \) with real coefficients.
(ii) Use the obvious fact that this polynomial takes values in \([0, \infty)\) for all \( t \in \mathbb{R} \), thus also \( \inf_{t \in \mathbb{R}} P(t) \geq 0 \), to prove the Cauchy-Schwarz inequality.

6.7 Proposition If \( \langle \cdot, \cdot \rangle \) is an inner product on \( V \) then \( \|x\| = +\sqrt{\langle x, x \rangle} \) is a norm on \( V \).

Proof. \( \|x\| \geq 0 \) holds by construction, and the third axiom in Definition 6.1 implies \( \|x\| = 0 \Rightarrow x = 0 \). We have
\[
\|cx\| = \sqrt{\langle cx, cx \rangle} = \sqrt{c^2 \langle x, x \rangle} = |c|\|x\|,
\]
thus \( \|cx\| = |c|\|x\| \) for all \( x \in V, c \in \mathbb{F} \). Finally,
\[
\|x + y\|^2 = \langle x + y, x + y \rangle = \langle x, x \rangle + \langle x, y \rangle + \langle y, x \rangle + \langle y, y \rangle = \|x\|^2 + \|y\|^2 + \langle x, y \rangle + \langle y, x \rangle.
\]
With \( \Re z \leq |z| \) for all \( z \in \mathbb{C} \) and the Cauchy-Schwarz inequality we have
\[
\langle x, y \rangle + \langle y, x \rangle = \langle x, y \rangle + \overline{\langle x, y \rangle} = 2 \Re \langle x, y \rangle \leq 2|\langle x, y \rangle| \leq 2\|x\||\|y\|, \]
thus
\[
\|x + y\|^2 \leq \|x\|^2 + \|y\|^2 + 2\|x\||\|y\| = (\|x\| + \|y\|)^2
\]
and therefore \( \|x + y\| \leq \|x\| + \|y\| \), i.e. subadditivity. \( \blacksquare \)

In terms of the norm, the Cauchy-Schwarz inequality just becomes \( |\langle x, y \rangle| \leq \|x\||\|y\| \).

6.8 Definition A pre-Hilbert space (or inner product space) is a pair \((V, \langle \cdot, \cdot \rangle)\), where \( V \) is an \( \mathbb{F} \)-vector space and \( \langle \cdot, \cdot \rangle \) an inner product on it. A Hilbert space is a pre-Hilbert space that is complete for the norm \( \| \cdot \| \) obtained from the inner product.

6.9 Remark 1. By the above, an inner product gives rise to a norm and therefore to a norm topology \( \tau \). Now the Cauchy-Schwarz inequality implies that the inner product \( \langle \cdot, \cdot \rangle \to \mathbb{F} \) is jointly continuous:
\[
|\langle x, y \rangle - \langle x', y' \rangle| = |\langle x, y \rangle - \langle x, y' \rangle + \langle x', y' \rangle - \langle x', y' \rangle|
\leq \|x\||\|y - y'\| + \|x - x'\||\|y'\|.
\]

2. If \( \langle \cdot, \cdot \rangle \) be an inner product on \( H \) and \( \| \cdot \| \) is the norm derived from it then
\[
\|x\| = \sup_{y \in H, \|y\| = 1} |\langle x, y \rangle| \quad \forall x \in H. \tag{6.2}
\]

31
(For $x = 0$ this is obvious, and for $x \neq 0$ it follows from $\langle x, \frac{x}{\|x\|} \rangle = \|x\|$)

3. The restriction of an inner product on $H$ to a linear subspace $K \subseteq H$ again is an inner product. Thus if $H$ is a Hilbert space and $K$ a closed subspace then $K$ again is a Hilbert space (with the restricted inner product).

4. All spaces considered in Example 6.3 are complete, thus Hilbert spaces. For $\ell^2(S)$ this was proven in Section 5., and the claim for $C^n$, thus also $M_{n \times n}(C)$, follows since $C^n \cong \ell^2(S, C)$ when $\#S = n$. For $L^2(X, \mathcal{A}, \mu)$ see books on measure theory like [10, 67].

6.10 Definition Let $(H_1, \langle \cdot, \cdot \rangle_1), (H_2, \langle \cdot, \cdot \rangle_2)$ be pre-Hilbert spaces. A linear map $A : H_1 \to H_2$ is called

- **isometry** if $\langle Ax, Ay \rangle_2 = \langle x, y \rangle_1 \ \forall x, y \in H_1$.
- **unitary** if it is a surjective isometry.

6.11 Remark Every unitary map is invertible and its inverse is also unitary. Two Hilbert spaces $H_1, H_2$ are called unitarily equivalent or isomorphic if there exists a unitary $U : H_1 \to H_2$.

If $(H_1, \langle \cdot, \cdot \rangle_1), (H_2, \langle \cdot, \cdot \rangle_2)$ are (pre)Hilbert spaces then

$$\langle (x_1, x_2), (y_1, y_2) \rangle = \langle x_1, y_1 \rangle_1 + \langle x_2, y_2 \rangle_2$$

defines an inner product on $H_1 \oplus H_2$ turning it into a (pre)Hilbert space. More generally, if $\{H_i, \langle \cdot, \cdot \rangle_i\}_{i \in I}$ is a family of (pre)Hilbert spaces then

$$\bigoplus_{i \in I} H_i = \{\{x_i\}_{i \in I} \mid \sum_{i \in I} \langle x_i, x_i \rangle_i < \infty\}$$

with

$$\langle \{x_i\}, \{y_i\} \rangle = \sum_{i \in I} \langle x_i, y_i \rangle_i$$

is a (pre)Hilbert space. (If $H_i = \mathbb{F}$ for all $i \in I$, this construction recovers $\ell^2(I, \mathbb{F})$, while the Banach space direct sum gives $\ell^1(I, \mathbb{F})$.)

6.12 Exercise Let $(V, \langle \cdot, \cdot \rangle)$ be a pre-Hilbert space and $\|\cdot\|$ the associated norm. Let $(V', \|\cdot\|')$ be the completion (as a normed space) of $(V, \|\cdot\|)$. Prove that $V'$ is a Hilbert space.

6.2 The parallelogram and polarization identities

Given a normed space $(V, \|\cdot\|)$, it is natural to ask whether there exists an inner product on $V$ giving rise to the given norm.

6.13 Exercise Let $(V, \langle \cdot, \cdot \rangle)$ be a pre-Hilbert space over $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$. Prove the parallelogram identity

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

and the polarization identities

$$\langle x, y \rangle = \frac{1}{4} \sum_{k=0}^{3} i^k \|x + i^k y\|^2 \text{ if } \mathbb{F} = \mathbb{C},$$

$$\langle x, y \rangle = \frac{1}{4} (\|x + y\|^2 - \|x - y\|^2) \text{ if } \mathbb{F} = \mathbb{R},$$

32
6.14 Remark The proof of (6.4) only uses the sesquilinearity of \( \langle \cdot, \cdot \rangle \), so that the polarization identity \( \langle x, y \rangle = \frac{1}{4} \sum_{k=0}^{3} i^k [x + i^k y, x + i^k y] \) holds for every sesquilinear form \( [\cdot, \cdot] \) over \( \mathbb{C} \). \( \square \)

For a map of (pre)Hilbert spaces we have two a priori different notions of isometry, but they are equivalent:

6.15 Exercise Let \((H_1, \langle \cdot, \cdot \rangle_1), (H_2, \langle \cdot, \cdot \rangle_2)\) be (pre)Hilbert spaces over \( \mathbb{C} \). Let \( \| \cdot \|_{1,2} \) be the norms induced by the inner products. Prove that if a linear map \( A : H_1 \to H_2 \) is an isometry of normed spaces then it is an isometry of pre-Hilbert spaces. (I.e. if \( \| Ax \|_2 = \| x \|_1 \) \( \forall x \in H_1 \) then \( \langle Ax, Ay \rangle_2 = \langle x, y \rangle_1 \) \( \forall x, y \in H_1 \).) Hint: Polarization.

The polarization identities actually characterize norms coming from inner products:

6.16 Exercise Let \((V, \| \cdot \|)\) be a normed space over \( \mathbb{C} \) whose norm satisfies (6.3). Take (6.4) as definition of \( \langle \cdot, \cdot \rangle \) and prove that this is an inner product. Thus (6.3) is necessary and sufficient for the norm to come from an inner product. (See the hints in [41, Exercise 1.14].)

The above is only one of very many characterizations of Hilbert spaces among the Banach spaces. There is a whole book [2] about them!

6.17 Exercise Let \( S \) be a set with \#\( S \geq 2 \). Prove that \((\ell^p(S, \mathbb{F}), \| \cdot \|_p)\), where \( p \in [1, \infty] \), satisfies the parallelogram identity if and only if \( p = 2 \).

6.3 Orthogonality, subspaces, orthogonal projections

6.3.1 Basic Hilbert space geometry

6.18 Definition If \( H \) is a (pre)Hilbert space then \( x, y \in H \) are called orthogonal, denoted \( x \perp y \), if \( \langle x, y \rangle = 0 \). If \( S, T \subseteq H \) then \( S \perp T \) means \( x \perp y \) \( \forall x \in S, y \in T \).

It should be obvious that \( x \perp y \) implies \( cx \perp dy \) for all \( c, d \in \mathbb{F} \).

6.19 Lemma (Pythagoras’ theorem) Let \( H \) be a (pre)Hilbert space and \( x_1, \ldots, x_n \in H \) mutually orthogonal, i.e. \( i \neq j \Rightarrow \langle x_i, x_j \rangle = 0 \). Let \( x = x_1 + \cdots + x_n \). Then
\[
\| x \|^2 = \| x_1 \|^2 + \cdots + \| x_n \|^2.
\]

Proof. We have
\[
\| x \|^2 = \langle x, x \rangle = \left( \sum_i x_i, \sum_j x_j \right) = \sum_{i,j} \langle x_i, x_j \rangle = \sum_i \langle x_i, x_i \rangle = \sum_i \| x_i \|^2,
\]
where we used \( i \neq j \Rightarrow \langle x_i, x_j \rangle = 0 \). \( \blacksquare \)

6.20 Remark If \( H \) is a Hilbert space, \( I \) is an infinite set and \( \{ x_i \}_{i \in I} \subseteq H \) is such that \( \sum_i \| x_i \| < \infty \) then we can make sense of \( x = \sum_{i \in I} x_i \in H \) by completeness and Proposition 4.2. If all \( x_i \) are mutually orthogonal then by taking the limit over finite subsets we again have \( \| x \|^2 = \sum_{i \in I} \| x_i \|^2 \). (This shows that \( \sum_i \| x_i \| < \infty \Rightarrow \sum_i \| x_i \|^2 < \infty \), which also follows from the inclusion \( \ell^1(S) \subseteq \ell^2(S) \) proven in Lemma 5.8.) \( \square \)

6.21 Definition Let \( V \) be an \( \mathbb{F} \)-vector space. Then \( C \subseteq V \) is called convex if \( x, y \in C, t \in [0, 1] \Rightarrow tx + (1-t)y \in C \).
6.22 Proposition (Riesz lemma) \(^{16}\) Let \(H\) be a Hilbert space and \(C \subseteq H\) a non-empty closed convex set. Then for each \(x \in H\) there is a unique \(y \in C\) minimizing \(\|x - y\|\), i.e. 
\[\|x - y\| = \inf_{z \in C} \|x - z\|\].

Proof. We will prove this for \(x = 0\), in which case the statement says that there is a unique element of \(C\) of minimal norm. For general \(x \in H\), let \(y'\) be the unique element of minimal norm in the convex set \(C' = C - x\). Then \(y = y' + x\) is the unique element in \(C\) minimizing \\
\[\|x - y\|\].

Let \(d = \inf_{z \in C} \|z\|\) and pick a sequence \(\{y_n\}\) in \(C\) such that \(\|y_n\| \to d\). Now with the parallelogram identity (6.3) we have
\[\|y_n - y_m\|^2 = 2\|y_n\|^2 + 2\|y_m\|^2 - 2\|y_n + y_m\|^2\]
\[= 2\|y_n\|^2 + 2\|y_m\|^2 - 4 \left(\frac{\|y_n + y_m\|^2}{2}\right).

As \(n, m \to \infty\) we have \(\lim_{n, m \to \infty} 2\|y_n\|^2 + 2\|y_m\|^2 = 4d^2\). Since \(C\) is convex, we have \(\frac{y_n + y_m}{2} \in C\), thus \(\|\frac{y_n + y_m}{2}\| \geq d\) for all \(n, m\), so that \(\limsup_{n, m \to \infty} \|y_n - y_m\|^2 \leq 0\). Since this \(\limsup\) also must be positive, it follows that \(\|y_n - y_m\| \to 0\) as \(n, m \to \infty\). Thus \(\{y_n\}\) is a Cauchy sequence and therefore converges to some \(y \in C\) by completeness of \(H\) and closedness of \(C\). By continuity of the norm, \(\|y\| = \lim \|y_n\| = d\).

If \(y, y' \in C\) with \(\|y\| = \|y'\| = d\) then the parallelogram identity gives \(\|y - y'\|^2 = 4d^2 - 4 \left(\frac{\|y + y'\|^2}{2}\right) \geq 0\). Since again \(\|(y + y')/2\|^2 \geq d^2\), we have \(\|y - y'\| = 0\), thus the uniqueness claim.

6.3.2 Closed subspaces, orthogonal complement, and orthogonal projections

6.23 Definition Let \((H, \langle \cdot, \cdot \rangle)\) be a Hilbert space and \(S \subseteq H\). The orthogonal complement \(S^\perp\) is defined as
\[S^\perp = \{y \in H \mid \langle y, x \rangle = 0 \ \forall x \in S\}.

Here some easy facts without proofs: For every \(S \subseteq H\) we have \(\overline{S}^\perp = S^\perp\), and \(S^\perp \subseteq H\) is a closed linear subspace. If \(S \subseteq T \subseteq H\) then \(T^\perp \subseteq S^\perp\).

A linear subspace of a vector space clearly is a convex subset. Now,

6.24 Theorem Let \(H\) be a Hilbert space and \(K \subseteq H\) a closed linear subspace. Define a map \(P : H \to K\) by \(Px = y\), where \(y \in K\) minimizes \(\|x - y\|\) as in Proposition 6.22. Also define \(Qx = x - Px\). Then

(i) \(Qx \in K^\perp \ \forall x\).

(ii) For each \(x \in H\) there are unique \(y \in K, z \in K^\perp\) with \(x = y + z\), namely \(y = Px, z = Qx\).

(iii) The maps \(P, Q\) are linear.

(iv) The map \(U : H \to K \oplus K^\perp, x \mapsto (Px, Qx)\) is an isomorphism of Hilbert spaces. In particular, \(\|x\|^2 = \|Px\|^2 + \|Qx\|^2\) \(\forall x\).

(v) The map \(P : H \to H\) satisfies \(P^2 = P\) and \(\langle Px, y \rangle = \langle x, Py \rangle\). The same holds for \(Q\).

\(^{16}\)Frigyes Riesz (1880-1956). Hungarian mathematician and one of the pioneers of functional analysis. (The same applies to his younger brother Marcel Riesz (1886-1969).)
Proof. (i) Let \( x \in H, v \in K \). We want to prove \( Qx \perp v \), i.e. \( \langle x - Px, v \rangle = 0 \). Since \( y = Px \) is the element of \( K \) minimizing \( \|x - y\| \), we have for all \( t \in \mathbb{C} \)
\[
\|x - Px\| \leq \|x - Px - tv\|.
\]
Taking squares and putting \( z = x - y = x - Px \), this becomes \( \langle z, z \rangle \leq \langle z - tv, z - tv \rangle \), equivalent to
\[
2 \Re\langle t(v, z) \rangle \leq |t|^2\|v\|^2.
\]
With the polar decomposition \( t = |t|e^{i\varphi} \), the above becomes \( 2 \Re(e^{i\varphi}\langle v, z \rangle) \leq |t|^2\|v\|^2 \). Taking \( |t| \to 0 \), we find \( \Re(e^{i\varphi}\langle v, z \rangle) = 0 \), and since \( \varphi \) was arbitrary, we conclude \( \langle v, z \rangle = 0 \). In view of \( z = x - y = x - Px \) this is what we wanted.

(ii) For each \( x \in H \) we have \( x = Px + Qx \) with \( Px \in K, Qx \in K^\perp \), proving the existence. If \( y, y' \in K, z, z' \in K^\perp \) such that \( y + z = y' + z' \) then \( y - y' = z' - z \in K \cap K^\perp = \{0\} \). Thus \( y - y' = z' - z = 0 \), proving the uniqueness.

(iii) If \( x, x' \in H, c, c' \in \mathbb{F} \) then \( cx + c'x' = P(cx + c'x') + Q(cx + c'x') \). But also
\[
 cx + c'x' = c(Px + Qx) + c'(Px' + Qx') = (cPx + c'Px') + (cQx + c'Qy')
\]
is a decomposition of \( cx + c'x' \) as a sum of vectors in \( K \) and \( K^\perp \), respectively. Since such a decomposition is unique by (ii), we have \( P(cx + c'x') = cPx + c'Px' \) and \( Q(cx + c'x') = cQx + c'Qy' \), thus \( P, Q \) are linear.

(iv) It is clear that \( U \) is a linear isomorphism. Furthermore, \( Px \perp Qy \) implies
\[
\langle x, y \rangle = \langle Px + Qx, Py + Qy \rangle = \langle Px, Py \rangle + \langle Qx, Qy \rangle = \langle Ux, Uy \rangle,
\]
so that \( U \) is an isometry of Hilbert spaces.

(v) In view of (iv), i.e. \( U : H \xrightarrow{\sim} K \oplus K^\perp \), it suffices to prove this for the map \( S : H_1 \oplus H_2 \to H_1 \oplus H_2, (x_1, x_2) \mapsto (x_1, 0) \). It is clear that \( S^2 = S \circ S = S \). And
\[
\langle S(x_1, x_2), (y_1, y_2) \rangle = \langle (x_1, 0), (y_1, y_2) \rangle = \langle x_1, y_1 \rangle_1 = \langle (x_1, x_2), (y_1, 0) \rangle = \langle (x_1, x_2), S(y_1, y_2) \rangle.
\]

\[\blacksquare\]

6.25 Remark The above theorem remains valid if \( H \) is only a pre-Hilbert space, provided \( K \subseteq H \) is finite dimensional. We first note that the proof of Lemma 6.22 only uses completeness of \( C \subseteq H \), not that of \( H \). And we recall that finite dimensional subspaces of normed spaces are automatically complete and closed, cf. Exercises 4.21 and 4.22. In the proof of Theorem 6.24 we use Lemma 6.22 with \( C = K \), which is complete as just noted.

6.26 Exercise Let \( H \) be a Hilbert space and \( V \subseteq H \) a linear subspace. Prove:

(i) \( V^\perp = \{0\} \) if and only if \( V = H \).

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

6.27 Definition Let \( V \) be a vector space and \( H \) a (pre)Hilbert space.

- A linear map \( P : V \to V \) is idempotent if \( P^2 = P \circ P = P \).
- A linear map \( P : H \to H \) is self-adjoint if \( \langle Px, y \rangle = \langle x, Py \rangle \) for all \( x, y \in H \).
- A bounded linear map \( P : H \to H \) is an orthogonal projection if it is a self-adjoint idempotent.
(In Theorem 9.9 we will prove that every self-adjoint $P : H \to H$ is automatically bounded, but this is not needed here.)

We have seen that every closed subspace $K$ of a Hilbert space gives rise to an orthogonal projection $P$ with $PH = K$. Conversely, we have:

6.28 Exercise Let $H$ be a Hilbert space and $P$ an idempotent on $H$ (i.e. in $B(H)$). Prove:

(i) $K = PH \subseteq H$ and $L = (1 - P)H$ are closed linear subspaces.

(ii) We have $K \perp L$ if and only if $P = P^*$, i.e. $P$ is an orthogonal projection.

(iii) If $P$ is an orthogonal projection then it equals the $P$ associated to $K$ by Theorem 6.24.

6.4 The dual space $H^*$ of a Hilbert space

If $H$ is a Hilbert space, every $y \in H$ gives rise to a bounded linear functional on $H$ via $\varphi_y : x \mapsto \langle x, y \rangle$. The Cauchy-Schwarz inequality implies $\|\varphi_y\| \leq \|y\|$. For every $y \neq 0$ we have $\varphi_y(y) = \langle y, y \rangle = \|y\|^2 > 0$, implying $\|\varphi_y\| = \|y\|$. As a consequence, for every non-zero $x \in H$ there is a $\varphi \in H^*$ with $\varphi(x) \neq 0$. Thus $H^*$ separates the points of $H$. But we have more:

6.29 Theorem (Riesz-Frèchet representation theorem) If $H$ is a Hilbert space and $\varphi \in H^*$ then there is a unique $y \in H$ such that $\varphi = \varphi_y = \langle \cdot, y \rangle$.

Proof. See [41, Theorem 1.29]. If $\varphi = 0$, put $y = 0$ (every other $y$ gives $\varphi_y \neq 0$). Now assume $\varphi \neq 0$. Let $K = \ker \varphi = \varphi^{-1}(0)$. Then $K \subseteq H$ is a proper (since $\varphi \neq 0$) closed (by continuity) linear subspace of $H$. By the preceding section, $K^\perp \neq \{0\}$. The dimension of $K^\perp$ is one. (If $y_1, y_2 \in K^\perp \setminus\{0\}$ then $\varphi(y_1) \neq 0 = \varphi(y_2)$ implies $\varphi(y_1 - y_2) = 0$ and therefore $y_1 - y_2 \in K \cap K^\perp = \{0\}$, thus $y_1$ and $y_2$ are linearly dependent.) Pick a non-zero $z \in K^\perp$ and put $y = \frac{\varphi(y_1)}{\|z\|^2} z$. Then $\varphi_y$ vanishes on $K$, and $\varphi_y(z) = \langle z, y \rangle = \varphi(z)$. Thus $\varphi = \varphi_y$ on both $K$ and $K^\perp = \mathbb{C}z$, thus on $H = K + K^\perp$. In view of the construction, $y$ is unique.

By the above, the map $H \to H^*$, $y \mapsto \varphi_y$ is an isometric bijection of Banach spaces. If we want, we can use it to put an inner product on $H^*$.

6.30 Exercise Let $H$ be a Hilbert space and $K \subseteq H$ a linear subspace.

(i) Prove that for every $\varphi \in K^*$ there exists $\hat{\varphi} \in H^*$ such that $\hat{\varphi} | K = \varphi$.

(ii) Prove that $\hat{\varphi} \in H^*$ with $\hat{\varphi} | K = \varphi$ is unique if and only $\overline{K} = H$.

(iii) Prove that there is a unique $\hat{\varphi} \in H^*$ satisfying $\hat{\varphi} | K = \varphi$ and $\|\hat{\varphi}\| = \|\varphi\|$.

6.5 Orthonormal sets. Bases

We begin by recalling the notion of bases from linear algebra: A finite subset $\{x_1, \ldots, x_n\}$ of a vector space $V$ over the field $k$ is called linearly independent if $\sum_{i=1}^n c_i x_i = 0$, where $c_1, \ldots, c_n \in k$, implies $c_1 = \cdots = c_n = 0$. (In particular, $x_i \neq 0 \forall i$.) An arbitrary subset $B \subseteq V$ is called linearly independent if every finite subset $S \subseteq B$ is linearly independent. A linearly independent subset $B \subseteq V$ is called a base (or Hamel base) if every $x \in V$ can be written as a linear combination of finitely many elements of $B$. This is equivalent to $B$ being maximal, i.e. non-existence of linearly independent sets $B' \supsetneq B$ properly containing $B$. One now proves that any two bases of $V$ have the same cardinality. All this is known from linear algebra, but the following possibly not:
6.31 Proposition Every vector space \( V \) has a base.

Proof. If \( V = \{0\} \), \( \emptyset \) is a base. Thus let \( V \) be non-zero and let \( \mathcal{B} \) be the set of linearly independent subsets of \( V \). The set \( \mathcal{B} \) is partially ordered by inclusion \( \subseteq \) and non-empty (since it contains \( \{x\} \) for all \( 0 \neq x \in V \)). We claim that every chain in (=totally ordered subset of) \((\mathcal{B}, \subseteq)\) has a maximal element: Just take the union \( \hat{\mathcal{B}} \) of all sets in the chain. Since any finite subset of the union over a chain of sets is contained in some element of the chain, every finite subset of \( \hat{\mathcal{B}} \) is linearly independent. Thus \( \hat{\mathcal{B}} \) is in \( \mathcal{B} \) and clearly is an upper bound of the chain. Thus the assumption of Zorn’s Lemma is satisfied, so that \((\mathcal{B}, \subseteq)\) has a maximal element \( M \).

We claim that \( M \) is a basis for \( V \): If this was false, we could find a \( v \in V \) not contained in the span of \( M \). But then \( M \cup \{v\} \) would be a linearly independent set strictly larger than \( M \), contradicting the maximality of \( M \). \( \blacksquare \)

The linear algebra notion of base is quite irrelevant as soon as we are concerned with topological vector spaces, like normed spaces since in the presence of a topology we can also talk about infinite linear combinations. This leads to the notion of a Hilbert space base and the above purely algebraic one is of little or no relevance.

6.32 Definition Let \( H \) be a (pre)Hilbert space. A subset \( E \subseteq H \) is orthonormal if

- \( \|x\| = 1 \) \( \forall x \in E \).
- If \( x, y \in E, x \neq y \) then \( \langle x, y \rangle = 0 \).

(Equivalently, \( \langle x, y \rangle = \delta_{x,y} \forall x, y \in E \).)

6.33 Exercise Prove that every orthonormal set is linearly independent.

6.34 Lemma Let \( H \) be a (pre)Hilbert space and \( E \subseteq H \) an orthonormal set. Then the Bessel inequality

\[
\sum_{e \in E} |\langle x, e \rangle|^2 \leq \|x\|^2 \quad \forall x \in H
\]  

(6.6)

holds.

Proof. Let \( E \) be a finite orthonormal set and \( x \in V \). Define \( y = x - \sum_{e \in E} \langle x, e \rangle e \). It is straightforward to check that \( \langle y, e \rangle = 0 \) for all \( e \in E \), so that \( E \cup \{y\} \) is an orthonormal set. In view of

\[
\|x\|^2 = \|y\|^2 + \sum_{e \in E} |\langle x, e \rangle|^2,
\]

which in view of \( \|y\|^2 \geq 0 \) implies (6.6) for all finite \( E \). If \( E \) is infinite, \( \sum_{e \in E} |\langle x, e \rangle|^2 \) equals the supremum of \( \sum_{e \in F} |\langle x, e \rangle|^2 \leq \|x\|^2 \) over the finite subsets \( F \subseteq E \), which thus also satisfies (6.6). \( \blacksquare \)

6.35 Definition An orthonormal base (ONB) in a (pre)Hilbert space is an orthonormal set \( E \) that is maximal. I.e. there is no orthonormal set \( E' \) properly containing \( E \).

6.36 Lemma For every orthonormal set \( E \) in a (pre)Hilbert space \( H \) there is an orthonormal base \( \hat{E} \) containing \( E \).
6.37 Theorem 17 Let $H$ be a Hilbert space and $E$ an orthonormal set in $H$. Then the following are equivalent:

(i) $E$ is an orthonormal base, i.e. maximal.
(ii) If $x \in H$ and $x \perp e$ for all $e \in E$ then $x = 0$.
(iii) $\text{span}_F E = H$.
(iv) For every $x \in H$, there are numbers $\{a_e\}_{e \in E}$ in $F$ such that $x = \sum_{e \in E} a_e e$.
(v) For every $x \in H$, the equality $x = \sum_{e \in E}^\perp \langle x, e \rangle e$ holds.
(vi) For every $x \in H$, we have $\|x\|^2 = \sum_{e \in E} |\langle x, e \rangle|^2$. (Abstract Parseval identity)
(vii) The map $H \to \ell^2(E, F), x \mapsto \{\langle x, e \rangle\}_{e \in E}$ is an isometric map of normed spaces, where $\ell^2(S)$ has the $\| \cdot \|_2$-norm.

Here all summations over $E$ are in the sense of the unordered summation of Appendix A.1 (with $V = H$ in (iv), (v) and $V = F$ in (vi), (vii)).

Proof. If (ii) holds then $E$ is maximal, thus (i). If (ii) is false then there is a non-zero $x \in H$ with $x \perp e$ for all $e \in E$. Then $E \cup \{x/\|x\|\}$ is an orthonormal set larger than $E$, thus $E$ is not maximal. Thus (i)$\iff$(ii). The equivalence (ii)$\iff$(ii) follows from the fact that a linear map is injective if and only if its kernel is $\{0\}$.

(iii)$\implies$(i) If $\text{span}_F E = H$ and $x \in H$ satisfies $x \perp E$ then also $x \perp (\text{span}_F E = H)$, thus $x = 0$. $E$ is maximal and therefore a base.

(ii) $\implies$(iii) $K = \text{span}_F E \subseteq H$ is a closed linear subspace. If $K \neq H$ then by Theorem 6.24 we can find a non-zero $x \in K^\perp$. In particular $x \perp e$ for all $e \in E$, contradicting (ii). Thus $K = H$.

It should be clear that the statements (vi b) and (vii b) are just high-brow versions of (vi a), (vii a), respectively, to which they are equivalent. That (vi a) implies (vi a) is seen by taking $x = y$. Since Exercise 6.15 gives (vi b)$\implies$(vii b), we have the mutual equivalence of (vi a), (vi b), (vii a), (vii b).

(v)$\implies$(iv) is trivial. If (iv) holds then continuity of the inner product, cf. Remark 6.9.1, implies $\langle x, y \rangle = \sum_{e \in E} a_e (e, y)$ for all $y \in H$. For $y \in E$, the r.h.s. reduces to $a_y$, implying (v).

(iv) means that every $x \in H$ is a limit of finite linear combinations of the $e \in E$, thus (iii) holds.

(v)$\implies$(vi a) For finite $F \subseteq E$ we define $x_F = \sum_{e \in F} \langle x, e \rangle e$. Pythagoras’ theorem gives $\|x_F\|^2 = \sum_{e \in F} |\langle x, e \rangle|^2$. As $F \nrightarrow E$, the l.h.s. converges to $\|x\|^2$ by (iii) and the r.h.s. to $\sum_{e \in E} |\langle x, e \rangle|^2$. Thus (vi a) holds.

---

I dislike the approach of [41] of restricting this statement to finite or countably infinite orthonormal sets. This in particular means that the hypotheses of [41, Theorem 1.33] can never be satisfied if $H$ is non-separable! I also find it desiable to understand how much of the theorem survives without completeness since the latter does not hold in situations like Example 6.43. See Remark 6.38.

Marc-Antoine Parseval (1755-1836). French mathematician.
If (vi a) holds then for each $\varepsilon > 0$ there is a finite $F \subseteq E$ such that $\sum_{e \in E \setminus F} |\langle x, e \rangle|^2 < \varepsilon$. Since $x-x_F$ is orthogonal to each $e \in F$, we have $x-x_F \perp x_F$, to that $\|x\|^2 = \|x-x_F\|^2 + \|x_F\|^2$. Combining this with (iv a) and $\|x_F\|^2 = \sum_{e \in F} |\langle x, e \rangle|^2$ we find $\|x-x_F\|^2 = \sum_{e \in E \setminus F} |\langle x, e \rangle|^2 < \varepsilon$. Since $\varepsilon$ was arbitrary, this proves that $\lim_{F,F \subseteq E, x_F \rightarrow x}$, thus (v).

It remains to prove (iii) $\Rightarrow$ (v). The fact $\overline{\text{span}_x E} = H$ means that for every $x \in H$ and $\varepsilon > 0$ there are a finite subset $F \subseteq E$ and coefficients $\{a_e\}_{e \in F}$ such that $\|x - \sum_{e \in F} a_e e\| < \varepsilon$. On the other hand, Theorem 6.24 tells us that for each finite $F \subseteq E$ there is a unique $P_F(x) \in K_F = \text{span}_x F$ minimizing $\|x - P_F(x)\|$. Clearly $P_F \upharpoonright K_F$ is the identity map and the zero map on $K_F^\perp$. Thus defining $P'_F(x) = \sum_{e \in F} \langle e, e \rangle e$, we have $P'_F = P_F$. Thus $\|x - \sum_{e \in F} \langle e, e \rangle e\| \leq \|x - \sum_{e \in F} a_e e\| < \varepsilon$. Since $\varepsilon > 0$ was arbitrary, this proves $\lim_{F,F \subseteq E, x_F \rightarrow x}$, which is nothing other than the statement $x = \sum_{e \in E} \langle x, e \rangle e$. Since the finite sums $\sum_{e \in F} \langle x, e \rangle e$ are in $H$, the identity $x = \sum_{e \in F} \langle x, e \rangle e$ also holds in $H$ for all $x \in H$. 

6.38 REMARK If $H$ is only a pre-Hilbert space, we still have (i) $\Leftrightarrow$ (ii $a/b$) $\Leftrightarrow$ (iii) $\Leftrightarrow$ (iv) $\Leftrightarrow$ (v) $\Leftrightarrow$ (vi $a/b$) $\Leftrightarrow$ (vii $a/b$). This follows from the fact that completeness is not used in proving these equivalences, except in (iii) $\Rightarrow$ (v) where it can be avoided by appealing to Remark 6.25.

Furthermore, we have equivalence of (iii), i.e. $\overline{\text{span}_x E} = H$ (in $H$), with $\overline{\text{span}_x E} = \hat{H}$ (in the completion $\hat{H}$) and with $E$ being an ONB for $\hat{H}$. The equivalence of the second and third statement comes from (i) $\Rightarrow$ (iii), applied to $\hat{H}$. If $\overline{\text{span}_x E}$ is dense in $H$ then it is dense in $\hat{H}$ since $H$ is dense in $\hat{H}$. And the converse follows from the general topology fact that the closure in $H$ of some $S \subseteq H \subseteq \hat{H}$ equals $\overline{S} \cap H$, where $\overline{S}$ is the closure in $\hat{H}$.

In Example 6.43 below, all statements (i)-(vii) hold despite the incompleteness of $H$. But in the absence of completeness the implication (i) $\Rightarrow$ (iii) can fail. For a counterexample see Exercise 6.39. (In view of this, maximal orthonormal sets in pre-Hilbert spaces should not be called bases.) In [25] it is even proven that a pre-Hilbert space in which every maximal orthonormal set $E$ has dense span actually is a Hilbert space. Equivalently, in every incomplete pre-Hilbert space there is a maximal orthonormal set $E$ whose span is non-dense! There even are pre-Hilbert spaces (called pathological) in which no orthonormal set has dense span!

Actually, most of the non-trivial results, like $H \cong K \oplus K^\perp$ for closed subspaces $K$ and Theorem 6.29, hold for a pre-Hilbert space if and only if it is a Hilbert space, see [25].

6.39 EXERCISE (COUNTEREXAMPLE) Let $H = \ell^2(\mathbb{N}, \mathbb{F})$, let $f = \sum_{n=1}^{\infty} \delta_n/n \in H$ (equivalently, $f(n) = 1/n$). Now $K = \text{span}_x \{f, \delta_2, \delta_3, \ldots\}$ (no closure!) is a pre-Hilbert space. Prove:

(i) $E = \{\delta_2, \delta_3, \ldots\}$ is a maximal orthonormal set in $K$.

(ii) $f \notin \overline{\text{span}_x E}$, thus $\overline{\text{span}_x E} \neq K$ (both closures in $K$).

6.40 THEOREM ((F.) RIESZ-FISCHER) 1920 Let $H$ be a pre-Hilbert space and $E$ an orthonormal set such that $\overline{\text{span}_x E} = H$. Then the following are equivalent:

(i) $H$ is a Hilbert space (thus complete).

(ii) The isometric map $H \to \ell^2(E, \mathbb{F})$, $x \mapsto \{\langle x, e \rangle\}_{e \in E}$ is surjective. I.e. for every $f \in \ell^2(E, \mathbb{F})$ there is an $x \in H$ such that $\langle x, e \rangle = f(e)$ for all $e \in E$.

20Also the completeness of $L^2(X, \mathcal{A}, \mu; \mathbb{F})$ (see Lemma 5.10 for $L^2(S)$) is sometimes called Riesz-Fischer theorem.
Proof. (ii)⇒(i) We know from (iii)⇒(vi b) in Theorem 6.37 that the map \( H \to \ell^2(E,F) \) is an isometry. If it is surjective then it is an isomorphism of pre-Hilbert spaces. Since \( \ell^2(E,F) \) is complete by Lemma 5.10, so is \( H \).

(i)⇒(ii) Let \( f \in \ell^2(E,F) \). For each finite subset \( F \subseteq E \) we define \( x_F = \sum_{e \in F} f(e) e \). For each \( \varepsilon > 0 \) there is a finite \( F \subseteq E \) such that \( \sum_{e \in E \setminus F} |f(e)|^2 < \varepsilon \). Whenever \( F \subseteq U \cap U' \), the identity \( x_U - x_U' = \sum_{e \in E} (\chi_U(e) - \chi_{U'}(e)) f(e) e \) implies

$$
\|x_U - x_{U'}\|^2 = \sum_{e \in E} |\chi_U(e) - \chi_{U'}(e)|^2 |f(e)|^2 \leq \varepsilon
$$

since \(|\chi_U - \chi_{U'}|\) vanishes on \( F \) and is bounded by one on \((U \cup U') \setminus F\). Thus \( \{x_F\} \) is a Cauchy net and therefore convergent to a unique \( x \in H \) by completeness, cf. Lemma A.11. By continuity of the inner product, \( \langle x_F, e \rangle \) converges to \( f(e) \), so that \( \langle x, e \rangle = f(e) \) for all \( e \in E \).

\[ \Box \]

6.41 Remark If \( E \) and \( E' \) are ONBs for a Hilbert space \( H \) then one can prove that \( E \) and \( E' \) have the same cardinality, i.e. there is a bijection between \( E \) and \( E' \), cf. [11, Proposition I.4.14]. (This does not follow from the linear algebra proof, since the latter uses a different notion of base, the Hamel bases.) The common cardinality of all bases of \( H \) is called the dimension of \( H \).

\[ \Box \]

6.42 Proposition For a Hilbert space \( H \), the following are equivalent:

(i) \( H \) is separable in the topological sense, i.e. there is a countable dense set \( S \subseteq H \).

(ii) \( H \) admits a countable orthonormal base.

Proof. (ii)⇒(i) Let \( E \) be a countable ONB for \( H \). Then by Theorem 6.40 we have a unitary equivalence \( H \cong \ell^2(E,F) \), and the claim follows from Proposition 5.14.

(i)⇒(ii) Let \( S \) be a countable dense set not containing zero and write it as \( S = \{x_1, x_2, \ldots\} \). Put \( y_1 = x_1/\|x_1\| \). Put \( z_2 = x_2 - \langle x_2, y_1 \rangle y_1 \). If \( z_2 \neq 0 \), put \( y_2 = z_2/\|z_2\| \), otherwise \( y_2 = 0 \). Now \( z_3 = x_3 - \langle x_3, y_1 \rangle y_1 - \langle x_3, y_2 \rangle y_2 \) etc. Now let \( E = \{y_n \mid n \in \mathbb{N}\} \setminus \{0\} \). We claim that \( E \) is an ONB. It is clear by construction that \( E \) is orthonormal. If \( z \) is orthogonal to all \( b \in E \) then \( z \) is orthogonal to all \( s \in S \). If \( x \in H \) is arbitrary, there is a sequence \( \{s_n\} \) in \( S \) such that \( s_n \to x \).

Now continuity of the inner product implies \( \langle x, s \rangle = \lim_{n \to \infty} \langle s_n, s \rangle = 0 \). Thus \( x \in H^\perp = \{0\} \).

\[ \Box \]

6.43 Example Here is an application of Theorem 6.37: Let

\[ H = \{f \in C([0, 2\pi], \mathbb{C}) \mid f(0) = f(2\pi)\} \cong C(S^1, \mathbb{C}). \]

One easily checks that \( \langle f, g \rangle = (2\pi)^{-1} \int_0^{2\pi} f(x) \bar{g}(x) dx \) (Riemann integral) is an inner product, so that \( (H, \langle \cdot, \cdot \rangle) \) is a pre-Hilbert space. (If a continuous function satisfies \( \int |f(x)|^2 dx = 0 \) then it is identically zero.) For \( n \in \mathbb{Z} \), let \( e_n(x) = e^{inx} \). It is straightforward to show that \( E = \{e_n \mid n \in \mathbb{Z}\} \) is an orthonormal set, thus Bessel’s inequality holds. For \( f \in H \) we have

\[ \langle f, e_n \rangle = \frac{1}{2\pi} \int_0^{2\pi} f(x) e^{-inx} dx, \]

which is the \( n \)-th Fourier coefficient \( \hat{f}(n) \) of \( f \), cf. e.g. [74, 34]. In fact, in Fourier analysis one proves, cf. e.g. [74, Corollary 5.4], that the finite linear combinations of the \( e_n \) (‘trigonometric polynomials’) are dense in \( H \), which is (iii) of Theorem 6.37. Thus all other statements in the
21 Unfortunately, this is often omitted from undergraduate linear algebra teaching. E.g., it does not appear in [21] despite the book’s > 500 pages. See however [35, 39] which, admittedly, are aiming higher.

\[
\frac{1}{2\pi} \int_0^{2\pi} |f(x)|^2 dx = \|f\|^2 = \sum_{n \in \mathbb{Z}} |\langle f, e_n \rangle|^2 = \sum_{n \in \mathbb{Z}} |\hat{f}(n)|^2.
\]

This is the original Parseval formula, cf. e.g. [74, Chapter 3, Theorem 1.3]. Note that \( H \) is not complete. Measure theory tells us that this completion is \( L^2([0,2\pi]) \), the measure being Lebesgue measure \( \lambda \) (defined on the \( \sigma \)-algebra of Borel sets). Now the map \( L^2([0,2\pi]) \to \ell^2(\mathbb{Z}, \mathbb{C}), f \mapsto \hat{f} \) is an isomorphism of Hilbert spaces. This nice situation shows that the Lebesgue integral is much more appropriate for the purposes of Fourier analysis than the Riemann integral (as for most other purposes).

6.44 EXERCISE Prove that the pre-Hilbert space \( H = C([0,1]) \) with inner product \( \langle f, g \rangle = \int_0^A f(t)g(t)dt \) is not complete.

6.6 * Tensor products of Hilbert spaces

In this optional section, referenced only in Section 15.4 but important well beyond that, I assume known\(^{21}\) the notion of (algebraic) tensor product \( V \otimes_k W \) of two vector spaces \( V, W \) over a field \( k \). (In two sentences: \( V \otimes_k W \) is the free abelian group spanned the pairs \( (v, w) \in V \times W \), divided by the subgroup generated by all elements of the form \( (v + v', w) - (v, w) - (v', w) \) and \( (v, w + w') - (v, w) - (v, w') \) and \( (cv, w) - (v, cw) \), where \( v, v' \in V, w, w' \in W, c \in k \), the quotient being a \( k \)-vector space in the obvious way. If \( v \in V, w \in W \) then the equivalence class \([ (v, w) ]\) is denoted \( v \otimes w \).) The crucial property is that given a bilinear map \( \alpha : V \times W \to Z \) (where \( V \times W \) is the Cartesian product) there is a unique linear map \( \beta : V \otimes_k W \to Z \) such that \( \beta(v \otimes w) = \alpha((v, w)) \).

6.45 LEMMA Let \( (H, \langle \cdot, \cdot \rangle_H), (H', \langle \cdot, \cdot \rangle_{H'}) \) be pre-Hilbert spaces over \( \mathbb{F} \in \{ \mathbb{R}, \mathbb{C} \} \). Then there is a unique inner product \( \langle \cdot, \cdot \rangle_Z \) on \( Z = H \otimes_{\mathbb{F}} H' \) such that \( \langle v \otimes w, v' \otimes w' \rangle = \langle v, v' \rangle_H \langle w, w' \rangle_{H'} \).

Proof. Every element \( z \in Z = H \otimes_{\mathbb{F}} H' \) has a representation \( z = \sum_{k=1}^K v_k \otimes w_k \) with \( K < \infty \). Given another \( z' = \sum_{l=1}^L v'_l \otimes w'_l \in H \otimes_{\mathbb{F}} H' \), we must define

\[
\langle z, z' \rangle_Z = \sum_{k=1}^K \sum_{l=1}^L \langle v_k, v'_l \rangle_H \langle w_k, w'_l \rangle_{H'}.
\]

Since an element \( z \in Z \) can have many representations of the form \( z = \sum_{k=1}^K v_k \otimes w_k \), we must show that this is well-defined. Let thus \( \sum_{k=1}^K v_k \otimes w_k = \sum_{l=1}^L v'_l \otimes w'_l \). Now, for fixed \( l \) the map \( H \times H \to \mathbb{F}, (x, y) \mapsto \langle x, v'_l \rangle_H \langle y, w'_l \rangle_{H'} \) clearly is bilinear, thus it gives rise to a unique linear map \( H \otimes_{\mathbb{F}} H' \to \mathbb{F} \). This implies

\[
\sum_{k=1}^K \sum_{l=1}^L \langle v_k, v'_l \rangle_H \langle w_k, w'_l \rangle_{H'} = \sum_{k=1}^K \sum_{l=1}^L \langle v'_l, v'_l \rangle_H \langle w'_l, w'_l \rangle_{H'}.
\]

The independence of \( \langle z, z' \rangle_Z \) of the representation of \( z' \) is shown in the same way.

\(^{21}\)Unfortunately, this is often omitted from undergraduate linear algebra teaching. E.g., it does not appear in [21] despite the book’s > 500 pages. See however [35, 39] which, admittedly, are aiming higher.
It is quite clear that $\langle \cdot, \cdot \rangle_Z$ is sesquilinear and satisfies $\langle z', z \rangle_Z = \langle z, z' \rangle_Z$.

In order to study $\langle z, z \rangle_Z$ we may assume that $z = \sum_k v_k \otimes w_k$, where the $w_k$ are mutually orthogonal. This leads to

$$\langle z, z \rangle_Z = \sum_k \langle v_k, v_k \rangle_H \langle w_k, w_k \rangle_{H'} = \sum_k \|v_k\|^2 \|w_k\|^2 \geq 0$$

and $\langle z, z \rangle_Z = 0 \Rightarrow z = 0$.

6.46 Definition If $H, H'$ are Hilbert spaces then $H \otimes H'$ is the Hilbert space obtained by completing the above pre-Hilbert space $(Z, \langle \cdot, \cdot \rangle_Z)$.

6.47 Remark 1. We usually write the completed tensor products $\otimes$ without subscript to distinguish them from the algebraic ones.

2. If $E, E'$ are ONBs in the Hilbert spaces $H, H'$, respectively, then it is immediate that $E \times E'$ is an orthonormal set in the algebraic tensor product $H \otimes_k H'$, thus also in $H \otimes H'$. In fact its span is dense in $E \otimes E'$, so that it is an ONB.

This leads to a pedestrian way of defining the tensor product $H \otimes H'$ of Hilbert spaces over $\mathbb{F}$: Pick ONBs $E \subseteq H, E' \subseteq H'$ and define $H \otimes H' = \ell^2(E \times E', \mathbb{F})$. By Remark 6.41, the outcome is independent of the chosen bases up to isomorphism. If $x \in H, x' \in H'$ then the map $E \times E' \to \mathbb{F}$, $(e, e') \mapsto \langle x, e \rangle_H \langle x', e' \rangle_{H'}$ is in $\ell^2(E \times E, \mathbb{F})$, thus defines an element $x \otimes x' \in H \otimes H'$. This map $H \times H' \to H \otimes H'$ is bilinear. But this definition is very ugly and unconceptual due to its reliance on a choice of bases.

7 Quotient spaces and complemented subspaces

In linear algebra, one has a notion of quotient spaces, cf. e.g. [21, Exercise 31 in Section 1.3]: If $V$ is an $\mathbb{F}$-vector space and $W \subseteq V$ is a linear subspace, one defines an equivalence relation on $V$ by $x \sim y \iff x - y \in W$ and then lets $V/W$ denote the quotient space $V/\sim$, i.e. the set of $\sim$-equivalence classes. One shows that $V/W$ again is a $\mathbb{F}$-vector space.

It is very natural to ask whether $V/K$ again is a Hilbert (or Banach) space if that is the case for $V$. We begin with Hilbert spaces.

7.1 Quotient spaces of Hilbert spaces

7.1 Exercise Let $H$ be a Hilbert space and $K \subseteq H$ a closed linear subspace. Prove that there is a linear isomorphism $H/K \to K^\perp$ of $\mathbb{F}$-vector spaces.

Conclude that the quotient space $H/K$ of a Hilbert space $H$ by a closed subspace $K$ admits an inner product turning it into a Hilbert space.

The above, which completes our first encounter with Hilbert spaces, shows that for a Hilbert space $H$ the notion of quotient $H/K$ by a closed subspace $K$ in a sense is quite superfluous since one has the orthogonal complement $K^\perp \subseteq H$ as a simpler substitute. The latter is no more available for general Banach spaces, so that we’ll have some more work to do.

7.2 Quotient spaces of Banach spaces

In a general Banach space, we don’t have the notion of orthogonal complement. But in most situations, having Banach quotient spaces is good enough. (For a different substitute for orthogonal complements see Section 7.3.)
7.2 Proposition If \( V \) is a normed space, \( W \subseteq V \) a linear subspace and \( V/W \) denotes the quotient vector space, we define \( \| \cdot \|' : V/W \to [0, \infty) \) by \( \|v+W\|' = \inf_{w \in W} \|v-w\|. \) Then

(i) \( \| \cdot \|' \) is a seminorm on \( V/W, \) and the quotient map \( p : V \to V/W \) satisfies \( \|p\| \leq 1. \)

(ii) \( \| \cdot \|' \) is a norm if and only if \( W \subseteq V \) is closed.

(iii) If \( W \subseteq V \) is closed, the topology on \( V/W \) induced by \( \| \cdot \|' \) coincides with the quotient topology, and the quotient map \( p : V \to V/W \) is open.

(iv) If \( V \) is a Banach space and \( W \subseteq V \) is closed then \( (V/W, \| \cdot \|') \) is Banach space.

(v) If \( V \) is a Banach space with closed subspace \( W ) \) and \( T \in B(V,E), \) where \( E \) is a normed space with \( W \subseteq \ker T \) then there is a unique \( T' \in B(V/W,E) \) such that \( T'|p = T. \)

Furthermore, \( \|T'\| = \|T\|. \) \( T' \) is surjective if and only if \( T \) is surjective and injective if and only if \( W = \ker T. \)

(vi) If \( A \) is a normed algebra and \( I \subseteq A \) is a closed two-sided ideal, then \( A/I \) is a normed algebra.

Proof. (i) It is clear that \( \|0\|' = 0 \) (where we denote the zero element of \( V/W \) by 0 rather than \( W \) ) For \( x \in V, c \in \mathbb{F}\{0\} \) we have

\[
\|c(x + W)\|' = \|cx + W\|' = \inf_{w \in W} \|cx - w\| = |c| \inf_{w \in W} \|x - w/c\| = |c| \inf_{w \in W} \|x - w\| = |c|\|x\|',
\]

where we used that \( W \to W, w \mapsto cw \) is a bijection. Now let \( x_1, x_2 \in V \) and \( \epsilon > 0. \) Then there are \( w_1, w_2 \in W \) such that \( \|x_i - w_i\| < \|x_i + W\|' + \epsilon/2 \) for \( i = 1, 2. \) Then

\[
\|x_1 + x_2 + W\|' = \inf_{w \in W} \|x_1 + x_2 + w\| \leq \|(x_1 - w_1) + (x_2 - w_2)\| \leq \|x_1 - w_1\| + \|x_2 - w_2\| < \|x_1 + W\|' + \|x_2 + W\|' + \epsilon.
\]

Since \( \epsilon > 0 \) was arbitrary, we have \( \|x_1 + x_2 + W\|' \leq \|x_1 + W\|' + \|x_2 + W\|', \) proving subadditivity of \( \| \cdot \|'. \) It is immediate that \( \|v + W\|' = \inf_{w \in W} \|v - w\| \leq \|v\|. \)

(ii) If \( v \in V, \) the definition of \( \| \cdot \|' \) readily implies that \( \|v + W\|' = 0 \) if and only if \( v \in W. \) Thus if \( W \) is closed then \( w = v + W \in V/W \) has \( \|w\|' = 0 \) only if \( w \) is the zero element of \( W. \) And if \( W \) is non closed then every \( v \in V/W \) satisfies \( \|v + W\|' = 0 \) even though \( v + W \in V/W \) is zero-non. Thus \( \| \cdot \|' \) is not a norm.

(iii) Continuity of \( p : (V, \| \cdot \|) \to (V/W, \| \cdot \|') \) follows from \( \|p\|' \leq 1, \) see (i). Since \( p \) is norm-decreasing, we have \( p(B(V(0,r))) \subseteq B(V/W(0,r)) \) for each \( r > 0. \) And if \( y \in V/W \) with \( \|y\| < r \) then there is an \( x \in V \) with \( p(x) = y \) and \( \|x\| < r \) (but typically larger than \( \|y\|). \) Thus \( p \) maps \( B(V(0,r)) \) onto \( B(V/W(0,r)) \) for each \( r. \) Similarly, \( p(B(V(x,r))) = B(V/W(p(x,r)), \) and from this it is easily deduced that \( p(U) \subseteq V/W \) is open for each open \( U \subseteq V. \) Thus \( p \) is open (w.r.t. the norm topologies on \( V, V/W \), which implies (cf. [46, Lemma 6.4.5]) that \( p \) is a quotient map, thus the topology on \( V/W \) coming from \( \| \cdot \|' \) is the quotient topology.

(iv) Let \( \{y_n\} \subseteq V/W \) be a Cauchy sequence. Then we can pass to a subsequence \( w_n = y_{n^m} \) such that \( \|w_n - w_{n+1}\| < 2^{-n}. \) Pick \( x_n \in V \) such that \( p(x_n) = w_n \) and \( \|x_n - x_{n+1}\| < 2^{-n}. \) (Why can this be done?) Then \( \{x_n\} \) is a Cauchy sequence converging to some \( x \in V \) by completeness of \( V. \) With \( y = p(x) \) we have \( \|y_n - y\| < \|x_n - x\| \to 0. \) Thus \( y_n \to y, \) and \( V/W \) is complete.

(v) Existence and uniqueness of \( T' \) as linear map are standard. And using \( p(B(V(0,1))) = B(V/W(0,1)) \) we have

\[
\|T'\| = \sup\{\|T'y\| \ | y \in B(V/W(0,1))\} = \sup\{\|T'p(x)\| \ | x \in B(V(0,1))\} = \sup\{\|Tx\| \ | x \in B(V(0,1))\} = \|T\|.
\]
Also the statements concerning injectivity and surjectivity of $T'$ are again pure algebra, but for completeness we give proofs: The statement about surjectivity follows from $T = T' \circ p$ together with surjectivity of $p$, which gives $T(V) = T'(V/W)$. If $W \subset \ker T$, pick $x \in (\ker T) \setminus W$ and put $y = p(x)$. Then $y \neq 0$, but $T'y = T'px = Tx = 0$, so that $T'$ is not injective. Now assume $W = \ker T$. If $y \in \ker T'$ then pick $x \in V$ with $y = p(x)$. Then $Tx = T'px = T'y = 0$, thus $x \in \ker T = W$, so that $y = p(x) = 0$, proving injectivity of $T'$.

(vi) It is known from algebra that $A/I$ is again an algebra. By the above, it is normed. It remains to prove that the quotient norm on $A/I$ is submultiplicative. Let $c, d \in A/I$ and $\varepsilon > 0$. Then there are $a, b \in A$ with $p(a) = c$, $p(b) = d$, $\|a\| < \|c\| + \varepsilon$, $\|b\| < \|d\| + \varepsilon$ (see the exercise below). Then $\|cd\| = \|p(ab)\| \leq \|ab\| \leq \|a\|\|b\| < (\|c\| + \varepsilon)(\|d\| + \varepsilon)$, and since this holds for all $\varepsilon > 0$, we have $\|cd\| \leq \|c\|\|d\|$. ■

7.3 Exercise  (i) If $V$ is a normed space and $W \subseteq V$ is a closed subspace, prove that for every $y \in V/W$ and every $\varepsilon > 0$ there is an $x \in V$ with $p(x) = y$ and $\|x\| \leq \|y\| + \varepsilon$.

(ii) Give an example of a normed space $V$, a closed subspace $W$ and $y \in V/W$ for which no $x \in V$ with $y = p(x)$, $\|x\| = \|y\|$ exists.

7.4 Exercise  Use the quotient space construction of Banach spaces to give a new proof for the difficult part of Exercise 4.16.

The following is closely related to the Hilbert space $\perp$, but not the same:

7.5 Definition  Let $V$ be a Banach space and $W \subseteq V$ a subspace. Then the annihilator of $W$ is $W^\perp = \{ \varphi \in V^* \mid \varphi \mid W = 0 \} \subseteq V^*$. One easily checks $W^\perp = W^\perp = W^\perp$.

7.6 Exercise  Let $V$ be a Banach space and $W \subseteq V$ a closed subspace. Let $p : V \to V/W$ be the quotient map. Prove that the map $\alpha : (V/W)^* \to V^*$, $\psi \mapsto \psi \circ p$ is injective and isometric and its image is $W^\perp \subseteq V^*$. Thus $W^\perp \cong (V/W)^*$ as Banach spaces.

7.7 Exercise  Let $V$ be a Banach space and $Z \subseteq V^*$ a closed subspace. Define $Z^\perp \subseteq V$ and prove $V^*/Z \cong (V/Z^\perp)^*$.

7.8 Exercise  Let $V$ be a Banach space, $W \subseteq V$ a closed subspace and $Z \subseteq V$ a finite dimensional subspace. Prove that $W + Z \subseteq V$ is closed. Hint: Use $V/W$.

7.9 Exercise  Give a counterexample showing that $W + Z = \{ w + z \mid w \in W, z \in Z \} \subseteq V$ need not be closed for all closed subspaces $W, Z$ of a Banach space. Hint: $V = l^2(\mathbb{N}, \mathbb{R})$.

7.3 Complemented subspaces

The following notion provides a partial substitute for orthogonal complements which we don’t have in Banach spaces:

7.10 Definition  Let $V$ be a Banach space. A closed subspace $W \subseteq V$ is called complemented if there is a closed subspace $Z \subseteq V$ such that $V = W + Z$ and $W \cap Z = \{0\}$.

If $V, W, Z$ are as in the definition (without closedness) then every $v \in V$ can be written as $v = w + z$ with $w \in W, z \in Z$ in a unique way. (Uniqueness follows from $w + z = w' + z' \Rightarrow w - w' = z' - z \in W \cap Z = \{0\}$.) One says $V$ is the internal direct sum of $W$ and $Z'$. Purely algebraically, every subspace $W$ has a complementary subspace $Z$: Pick a (Hamel) base $E$ for
W, extend to a base $E'$ of $V$ and put $Z = \text{span}_F E' \setminus E$. But here we want $Z$ to be closed! In Exercise 10.8 we will prove that with closedness of $W, Z$ we have $V \cong W \oplus Z$ also topologically.

7.11 Exercise Let $V = C([0,2], \mathbb{R})$ with the $\| \cdot \|_\infty$-norm. Let $W = \{ f \in V \mid f_{(1,2]} = 0 \}$.

(i) Prove that $W$ is complemented.

(ii) Can you ‘classify’ all possible complements, i.e. put them in bijection with a simpler set?

7.12 Exercise Let $V$ be a Banach space and $P \in B(V)$ satisfying $P^2 = P$. Prove that $W = PV$ is a complemented subspace. (The converse is also true, as you will prove later.)

7.13 Exercise Let $V$ be a Banach space and $W \subseteq V$ a closed subspace such that $\dim V/W < \infty$. Prove that $W$ is complemented.

7.14 Proposition Every finite dimensional subspace of a Banach space is complemented.

Proof. The proof will be given in Section 8.3 since it requires tools to which we now turn. ■

Not every closed subspace of a Banach space is complemented! In view of Exercise 7.13 and Proposition 7.14, a non-complemented subspace $W \subseteq V$ must have infinite dimension and codimension. And indeed, $c_0(\mathbb{N}, F) \subseteq \ell_\infty(\mathbb{N}, F)$ is non-complemented, as we prove in Appendix B.3. See also [42] for more on the subject of complemented subspaces.

In fact, a Banach space has complementary subspaces for all closed subspaces if and only if it is isomorphic to a Hilbert space, i.e. it admits an inner product whose associated norm is equivalent to the original one! See [40].

In the process of returning from Hilbert to Banach spaces, the above discussion of quotient spaces and complements was the easiest part. The question of bases is much harder for Banach spaces, as the existence of the two volume treatment [72] of the subject, having 680+888 pages, might suggest. (Then again, the basics are quite accessible, cf. e.g. [42, 26, 8, 1], but unfortunately we don’t have the time.) The same is true for the formidable subject of tensor products of Banach spaces, see e.g. [65]. Going into that would be pointless given that we already slighted the much simpler tensor products of Hilbert spaces.

A more tractable problem is the fact that in the absence of an inner product, the existence of non-zero bounded linear functionals is rather non-trivial and can in general only be proven non-constructively, as we will do in the next section. (Of course, for spaces that are given very explicitly like $\ell^p(S, \mathbb{F})$, we may well have more concrete approaches as in Section 5.5.)

8 Hahn-Banach theorem and its applications

We have seen that every bounded linear functional $\varphi \in H^*$, where $H$ is a Hilbert space, is of the form $\varphi = \varphi_y$ for a certain (unique) $y \in H$. Thus dual spaces of Hilbert spaces are completely understood. (The map $H \to \overline{H}, y \mapsto \varphi_y$ is an anti-linear bijection.) For a general Banach space $V$, matters are much more complicated. The point of the Hahn$^{22}$-Banach theorem (which comes in many versions)$^{23}$ is to show that there many linear functionals.

---

$^{22}$Hans Hahn (1879-1934). Austrian mathematician who mostly worked in analysis and topology.

$^{23}$Important early results are due to Eduard Helly (1884-1943), another Austrian mathematician. See [41, p. 54-55].
8.1 First version of Hahn-Banach over \( \mathbb{R} \)

8.1 Definition. If \( V \) is a real vector space, a map \( p : V \to \mathbb{R} \) is called sublinear if it satisfies

- Positive homogeneity: \( p(cv) = cp(v) \) for all \( v \in V \) and \( c > 0 \).
- Subadditivity: \( p(x + y) \leq p(x) + p(y) \) for all \( x, y \in V \).

8.2 Theorem. Let \( V \) be a real vector space and \( p : V \to \mathbb{R} \) a sublinear function. Let \( W \subseteq V \) be a linear subspace and \( \wp : W \to \mathbb{R} \) a linear functional such that \( \wp(w) \leq p(w) \) for all \( w \in W \). Then there is a linear functional \( \wp' : V \to \mathbb{R} \) such that \( \wp' \mid W = \wp \) and \( \wp(v) \leq p(v) \) for all \( v \in V \).

The heart of the proof of the theorem is proving it in the case where we extend \( \wp \) from \( W \) to \( W + \mathbb{R}v' \).

8.3 Lemma. Let \( V, p, W, \wp \) be as in Theorem 8.2 and \( v' \in V \). Then there is a linear functional \( \wp' : W + \mathbb{R}v' \to \mathbb{R} \) such that \( \wp' \mid W = \wp \) and \( \wp(v) \leq p(v) \) for all \( v \in V \).

Proof. If \( v \in W \), there is nothing to do so that we may assume \( v' \in V \setminus W \). Then every \( x \in W + \mathbb{R}v' \) can be written as \( x = w + cv' \) with unique \( w \in W, c \in \mathbb{R} \). Thus if \( d \in \mathbb{R} \), we can define \( \wp : W + \mathbb{R}v' \to \mathbb{R} \) by \( \wp(w + cv') = \wp(w) + cd \) for all \( w \in W \) and \( c \in \mathbb{R} \). Since \( \wp \) is linear and trivially satisfies \( \wp \mid W = \wp \), it remains to show that \( d \) can be chosen such that

\[
\wp(w + cv') = \wp(w) + cd \leq p(w + cv') \quad \forall w \in W, c \in \mathbb{R}.
\] (8.1)

For \( c = 0 \), this holds by assumption. If (8.1) holds for all \( w \in W \) and \( c \in \{1, -1\} \), i.e.

\[
\wp(w) \pm d \leq p(w \pm v'),
\] (8.2)

then for all \( e > 0 \) we have

\[
\wp(w \pm ev') = e\wp(e^{-1}w \pm v') \leq ep(e^{-1}w \pm v') = p(w \pm ev'),
\]

thus the desired inequality (8.1) holds for all \( w \in W, c \in \mathbb{R} \). Now \( d \in \mathbb{R} \) satisfies (8.2) for all \( w \in W \) and both signs if and only if

\[
\wp(w) - p(w - v') \leq d \leq p(w' + v') - \wp(w') \quad \forall w, w' \in W.
\]

Clearly this is possible if and only if \( \wp(w) - p(w - v') \leq p(w' + v') - \wp(w') \) for all \( w, w' \in W \), which is equivalent to \( \wp(w) + \wp(w') \leq p(w - v') + p(w' + v') \forall w, w' \). This is indeed satisfied for all \( w, w' \in W \) since \( w + w' \in W \) so that

\[
\wp(w) + \wp(w') = \wp(w + w') \leq p(w + w') \leq p(w - v') + p(w' + v')
\]

holds since \( \wp \) is linear and bounded by \( p \) and since \( p \) is subadditive.

Proof of Theorem 8.2. If \( W = V \), there is nothing to do, so assume \( W \subsetneq V \). Let \( \mathcal{E} \) be the set of pairs \( (Z, \psi) \), where \( Z \subseteq V \) is a linear subspace containing \( W \) and \( \psi : Z \to \mathbb{R} \) is a linear map extending \( \wp \) such that \( \psi(z) \leq p(z) \) \( \forall z \in Z \). Since \( W \neq V \), Lemma 8.3 implies \( \mathcal{E} \neq \emptyset \).

We define a partial ordering on \( \mathcal{E} \) by \( (Z, \psi) \leq (Z', \psi') \iff Z \subseteq Z', \psi' \mid Z = \psi \). If \( \mathcal{C} \subseteq \mathcal{E} \) is a chain, i.e. totally ordered by \( \leq \), let \( Y = \bigcup_{(Z, \psi) \in \mathcal{C}} Z \) and define \( \psi_Y : Y \to \mathbb{R} \) by \( \psi_Y(v) = \psi(v) \) for any \( (Z, \psi) \in \mathcal{C} \) with \( v \in Z \). This clearly is consistent and gives a linear map. Now \( (Y, \psi_Y) \) is in element of \( \mathcal{E} \) and an upper bound for \( \mathcal{C} \). Thus by Zorn’s lemma there is a maximal element \( (Y_M, \psi_M) \) of \( \mathcal{E} \). Now \( \psi_M : Y_M \to \mathbb{R} \) is an extension of \( \wp \) satisfying \( \psi_M(y) \leq p(y) \) for all \( y \in Y_M \), so we are done if we prove \( Y_M = V \). If this is not the case, we can pick \( v' \in V \setminus Y_M \) and use Lemma 8.3 to extend \( \psi_Y \) to \( Y_M + \mathbb{R}v' \), but this contradicts the maximality of \( (Y_M, \psi_M) \).
8.4 Remark The above proof used Zorn’s lemma, which is equivalent to the Axiom of Choice (AC), and therefore very non-constructive\textsuperscript{24}. There is nothing much to be done about this, but we mention that the Hahn-Banach theorem can be deduced from the ‘ultrafilter lemma’, which is strictly weaker than AC. For separable spaces, the Hahn-Banach theorem can be proven using only the axiom DC\textsubscript{\omega} of countable dependent choice. For proofs of these claims see [46, Appendix G].

8.2 Hahn-Banach theorem for (semi)normed spaces

With the exception of Section B.5.2 we will not use Theorem 8.2 directly, but only the following consequence:

8.5 Theorem (Hahn-Banach Theorem) If \( V \) be a vector space over \( \mathbb{F} \in \{\mathbb{R}, \mathbb{C}\} \), \( p \) a seminorm on it, \( W \subseteq V \) a linear subspace and \( \varphi : W \rightarrow \mathbb{C} \) a linear functional such that \( |\varphi(w)| \leq p(w) \) for all \( w \in W \). Then there is a linear functional \( \hat{\varphi} : V \rightarrow \mathbb{C} \) such that \( \hat{\varphi} \upharpoonright W = \varphi \) and \( |\hat{\varphi}(v)| \leq p(v) \) for all \( v \in V \).

Proof. \( \mathbb{F} = \mathbb{R} \): This is an immediate consequence of Theorem 8.2 since a seminorm \( p \) is sublinear with the additional properties \( p(-v) = p(v) \geq 0 \) for all \( v \). In particular, \( -\hat{\varphi}(v) \leq \hat{\varphi}(-v) \) \( \leq p(-v) = p(v) \), so that \( -p(v) \leq \varphi(v) \leq p(v) \) for all \( v \in V \), which is equivalent to \( |\hat{\varphi}(v)| \leq p(v) \) \( \forall v \).

\( \mathbb{F} = \mathbb{C} \): Assume \( V \supseteq W \) \( \hat{\varphi} \mathbb{C} \) satisfies \( |\varphi(w)| \leq p(w) \) \( \forall w \in W \). Define \( \psi : W \rightarrow \mathbb{R}, w \mapsto \Re(\varphi(w)) \), which clearly is \( \mathbb{R} \)-linear and satisfies the same bounds. Thus by the real case just considered, there is an \( \mathbb{R} \)-linear functional \( \hat{\psi} : V \rightarrow \mathbb{R} \) such that \( |\hat{\psi}(v)| \leq p(v) \) for all \( v \in V \). Define \( \hat{\varphi} : V \rightarrow \mathbb{C} \) by

\[
\hat{\varphi}(v) = \hat{\psi}(v) - i\hat{\psi}(iv).
\]

Again it is clear that \( \hat{\varphi} \) is \( \mathbb{R} \)-linear. Furthermore

\[
\hat{\varphi}(iv) = \hat{\psi}(iv) - i\hat{\psi}(-v) = \hat{\psi}(iv) + i\hat{\psi}(v) = i(\hat{\psi}(v) - i\hat{\psi}(iv)) = i\hat{\varphi}(v),
\]

proving that \( \hat{\varphi} : V \rightarrow \mathbb{C} \) is \( \mathbb{C} \)-linear. If \( w \in W \) then

\[
\hat{\varphi}(w) = \hat{\psi}(w) - i\hat{\psi}(iw) = \psi(w) - i\hat{\psi}(iw) = \Re(\varphi(w)) - i\Re(\varphi(iv)) = \Re(\varphi(w)) - i\Re(\varphi(iv)) = \Re(\varphi(w)) + i\Im(\varphi(w)) = \varphi(w),
\]

so that \( \hat{\varphi} \) extends \( \varphi \).

Given \( v \in V \), let \( \alpha \in \mathbb{C}, |\alpha| = 1 \) be such that \( \alpha\hat{\varphi}(v) \geq 0 \). Then \( \alpha\hat{\varphi}(v) = \hat{\varphi}(\alpha v) = \Re(\hat{\varphi}(\alpha v)) = \hat{\psi}(\alpha v) \), so that \( |\hat{\varphi}(v)| = |\alpha\hat{\varphi}(v)| = \hat{\psi}(\alpha v) \leq p(\alpha v) = p(v). \)

\[\square\]

\[\text{\textsuperscript{24}}\text{"Such reliance on awful non-constructive results is unfortunately typical of traditional functional analysis." [37]\]
8.3 First applications

We are now in a position to give the

Proof of Proposition 7.14. To begin with, finite dimensional subspaces are automatically closed by Exercise 4.22. Let \( W \subseteq V \) be finite dimensional and let \( \{e_1, \ldots, e_n\} \) be a base for \( W \). Since every \( w \in W \) can be written as \( \sum_{i=1}^{n} c_i e_i \) in a unique way, there are linear functionals \( \varphi_i : W \to \mathbb{C} \) such that \( w = \sum_{i=1}^{n} \varphi_i(w)e_i \) for each \( w \in W \). Since \( W \) is finite dimensional, the \( \varphi_i \) are automatically bounded. Now by the Hahn-Banach Theorem 8.5 there are continuous linear functionals \( \hat{\varphi}_i : V \to \mathbb{C} \) extending the \( \varphi_i \). Then \( Z = \bigcap_{i=1}^{n} \ker \hat{\varphi}_i \) is a closed linear subspace of \( V \). It should be clear that \( W \cap Z = \{0\} \). Define \( P : V \to W, v \mapsto \sum_{i=1}^{n} \hat{\varphi}_i(v)e_i \). We have \( P \upharpoonright W = \text{id}_W \), thus \( P^2 = P \). Now apply Exercise 7.12.

8.7 Proposition Let \( E \) be a normed space over \( F \in \{\mathbb{R}, \mathbb{C}\} \).

(i) For every \( 0 \neq x \in E \) there is a \( \varphi \in E^* \) with \( \|\varphi\| = 1 \) such that \( \varphi(x) = \|x\| \). Thus \( E^* \) separates the points of \( E \).

(ii) If \( x \in E \) then \( \hat{x} : E^* \to F, \varphi \mapsto \varphi(x) \) is in \( E^{**} \) with \( \|\hat{x}\| = \|x\| \). The map \( \iota_E : E \to E^{**}, x \mapsto \hat{x} \) is an isometric embedding.

(iii) The image\(^{25} \) \( \iota_E(E) \subseteq E^{**} \) is closed if and only if \( E \) is complete (i.e. Banach).

Proof. (i) Let \( W = Fx \subseteq E \). The linear functional \( \varphi : W \to F, cx \mapsto c\|x\| \) is isometric since \( |\varphi(x)| = \|x\| \), thus \( \|\varphi\| = 1 \). By the Hahn-Banach Theorem 8.5 there exists a \( \hat{\varphi} \in E^* \) with \( \hat{\varphi}(x) = \varphi(x) = \|x\| \) and \( \|\hat{\varphi}\| = \|\varphi\| = 1 \).

(ii) It is clear that \( \hat{x} : E^* \to F, \varphi \mapsto \varphi(x) \) is a linear functional. If \( x \in E \), \( \varphi \in E^* \) then \( |\hat{x}(\varphi)| = |\varphi(x)| \leq \|x\|\|\varphi\| \). Thus \( \|\hat{x}\| \leq \|x\| \). By (i) there is \( \varphi \in E^* \) with \( \|\varphi\| = 1 \) such that \( \varphi(x) = \|x\| \). This gives \( \|\hat{x}\| = \|x\| \). Thus the map \( \iota_E : E \to E^{**}, x \mapsto \hat{x} \), which clearly is linear, is an isometric embedding.

(iii) If \( E \) is complete then \( \iota_E(E) \subseteq E^{**} \) is closed by Corollary 4.6 since \( \iota_E \) is an isometry by (ii). Conversely, if \( \iota_E(E) \subseteq E^{**} \) is closed then completeness of \( E^{**} \) (Proposition 4.24(ii)) implies that \( \iota_E(E) \) is complete, thus also \( E \) since \( \iota_E \) is an isometry.

It is customary to write simply \( \iota \) instead of \( \iota_E \) or to drop it from the notation entirely, identifying \( E \) with its image \( \iota_E(E) \) in \( E^{**} \), so that \( E \subseteq E^{**} \).

8.8 Corollary Every normed space \( E \) embeds isometrically into a Banach space \( \hat{E} \) as a dense subspace. That space \( \hat{E} \) is unique up to isometric isomorphism and is called the completion of \( E \).

Proof. This can be proven by completing the metric space \( (E, d) \), where \( d(x, y) = \|x - y\| \) and showing that the completion is a linear space, which is easy. Alternatively, using the above result that \( \iota_E : E \to E^{**} \) is an isometry, we can take \( \hat{E} = \iota_E(E) \subseteq E^{**} \) as the definition of \( \hat{E} \) since this is a closed subspace of the complete space \( E^{**} \) and therefore complete.

Uniqueness of the completion follows with the same proof as for metric spaces, cf. [46].

\(^{25}\)If \( f : X \to Y \) is any function, from a category theory point of view one would call \( X \) the source (or domain) and \( Y \) the target (or codomain) of \( f \) and call the subset \( f(X) \subseteq Y \) the image of \( f \). I prefer to avoid the term ‘range’ since some authors use it for ‘target’ (thus \( Y \)) and others for ‘image’ (thus \( f(X) \)). The term ‘image’ is unambiguous since no reasonable person would use it intending \( Y \).
8.9 Exercise Let $V$ be a Banach space and $W \subseteq V$ a subspace. Prove: $W = V \iff W^\perp = \{0\}$. (This is a Banach space analogue of Exercise 6.26(i), but now $W^\perp \subseteq V^*$, not $W^\perp \subseteq V$!)

8.10 Exercise Let $V$ be a Banach space and $x \in V$, $\varphi \in V^*$. Prove that $i_{V^*}(\varphi)(i_V(x)) = \varphi(x)$.

8.4 Reflexivity of Banach spaces

8.11 Definition A Banach space $E$ is called reflexive if the map $i_E : E \to E^{**}$ is surjective (thus an isometric bijection).

8.12 Remark While reflexivity of $E$ implies $E \cong E^{**}$, there are Banach spaces $E$ that are not reflexive, yet satisfy $E \cong E^{**}$ non-canonically! An example is the James space, see e.g. [42, Section 4.5], which is also interesting since $E^{**}/i_E(E)$ is one-dimensional! For most non-reflexive spaces this quotient is infinite dimensional.

8.13 Exercise Prove:

(i) Every finite dimensional Banach space is reflexive.

(ii) Every Hilbert space is reflexive.

(iii) If $1 < p < \infty$ then $\ell^p(S, \mathbb{F})$ is reflexive.

(iv) If $S$ is infinite then $c_0(S, \mathbb{F})$ and $\ell^1(S, \mathbb{F})$ are not reflexive.

8.14 Exercise (i) Prove that if $E$ is reflexive then for each $\varphi \in E^*$ there is $x \in E$ such that $\|x\| = 1$ and $|\varphi(x)| = \|\varphi\|$.

(ii) Use (i) and Theorem 5.16 to prove (again) that $c_0(\mathbb{N}, \mathbb{C})$ is not reflexive.

8.15 Remark 1. The converse of the statement in Exercise 8.14(i) is also true, but the proof is much harder and more than 10 pages long! (See [42, Section 1.13].)

2. See Appendix B.6 for the notion of uniform convexity, which is stronger than the strict convexity encountered earlier, and a proof of the fact that uniformly convex spaces are reflexive.

We will also prove prove that $L^p(X, \mathcal{A}, \mu)$ is uniformly convex for each measure space $(X, \mathcal{A}, \mu)$ and $1 < p < \infty$. This provides a proof of reflexivity of these spaces that does not use the relation between $L^p$ and $L^q$. This in turn leads to a simple proof of surjectivity of the isometric map $L^q \to (L^p)^*$ known from Section 5.6 (reversing the logic of Exercise 8.13(iii)).

3. If $E$ is a Banach space and $F \subseteq E$ is a closed subspace then $E$ is reflexive if and only both $F$ and $E/F$ are reflexive. The proof uses only Hahn-Banach. See [83] for a nice exposition.

8.16 Theorem Let $V$ be a Banach space. Then $V$ is reflexive if and only if $V^*$ is reflexive.

Proof. $\Rightarrow$ Given surjectivity of the canonical map $i_V : V \to V^{**}$, we want to prove surjectivity of $i_{V^*} : V^* \to V^{***}$. Let thus $\varphi \in V^{***} = (V^{**})^*$. Putting $\varphi' = \varphi \circ i_V \in V^*$, the implication is proven if we show $\varphi = i_{V^*}(\varphi')$, which means $\varphi(x^{**}) = i_{V^*}(\varphi')(x^{**})$ for all $x^{**} \in V^{**}$. By surjectivity of $i_V : V \to V^{**}$, this is equivalent to $\varphi(i_V(x)) = i_{V^*}(\varphi')(i_V(x))$ for all $x \in V$. This is true since the l.h.s. is $\varphi'(x)$ by definition of $\varphi'$ and the r.h.s. equals $\varphi'(x)$ by Exercise 8.10.

$\Leftarrow$ Assume that $V$ is not reflexive. Then $i_V(V) \subseteq V^{**}$ is a proper closed subspace, so that $i_V(V)^\perp \neq \{0\}$ by Exercise 8.9. Let thus $0 \neq \varphi \in i_V(V)^\perp \subseteq V^{***}$. Since $V^*$ is reflexive, we have $\varphi = i_{V^*}(\varphi')$ for some $\varphi' \in V^*$. Using Exercise 8.10 again, for each $x \in V$ we have $\varphi'(x) = i_{V^*}(\varphi')(i_V(x)) = \varphi(i_V(x)) = 0$. But this means $\varphi' = 0$, thus $\varphi = 0$, a contradiction. ■
8.17 Remark 1. Since $\ell^\infty(S, F) \cong \ell^1(S, F)^*$, the theorem implies that also $\ell^\infty(S, F)$ is not reflexive for infinite $S$.

2. More generally, for non-reflexive $E$ none of the spaces $E^*, E^{**}, \ldots$ is reflexive, so that $E \subsetneq E^{**} \subsetneq E^{***} \subsetneq \cdots$ and $E^* \subsetneq E^{**} \subsetneq E^{***} \subsetneq \cdots$, and we have two somewhat mysterious successions of ever larger spaces! There do not seem to be many general results about this, but see Lemma B.10(iv). Even understanding $C(X, \mathbb{R})^{**}$ for compact $X$ is complicated, cf. [33]. □

8.18 Exercise Let $V$ be a Banach space. Prove:

(i) If $V^*$ is separable then $V$ is separable.

(ii) If $V$ is separable then $V^*$ can be separable or non-separable. (Examples!)

(iii) If $V$ is separable and reflexive then $V^*$ is separable.

While the material of the present section almost trivializes for Hilbert spaces, the results in the next two sections remain equally non-trivial when restricted to Hilbert spaces.

9 Uniform boundedness theorem: Two versions

9.1 The weak version, using only countable choice

9.1 Definition Let $E$, $F$ be normed spaces and $F \subseteq B(E, F)$ a family of bounded linear maps.

(i) $F$ is called pointwise bounded if $\sup_{A \in F} \|Ax\| < \infty$ for each $x \in E$.

(ii) $F$ is called uniformly bounded if $\sup_{A \in F} \|A\| < \infty$.

It is trivial that uniform boundedness of $F$ implies pointwise boundedness. Remarkably:

9.2 Theorem [Helly 1912, Hahn, Banach 1922] Let $E$ be a Banach space, $F$ a normed space and $F \subseteq B(E, F)$ pointwise bounded. Then $F$ is uniformly bounded.

Proof. Assume that $F$ is not uniformly bounded. Then the sets $F_n = \{ A \in F \mid \|A\| \geq 4^n \}$ are all non-empty, so that using AC (axiom of countable choice), we can pick an $A_n \in F_n$ for each $n \in \mathbb{N}$. By definition of $\|A_n\|$, the sets $X_n = \{ x \in E \mid \|x\| \leq 1, \|A_n x\| \geq \frac{1}{2} \|A_n\| \}$ are all non-empty, to that using AC again, we can choose an $x_n \in X_n$ for each $n \in \mathbb{N}$.

Applying the triangle inequality to $Az = \frac{1}{2}(A(y + z) - A(y - z))$ gives

$$\|Az\| = \frac{1}{2}\|A(y + z) - A(y - z)\| \leq \frac{1}{2}\|A(y + z)\| + \|A(y - z)\| \leq \max(\|A(y + z)\|, \|A(y - z)\|).$$

Applying this inequality to $A = A_{n+1}$, $y = y_n$, $z = \pm 3^{-(n+1)}x_{n+1}$, recalling $\|A_n x_n\| \geq \frac{2}{3} \|A_n\|$, we see that for at least one of the signs $\pm$ we have

$$\|A_{n+1}(y_n \pm 3^{-(n+1)}x_{n+1})\| \geq 3^{-(n+1)}\|A_{n+1}x_{n+1}\| \geq 3^{-(n+1)}\frac{2}{3} \|A_{n+1}\|.$$ 

Thus defining a sequence $\{y_n\} \subseteq E$ by $y_1 = x_1$ and

$$y_{n+1} = \begin{cases} 
   y_n + 3^{-(n+1)}x_{n+1} & \text{if } \|A_{n+1}(y_n + 3^{-(n+1)}x_{n+1})\| \geq 3^{-(n+1)}\frac{2}{3} \|A_{n+1}\| \\
   y_n - 3^{-(n+1)}x_{n+1} & \text{otherwise}
\end{cases} \quad (9.1)$$

we have $\|A_n y_n\| \geq \frac{2}{3} 3^{-n} \|A_n\|$ for all $n$. (For $n = 1$ this is true since $y_1 = x_1$.) Since (9.1) involves no further free choices, this inductive definition can be formalized in ZF (which we don’t do here, see [20]).
With (9.1) and \( \|x_n\| \leq 1 \) for all \( n \), we have \( \|y_{n+1} - y_n\| \leq 3^{-(n+1)} \forall n \). Now for all \( m > n \):
\[
\|y_m - y_n\| = \left\| \sum_{k=n}^{m-1} y_{k+1} - y_k \right\| \leq \sum_{k=n}^{\infty} 3^{-(k+1)} = 3^{-(n+1)} \frac{1}{1 - \frac{1}{3}} = \frac{1}{2} 3^{-n},
\]
so that \( \{y_n\} \) is a Cauchy sequence. By completeness of \( E \) we have \( y_n \to y \in E \) with \( \|y - y_n\| \leq \frac{1}{2} 3^{-n} \). Another use of the triangle inequality gives
\[
\|A_n y_n\| = \|A_n (y - y + y_n)\| \leq \|A_n y\| + \|A_n (y - y_n)\| \leq \|A_n y\| + \|A_n\| \|y - y_n\|,
\]
so that with \( \|y - y_n\| \leq \frac{1}{2} 3^{-n}, \|A_n y_n\| \geq \frac{2}{3} 3^{-n} \|A_n\| \) and \( \|A_n\| \geq 4^n \) for all \( n \) we finally have
\[
\|A_n y\| \geq \|A_n y_n\| - \|A_n\| \|y - y_n\| \geq \|A_n\| \left( \frac{2}{3} 3^{-n} - \frac{1}{2} 3^{-n} \right) = \frac{1}{6} 3^{-n} \|A_n\| \geq \frac{1}{6} \left( \frac{4}{3} \right)^n \to \infty.
\]
Thus \( y \in E \) is a witness for the failure of pointwise boundedness of \( F \).  

9.3 Remark The above method of proof is called the gliding (or sliding) hump method and is more than 100 years old. (See also Section B.7 for another use of this method.) Nowadays, the above theorem is usually deduced from Baire’s theorem, cf. Appendix A.5. As mentioned there, the latter is equivalent to the axiom DC\(_\omega\) of countable dependent choice, whereas above we only used the weaker axiom AC\(_\omega\) of countable choice. The above argument was discovered only a few years ago and published [20] in 2017!

\[\square\]

9.2 Applications: Banach-Steinhaus, Hellinger-Toeplitz

9.4 Definition Let \( E, F \) be normed spaces. A sequence (or net) \( \{A_n\} \subseteq B(E, F) \) is strongly convergent if \( \lim_{n \to \infty} A_n x \) exists for every \( x \in E \).

Under the above assumption, the map \( A : E \to F, x \mapsto \lim_{n \to \infty} A_n x \) is easily seen to be linear. Now we write \( A_n \xrightarrow{\ast} A \) or \( A = s\text{-}\lim A_n \).

9.5 Corollary (Banach-Steinhaus) \(^{26,27}\) If \( E \) is a Banach space, \( F \) a normed space and the sequence \( \{A_n\} \subseteq B(E, F) \) is strongly convergent then the map \( A = s\text{-}\lim A_n \) is bounded, thus in \( B(E, F) \).

Proof. The convergence of \( \{A_n x\} \subseteq F \) for each \( x \in E \) implies boundedness of \( \{A_n x \mid n \in \mathbb{N}\} \) for each \( x \), so that \( F = \{A_n \mid n \in \mathbb{N}\} \subseteq B(E, F) \) is pointwise bounded and therefore uniformly bounded by Theorem 9.2. Thus there is \( T \) such that \( \|A_n\| \leq T \forall n \), so that \( \|A_n x\| \leq T \|x\| \forall x \in E, n \in \mathbb{N} \). With \( A_n x \to Ax \) this implies \( \|Ax\| \leq T \|x\| \) for all \( x \), thus \( \|A\| \leq T < \infty \).  

\[\square\]

9.6 Remark Clearly \( A_n \xrightarrow{\ast} A \) is equivalent to \( \|A_n - A\|_x \to 0 \) for all \( x \in E \), where \( \|A\|_x := \|Ax\| \) is a seminorm on \( B(E, F) \) for each \( x \in E \). If \( \|A\|_x = 0 \) for all \( x \in E \) then \( Ax = 0 \forall x \in E \), thus \( A = 0 \). Thus the family \( F = \{\|\cdot\|_x \mid x \in E\} \) is separating and induces a locally convex topology on \( B(E, F) \), the strong operator topology \( \tau_{\text{so}} \). Norm convergence \( \|A_n - A\| \to 0 \) clearly implies strong convergence \( A_n \xrightarrow{\ast} A \), but usually the strong (operator) topology is strictly weaker (despite its name) than the norm topology. See the following exercise for an example.

\(^{26}\)Hugo Steinhaus (1887-1972). Polish mathematician

\(^{27}\)In the literature, one can find either this result or Theorem 9.2 denoted as ‘Banach-Steinhaus theorem’.
9.7 Exercise Let \( 1 \leq p < \infty \) and \( V = \ell^p(\mathbb{N}, \mathbb{F}) \). For each \( m \in \mathbb{N} \) define \( P_m \in B(V) \) by \( (P_m f)(n) = f(n) \) for \( n \geq m \) and \( (P_m f)(n) = 0 \) if \( n < m \). Prove \( P_m \xrightarrow{\text{sot}} 0 \), but \( \| P_m \| = 1 \forall m \), thus \( P_m \xrightarrow{\text{F}} 0 \).

9.8 Exercise Let \( V \) be a separable Banach space and \( B \subseteq B(V) \) a bounded subset.

(i) Prove: If \( S \subseteq V \) is dense and a net \( \{ A_i \} \subseteq B \) satisfies \( \| A_i x \| \to 0 \) for all \( x \in S \) then \( \| A_i x \| \to 0 \) for all \( x \in V \), thus \( A_i \to 0 \) in the strong operator topology.

(ii) Prove that the topological space \((B, \tau_{\text{sot}})\) is metrizable.

(iii) BONUS: Prove that \((V, \tau_{\text{sot}})\) is not metrizable if \( V \) is infinite dimensional.

9.9 Corollary (Hellinger-Toeplitz theorem) \(^{28}\) If \( H \) is a Hilbert space and a linear map \( A : H \to H \) is self-adjoint (i.e. \( \langle Ax, y \rangle = \langle x, Ay \rangle \) for all \( x, y \in H \)) then \( A \) is bounded.

Proof. The set \( \mathcal{F} = \{ x \mapsto \langle x, Ay \rangle \mid y \in H, \| y \| \leq 1 \} \) clearly is contained in \( H^* = B(H, \mathbb{C}) \). For each \( x \in H \) we have
\[
\mathcal{F} x = \{ \langle x, Ay \rangle \mid y \in H, \| y \| \leq 1 \} = \{ \langle Ax, y \rangle \mid y \in H, \| y \| \leq 1 \}.
\]
With Cauchy-Schwarz and \( \| y \| \leq 1 \) we have \( |\langle Ax, y \rangle| \leq \| Ax \| \). Thus \( \mathcal{F} \) is pointwise bounded and therefore uniformly bounded by Theorem 9.2. Thus there is an \( M \in [0, \infty) \) such that \( |\langle Ax, y \rangle| = |\langle x, Ay \rangle| \leq M \| x \| \) for all \( y \in H \) with \( \| y \| \leq 1 \), and this implies \( \| A \| \leq M \). \( \Box \)

9.10 Remark The Hellinger-Toeplitz Theorem shows that on a Hilbert space \( H \) there are no unbounded linear operators \( A : H \to H \) satisfying \( \langle Ax, y \rangle = \langle x, Ay \rangle \) \( \forall x, y \). This is a typical example of a ‘no-go-theorem’. Occasionally such results are a nuisance. After all, the operator of multiplication by \( n \) on \( \ell^2(\mathbb{N}) \) ‘obviously’ is self-adjoint. What Hellinger-Toeplitz really says is that such an operator cannot be defined everywhere, i.e. on all of \( H \). This leads to the notion of symmetric operators, and also illustrates that no-go theorems often can be circumvented by generalizing the setting. This is the case here, since the Hellinger-Toeplitz theorem only applies to operators that are defined everywhere. \( \Box \)

9.11 Definition A symmetric operator on a Hilbert space \( H \) is a linear map \( A : D \to H \), where \( D \subseteq H \) is a dense linear subspace, that satisfies \( \langle Ax, y \rangle = \langle x, Ay \rangle \) \( \forall x, y \). This is a typical example of a ‘no-go-theorem’. Occasionally such results are a nuisance. After all, the operator of multiplication by \( n \) on \( \ell^2(\mathbb{N}) \) ‘obviously’ is self-adjoint. What Hellinger-Toeplitz really says is that such an operator cannot be defined everywhere, i.e. on all of \( H \). This leads to the notion of symmetric operators, and also illustrates that no-go theorems often can be circumvented by generalizing the setting. This is the case here, since the Hellinger-Toeplitz theorem only applies to operators that are defined everywhere. \( \Box \)

9.12 Exercise Let \( H = \ell^2(\mathbb{N}, \mathbb{C}), D = \{ f \in \ell^2(\mathbb{N}, \mathbb{C}) \mid \sum_n |n f(n)|^2 < \infty \} \subseteq H \) and \( A : D \to H, (Af)(n) = nf(n) \). Prove:

(i) \( D \subseteq H \) is a dense proper linear subspace.

(ii) \( A : D \to H \) is symmetric and unbounded.

There is an extensive theory of unbounded linear operators defined on dense subspaces of a Hilbert space. Most books on (linear) functional analysis have a chapter on them, e.g., [54, 58, 11, 62]. This theory is quite important for applications to differential equations and quantum mechanics, but since it is quite technical one should not approach it before one has mastered the material of this course.

\(^{28}\) Ernst David Hellinger (1883-1950), Otto Toeplitz (1881-1940). German mathematicians. Both were forced into exile in 1939.
9.3 The strong version, using Baire’s theorem

We have seen two applications of the statement $\mathcal{F} \subseteq B(E, F)$ pointwise bounded $\Rightarrow \mathcal{F}$ uniformly bounded. Some applications of the uniform boundedness theorem use the contraposition: If $\mathcal{F}$ is not uniformly bounded then it is not pointwise bounded, thus there exists $x \in E$ with $\sup_{A \in \mathcal{F}} \|Ax\| = \infty$. For some of these applications the following statement, which clearly implies Theorem 9.2, is a very definite improvement of the latter:

9.13 Theorem Let $E$ be a Banach space, $F$ a normed space and $\mathcal{F} \subseteq B(E, F)$. Then either $\mathcal{F}$ is uniformly bounded or the set $\{x \in E \mid \sup_{A \in \mathcal{F}} \|Ax\| = \infty\} \subseteq E$ is dense $G_δ$.

Proof. The map $F \to \mathbb{R}_{\geq 0}$, $x \mapsto \|x\|$ is continuous and each $A \in \mathcal{F}$ is bounded, thus continuous. Therefore the map $f_A : E \to \mathbb{R}_{\geq 0}$, $x \mapsto \|Ax\|$ is continuous for every $A \in \mathcal{F}$. Defining for each $n \in \mathbb{N}$

$$V_n = \{x \in E \mid \sup_{A \in \mathcal{F}} \|Ax\| > n\},$$

the definition of sup implies

$$V_n = \{x \in E \mid \exists A \in \mathcal{F} : \|Ax\| > n\} = \bigcup_{A \in \mathcal{F}} \{x \in E \mid \|Ax\| > n\} = \bigcup_{A \in \mathcal{F}} f_A^{-1}((n, \infty)),
$$

which is open by continuity of the $f_A$.

If $V_n$ is non-dense for some $n \in \mathbb{N}$, there exists $x_0 \in E$ and $r > 0$ such that $B(x_0, r) \cap V_n = \emptyset$. This means $\sup_{A \in \mathcal{F}} \|A(x_0 + x)\| \leq n$ for all $x$ with $\|x\| < r$. With $x = (x_0 + x) - x_0$ and the triangle inequality we have

$$\|Ax\| \leq \|A(x_0 + x)\| + \|Ax_0\| \leq 2n \quad \forall A \in \mathcal{F}, \ x \in B(0, r).$$

This implies $\|Ax\| \leq 2n/r$ for all $A \in \mathcal{F}$, thus $\mathcal{F}$ is uniformly bounded.

If $V_n \subseteq E$ is dense for all $n \in \mathbb{N}$ then Baire’s Theorem A.20 gives that the $G_δ$-set $X = \bigcap_{n \in \mathbb{N}} V_n$ is dense. Since the definition of the $V_n$ gives $X = \{x \in E \mid \sup_{A \in \mathcal{F}} \|Ax\| = \infty\}$, the claim is proven.

9.4 Application: A dense set of continuous functions with divergent Fourier series

Let $f : \mathbb{R} \to \mathbb{C}$ be $2\pi$-periodic, i.e. $f(x + 2\pi) = f(x) \ \forall x$, and integrable over finite intervals. Define

$$c_n(f) = \frac{1}{2\pi} \int_0^{2\pi} f(x) e^{-inx} \, dx \quad (9.2)$$

and

$$S_n(f)(x) = \sum_{k=-n}^{n} c_k(f) e^{ikx}, \quad n \in \mathbb{N}. \quad (9.3)$$

The fundamental problem of the theory of Fourier series is to find conditions for the convergence $S_n(f)(x) \to f(x)$ as $n \to \infty$, where convergence can be understood as (possibly almost) everywhere pointwise or w.r.t. some norm, like $\| \cdot \|_2$ (as in Example 6.43) or $\| \cdot \|_\infty$. Here we will discuss only continuous functions and we identify continuous $2\pi$-periodic functions with continuous functions on $S^1$. It is not hard to show that $S_n(f)(x) \to f(x)$ if $f$ is differentiable

---

29Mystifyingly, not many authors state this better result, even though using Baire it comes out without extra effort.
at \( x \) (or just Hölder continuous: \( |f(x') - f(x)| \leq C|x' - x|^D \) with \( C, D > 0 \) for \( x' \) near \( x \)) and that convergence is uniform when \( f \) is continuously differentiable (or the Hölder condition holds uniformly in \( x, x' \)). (See any number of books on Fourier analysis, e.g. [74, 34].)

Assuming only continuity of \( f \) one can still prove that \( \lim_{n \to \infty} S_n(f)(x) = f(x) \) if the limit exists, but there actually exist continuous functions \( f \) such that \( S_n(f)(x) \) diverges at some \( x \). Such functions were first constructed in the 1870s using ‘condensation of singularities’, a relative and precursor of the gliding hump method. Nowadays, most textbook presentations of such functions are based on Lemma 9.15 below combined with either the uniform boundedness theorem or constructions ‘by hand’, see e.g. [34, Section II.2], that are quite close in spirit to the uniform boundedness method.

However, individual examples of continuous functions with divergent (in a point) Fourier series can be produced in a totally constructive fashion, avoiding all choice axioms! (See [48] for a very classical example.) But using non-constructive arguments seems unavoidable if one wants to prove that there are many such functions as in the following:

9.14 Theorem There is a dense \( G_\delta \)-set \( X \subseteq C(S^1) \) such that \( \{S_n(f)(0)\}_{n \in \mathbb{N}} \) diverges for each \( f \in X \).

Proof. Inserting (9.2) into (9.3) we obtain

\[
S_n(f)(x) = \frac{1}{2\pi} \sum_{k=-n}^{n} e^{ikx} \int_{0}^{2\pi} f(t)e^{-ikt} dt = \frac{1}{2\pi} \int_{0}^{2\pi} f(t) \left( \sum_{k=-n}^{n} e^{ik(x-t)} \right) dt = (D_n \ast f)(x),
\]

where \( \ast \) denotes convolution, defined for \( 2\pi \)-periodic \( f,g \) by \( (f \ast g)(x) = \frac{1}{2\pi} \int_{0}^{2\pi} f(t)g(x-t) dt \), and

\[
D_n(x) := \sum_{k=-n}^{n} e^{ikx} = \frac{\sin(n + \frac{1}{2})x}{\sin \frac{x}{2}}
\]

is the Dirichlet kernel. The quickest way to check the last identity is the telescoping calculation

\[
(e^{ix/2} - e^{-ix/2})D_n(x) = \sum_{k=-n}^{n} e^{ix(k+1/2)} - \sum_{k=-n}^{n} e^{ix(k-1/2)} = e^{ix(n+1/2)} - e^{-ix(n+1/2)},
\]

together with \( e^{ix} - e^{-ix} = 2i \sin x \). Since \( D_n(x) \) is an even function, we have

\[
\varphi_n(f) := S_n(f)(0) = \frac{1}{2\pi} \int_{0}^{2\pi} f(x)D_n(x) dx.
\]

It is clear that the norm of the map \( \varphi_n : (C(S^1), \| \cdot \|_\infty) \to \mathbb{C} \) is bounded above by \( \|D_n\|_1 \).

For \( g_n(x) = \text{sgn}(D_n(x)) \) we have \( \varphi_n(g_n) = (2\pi)^{-1} \int_{0}^{2\pi} |D_n(x)| dx =: \|D_n\|_1 \). While \( g_n \) is not continuous, we can find a sequence of continuous \( g_{n,m} \) bounded by 1 such that \( g_{n,m} \xrightarrow{m \to \infty} g_n \) pointwise. Now Lebesgue’s dominated convergence theorem implies \( \varphi_n(g_{n,m}) \to \varphi_n(g_n) = \|D_n\|_1 \), which implies \( \|\varphi_n\| = \|D_n\|_1 \). By Lemma 9.15 below, \( \|D_n\|_1 \to \infty \) as \( n \to \infty \). Thus the family \( \mathcal{F} = \{\varphi_n\} \subseteq B(C(S^1), \mathbb{C}) \) is not uniformly bounded. Now Theorem 9.13 implies that the set \( X = \{f \in C(S^1, \mathbb{C}) \mid \{S_n(f)(0)\} \text{ is unbounded}\} \) is dense \( G_\delta \).

9.15 Lemma We have \( \|D_n\|_1 \geq \frac{4}{\pi^2} \log n \) for all \( n \in \mathbb{N} \).
Proof. Using $|\sin x| \leq |x|$ for all $x \in \mathbb{R}$, we compute

$$
\|D_n\|_1 = \frac{1}{2\pi} \int_{-\pi}^{\pi} |D_n(x)|dx \geq \frac{2}{\pi} \int_0^{\pi} \left| \sin \left( n + \frac{1}{2} \right) x \right| \frac{dx}{x}
$$

$$
= \frac{2}{\pi} \int_0^{(n+1/2)\pi} |\sin x| \frac{dx}{x} \geq \frac{2}{\pi} \sum_{k=1}^{n} \int_{(k-1)\pi}^{k\pi} \frac{|\sin x|}{x} dx
$$

$$
\geq \frac{2}{\pi} \sum_{k=1}^{n} \frac{1}{k} \int_0^{\pi} \sin x dx = \frac{4}{\pi^2} \sum_{k=1}^{n} \frac{1}{k} \geq \frac{4}{\pi^2} \log n,
$$

where we used $\sum_{k=1}^{n} 1/k \geq \int_1^{n+1} dx/x = \log(n+1) > \log n$. \hfill \blacksquare

10 Open mappings and closed graphs

10.1 The Open Mapping Theorem

10.1 Theorem (Schauder 1930) \(^{30}\) Let $E, F$ be Banach spaces and let $T \in B(E, F)$ (thus linear and bounded) be surjective. Then $T$ is an open map. (I.e. $TU \subseteq F$ is open for every open $U \subseteq E$.)

Proofs of the open mapping theorem tend to be longish and monolithic. I try to give a more accessible presentation. Following [24], we begin with a lemma that deduces surjectivity of a linear map from a certain approximate surjectivity property (and also leads to a quick proof of the Tietze extension theorem in topology, see Appendix A.6):

10.2 Lemma Let $E$ be a Banach space, $F$ a normed space (real or complex) and $T : E \to F$ a linear map. Assume also that there are $m > 0$ and $r \in (0, 1)$ such that for every $y \in F$ there is an $x_0 \in E$ with $\|x_0\|_E \leq m \|y\|_F$ and $\|y - Tx_0\|_F \leq r \|y\|_F$. Then for every $y \in F$ there is an $x \in E$ such that $\|x\|_E \leq \frac{m}{1-r} \|y\|_F$ and $Tx = y$. In particular, $T$ is surjective.

Proof. It suffices to consider the case $\|y\| = 1$. By assumption, there is $x_0 \in E$ such that $\|x_0\| \leq m$ and $\|y - Tx_0\| \leq r$. Now, applying the hypothesis to $y - Tx_0$ instead of $y$, we find an $x_1 \in E$ with $\|x_1\| \leq rm$ and $\|y - T(x_0 + x_1)\| \leq r^2$. Continuing this inductively (thus using $DC_\omega$!) we obtain a sequence $\{x_n\}$ such that for all $n \in \mathbb{N}$

$$
\|x_n\| \leq r^n m, \quad (10.1)
$$

$$
\|y - T(x_0 + x_1 + \cdots + x_n)\| \leq r^{n+1}. \quad (10.2)
$$

Now, (10.1) together with completeness of $E$ implies, cf. Proposition 4.2, that $\sum_{n=0}^{\infty} x_n$ converges to an $x \in E$ with

$$
\|x\| \leq \sum_{n=0}^{\infty} \|x_n\| \leq \sum_{n=0}^{\infty} r^n m = \frac{m}{1-r},
$$

and taking $n \to \infty$ in (10.2) gives $y = Tx$. \hfill \blacksquare

10.3 Proposition Let $E$ be a Banach space, $F$ a normed space and $T \in B(E, F)$ such that $B^F(0, \beta) \subseteq T(B^E(0, \alpha))$ for certain $\alpha, \beta > 0$. Then $B^F(0, \beta') \subseteq T(B^E(0, \alpha))$ if $0 < \beta' < \beta$.

Proof. If $0 < \beta' < \beta'' < \beta$ then $\overline{B}^F(0, \beta'') \subseteq B^F(0, \beta) \subseteq \overline{T B^E(0, \alpha)}$. Equivalently (since $x \mapsto \lambda x$ is a homeomorphism for every $\lambda > 0$), $\overline{B}^F(0, 1) \subseteq \overline{T B^E(0, \alpha/\beta''')}$.

With the definition of the closure, this means that for every $y \in F$, $\|y\| \leq 1$ and $\varepsilon > 0$ there exists $x \in E$ with $\|x\| < \alpha/\beta'$ and $\|Tx - y\| < \varepsilon$. Equivalently, for every $y \in F$ and $\varepsilon > 0$ there is $x \in E$ with $\|x\| < \frac{\varepsilon}{\beta'}$ and $\|Tx - y\| < \varepsilon\|y\|$. Now Lemma 10.2 gives (assuming $\varepsilon < 1$) that for every $y \in F$ there is $x \in E$ with $Tx = y$ and $\|x\| \leq \frac{\alpha/\beta'}{1-\varepsilon} \|y\|$. If we choose $\varepsilon \in (0, 1 - \frac{\beta'}{\beta''})$ then $\frac{\beta'/\beta''}{1-\varepsilon} < 1$, so that for every $y \in F$ with $\|y\| \leq \beta'$ there is $x \in E$ with $Tx = y$ and $\|x\| < \alpha$. Thus $\overline{B}^F(0, \beta') \subseteq \overline{T B^E(0, \alpha)}$. ■

Proof of Theorem 10.1. Since $T$ is surjective, we have

$$F = T(E) = \bigcup_{n=1}^{\infty} \overline{T(B^E(0, n))}.$$ 

Since $F$ is a complete metric space and has non-empty interior $F^0 = F \neq \emptyset$, Corollary A.21 of Baire’s theorem implies that at least one of the closed sets $\overline{T(B^E(0, n))}$ has non-empty interior. Thus there are $n \in \mathbb{N}, y \in F, \varepsilon > 0$ such that $B^F(y, \varepsilon) \subseteq \overline{T(B^E(0, n))}$. If $x \in B^F(0, \varepsilon)$ then $2x = (y + x) - (y - x)$, thus $2B^F(0, \varepsilon) \subseteq B^F(y, \varepsilon) - B^F(y, \varepsilon)$ and thus

$$B^F(0, \varepsilon) \subseteq \frac{1}{2}(B^F(y, \varepsilon) - B^F(y, \varepsilon)) \subseteq \frac{1}{2}(\overline{T(B^E(0, n))} - \overline{T(B^E(0, n))}) = \overline{T(B^E(0, n))}.$$ 

Now Proposition 10.3 implies that $B^F(0, \varepsilon') \subseteq T(B^E(0, n))$ for some $\varepsilon' > 0$ (actually every $\varepsilon' \in (0, \varepsilon)$, but we don’t need this). By linearity we have that for every $\delta > 0$ there is a $\delta' > 0$ such that $B^F(0, \delta') \subseteq TB^E(0, \delta)$. Now using the linearity of $T$, proving its openness is routine. ■

10.4 Exercise. Let $E, F$ be normed spaces and $T : E \rightarrow F$ linear such that for every $\delta > 0$ there is $\delta' > 0$ for which $B^F(0, \delta') \subseteq TB^E(0, \delta)$. Prove that $U$ is open.

10.2 The Bounded Inverse Theorem

Now we can prove Theorem 3.14:

10.5 Corollary (Banach 1929) If $E, F$ are Banach spaces and $T \in B(E, F)$ (thus linear and bounded) is a bijection then also $T^{-1}$ is bounded. (Thus $T$ is a homeomorphism.)

Proof. By Theorem 10.1, $T$ is open. Thus the inverse $T^{-1}$ that exists by bijectivity (and clearly is linear) is continuous, thus bounded by Lemma 4.13. ■

10.6 Definition. A linear map $A : E \rightarrow F$ between normed spaces that is a bijection and a homeomorphism is called an isomorphism of normed spaces. (Not to be confused with isometric isomorphisms, for which $\|Ax\| = \|x\| \forall x \in E$.) If an (isometric) isomorphism $A : E \rightarrow F$ exists, we write $E \simeq F$ ($E \cong F$).

10.7 Remark. 1. If $\| \cdot \|_1, \| \cdot \|_2$ are norms on $V$ then $(V, \| \cdot \|_1) \cong (V, \| \cdot \|_2)$ if and only if the two norms are equivalent.

2. The Bounded Inverse Theorem is a special case of the Open Mapping Theorem, but it also implies the latter: Assume that the former holds, that $E, F$ are Banach spaces and that $T \in B(E, F)$ is surjective. The kernel $\ker T \subseteq E$ is closed, so that the quotient space $E/\ker T$ is a Banach space, and the quotient map $p : E \rightarrow E/\ker T$ is continuous and open by Proposition
7.2. Since $T$ is surjective, the induced map $\tilde{T} : E/\ker T \to F$ is a continuous bijection, so that $\tilde{T}^{-1} : F \to E/\ker T$ is continuous by the Bounded Inverse Theorem. Equivalently, $\tilde{T}$ is open, so that the $T = \tilde{T} \circ p$ is open as the composite of two open maps.

3. Also the Bounded Inverse Theorem has an interesting application to Fourier analysis: For $f \in L^1([0,2\pi])$, we define the Fourier coefficients $\hat{f}(n) = (2\pi)^{-1} \int_0^{2\pi} f(t)e^{-int}dt$ for all $n \in \mathbb{N}$. It is immediate that $\|\hat{f}\|_\infty \leq \|f\|_1$, and is not hard to prove the Riemann-Lebesgue theorem $\hat{f} \in c_0(\mathbb{Z}, \mathbb{C})$ and injectivity of the resulting map $L^1([0,2\pi]) \to c_0(\mathbb{Z}, \mathbb{C})$, $f \mapsto \hat{f}$, see e.g. [61, Theorem 5.15] or [34]. If this map was surjective, the Bounded Inverse Theorem would give $\|f\|_1 \leq C\|\hat{f}\|_\infty$. For the Dirichlet kernel it is immediate that $\hat{D}_n(m) = \chi_{[-n,n]}(m)$, thus $\|\hat{D}_n\|_\infty = 1$ for all $n \in \mathbb{N}$. Since we know that $\|D_n\|_1 \to \infty$, we would have a contradiction. Thus $L^1([0,2\pi]) \to c_0(\mathbb{Z}, \mathbb{C})$, $f \mapsto \hat{f}$ is not surjective.

4. The Open Mapping Theorem can be generalized to the case where $E$ is an $F$-space, i.e. a TVS admitting a complete translation-invariant metric. See [62, Theorem 2.11].

10.8 Exercise Let $V$ be a Banach space.

(i) Let $W, Z \subseteq V$ be closed subspaces such that $W + Z = V$ and $W \cap Z = \{0\}$. Give $W \oplus Z$ the norm $\|(w, z)\| = \|w\| + \|z\|$. Prove that $\alpha : W \oplus Z \to V, (w, z) \mapsto w + z$ is a homeomorphism, thus an isomorphism of Banach spaces.

(ii) If $W \subseteq V$ is complemented then:

- There is a bounded linear map $P \in B(V)$ with $P^2 = P$ and $W = PV$. (The converse was proven in Exercise 7.12.)
- Every closed $Z \subseteq V$ complementary to $W$ is isomorphic to $V/W$ as Banach spaces.

10.9 Exercise Let $V$ be a Banach space and $W, Z \subseteq V$ closed linear subspaces satisfying $W \cap Z = \{0\}$, so that $W + Z \cong W \oplus Z$ algebraically. Prove that $W + Z \subseteq V$ is closed if and only if the projection $W + Z \to W : w + z \mapsto w$ is continuous.

[There is a generalization without the assumption $W \cap Z = \{0\}$, but we don’t pursue this.]

10.10 Exercise Let $V, W$ be Banach spaces and $A \in B(V, W)$ such that $\dim(W/AV) < \infty$. (The quotient $W/AV$ is called the cokernel of $A$. Some authors define the cokernel as $W/AV$, but we don’t!)

(i) Prove that $AW \subseteq V$ is closed, assuming injectivity of $A$.

(ii) Remove the injectivity assumption.

10.11 Exercise It is not true that every subspace $W \subseteq V$ with $\dim(V/W) < \infty$ of a Banach space $V$ is closed! Find a counterexample! (Hint: codimension one.)

10.2.1 The Closed Graph Theorem

We quickly look at an interesting result equivalent to the Bounded Inverse Theorem, but we will not need it afterwards.

If $f : X \to Y$ is a function, the graph of $f$ is the set $\mathcal{G}(f) = \{(x, f(x)) \mid x \in X\} \subseteq X \times Y$.

10.12 Exercise Let $X$ be a topological space, $Y$ a Hausdorff space and $f : X \to Y$ continuous. Prove that $\mathcal{G}(f) \subseteq X \times Y$ is closed.
If $E, F$ are normed spaces, we know that $\|(x, y)\| = \|x\| + \|y\|$ is a norm on $E \oplus F$, complete if $E$ and $F$ are. The projections $p_1 : E \oplus F \to E$, $p_2 : E \oplus F \to F$ are bounded.

10.13 **Lemma** Let $E, F$ be normed spaces and $T \in B(E, F)$. Then the following are equivalent:

(i) The graph $\mathcal{G}(T) = \{(x, Tx) \mid x \in E\} \subseteq E \oplus F$ of $T$ is closed.

(ii) Whenever $\{x_n\}_{n \in \mathbb{N}} \subseteq E$ is a sequence such that $x_n \to x \in E$ and $Tx_n \to y \in F$, we have $y = Tx$.

**Proof.** Since $E \oplus F$ is a metric space, $\mathcal{G}(T)$ is closed if and only if it contains the limit $(x, y)$ of every sequence $\{(x_n, y_n)\}$ in $\mathcal{G}(T)$ that converges to some $(x, y) \in E \oplus F$. But a sequence in $\mathcal{G}(T)$ is of the form $\{(x_n, Tx_n)\}$, and $(x, y) \in \mathcal{G}(T) \iff y = Tx$. □

10.14 **Definition** If $E, F$ are normed spaces, a linear map $T \in B(E, F)$ satisfying the equivalent statements in the lemma is called closed.

10.15 **Theorem (Banach 1929)** If $E, F$ are Banach spaces, then a linear map $T : E \to F$ is bounded if and only if it is closed.

**Proof.** Let $E, F$ be Banach spaces, and let $T : E \to F$ be linear. If $T$ is bounded then it is continuous, thus $\mathcal{G}(T)$ is closed by Exercise 10.12. Now assume $T$, thus $\mathcal{G}(T)$, is closed. The Cartesian product $E \oplus F$ with norm $\|(e, f)\| = \|e\| + \|f\|$ is a Banach space. The linear subspace $\mathcal{G}(T) \subseteq E \oplus F$ is closed by assumption, thus a Banach space. Since the projection $p_1 : \mathcal{G}(T) \to E$ is a bounded bijection, by Corollary 10.5 it has a bounded inverse $p_1^{-1} : E \to \mathcal{G}(T)$. Then also $T = p_2 \circ p_1^{-1}$ is bounded. □

10.16 **Exercise** Show that the Bounded Inverse Theorem (Corollary 10.5) can be deduced from the Closed Graph Theorem. (Thus the three main results of this section are ‘equivalent’.)

10.17 **Remark** 1. The Hellinger-Toeplitz Theorem (Corollary 9.9) can also be deduced from the Closed Graph Theorem: Let $\{x_n\} \subseteq H$ be a sequence converging to $x \in H$ and assume that $Ax_n \to y$. Then

$$\langle Ax, z \rangle = \langle x, Az \rangle = \lim_n \langle x_n, Az \rangle = \lim_n \langle Ax_n, z \rangle = \langle y, z \rangle \quad \forall z \in H,$$

thus $Ax = y$. Thus $A$ is closed and therefore bounded by Theorem 10.15.

2. Since we deduced the Hellinger-Toeplitz theorem from the weak version of the uniform boundedness theorem, it is moderately interesting [But not too much: We needed $DC_\omega$ to prove the closed graph theorem, whereas we know that $AC_\omega$ suffices for proving the weak version of the uniform boundedness theorem! And with $DC_\omega$ one has the better Theorem 9.13.] that also the latter can also be deduced from the closed graph theorem: □

10.18 **Exercise** Let $E, F$ be Banach spaces and $\mathcal{F} \subseteq B(E, F)$ a pointwise bounded family. Use the Closed Graph Theorem to prove that $\mathcal{F}$ is uniformly bounded, as follows:

(i) Prove that $F_\mathcal{F} = \{y_A \mid A \in \mathcal{F}, y_A \in F \} = \text{Fun}(\mathcal{F}, F) \mid \sup_{A \in \mathcal{F}} \|y_A\| < \infty \}$ is a Banach space.

(ii) Show that pointwise boundedness of $\mathcal{F}$ is equivalent to $T(E) \subseteq F_\mathcal{F}$.

(iii) Prove that the graph $\mathcal{G}(T) \subseteq E \oplus F_\mathcal{F}$ of $T$ is closed. (Thus $T$ is bounded by Theorem 10.15.)

(iv) Deduce uniform boundedness of $\mathcal{F}$ from the boundedness of $T$.

(v) Remove the requirement that $F$ be complete.
10.3 Boundedness below. Invertibility

10.19 Definition Let \( E, F \) be normed spaces and let \( A : E \to F \) be a linear map. Then \( A \) is called bounded below\(^{31}\) if there is a \( \delta > 0 \) such that \( \|Ax\| \geq \delta \|x\| \) \( \forall x \in E \).

(Equivalently, \( \inf_{\|x\|=1} \|Ax\| > 0 \).)

It is obvious that boundedness below of a map implies injectivity, but the converse is not true. Furthermore, the image \( AE = \{Ax \mid x \in E\} \) of a linear map \( A : E \to F \) need not be closed. In particular, the image can be dense without \( A \) being surjective. The operator \( A \in B(E), \) where \( E = \ell^2(\mathbb{N}, \mathbb{C}) \), defined by \( (Af)(n) = f(n)/n \) exemplifies both phenomena.

10.20 Exercise Let \( E, F \) be be normed spaces, where \( E \) is finite dimensional, and let \( A : E \to F \) be an injective linear map. Prove that \( A \) is bounded below.

10.21 Lemma Let \( E, F \) be normed spaces and \( A : E \to F \) a linear bijection. Then

\[
\inf_{\|x\|=1} \|Ax\| = \|A^{-1}\|^{-1}.
\]

In particular, \( A \) is bounded below if and only if its (set-theoretic) inverse \( A^{-1} \) is bounded.

Proof. Using the invertibility of \( A \), thus bijectivity of \( x \mapsto Ax \), we have

\[
\|A^{-1}\| = \sup_{y \in F \setminus \{0\}} \frac{\|A^{-1}y\|}{\|y\|} = \sup_{x \in E \setminus \{0\}} \frac{\|x\|}{\|Ax\|} = \left( \inf_{x \in E \setminus \{0\}} \frac{\|Ax\|}{\|x\|} \right)^{-1} = \left( \inf_{\|x\|=1} \|Ax\| \right)^{-1}.
\]

The second statement follows immediately. \( \blacksquare \)

The following generalizes Corollary 4.6:

10.22 Lemma If \( E \) is a Banach space, \( F \) is a normed space and \( A : E \to F \) is a linear map that is bounded and bounded below then \( AE \subseteq F \) is closed.

Proof. Since \( A \) is bounded below, it is injective, thus \( \widetilde{A} : E \to AE \) (the map \( A \), with the codomain replaced by \( AE \)) is a bijection. Now \( \widetilde{A}^{-1} : AE \to E \) is bounded by Lemma 10.21. Thus if \( \{f_n\} \) is a Cauchy sequence in \( AE \), \( \{\widetilde{A}^{-1}f_n\} \) is a Cauchy sequence in \( E \). Since \( E \) is complete, there is \( e \in E \) such that \( \widetilde{A}^{-1}f_n \to e \). Since \( A \) is bounded, \( \{f_n = A(\widetilde{A}^{-1}f_n)\} \) converges to \( Ae \in AE \). Thus \( AE \) is complete, thus closed. \( \blacksquare \)

10.23 Definition If \( E, F \) are normed spaces then \( A \in B(E,F) \) is called invertible if there is a \( B \in B(F,E) \) such that \( BA = \text{id}_E \) and \( AB = \text{id}_F \).

10.24 Proposition Let \( E, F \) be a Banach spaces and \( A \in B(E,F) \). Then the following are equivalent:

(i) \( A \) is invertible.

(ii) \( A \) is injective and surjective.

(iii) \( A \) is bounded below and has dense image.

\(^{31}\)This terminology clashes with another one according to which a self-adjoint operator \( A \) is bounded below if \( \sigma(A) \subseteq [c, \infty) \) for some \( c \in \mathbb{R} \). Since we consider only bounded operators, we’ll have no use for this notion. The problem could be avoided by writing ‘bounded away from zero’, as some authors do, but this is a bit tedious.
Proof. It is clear that invertibility implies injectivity and surjectivity, thus in particular dense image. Since $A^{-1}$ is bounded, Lemma 10.21 gives that $A$ is bounded below.

If (ii) holds then the the set-theoretic inverse, clearly linear, is bounded by the bounded inverse theorem, (Corollary 10.5). Thus $A$ is invertible in the sense of Definition 10.19.

Assume (iii). By boundedness below, $A$ is injective. And $AE \subseteq F$ is dense by assumption and closed by Lemma 10.22, thus $AE = F$. Thus $A$ is injective and surjective. Now boundedness of the inverse $A^{-1}$ follows from boundedness below of $A$ and Lemma 10.21. ■

10.25 Remark 1. Note that dense image is weaker than surjectivity, while boundedness below is stronger than injectivity. The point of criterion (iii) is that it can be quite hard to verify surjectivity of $A$ directly, while density of the image usually is easier to establish.

2. The material on bounded below maps discussed so far, including (i)$\Leftrightarrow$(iii) in Proposition 10.24, was entirely elementary and could be moved to Section 4.

10.26 Exercise Let $E, F$ be Banach spaces and $A \in B(E, F)$. Prove:

(i) If $A$ is injective then $AE \subseteq F$ is closed $\Leftrightarrow A$ is bounded below.

(ii) If ker $A$ has a complement $W$ then $AE \subseteq F$ is closed $\Leftrightarrow A|_{W}$ is bounded below.

(iii) If $E = H$ is a Hilbert space then $AH \subseteq F$ is closed if and only if $A \upharpoonright (\ker A)^{\perp}$ is bounded below.

10.27 Exercise Let $H$ be a Hilbert space and $A \in B(H)$ such that $|\langle Ax, x \rangle| \geq C\|x\|^{2}$ for some $C > 0$. Prove that $A$ is invertible and $\|A^{-1}\| \leq C^{-1}$.

11 Spectrum of bounded operators and of (elements of) Banach algebras

11.1 The spectra of $A \in B(E)$

We now specialize from $B(E, F)$ to $B(E)$, i.e. linear maps from a normed space to itself. If $E$ is a finite dimensional vector space and $A \in \text{End } E$, it is well known that one has: $A$ is injective $\Leftrightarrow A$ is surjective $\Leftrightarrow A$ is invertible.

It is extremely important that this fails in infinite dimensions:

11.1 Definition Let $E = \ell^{p}(\mathbb{N}, \mathbb{F})$, where $1 \leq p \leq \infty$. Define $L, R \in B(E)$ by

$$(Lf)(n) = f(n + 1), \quad (Rf)(n) = \begin{cases} 0 & \text{if } n = 1 \\ f(n - 1) & \text{if } n \geq 2 \end{cases}$$

Equivalently: $R\delta_{n} = \delta_{n+1}$, $L\delta_{1} = 0$, $L\delta_{n} = \delta_{n-1}$ if $n \geq 2$, which is why we call $L, R$ the left and right, respectively, shift operators on $E$.

It is immediate that $R$ is injective (in fact isometric), but not surjective (since $(Rf)(1) = 0 \ \forall f \in E$) while $L$ is surjective, but not injective (since $Lf$ does not depend on $f(1)$). One easily checks $LR = \text{id}_{E}$, while $RL \neq \text{id}_{E}$ since $RL = P_{2}$ (notation from Exercise 9.7).

Let $E$ be a finite dimensional $\mathbb{F}$-vector space, $A \in \text{End } E$ and $\lambda \in \mathbb{F}$. Then failure of $A - \lambda 1_{E}$ to be invertible is equivalent to ker$(A - \lambda 1_{E}) \neq \{0\}$, thus the existence of a non-zero $x \in E$ with $Ax = \lambda x$. Thus $\lambda$ is an eigenvalue of $A$ and $x$ is a corresponding eigenvector.
Passing to an infinite dimensional Banach space $E$, there can be $A \in B(E)$ and $\lambda \in \mathbb{F}$ for which $A - \lambda \mathbf{1}$ is injective, but not surjective, thus not invertible. Such $\lambda$ are not eigenvalues, but they turn out to be equally important as the former. This motivates the following:

11.2 Definition Let $E$ be a Banach space over $\mathbb{F}$ and $A \in B(E)$. Then

- The spectrum\(^{32}\) $\sigma(A)$ is the set of $\lambda \in \mathbb{F}$ for which $A - \lambda \mathbf{1}_E$ is not invertible.
- The point spectrum $\sigma_p(A)$ consists of those $\lambda \in \mathbb{F}$ for which $A - \lambda \mathbf{1}_E$ is not injective. Equivalently, $\sigma_p(A)$ consists of the eigenvalues of $A$.
- The continuous spectrum $\sigma_c(A)$ consists of those $\lambda \in \mathbb{F}$ for which $A - \lambda \mathbf{1}_E$ is injective, but not surjective, while it has dense image, i.e. $(A - \lambda \mathbf{1}_E)E = E$.
- The residual spectrum $\sigma_r(A)$ consists of those $\lambda \in \mathbb{F}$ for which $A - \lambda \mathbf{1}_E$ is injective and $(A - \lambda \mathbf{1}_E)E \neq E$.

We have some immediate observations:

- It is obvious by construction that the sets $\sigma_p(A), \sigma_c(A), \sigma_r(A)$ are mutually disjoint and have $\sigma(A)$ as their union.
- Clearly $0 \in \sigma(A)$ is equivalent to non-invertibility of $A$ and $0 \in \sigma_p(A)$ to $\ker A \neq \{0\}$.
- If $E$ is finite dimensional then we know from linear algebra that injectivity and surjectivity of any $A \in B(E)$ are equivalent. Thus for all operators on a finite dimensional space we have $\sigma_c(A) = \sigma_r(A) = \emptyset$, thus $\sigma(A) = \sigma_p(A)$.
- If $E$ is infinite dimensional, the situation is much more complicated, thus more interesting. For example, the right shift $R$ on $\ell^2(\mathbb{N})$ is injective, but not surjective. Thus $0 \in \sigma(R)$, while $0 \notin \sigma_p(R)$.
- If $\lambda \in \sigma_p(A)$ then there is non-zero $x \in E$ with $Ax = \lambda x$. Then $A^n x = \lambda^n x \forall n \in \mathbb{N}$. With the definition of $\|A\|$ it follows that $|\lambda| \leq \inf_{n \in \mathbb{N}} \|A^n\|^{1/n}$. (This can be smaller than $\|A\|$, e.g. if $A$ is nilpotent, i.e. $A^n = 0$ for some $n \in \mathbb{N}$.)
- There are other interesting subsets of $\sigma(A)$, motivated by Proposition 10.24(iii):
  - The approximate point spectrum $\sigma_{\mathrm{app}}(A) = \{ \lambda \in \mathbb{F} \mid A - \lambda \mathbf{1} \text{ not bounded below} \}$. Clearly $\sigma_p(A) \subseteq \sigma_{\mathrm{app}}(A)$.
  - The compression spectrum $\sigma_{\mathrm{cp}}(A) = \{ \lambda \in \mathbb{F} \mid (A - \lambda \mathbf{1})E \neq F \}$. Obviously $\sigma(A) = \sigma_{\mathrm{app}}(A) \cup \sigma_{\mathrm{cp}}(A)$ (but the two need not be disjoint). And $\sigma_r(A) = \sigma_{\mathrm{cp}}(A) \setminus \sigma_p(A)$.
  - The discrete spectrum $\sigma_d(A) \subseteq \sigma_p(A)$, cf. Exercise 11.31. The essential spectrum $\sigma_{\mathrm{ess}}(A)$ (or rather one version of it – there are others) is $\sigma(A) \setminus \sigma_d(A)$.

11.3 Exercise Let $V$ be a Banach space, $A \in B(V)$, and let $W, Z \subseteq V$ closed subspaces such that $W + Z = V, W \cap Z = \{0\}$ and $AW \subseteq W, AZ \subseteq Z$. Prove $\sigma(A) = \sigma(A|_W) \cup \sigma(A|_Z)$ and $\sigma_t(A) = \sigma_t(A|_W) \cup \sigma_t(A|_Z)$ for all $t \in \{p, c, r\}$.

Before we try to compute the spectra of some interesting operators, it is better to first prove some general results, since they will be helpful also for studying examples. Remarkably one can get rather far using only the fact that $B(E)$ is a Banach algebra.

\(^{32}\)The choice of this term by Hilbert was nothing less than a stroke of genius since it turned out to fit exactly its later use in quantum theory.
11.2 The spectrum in a unital Banach algebra

Since $B(E)$ is a unital Banach algebra for every Banach space, all results proven here apply to bounded operators on Banach spaces. Restricting these results to $B(E)$ does not significantly simplify their proofs! In the beginning it does not matter whether $F$ is $\mathbb{R}$ or $\mathbb{C}$.

11.4 Definition If $A$ is a unital algebra over $F$ then $\text{Inv}A = \{ a \in A \mid \exists b \in A : ab = ba = 1 \}$ is the set of invertible elements of $A$. The spectrum of $a \in A$ is defined as

$$\sigma(a) = \{ \lambda \in F \mid a - \lambda 1 \notin \text{Inv}A \}.$$ 

The spectral radius of $a$ is $r(a) = \sup\{ |\lambda| \mid \lambda \in \sigma(a) \}$, where $r(a) = 0$ if $\sigma(a) = \emptyset$. (But we will soon prove $\sigma(a) \neq \emptyset$ for all $a \in A$ if $A$ is normed.)

11.5 Remark It is clear that for an element of the Banach algebra $B(E)$, where $E$ is a Banach space, this definition is equivalent to Definition 11.2. But in the present abstract setting there is no distinction between point, continuous and residual spectrum.

As to our standard example of a Banach algebra not of the form $B(V)$ with $V$ Banach:

11.6 Exercise (i) Let $X$ be a compact Hausdorff space. Recall that $(C(X,F), \| \cdot \|_\infty)$ is a Banach algebra. For $f \in C(X,F)$, prove $\sigma(f) = f(X) \subseteq F$.

(ii) If $S$ is any set, $\ell^\infty(S,F)$ is a Banach algebra w.r.t. pointwise multiplication. If $f \in \ell^\infty(S,F)$, prove $\sigma(f) = f(S)$.

There is another case where the spectrum is easy to determine:

11.7 Definition An element $a \in A$ of an algebra is called nilpotent if $a^n = 0$ for some $n \in \mathbb{N}$.

11.8 Lemma If $A$ is a unital algebra and $a \in A$ is nilpotent then $\sigma(a) = \{0\}$, thus $r(a) = 0$.

Proof. If $a \in A$ is nilpotent then the series $b = \sum_{n=0}^\infty a^n$ converges since it breaks off after finitely many terms. Now $(1-a)b = b(1-a) = \sum_{n=0}^\infty a^n - \sum_{n=1}^\infty a^n = 1$, thus $a - 1 \in \text{Inv}A$. Since the same holds for $a/\lambda$ whenever $\lambda \neq 0$, we have $\sigma(a) \subseteq \{0\}$. Since no nilpotent $a$ is invertible (why?) we have $\sigma(a) = \{0\}$. ■

11.9 Definition If $A$ is a unital Banach algebra, $a \in A$ is called quasi-nilpotent if $r(a) = 0$. As just proven, nilpotent $\Rightarrow$ quasi-nilpotent.

11.10 Lemma Let $A$ be a unital Banach algebra. Then

(i) If $a \in A$, $\|a\| < 1$ then $1 - a \in \text{Inv}A$.

(ii) $\text{Inv}A \subseteq A$ is open.

Proof. (i) If $\|a\| < 1$ then $\sum_{n=0}^\infty \|a^n\| \leq \sum_{n=0}^\infty \|a\|^n < \infty$, so that the series $\sum_{n=0}^\infty a^n$ converges to some $b \in A$ by completeness and Proposition 4.2. Now again $(1-a)b = b(1-a) = \sum_{n=0}^\infty a^n - \sum_{n=1}^\infty a^n = 1$, so that $1-a \in \text{Inv}A$.

(ii) If $a \in \text{Inv}A$ and $a' \in A$ with $\|a - a'\| < \|a^{-1}\|^{-1}$ then

$$\|1 - a^{-1}a'\| = \|a^{-1}(a-a')\| \leq \|a^{-1}\| \|a-a'\| < 1 \text{ so that } a^{-1}a' = 1 - (1 - a^{-1}a') \in \text{Inv}A,$$

thus $a' = a(a^{-1}a') \in \text{Inv}A$. This proves that $\text{Inv}A$ is open. ■

11.11 Proposition If $A$ is a unital Banach algebra and $a \in A$ then
(i) \( \sigma(a) \) is closed.

(ii) \( r(a) \leq \inf_{n \in \mathbb{N}} \|a^n\|^{1/n} \leq \|a\| \).

**Proof.** (i) If \( a \in \mathcal{A} \) then \( f_a : \mathbb{F} \to \mathcal{A} \), \( \lambda \mapsto a - \lambda \mathbf{1} \) is continuous, thus \( f_a^{-1}(\text{Inv}\mathcal{A}) \subseteq \mathbb{F} \) is open by Lemma 11.10(ii). Now \( \sigma(a) = \mathbb{F} \setminus f_a^{-1}(\text{Inv}\mathcal{A}) \) is closed.

(ii) If \( \lambda \in \mathbb{F} \), \( |\lambda| > \|a\| \) then \( \|a/\lambda\| < 1 \) so that \( \mathbf{1} - a/\lambda \in \text{Inv}\mathcal{A} \) by Lemma 11.10(i). Thus \( \lambda \mathbf{1} - a \in \text{Inv}\mathcal{A} \), so that \( \lambda \notin \sigma(a) \). This proves \( r(a) \leq \|a\| \).

In each associative unital algebra, a simple telescoping argument gives the formula

\[
z^n - 1 = (z - 1)(1 + z + z^2 + \cdots + z^{n-1}),
\]

known from finite geometric sums. If \( 0 \neq \lambda \in \sigma(a) \) then \( a/\lambda - \mathbf{1} \) is not invertible, thus putting \( z = a/\lambda \) in (11.1) gives that \( (a/\lambda)^n - \mathbf{1} \) is not invertible (since a product of two **commuting** elements is invertible if and only if both are invertible, cf. Exercise 11.21(i)). Thus \( \lambda^n \in \sigma(a^n) \), so that \( r(a^n) \leq r(a) \). Since this holds for all \( n \in \mathbb{N} \), using \( r(b) \leq \|b\| \) just proven, we have \( r(a) \leq \inf_{n \in \mathbb{N}} r(a^n)^{1/n} \leq \inf_{n \in \mathbb{N}} \|a^n\|^{1/n} \leq \|a\| \).

All this was not too difficult, but open questions remain. Settling them will be more work. We begin with a number of fairly elementary observations:

11.12 Lemma Let \( \mathcal{A} \) be a unital normed algebra and \( a \in \mathcal{A} \). Put \( \nu = \inf_{n \in \mathbb{N}} \|a^n\|^{1/n} \). Then

(i) \( \lim_{n \to \infty} \|a^n\|^{1/n} = \nu \).

(ii) For all \( \mu > \nu \) we have \( \left( \frac{a}{\mu} \right)^n \to 0 \) as \( n \to \infty \), but \( \left( \frac{a}{\nu} \right)^n \not\to 0 \) provided \( \nu > 0 \).

(This is of course trivial if \( \mu > \|a\| \), but our hypothesis is weaker when \( \nu < \|a\| \)).

(iii) If \( \nu = 0 \) then \( a \notin \text{Inv}\mathcal{A} \), thus \( 0 \in \sigma(a) \).

**Proof.** (i) With \( \|a^n\| \leq \|a\|^n \) we trivially have

\[
0 \leq \nu = \inf_{n \in \mathbb{N}} \|a^n\|^{1/n} \leq \liminf_{n \to \infty} \|a^n\|^{1/n} \leq \limsup_{n \to \infty} \|a^n\|^{1/n} \leq \|a\| < \infty. \tag{11.2}
\]

By definition of \( \nu \), for every \( \varepsilon > 0 \) there is a \( k \) such that \( \|a^k\|^{1/k} < \nu + \varepsilon \). Every \( m \in \mathbb{N} \) is of the form \( m = sk + r \) with unique \( s \in \mathbb{N}_0 \) and \( 0 \leq r < k \) (division with remainder). Then

\[
\|a^m\| = \|a^{sk+r}\| \leq \|a^k\|^s \|a\|^r < (\nu + \varepsilon)^s \|a\|^r,
\]

\[
\|a^m\|^{1/m} \leq (\nu + \varepsilon)^{sk} \|a\|^{r/ sk + r}.
\]

Now \( m \to \infty \) means \( \frac{sk}{sk+r} \to 1 \) and \( \frac{r}{sk+r} \to 0 \), so that \( \limsup_{m \to \infty} \|a^m\|^{1/m} \leq \nu + \varepsilon \). Since this holds for every \( \varepsilon > 0 \), we have \( \limsup_{m \to \infty} \|a^m\|^{1/m} \leq \inf_{n \in \mathbb{N}} \|a^n\|^{1/n} \). Together with (11.2) this implies that \( \lim_{m \to \infty} \|a^m\|^{1/m} \) exists and equals \( \inf_{n \in \mathbb{N}} \|a^n\|^{1/n} \). (Compare Exercise 11.25.)

(ii) Let \( \mu > \nu \), and choose \( \mu' \) such that \( \nu < \mu' < \mu \). Since \( \|a^n\|^{1/n} \to \nu \) by (i), there is a \( n_0 \) such that \( n \geq n_0 \Rightarrow \|a^n\|^{1/n} < \mu' \). For such \( n \) we have

\[
\left( \frac{a}{\mu} \right)^n = \frac{a^n}{\mu^n} \leq \left( \frac{\mu'}{\mu} \right)^n \xrightarrow{n \to \infty} 0.
\]

This proves the first claim. On the other hand, for all \( n \in \mathbb{N} \) we have \( \|a^n\|^{1/n} \geq \nu \). With \( \nu > 0 \) this implies \( \|(a/\nu)^n\|^{1/n} \geq 1 \forall n \), and therefore \( (a/\nu)^n \not\to 0 \).

(iii) Assume \( a \in \text{Inv}\mathcal{A} \). Then there is \( b \in \mathcal{A} \) such that \( ab = ba = \mathbf{1} \). Then \( 1 = a^n b^n \), thus with Remark 4.28 we have \( 1 \leq \|1\| = \|a^n b^n\| \leq \|a^n\| \|b^n\| \leq \|a^n\| \|b^n\|^\nu \). Taking \( n \)-th roots, we have \( 1 \leq \|a^n\|^{1/n} \|b^n\|^{\nu/n} \), and taking the limit gives the contradiction \( 1 \leq \nu \|b\| = 0 \). Thus if \( \nu = 0 \) then \( a \) is not invertible, so that \( 0 \in \sigma(a) \).
11.13 Lemma Let $\mathcal{A}$ be a unital normed algebra. Then $\text{Inv}\mathcal{A}$ is a topological group (w.r.t. the norm topology).

Proof. It is clear that $\text{Inv}\mathcal{A}$ is a group and that multiplication is continuous, since multiplication $\mathcal{A} \times \mathcal{A} \to \mathcal{A}$ is jointly continuous (Remark 4.28). It remains to show that the inverse map $\sigma : \text{Inv}\mathcal{A} \to \text{Inv}\mathcal{A}, a \mapsto a^{-1}$ is continuous. To this purpose, let $r, r + h \in \text{Inv}\mathcal{A}$ and put $(r + h)^{-1} = r^{-1} + k$. We must show that $\|h\| \to 0$ implies $\|k\| \to 0$. From $1 = (r^{-1} + k)(r + h) = 1 + r^{-1}h + kr + khr$ we obtain $r^{-1}h + kr + khr = 0$. Multiplying this on the right by $r^{-1}$ we have $r^{-1}hr^{-1} + k + rkrh = 0$, thus $k = -r^{-1}hr^{-1} - khr^{-1}$. Therefore $\|k\| \leq \|r^{-1}\|\|h\| + \|k\|\|r^{-1}\|\|r^{-1}\|$, which is equivalent to $\|k\|(1 - \|h\|\|r^{-1}\|) \leq \|r^{-1}\|^2\|h\|$ and, for $\|h\| < \|r^{-1}\|^{-1}$, to

$$\|k\| \leq \frac{\|r^{-1}\|^2\|h\|}{1 - \|h\|\|r^{-1}\|}\|h\|.$$  

From this it is clear that $\|h\| \to 0$ implies $\|k\| \to 0$.

Everything we did so far works over $\mathbb{R}$ and over $\mathbb{C}$. (With the obvious exception of the ONB $\{e_n : x \mapsto e^{inx}\} \subseteq L^2([0, 2\pi], \lambda; \mathbb{C})$ in the discussion of Fourier series. But over $\mathbb{R}$ that can be replaced by $\{\cos nx \mid n \in \mathbb{N}_0\} \cup \{\sin nx \mid n \in \mathbb{N}\}$.) The rest of this section requires $\mathbb{F} = \mathbb{C}$, and the same applies whenever we use Theorem 11.14 or Corollaries 11.17, 11.19.

11.14 Theorem (Beurling 1938-Gelfand 1939) Let $\mathcal{A}$ be a unital normed algebra over $\mathbb{C}$ (not necessarily complete) and $a \in \mathcal{A}$. Then $\sigma(a) \neq \emptyset$, and

$$r(a) \geq \inf_{n \in \mathbb{N}} \|a^n\|^{1/n} = \lim_{n \to \infty} \|a^n\|^{1/n}. \quad (11.3)$$

If $\mathcal{A}$ is complete, equality holds in (11.3), which then is called the spectral radius formula.

Proof. The equality of infimum and limit was Lemma 11.12(i). Once the $\geq$ is proven, combining it with Proposition 11.11(ii) in the complete case gives $r(a) = \lim_{n \to \infty} \|a^n\|^{1/n}$.

For $a \in \mathcal{A}$, define $\nu$ as before. If $\nu = 0$ then $0 \in \sigma(a)$ by Lemma 11.12(iii). Thus $\sigma(a) \neq \emptyset$ and (11.3) is trivially true.

From now on assume $\nu > 0$. Assume that there is no $\lambda \in \sigma(a)$ with $|\lambda| \geq \nu$. This implies that $(a - \lambda 1)^{-1}$ exists for all $|\lambda| \geq \nu$ and depends continuously on $\lambda$ by Lemma 11.13. The same holds (since $|\lambda| \geq \nu > 0$) for the slightly more convenient function

$$\phi : \{\lambda \in \mathbb{C} \mid |\lambda| \geq \nu\} \to \mathcal{A}, \quad \lambda \mapsto \left(\frac{a}{\lambda} - 1\right)^{-1}.$$  

Claim: For all $\lambda$ with $|\lambda| \geq \nu$ and $n \in \mathbb{N}$ we have $(\frac{a}{\lambda})^n - 1 \in \text{Inv}\mathcal{A}$ and

$$\left(\left(\frac{a}{\lambda}\right)^n - 1\right)^{-1} = \frac{1}{n} \sum_{k=1}^{n} \phi(\lambda_k), \quad \text{where} \quad \lambda_k = e^{\frac{2\pi i k}{n}} \lambda. \quad (11.4)$$

Before we give the proof, which is elementary algebra, we show how this implies the theorem.

Pick any $\eta > \nu$. Since the annulus $\Lambda = \{\lambda \in \mathbb{C} \mid \nu \leq |\lambda| \leq \eta\}$ is compact, the continuous map $\phi : \Lambda \to \mathcal{A}$ is uniformly continuous. I.e., for every $\varepsilon > 0$ we can find $\delta > 0$ such that $|\lambda, \lambda' \in \Lambda, |\lambda - \lambda'| < \delta \Rightarrow ||\phi(\lambda) - \phi(\lambda')|| < \varepsilon$. If $\nu < \mu < \nu + \delta$, we have $|\nu_k - \mu_k| = |\nu - \mu| < \delta$

34Israel Moiseevich Gelfand (1913-2009). Outstanding Soviet mathematician. Many important contributions to many areas of mathematics, among which functional analysis and Banach algebras.
and therefore \( \| \phi(\nu_k) - \phi(\mu_k) \| < \varepsilon \) for all \( n \in \mathbb{N} \) and \( k = 1, \ldots, n \). Combining this with (11.4) we have \( \| ((\frac{a}{\mu})^n - 1)^{-1} - ((\frac{a}{\mu})^n - 1)^{-1} \| \leq \frac{1}{n} \sum_{k=1}^{n} \| \phi(\nu_k) - \phi(\mu_k) \| < \varepsilon \quad \forall n \in \mathbb{N} \). Thus:

\[
\forall \varepsilon > 0 \exists \mu > \nu \forall n \in \mathbb{N} : \| ((\frac{a}{\mu})^n - 1)^{-1} - ((\frac{a}{\mu})^n - 1)^{-1} \| < \varepsilon. \tag{11.5}
\]

By Lemma 11.12(ii), \( \mu > \nu \) implies \( (a/\mu)^n \to 0 \) as \( n \to \infty \). With continuity of the inverse map, \( ((a/\mu)^n - 1)^{-1} \to -1 \). Thus for \( n \) large enough we have \( \| ((a/\nu)^n - 1)^{-1} + 1 \| < \varepsilon \), and combining this with (11.5) we have \( \| ((a/\nu)^n - 1)^{-1} + 1 \| < 2\varepsilon \). Since \( \varepsilon > 0 \) was arbitrary, we have \( ((a/\nu)^n - 1)^{-1} \to -1 \) as \( n \to \infty \) and therefore \( (a/\nu)^n \to 0 \). This contradicts the other part of Lemma 11.12(ii), so that our assumption that there is no \( \lambda \in \sigma(a) \) with \( |\lambda| \geq \nu \) is false. Existence of such a \( \lambda \) obviously gives \( \sigma(a) \neq \emptyset \) and \( r(a) \geq \nu \), completing the proof. We emphasize that completeness of \( A \) was not needed!

It remains to prove (11.4): For \( 0 \neq \lambda \in \mathbb{C} \) and \( n \in \mathbb{N} \), put \( \lambda_k = \lambda e^{\frac{2\pi i k}{n}} \), where \( k = 1, \ldots, n \). (One should really write \( \lambda_{n,k} \), but we suppress the \( n \).) Then \( \lambda_1, \ldots, \lambda_n \) are the solutions of \( z^n = \lambda^n \), and we have \( z^n - \lambda^n = \prod_k (z - \lambda_k) \). This is an identity in \( \mathbb{C}[z] \), thus is also holds in every unital \( \mathbb{C} \)-algebra \( A \) with \( z \) replaced by \( a \in A \). Let \( |\lambda| \geq \nu > 0 \) and \( n \in \mathbb{N} \). Then our assumption \( |\lambda| \geq \nu \implies \lambda \not\in \sigma(a) \) implies \( \lambda_k \not\in \sigma(a) \) for all \( k = 1, \ldots, n \). Thus all \( a - \lambda_k 1 \) are invertible, and so is \( a^n - \lambda^n 1 = \prod_k (a - \lambda_k 1) \). Thus also \( (\frac{a}{\lambda})^n - 1 \in \text{Inv}_A \), our first claim.

Putting \( z = a/\lambda_k \) in (11.1) and observing \( \lambda_k^n = \lambda^n \), we have

\[
\left( \frac{a}{\lambda_k} \right)^n - 1 = \left( \frac{a}{\lambda_k} \right)^n - 1 = \left( \frac{a}{\lambda_k} - 1 \right) \left( 1 + \frac{a}{\lambda_k} + \cdots + \left( \frac{a}{\lambda_k} \right)^{n-1} \right).
\]

Using the invertibility of \( \frac{a}{\lambda_k} - 1 \forall k \) and \( (\frac{a}{\lambda})^n - 1 \), we can rewrite this as

\[
\phi(\lambda_k) = (\frac{a}{\lambda_k} - 1)^{-1} = (\frac{a}{\lambda} - 1)^{-1} \left( 1 + \frac{a}{\lambda_k} + \cdots + \left( \frac{a}{\lambda_k} \right)^{n-1} \right) = (\frac{a}{\lambda} - 1)^{-1} \sum_{l=0}^{n-1} \left( \frac{a}{\lambda} \right)^l e^{-\frac{2\pi i kl}{n}}.
\]

Summing over \( k \in \{1, \ldots, n\} \), we obtain

\[
\sum_{k=1}^{n} \phi(\lambda_k) = (\frac{a}{\lambda} - 1)^{-1} \sum_{l=0}^{n-1} \left( \frac{a}{\lambda} \right)^l \sum_{k=1}^{n} e^{-\frac{2\pi i kl}{n}}.
\tag{11.6}
\]

If \( l \in \{1, \ldots, n - 1\} \) then \( z = e^{-\frac{2\pi i l}{n}} \) satisfies \( z 
eq 1 \) and \( z^n = 1 \), so that (11.1) gives

\[
\sum_{k=1}^{n} e^{-\frac{2\pi i kl}{n}} = z \sum_{k=0}^{n-1} z^k = z \frac{z^n - 1}{z - 1} = 0.
\]

Thus only \( l = 0 \) contributes to (11.6), and the r.h.s. equals \( n((\frac{a}{\lambda})^n - 1)^{-1} \), yielding (11.4). \( \blacksquare \)

11.15 Remark 1. The standard proof of the above theorem, see e.g. [49, p. 7-10], strengthens Lemma 11.13 to differentiability of the map \( \text{Inv}_A \to \text{Inv}_A, a \mapsto a^{-1} \), which implies differentiability of the 'resolvent map' \( R : \mathbb{C} \setminus \sigma(a) \to A, \lambda \mapsto (a - \lambda 1)^{-1} \). Now one appeals to a certain amount of complex analysis. The latter, however, is inessential to the problem, as shown by the above much more elementary (which does not mean simple) proof due to Rickart\textsuperscript{35} (1958).

2. Even though we avoided complex analysis (holomorphicity etc.), it is clear that the proof only works over \( \mathbb{C} \). In fact, \( \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \in M_{2 \times 2}(\mathbb{R}) \) has empty spectrum over \( \mathbb{R} \). \( \square \)

\textsuperscript{35}Charles Earl Rickart (1913-2002). American mathematician, mostly operator algebraist.
11.16 Corollary (‘Fundamental Theorem of Algebra’) Let \( P \in \mathbb{C}[z] \) be a polynomial of degree \( d \geq 1 \). Then there is \( \lambda \in \mathbb{C} \) with \( P(\lambda) = 0 \).

Proof. We may assume that \( P \) is monic, i.e. the coefficient of the highest power \( z^d \) is 1. It is not hard to construct a matrix \( a_P \in M_{d \times d}(\mathbb{C}) \) such that \( P(\lambda) = \det(a_P - \lambda I) \) (do it!). Now Theorem 11.14 gives \( \sigma(a_P) \neq \emptyset \), and for every \( \lambda \in \sigma(a_P) \) we have \( P(\lambda) = 0 \). \( \blacksquare \)

The above is not a joke! This proof is certainly more elementary than those using complex analysis (Liouville’s theorem) or topological arguments based on \( \pi_1(S^1) \neq 0 \). And the ‘standard’ proof using compactness and \( n \)-th roots of complex numbers, cf. e.g. [46, Theorem 7.7.57], has more than a little in common with the above argument.

11.17 Corollary (Gelfand-Mazur) \(^{36}\)

(i) Every unital normed algebra over \( \mathbb{C} \) other than \( \mathbb{C}1 \) has non-zero non-invertible elements.

(ii) If \( \mathcal{A} \) is a normed division algebra (i.e. unital with \( \text{Inv} \mathcal{A} = \mathcal{A}\setminus\{0\} \)) over \( \mathbb{C} \) then \( \mathcal{A} = \mathbb{C}1 \).

Proof. (i) Let \( a \in \mathcal{A}\setminus\mathbb{C}1 \). By Theorem 11.14 we can pick \( \lambda \in \sigma(a) \). Then \( a - \lambda 1 \) is non-zero and non-invertible. Now (ii) is immediate. \( \blacksquare \)

11.18 Remark 1. That there are no finite dimensional division algebras over \( \mathbb{C} \) is an easy consequence of algebraic closedness. (Why?) There are infinite dimensional ones (like the field of rational functions over \( \mathbb{C} \)), but they do not admit norms by the above corollary, which does not assume finite dimensionality of \( \mathcal{A} \).

2. Over \( \mathbb{R} \) a theorem of Hurwitz\(^{37}\) says that there are precisely four division algebras admitting a norm, namely \( \mathbb{R}, \mathbb{C}, \mathbb{H} \) (Hamilton’s quaternions, which everyone should know) and \( \mathbb{O} \), the octonions of Graves\(^{38}\). But of these only \( \mathbb{C} \) is an algebra over \( \mathbb{C} \). For more on the fascinating subject of real division algebras see the 120 pages on the subject in [17]. \( \square \)

11.19 Corollary Let \( \mathcal{A} \) be a unital Banach algebra and \( \mathcal{B} \subseteq \mathcal{A} \) a closed subalgebra containing \( 1 \). Then \( \sigma(\mathcal{A}) b \subseteq \sigma(\mathcal{B}) b \) and \( r(\mathcal{A}) b = r(\mathcal{B}) b \) for all \( b \in \mathcal{B} \).

Proof. If \( b - \lambda 1 \) has an inverse in \( \mathcal{B} \) then the latter also is an inverse in \( \mathcal{A} \). Thus \( \lambda \not\in \sigma(\mathcal{B}) b \Rightarrow \lambda \not\in \sigma(\mathcal{A}) b \), whence the first claim.

Since the norm of \( \mathcal{B} \) is the restriction to \( \mathcal{B} \) of the norm of \( \mathcal{A} \), the spectral radius formula gives \( r(\mathcal{B}) b = \lim_{n \to \infty} \|b^n\|^{1/n} = r(\mathcal{A}) b \). \( \blacksquare \)

In the situation of Corollary 11.19, \( \sigma(\mathcal{A}) b \not\subseteq \sigma(\mathcal{B}) b \) is possible:

11.20 Example Consider the Banach space \( \mathcal{A} = \ell^1(\mathbb{Z}, \mathbb{C}) \) with norm \( \| \cdot \|_1 \). For \( f, g \in \mathcal{A} \), define the convolution product \( (f * g)(n) = \sum_{m \in \mathbb{Z}} f(m)g(n-m) = \sum_{r + s = n} f(r)g(s) \). Then

\[
\|f * g\|_1 = \sum_{n \in \mathbb{Z}} \left| \sum_{m \in \mathbb{Z}} f(m)g(n-m) \right| \leq \sum_{n \in \mathbb{Z}} \sum_{m \in \mathbb{Z}} |f(m)g(n-m)| = \|f\|_1 \|g\|_1,
\]


\(^{37}\)Adolf Hurwitz (1859-1919). German mathematician who worked on many subjects.

\(^{38}\)Sir William Rowan Hamilton (1805-1865). Irish mathematician. Known particularly for quaternions and Hamiltonian mechanics. It was he who advocated the modern view of complex numbers as pairs of real numbers.

thus $f \ast g \in \mathcal{A}$. It is clear that $\ast$ is bilinear with $1 = \delta_0$ as unit, and associativity is easy to check. Thus $(\mathcal{A}, \| \cdot \|_1, \ast, 1)$ is a unital Banach algebra. The functions $\delta_n(m) = \delta_{n,m}$ satisfy $\delta_n \ast \delta_m = \delta_{n+m}$. In particular $\delta^{-1}_n = \delta_{-n}$ for each $n \in \mathbb{Z}$. 

Let $\mathcal{B} = \{ f \in \mathcal{A} \mid f(n) = 0 \ \forall n < 0 \} = \text{span}_{\mathbb{C}} \{ \delta_n \mid n \geq 0 \} \subseteq \mathcal{A}$. It is immediate that $\mathcal{B}$ is a closed subalgebra containing $1$. Now $b = \delta_1 \in \mathcal{B}$ has an inverse in $\mathcal{A}$, namely $\delta_{-1}$, but not in $\mathcal{B}$: If there was an inverse in $\mathcal{B}$, it would also be an inverse in $\mathcal{A}$ and would have to equal $\delta_{-1} \notin \mathcal{B}$, which is a contradiction. Thus $0 \notin \sigma_{\mathcal{B}}(b)$ and $0 \notin \sigma_{\mathcal{A}}(b)$, so that $\sigma_{\mathcal{A}}(b) \subseteq \sigma_{\mathcal{B}}(b)$.

(The Banach algebra $\ell^1(\mathbb{Z}, \mathbb{C})$ has other interesting applications, cf. Section 18.2. Its construction generalizes to all discrete groups, and in fact to all locally compact groups if one replaces summation by integration w.r.t. the Haar measure $\mu$, obtaining $L^1(G, \mu; \mathbb{C})$.)

11.21 Exercise Let $\mathcal{A}$ be a unital normed algebra.

(i) If $a,b \in \mathcal{A}$ and $ab = ba \in \text{Inv}\mathcal{A}$ with $c = (ab)^{-1}$, prove that $a,b \in \text{Inv}\mathcal{A}$ and $a^{-1} = cb = bc$, $b^{-1} = ca = ac$.

(ii) Give an example of a unital algebra $\mathcal{A}$ and $a,b \in \mathcal{A}$ such that $ab \neq ba \in \text{Inv}\mathcal{A}$ with $a,b \notin \text{Inv}\mathcal{A}$.

(iii) If $a,b \in \mathcal{A}$ and $1 - ab \in \text{Inv}\mathcal{A}$, prove $1 - ba \in \text{Inv}\mathcal{A}$.

Hint: Assuming that $\mathcal{A}$ is Banach and $\|a\|\|b\| < 1$, find a formula for $(1 - ba)^{-1}$ in terms of $(1 - ab)^{-1}$. Now prove that the latter holds without the mentioned assumptions.

(iv) Deduce that $\sigma(ab) \cup \{0\} = \sigma(ba) \cup \{0\}$ and $r(ab) = r(ba)$.

11.22 Exercise Let $\mathcal{A}$ be unital Banach algebra and $a \in \text{Inv}\mathcal{A}$. Prove:

(i) $\sigma(a^{-1}) = \{ \lambda^{-1} \mid \lambda \in \sigma(a) \}$.

(ii) If $\|a\| \leq 1$ and $\|a^{-1}\| \leq 1$ then $\sigma(a) \subseteq S^1 = \{ z \in \mathbb{C} \mid |z| = 1 \}$.

11.23 Exercise Let $\mathcal{A}$ be a unital Banach algebra over $\mathbb{C}$ and $a,b \in \mathcal{A}$ commuting elements. Prove $r(ab) \leq r(a)r(b)$. (I.e. $r$ is submultiplicative. We will soon prove subadditivity.)

11.24 Exercise Let $\mathcal{A}$ be a unital Banach algebra and $\mathcal{B} \subseteq \mathcal{A}$ a maximal abelian Banach subalgebra with $1 \in \mathcal{B}$. (Maximality means that we cannot have $\mathcal{B} \subseteq \mathcal{C} \subseteq \mathcal{A}$ with $\mathcal{C}$ commutative Banach.) Prove $\text{Inv}\mathcal{B} = \mathcal{B} \cap \text{Inv}\mathcal{A}$ and conclude that $\sigma_{\mathcal{B}}(b) = \sigma_{\mathcal{A}}(b)$ $\forall b \in \mathcal{B}$.

11.25 Exercise Let $\mathcal{H} = \ell^2(\mathbb{N}, \mathbb{C})$ and define $A \in B(\mathcal{H})$ by $Ae_k = \alpha_k e_{k+1}$, where $\alpha_k = 2$ for odd $k$ and $\alpha_k = 1/2$ for even $k$. Compute $\|A^n\|$ for all $n$ and show that $n \mapsto \|A^n\|^{1/n}$ is not monotonously decreasing. (This $A$ is an example of a weighted shift operator, generalizing $R$.)

Since we need a unit in order to define $\sigma(a)$, the following construction is quite important (but we won’t use it):

11.26 Exercise (Unitization of Banach algebras) Let $\mathcal{A}$ be a Banach algebra over $\mathbb{F}$, possibly without unit. Define $\tilde{\mathcal{A}} = \mathcal{A} \oplus \mathbb{F}$, which is an $\mathbb{F}$-vector space in the obvious way. For $(a, \alpha), (b, \beta) \in \tilde{\mathcal{A}}$ define $\|(a, \alpha)\|\tilde{\mathcal{A}} = \|a\| + |\alpha|$ and $(a, \alpha)(b, \beta) = (ab + \alpha b + \beta a, \alpha \beta)$. Prove:

(i) $\tilde{\mathcal{A}}$ is an associative algebra with unit $(0,1)$.

(ii) $(\tilde{\mathcal{A}}, \|\cdot\|\tilde{\mathcal{A}})$ is a Banach algebra.

(iii) The map $\iota : \mathcal{A} \rightarrow \tilde{\mathcal{A}}, a \mapsto (a,0)$ is an isometric algebra homomorphism, and $\iota(\mathcal{A}) \subseteq \tilde{\mathcal{A}}$ is a two-sided ideal.
11.3 Examples of spectra of operators

We summarize what we have proven: If $A \in B(E)$ then $\sigma(A) \subseteq \mathbb{C}$ is closed, non-empty, and bounded by $r(A) = \lim_{n \to \infty} \|A^n\|^{1/n} \leq \|A\|$. Note that we do not yet have good general results on $\sigma_p, \sigma_c, \sigma_r$. We will later prove some, for special classes of operators.

11.27 Exercise Compute $\sigma_p(L)$ and $\sigma_p(R)$ for the shift operators $L, R$ on $\ell^p(\mathbb{N}, \mathbb{C})$ for all $p \in [1, \infty]$. (Of course, the $p$ in $\sigma_p$ has nothing to do with the $p$ in $\ell^p$.)

11.28 Exercise Let $H = \ell^2(\mathbb{N}, \mathbb{C})$ and define $A \in B(H)$ by $(Af)(n) = f(n)/n$. Determine $\sigma_p(A), \sigma_c(A), \sigma_r(A)$.

11.29 Exercise (Assuming some measure theory) Let $H = L^2([a, b])$, where $-\infty < a < b < \infty$, and define $A \in B(H)$ by $(Af)(x) = xf(x)$. Prove $\sigma_c(A) = [a, b]$ and $\sigma_p(A) = \sigma_r(A) = \emptyset$.

11.30 Exercise Prove that for every compact set $C \subseteq \mathbb{C}$ there is an operator $A \in B(H)$, where $H$ is a separable Hilbert space, such that $\sigma(A) = C$.

Hint: Prove and use that $C$ has a countable dense subset.

11.31 Exercise (Discrete spectrum) Let $V$ be a Banach space and $A \in B(V)$. For $\lambda \in \mathbb{C}$ define the generalized eigenspace by $L_\lambda(A) = \bigcup_{n \in \mathbb{N}} \ker(A - \lambda 1)^n$.

(i) Prove that the following are equivalent for $\lambda \in \sigma(A)$:

(a) $L_\lambda(A)$ is at most finite dimensional and has a complementary subspace $W \subseteq V$ such that $(A - \lambda 1)W \subseteq W$ and $(A - \lambda 1) \upharpoonright W$ is invertible. ($W$ is automatically unique.]

(b) $L_\lambda(A)$ is at most finite dimensional, $A - \lambda 1$ has closed image, and $\lambda$ is an isolated point in $\sigma(A)$ (i.e. there is $\varepsilon > 0$ such that $\sigma(A) \cap B(\lambda, \varepsilon) = \{\lambda\}$).

(c) $\lambda$ is isolated and $A - \lambda 1$ is Fredholm, i.e. the image $(A - \lambda 1)V$ is closed and $\ker(A - \lambda 1)$ and $V/(A - \lambda 1)V$ are finite dimensional.

(ii) Let $\sigma_d(A)$ be the set of $\lambda \in \sigma(A)$ for which the above equivalent statements hold. Prove that $\sigma_d(A) \subseteq \sigma_p(A)$. (The $\lambda \in \sigma_d(A)$ are called normal eigenvalues and $\sigma_d(A)$ the discrete spectrum.)

Hint: You may want to use Lemma 11.8. For the proof of (ii)⇒(i) one needs the following fact that we can’t prove here: If $\lambda \in \sigma(A)$ is isolated then there is a non-zero idempotent $P \in B(V)$ such that $PA = AP, (A - \lambda 1)P = 0$ and $(A - \lambda 1)(1 - P)V$ is injective. (In Proposition 14.10 we prove this for normal operators on Hilbert spaces.)

11.4 Characters. Spectrum of a Banach algebra

We now develop a new perspective on the spectrum that will prove very powerful, allowing to obtain results that would be hard to reach in other ways. For example: If $A$ is a unital Banach algebra and $a, b \in A$. What can we say about $\sigma(a + b)$ or $\sigma(ab)$? Using only the definition of the spectrum this seems quite difficult. In this section we require $\|1\| = 1$.

11.32 Definition If $A, B$ are $\mathbb{F}$-algebras, an (algebra) homomorphism $\alpha : A \to B$ is a linear map such that also $\alpha(aa') = \alpha(a)\alpha(a') \forall a, a' \in A$. If $A, B$ are unital, $\alpha$ is called unital if $\alpha(1_A) = 1_B$. Algebra homomorphisms from an $\mathbb{F}$-algebra to $\mathbb{F}$ are called characters.

11.33 Lemma If $A, B$ are unital algebras and $\alpha : A \to B$ is a unital algebra homomorphism then $\sigma_B(\alpha(a)) \subseteq \sigma_A(a)$. 

68
Proof. If \( \lambda \not\in \sigma_A(a) \) then \( a - \lambda 1_A \in A \) has an inverse \( b \). Then \( \alpha(b) \) is an inverse for \( \alpha(a) - \lambda 1_B \in B \), thus \( \lambda \not\in \sigma_B(\alpha(a)) \).  

\[\text{11.34 Lemma} \quad \text{Let } A \text{ be a unital Banach algebra. Then every non-zero character } \varphi : A \to F \text{ satisfies } \varphi(1) = 1, \varphi(a) \in \sigma(a) \quad \forall a \in A \text{ and } \|\varphi\| = 1, \text{ thus } \varphi \text{ is continuous.} \]

\[\text{Proof. If } \varphi(1) = 0 \text{ then } \varphi(a) = \varphi(a1) = \varphi(a)\varphi(1) = 0 \text{ for all } a \in A, \text{ thus } \varphi = 0. \text{ Thus } \varphi \neq 0 \Rightarrow \varphi(1) \neq 0. \text{ Now } \varphi(1) = \varphi(1^2) = \varphi(1)^2 \text{ implies } \varphi(1) = 1. \]

We have just proven that every non-zero character is a unital homomorphism. Thus by Lemma 11.33, \( \sigma(\varphi(a)) \subseteq \sigma(a) \). Since the spectrum of \( z \in F \) clearly is \( \{z\} \), this means \( \varphi(a) \in \sigma(a) \), thus \( |\varphi(a)| \leq r(a) \leq \|a\| \) by Proposition 11.11, whence \( \|\varphi\| \leq 1. \) Since we require \( \|1\| = 1 \), we also have \( \|\varphi\| \geq |\varphi(1)|/\|1\| = 1. \)

\[\text{11.35 Definition} \quad \text{If } A \text{ is a unital Banach algebra, the spectrum } \Omega(A) \text{ of } A \text{ is the set of non-zero characters } \varphi : A \to F. \]

\[\text{11.36 Exercise} \quad \text{Let } X \text{ be a compact Hausdorff space and } A = C(X,F). \text{ For every } x \in X \text{ define } \varphi_x : A \to F, f \mapsto f(x). \text{ Prove:} \]

(i) \( \varphi_x \) is a non-zero character of \( A \), thus \( \varphi_x \in \Omega(A) \), for each \( x \in X \).

(ii) The map \( \iota : X \to \Omega(A), \ x \mapsto \varphi_x \) is injective.

(iii) For each \( f \in A \) we have \( \sigma(f) = \{\varphi(f) \mid \varphi \in \Omega(A)\} \).

(Later we will define a topology on \( \Omega(A) \) and see that \( \iota \) is a homeomorphism.)

One could hope that \( \sigma(a) = \{\varphi(a) \mid \varphi \in \Omega(A)\} \) holds for every unital Banach algebra \( A \) and \( a \in A \). But this is too much to ask since a non-commutative algebra \( A \) may well have \( \Omega(A) = \emptyset. \) E.g., this holds for all matrix algebras \( M_{n \times n}(F), \ n \geq 2 \) since these are simple (no proper two-sided ideals) so that a homomorphism to another algebra \( B \) must be zero or injective, the latter being impossible for \( B = F \) for dimensional reasons.

\[\text{11.37 Proposition} \quad \text{Let } A \text{ be a commutative unital Banach algebra over } \mathbb{C}. \text{ Then} \]

(i) If \( \varphi \in \Omega(A) \) then \( \ker \varphi \subseteq A \) is a maximal ideal (i.e. not contained in a larger proper ideal).

(ii) Every maximal ideal in \( A \) is the kernel of a unique \( \varphi \in \Omega(A) \). In particular, \( \Omega(A) \neq \emptyset. \)

(iii) For each \( a \in A \) we have

\[\sigma(a) = \{\varphi(a) \mid \varphi \in \Omega(A)\}. \quad \text{(11.7)}\]

Proof. (i) Every \( \varphi \in \Omega(A) \) is continuous, thus \( M = \ker \varphi \) is a closed ideal. We have \( M \neq A \) since \( \varphi \neq 0 \). This ideal has codimension one since \( A/M \cong \mathbb{C} \) and therefore is maximal.

(ii) Now let \( M \subseteq A \) be a maximal ideal. Since maximal ideals are proper, no element of \( M \) is invertible. For each \( b \in M \) we have \( \|1 - b\| \geq 1 \) since otherwise \( b = 1 - (1 - b) \) would be invertible by Lemma 11.10(i). (This is the only place where completeness is used.) Thus \( 1 \notin M \), so that \( M \) is a proper ideal containing \( M \). Since \( M \) is maximal, we have \( M = M \), thus \( M \) is closed. Now by Proposition 7.2(vi), \( A/M \) is a normed algebra, and by a well-known algebraic argument the maximality of \( M \) implies that \( A/M \) is a division algebra. Thus \( A/M \cong \mathbb{C} \) by Gelfand-Mazur (Corollary 11.17, which holds only over \( \mathbb{C} \)), so that there is a unique isomorphism \( \alpha : A/M \to \mathbb{C} \) sending \( 1 \in A/M \) to \( 1 \in \mathbb{C} \). If \( p : A \to A/M \) is the quotient homomorphism then \( \varphi = \alpha \circ p : A \to \mathbb{C} \) is a non-zero character with \( \ker \varphi = M \). This \( \varphi \) clearly is unique. The
This defines a bilinear map (iii) of Proposition 11.37 (with 11.39 Exercise Hint: For (iii), use an abelian subalgebra, for (iv) a maximal one.

12 Transpose and adjoint of bounded operators. \( C^* \)-algebras

12.1 The transpose of a bounded Banach space operator

Let \( E, F \) be normed spaces, \( A \in B(E, F) \) and \( \varphi \in F^* = B(F, \mathbb{F}) \). Then \( \varphi \circ A \in B(E, \mathbb{F}) = E^* \). This defines a bilinear map \( B(E, F) \times F^* \to E^* \), and keeping \( A \) fixed, we have a linear map \( A^t : F^* \to E^* \), \( \varphi \mapsto \varphi \circ A \), which is called the transpose (or adjoint) of \( A \). We will stick to ‘transpose’ to avoid confusion with the Hilbert space adjoint. Note that the transpose goes in the ‘opposite direction’! In fact, if \( E, F, G \) be normed spaces and \( A \in B(E, F), B \in B(F, G) \) then \( (B \circ A)^t = A^t \circ B^t \) in \( B(G^*, E^*) \).

12.1 Lemma The linear map \( B(E, F) \to B(F^*, E^*) \) is isometric, i.e. \( \|A^t\| = \|A\| \).

Proof. By Proposition 8.7(i), for \( f \in F \) we have \( \|f\| = \sup_{\varphi \in F^*, \|\varphi\| = 1} |\varphi(f)| \). Thus

\[
\|A\| = \sup_{\|e\| = 1} \|Ae\| = \sup_{\|e\| = 1} \sup_{\|\varphi\| = 1} |\varphi(Ae)| = \sup_{\varphi \in F^*, \|\varphi\| = 1} \sup_{\|e\| = 1} |\varphi(Ae)| = \sup_{\|\varphi\| = 1} \|A^t \varphi\| = \|A^t\|.
\]

The transposition operation can be iterated, giving \( A^{tt} \in B(E^{**}, F^{**}) \), etc.
12.2 Lemma Let $E, F$ be Banach spaces and $A \in B(E, F)$. If $\iota_E : E \to E^{**}$ and $\iota_F : F \to F^{**}$ are the canonical inclusions, then $A^{tt} \circ \iota_E = \iota_F \circ A$.

Equivalently, $A^{tt} : E^{**} \to F^{**}$ maps $E \subseteq E^{**}$ to $F \subseteq F^{**}$ and $A^{tt} \mid E = A$.

Proof. Let $x \in E, \varphi \in F^*$. Then using the definition of $\iota_E, \iota_F$ and of the transpose, we have

$$
\iota_F(Ax)(\varphi) = \varphi(Ax) = (A^t\varphi)(x) = \iota_E(x)(A^t\varphi) = (A^{tt}\iota_E(x))(\varphi).
$$

Now, $\iota_F(Ax)$ and $(A^{tt}\iota_E(x))$ are in $F^{**}$, and the fact that they coincide on all $\varphi \in F^*$ means $\iota_F(Ax) = A^{tt}\iota_E(x)$. And since this holds for all $x \in E$, we have $\iota_F A = A^{tt}\iota_E$, as claimed. ■

12.3 Exercise Let $E, F$ be Banach spaces and $A \in B(E, F)$. Prove:

(i) Let $E, F$ be Banach spaces, $A \in B(E, F)$. Prove $\ker A^t = (AE)^\perp$.

(ii) If $A \in B(E, F)$ is invertible then $A^t \in B(F^*, E^*)$ is invertible.

(iii) If $A^t \in B(F^*, E^*)$ then $A \in B(E, F)$ is invertible. (Warning: We don’t assume reflexivity of the spaces involved!)

(iv) $\sigma(A) = \sigma(A^t)$ for each $A \in B(E)$.

Hint: (i),(ii),(iv) are very easy. The proof of (iii) uses (i) and (ii).

12.2 The adjoint of a bounded Hilbert space operator

In the rest of this section we will study aspects of bounded (linear) operators that are specific to bounded operators between Hilbert spaces, as well as the closely related $C^*$-algebras.

If $H$ is a Hilbert space then we have a canonical map $\gamma_H : H \to H^*$ given by $y \mapsto \varphi_y = (\cdot, y)$. This map is antilinear, and by Riesz’ representation theorem it is a bijection. This bijection in a sense makes the dual spaces of Hilbert spaces redundant to a large extent, so that it is desirable to eliminate them from considerations of the transpose:

12.4 Proposition Let $H_1, H_2$ be Hilbert spaces. For $A \in B(H_1, H_2)$, define $A^* := \gamma_{H_1}^{-1} \circ A^t \circ \gamma_{H_2} : H_2 \to H_1$.

(i) The map $A^* : H_2 \to H_1$ is linear.

(ii) The map $B(H_1, H_2) \to B(H_2, H_1)$, $A \mapsto A^*$ is anti-linear.

(iii) For all $x \in H_1, y \in H_2$ we have $\langle Ax, y \rangle_2 = \langle x, A^*y \rangle_1$.

Proof. (i) Linearity of $A^* : H_2 \to H_1$ follows from its being the composite of the linear map $A^t$ with the two anti-linear maps $\gamma_{H_2}$ and $\gamma_{H_1}^{-1}$.

(ii) Additivity of $A \mapsto A^*$ is obvious. Let $A \in B(H_1, H_2), c \in \mathbb{C}, x \in H_2$. Then

$$(cA)^*(x) = \gamma_{H_1}^{-1} \circ (cA)^t \circ \gamma_{H_2}(x) = \gamma_{H_1}^{-1}(cA^t(\gamma_{H_2}(x))) = \overline{c}\gamma_{H_1}^{-1}(A^t(\gamma_{H_2}(x))) = \overline{c}A^*(x),$$

where we used the linearity of $A \mapsto A^t$ and anti-linearity of $\gamma_{H_1}^{-1}$, shows $(cA)^* = \overline{c}A^*$. (The anti-linearity of $\gamma_{H_2}$ is irrelevant here.)

(iii) If $y \in H_2$ then $\gamma_{H_2}(y) \in H_2^*$ is the functional $(\cdot, y)_2$. Then $(A^t \circ \gamma_{H_2})(y) \in H_1^*$ is the functional $x \mapsto (Ax, y)_2$. Thus $z = A^*y = (\gamma_{H_1}^{-1} \circ A^t \circ \gamma_{H_2})(y) \in H_1$ is a vector such that $\langle x, z \rangle_1 = \langle Ax, y \rangle_2$ for all $x \in H_1$. This means $\langle x, A^*y \rangle_1 = \langle Ax, y \rangle_2 \forall x \in H_1, y \in H_2$, as claimed. ■
There is a useful bijection between bounded operators and bounded sesquilinear forms. It can be used to give an alternative (at least in appearance) construction of the adjoint $A^\ast$ (and for many other purposes). It is based on the following observation: If $A \in B(H)$ satisfies $\langle Ax, y \rangle = 0$ for all $x, y \in H$ then $Ax = 0$ for all $x$, thus $A = 0$. Applying this to $A - B$ shows that $\langle Ax, y \rangle = \langle Bx, y \rangle \forall x, y$ implies $A = B$. Thus bounded operators are determined by their ‘matrix elements’. This motivates the following developments.

12.5 Definition Let $V$ be an $F$-vector space. A map $V \times V \to V, (x, y) \mapsto [x, y]$ is called sesquilinear if it is linear w.r.t. $x$ and anti-linear w.r.t. $y$. A sesquilinear form $[\cdot, \cdot]$ is bounded if $\sup_{\|x\|=\|y\|=1} |[x, y]| < \infty$.

12.6 Remark Recall that the inner product $\langle \cdot, \cdot \rangle$ on a (pre)Hilbert space is sesquilinear and bounded by Cauchy-Schwarz. If $F = \mathbb{R}$, the definition of course reduces to bilinearity. \qed

12.7 Proposition Let $H$ be a Hilbert space. Then there is a bijection between $B(H)$ and the set of bounded sesquilinear forms on $H$, given by $B(H) \ni A \mapsto [\cdot, \cdot]_A$, where $[x, y]_A = \langle Ax, y \rangle$.

Proof. Let $A \in B(H)$. Sesquilinearity of $[\cdot, \cdot]_A = \langle Ax, y \rangle$ is an obvious consequence of sesquilinearity of $\langle \cdot, \cdot \rangle$ and linearity of $A$, and boundedness follows from Cauchy-Schwarz:

$$[x, y]_A = \langle Ax, y \rangle \leq \|Ax\|\|y\| \leq \|A\|\|x\|\|y\| \forall x, y.$$ 

Now let $[\cdot, \cdot]$ be a sesquilinear form bounded by $M$. Then for each $x \in H$, the map $\psi_x : H \to \mathbb{C}$, $y \mapsto [x, y]$ is linear (thanks to the complex conjugation) and satisfies $|\psi_x(y)| \leq M\|y\|\|x\|$, thus $\psi_x \in H^\ast$. Thus by Theorem 6.29 there is a unique vector $z_x \in H$ such that $\psi_x = \varphi_{z_x}$, thus $[x, y] = \psi_x(y) = \varphi_{z_x}(y) = \langle y, z_x \rangle \forall y$ and, taking complex conjugates, $\langle z_x, y \rangle = [x, y] \forall y$. Thus defining $A : H \to H$ by $Ax = z_x \forall x$ we have $\langle Ax, y \rangle = [x, y] \forall x, y$. Since the maps $x \mapsto \psi_x$ and $\psi_x \mapsto z_x$ are both anti-linear, their composite $A$ is linear. And since $H \to H^\ast$, $z \mapsto \varphi_z$ is an isometry, we have $\|Ax\| = \|z_x\| = \|\varphi_{z_x}\| = \|\psi_x\| \leq M\|x\|$, thus $A \in B(H)$. \qed

12.8 Proposition Let $H$ be a Hilbert space and $A \in B(H)$. Then there is a unique $B \in B(H)$ such that

$$\langle Ax, y \rangle = \langle x, By \rangle \forall x, y \in H.$$ 

This $B$ is denoted $A^\ast$ and called the adjoint of $A$.

Proof. The map $(y, x) \mapsto \langle y, Ax \rangle$ is sesquilinear and bounded (by $\|A\|$). Thus by Proposition 12.7 there is a bounded $B \in B(H)$ such that $\langle By, x \rangle = \langle y, Ax \rangle \forall x, y$. Taking the complex conjugate gives $\langle x, By \rangle = \langle By, x \rangle = \langle y, Ax \rangle = \langle Ax, y \rangle$, which is the wanted identity. \qed

12.9 Remark 1. In view of the identity $\langle Ax, y \rangle = \langle x, A^\ast y \rangle$ satisfied by the adjoint as defined above and Proposition 12.4(iii) (with $H_1 = H_2 = H$), it is clear that the two constructions of $A^\ast$ give the same result (and in a sense are the same construction since both use Theorem 6.29).

2. It is obvious that $A \in B(H)$ is self-adjoint as defined earlier ($\langle Ax, y \rangle = \langle x, Ay \rangle \forall x, y \in H$) if and only if $A^\ast = A$.

3. If $[\cdot, \cdot]$ is a sesquilinear form then also $[x, y]^\prime := [\overline{y}, x]$ is a sesquilinear form, called the adjoint form. Looking at the above definition of $A^\ast$, one finds that $A^\ast$ is the bounded operator associated with the form $[\cdot, \cdot]^\prime$. Thus self-adjointness of $A$ is equivalent to $[\cdot, \cdot]_A = [\cdot, \cdot]^\ast_A$, i.e. self-adjointness of $[\cdot, \cdot]_A$. (Self-adjoint forms and operators are also called hermitian.) \qed

72
12.10 Lemma The map \( B(H) \rightarrow B(H) \), \( A \rightarrow A^* \) satisfies

(i) \( (cA + dB)^* = \overline{c}A^* + \overline{d}B^* \) \( \forall A, B \in B(H), c, d \in \mathbb{F} \) (antilinearity).

(ii) \( (AB)^* = B^*A^* \) (anti-multiplicativity).

(iii) \( A^{**} = A \) (involutivity).

(iv) \( 1^* = 1 \).

Proof. (i) Follows from

\[
\langle x, (\overline{c}A^* + \overline{d}B^*)y \rangle = c\langle x, A^*y \rangle + d\langle x, B^*y \rangle = c\langle Ax, y \rangle + d\langle Bx, y \rangle = \langle (cA + dB)x, y \rangle = \langle x, (cA + dB)^*y \rangle.
\]

(ii) \( \langle x, (AB)^*y \rangle = \langle (AB)x, y \rangle = \langle Bx, A^*y \rangle = \langle x, B^*A^*y \rangle \).

(iii) Complex conjugating \( \langle Ax, y \rangle = \langle x, A^*y \rangle \) gives

\[
\langle y, Ax \rangle = \langle A^*x, y \rangle = \langle x, A^{**}y \rangle = \langle A^*y, x \rangle,
\]

which shows that \( A \) is an adjoint of \( A^* \). Uniqueness of the adjoint now implies \( A^{**} = A \).

(iv) Obvious. 

12.11 Proposition Let \( H \) be a Hilbert space. Then for all \( A \in B(H) \) we have

(i) \( \|A^*\| = \|A\| \). (The *-operation is isometric.)

(ii) \( \|A^*A\| = \|A\|^2 \). ("\( C^* \)-identity")

Proof. (i) Similarly to Lemma 12.1, using (6.2) we have

\[
\|A^*\| = \sup_{\|x\|=\|y\|=1} |\langle A^*x, y \rangle| = \sup_{\|x\|=\|y\|=1} |\langle x, Ay \rangle| = \sup_{\|x\|=\|y\|=1} |\langle Ay, x \rangle| = \|A\|.
\]

(ii) On the one hand, \( \|A^*A\| \leq \|A^*\|\|A\| = \|A\|^2 \), where we used (i). On the other, using (6.2) we have

\[
\|A^*A\| = \sup_{\|x\|=1} \|A^*Ax\| = \sup_{\|x\|=\|y\|=1} |\langle A^*Ax, y \rangle| = \sup_{\|x\|=\|y\|=1} |\langle Ax, Ay \rangle| \geq \sup_{\|x\|=1} \langle Ax, Ax \rangle = \sup_{\|x\|=1} \|Ax\|^2 = (\sup_{\|x\|=1} \|Ax\|)^2 = \|A\|^2.
\]

The above construction of \( A^* \) for \( A \in B(H) \) can be generalized to bounded linear maps \( A : H_1 \rightarrow H_2 \), so as to give \( A^* : H_2 \rightarrow H_1 \) satisfying

\[
\langle Ax, y \rangle_2 = \langle x, A^*y \rangle_1 \quad \forall x \in H_1, y \in H_2.
\]

We refrain from doing so explicitly since, as in the case \( H_1 = H_2 \), the result would be the same as that of the construction in Proposition 12.4. Instead we take the latter as definition of \( A^* \) in the general case. Now one has:

12.12 Lemma Let \( H_1, H_2 \) be Hilbert spaces and \( A \in B(H_1, H_2) \). Then

(i) \( A \) is an isometry if and only if \( A^*A = \text{id}_{H_1} \).

(ii) \( A \) is unitary if and only if \( A^*A = \text{id}_{H_1} \) and \( AA^* = \text{id}_{H_2} \).
Proof. (i) By definition, $A$ is an isometry if $\langle Ax, Ay \rangle_2 = \langle x, y \rangle_1$ for all $x, y \in H_1$. Since the l.h.s. equals $\langle x, A^*Ay \rangle_1$, $A$ is an isometry if and only if $\langle A^*Ax, y \rangle_1 = \langle x, y \rangle_1$ for all $x, y \in H_1$, which is equivalent to $A^*A = \text{id}_{H_1}$ by the observations at the beginning of the section.

(ii) By definition, a unitary is a surjective isometry, which is equivalent to being an invertible isometry, or an isometry that has an isometry as inverse. Now apply (i). $lacksquare$

12.13 Exercise Consider the left and right shift operators $L, R$ of Definition 11.1 in the Hilbert space $l^2(\mathbb{N}, \mathbb{C})$.

(i) Prove $L^* = R$ and $R^* = L$.

(ii) Prove $\sigma(L) = \sigma(R) = \overline{B}(0, 1)$ (closed unit disk). Hint: Use Exercise 11.27.

12.3 Involutions. Definition of $C^*$-algebras

The properties of the adjoint map $A \mapsto A^*$ on $B(H)$ motivate some definitions:

12.14 Definition Let $\mathcal{A}$ be a $\mathbb{C}$-algebra. A map $*: \mathcal{A} \to \mathcal{A}$ satisfying antilinearity, antimultiplicativity and involutivity ((i)-(iii) in Lemma 12.10) is called an involution or $^*$-operation. An algebra with a chosen $^*$ is called a $^*$-algebra. A $^*$-homomorphism $\alpha: \mathcal{A} \to \mathcal{B}$ of $^*$-algebras is a homomorphism satisfying $\alpha(a^*) = \alpha(a)^*$ $\forall a \in \mathcal{A}$.

If a $^*$-algebra has a unit $1$ we automatically have $1^* = 11^* = 1^*1^* = (11^*)^* = (1^*)^* = 1$.

12.15 Definition If $\mathcal{A}$ is a Banach algebra and $*: \mathcal{A} \to \mathcal{A}$ an involution then $\mathcal{A}$ is called a

- Banach $^*$-algebra if $\|a^*\| = \|a\|$ $\forall a \in \mathcal{A}$.
- $C^*$-algebra if $\|a^*a\| = \|a\|^2$ $\forall a \in \mathcal{A}$.

12.16 Lemma Every $C^*$-algebra is a Banach $^*$-algebra. If it has a unit $1$ then $\|1\| = 1$.

Proof. With the $C^*$-identity and submultiplicativity we have $\|a\|^2 = \|a^*a\| \leq \|a^*\|\|a\|$, thus $\|a\| \leq \|a^*\|$ for all $a \in \mathcal{A}$. Replacing $a$ by $a^*$ herein gives the converse inequality, thus $\|a^*\| = \|a\|$.

If $1$ is a unit then $\|1\|^2 = \|1^*1\| = \|1\| = \|1\|$, and since $\|1\| \neq 0$ this implies $\|1\| = 1$. $lacksquare$

12.17 Remark 1. Clearly $B(H)$ is a $C^*$-algebra for each Hilbert space $H$. Since this holds also for real Hilbert spaces, is shows that one can discuss Banach $^*$-algebras and $C^*$-algebras over $\mathbb{R}$. But we will consider only complex ones.

2. There is no special name for the non-complete variants of the above definitions. But a submultiplicativc norm on a $^*$-algebra satisfying the $C^*$-identity is called a $C^*$-norm, whether $\mathcal{A}$ is complete w.r.t. it or not. Completion of a $^*$-algebra w.r.t. a $C^*$-norm gives a $C^*$-algebra, and this is an important way of constructing new $C^*$-algebras. $lacksquare$

12.18 Exercise Recall the Banach algebra $\mathcal{A} = \ell^1(\mathbb{Z}, \mathbb{C})$ from Example 11.20. Show that both $f^*(n) = \overline{f(n)}$ and $f^*(n) = \overline{f(-n)}$ are involutions on $\mathcal{A}$ making it a Banach $^*$-algebra. Show that neither of them satisfies the $C^*$-identity. (Thus Banach-$^*$ $\not\equiv C^*$.)

12.19 Lemma Let $X$ be a compact space. For $f \in C(X, \mathbb{C})$, define $f^*$ by $f^*(x) = \overline{f(x)}$. Then $C(X, \mathbb{C})$ is a $C^*$-algebra. The same holds for $C_b(X)$, where $X$ is arbitrary, thus also for $\ell^\infty(S, \mathbb{C})$. 

74
Proof. We know that \(C(X, \mathbb{C})\) equipped with the norm \(\|f\| = \sup_x |f(x)|\) is a Banach algebra. It is immediate that \(\ast\) is an involution. The computation
\[
\|f^* f\| = \sup_x |\overline{f(x)} f(x)| = \sup_x |f(x)|^2 = \left(\sup_x |f(x)|\right)^2 = \|f\|^2
\]
proves the \(C^\ast\)-identity. It is clear that this generalizes to the bounded continuous functions on any space \(X\).

In a sense, the examples \(B(H)\) and \(C(X, \mathbb{C})\) for compact \(X\) are all there is: One can prove, as we will do in Theorem 18.11, that every commutative unital \(C^\ast\)-algebra is isometrically \(*\)-isomorphic to \(C(X, \mathbb{C})\) for some compact Hausdorff space \(X\), determined uniquely up to homeomorphism. (For example one has \(\ell^\infty(S, \mathbb{C}) \cong C(\beta S, \mathbb{C}), \) where \(\beta S\) is the Stone-Čech compactification of \((S, \tau_{\text{disc}})\).) And every \(C^\ast\)-algebra is isometrically \(*\)-isomorphic to a norm-closed \(*\)-subalgebra of \(B(H)\) for some Hilbert space \(H\). Cf. e.g. [49].

12.20 Definition Let \(\mathcal{A}\) be a \(\mathbb{C}\)-algebra with an involution \(\ast\). Then \(a \in \mathcal{A}\) is called
- self-adjoint if \(a = a^\ast\).
- unitary if \(aa^\ast = a^\ast a = 1\). (Obviously \(\mathcal{A}\) needs to be unital.)
- normal if \(aa^\ast = a^\ast a\).
- orthogonal projection if \(a^2 = a = a^\ast\).

A subset \(S\) of a \(*\)-algebra is called self-adjoint if \(S = S^\ast := \{s^\ast \mid s \in S\}\).

12.21 Remark 1. If \(\mathcal{A}\) is a \(*\)-algebra and \(a \in \mathcal{A}\), we define \(\text{Re}(a) = \frac{a + a^\ast}{2}, \) \(\text{Im}(a) = \frac{a - a^\ast}{2i}\). Now it is immediate that \(\text{Re}(a), \text{Im}(a)\) are self-adjoint and \(a = \text{Re}(a) + i \text{Im}(a)\). Furthermore, \(a\) is self-adjoint \(\iff \text{Im}(a) = 0\), and \(a\) is normal \(\iff \text{Re}(a)\) and \(\text{Im}(a)\) commute.

2. Obviously all self-adjoint and all unitary elements of a \(C^\ast\)-algebra are normal. On the other hand, non-unitary isometries \((a^\ast a = 1 \neq aa^\ast)\) are not normal. Every element of a commutative \(C^\ast\)-algebra is normal. Conversely, if \(a \in \mathcal{A}\) is normal then the \(C^\ast\)-subalgebras \(C^\ast(a)\) and \(C^\ast(1, a)\) of \(\mathcal{A}\), i.e. the smallest (unital) \(C^\ast\)-subalgebra containing \(a\), are commutative.

Later we will see that normal elements of abstract \(C^\ast\)-algebras and normal operators on Hilbert spaces have very nice properties and closely correspond to continuous functions and multiplication operators, respectively, cf. Sections 16 and 18.

12.4 Spectrum of elements of a \(C^\ast\)-algebra

12.22 Lemma Let \(\mathcal{A}\) be a unital \(C^\ast\)-algebra.

(i) If \(a \in \mathcal{A}\) is invertible then \(a^\ast\) is invertible and \((a^\ast)^{-1} = (a^{-1})^\ast\).

(ii) \(\sigma(a^\ast) = \sigma(a)^\ast := \{\overline{\lambda} \mid \lambda \in \sigma(a)\}\).\(^{40}\)

Proof. (i) Taking the adjoint of the equation \(aa^{-1} = 1 = a^{-1}a\) gives \((a^{-1})^\ast a^\ast = 1 = a^\ast(a^{-1})^\ast\), thus \((a^\ast)^{-1} = (a^{-1})^\ast\). (ii) By (i), \(a - \lambda 1\) is invertible if and only if \(a^\ast - \overline{\lambda} 1\) is invertible.

12.23 Proposition Let \(\mathcal{A}\) be a unital \(C^\ast\)-algebra. Then

(i) If \(a \in \mathcal{A}\) is normal then \(r(a) = \|a\|\).

\(^{40}\)If \(S \subseteq \mathbb{C}\) we write \(S^\ast\) for \(\{\overline{s} \mid s \in S\}\) since \(S\) could be confused with the closure.
(ii) If \( u \in \mathcal{A} \) is unitary then \( \sigma(u) \subseteq S^1 \).

(iii) If \( a \in \mathcal{A} \) is self-adjoint then \( \sigma(a) \subseteq \mathbb{R} \).

Proof. (i) If \( b = b^* \) then \( \|b\|^2 = \|b^*b\| = \|b\|^2 \), and induction gives \( \|b^n\| = \|b\|^n \) \( \forall n \). If \( a \) is normal, then
\[
\|a^{2n}\| = \|(a^*)^{2n}a^{2n}\|^{1/2} = \|(a^*)^{2n}\|^{1/2}(\|a^{2n}\|)^{1/2} = \|a\|^{2n},
\]

since \( a^*a \) is self-adjoint. Now Theorem 11.14 gives
\[
r(a) = \lim_{n \to \infty} \|a^{2n}\|^{1/2n} = \lim_{n \to \infty} (\|a\|^{2n})^{1/2n} = \|a\|. \quad (12.1)
\]

(ii) We have \( \|u\|^2 = \|u^*u\| = \|1\| = 1 \), and in same way \( \|u^{-1}\| = \|u^*\| = 1 \). Now Exercise 11.22 gives \( \sigma(u) \subseteq S^1 \).

(iii) Given \( \lambda \in \sigma(a) \), write \( \lambda = \alpha + i\beta \) with \( \alpha, \beta \in \mathbb{R} \). Since \( \sigma(a+z1) = \sigma(a) + z \forall z \in \mathbb{C} \) (why?), we have
\[
i\beta(n+1) = \alpha + i\beta - \alpha + in\beta \in \sigma(a - \alpha1 + in\beta1).
\]
Thus with \( r(c) \leq \|c\| \) (Proposition 11.11), the \( C^* \)-identity and \( \|1\| = 1 \) we have
\[
(n^2 + 2n + 1)\beta^2 = \|i\beta(n+1)^2 \leq r(a - \alpha1 + in\beta1)^2 \leq \|a - \alpha1 + in\beta1\|^2
\]
\[
= \|\ (a - \alpha1 - in\beta1)(a - \alpha1 + in\beta1) \| = \| (a - \alpha1)^2 + n^2\beta^21 \| \leq \|a - \alpha1\|^2 + n^2\beta^2,
\]

which simplifies to \( (2n + 1)\beta^2 \leq \|a - \alpha1\|^2 \) \( \forall n \in \mathbb{N} \). Thus \( \beta = 0 \) and \( \lambda \in \mathbb{R} \).

12.24 REMARK 1. Since (i) implies \( \|a\| = \|a^*a\|^{1/2} = r(a^*a)^{1/2} \) for all \( a \in \mathcal{A} \) and the spectral radius \( r(a) \) by definition depends only on the algebraic structure of \( \mathcal{A} \), the latter also determines the norm, which therefore is unique in a \( C^* \)-algebra! But note that two \( C^* \)-norms on a \( * \)-algebra \( \mathcal{A} \) can be very different if \( \mathcal{A} \) fails to be complete w.r.t. one of the two norms!

2. An alternative and perhaps more insightful proof for (iii) goes like this: Since \( e^z \equiv \exp(z) = \sum_{n=0}^{\infty} z^n/n! \) converges absolutely for all \( z \in \mathbb{C} \), Proposition 4.2 gives convergence of \( \exp(a) \) for all \( a \in \mathcal{A} \). It is easy to verify \( (e^a)^* = e^{(a)^*} \) and \( e^{a+b} = e^a e^b \) provided \( ab = ba \).
In particular we have \( e^a \in \text{Inv}\mathcal{A} \) for all \( a \) with \( (e^a)^{-1} = e^{-a} \). If now \( a = a^* \) then \( u = e^{ia} \satisfies u^* = e^{-ia}, \) thus \( uu^* = u^*u = 1 \), so that \( u \) is unitary and therefore \( \sigma(e^{ia}) \subseteq S^1 \) by (ii).

Now the holomorphic spectral mapping theorem, cf. Section 13.1, in particular (13.1), gives \( \{e^{i\lambda} \mid \lambda \in \sigma(a)\} = \sigma(e^{ia}) \subseteq S^1 \), and this implies \( \sigma(a) \subseteq \mathbb{R} \). [The above argument only needs \( \{e^{i\lambda} \mid \lambda \in \sigma(a)\} \subseteq \sigma(e^{ia}) \), which can be proven more directly: For all \( \lambda \in \mathbb{C} \) we have
\[
e^{ia} - e^{i\lambda}1 = (e^{i(a-\lambda1)} - 1)e^{i\lambda} = \left( \sum_{k=1}^{\infty} \frac{(i(a - \lambda1))^k}{k!} \right) e^{i\lambda} = (a - \lambda1)b e^{i\lambda},
\]
where \( b = i \sum_{k=1}^{\infty} \frac{(i(a - \lambda1))^{k-1}}{k!} \in \mathcal{A} \). Since \( a - \lambda1 \) and \( b \) commute, we have \( \lambda \in \sigma(a) \Rightarrow e^{i\lambda} \in \sigma(e^{ia}) \).] For another (quite striking) application of \( \exp \) to \( C^* \)-algebras see Section B.8.

12.25 EXERCISE Let \( \mathcal{A} \) be a unital \( C^* \)-algebra and \( a \in \mathcal{A} \) normal. Prove:

(i) \( \|a^n\| = \|a\|^n \) \( \forall n \).
(ii) There is \( \lambda \in \sigma(a) \) with \( |\lambda| = \|a\| \).
(iii) If \( \sigma(a) = \{\lambda\} \) then \( a = \lambda1 \).

The following improvement over Corollary 11.19 and Exercise 11.24 illustrates that \( C^* \)-algebras are better behaved than general Banach algebras:
12.26 Theorem  Let \( \mathcal{A} \) be a unital \( C^* \)-algebra, \( \mathcal{B} \subseteq \mathcal{A} \) a \( C^* \)-subalgebra containing 1. Then

(i) \( \text{Inv}\mathcal{B} = \mathcal{B} \cap \text{Inv}\mathcal{A} \).

(ii) \( \sigma_\mathcal{B}(b) = \sigma_\mathcal{A}(b) \) for all \( b \in \mathcal{B} \).

Proof. (i) The inclusion \( \sigma_\mathcal{A}(b) \subseteq \sigma_\mathcal{B}(b) \) was part of Corollary 11.19, so that we are left with proving \( \supseteq \). Let first \( b = b^* \in \mathcal{B} \cap \text{Inv}\mathcal{A} \). Proposition 12.23(iii) then gives that \( b - it1 \) is invertible in \( \mathcal{B} \) for all \( t \in \mathbb{R} \setminus \{0\} \) and in \( \mathcal{A} \) for all \( t \in \mathbb{R} \). Lemma 11.13 implies that the function \( f : \mathbb{R} \to \mathcal{A} \), \( t \mapsto (b - it1)^{-1} \) is continuous. For \( t \in \mathbb{R} \setminus \{0\} \), \( b - it1 \) is invertible in \( \mathcal{B} \), so that uniqueness of inverses gives \( f(t) \in \mathcal{B} \) for all \( t \neq 0 \). Now continuity of \( f \) and closedness of \( \mathcal{B} \subseteq \mathcal{A} \) imply \( f(0) \in \mathcal{B} \). Since \( f(0) = b^{-1} \), we have \( b^{-1} \in \mathcal{B} \), as claimed.

Let now \( b \in \mathcal{B} \cap \text{Inv}\mathcal{A} \). Then \( b \) has an inverse \( a \in \mathcal{A} \). Furthermore, \( b^* \) is invertible in \( \mathcal{A} \) (Lemma 12.22), thus also \( bb^* \). Since \( bb^* \) is self-adjoint, it has an inverse \( c \in \mathcal{B} \) by the above. Thus \( bb^*c = 1 = cbb^* \), so that \( b^*c \in \mathcal{B} \) is a right inverse for \( b \). Now \( a = abb^*c = b^*c \) proves that \( a = b^*c \), and with \( b, c \in \mathcal{B} \) we have \( b^{-1} = a \in \mathcal{B} \).

(ii) This is immediate by (i) and the definition of the spectrum. ■

12.27 Definition  If \( \mathcal{A} \) is a unital \( C^* \)-algebra, \( a \in \mathcal{A} \) is called positive, or \( a \geq 0 \), if \( a = a^* \) and \( \sigma(a) \subseteq [0, \infty) \).

12.28 Exercise  Give an example of a unital \( C^* \)-algebra \( \mathcal{A} \) and \( a \in \mathcal{A} \) showing that \( \sigma(a) \subseteq [0, \infty) \) does not imply \( a = a^* \! \).

12.29 Exercise  Prove: If \( \mathcal{A} \) is a unital \( C^* \)-algebra and \( a, b \in \mathcal{A} \) are positive and \( ab = ba \) then \( a + b \) is positive. (Later we will use different methods to remove the condition \( ab = ba \).)

13  Functional calculus in Banach and \( C^* \)-algebras

13.1 Some functional calculus in Banach algebras

Let \( A \in B(E) \) be a bounded operator and \( f \) a function (we will soon be more specific). It is natural to ask how to define \( f(A) \). The next question is: Determine \( \sigma(f(A)) \). Does it equal \( f(\sigma(A)) \)? These are the basic questions addressed by the many different ‘functional calculi’ that there are: holomorphic, continuous, Borel, etc.

We immediately generalize the above questions to elements of unital Banach algebras, but mostly we will (later) focus on \( C^* \)-algebras, to which \( B(H) \) belongs for each Hilbert space.

Defining \( f(a) \) poses no problem in the simplest case, which surely is \( f = P \), a polynomial:

13.1 Definition  If \( \mathcal{A} \) is a unital algebra, \( a \in \mathcal{A} \) and \( P(x) = c_nx^n + \cdots + c_1x + c_0 \) is a polynomial, we put \( P(a) = c_na^n + \cdots + c_1a + c_01 \).

13.2 Exercise  Let \( \mathcal{A} \) be a unital algebra and \( a \in \mathcal{A} \). Prove that the map \( \mathbb{C}[x] \to \mathcal{A}, P \mapsto P(a) \) is a homomorphism of unital \( \mathbb{C} \)-algebras.

13.3 Lemma  Let \( \mathcal{A} \) be a unital Banach algebra, \( a \in \mathcal{A} \) and \( P \in \mathbb{C}[x] \) a polynomial. Then

\[ \sigma(P(a)) = P(\sigma(a)) := \{ P(\lambda) \mid \lambda \in \sigma(a) \} \].
Proof. Choose a maximal abelian Banach subalgebra $B \subseteq A$ containing $a$. Since every $\varphi \in \Omega(B)$ is a unital homomorphism, we have

$$\varphi(P(a)) = \varphi(c_n a^n + \cdots + c_1 a + c_0 1) = c_n \varphi(a)^n + \cdots + c_1 \varphi(a) + c_0 = P(\varphi(a)).$$

Now by (11.7) we have

$$\sigma_B(P(a)) = \{ \varphi(P(a)) \mid \varphi \in \Omega(B) \} = \{ P(\varphi(a)) \mid \varphi \in \Omega(B) \} = \{ P(\lambda) \mid \lambda \in \sigma(a) \} = P(\sigma(a)).$$

Now appeal to Exercise 11.24 to have $\sigma_A(P(a)) = \sigma_B(P(a)).$ \hfill \Box

There is a more elementary proof of Lemma 13.3, which works in every normed algebra (but does not lend itself to generalizations like (13.1) or Proposition 13.19(ii)):  

13.4 Exercise Let $A$ be a unital normed algebra and $P \in \mathbb{C}[z]$ with $n = \deg P$.

(i) Prove $\sigma(P(a)) = P(\sigma(a))$ when $n = 0$.

(ii) Assume $n \geq 1$ and $\lambda \in \mathbb{C}$. Use a factorization $P(z) - \lambda = c_n \prod_{k=1}^{n} (z - z_k)$ to prove $\sigma(P(a)) = P(\sigma(a))$ without using characters.

(iii) Why did we assume $A$ to be normed?

The above ‘polynomial functional calculus’ can be generalized: If the power series $f(z) = \sum_{n=0}^{\infty} c_n z^n$ has convergence radius $R$ and $a \in A$ satisfies $\|a\| < R$ then

$$\sum_{n=0}^{\infty} |c_n a^n| = \sum_{n=0}^{\infty} |c_n r^n| \leq \sum_{n=0}^{\infty} |c_n| \|a^n\| < \infty,$$

where we used that the power series $\sum c_n z^n$ and $\sum |c_n| z^n$ have the same convergence radius. Thus we can define $f(a)$ as $\sum_{n=0}^{\infty} c_n a^n$. Furthermore, if $\varphi \in \Omega(A)$ then by continuity of $\varphi$ we have

$$\varphi(f(a)) = \varphi \left( \lim_{N \to \infty} \sum_{n=0}^{N} c_n a^n \right) = \lim_{N \to \infty} \varphi \left( \sum_{n=0}^{N} c_n a^n \right) = \lim_{N \to \infty} \sum_{n=0}^{N} c_n \varphi(a)^n = \sum_{n=0}^{\infty} c_n \varphi(a)^n = f(\varphi(a)).$$

Thus if $A$ is commutative we have

$$\sigma(f(a)) = \{ \varphi(f(a)) \mid \varphi \in \Omega(A) \} = \{ f(\varphi(a)) \mid \varphi \in \Omega(A) \} = f(\sigma(a)). \quad (13.1)$$

If $A$ is non-commutative, we can find a maximal abelian Banach subalgebra $B \subseteq A$ containing $a$. Then $f(a) \in B$, the above gives $\sigma_B(f(a)) = f(\sigma_B(a))$, and Exercise 11.24 allows us to drop the subscript.

If one defines $\mathcal{H}_R$ to be the set of functions defined by power series with convergence radius $\geq R$ then $\mathcal{H}_R$ is easily checked to be a commutative algebra (which coincides with the algebra of functions holomorphic on $B(0, R)$). Now for every $a \in A$ with $\|a\| < R$ one has a unital homomorphism $\mathcal{H}_R \to A$, $f \mapsto f(a)$. This can be generalized quite a bit, leading to the fully fledged holomorphic functional calculus, which we do not discuss. See e.g. [32, 38, 71].

Of course every power series $f(z) = \sum_{n=0}^{\infty} c_n z^n$ with infinite convergence radius is in all $\mathcal{H}_R$, thus it can be ‘applied’ to every $a \in A$. For example $\exp(a) = e^a = \sum_{n=0}^{\infty} \frac{a^n}{n!}$ converges for every $a \in A$. Whenever $ab = ba$ one has $e^a e^b = e^{a+b} = e^b e^a$ by essentially the same argument as for complex numbers. But the commutativity $ab = ba$ is essential! (Why?) Now it follows that $e^{a-a} = e^{a-a} = e^0 = 1 = e^{-a} e^a$, thus $e^a \in \text{Inv} A$ with $(e^a)^{-1} = e^{-a}$. 

78
13.2 Continuous functional calculus for self-adjoint elements in a $C^*$-algebra

Our goal is to make sense of $f(a)$, where $a$ is a normal element of some arbitrary $C^*$-algebra $\mathcal{A}$, for all functions $f \in C(\sigma(a), \mathbb{C})$, in such a way that $f \mapsto f(a)$ is a $*$-homomorphism. (If you don’t care for this generality, you may substitute $\mathcal{A} = B(H)$.) We will first do this for self-adjoint elements and then generalize to normal ones. We cannot hope to go beyond this: If $f(z) = z^2$ then it is not clear whether to define $f(a)$ as $a^2a^*$ or $aa^*a$ or $a^2a^*$ when $aa^* \neq a^*a$. [Yet ‘quantization theory’, motivated by quantum theory, tries to do it.] For normal $a$, this problem does not arise.

13.5 PROPOSITION Let $\mathcal{A}$ be a unital $C^*$-algebra, $a \in \mathcal{A}$ normal and $P$ a polynomial. Then

$$\|P(a)\| = \sup_{\lambda \in \sigma(a)} |P(\lambda)| = \|P_{|\sigma(a)}\|_\infty.$$ 

Proof. Normality of $a$ implies that $P(a)$ is normal. Thus

$$\|P(a)\| = r(P(a)) = \sup_{\lambda \in \sigma(P(a))} |\lambda| = \sup_{\lambda \in \sigma(a)} |P(\lambda)|.$$ 

The first equality is due to Proposition 12.23(i), the second is the definition of $r$ and the third comes from the Lemma 13.3. □

Even though we are after a result for all normal operators, we first consider self-adjoint operators:

13.6 THEOREM Let $\mathcal{A}$ be a unital $C^*$-algebra and $a = a^* \in \mathcal{A}$. Then there is a unique continuous $*$-homomorphism $\alpha_a : C(\sigma(a), \mathbb{C}) \to \mathcal{A}$ such that $\alpha_a(P) = P(a)$ for all polynomials. (Usually we will write $f(a)$ instead of $\alpha_a(f)$.) It satisfies

(i) $\|\alpha_a(f)\| = \sup_{\lambda \in \sigma(a)} |f(\lambda)|$. (Thus $\alpha_a$ is an isometry.)

(ii) The image of $\alpha_a$ is the smallest $C^*$-subalgebra $\mathcal{B} \subseteq \mathcal{A}$ containing $1$ and $a$, and $\alpha : C(\sigma(a), \mathbb{C}) \to \mathcal{B}$ is a $*$-isomorphism.

(iii) $\sigma(\alpha_a(f)) = \{f(\lambda) \mid \lambda \in \sigma(a)\}$. (Spectral mapping theorem)

(iv) If $g \in C(f(\sigma(a)), \mathbb{C})$ then $\alpha_a(g \circ f) = \alpha_a(f)(g)$, or just $g(f(a)) = (g \circ f)(a)$.

Proof. (i) By Propositions 11.11 and 12.23(iii), we have $\sigma(a) \subseteq [-\|a\|, \|a\|]$. By the classical Weierstrass approximation theorem, cf. Theorem A.25, for every continuous function $f : [c, d] \to \mathbb{C}$ and $\varepsilon > 0$ there is a polynomial $P$ such that $|f(x) - P(x)| \leq \varepsilon$ for all $x \in [c, d]$. We cannot apply this directly since $\sigma(a)$, while contained in an interval, need not be an entire interval. But using Tietze’s extension theorem, cf. Appendix A.6, we can find (very non-uniquely) a continuous function $g : [-\|a\|, \|a\|] \to \mathbb{C}$ that coincides with $f$ on $\sigma(a)$. Now this $g$ can be approximated uniformly by polynomials thanks to Weierstrass’ theorem. (Alternatively, apply the more abstract Stone-Weierstrass theorem directly to $f$.) In any case, the restriction of the polynomials to $\sigma(a)$ is dense in $C(\sigma(a), \mathbb{C})$ w.r.t. $\| \cdot \|_\infty$. By Proposition 13.5, the map $C(\sigma(a), \mathbb{C}) \ni C[\sigma(a)] \to \mathcal{A}, P \mapsto P(a)$ is an isometry. Thus applying Lemma 4.17 we obtaining a unique isometry $\alpha_a : C(\sigma(a), \mathbb{C}) \to \mathcal{A}$ extending $P \mapsto P(a)$. Thus (i) is proven up to the claim that $\alpha_a$ is a $*$-homomorphism. This is left as an exercise.

(ii) Since $\alpha_a$ is a $*$-homomorphism, $\mathcal{B} := \alpha_a(C(\sigma(a), \mathbb{C})) \subseteq \mathcal{A}$ is a $*$-subalgebra. And since $\alpha_a$ is an isometry by (i) and $C(\sigma(a), \| \cdot \|_\infty)$ is complete, $\mathcal{B}$ is closed, thus a $C^*$-algebra.
Since $\alpha_a$ maps the constant-one function to $1 \in A$ and the inclusion map $\sigma(a) \hookrightarrow C$ to $a$, $B$ contains $1, a$. Conversely, the smallest $C^*$-subalgebra of $A$ containing $1$ and $a$ clearly is obtained by taking the norm-closure of the set \{ $P(a) \mid P \in \mathbb{C}[z]$ \}, which is contained in the image of $\alpha_a$.

(iii) Let $f \in C(\sigma(a), \mathbb{C})$. Then clearly $\alpha_a(f) \in B$. Now

$$\sigma_A(\alpha_a(f)) = \sigma_B(\alpha_a(f)) = \sigma_{C(\sigma(a), \mathbb{C})}(f) = f(\sigma(a)),$$

where the equalities come from Theorem 12.26, from the fact that $\alpha_a : C(\sigma(a), \mathbb{C}) \to B$ is a $*$-isomorphism, and from Exercise 11.6, respectively.

(iv) If \{ $P_n$ \} is a sequence of polynomials converging to $f$ uniformly on $\sigma(a)$ and \{ $Q_n$ \} is a sequence of polynomials converging to $g$ uniformly on $\sigma(f(a))$, then $Q_n \circ P_n$ converges uniformly to $g \circ f$, thus $Q_n(P_n(a)) = (Q_n \circ P_n)(a)$ converges to $(g \circ f)(a)$. On the other hand, \{ $Q_n(P_n(a))$ \} converges uniformly to $g(f(a))$.

\[ 13.7 \text{ EXERCISE} \quad \text{Prove that $\alpha_a$ is a $*$-homomorphism.} \]

\[ 13.8 \text{ REMARK} \quad \text{The isometric $*$-isomorphism $\alpha_a : C(\sigma(a), \mathbb{C}) \to B = C^*(1, a)$ is a special case of the Gelfand isomorphism for commutative unital $C^*$-algebras proven in Section 18 (which will go in the opposite direction $\pi : A \to C(\Omega(A), \mathbb{C})$). A general commutative $C^*$-algebra $A$ is not generated by a single element $a$, so that we'll need to find a substitute for $\sigma(a)$. It shouldn't be surprising that this will be $\Omega(A)$ (with a suitable topology).} \]

Here are a few of the many applications of the preceding results:

\[ 13.9 \text{ EXERCISE} \quad \text{(i) Define $f_+, f_- : \mathbb{R} \to \mathbb{R}$ by $f_+(x) = \max(x, 0)$, $f_-(x) = -\min(x, 0)$. Prove the alternative formulae $f_\pm(x) = (|x| \pm x)/2$ and $f_+ f_- = 0$ and $f_\pm \in C(\mathbb{R}, \mathbb{R})$.} \]

(ii) Let now $A$ be a unital $C^*$-algebra and $a = a^* \in A$. Define $a_\pm \in A$ by functional calculus as $a_\pm = f_\pm(a)$. Prove: 1. $a_+ - a_- = a$ and $a_+ + a_- = |a|$, 2. $a_+ a_- = a_- a_+ = 0$, 3. $a_+ \geq 0, a_- \geq 0$.

\[ 13.10 \text{ PROPOSITION} \quad \text{Let $A$ be a unital $C^*$-algebra.} \]

(i) If $a = a^* \in A$ then $a^2$ is positive.

(ii) If $a \in A$ is positive then there is a positive $b \in A$ such that $b^2 = a$, unique in $C^*(1, a)$ (and in $A$). We write $b = \sqrt{a}$.

\textbf{Proof.} (i) It is clear that $a^2$ is self-adjoint. Since $\sigma(a) \subseteq \mathbb{R}$ by Proposition 12.23(iii), the spectral mapping theorem (Lemma 13.3 suffices) gives $\sigma(a^2) = \{ \lambda^2 \mid \lambda \in \sigma(a) \} \subseteq [0, \infty)$, thus $a^2 \geq 0$.

(ii) In view of $a \geq 0$ we have $\sigma(a) \subseteq [0, \infty)$. Now continuity of the function $[0, \infty) \to [0, \infty)$, $x \mapsto \sqrt{x}$ allows us to define $b = \sqrt{a}$ by the continuous functional calculus. It is immediate by construction that $b = b^*$, and the spectral mapping theorem gives $\sigma(b) \subseteq [0, \infty)$, thus $b \geq 0$. Now $b^2 = (a^{1/2})^2 = a$ since $\sqrt{x^2} = x$. If $c \in C^*(1, a)$ is positive and $c^2 = a$ then $c = b$. This follows from the $*$-isomorphism $C^*(1, a) \cong C(\sigma(a), \mathbb{C})$ and the fact that positive square roots are unique in the function algebra $C(\sigma(a), \mathbb{C})$ (why?).

One has the stronger result that if $c \in A$ is positive and $c^2 = a$ then $c = b$. The idea is this: We have $ca = c^3 = ac$, so that $c$ commutes with $a$. Since $b$ is a norm-limit of polynomials in $a$, it also commutes with $c$. Now $C^*(1, a, c)$ is a commutative $C^*$-algebra, thus isomorphic to a function algebra as we will prove in Theorem 18.11. Now $c^2 = a = b^2$ implies $c = b$ as before.
Exercise 13.11 Prove: If $\mathcal{A}$ is a unital $C^*$-algebra and $c \in \mathcal{A}$ is normal then $c^*c$ is positive. 

Hint: Remark 12.21.1.

The next two results, not proven during the lecture, generalize Exercises 12.29 and 13.11, respectively, and show the usefulness of the preceding theory even in very non-commutative situations:

Proposition 13.12 Let $\mathcal{A}$ be a unital $C^*$-algebra.

(i) $a \in \mathcal{A}$ is positive if and only if $a = a^*$ and there is a $t \geq 0$ such that $\|a - t1\| \leq t$.

(ii) If $a, b \in \mathcal{A}$ are positive then $a + b$ is positive. (No assumption that $ab = ba$!)

Proof. (i) Under either assumption we have $a = a^*$. Let $\alpha_a : C(\sigma(a), \mathbb{C}) \to C^*(a, 1) \subseteq \mathcal{A}$ be the $*$-isomorphism from Theorem 13.6. We have $\alpha_a(a) = a$, where $f : \sigma(a) \to \mathbb{C}$ is the inclusion map.

If $a \geq 0$ then the function $f$ takes values in $[0, \|a\|]$. Putting $t = \|a\|, f - t1$ takes values in $[t, 0]$, thus $\|a - t1\| = \|f - t1\| \leq t$.

Now assume $\|a - t1\| \leq t$ for some $t \geq 0$. Then the function $f - t1$ takes values in $\sigma(a) - t1 \subseteq [-t, t]$, thus $f$ takes values in $[0, 2t]$. Thus $f$ is positive, and so is $a = \alpha_a(a)$.

(ii) If $a, b \geq 0$ then $a = a^*, b = b^*$, so that $(a + b)^* = a + b$. And by (i) there are $s, t \geq 0$ such that $\|a - s1\| \leq s$ and $\|b - t1\| \leq t$. This implies $\|(a + b) - (s + t)1\| \leq \|a - s1\| + \|b - t1\| \leq s + t$, so that $a + b \geq 0$ by (i).

Proposition 13.13 If $\mathcal{A}$ is a unital $C^*$-algebra and $a \in \mathcal{A}$ then $a^*a$ is positive.

Proof. First a preparatory argument: Assume $c \in \mathcal{A}$ is such that $-c^*c$ is positive. Then by Exercise 11.21(iv) we have $\sigma(-cc^*) \setminus \{0\} = \sigma(-c^*c) \setminus \{0\}$, thus $-c^*c$ is positive. Writing $c = a + ib$ with $a, b$ self-adjoint, we have $c^*c + cc^* = (a - ib)(a + ib) + (a + ib)(a - ib) = 2a^2 + 2b^2$, thus $c^*c = 2a^2 + 2b^2 - cc^*$. Using $-cc^* \geq 0$ just proven and Propositions 13.10(i) and 13.12(ii) this implies $c^*c \geq 0$. Combining $-c^*c \geq 0$ and $c^*c \geq 0$ gives $\sigma(c^*c) \subseteq [0, \infty) \cap (\infty, 0] = \{0\}$. This implies $\|c\|^2 = \|c^*c\| = r(c^*c) = 0$, thus $c = 0$.

We turn to the proof of the claim. Let $a \in \mathcal{A}$ be arbitrary. Then $b = a^*a$ is self-adjoint, thus with Exercise 13.9(ii) we have $b = b_+ - b_-$ with $b_+ \geq 0$ and $b_-b_- = 0$. Putting $c = ab_-$ we have $-c^*c = -b_-a^*a b_- = -b_-(b_+ - b_-)b_- = b_-^2$, which is positive (spectral mapping theorem). Now the preparatory step gives $ab_- = c = 0$. This implies $-b_-^2 = (b_+ - b_-)b_- = b_-^2 = a^*a b_- = 0$, thus $b_- = 0$. (Since $d = d^*$, $d^2 = 0$ implies $d = 0$.) Now we have $a^*a = b = b_+ \geq 0$.

Exercise 13.14 Let $\mathcal{A}$ be a unital $C^*$-algebra and $a, b \in \mathcal{A}$ with $a \geq 0$. Prove that $bab^* \geq 0$.

Definition 13.15 If $\mathcal{A}$ is a unital $C^*$-algebra and $a \in \mathcal{A}$, we define $|a| = (a^*a)^{1/2}$.

By construction, $|a|$ is positive and is similar to $|z|$ for $z \in \mathbb{C}$. But some care is required since $|a^*| = |a|$ holds if and only if $a$ is normal. The ‘if’ part is obvious, and the converse follows by $a^*a = \sqrt{a^*a^2} = |a|^2 = |a^*|^2 = \sqrt{aa^*} = aa^*$, where we used that $(\sqrt{b})^2$ for $b \geq 0$.

It is not unreasonable to ask whether, as for complex numbers, we have a factorization $a = b|a|$ for each $a \in \mathcal{A}$. In Section 14.3 we will prove this for $\mathcal{A} = B(H)$, but:

Exercise 13.16 Give an example of a unital $C^*$-algebra $\mathcal{A}$ and $a \in \mathcal{A}$ such that there is no $b \in \mathcal{A}$ with $a = b|a|$.
13.3 Continuous functional calculus for normal elements in a $C^*$-algebra

13.17 Theorem  Theorem 13.6 literally extends to all normal elements of a unital $C^*$-algebra.

The proof of Theorem 13.6 does not generalize immediately. The reason is that the spectrum of a normal operator need not be contained in $\mathbb{R}$. (In fact, for normal $a$ we have $\sigma(a) \subseteq \mathbb{R} \Leftrightarrow a = a^*$, cf. Exercise 13.21.) If that happens, the polynomials, restricted to $\sigma(a)$, fail to be uniformly dense in $C(\sigma(a), \mathbb{C})$. (All functions that are uniform limits of polynomials on sufficiently large subsets of $\mathbb{C}$ are holomorphic so that, e.g. $f(z) = \text{Re} z$ cannot be approximated by polynomials in $z = x + iy$.) But with $\sigma(a) \subseteq \mathbb{C} \cong \mathbb{R}^2$ and considering functions on (a subset of) $\mathbb{C}$ as functions of two real variables, the polynomials in $x, y$ are dense in $C(\sigma(a), \mathbb{C})$ by the higher dimensional version of the classical Weierstrass theorem, cf. Theorem A.31. Thus also the polynomials in $z = x + iy$ and $\bar{z} = x - iy$ are dense\(^{41}\). Now there is a unique unital homomorphism $\alpha_a$ from $\mathbb{C}[z, \bar{z}]$ to $\mathcal{A}$ sending $z$ to $a$ and $\bar{z}$ to $a^*$, and we need to adapt Proposition 13.5 to this setting. For this we need another lemma:

13.18 Lemma  Let $\mathcal{A}$ be a unital $C^*$-algebra. Then every character $\varphi \in \Omega(\mathcal{A})$ satisfies $\varphi(a^*) = \overline{\varphi(a)}$ for all $a \in \mathcal{A}$, i.e. $\varphi$ is a $*$-homomorphism.

Proof. We have $c = a + ib$, where $a = \text{Re}(c), b = \text{Im}(c)$ are self-adjoint. Now $\sigma(a) \subseteq \mathbb{R}$ by Proposition 12.23(iii), thus $\varphi(a) \in \sigma(a) \subseteq \mathbb{R}$ by Lemma 11.34. Similarly $\varphi(b) \in \mathbb{R}$. Thus

$$\varphi(a^*) = \varphi(a - ib) = \varphi(a) - i\varphi(b) = \overline{\varphi(a)} + i\overline{\varphi(b)} = \overline{\varphi(a + ib)} = \overline{\varphi(c)},$$

where the third equality used that $\varphi(a), \varphi(b) \in \mathbb{R}$ as shown before. □

13.19 Proposition  Let $\mathcal{A}$ be a unital $C^*$-algebra and $a \in \mathcal{A}$ normal. Then

(i) If $P \in \mathbb{C}[z, \bar{z}]$, define $P(a, a^*)$ by replacing $z$ and $\bar{z}$ in $P$ by $a, a^*$, respectively. Then $\alpha_a : P \mapsto P(a, a^*)$ is a $*$-homomorphism (extending the $\alpha_a : \mathbb{C}[z] \to \mathcal{A}$ defined earlier).

(ii) For every $P \in \mathbb{C}[z, \bar{z}]$ we have $\sigma(P(a, a^*)) = \{P(\lambda, \bar{\lambda}) \mid \lambda \in \sigma(a)\}$ and

$$\|P(a, a^*)\| = \sup_{\lambda \in \sigma(a)} |P(\lambda, \bar{\lambda})|.$$

Proof. (i) If $P(z, \bar{z}) = \sum_{i,j=0}^N c_{ij} z^i \bar{z}^j$ we put $P(a, a^*) = \sum_{i,j=0}^N c_{ij} a^i a^j \bar{\lambda}^j$. Using the normality of $a$ it is very easy to see that $P \mapsto P(a, a^*)$ is an algebra homomorphism. It also is a $*$-homomorphism if we define $P^*(z, \bar{z}) = \sum_{i,j=0}^N \overline{c_{ij}} z^i \bar{z}^j$.

(ii) Since $a$ is normal, $\mathcal{B} = C^*(1, a) \subseteq \mathcal{A}$ is commutative, so that Proposition 11.37 applies, and using that the $\varphi \in \Omega(\mathcal{B})$ are $*$-homomorphisms thanks to the preceding lemma, we have

$$\sigma_{\mathcal{B}}(P(a, a^*)) = \{\varphi(P(a, a^*)) \mid \varphi \in \Omega(\mathcal{B})\} = \left\{\varphi\left(\sum_{i,j=0}^N c_{ij} a^i (a^*)^j\right) \mid \varphi \in \Omega(\mathcal{B})\right\}$$

$$= \left\{\sum_{i,j=0}^N c_{ij} \varphi(a)^i \overline{\varphi(a)}^j \mid \varphi \in \Omega(\mathcal{B})\right\} = \{P(\lambda, \bar{\lambda}) \mid \lambda \in \sigma(a)\}.$$

Now we appeal to Theorem 12.26 to get $\sigma_{\mathcal{A}}(P(a, a^*)) = \sigma_{\mathcal{B}}(P(a, a^*))$.

\(^{41}\)We allow ourselves the harmless sloppiness of not distinguishing between elements of the ring $\mathbb{C}[z, \bar{z}]$ (where $z, \bar{z}$ are independent variables) and the functions $\mathbb{C} \to \mathbb{C}$ induced by them.
Since $P(a, a^*)$ is normal, the last claim follows from (ii) and Proposition 12.23(i).

Now the proof of Theorem 13.6 becomes a proof of Theorem 13.17 if we replace the invocation of Proposition 13.5 by one of Proposition 13.19 and use the density of the $P(z, z) \uparrow \sigma(a)$ in $\mathcal{C}(\sigma(a), \mathbb{C})$ as explained before.

13.20 Exercise Let $\mathcal{A}$ be a unital $C^*$-algebra and $a \in \mathcal{A}$ normal. For $t \notin \sigma(a)$, prove that $\|(a - t1)^{-1}\| = (\text{dist}(t, \sigma(a)))^{-1}$.

13.21 Exercise Let $\mathcal{A}$ be a unital $C^*$-algebra and $a \in \mathcal{A}$ normal. Prove

(i) If $\sigma(a) \subseteq \mathbb{R}$ then $a$ is self-adjoint.
(ii) If $\sigma(a) \subseteq S^1$ then $a$ is unitary.

The following will have applications to normal operators on Hilbert space:

13.22 Proposition Let $\mathcal{A}$ be a unital $C^*$-algebra and $a \in \mathcal{A}$ normal. If $\Sigma \subseteq \sigma(a)$ is clopen (closed and open) and $\emptyset \neq \Sigma \neq \sigma(a)$ then $p = \chi_\Sigma(a) \in \mathcal{A}$ is a non-trivial orthogonal projection that commutes with $a$.

Proof. The function $f = \chi_\Sigma$ is continuous since $\Sigma$ is clopen, thus we can use continuous functional calculus to put $p = f(a)$. Since $f$ takes values in $\{0, 1\}$, the spectral mapping theorem gives $\sigma(p) = f(\sigma(a)) \subseteq \{0, 1\}$. Thus $p^* = \overline{f}(a) = f(a) = p$ and $p^2 = f^2(a) = f(a) = p$, so that $p$ is an orthogonal projection. It is non-zero since $\Sigma \neq \emptyset$ and $f|_{\Sigma} = 1$. Similarly $p \neq 1$ since $\Sigma \neq \sigma(a)$. Now $ap = pa$ follows from the fact that $p \in C^*(1, a)$, which is commutative.

We now leave the discussion of abstract Banach and $C^*$-algebras (to which we will briefly return at the end) and return to operator theory.

14 More on Hilbert space operators

We now return to bounded operators on Hilbert spaces. Of course all results of the preceding section also apply to the $C^*$-algebra $B(H)$. But there is more to say about operators on Hilbert space than about general $C^*$-algebra elements.

If $H$ is a finite dimensional Hilbert space and $A \in B(H)$ then one easily checks that $A$ is surjective if and only if $A^*$ is injective. In infinite dimensions, this becomes modified:

14.1 Lemma Let $H$ be a Hilbert space and $A \in B(H)$. Then $\ker A^* = (AH)^\perp$. Thus $A$ has dense image ($\overline{A}H = H$) if and only if $A^*$ is injective. (Compare Exercise 12.3(i).)

Proof. We have $x \in (AH)^\perp \iff \langle Ay, x \rangle = 0$ $\forall y \iff \langle y, A^*x \rangle = 0$ $\forall y \iff A^*x = 0$. Thus $A^*$ is injective if and only if $(AH)^\perp = \{0\}$, which is equivalent to $\overline{A}H = H$ by Exercise 6.26(i).

14.2 Exercise Let $A \in B(H)$. Prove:

(i) If $\lambda \in \sigma_c(A)$ then $\overline{\lambda} \in \sigma_p(A^*)$.
(ii) If $\lambda \in \sigma_p(A)$ then $\overline{\lambda} \in \sigma_p(A^*) \cup \sigma_c(A^*)$.
(iii) Use the general theory proven so far and the results of Exercise 11.27 to determine $\sigma_c(L), \sigma_c(L), \sigma_c(R), \sigma_c(R)$ for the shift operators $L, R$ on $\ell^2(\mathbb{N}, \mathbb{C})$.

83
14.3 Remark Using Exercise 12.3(i) instead of Lemma 14.1, one has analogous results for the transpose $A^t \in B(V^*)$ of a Banach space operator $A \in B(V)$: $\sigma_r(A) \subseteq \sigma_p(A^t)$ and $\sigma_p(A) \subseteq \sigma_p(A^t) \cup \sigma_r(A^t)$. (As in Exercise 12.3(iv) there is no complex conjugation since $A \mapsto A^t$ is linear.) □

14.1 Normal operators

We proved a bijection between bounded operators and bounded sesquilinear forms. In fact, an operator $A \in B(H)$ is already determined by the ‘diagonal’ elements (or ‘expectation values’ in quantum theory) $[x,x]_A = \langle Ax, x \rangle$ of the associated form:

14.4 Lemma Let $H$ be a Hilbert space over $\mathbb{C}$ and $A,B \in B(H)$.

(i) If $\langle Ax, x \rangle = 0$ for all $x \in H$ (‘$A$ has vanishing diagonal elements’) then $A = 0$.

(ii) If $\langle Ax, x \rangle = \langle Bx, x \rangle$ for all $x \in H$ then $A = B$.

(The converse statements are trivially true.)

Proof. (i) The hypothesis implies $\langle A(x + y), x + y \rangle = 0$ $\forall x,y$, and expanding this, using $\langle Ax, x \rangle = \langle Ay, y \rangle = 0$ gives $\langle Ax, y \rangle + \langle Ay, x \rangle = 0$. Replacing $x$ by $ix$ gives $\langle Ax, y \rangle - \langle Ay, x \rangle = 0$. Adding the two equations, we obtain $\langle Ax, y \rangle = 0$ $\forall x,y$. This implies $Ax = 0$ $\forall x$, thus $A = 0$.

(ii) follows by applying (i) to $A - B$. ■

14.5 Remark There is something to be said for the above simple direct argument, but the result also follows from Remark 6.14, which even allows to recover $A$ (more precisely $[\cdot, \cdot]_A$) from the map $x \mapsto \langle Ax, x \rangle$.

Normality of $A \in B(H)$ is defined as in general $C^*$-algebras, i.e. as $AA^* = A^*A$.

14.6 Proposition Let $H$ be a Hilbert space and $A \in B(H)$. Then

(0) $A$ is normal if and only if $A^*$ is normal.

(i) $A$ is normal if and only if $\|Ax\| = \|A^*x\|$ for all $x \in H$.

(ii) If $A$ is normal then $(AH)\perp = \ker A^* = \ker A$.

(iii) For a normal operator, injectivity ⇔ dense image.

(iv) For a normal operator, invertibility ⇔ boundedness below ⇔ surjectivity.

Proof. (0) This is trivial, but nevertheless worth pointing out.

(i) If $A$ is normal then for all $x \in H$ we have

$$\|A^*x\|^2 = \langle A^*x, A^*x \rangle = \langle AA^*x, x \rangle = \langle A^*Ax, x \rangle = \langle Ax, Ax \rangle = \|Ax\|^2.$$  

This computation holds both ways, thus $\|Ax\| = \|A^*x\|$ $\forall x$ implies $\langle AA^*x, x \rangle = \langle AA^*x, x \rangle$ $\forall x$.

By Lemma 14.4, this implies $AA^* = A^*A$.

(ii) The first equality is Lemma 14.1, and the second is immediate from (i).

(iii) In view of $(AH)\perp = \ker A$ established in (ii), injectivity of $A$ is equivalent to $(AH)\perp = \{0\}$, which is equivalent to $AHH = H$ by Exercise 6.26(i).

(iv) Invertibility implies boundedness below and surjectivity, cf. Proposition 10.24. If a normal operator is bounded below, it is injective, so that it has dense image by (iii). Now boundedness below and dense image imply invertibility by Proposition 10.24. And surjectivity
imply dense image, thus injectivity by (iii). Now injectivity and surjectivity give invertibility.

Normal operators have very nice spectral properties, which foreshadows the spectral theorem:

14.7 Exercise Let $A \in B(\mathcal{H})$ be normal. Prove:

(i) If $A^n x = 0$ for some $n \in \mathbb{N}$ then $Ax = 0$.

(ii) $L_{\lambda}(A) = \ker(A - \lambda 1)$. (Thus generalized eigenvectors are eigenvectors.)

14.8 Exercise Let $A \in B(\mathcal{H})$ be normal. Let $x, x'$ be (non-zero) eigenvectors for the eigenvalues $\lambda, \lambda'$, respectively. Prove:

(i) $A^*x = \overline{\lambda}x$, thus $x$ is eigenvector for $A^*$ with eigenvalue $\overline{\lambda}$.

(ii) $\sigma_p(A^*) = \sigma_p(A)^*$.

(iii) If $\lambda \neq \lambda'$ then $x \perp x'$.\(^{42}\)

14.9 Exercise (Spectra of normal operators) Let $A \in B(\mathcal{H})$ be normal. Prove:

(i) $\sigma_t(A) = \emptyset$. (No residual spectrum)

(ii) $\sigma_c(A^*) = \sigma_c(A)^*$.

(iii) $\sigma(A) = \sigma_{app}(A) = \{\lambda \in \mathbb{C} \mid \forall \varepsilon > 0 \exists x \in \mathcal{H} : \|x\| = 1, \|\lambda x - \lambda x\| < \varepsilon\}$.

The next result is less straightforward in that its proof requires some machinery:

14.10 Proposition Let $A \in B(\mathcal{H})$ be normal.

(i) If there is a clopen $\Sigma \subseteq \sigma(a)$ with $\emptyset \neq \Sigma \neq \sigma(a)$ then there are non-zero closed subspaces $H_1 \subseteq \mathcal{H}$ and $H_2 = H_1^\perp$ such that $AH_i \subseteq H_i, i = 1, 2$. Thus $A = A|_{H_1} \oplus A|_{H_2}$.

(ii) If $\lambda \in \sigma(A)$ is isolated then $\lambda \in \sigma_p(A)$, thus $\lambda$ is an eigenvalue.

Proof. (i) Define an orthogonal projection $P = \chi_\Sigma \in B(\mathcal{H})$ as in Proposition 13.22. Then $H_1 = PH, H_2 = (1 - P)\mathcal{H}$ are mutually orthogonal closed subspaces. Since $P$ commutes with $A$, the subspaces $H_1, H_2$ are $A$-invariant, i.e. $AH_i \subseteq H_i$, thus $A = A|_{H_1} \oplus A|_{H_2}$.

(ii) Since $\lambda$ is isolated, $\{\lambda\} \subseteq \sigma(A)$ is clopen. By Proposition 13.22 there is an orthogonal projection $P = \chi_{\{\lambda\}} \in B(\mathcal{H})$. Then $P \neq 0$, thus $PH \subseteq \mathcal{H}$ is a non-zero closed subspace. If $0 \neq x \in PH$ then $x = Px$, and

$$(A - \lambda 1)x = (A - \lambda 1)Px = (z - \lambda)(A)\chi_{\{\lambda\}}(A)x = ((z - \lambda)\chi_{\{\lambda\}})(A)x = 0,$$

where $z$ is the inclusion map $\sigma(A) \hookrightarrow \mathbb{C}$ and we used the homomorphism property of the functional calculus and the fact that the function $z \mapsto (z - \lambda)\chi_{\{\lambda\}}(z)$ is identically zero. This proves that $x \in \ker(A - \lambda 1)$, so that $\lambda \in \sigma_p(A)$.

Analogue of Propositions 13.22 and 14.10 hold for all Banach algebras and Banach space operators (thus also non-normal Hilbert space operators), respectively, but the proofs require the holomorphic functional calculus mentioned before.

\(^{42}\)You may have seen this before, but probably only for self-adjoint operators.
14.2 Self-adjoint operators

Self-adjoint operators are normal, thus the results of Section 14.1 all apply!

14.11 Lemma Let $A \in B(H)$. Then

(i) $A = A^* \iff \langle Ax, x \rangle \in \mathbb{R}$ \(\forall x \in H\).

(ii) If $A = A^*$ then $\sigma_p(A) \subseteq \mathbb{R}$. (We already know $\sigma(A) \subseteq \mathbb{R}$, but with a longer proof.)

Proof. (i) By Lemma 14.4, $A = A^*$ is equivalent to $\langle Ax, x \rangle = \langle A^*x, x \rangle \forall x$. In view of $\langle A^*x, x \rangle = \langle x, Ax \rangle = \langle Ax, x \rangle$, we find that $A = A^*$ is equivalent to $\langle Ax, x \rangle = \langle Ax, x \rangle \forall x$, which in turn is equivalent to $\langle Ax, x \rangle \in \mathbb{R} \forall x$.

(ii) If $\lambda \in \sigma_p(A)$ then $\lambda$ is an eigenvalue, so that there is a corresponding eigenvector $x \neq 0$. Now $\lambda \|x\|^2 = \lambda \langle x, x \rangle = \langle Ax, x \rangle \in \mathbb{R}$ by (i). With $\|x\| \neq 0$ this implies $\lambda \in \mathbb{R}$. ■

14.12 Exercise Use Exercise 14.9(iii) to give simple(r) proofs for $\sigma(A) \subseteq \mathbb{R}$ and $\sigma(A) \subseteq S^1$ for $A \in B(H)$ self-adjoint or unitary, respectively.

14.13 Exercise (i) Let $A \in B(H)$ and $K \subseteq H$ a closed subspace such that $AK \subseteq K$. (Thus $K$ is $A$-invariant.) Prove that $AK^\perp \subseteq K^\perp$ is equivalent to $A^*K \subseteq K$.

In this situation, $K$ is called reducing since then $A \cong A|_K \oplus A|_{K^\perp}$.

(ii) Deduce that every invariant subspace of a self-adjoint operator is reducing.

(iii) Show by example that a normal operator can have invariant but non-reducing subspaces.

For every $A \in B(H)$, using (6.2) we have

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

For self-adjoint $A$, the norm is determined already by the ‘diagonal’ elements:

14.14 Proposition If $H$ is a Hilbert space and $A = A^* \in B(H)$ then

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

Proof. Putting $M = \sup_{\|x\| = 1} |\langle Ax, x \rangle|$, Cauchy-Schwarz gives $M \leq \|A\|$. (We also note for later use that $|\langle Ax, x \rangle| \leq M \|x\|^2 \forall x$.) It remains to prove $\|A\| \leq M$, which in view of $\|A\| = \sup_{\|x\| = 1} \|Ax\|$ follows if we have $\|Ax\| \leq M$ whenever $\|x\| = 1$. This inequality is trivially true if $Ax = 0$. If not, put $y = \frac{Ax}{\|Ax\|}$. Using $A = A^*$ and $\langle Ax, y \rangle = \|Ax\|^{-1} \langle Ax, Ax \rangle \in \mathbb{R}$, we have $\langle Ay, x \rangle = \langle y, Ax \rangle = \langle y, \overline{Ax} \rangle = \langle Ax, y \rangle$. Using this, we have

$$\|Ax\| = \langle Ax, y \rangle = \frac{1}{4}(|\langle Ax, x + y \rangle - \langle Ax, x - y \rangle| - \langle Ax, x - y \rangle|) \leq \frac{1}{4} \left(|\langle Ax, x + y \rangle| + |\langle Ax, x - y \rangle|\right) \leq \frac{M}{4} \left(\|x + y\|^2 + \|x - y\|^2\right) \leq \frac{M}{2} (\|x\|^2 + \|y\|^2) = M,$$

where in the last steps we used the parallelogram identity (6.3) and $\|x\| = \|y\| = 1$. ■
14.15 REMARK 1. The set \( \{ \langle Ax, x \rangle \mid \|x\| = 1 \} \) is called the numerical range of \( A \). In quantum mechanics [37] it is the set of expectation values of \( A \).

2. The number \( \sup_{\|x\|=1} |\langle Ax, x \rangle| \) is the numerical radius of \( A \). The identity \( \|A\| = \|A^*\| \) generalizes to all normal operators, but the proof is a bit trickier, see [54, Proposition 3.2.25] or [28]. (It also follows from the spectral theorem, cf. Section 16.1.) □

14.3 Positive operators. Polar decomposition

In an abstract \( C^*-\)algebra \( A \) we called \( a \in A \) positive if \( a = a^* \) and \( \sigma(a) \subseteq [0, \infty) \). In the \( C^*-\)algebra \( B(H) \), there is another notion of positivity:

14.16 DEFINITION Let \( H \) be a Hilbert space and \( A \in B(H) \). Then \( A \) is called operator positive, \( A \geq_0 0 \), if \( \langle Ax, x \rangle \geq 0 \) for all \( x \in H \). (I.e., the numerical range of \( A \) is contained in \( [0, \infty) \).

14.17 PROPOSITION Let \( A, B \in B(H) \). Then

(i) If \( A \geq_0 0 \), \( B \geq_0 0 \) then \( A + B \geq_0 0 \).
(ii) \( A^*A \geq_0 0 \).
(iii) If \( A \geq_0 0 \) then \( BAB^* \geq_0 0 \).
(iv) If \( A \geq_0 0 \) then \( A = A^* \) and \( \sigma(A) \subseteq [0, \infty) \).
(v) If \( A = A^* \) and \( \sigma(A) \subseteq [0, \infty) \) then \( A \geq_0 0 \).
(vi) Thus \( A \geq_0 0 \iff A \geq 0 \) in the \( C^*-\)sense. We therefore drop the notation \( \geq_0 \).

Proof. (i) Obvious.

(ii) For every \( x \in H \) one has \( \langle A^*Ax, x \rangle = \langle Ax, Ax \rangle \geq 0 \) so that \( A^*A \geq 0 \).

(iii) \( \langle BAB^*x, x \rangle = \langle AB^*x, B^*x \rangle \geq 0 \).

(Note how easy these three proofs were compared to the analogous statements for \( C^*-\)algebras.)

(iv) If \( A \geq 0 \) then Lemma 14.11(i) gives \( A = A^* \) so that \( \sigma(A) \subseteq \mathbb{R} \) by Proposition 12.23(iii). It remains to prove that \( \lambda < 0 \) implies \( \lambda \not\in \sigma(A) \). For all \( x \in H \) we have, using \( \lambda < 0 \) and \( A \geq 0 \),

\[
\langle (A - \lambda \mathbf{1})x, (A - \lambda \mathbf{1})x \rangle = \|Ax\|^2 + |\lambda|^2\|x\|^2 - 2\lambda\langle Ax, x \rangle \geq |\lambda|^2\|x\|^2,
\]

(note that \( -2\lambda \langle Ax, x \rangle \geq 0 \), thus \( \|A - \lambda \mathbf{1}\| \geq \|\lambda\| \|x\| \), so that \( A - \lambda \mathbf{1} \) is bounded below. Since it is also normal, Proposition 14.6(iv) implies that \( A - \lambda \mathbf{1} \) is invertible. Thus \( \sigma(A) \cap (-\infty, 0) = \emptyset \).

(v) \( A \) satisfies the hypotheses of Proposition 13.10, so that there is a \( B = B^* \in B(H) \) such that \( A = B^2 = B^*B \). Now the claim follows from (ii).

(vi) Combine (iv) and (v).

If \( A \in B(H) \), we put \( |A| = (A^*A)^{1/2} \), as in Definition 13.15.

14.18 DEFINITION \( V \in B(H) \) is a partial isometry if \( V|_{\ker V^\perp} : (\ker V)^\perp \to H \) is an isometry.

14.19 EXERCISE Prove that for \( V \in B(H) \), the following are equivalent.

(i) \( V \) is a partial isometry.
(ii) \( V^* \) is a partial isometry.
(iii) \( V^*V \) is an orthogonal projection.
(iv) \( VV^* \) is an orthogonal projection.

87
In Exercise 13.16 we saw a $C^*$-algebra not admitting polar decomposition. But:

14.20 Proposition (Polar Decomposition) Let $H$ be a Hilbert space and $A \in B(H)$.

(i) There exists a unique partial isometry $V$ such that $A = V|A|$ and $\ker A = \ker V$.

(ii) If $A$ is injective (invertible) then $V$ is an isometry (unitary).

(iii) In addition, we have $|A| = V^* A$. [This follows trivially from (i) only if $V$ is unitary.]

Proof. (i) For each $x \in X$ we have

$$\|Ax\|^2 = \langle Ax, Ax \rangle = \langle x, A^*Ax \rangle = \langle (A^*A)^{1/2}x, (A^*A)^{1/2}x \rangle = \langle |A|x, |A|x \rangle = \|A|x\|^2.$$  \hfill (14.1)

Thus we can define a linear map $V : |A|H \to H$ by $V|A|x = Ax$. (If $|A|x = |A|x'$ then $|A|(x - x') = 0$, thus $A(x - x') = 0$ by (14.1), so that $V$ is well defined.) This map is isometric, thus it extends continuously (Lemma 4.17) to $|A|H$. We extend $V$ to all of $H$ by having it send $|A|H^\perp$ to zero, obtaining a partial isometry. We have

$$\ker V = |A|H^\perp = \ker |A|^* = \ker |A| = \ker A,$$

where we used Lemma 14.1, the self-adjointness of $|A|$ and (14.1). It is clear from the definition that $V|A| = A$.

That $V$ is uniquely determined by its properties is quite clear: If must send $|A|x$ to $Ax$, which determines it on $|A|H$. And in view of $\ker A = \ker |A| = \ker |A|^* = |A|H^\perp$, the requirement $\ker V = \ker A$ forces $V$ to be zero on $|A|H^\perp$.

(ii) It is trivial that an injective (bijective) partial isometry is an isometry (unitary).

(iii) Using (14.1) as in (i) one can define a partial isometry $W$ such that $W|x = |A|x \forall x \in H$ and $W \upharpoonright (AH)^\perp = 0$. Now it is immediate that $WV \upharpoonright |A|H = \id$ and $VW \upharpoonright AH = \id$, while $WV$ and $VW$ vanish on $|A|H^\perp$ and $AH^\perp$ respectively. Thus $W = V^*$, and we are done. \hfill $\Box$

14.21 Remark If $\mathcal{A} \subseteq B(H)$ is a von Neumann algebra (a class of particularly nice $C^*$-subalgebras of $B(H)$) and $A \in \mathcal{A}$ one has not only $|A| \in \mathcal{A}$, but also $V \in \mathcal{A}$. \hfill $\Box$

At this point, we could go on to Section 16, where the various versions of the spectral theorem for normal operators are proven. But it is customary to first study compact operators, since their spectral theory is much simpler and quite similar to that in finite dimensions.

15 Compact operators

15.1 Compact Banach space operators

We have met compact topological spaces many times in this course. A subset $Y$ of a topological space $(X, \tau)$ is compact if it is compact when equipped with the induced (=subspace) topology $\tau_Y$. And $Y \subseteq X$ is called precompact (or relatively compact) if its closure $\overline{Y}$ is compact. Recall that a metric space $X$, thus also a subset of a normed space, is compact if and only if every sequence $\{x_n\}$ in $X$ has a convergent subsequence.

A subset $Y$ of a normed space $(V, \| \cdot \|)$ is called bounded if there is an $M$ such that $\|y\| \leq M \forall y \in Y$. A compact subset of a normed space is closed and bounded, but the converse, while
true for finite dimensional spaces by the Heine-Borel theorem, is false in infinite dimensional spaces. This is particularly easy to see for a Hilbert space: Any ONB $B \subseteq H$ clearly is bounded. For any $e,e' \in B$, $e \neq e'$ we have $\|e - e'\| = \langle e - e', e - e' \rangle^{1/2} = \sqrt{2}$. Thus $B \subseteq H$ is closed and discrete. Since it is infinite, it is not compact.

For normed spaces, one needs the following easy, but important lemma:

15.1 Lemma (F. Riesz) Let $(V, \| \cdot \|)$ be a normed space and $W \subseteq V$ a closed proper subspace. Then for each $\delta \in (0,1)$ there is an $x_\delta \in V$ such that $\|x_\delta\| = 1$ and $\text{dist}(x_\delta, W) \geq \delta$, i.e. $\|x_\delta - x\| \geq \delta \ \forall x \in W$.

Proof. If $x_0 \in V \setminus W$ then closedness of $W$ implies $\lambda = \text{dist}(x_0, W) > 0$. In view of $\delta \in (0,1)$, we have $\frac{\lambda}{\delta} > \lambda$, so that we can find $y_0 \in W$ with $\|x_0 - y_0\| < \frac{\lambda}{\delta}$. Putting

$$x_\delta = \frac{y_0 - x_0}{\|y_0 - x_0\|},$$

we have $\|x_\delta\| = 1$. If $x \in W$ then

$$\|x - x_\delta\| = \left\|x - \frac{y_0 - x_0}{\|y_0 - x_0\|} \right\| = \frac{\|y_0 - x_0\| \cdot \|x - y_0 + x_0\|}{\|y_0 - x_0\|} \geq \text{dist}(x_\delta, W) \geq \frac{\lambda}{\delta} = \delta,$$

where the $\geq$ is due to $\|y_0 - x_0\| \cdot \|x - y_0 \| < \|x_0 - y_0\| < \frac{\lambda}{\delta}$. Since this holds for all $x \in W$, we have $\text{dist}(x_\delta, W) \geq \delta$. □

15.2 Proposition If $(V, \| \cdot \|)$ is an infinite dimensional normed space then:

(i) Each closed ball $\overline{B}(x, r) = \{y \in V \ | \ \|x - y\| \leq r\}$ (with $r > 0$) is non-compact.

(ii) Every subset $Y \subseteq V$ with non-empty interior $Y^0$ is non-compact.

Proof. (i) Choose $x_1 \in V$ with $\|x_1\| = 1$. Then $C_1$ is a closed proper subspace, thus there exists $x_2 \in V$ with $\|x_2\| = 1$ and $\|x_1 - x_2\| \geq \frac{1}{2}$. Since $V$ is infinite dimensional, $V_2 = \text{span}\{x_1, x_2\}$ is a closed proper subspace, thus there exists $x_3 \in V$ with $\text{dist}(x_3, V_2) \geq \frac{1}{2}$, thus in particular $\|x_3 - x_i\| \geq \frac{1}{2}$ for $i = 1, 2$. Continuing in this way we can construct a sequence of $x_i \in V$ with $\|x_i\| = 1$ and $\|x_i - x_j\| \geq \frac{1}{2} \ \forall i \neq j$. The sequence $\{x_i\}$ clearly cannot have a convergent subsequence, thus the closed unit ball $\overline{B}(0, 1)$ is non-compact. Since $x \mapsto \lambda x + x_0$ with $\lambda > 0$ is a homeomorphism, all closed balls are non-compact.

(ii) If $Y \subseteq V$ and $Y^0 \neq \emptyset$ then $Y$ contains some open ball $B(x, r)$, thus also $\overline{B}(x, r/2)$, which is non-compact. Thus neither $Y$ nor $Y^c$ are compact. □

In view of the above, it is interesting to look at linear operators that send sets $S \subseteq V$ to sets $AS$ with ‘better compactness properties’. There are several such notions:

15.3 Exercise Let $V$ be a normed space and $A : V \rightarrow V$ a linear map. Prove that the following conditions are equivalent and imply boundedness of $A$:

(i) The image $A\overline{B} \subseteq V$ of the closed unit ball $\overline{B} = V_{\leq 1}$ is precompact.

(ii) $AS$ is precompact whenever $S \subseteq V$ is bounded.

(iii) Given any bounded sequence $\{x_n\} \subseteq V$, the sequence $\{Ax_n\}$ has a convergent subsequence.

15.4 Definition Operators $A \in B(V)$ satisfying the above equivalent conditions are called compact (or completely continuous). The set of compact operators on $V$ is denoted $K(V)$. 89
15.5 Remark 1. Compactness for $A \in B(V,W)$ with $V \neq W$ is defined completely analogously.
2. Some authors write $B_0(V)$ rather than $K(V)$, motivated by Exercise 15.11(iii) below.
3. If $A \in B(V)$ is compact and $W \subseteq V$ is a closed $A$-invariant subspace, thus $AW \subseteq W$, then the restriction $A|_W$ is compact.
4. The Heine-Borel theorem implies that every linear operator on a finite dimensional normed space (automatically bounded by Exercise 4.14) is compact. For infinite dimensional spaces this is false since every closed ball is bounded but non-compact by Proposition 15.2. In particular the unit operator $1_V$ is compact if and only if $V$ is finite dimensional.
5. Compactness can also be defined for non-linear maps between Banach spaces. But then the three versions above a no more equivalent and continuity is no more automatic. See Section B.9. \hfill \Box

Before we develop further theory, we should prove that (non-zero) compact operators on infinite dimensional spaces exist.

15.6 Definition Let $V$ be a normed space and $A \in B(V)$. Then $A$ has finite rank if its image $AV$ is finite dimensional. The set of finite rank operators on $V$ is denoted $F(V)$.

For example, if $\varphi \in V^*$, $y \in V$ then $A \in B(V)$ define by $A : x \mapsto \varphi(x)y$ has finite rank.

15.7 Lemma For each Banach space $V$, $F(V) \subseteq K(V)$.

Proof. Let $A \in F(V)$. If $S \subseteq V$ is bounded then $AS \subseteq AV$ is bounded by boundedness of $V$. Since $AV$ is finite dimensional, is has the Heine-Borel property so that $\overline{AS} \subseteq AV$ is compact. Thus $A$ is compact. \hfill \Box

15.8 Lemma $K(V) \subseteq B(V)$ is a two-sided ideal (thus a linear subspace, and if $A \in B(V)$, $B \in K(V)$ then $AB, BA \in K(V)$).

Proof. Let $\{x_n\}$ be a bounded sequence in $V$. Since $A, B$ are compact, we can find a subsequence $\{x_{n_k}\}_{k \in \mathbb{N}}$ such that $Ax_{n_k}$ and $Bx_{n_k}$ converge as $k \to \infty$. Then also $(cA + dB)x_{n_k}$ converges, thus $cA + dB$ is compact. Thus $K(V) \subseteq B(V)$ is a linear subspace.

Alternative argument: Let $A, B \in K(V)$ and $S \subseteq V$ bounded. Then $\overline{AS}$ and $\overline{BS}$ are compact, and so are $c\overline{AS}$, $d\overline{BS}$ if $c, d \in \mathbb{F}$. Thus also $c\overline{AS} + d\overline{BS}$ is compact (by joint continuity of the map $+: V \times V \to V$), thus also $(cA + dB)S \subseteq c\overline{AS} + d\overline{BS} = \overline{cAS + dBS}$.

Now let $A \in B(V)$, $B \in K(V)$ and $S \subseteq V$ bounded. Then $\overline{BS}$ and $\overline{ABS} = \overline{ABS}$ are compact by compactness of $B$ and continuity of $A$, respectively. And boundedness of $A$ implies boundedness of $AS$, so that $B\overline{AS}$ has compact closure by compactness of $B$. Thus $AB$ and $BA$ are compact. \hfill \Box

For the proof of the next result, we need the notion of total boundedness in metric spaces, see Appendix A.8. In particular we will use Exercise A.36(iii).

15.9 Proposition For each Banach space $V$, $K(V) \subseteq B(V)$ is $\| \cdot \|$-closed.

Proof. Since $B(V)$ is a metric space, it suffices to prove that the limit $A$ of every norm-convergent sequence $\{A_n\}$ in $K(V)$ is in $K(V)$. Let thus $\{A_n\} \subseteq K(V)$ and $A \in B(V)$ such that $\|A_n - A\| \to 0$. We want to prove that $AB \subseteq V$ is precompact, where $\overline{B}$ is the closed unit ball. Since $(B(V), \| \cdot \|)$ is complete, by Exercise A.36(iii) this is equivalent to $AB\overline{B}$ being totally bounded. To show this, let $\varepsilon > 0$. Then there is an $n$ such that $\|A_n - A\| < \varepsilon/3$. Since $A_n$ is
compact, $A_n \overline{B}$ is precompact, thus totally bounded. Thus there are $x_1, \ldots, x_n \in \overline{B}$ such that  
\[ \bigcup_{i=1}^{n} B(A_n x_i, \varepsilon/3) \supseteq A_n \overline{B}. \]
If now $x \in \overline{B}$ (thus $\|x\| \leq 1$) then there is $i \in \{1, \ldots, n\}$ such that  
\[ \|A_n x - A_n x_i\| < \varepsilon/3. \]
Thus  
\[ \|Ax - Ax_i\| \leq \|Ax - A_n x\| + \|A_n x - A_n x_i\| + \|A_n x_i - Ax_i\| < \varepsilon/3 + \varepsilon/3 + \varepsilon = \varepsilon, \]
where we used $\|A - A_n\| < \varepsilon/3$ and $x, x_i \in \overline{B}$. This proves that $A \overline{B}$ is totally bounded, thus precompact. Thus $A$ is compact. \qed

15.10 Corollary For each Banach space $V$, we have $\overline{F(V)} \subseteq K(V)$.

15.11 Exercise Let $V = \ell^p(S, \mathbb{C})$, where $S$ is an infinite set and $1 \leq p < \infty$. If $g \in \ell^\infty(S, \mathbb{C})$ and $f \in \ell^p(S, \mathbb{C})$ then $M_g(f) = gf$ (pointwise product) is in $\ell^p(S)$. This defines a linear map $\ell^\infty(S, \mathbb{C}) \to \ell^p(V)$, $g \to M_g$. Prove:

(i) $g \mapsto M_g$ is an algebra homomorphism.
(ii) $\|M_g\| = \|g\|_\infty$.
(iii) $M_g \in K(V)$ if and only if $g \in c_0(S, \mathbb{C})$.

15.12 Remark 1. We now have two classes of compact operators: The (rather commutative) one of multiplication by $c_0$-functions, and the operators that are norm-limits of finite rank operators. Actually, the first class is contained in the second. Why?

2. It is quite natural to ask whether in fact $\overline{F(V)} = K(V)$, i.e. whether all compact operators on $V$ are norm-limits of finite rank operators. When this holds, $V$ is said to have the approximation property. We will later see that this is true for all Hilbert spaces. Whether all Banach spaces have the approximation property was an open problem until Enflo\textsuperscript{43} in 1973 [18] constructed a counterexample. His construction was very complicated and his spaces were not very ‘natural’ (in the sense of having a simple definition and/or having been encountered previously). A simpler example, but still tricky and not natural, can be found in [12]. Somewhat later, very natural examples were found: The Banach space $B(H)$ does not have the approximation property whenever $H$ is an infinite dimensional Hilbert space, cf. [76]. (Note that this is about compact operators on $B(H)$, not compact operators in $B(H)$!) All this is well beyond the level of this course, but you should be able to understand [12]. \qed

None of the above examples of compact operators seems very relevant for applications, even within mathematics. Indeed the most useful compact operators perhaps are integral operators. We will briefly look at a class of them in Exercise 15.35. But there are very simple examples:

15.13 Definition Let $V = C([0, C], \mathbb{F})$ for some $C > 0$, equipped with the norm $\|f\| = \sup_{x \in [0, C]} |f(x)|$. As we know, $(V, \|\cdot\|)$ is a Banach space. Define a linear operator, the Volterra\textsuperscript{44} operator, by  
\[ A : V \to V, \quad (Af)(x) = \int_0^x f(t)dt. \]

We have $\|Af\| = \sup_x |\int_0^x f(t)dt| \leq \int_0^C |f(t)|dt \leq C\|f\|$, thus $\|A\| \leq C < \infty$.

15.14 Proposition The Volterra operator $A : V \to V$ is compact.

\textsuperscript{43}Per H. Enflo (1944-). Swedish mathematician, working mostly in functional analysis.
\textsuperscript{44}Vito Volterra (1860-1940). Italian mathematician and one of the early pioneers of functional analysis.
The proof of this result makes essential use of the Arzelà-Ascoli Theorem\textsuperscript{45} which characterizes the (pre)compact subsets of $(C(X, \mathbb{F}), \| \cdot \|_\infty)$ for compact $X$.

\textbf{Proof.} We will prove that $\mathcal{F} = \overline{A B} \subseteq V$ is precompact by showing that it satisfies the hypotheses of Theorem A.38. If $x \in [0, C]$ and $f \in C([0, C])$ with $\|f\| \leq 1$ then

\[ |(Af)(x)| = \left| \int_0^x f(t)dt \right| \leq C \|f\| \leq C < \infty, \]

showing that $\mathcal{F}$ is pointwise bounded. For each $f \in V$ with $\|f\| \leq 1$ we have

\[ |(Af)(x) - (Af)(y)| = \left| \int_x^y f(t)dt \right| \leq |x - y|. \]

Since this is uniform in $f$ it shows that $\mathcal{F}$ is equicontinuous.

\[ \Box \]

The above proof is very easily generalized to give compactness of $A : V \to V$ with $V = C([a, b], \mathbb{F})$ given by $(Af)(x) = \int_a^x K(x, y)f(y)dy$ for any $K \in C([a, b] \times [a, b], \mathbb{F})$.

15.15 \textbf{PROPOSITION} Let $A \in B(V)$. Then $A^t \in B(V^*)$ is compact if and only if $A$ is compact.

\textbf{Proof.} As in the proof of Proposition, we use the equivalence of precompactness and total boundedness from Exercise A.36(iii). Assume that $A \in B(V)$ is compact. Then $X = \overline{AV_{\leq 1}} \subseteq V$ is compact. Let

\[ \mathcal{F} = \{ \varphi : X \mid \varphi \in (V^*)_\leq 1 \} \subseteq C(X, \mathbb{F}). \]

For each $x \in X$ we have $\mathcal{F}x = \{ \varphi(x) \mid \varphi \in (V^*)_\leq 1 \}$, which clearly is bounded. Thus $\mathcal{F}$ is pointwise bounded. If $x, x' \in X$ then for each $\varphi \in \mathcal{F}$ we have $|\varphi(x) - \varphi(x')| \leq \|\varphi\| \|x - x'\| \leq \|x - x'\|$, so that $\mathcal{F} \subseteq C(X, \mathbb{F})$ is equicontinuous. Thus by the Arzelà-Ascoli Theorem A.38, $\mathcal{F}$ is totally bounded. This means that for each $\varepsilon > 0$ there are $\varphi_1, \ldots, \varphi_N \in (V^*)_{\leq 1}$ such that for every $\varphi \in (V^*)_{\leq 1}$ there is an $i$ such that $\|\varphi - \varphi_i\|_{C(X, \mathbb{F})} = \sup_{x \in X} |\varphi(x) - \varphi_i(x)| < \varepsilon$. In view of $X = \overline{AV_{\leq 1}}$, this implies: For every $\varepsilon > 0$ there are $\varphi_1, \ldots, \varphi_N \in (V^*)_{\leq 1}$ such that for each $\varphi \in (V^*)_{\leq 1}$ there is an $i$ such that $\sup_{x \in V_{\leq 1}} |\varphi(Ax) - \varphi_i(Ax)| < \varepsilon$. In view of

\[ \sup_{x \in V_{\leq 1}} |\varphi(Ax) - \varphi_i(Ax)| \leq \sup_{x \in V_{\leq 1}} |(A^t \varphi - A^t \varphi_i)(x)| = \|A^t \varphi - A^t \varphi_i\|, \]

we have proven the total boundedness (=precompactness) of the set $A^t(V^*)_{\leq 1} \subseteq V^*$ and therefore compactness of $A^t \in B(V^*)$.

Now assume that $A^t \in B(V^*)$ is compact. Then by the above, $A^{tt} \in B(V^{**})$ is compact. Since $V \subseteq V^{**}$ is a closed subspace, the restriction $A^{tt}|_V$ is compact. But by Lemma 12.2, the latter equals $A$, so that $A$ is compact.

\[ \Box \]

15.16 \textbf{REMARK} The above result (due to Schauder) can be proven in different ways: One can give essentially the same proof avoiding invocation of Arzelà-Ascoli, cf. [49, Theorem 1.4.4], or use the circle of ideas in Section 17 as in [11, Theorem VI.3.4]. For another functional analysis proof see [64]. Cf. also Remark A.40.2 on different proofs of the Arzelà-Ascoli theorem.

\[ \Box \]

For Hilbert spaces, we have a simple proof using polar decomposition:

\textsuperscript{45}You should have seen this theorem in Analysis 2 or Topology. See e.g. Appendix A.9 or [22, Vol. 2, Theorem 15.5.1]. It has many applications in classical analysis, for example Peano’s existence theorem on differential equations.
15.17 Proposition If $H$ is a Hilbert space and $A \in B(H)$ is compact then $A^*$ and $|A|$ are compact.

Proof. Compactness of $A^*$ follows from Proposition 15.15 and $A^* = \gamma^{-1} \circ A^\dagger \circ \gamma$, where $\gamma : H \rightarrow H^*$, $y \mapsto \langle \cdot, y \rangle$.

For $|A|$ (and also $A^*$, if we want to avoid Proposition 15.15) we argue as follows: By the polar decomposition, there is a partial isometry $V$ such that $A = V |A|$ and $|A| = V^* A$. The second identity together with compactness of $A$ and Lemma 15.8 gives compactness of $|A|$. Since the first identity is equivalent to $A^* = |A| V^*$, the compactness of $A^*$ follows.

Yet another proof: If $A \in K(H)$ and $\varepsilon > 0$ then by Corollary 15.30 proven below there is $F \in F(H)$ with $\|A - F\| < \varepsilon$, thus $\|A^* - F^*\| < \varepsilon$. By the following exercise, $F^*$ is finite rank. Since $\varepsilon > 0$ was arbitrary, Corollary 15.10 gives $A^* \in K(H)$. ■

15.18 Exercise Prove: If $H$ is a Hilbert space and $A \in F(H)$ then $A^* \in F(H)$.

Thus the closed ideal $K(H) \subseteq B(H)$ is closed under the $*$-operation, i.e. self-adjoint, and of course a $C^*$-algebra. One can prove, see e.g. [49], that $K(H)$ it is the smallest self-adjoint closed ideal in $B(H)$. If $H$ is separable, it is the only one.

15.2 Fredholm alternative. The spectrum of compact operators

We now begin studying the spectrum of compact operators.

15.19 Lemma If $A \in B(V)$ is compact and $\lambda \in \mathbb{C} \setminus \{0\}$ then $\ker(A - \lambda I)$ is finite dimensional.

Proof. If $\lambda \notin \sigma(A)$ then this is trivial since $A - \lambda I$ is invertible. In general, $V_\lambda = \ker(A - \lambda I)$ is the space of eigenvectors of $A$ with eigenvalue $\lambda$. Clearly $A|_{V_\lambda} = \lambda id_{V_\lambda}$, so that $V_\lambda$ is an invariant subspace. Since $V_\lambda$ is closed and $A|_{V_\lambda}$ is compact by Remark 15.5.3, $V_\lambda$ must be finite-dimensional by Remark 15.5.4. ■

15.20 Proposition (Fredholm alternative) 46 Let $V$ be a Banach space, $A \in B(V)$ compact and $\lambda \in \mathbb{C} \setminus \{0\}$. Then the following are equivalent:

(i) $A - \lambda I$ is invertible. (I.e. $\lambda \notin \sigma(A)$.)

(ii) $A - \lambda I$ is injective.

(iii) $A - \lambda I$ is surjective.

Proof. We know from Proposition 10.24 that (i) is equivalent to the combination of (ii) and (iii). It therefore suffices to prove (ii)$\iff$(iii).

(iii)$\implies$(ii): It suffices to do this for $\lambda = 1$. (Why?) Assume that $A - 1$ is not injective, but surjective. Clearly $(A - 1)^n$ is surjective for all $n$. In view of $(A - 1)^{n+1} = (A - 1)(A - 1)^n$ we have $\ker(A - 1)^n = \ker(A - 1) \setminus \{0\}$, where the $-1$ stands for ‘preimage’. This space clearly contains $\ker(A - 1)^n$, but it is strictly larger since it contains vectors $x$ such that $(A - 1)^nx \in \ker(A - 1) \setminus \{0\}$. Thus $\ker(A - 1)^{n+1} \supseteq \ker(A - 1)^n$ for all $n$. Now by Riesz’ Lemma 15.1, for each $n$ we can find an $x_n \in \ker(A - 1)^{n+1}$ such that $\|x_n\| = 1$ and $\|Ax_n - Ax_m\| \geq \frac{1}{2}$. If $n > m$ then $(A - 1)x_n - Ax_m \in \ker(A - 1)^n$ (note that $(A - 1)^n$ commutes with $A$ and $x_m \in \ker(A - 1)^n$ since $n > m$), so that

$$\|Ax_n - Ax_m\| = \|x_n + ((A - 1)x_n - Ax_m)\| \geq \frac{1}{2}.$$ 

---


93
Thus \( \{Ax_n\} \) has no convergent subsequence, contradicting the compactness of \( A \).

(ii) \( \Rightarrow \) (iii): If \( A - \lambda \mathbf{1} \) is injective but not surjective, one similarly proves \( (A - \lambda \mathbf{1})^{n+1}H \subseteq (A - \lambda \mathbf{1})^nH \) for all \( n \), which again leads to a contradiction with compactness of \( A \).  

\[ \square \]

15.21 Remark In the above proof we have shown that if \( A \in B(V) \) is compact and \( \lambda \in \mathbb{C}\setminus\{0\} \) then we cannot have \( \ker((A - \mathbf{1})^{n+1}) \nsubseteq \mathbb{C}\{0\} \) for all \( n \) or \( (A - \lambda \mathbf{1})^{n+1}H \nsubseteq (A - \lambda \mathbf{1})^nH \) for all \( n \). One says that \( A - \lambda \mathbf{1} \) has finite ascent and descent.  

\[ \square \]

Fredholm’s alternative has far-reaching consequences for the spectrum of \( A \):

15.22 Corollary If \( A \in B(V) \) is compact then \( \sigma(A) \subseteq \sigma_p(A) \cup \{0\} \).

Proof. For \( \lambda \neq 0 \), with Proposition 15.20 we have: \( \lambda \in \sigma(A) \iff A - \lambda \mathbf{1} \) not invertible \( \iff A - \lambda \mathbf{1} \) not injective \( \iff \lambda \in \sigma_p(A) \).

\[ \square \]

15.23 Proposition Let \( H \) be a non-zero complex Hilbert space and \( A \in B(H) \) a compact normal operator. Then there is an eigenvalue \( \lambda \in \sigma_p(A) \) such that \( |\lambda| = \|A\| \).

Proof. If \( A = 0 \) then it is clear that \( \lambda = 0 \) does the job. Now assume \( A \neq 0 \). By Exercise 12.25, there is \( \lambda \in \sigma(A) \) with \( |\lambda| = \|A\| \). Since \( \lambda \neq 0 \), Corollary 15.22 gives \( \lambda \in \sigma_p(A) \).

\[ \square \]

The rest of this subsection is not needed for the proof of the spectral theorem, but puts the Fredholm alternative into perspective.

If \( A : V \to W \) is a linear map, the cokernel of \( A \) by definition is the linear quotient space \( W/AV \). (If \( V,W \) are Banach spaces and \( AV \subseteq W \) is closed, then \( W/AV \) is Banach.)

If \( V,W \) are Hilbert spaces and \( AV \subseteq W \) is closed then, recalling Exercise 7.1 we may alternatively define the cokernel of \( A \) to be \( (AV)^\perp \subseteq W \).

15.24 Proposition Let \( A \in B(V) \) be compact and \( \lambda \in \mathbb{C}\setminus\{0\} \). Then

(i) \( (A - \lambda \mathbf{1})V \subseteq V \) is closed.

(ii) \( \text{coker}(A - \lambda \mathbf{1}) \) is finite dimensional.

Proof. (i) By Lemma 15.19, \( K = \ker(A - \lambda \mathbf{1}) \) is finite dimensional, thus closed by Exercise 4.22. Thus there is a closed subspace \( S \subseteq V \) such that \( V = K \oplus S \). (If \( V \) is a Hilbert space, we can just take \( S = K^\perp \). For general Banach spaces this is the statement of Proposition 7.14.) The restriction \( (A - \lambda \mathbf{1})|_S : S \to H \) is compact and injective. We will prove that it is bounded below. Lemma 10.22 then gives that \( (A - \lambda \mathbf{1})H = (A - \lambda \mathbf{1})S \) is closed. If \( (A - \lambda \mathbf{1})|_S \) is not bounded below, we can find a sequence \( \{x_n\} \) in \( S \) with \( \|x_n\| = 1 \) for all \( n \) and \( \|(A - \lambda \mathbf{1})x_n\| \to 0 \). Since \( A \) is compact, we can find a subsequence \( \{x_{n_k}\} \) such that \( \{Ax_{n_k}\} \) converges. We relabel, so that now \( \{Ax_n\} \) converges. Now

\[
x_n = \lambda^{-1}[Ax_n - (A - \lambda \mathbf{1})x_n].
\]

Since \( \{Ax_n\} \) converges and \( \{(A - \lambda \mathbf{1})x_n\} \) converges to zero by choice of \( \{x_n\} \), \( \{x_n\} \) converges to some \( y \in S \) (since \( x_n \in S \forall n \) and \( S \) is closed). From \( (A - \lambda \mathbf{1})x_n \to 0 \) and \( x_n \to y \) we obtain \( (A - \lambda \mathbf{1})y = 0 \), so that \( y \in \ker(A - \lambda \mathbf{1}) = K \). Thus \( y \in K \cap S = \{0\} \), which is impossible since \( y = \lim_n x_n \) and \( \|x_n\| = 1 \forall n \). This contradiction shows that \( (A - \lambda \mathbf{1})|_S \) is bounded below.

(ii) By Proposition 15.15, \( A^t \) is compact, thus \( \ker(A^t - \lambda \mathbf{1}_{V^*}) \) is finite dimensional by Lemma 15.19. Now

\[
\ker(A^t - \lambda \mathbf{1}_{V^*}) = \ker(A - \lambda \mathbf{1}_V)^t = ((A - \lambda \mathbf{1}_V)V^\perp \cong (V/((A - \lambda \mathbf{1}_V)V))^*,
\]

94
where the second equality and the isomorphism come from Exercise 12.3(i) and Exercise 7.6, respectively. Thus \((V/(A - \lambda_1 V)V)^*\) is finite dimensional, implying (why?) finite dimensionality of \(V/(A - \lambda_1 V)V = \ker(A - \lambda_1 V)\).

If \(H\) is a Hilbert space and we define \(\text{coker}A\) as \((A^H)^\perp\), we can also argue as follows:

\[
\ker(A - \lambda 1) = \ker((A - \lambda 1)H)^\perp = \ker(A - \lambda 1)^* = \ker(A^* - \overline{\lambda} 1),
\]

where the second equality is Lemma 14.1. Now \(A^*\) is compact by Proposition 15.17, thus Lemma 15.19 gives finite dimensionality of \(\ker(A^* - \overline{\lambda} 1)\).

\[
\text{dim ker}(A - \lambda 1) = \text{dim coker}(A - \lambda 1). \tag{15.1}
\]

Compare this with the equivalence (ii)\(\iff\)(iii) in Proposition 15.20, which amounts to the weak statement \(\text{dim ker}(A - \lambda 1) = 0 \iff \text{dim coker}(A - \lambda 1) = 0\). For proofs of (15.1) in a Banach space context see any of [38, 49, 42].

2. If \(V\) is a Banach space then \(A \in B(V)\) is called a Fredholm operator if \(\ker A\) and \(\text{coker} A\) are finite dimensional. (Often it is also assumed that \(A\) has closed image, but by Exercise 10.10 this follows from finite dimensionality of the cokernel.) In this case, one calls \(\text{ind}(A) = \text{dim ker} A - \text{dim coker} A \in \mathbb{Z}\) the (Fredholm) index of \(A\). If \(A, B\) are both Fredholm then so is \(AB\) and \(\text{ind}(AB) = \text{ind}(A) + \text{ind}(B)\). (This can be used for proving (15.1).)

Thus if \(A\) is compact and \(\lambda \neq 0\) then \(A - \lambda 1\) is Fredholm with index zero.

Since \(1\) is Fredholm with index zero, this result is a very special case of the following: If \(F\) is Fredholm and \(K\) is compact then \(F + K\) is Fredholm and \(\text{ind}(F + K) = \text{ind}(F)\).

Another important connection between compact and Fredholm operators is Atkinson’s theorem: \(A \in B(V)\) is Fredholm if and only there exists \(B \in B(V)\) such that \(AB - 1\) and \(BA - 1\) are compact. (Equivalently, the image of \(A\) in the quotient algebra \(B(V)/K(V)\) is invertible.) For more on Fredholm operators see [54, p.110-112] or [49, Section 1.4].

Finally, \(\lambda \in \sigma_d(A)\) is equivalent to \(\lambda \in \sigma(A)\) being isolated and \(A - \lambda 1\) being Fredholm (equivalently, Fredholm of index zero).

\[\square\]

15.26 Exercise Let \(H\) be a Hilbert space, \(A \in K(H)\) and \(\lambda \in \mathbb{C}\backslash\{0\}\). Show that each of the implications (ii)\(\Rightarrow\)(iii) and (iii)\(\Rightarrow\)(ii) in Proposition 15.20 can be deduced from the other.

15.3 Spectral theorems for compact Hilbert space operators

It is well-known that self-adjoint \(n \times n\)-matrices can be diagonalized. The following beautiful result generalizes this to compact normal operators:

15.27 Theorem (Spectral theorem for compact normal operators) Let \(H\) be a Hilbert space and \(A \in B(H)\) compact normal. Then

(i) \(H\) is spanned by the eigenvectors of \(A\).

(ii) There is an ONB \(E\) of \(H\) consisting of eigenvectors, thus \(A = \sum_{e \in E} \lambda_e P_e\), where \(P_e : x \mapsto (x, e)e\).

(iii) For each \(\varepsilon > 0\) there are at most finitely many \(\lambda \in \sigma_p(A)\) with \(|\lambda| \geq \varepsilon\).

(iv) \(\sigma_p(A)\) is at most countable and has no accumulation points except perhaps 0.

(v) We have \(\sigma(A) \subseteq \sigma_p(A) \cup \{0\}\). Furthermore,

95
- $0 \in \sigma_p(A)$ if and only if $A$ is not injective.
- $0 \in \sigma_c(A)$ if and only if $A$ is injective and $H$ is infinite dimensional.

Proof. (i) Let $K \subseteq H$ be the smallest closed subspace containing $\bigcup_{\lambda \in \sigma_p(A)} H_\lambda$, where $H_\lambda = \ker(A - \lambda I)$. Clearly $K$ is an invariant subspace: $AK \subseteq K$. Exercise 14.8(i) implies that also $A^*K \subseteq K$. Now Exercise 14.13(i) gives that also $K^\perp$ is $A$-invariant: $AK^\perp \subseteq K^\perp$. Clearly $A_{|K^\perp}$ is compact, thus if $K^\perp \neq \{0\}$ then by Proposition 15.23 it contains an eigenvector of $A$. Since this contradicts the definition of $K$, we have $K^\perp = 0$, proving that $H$ is spanned by the eigenvectors of $A$.

(ii) By Exercise 14.8(ii), the eigenspaces for different eigenvalues of $A$ are mutually orthogonal. Now the claim follows from (i) by choosing ONBs $E_\lambda$ for each $H_\lambda$ and putting $E = \bigcup E_\lambda$.

(iii) Taking into account the unitary equivalence $H \cong \ell^2(E, \mathbb{C})$, cf. Theorem 6.40, this essentially is Exercise 15.11(iii).

(iv) This is an immediate consequence of (iii).

(v) The first statement was already proven in Corollary 15.22. The second is immediate. As to the last, if $H$ is finite dimensional then $0 \notin \sigma_p(A) = \sigma(A)$. If $H$ is infinite dimensional and $A$ injective then $\sigma_p$ is infinite since the eigenspaces for the $\lambda \neq 0$ are finite dimensional and span $H$. Thus in view of (iii), we have $0 \in \sigma_p(A)$. Now $0 \in \sigma_c(A)$ follows from the fact that $A$ is not bounded below or from the closedness of $\sigma(A)$. (And recall that $\sigma(A) = \emptyset$ by normality.) □

15.28 REMARK 1. The common theme of `spectral theorems' is that normal operators can be diagonalized, i.e. be interpreted as multiplication operators, compactness simplifying statement and proof considerably.

2. The statements about $\sigma(A)$ actually hold for all compact operators on Banach spaces. (Instead of the orthogonality of eigenvectors for different eigenvalues, it suffices to use their linear independence.) □

For non-normal operators one can only prove a weaker statement:

15.29 PROPOSITION Let $A \in K(H)$. Then there are ONBs $E$ and $F = \{f_e\}_{e \in E}$ and non-negative numbers $\{\beta_e\}$, called the singular values of $A$, such that $e \mapsto \beta_e$ is in $c_0(E, \mathbb{C})$ and

$$A = \sum_{e \in E} \beta_e f_e \langle \cdot, e \rangle.$$ 

The $\beta_e$ are the precisely the eigenvalues of $|A|$.

Proof. $B = A^*A$ is compact and self-adjoint, so that there is an ONB $E$ diagonalizing $B$, thus $A = \sum_{e \in E} \lambda_e P_e$. Let $E' = \{e \in E \mid Ae \neq 0\}$. For $e \in E'$ put $f_e = \frac{Ae}{\|Ae\|}$. Now let $F' = \{f_e \mid e \in E'\}$. Clearly

$$Ax = A \sum_{e \in E} \langle x, e \rangle e = \sum_{e \in E} \langle x, e \rangle Ae = \sum_{e \in E'} \langle x, e \rangle Ae = \sum_{e \in E'} \|Ae\| \langle x, e \rangle f_e$$

for all $x \in H$. If $e, e' \in E, e \neq e'$ then $\langle Ae, Ae' \rangle = \langle e, A^*Ae' \rangle = 0$ since $E$ diagonalizes $A^*A$. Thus the $f_e = \frac{Ae}{\|Ae\|}$ are mutually orthogonal, and they are obviously normalized. Thus $F'$ is orthonormal. Since $E'$ and $F'$ have the same cardinality, an ONB $F''$ for $(\text{span}_F F')^\perp$ has the same cardinality as $E \setminus E'$. Thus we can extend the bijection $E' \rightarrow F'$, $e \rightarrow f_e$ to a bijection $E \rightarrow F = F' \cup F''$ (in arbitrary fashion). For all $e \in E$ put $\beta_e = \|Ae\|$. Now $A = \sum_{e \in E} \beta_e f_e \langle \cdot, e \rangle$. 

96
where the exchange of summation is justified since all summands are non-negative.

15.30 Corollary Let $H$ be a Hilbert space, $A \in K(H)$ and $\varepsilon > 0$. Then there is a $B \in F(H)$ (finite rank) with $\|A - B\| \leq \varepsilon$. Thus $K(H) = \overline{F(H)}^{\|\cdot\|}$.

Proof. Pick a representation $A = \sum_{e \in E} \lambda_e \langle \cdot, f_e \rangle e$ as in the preceding proposition. Since $E \rightarrow \mathbb{C}$, $e \mapsto \lambda_e$ is in $c_0(E)$, there is a finite subset $F \subseteq E$ such that $|\lambda_e| < \varepsilon$ for all $e \in E \setminus F$. Define

$$B = \sum_{e \in F} \lambda_e \langle \cdot, f_e \rangle e,$$

which clearly has finite rank. If $x \in H$ then using the orthonormality of $E$, we have

$$\| (A - B)x \|^2 = \| \sum_{e \in E \setminus F} \lambda_e \langle x, f_e \rangle e \|^2 = \sum_{e \in E \setminus F} |\lambda_e \langle x, f_e \rangle|^2 \leq \varepsilon^2 \|x\|^2.$$

Thus $\|A - B\| \leq \varepsilon$, so that $K(H) \subseteq \overline{F(H)}$. The converse inclusion was Corollary 15.10. □

15.31 Remark 1. In the above, bases played a crucial role. Even though there is no notion of orthogonality in general Banach spaces, it turns out that Banach spaces having suitable bases do satisfy $K(H) = \overline{F(H)}^{\|\cdot\|}$, i.e. the approximation property.

2. If you like applications of complex analysis to functional analysis, see [58, Section VI.5] for an interesting alternative approach to compact operators. \qed

15.4 ** Hilbert-Schmidt operators

Generalizing a well-known construction in (finite dimensional) linear algebra, if $A \in B(H)$ is positive and $E$ is an orthonormal base of $H$, we define the ‘trace’ of $A$ w.r.t. $E$ by

$$\text{Tr}_E(A) = \sum_{e \in E} \langle Ae, e \rangle \in [0, \infty].$$

15.32 Lemma (i) For every $A \in B(H)$ we have $\text{Tr}_E(A^*A) = \text{Tr}_E(AA^*)$.

(ii) If $A \geq 0$ and $U$ is unitary then $\text{Tr}_E(UAU^*) = \text{Tr}_E(A)$.

(iii) If $A \geq 0$ then $\text{Tr}_E(A)$ is independent of the ONB $E$. We therefore just write $\text{Tr}(A)$.

(iv) For $A, B \in B(H)^+$ and $\lambda \geq 0$ we have $\text{Tr}(A + B) = \text{Tr}(A) + \text{Tr}(B)$ and $\text{Tr}(\lambda A) = \lambda \text{Tr}(A)$.

Proof. (i) Using Parseval ($\|x\|^2 = \sum_{e' \in E} |\langle x, e' \rangle|^2$), we have

$$\text{Tr}_E(A^*A) = \sum_{e \in E} \langle A^*Ae, e \rangle = \sum_{e \in E} \langle Ae, Ae \rangle = \sum_{e \in E} \|Ae\|^2 = \sum_{e \in E} \sum_{e' \in E} |\langle Ae, e' \rangle|^2 = \sum_{e' \in E} \sum_{e \in E} |\langle A^*e', e \rangle|^2 = \sum_{e' \in E} \|A^*e'\|^2 = \text{Tr}_E(AA^*),$$

where the exchange of summation is justified since all summands are non-negative.
(ii) Put $B = U A^{1/2}$. By (i), $\text{Tr}_E(A) = \text{Tr}_E(B^* B) = \text{Tr}_E(B B^*) = \text{Tr}_E(U A U^*)$.

(iii) Let $A \geq 0$ and let $E, F$ be ONBs for $H$. Since $E, F$ have the same cardinality, we can pick a bijection $\alpha : F \to E$. The latter extends to a unitary operator $U : H \to H$. Thus by (ii),

$$\text{Tr}_E(A) = \text{Tr}_E(U A U^*) = \sum_{e \in E} \langle A U^* e, U^* e \rangle = \sum_{f \in F} \langle A f, f \rangle = \text{Tr}_F(A).$$

(iv) The second statement is evident, and the first follows from the fact that a sum of non-negative numbers is independent of the order or bracketing. 

We have seen that $A^* A$ is positive for any $A \in B(H)$, so that $\text{Tr}(A^* A) \in [0, \infty]$ is well-defined. For each $A \in B(H)$ we define

$$\|A\|_2 = (\text{Tr}(A^* A))^{1/2} \in [0, \infty].$$

15.33 Definition An operator $A \in B(H)$ is called Hilbert-Schmidt\footnote{Erhard Schmidt (1876-1959). Baltic German mathematician, contributions to functional analysis like Gram-Schmidt orthogonalization.} operator if $\|A\|_2 < \infty$. The set of Hilbert-Schmidt operators is denoted $L^2(H)$.

15.34 Proposition Let $H$ be a Hilbert space. Then

(i) $\|A\| \leq \|A\|_2 = \|A^*\|_2$ for all $A \in B(H)$. Thus $L^2(H)$ is self-adjoint.

(ii) For every $A, B \in L^2(H)$,

$$\langle A, B \rangle_{HS} = \text{Tr}(B^* A) = \sum_e \langle B^* A e, e \rangle$$

is absolutely convergent and independent of the ONB $E$. Now $\langle \cdot, \cdot \rangle_{HS}$ is an inner product on $L^2(H)$ such that $\langle A, A \rangle_{HS} = \|A\|_2^2$. And $(L^2(H), \langle \cdot, \cdot \rangle_{HS})$ is complete, thus a Hilbert space.

(iii) For all $A, B \in B(H)$ we have $\|A B\|_2 \leq \|A\| \|B\|_2$ and $\|A B\|_2 \leq \|A\|_2 \|B\|_2$. Thus $L^2(H) \subseteq B(H)$ is a two-sided ideal.

(iv) We have $F(H) \subseteq L^2(H) \subseteq K(H)$ and $\overline{F(H)} \|_{\| \cdot \|_2} = L^2(H)$.

Proof. (i) If $x \in H$ is a unit vector, pick an ONB $E$ containing $x$. Then $\|A x\|^2 = (A^* A x, x) \leq \text{Tr}_E(A^* A) = \|A\|_2^2$. Thus $\|A x\| \leq \|A\|_2$ whenever $\|x\| = 1$, proving the inequality. And Lemma 15.32(i) gives $\|A^*\|_2^2 = \text{Tr}(A^* A^*) = \text{Tr}(A^* A) = \|A\|_2^2$.

(ii) If $E$ is any ONB (whose choice does not matter) for $H$, we have (as before)

$$\text{Tr}(A^* A) = \sum_{e \in E} \langle A^* A e, e \rangle = \sum_{e \in E} \langle A e, A e \rangle = \sum_{e \in E} \|A e\|^2 = \sum_{e \in E} \sum_{e' \in E} |\langle A e, e' \rangle|^2.$$

Thus $L^2(H)$ is the set of $A \in B(H)$ for which the matrix elements $\langle A e, e' \rangle$ (w.r.t. the ONB $E$) are absolutely square summable. We therefore have a map

$$\alpha : L^2(H) \to \ell^2(E \times E), \ A \mapsto \{\langle A e, e' \rangle\}_{(e, e') \in E^2}$$
that clearly is injective. (Recall that $\ell^2(S) = L^2(S, \mu)$, where $\mu$ is the counting measure.) To show surjectivity of $\alpha$, let $f = \{f_{e'}\} \in \ell^2(E \times E)$. Define a linear operator $A : H \to H$ by $A : e \mapsto \sum_{e'} f_{ee'} e'$. For each $e$, the r.h.s. is in $H$ by square summability of $f$. If $x \in H$ then

$$\|Ax\|^2 = \sup_{E'} \left\| \sum_e \langle x, e \rangle \sum_{e' \in E'} f_{ee'} e' \right\|^2 = \sup_{E'} \left\| \sum_{e' \in E'} \left( \sum_e \langle x, e \rangle f_{ee'} \right) e' \right\|^2$$

$$= \sup_{E'} \sum_{e'} \left\| \sum_e \langle x, e \rangle f_{ee'} \right\|^2 \leq \|x\|^2 \sum_{e, e'} |f_{ee'}|^2,$$

where the supremum is over the finite subsets $E' \subseteq E$, we used $|\langle x, e \rangle| \leq \|x\|$ and the change of summation order is allowed due to the finiteness of $E'$. This computation shows that $\|A\| \leq (\sum_{e, e'} |f_{ee'}|^2)^{1/2} < \infty$. Thus $A \in B(H)$ and $\alpha(A) = f$, so that $\alpha$ is surjective. Thus $\alpha : L^2(H) \to \ell^2(E \times E)$ is a linear bijection. Now $\ell^2(E \times E)$ is a Hilbert space (in particular complete) with inner product $\langle f, g \rangle = \sum_{e, e'} f_{ee'} g_{ee'}$, and pulling this inner product back to $L^2(H)$ along $\alpha$ we have

$$\langle A, B \rangle_{HS} = \sum_{(e, e') \in E^2} \langle Ae, e' \rangle \langle Be, e' \rangle = \sum_{(e, e') \in E^2} \langle Ae, e' \rangle \langle e', Be \rangle$$

$$= \sum_e \langle Ae, Be \rangle = \sum_e \langle B^* Ae, e \rangle = \text{Tr}(B^* A),$$

where all sums converge absolutely. Lemma 15.32(ii) implies that $\langle A, A \rangle = \text{Tr}(A^* A)$ is independent of the chosen ONB, and for general $(A, B)$ this follows by the polarization identity.

From the above it is clear that $(L^2(H), \langle \cdot, \cdot \rangle_{HS})$ is isomorphic to the Hilbert space $(\ell^2(E \times E), \langle \cdot, \cdot \rangle)$, thus a Hilbert space. And the norm associated to $\langle \cdot, \cdot \rangle_{HS}$ is nothing other than $\|\cdot\|_2$, (iii) For any ONB $E$ we have

$$\|AB\|_2^2 = \text{Tr}(B^* A^* AB) = \sum_{e \in E} \|ABe\|^2 \leq \|A\|^2 \sum_{e \in E} \|Be\|^2 = \|A\|^2 \text{Tr}(B^* B) = \|A\|^2 \|B\|^2_2,$$

proving $\|AB\|_2 \leq \|A\| \|B\|_2$. And $\|AB\|_2 = \|(AB)^*\|_2 = \|B^* A^*\|_2 \leq \|B^*\| A\|_2^2 = \|A\|_2 \|B\|$, where we used the fact just proven and $\|A\|_2^2 = \|A\|_2$. The conclusion is obvious.

(iv) The inclusion $F(H) \subseteq L^2(H)$ is very easy and is left as an exercise. If $A \in L^2(H)$ and $F$ is a finite subset of the ONB $E$, define $p_F = \sum_{e \in E} \langle \cdot, e \rangle e$ and $A_F = Ap_F$. Then $A_F \in F(H)$ and $A$

$$\|A - A_F\|_2^2 = \|A(1 - p_F)\|_2^2 = \sum_{(e, e') \in (E \setminus F) \times E} |\langle Ae, e' \rangle|^2,$$

This implies $\|A - A_F\|_2 \to 0$ as $F \nearrow E$, so that $L^2(H) = \overline{F(H)}_{\|\cdot\|_2}$. Finally, by (i) we have $\|A - A_F\| \leq \|A - A_F\|_2 \to 0$, thus $A \in \overline{F(H)}_{\|\cdot\|} = K(H)$, where we used Corollary 15.10. This proves $L^2(H) \subseteq K(H)$. $\blacksquare$

15.35 Example ($L^2$-Integral operators) Let $(X, \mathcal{A}, \mu)$ be a measure space, and put $H = L^2(X, \mathcal{A}, \mu)$. Let $K : X \times X \to \mathbb{C}$ be measurable (w.r.t. the product $\sigma$-algebra $\mathcal{A} \times \mathcal{A}$) and assume $\int \int |K(x, y)|^2 \, d\mu(x) \, d\mu(y) < \infty$. (Thus $K \in L^2(X \times X, \mathcal{A} \times \mathcal{A}, (\mu \times \mu)$.) Then

$$\langle Kf, g \rangle = \int_X K(x, y) f(y) \, d\mu(y)$$

defines a linear operator $K : H \to H$ whose Hilbert-Schmidt norm $\|K\|_2$ coincides with the norm $\|K\|_{L^2}$ of $K \in L^2(X \times X)$. Thus $K$ is Hilbert-Schmidt, and in particular compact.
15.36 Exercise Prove the equality \( \|K\|_{L^2} = \|K\|_2 \) of norms claimed in the above example.

If \( V, W \) are vector spaces over any field \( k \) then there is a canonical linear map \( W \otimes_k V^* \to \text{Hom}_k(V, W) \) sending \( w \otimes_k \varphi \) to the linear map \( v \mapsto w\varphi(v) \). (Here \( V^* \) is the algebraic dual space and \( \otimes_k \) is the algebraic tensor product.) If \( V \) or \( W \) is finite dimensional, this map is a bijection, but otherwise it is not. For Hilbert spaces, one has a statement that works irrespective of the dimensions:

15.37 Exercise Let \( H \) be a Hilbert space. Make sense of the statement that \( L^2(H) \cong H \otimes H^* \) as Hilbert spaces. (This is quite similar to the above discussion of integral operators.)

15.38 Remark If \( A \in B(H) \) and \( 1 \leq p < \infty \) one puts \( \|A\|_p = (\text{Tr}(|A|^p))^{1/p} \). For \( p = 2 \) this agrees with our previous definition since \( |A|^2 = A^*A \), while for \( p = 1 \) one has \( \|A\|_1 = \text{Tr}|A| \). Now each space \( L^p(H) = \{ A \in B(H) \mid \|A\|_p < \infty \} \), the ‘\( p \)-th Schatten class’, is a two-sided ideal in \( B(H) \) and in fact \( L^p(H) \subseteq K(H) \) for all \( p \), see e.g. \([70]\). In particular \( L^1(H) \), the ‘trace-class operators’, play an important role in von Neumann algebra theory. The treatments of them in \([58]\) and \([37]\) are quite good. See also \([47]\). If \( 1 \leq p \leq q < \infty \), it is not hard to show that \( \|A\| := \|A\|_{\infty} \leq \|A\|_q \leq \|A\|_p \), thus \( L^q(H) \subseteq L^p(H) \subseteq K(H) \). Thus the spaces \( L^p(H) \) behave quite similarly to the \( \ell^p(S, \mathbb{F}) \), as also this exercise shows: \( \square \)

15.39 Exercise Given a set \( S \) and \( f \in \ell^\infty(S, \mathbb{F}) \), define \( H = \ell^2(S, \mathbb{C}) \) and the multiplication operator \( M_g : H \to H, \ f \mapsto gf \) (known from Exercise 15.11, where we saw \( M_g \in K(H) \iff g \in c_0(S) \)). Prove \( |M_g| = M_{|g|} \) and \( \|M_g\|_p = \|g\|_p \) for all \( p \in [1, \infty) \). (Thus \( M_g \in L^p(H) \iff g \in \ell^p(S, \mathbb{C}) \).)

16 Spectral theorems for normal Hilbert space operators

16.1 Spectral theorem: Multiplication operator version

In the remainder of this section we assume some knowledge of measure theory or willingness to learn some.

Let \( H \) be a Hilbert space, \( A \in B(H) \) normal and \( x \in H \). Then the map \( \varphi_{A,x} : C(\sigma(A), \mathbb{C}) \to \mathbb{C}, \ f \mapsto \langle f(A)x, x \rangle \) is a bounded linear functional on \( (C(\sigma(A), \mathbb{C}), \| \cdot \|) \). If \( f \) is positive (i.e. takes values in \([0, \infty)\)) then \( \sigma(f(A)) \subseteq [0, \infty) \) by the spectral mapping theorem, so that \( f(A) \geq 0 \) and \( \langle f(A)x, x \rangle \geq 0 \) by Proposition 14.17. Thus \( \varphi_{A,x} \) is a bounded positive linear functional on \( C(\sigma(A), \mathbb{C}) \). Thus by the Riesz-Markov-Kakutani theorem, cf. \([41]\, \text{Appendix A.5}\) for the statement and, e.g., \([10, \text{Theorem 7.2.8}] \) or \([61, \text{Theorem 2.14}] \) for proofs, there is a unique finite regular positive measure \( \mu_{A,x} \) on the Borel \( \sigma \)-algebra of \( C(\sigma(A)) \) such that

\[
\int f \, d\mu_{A,x} = \varphi_{A,x}(f) = \langle f(A)x, x \rangle \quad \forall f \in C(\sigma(A), \mathbb{C}). \tag{16.1}
\]

Taking \( f = 1 = \text{const.} \), we have \( f(A) = 1 \), so that \( \mu_{A,x}(\sigma(A)) = \|x\|^2 < \infty \). Since all measures will be Borel measures, we omit the \( \sigma \)-algebra from the notation and just write \( L^2(\sigma(A), \mu_{A,x}) \).
16.1 Definition Let $H$ be a Hilbert space, $A \in B(H)$ and $x \in H$. Then $x$ is called *-cyclic for $A$ if $\text{span}_C\{A^n(A^*)^m x \mid n,m \in \mathbb{N}_0\} = H$.

16.2 Remark A vector $x$ is cyclic for $A$ if $\text{span}_C\{A^n x \mid n \in \mathbb{N}_0\} = H$. Clearly the two notions are equivalent for self-adjoint $A$, but in general they differ. For the present purpose, *-cyclicity is the right notion. □

16.3 Proposition Let $A$ be a Hilbert space, $A \in B(H)$ normal and $x \in H$ *-cyclic for $A$. Then there is a unique unitary $U : H \to L^2(\sigma(A),\mu_{A,x})$ such that $UAU^* = M_z$, where $(M_z f)(z) = zf(z)$ for all $f \in L^2(\sigma(A),\mu_{A,x})$, $z \in \sigma(A)$.

Thus $A$ is unitarily equivalent to a multiplication operator.

Proof. The computation

$$
\|f(A)x\|^2 = \langle f(A)x, f(A)x \rangle = \langle f(A)^* f(A)x, x \rangle = \langle (\overline{f} f)(A)x, x \rangle = \int |f|^2 d\mu_{A,x}
$$

shows that the map

$$
D = \{f(A)x \mid f \in C(\sigma(A), C)\} \to L^2(\sigma(A), \mu_{A,x}), \quad f(A)x \mapsto f
$$

is well-defined and isometric. Since the domain $D \subseteq H$ is dense by *-cyclicity of $x$, the map extends by continuity (Lemma 4.17) to an isometric map $U : H \to L^2(\sigma(A), \mu_{A,x})$. Since $UH \subseteq L^2(\sigma(A), \mu_{A,x})$ is closed and $C(\sigma(A), C)$ is dense in $L^2(\sigma(A), \mu_{A,x})$, we have $UH = L^2(\sigma(A), \mu_{A,x})$. Thus $U$ is unitary. The uniqueness of $U$ (for given $x$) is clear from the construction. Now for $f \in C(\sigma(A), C)$ we have

$$(UAU^*)(f)(z) = (UAf(A)x)(z) = (U(xf)(A))(z) = zf(z),$$

and by density of $C(\sigma(A), C)$ in $L^2(\sigma(A), \mu_{A,x})$, this holds for all $f \in L^2(\sigma(A), \mu_{A,x})$. □

Not every normal operator $A \in B(H)$ admits a *-cyclic vector. (If $H$ is separable, $A$ has a *-cyclic vector if and only if the algebra $\{B \in B(H) \mid AB = BA, A^*B = BA^*\}$ is commutative.) In this case we say that $A$ is multiplicity free. In general we have:

16.4 Theorem (Spectral theorem for normal operators) Let $H$ be a Hilbert space and $A \in B(H)$ normal. Then there exists a family $\{\mu_i\}_{i \in I}$ of finite Borel measures on $\sigma(A)$ and unitary $U : H \to \bigoplus_{i \in I} L^2(\sigma(A), \mu_i)$\footnote{Here $\bigoplus$ is the Hilbert space direct sum defined at the end of Section 6.1.} such that $UAU^* = \bigoplus_{i \in I} M_z$, i.e.

$$(UAU^*) f_i(z) = zf_i(z) \quad \forall f = \{f_i\} \in \bigoplus_{i \in I} L^2(\sigma(A), \mu_i), \ z \in \sigma(A).\quad (16.2)$$

Proof. Let $\mathcal{F}$ be the family of subsets $F \subseteq H$ such that for $x, y \in F, x \neq y$ we have $f(A)x \perp f'(A)y$ for all $f, f' \in C(\sigma(A), C)$. We partially order $\mathcal{F}$ by inclusion. One easily checks that $\mathcal{F}$ satisfies the hypothesis of Zorn’s lemma. (Given a totally ordered subset $\mathcal{C} \subseteq \mathcal{F}$, $\bigcup \mathcal{C}$ is in $\mathcal{F}$, thus an upper bound for $\mathcal{C}$.) Thus there is a maximal element $M \in \mathcal{F}$. For each $x \in M$ we put $H_x = \{f(A)x \mid f \in C(\sigma(A), C)\}$. By construction these $H_x$ are mutually orthogonal. Let $K = \bigoplus_{x \in M} H_x$. By construction, we have $f(A)K \subseteq K$ for all $f \in C(\sigma(A), \sigma)$, thus also $f(A)^* K \subseteq K$ since $f(A)^* = \overline{f}(A)$. Thus $K^\perp$ is invariant under all $f(A)$. Picking a non-zero $y \in K^\perp$, we have $M \cup \{y\} \in \mathcal{F}$, which is a contradiction. Thus $K = H$.\footnote{Here $\bigoplus$ is the Hilbert space direct sum defined at the end of Section 6.1.}
Since clearly $x \in M$ is $*$-cyclic for the restriction of $A$ to $H_x$, we can use Proposition 16.3 to obtain unitaries $U_x : H_x \to L^2(\sigma(A), \mu_{A,x})$. Defining $U : H \to \bigoplus_{x \in M} L^2(\sigma(A), \mu_{A,x})$ by sending $x \in H_x$ to $U_x x \in L^2(\sigma(A), \mu_{A,x})$ and extending linearly, $U$ is unitary, and we are done. (Of course we have identified $I = M$ and $\mu_i = \mu_{A,x}$.)

16.5 Remark 1. Once the maximal family $M$ of vectors has been picked, the construction is canonical. But there is no uniqueness in the choice of that family. (This is similar to the non-uniqueness of the choices of ONBs in the eigenspace $\ker(A - \lambda I)$ that we make in proving Theorem 15.27.) For much more on this (in the self-adjoint case) see [58, Section VII.2].

2. Theorem 16.4 is perfectly compatible with Theorem 15.27: If $A$ is compact normal and $E$ is an ONB diagonalizing it then the $H_i$ in Theorem 16.4 are precisely the one-dimensional spaces $Ce$ for $e \in E$ and the measure $\mu_i$ corresponding to $H_i = Ce$ is the $\delta$-measure on $P(\sigma(A))$ defined by $\mu(S) = 1$ if $\lambda_e \in S$ and $\mu(S) = 0$ otherwise. (To be really precise, one should take the non-uniqueness in both theorems into account.)

3. If $A$ is as in the theorem and $g \in C(\sigma(A), \mathbb{C})$ then the continuous functional calculus gives us a normal operator $g(A)$. We now have

$$U g(A) U^* = \bigoplus_{i \in I} M_g.$$

(This is an obvious consequence of (16.2) when $g$ is a polynomial and follows by a density argument in general.) If one took Theorem 16.4 as given, this could even be used to define the continuous functional calculus. This would be circular since we used the continuous functional calculus to prove the theorem, or rather Proposition 16.3 on which it relied, but it shows that the continuous functional calculus and the spectral theorem are ‘equivalent’ in the sense of being easily deducible from each other.

4. The statement of Theorem 16.4 may not quite be what we expected, given the slogan ‘normal operators are multiplication operators’, since there is a direct sum involved. But this can be fixed when $H$ is separable: 

16.6 Corollary Let $A$ be a separable Hilbert space and $A \in B(H)$ normal. Then there exists a finite measure space $(X, \mathcal{A}, \mu)$, a function $g \in L^\infty(X, \mathcal{A}, \mu; \mathbb{C})$ and a unitary $W : H \to L^2(X, \mathcal{A}, \mu; \mathbb{C})$ such that $W A W^* = M_g$.

Proof. We apply Theorem 16.4. Since $H$ is separable, the index set $I$ is at most countable, and we write $I = \{1, \ldots, N\}$ where $N \in \mathbb{N} \cup \{\infty\}$ with $\infty = \# \mathbb{N}$. Now we put $X = I \times \sigma(A) = \bigoplus_{i \in I} \sigma(A)$ and for $Y \subseteq X$ we put $Y_i = p_2(p_1^{-1}(i)) = \{x \in \sigma(A) \mid (i, x) \in Y\} \subseteq \sigma(A)$. We define $\mathcal{A} \subseteq P(X)$ and $\mu : \mathcal{A} \to [0, \infty]$ by

$$\mathcal{A} = \{Y \subseteq X \mid Y_i \in \mathcal{B}(\sigma(A)) \forall i \in I\},$$

$$\mu(Y) = \sum_{i \in I} \mu_i(Y_i).$$

Using the countability of $I$ it is straightforward to check that $\mathcal{A}$ is a $\sigma$-algebra on $X$ and $\mu$ a (positive) measure on $(X, \mathcal{A})$. With (16.1) we have $\mu_i(\sigma(A)) = \|x_i\|^2$. Thus if we choose the cyclic vectors $x_i$ such that $\|x_i\| = 2^{-i}$ then $\mu(X) = \sum_i \mu_\{i\} \times \sigma(A) = \sum_i \mu_i(\sigma(A)) < \infty$, so that the measure space $(X, \mathcal{A}, \mu)$ is finite. Now we define a linear map

$$V : \bigoplus_{i \in I} L^2(\sigma(A), \mu_i) \to L^2(X, \mathcal{A}, \mu), \quad \{f_i\}_{i \in I} \mapsto f \quad \text{where} \quad f((i, x)) = f_i(x).$$
From the way \((X,\mathcal{A},\mu)\) was constructed, it is quite clear that \(V\) is unitary. (Check this!) Now \(W = VU : \mathcal{H} \rightarrow L^2(X,\mathcal{A},\mu)\), where \(U\) comes from Theorem 16.4, is unitary. In view of \(\langle UAU^*f\rangle_i(\lambda) = \lambda f_i(\lambda)\), defining \(g : X \rightarrow \mathbb{C}, (i,x) \mapsto x\) (which is bounded by \(r(A) = \|A\|\)), we have \(WAW^* = M_g\).

16.7 Exercise  
(i) Let \(\Sigma \subseteq \mathbb{C}\) be compact and non-empty and \(\mu\) be a finite positive Borel measure on \(\Sigma\). Put \(H = L^2(\Sigma,\mu)\) and define \(A \in B(H)\) by \((Af)(x) = xf(x)\) for \(f \in H\). Prove:

\[
\sigma(A) = \{\lambda \in \Sigma \mid \forall \varepsilon > 0 : \mu(B(\lambda,\varepsilon)) > 0\},
\]

\[
\sigma_p(A) = \{\lambda \in \Sigma \mid \mu(\{\lambda\}) > 0\},
\]

where \(B(\lambda,\varepsilon)\) denotes the open \(\varepsilon\)-disc around \(\lambda\).

(ii) Let \(A \in B(H)\) be normal. Use (i) to prove that \(\lambda \in \sigma_p(A)\) if and only if \(\mu(\{\lambda\}) > 0\) holds for at least one \(i \in I\) with \(\mu_i\) as in Theorem 16.4.

### 16.2 Borel functional calculus for normal operators

In the preceding section we used the continuous functional calculus to prove the spectral theorem for normal operators. Now we will turn the logic around and use the spectral theorem to extend the functional calculus to a larger class of functions!

16.8 Definition  If \((X,\tau)\) is a topological space, \(B^\infty(X,\mathbb{C})\) denotes the set of bounded functions \(X \rightarrow \mathbb{C}\) that are measurable with respect to the Borel \(\sigma\)-algebra \(B(X,\tau)\).

16.9 Lemma  Let \((X,\tau)\) be a topological space. Then

(i) If \(\{f_n\}_{n \in \mathbb{N}}\) is a sequence of Borel measurable functions \(X \rightarrow \mathbb{C}\) converging pointwise to \(f\) then is Borel measurable.

(ii) \((B^\infty(X,\mathbb{C}),\|\cdot\|_\infty)\), equipped with pointwise multiplication and \(*\)-operation is a \(C^*\)-algebra.

Proof. (i) It is an elementary fact of measure theory, cf. e.g. [10, Proposition 2.1.5], that the pointwise limit of a sequence of measurable functions (whatever the \(\sigma\)-algebra) is measurable.

(ii) Every sequence in \(B^\infty(S,\mathbb{C})\) that is Cauchy w.r.t. \(\|\cdot\|_\infty\) converges pointwise everywhere, thus is measurable by (i), and clearly bounded. Thus \(B^\infty(S,\mathbb{C})\) is complete. It is a \(C^*\)-algebra since product and \(*\)-operation satisfy submultiplicativity and the \(C^*\)-identity.

For a normal element \(a \in \mathcal{A}\) of a \(C^*\)-algebra, we cannot make sense of \(f(a)\) is \(f\) is not continuous. But the \(C^*\)-algebra \(B(H)\) has much more structure, and it turns out there is a Borel functional calculus extending the continuous functional calculus:

16.10 Theorem  Let \(H\) be a Hilbert space and \(A \in B(H)\) normal. Then:

(i) There is a unique unital \(*\)-homomorphism \(\alpha_A : B^\infty(\sigma(A),\mathbb{C}) \rightarrow B(H)\) extending the continuous functional calculus \(C(\sigma(A),\mathbb{C}) \rightarrow B(H)\) and satisfying \(\|\alpha_A(f)\| \leq \|f\|_\infty\). Again we write more suggestively \(f(A) = \alpha_A(f)\).

(ii) If \(B \in B(H)\) commutes with \(A\) and \(A^{*19}\) then \(B\) commutes with \(g(A)\) for all \(g \in B^\infty(\sigma(A),\mathbb{C})\).

\(^{19}\)Since \(A\) is normal, \(AB = BA\) actually implies \(A^*B = BA^*\) by Fuglede’s theorem, cf. Section B.8!
(iii) If \( \{f_n\}_{n \in \mathbb{N}} \subseteq B^\infty(\sigma(A), \mathbb{C}) \) is a bounded sequence converging pointwise to \( f \) then \( f \in B^\infty(\sigma(A), \mathbb{C}) \) and \( f_n(A) \xrightarrow{w} f(A) \), i.e. w.r.t. \( \tau_{wot} \), cf. Definition 17.8. (And \( \|f_n - f\|_\infty \to 0 \Rightarrow \|f_n(A) - f(A)\| \to 0 \).)

**Proof.** (i) For all \( x, y \in H \), the map

\[
\varphi_{x,y} : C(\sigma(A), \mathbb{C}) \to \mathbb{C}, \quad f \mapsto \langle f(A)x, y \rangle
\]

is a linear functional on \( C(\sigma(A), \mathbb{C}) \) that is bounded since \( \|f(A)\| \leq \|f\|_\infty \). Thus by the Riesz-Markov-Kakutani theorem there exists a unique complex Borel measure \( \mu_{x,y} \) on \( \sigma(A) \) such that

\[
\varphi_{x,y}(f) = \int f \, d\mu_{x,y} \quad \text{for all } f \in C(\sigma(A), \mathbb{C}).
\]

Since \( \varphi_{x,y} \) depends in a sesquilinear way on \( (x, y) \), the same holds for \( \mu_{x,y} \), and \( |\mu_{x,y}(\sigma(A))| = |\langle x, y \rangle| \leq \|x\| \|y\| \). Thus if \( f \in B^\infty(\sigma(A), \mathbb{C}) \), the map \( \psi_f : H^2 \to \mathbb{C} \) defined by \( (x, y) \mapsto \int f \, d\mu_{x,y} \) is a sesquilinear form that is bounded since

\[
|\psi_{x,y}(f)| \leq \|f\|_\infty \|x\| \|y\|.
\]

Thus by Proposition 12.7 there is a unique \( A_f \in B(H) \) such that

\[
\langle A_f x, y \rangle = \psi_{x,y}(f) \quad \text{for all } x, y \in H.
\]

It satisfies \( \|A_f\| \leq \|f\|_\infty \). Define \( \alpha : B^\infty(\sigma(A), \mathbb{C}) \to B(H) \) by \( f \mapsto A_f \). If \( f \in C(\sigma(A), \mathbb{C}) \) then \( \psi_{x,y}(f) = \langle f(A)x, y \rangle \) \( \forall x, y \), implying \( A_f = f(A) \). Thus \( \alpha_A \) extends the continuous functional calculus.

It remains to be shown that \( \alpha_A \) is a *-homomorphism. Linearity is quite obvious. Since the continuous functional calculus is a *-homomorphism, for \( f \in C(\sigma(A), \mathbb{C}) \) we have \( \overline{f}(A) = f(A)^* \), thus

\[
\int f \, d\mu_{x,y} = \langle f(A)x, y \rangle = \langle x, f(A)^* y \rangle = \langle x, \overline{f}(A)y \rangle = \int \overline{f}(A)y \, d\mu_{y,x} = \int f \, d\overline{\mu}_{y,x},
\]

thus \( \mu_{y,x} = \overline{\mu}_{x,y} \). Reading the above computation backwards, this implies \( \alpha_A(\overline{f}) = \alpha_A(f)^* \) for all \( f \in B^\infty(\sigma(A), \mathbb{C}) \). Since the continuous functional calculus is a homomorphism, for all \( f, g \in C(\sigma(A), \mathbb{C}) \) we have

\[
\int (fg) \, d\mu_{x,y} = \langle (fg)(A)x, y \rangle = \langle f(g(A)x, y \rangle = \langle g(A)x, f(A)^* y \rangle = \int g \, d\mu_{x,y}(A)g(A)y;
\]

The fact that this holds for all \( f, g \in C(\sigma(A), \mathbb{C}) \) implies \( f \mu_{x,y} = x, \overline{\mu}(A)y \). Thus for all \( f \in C(\sigma(A), \mathbb{C}) \), \( g \in B^\infty(\sigma(A), \mathbb{C}) \) we have

\[
\langle (fg)(A)x, y \rangle = \int f \, d\mu_{x,y} = \int g \, d\mu_{x,\overline{f}(A)y} = \langle g(A)x, \overline{f}(A)y \rangle = \langle f(A)g(A)x, y \rangle,
\]

so that \( (fg)(A) = f(A)g(A) \). As above, we deduce from this that \( f \mu_{x,y} = x, \overline{\mu}(A)y \) for all \( f \in B^\infty(\sigma(A), \mathbb{C}) \), and then \( (fg)(A) = f(A)g(A) \) for all \( f, g \in B^\infty(\sigma(A), \mathbb{C}) \).

(ii) The assumption implies \( Bf(A) = f(A)B \) for all \( f \in C(\sigma(A), \mathbb{C}) \). Thus

\[
\varphi_{Bx,y}(f) = \langle f(A)Bx, y \rangle = \langle Bf(A)x, y \rangle = \langle f(A)x, B^* y \rangle = \varphi_{x,B^* y}(f) \quad \forall x, y, f.
\]

This implies \( \mu_{Bx,y} = x, \overline{B^* y} \) for all \( x, y \), whence

\[
\langle f(A)Bx, y \rangle = \int f \, d\mu_{Bx,y} = \int f \, d\mu_{x,B^* y} = \langle f(A)x, B^* y \rangle \quad \forall x, y \in H, \ f \in B^\infty(\sigma(A), \mathbb{C}),
\]

thus \( f(A)B = Bf(A) \) for all \( f \in B^\infty(\sigma(A), \mathbb{C}) \).

(iii) Measurability of the limit function \( f \) follows from Lemma 16.9(i). If \( \|f_n\| \leq M \ \forall n \) then clearly \( \|f\| \leq M \). Thus \( f \in B^\infty(\sigma(A)) \). For all \( x, y \in H \) we have

\[
\langle \alpha_A(f_n)x, y \rangle = \int f_n \, d\mu_{x,y} \longrightarrow \int f \, d\mu_{x,y} = \langle \alpha_A(f)x, y \rangle,
\]

104
where convergence in the center is a trivial application of the dominated convergence theorem, using boundedness of \( \mu_{x,y} \) and \( \|f_n\|_\infty \leq M \) for all \( n \). This proves \( \alpha_A(f_n) \xrightarrow{u} \alpha_A(f) \). The final claim clearly follows from \( \|f_n(A) - f(A)\| = \|(f_n - f)(A)\| \leq \|f_n - f\|_\infty \). 

We now wish to connect the above construction of the Borel functional calculus with the spectral theorem. This is the first step:

**16.11 Exercise** Let \( \Sigma \subseteq \mathbb{C} \) be compact and \( \lambda \) a finite positive Borel measure on \( \Sigma \). Let \( H = L^2(\Sigma, \lambda; \mathbb{C}) \) and \( g \in B^\infty(\Sigma, \mathbb{C}) \).

(i) Prove that the multiplication operator \( M_g : H \to H, [f] \mapsto [gf] \) satisfies

\[
\|M_g\| = \text{ess sup}_\mu |g| = \inf \{ t \geq 0 \mid \lambda(\{x \in X \mid |g(x)| > t\}) = 0\} \leq \|g\|_\infty.
\]

(ii) Let \( A = M_z \in B(H) \), where \( z : \Sigma \to \mathbb{C} \). Prove that \( g(A) \) as defined by the Borel functional calculus coincides with \( M_g \).

**16.12 Corollary** Let \( A \in B(H) \) be normal and \( g \in B^\infty(\sigma(A), \mathbb{C}) \). Then

(i) \( \sigma(g(A)) = \overline{g(\sigma(A))} \).

(ii) If \( h \in B^\infty(g(\sigma(A)), \mathbb{C}) \) then \( h(g(A)) = (h \circ g)(A) \).

**Proof.** Let \( U : H \to \bigoplus_{i \in I} L^2(\sigma(A), \mu_i) \) as in Theorem 16.4, so that \( UAU^* = \bigoplus_i M_{\xi} \). The projectors \( P_i \) onto the subspaces \( L^2(\sigma(A), \mu_i) \) of the direct sum commute with \( A \) (and \( A^* \)), thus also with \( g(A) \) for each \( g \in B^\infty(\sigma(A), \mathbb{C}) \) by Theorem 16.10(ii). Thus the Borel functional calculus respects the direct sum decomposition of \( A \) (no matter how the maximal set \( M \) was chosen). It is a pure formality to show that if \( V : H \to H' \) is unitary then \( Vg(A)V^* = g(VA) \).

Thus with the direct sum decomposition \( UAU^* = \bigoplus_i M_{\xi} \) we have \( Ug(A)U^* = \bigoplus_i g(M_{\xi}) = \bigoplus_i M_g \), where the second equality comes from Exercise 16.11(ii).

(i) If \( \lambda \notin \overline{g(\sigma(A))} \) then \( M_{\lambda \xi} \) has a bounded inverse with norm \( \leq \text{dist}(\lambda, \overline{g(\sigma(A))})^{-1} \). Thus all \( M_{\lambda \xi} \) in the direct sum decomposition of \( A - \lambda I \) have inverses with uniformly bounded norms. Thus \( A - \lambda I \) has a bounded inverse.

(ii) Under the assumption on \( h \), we have

\[
Uh(g(A))U^* = \bigoplus_i h(M_g) = \bigoplus_i M_{hg} = U(h \circ g)(A)U^*.
\]

(This is too sloppy, but the reader should be able to make it precise.)

**16.13 Remark 1.** Since it turns out that \( g(A) = U^*(\bigoplus_i M_g)U \) for all \( g \in B^\infty(\sigma(A), \mathbb{C}) \), one might try to take this as the definition of \( g(A) \). But apart from being very inelegant, it has the problem that one must prove the independence of \( g(A) \) thus defined from the choice of the maximal set \( M \subseteq H \) in the proof of the spectral theorem. This would not be difficult if every Borel measurable function was a pointwise limit of a sequence of continuous functions. But this is false, making such an approach quite painful. (Compare Lusin’s theorem in, e.g., \([61]\).)

2. We cannot hope to prove \( \|g(A)\| = \|g\|_\infty \) for all \( g \in B^\infty(\sigma(A), \mathbb{C}) \) since it is true only if \( \sigma(A) = \sigma_p(A) \). Since singletons in \( \mathbb{C} \) are closed, thus Borel measurable, we can change \( g \) arbitrarily for some \( \lambda \in \sigma(A) \) without destroying the measurability of \( g \), making \( \|g\|_\infty \) as large as we want. But if \( \lambda \in \sigma(A) \backslash \sigma_p(A) \), Exercise 16.7 gives \( \mu_i(\{\lambda\}) = 0 \forall i \in I \), so that this change of \( g \) does not affect the norms, cf. Exercise 16.11, \( \text{ess sup}_\mu |g| \) of the multiplication operators making up \( g(A) \) and therefore does not affect \( \|g(A)\| \).
3. Let \( A \in B(H) \) be normal and consider the \( C^* \)-algebra \( \mathcal{A} = C^*(1, A) \subseteq B(H) \). Then \( g(A) \in \mathcal{A} \) for continuous \( g \), but for most non-continuous \( g \) we have \( g(A) \not\in \mathcal{A} \). For this reason there is no Borel functional calculus in abstract \( C^* \)-algebras. (But \( g(A) \) is always contained in the von Neumann algebra \( \text{vN}(A) = C^*(A, 1)^{\text{not}} \) generated by \( A \). This follows from Theorem 16.10(ii) and von Neumann’s ‘double commutant theorem’.)

\[
\square
\]

16.3 Normal operators vs. projection-valued measures

There is yet another perspective on the spectral theorem/functional calculus, provided by projection-valued measures:

16.14 Definition Let \( H \) be a Hilbert space and \( \Sigma \subseteq \mathbb{C} \) a compact subset. Let \( \mathcal{B}(\Sigma) \) be the Borel \( \sigma \)-algebra on \( \Sigma \). A projection-valued measure relative to \( (H, \Sigma) \) is a map \( P : \mathcal{B}(\Sigma) \to B(H) \) such that

\begin{enumerate}
\item \( P(S) \) is an orthogonal projection for all \( S \in \mathcal{B}(\Sigma) \).
\item \( P(\emptyset) = 0 \), \( P(\Sigma) = 1 \).
\item \( P(S \cap S') = P(S)P(S') \) for all \( S, S' \in \mathcal{B}(\Sigma) \).
\item For all \( x, y \in H \), the map \( E_{x,y} : \mathcal{B}(\sigma) \to \mathbb{C} \), \( S \mapsto \langle P(S)x, y \rangle \) is a complex measure. (Equivalently, if the \( \{S_n\}_{n \in \mathbb{N}} \subseteq \mathcal{B}(\Sigma) \) are mutually disjoint then \( \sum_n P(S_n) \) converges weakly to \( P(\bigcup_n S_n) \).)
\end{enumerate}

Note that (iii) implies \( P(S)P(S') = P(S')P(S) \) for all \( S, S' \in \mathcal{B}(\Sigma) \).

16.15 Proposition Let \( H \) be a Hilbert space and \( A \in B(H) \) normal. Put \( \Sigma = \sigma(A) \). For each \( S \in \mathcal{B}(\Sigma) \), define \( P(S) = \chi_S(A) \). Then \( S \mapsto P(S) \) is a projection-valued measure relative to \( (H, \Sigma) \), also called the spectral resolution of \( A \).

\begin{proof}
If \( g = \chi_S \) for \( S \in \mathcal{B}(\Sigma) \), \( g(A) \) is a direct sum of operators of multiplication by \( \chi_S \), which clearly all are idempotent. And since \( g = \chi_S \) is real-valued, \( g(A) \) is self-adjoint. Thus each \( P(S) = \chi_S(A) \) is an orthogonal projection. \( P(\emptyset) = 0 \) is clear, and \( P(\Sigma) = 1(A) = 1_H \) (since the constant 1 function is continuous). Property (iii) is immediate from \( \chi_{S \cap S'} = \chi_S \chi_{S'} \). Finally, if \( x, y \in H \) let \( Ux = \{f_i\}_{i \in I} \), \( Uy = \{g_i\}_{i \in I} \). Then

\[
E_{x,y}(S) = \langle P(S)x, y \rangle = \sum_{i \in I} \int_{\sigma(A)} \chi_S(z)f_i(z)g_i(z) \ d\mu_i(z).
\]

From this it is clear that \( S \mapsto E_{x,y}(S) \) is additive on countable families \( \{S_n\}_{n \in \mathbb{N}} \subseteq \mathcal{B}(\Sigma) \) by absolute convergence of \( \sum_i \int f_i g_i \ d\mu_i \).
\end{proof}

16.16 Exercise Let \( A \in B(H) \) be a normal operator and let \( P \) be the corresponding spectral measure. Prove:

\begin{enumerate}
\item \( \lambda \in \sigma(A) \) if and only if \( P(\sigma(A) \cap B(\lambda, \varepsilon)) \neq 0 \) for each \( \varepsilon > 0 \).
\item \( \lambda \in \sigma(A) \) is an eigenvalue if and only if \( P(\{\lambda\}) \neq 0 \).
\end{enumerate}

We have thus seen that every normal operator gives rise to a projection valued measure. The converse is also true, and we have a bijection between normal operators and projection-valued measures:
16.17 Proposition Let $H$ be a Hilbert space, $\Sigma \subseteq \mathbb{C}$ a compact subset and $P$ a projection-valued measure relative to $(H, \Sigma)$. Then

(i) For every $f \in B^\infty(\Sigma, \mathbb{C})$ there is a unique $\alpha(f) \in B(H)$ such that

$$\langle \alpha(f)x, y \rangle = \int f dE_{x,y} \quad \forall x, y \in H$$

and $\|\alpha(f)\| \leq \|f\|_\infty \forall f$. We also write, somewhat symbolically, $\alpha(f) = \int f(z) dP(z)$.

(ii) $\alpha : B^\infty(\Sigma, \mathbb{C}) \to B(H)$ is a unital $*$-homomorphism, and $\alpha(f)$ is normal for each $f$.

(iii) Put $A = \alpha(z) \in B(H)$, where $z : \Sigma \mapsto \mathbb{C}$ is the inclusion map. Then $\sigma(A) \subseteq \Sigma$ and $\alpha(f) = f(A)$ for each $f \in C(\Sigma, \mathbb{C})$.

(iv) The maps from normal elements to projection-valued measures (Proposition 16.15) and conversely (i)-(iii) above) are mutually inverse.

Proof. (i) It is clear that the map $(x, y) \mapsto E_{x,y}(S) = \langle P(S)x, y \rangle$ is sesquilinear for each $S \in \mathcal{B}(\sigma)$. For each $f \in B^\infty(\Sigma, \mathbb{C})$, we have $|\int f dE_{x,y}| \leq \|f\|_\infty \|x\| \|y\|$. Thus $[x, y]_f = \int f dE_{x,y}$ is a sesquilinear form with norm $\leq \|f\|_\infty$. Thus by Proposition 12.7 there is a unique $A_f \in B(H)$ such that $(A_f x, y) = [x, y]_f \forall x, y \in H$ and $\|A_f\| \leq \|f\|_\infty$. Now put $\alpha(f) := A_f$.

(ii) The inequality has already been proven. It is clear that $f \mapsto A_f = \alpha(f)$ is linear. If $1$ is the constant one function, we have $\langle \alpha(1)x, y \rangle = \int 1 dE_{x,y} = E_{x,y}(\Sigma) = \langle x, y \rangle$ since $P(\Sigma) = 1$. Thus $\alpha(1) = 1$. It remains to prove $\alpha(fg) = \alpha(f)\alpha(g)$ and $\alpha(f^*) = \alpha(f^*)$. We first do this for characteristic functions of measurable sets $S, T$: $f = \chi_S, g = \chi_T$. Now

$$\langle \alpha(\chi_S)x, y \rangle = \int_S dE_{x,y} = \langle P(S)x, y \rangle.$$ 

Thus $\alpha(\chi_S\chi_T) = \alpha(\chi_{S\cap T}) = E(S \cap T) = E(S)E(T) = \alpha(\chi_S)\alpha(\chi_T)$. By linearity of $\alpha$ we now have $\alpha(fg) = \alpha(f)\alpha(g)$ for all simple functions, i.e. finite linear combinations of characteristic functions. The latter are $\|\cdot\|_\infty$-dense in $B^\infty(\Sigma, \mathbb{C})$. (By a proof very similar to that of Lemma 5.13 in the case of $\ell^\infty$. Note that a measurable function is simple if and only if it assumes only finitely many values.) Now the identity follows for all $f, g$ by continuity of $\alpha$.

Furthermore, $\alpha(\chi_S)^* = P(S)^* = P(S) = \alpha(\chi_S)$, so that $\alpha(f^*) = \alpha(f^*)$ for simple functions. Now apply the same density and continuity argument as above.

In view of $f(A) = \alpha(f)$, the normality of $f(A)$ follows from

$$f(A)(A)^* = \alpha(f)\alpha(f)^* = \alpha(f^*f) = \alpha(f^*)\alpha(f) = f(A^*)f(A).$$

(iii) $\sigma(A) \subseteq \Sigma$ is clear. Since $\alpha(1) = 1$ and $\alpha(z) = A$ by definition, we have $\alpha(P) = P(A)$ for each polynomial. More generally, since $\alpha$ is a $*$-homomorphism, a polynomial in $z, \overline{z}$ is sent by $\alpha$ to the corresponding polynomial in $A, A^*$. These polynomials are $\|\cdot\|_\infty$-dense in $C(\Sigma, \mathbb{C})$ by Weierstrass Theorem A.31, so that the continuity proven in (ii) implies that $\alpha(f) = f(A)$ as produced by the continuous functional calculus.

(iv) Left as an exercise. ■

We close the discussion of spectral theorems with the advice of looking at the paper [28] and [71, Chapter 5] by two masters.
17 Weak and weak *-topologies. Alaoglu’s theorem

17.1 The weak topology of a Banach space

17.1 Definition If $V$ is a Banach space, the weak topology $\tau_w$ is the topology on $V$ induced by the family of seminorms $F = \{\| \cdot \|_\varphi = |\varphi(\cdot)| \mid \varphi \in V^*\}$. Thus a net $\{x_i\} \subseteq V$ converges weakly to $x \in V$ if and only if $\varphi(x_i) \to \varphi(x)$ for all $\varphi \in V^*$.

The weak topology is also called the $\sigma(V,V^*)$-topology (the topology on $V$ induced by the linear functionals in $V^*$). The Hahn-Banach theorem immediately gives that $F$ is separating, so that this topology is locally convex. It is clear that a norm-convergent net is weakly convergent since $|\varphi(x_i) - \varphi(x)| \leq \|\varphi\| \|x_i - x\|$. This implies $\overline{S}_\varphi \subseteq \overline{S}_w$ for every $S \subseteq V$ and $\tau_w \subseteq \tau_{\|\cdot\|}$.

If $(H,\langle \cdot,\cdot \rangle)$ is a Hilbert space, Theorem 6.29 implies $F = \{\langle \cdot, y \rangle \mid y \in H\}$, so that weak convergence of a net $\{x_i\}$ in a Hilbert space $H$ is equivalent to convergence of the nets $\{\langle x_i, y \rangle\}$ for each $y \in H$.

17.2 Proposition The weak topology on every infinite dimensional Banach space is strictly weaker than the norm-topology.

Proof. By the definition of $\tau_w$, for every weakly open neighborhood $U$ of $0$ there are $\varphi_1, \ldots, \varphi_n \in V^*$ such that $\{x \in V \mid |\varphi_i(x)| < 1 \forall i = 1, \ldots, n\} \subseteq U$. Thus $U$ contains the linear subspace $W = \bigcap_{i=1}^n \varphi_i^{-1}(0) \subseteq V$, whose codimension is $\leq n$. Thus if $V$ is infinite dimensional then $\dim W$ is infinite, thus non-zero. On the other hand, it is clear that the (norm-)open ball $B(0,1)$ contains no linear subspace of dimension $> 0$. Thus $B(0,1) \notin \tau_w$. Since $\tau_w \subseteq \tau_{\|\cdot\|}$ was clear, we have $\tau_w \subsetneq \tau_{\|\cdot\|}$. ■

17.3 Exercise (i) Prove that the sequence $\{\delta_n\}_{n \in \mathbb{N}}$ has no weak limit in $\ell^1(\mathbb{N}, \mathbb{F})$.

(ii) Let $1 < p < \infty$. Prove that the sequence $\{\delta_n\}_{n \in \mathbb{N}} \subseteq \ell^p(\mathbb{N}, \mathbb{F})$ converges to zero weakly, but not in norm.

(iii) (Bonus) Let $1 \leq p < \infty$ and $g$, $\{f_n\}_{n \in \mathbb{N}} \subseteq \ell^p(\mathbb{N}, \mathbb{F})$. Prove that if $f_n \rightharpoonup g$ and $\|f_n\|_p \to \|g\|_p$ then $\|f_n - g\|_p \to 0$.

The deviant behavior of $\ell^1$ in the preceding exercise can be understood as a consequence of the following remarkable result:

17.4 Theorem (I. Schur 1920) \footnote{Issai Schur (1874-1941). Russian mathematician. Studied and worked in Germany up to his emigration to Israel in 1939. Mostly known for his work in group and representation theory.} If $g$, $\{f_n\}_{n \in \mathbb{N}} \subseteq \ell^1(\mathbb{N}, \mathbb{F})$ and $f_n \rightharpoonup g$ then $\|f_n - g\|_1 \to 0$.

Like the uniform boundedness theorem, this result can be proven using the gliding hump method or using Baire’s theorem, cf. Section B.7.

Theorem 17.4 does not generalize to nets since the weak and norm topologies on $\ell^1(\mathbb{N}, \mathbb{F})$ differ by Proposition 17.2 and nets can distinguish topologies, cf. [46, Section 5.1].

17.5 Exercise Prove that every weakly convergent sequence in a Banach space is norm-bounded. Hint: Uniform boundedness theorem.

17.6 Exercise Let $V$ be a Banach space. Prove that the (norm) closed unit ball $V_{\leq 1}$ is also weakly closed. Hint: Hahn-Banach.
In Section 15.1 we saw that $V_{<1}$ is compact w.r.t. the norm topology if and only if $V$ is finite-dimensional. But the weak topology is weaker than the norm topology, so that a set can be weakly compact even though it is not norm compact. Indeed, after some further preparations we will prove the following theorem:

17.7 Theorem Let $V$ be a Banach space. Then the following are equivalent:

(i) $V$ is reflexive. ($\iff V^*$ is reflexive by Theorem 8.16.)

(ii) $V_{<1}$ is compact w.r.t. the weak topology.

In Remark 9.6 we have encountered the strong (operator) topology on $B(V)$: A net $\{A_i\} \subseteq B(V)$ converges strongly to $A \in B(V)$ if $\|(A_i - A)x\| \to 0$ for all $x \in V$. Now we can have a brief look at the weak operator topology:

17.8 Definition Let $V$ be a Banach space. The weak operator topology $\tau_{wot}$ on $B(V)$ is generated by the family $F = \{\| \cdot \|_{x,\varphi} : A \mapsto \|\varphi(Ax)\| \mid x \in V, \varphi \in V^*\}$ of seminorms. Thus $\{A_i\} \subseteq B(V)$ converges w.r.t. $\tau_{wot}$ to $A \in B(V)$ if and only if $\{A_i x\} \subseteq V$ converges weakly to $Ax$ for all $x$, i.e. $\varphi((A_i - A)x) \to 0$ for all $x \in X, \varphi \in V^*$. The family $F$ is separating, so that $\tau_{wot}$ is Hausdorff. We write $A_i \xrightarrow{\text{wot}} A$ or $A = w\text{-}\lim A_i$.

If $H$ is a Hilbert space, we have $A_i \xrightarrow{\text{wot}} A$ if and only if $\langle A_i x, y \rangle \to \langle Ax, y \rangle$ for all $x, y \in H$.

There is little risk of confusing the weak topology on $V$ with the weak operator topology on $B(V)$. But one might confuse the latter with the weak topology that $B(V)$ has as a Banach space, in particular since the above $\|\cdot\|_{x,\varphi}$ are in $B(V)^*$! However, when $V$ is infinite dimensional these seminorms do not exhaust (or span) the bounded linear functionals on $B(V)$, so that the weak operator topology on $B(V)$ is strictly weaker than the weak topology!

17.9 Exercise Let $H$ be a Hilbert space.

(i) Prove that the map $(B(H), \tau_{wot}) \to (B(H), \tau_{wot}), A \mapsto A^*$ is continuous.

(ii) Prove that the map $(B(H), \tau_{sot}) \to (B(H), \tau_{sot}), A \mapsto A^*$ is not continuous if $\dim H = \infty$.

The strong and weak operator topologies on $B(H)$ are quite important for the theory of von Neumann algebras, an important special class of $C^*$-algebras. See [32, 49] for the basics.

17.2 The weak-* topology on the dual space of a Banach space

17.10 Definition If $V$ is a Banach space, the weak-* topology $\tau_{w^*}$ (or $\sigma(V^*, V)$-topology) is the topology on the dual space $V^*$ defined by the family $F = \{\|\cdot\|_x \mid x \in X\}$ of seminorms, where $\|\varphi\|_x = |\hat{x}(\varphi)| = |\varphi(x)|$. Thus a net $\{\varphi_i\}$ in $V^*$ converges to $\varphi \in V^*$ if and only if $\varphi_i(x) \to \varphi(x)$ for every $x \in V$.

17.11 Remark 1. Since $\varphi(x) = 0$ for all $x \in V$ means $\varphi = 0$, $F$ is separating, thus the $\sigma(V^*, V)$-topology is Hausdorff and therefore locally convex.

2. If $V$ is infinite dimensional, the weak-* topology $\tau_{w^*}$ is neither normable nor metrizable.

3. Since the weak-* topology is induced by the linear functionals on $V^*$ of the form $\hat{x}$, which constitute a subset of $V^{**}$, it is weaker than the weak topology, thus also weaker than the norm topology: $\tau_{w^*} \subseteq \tau_w \subseteq \tau_{\|\cdot\|}$. As we know, the second inclusion is proper whenever $V$ is infinite dimensional. For the first, we have: 

\[\square\]
17.12 Proposition If $V$ is a Banach space, the weak-* topology $\sigma(V^*, V)$ on $V^*$ coincides with the weak topology $\sigma(V^*, V^{**})$ if and only if $V$ is reflexive.

Proof. If $V$ is reflexive then $V^{**} \cong V$, so that the weak-* topology $\sigma(V^*, V)$ on $V^*$ clearly coincides with the weak topology $\sigma(V^*, V^{**})$. If $V$ is not reflexive, we have $V \not\subseteq V^{**}$. Now for $\psi \in V^{**} \setminus V$ it is clear that the linear functional $\psi$ on $V^*$ is $\sigma(V^*, V^{**})$-continuous, whereas Exercise 17.13 gives that it is not $\sigma(V^*, V)$-continuous. This proves $\sigma(V^*, V) \neq \sigma(V^*, V^{**})$. ■

17.13 Exercise Let $V$ be an $\mathbb{F}$-vector space with algebraic dual space $V^*$.

(i) For $\varphi, \psi_1, \ldots, \psi_n \in V^*$ prove that $\varphi \in \text{span}_\mathbb{F}\{\psi_1, \ldots, \psi_n\} \iff \bigcap_{i=1}^n \ker \psi_i \subseteq \ker \varphi$.

(ii) Let $W \subseteq V^*$ be a linear subspace. Prove that a linear functional $\varphi : V \to \mathbb{F}$ is $\sigma(V, W)$-continuous if and only if $\varphi \in W$. Hint: Use (i).

17.14 Remark Before we proceed, some comments are in order: While the norm and weak topologies are defined for each Banach space, the weak-* topology is defined only on spaces that are the dual space $V^*$ of a given space $V$. There are Banach spaces, like $c_0(\mathbb{N}, \mathbb{F})$, that are not isomorphic (isometrically or not) to the dual space of any Banach space, cf. Corollary B.11. And there are non-isomorphic Banach spaces with isomorphic dual spaces, cf. Corollary B.12. Thus to define the weak-* topology on a Banach space, it is not enough just to know that the latter is a dual space. We must choose a ‘pre-dual’ space. □

The following is the reason for the importance of the weak-* topology:

17.15 Theorem (Alaoglu’s Theorem)\(^{51}\) If $V$ is a Banach space then the (norm)closed unit ball $(V^*)_{\leq 1} = \{\varphi \in V^* \mid \|\varphi\| \leq 1\}$ is compact in the $\sigma(V^*, V)$-topology.

Proof. Define

$$Z = \prod_{x \in V} \{z \in \mathbb{C} \mid |z| \leq \|x\|\},$$

equipped with the product topology. Since the closed discs in $\mathbb{C}$ are compact, $Z$ is compact by Tychonov’s theorem. If $\varphi \in (V^*)_{\leq 1}$ then $|\varphi(x)| \leq \|x\| \forall x$, so that we have a map

$$f : (V^*)_{\leq 1} \to Z, \quad \varphi \mapsto \prod_{x \in V} \varphi(x).$$

Since the map $\varphi \mapsto \varphi(x)$ is continuous for each $x$, $f$ is continuous (w.r.t. the weak-* topology on $(V^*)_{\leq 1}$). It is trivial that $V$ separates the points of $V^*$, thus $f$ is injective. By definition, a net $\{\varphi_i\}$ in $(V^*)_{\leq 1}$ converges in the $\sigma(V^*, V)$-topology if and only if $\varphi_i(x)$ converges for all $x \in V$, and therefore if and only if $f(\varphi_i)$ converges. Thus $f : (V^*)_{\leq 1} \to f((V^*)_{\leq 1}) \subseteq Z$ is a homeomorphism.

Now let $z \in f((V^*)_{\leq 1}) \subseteq Z$. Clearly, $|z_x| \leq \|x\| \forall x \in X$. By Proposition A.9.2 there is a net in $f((V^*)_{\leq 1})$ converging to $z$ and therefore a net $\{\varphi_i\}$ in $(V^*)_{\leq 1}$ such that $f(\varphi_i) \to z$. This means $\varphi_i(x) \to z_x \forall x \in V$. In particular $\varphi_i(\alpha x + \beta y) \to z_{\alpha x + \beta y}$, while also $\varphi_i(\alpha x + \beta y) = \alpha \varphi_i(x) + \beta \varphi_i(y) \to \alpha z_x + \beta z_y$. Thus the map $\psi : V \to \mathbb{C}, x \mapsto z_x$ is linear with $\|\psi\| \leq 1$, to wit $\psi \in (V^*)_{\leq 1}$ and $z = f(\psi)$. Thus $f((V^*)_{\leq 1}) \subseteq f(\psi((V^*)_{\leq 1}))$, so that $f((V^*)_{\leq 1}) \subseteq Z$ is closed.

Now we have proven that $(V^*)_{\leq 1}$ is homeomorphic to the closed subset $f((V^*)_{\leq 1})$ of the compact space $Z$, and therefore compact. ■

\(^{51}\)Leonidas Alaoglu (1914-1981). Greek mathematician. (Earlier versions due to Helly and Banach.)
17.16 Remark We deduced Alaoglu’s theorem from Tychonov’s theorem. The latter is known to be equivalent to the axiom of choice (AC). However, we only needed Tychonov’s theorem as restricted to Hausdorff spaces. The latter can be proven from a weaker axiom than AC, to which it actually is equivalent (namely the ‘ultrafilter lemma’, which also implies the Hahn-Banach theorem). Cf. [46].

17.17 Exercise Use Alaoglu’s theorem to prove that every Banach space $V$ over $F$ admits an isometric embedding into $C(X,F)$ for some compact Hausdorff space $X$.

17.18 Exercise (i) Use Alaoglu’s theorem to prove (i)$\Rightarrow$(ii) in Theorem 17.7.

(ii) Conclude that the closed unit ball of every Hilbert space is weakly compact.

(iii) Prove $\sigma(V,V^*) = \sigma(V^{**},V^*) \upharpoonright V$.

(iv) Use Theorem 17.19 and (iii) to prove (ii)$\Rightarrow$(i) in Theorem 17.7.

17.19 Theorem (Goldstine) 52 If $V$ is Banach then $V_{\leq 1} \subseteq (V^{**})_{\leq 1}$ is $\sigma(V^{**},V^*)$-dense.

The fairly non-trivial proof will be given in the supplementary Section B.5.2.

18 The Gelfand homomorphism for commutative Banach and $C^*$-algebras

We now pick up the discussion begun in Section 11.4.

18.1 The topology of $\Omega(A)$. The Gelfand homomorphism

The following is one of the most important applications of Alaoglu’s Theorem 17.15:

18.1 Proposition  Let $A$ be a unital Banach algebra and $\Omega(A)$ its spectrum. For each $a \in A$ define $\hat{a} : \Omega(A) \to \mathbb{C}$, $\varphi \mapsto \varphi(a)$. Let $\tau$ be the initial topology on $\Omega(A)$ defined by $\{\hat{a} \mid a \in A\}$, i.e. the weakest topology making all $\hat{a}$ continuous. Then $(\Omega(A),\tau)$ is compact Hausdorff.

Proof. We have just proven that (non-zero) characters are automatically continuous with norm one, so that $\Omega(A) \subseteq (A^*)_{\leq 1}$. By definition, $\hat{a}(\varphi) = \varphi(a)$. Thus the topology generated by the $\hat{a}$ is the restriction to $\Omega(A) \subseteq A$ of the $\sigma(A^*,A)$-topology, thus Hausdorff. Let $\{\varphi_i\}$ be a net in $\Omega(A)$ that converges to $\psi \in A^*$ w.r.t. the $\sigma(A^*,A)$-topology. Then for all $a, b \in A$ we have $\psi(ab) = \lim_i \varphi_i(ab) = \lim_i \varphi_i(a) \varphi_i(b) = \psi(a) \psi(b)$, so that $\psi \in \Omega(A)$. Thus $\Omega(A) \subseteq (A^*)_{\leq 1}$ is $\sigma(A^*,A)$-closed, thus compact since $(A^*)_{\leq 1}$ is $\sigma(A^*,A)$-compact by Alaoglu’s theorem. \[\square\]

The above works whether or not $A$ is commutative, but we’ll now restrict to commutative $A$ since $\Omega(A)$ can be very small otherwise. We begin by completing Exercise 11.36:

18.2 Proposition Let $X$ be a compact Hausdorff space and $A = C(X,F)$. Then the map $X \to \Omega(A)$, $x \mapsto \varphi_x$ is a homeomorphism.

52Herman Heine Goldstine (1913-2004). American mathematician and computer scientist. Worked on very pure and very applied mathematics, like John von Neumann with whom he collaborated on computers.
Proof. Injectivity was already proven in Exercise 11.36. In order to prove surjectivity, let \( \varphi \in \Omega(A) \) and put \( M = \ker \varphi \). Then \( M \subseteq A \) is a proper closed subalgebra (in fact an ideal), and it is self-adjoint by Lemma 13.18 since \( A \) is a \( C^* \)-algebra. If \( x, y \in X, \ x \neq y, \) pick \( f \in A \) with \( f(x) \neq f(y) \). With \( g = f - \varphi(f)1 \) we have \( \varphi(g) = 0 \), thus \( g \in M \). This proves that \( M \) separates the points of \( X \), yet it is not dense in \( A \). Now the incarnation Corollary A.34 of the Stone-Weierstrass theorem implies that there must be an \( x \in X \) at which \( M \) vanishes identically, i.e. \( \varphi_x(f) = 0 \) for all \( f \in M \). Now for every \( f \in A \) we have \( f - \varphi(f)1 \in M \), thus 

\[
\varphi_x(f - \varphi(f)1) = 0,
\]

which is equivalent to \( \varphi_x(f) = \varphi(f) \). Thus \( \iota : X \rightarrow \Omega(A) \) is surjective.

If \( \{x_i\} \subseteq X \) such that \( x_i \rightarrow x \) then \( \varphi_{x_i}(f) = f(x_i) \rightarrow f(x) = \varphi_x(f) \) for every \( f \in A \) by continuity of \( f \). But this precisely means that \( \varphi_{x_i} \rightarrow \varphi_x \) w.r.t. the weak*-topology. Thus \( \iota \) is continuous. As a continuous bijection of compact Hausdorff spaces it is a homeomorphism. ■

18.3 Definition Let \( A \) be a unital Banach algebra. Then its radical is the set of quasi-nilpotent elements: \( \text{rad} A = \{ a \in A \mid r(a) = 0 \} \). We call \( A \) semisimple if \( \text{rad} A = \{0\} \).

18.4 Proposition If \( A \) is a unital commutative Banach algebra, the map

\[
\pi : A \rightarrow C(\Omega(A), \mathbb{C}), \ a \mapsto \hat{a}
\]

is a unital homomorphism, called the Gelfand homomorphism (or representation) of \( A \), and \( \| \pi(a) \| = r(a) \leq \| a \| \) for all \( a \in A \). Thus \( \ker \pi = \text{rad} A \), and \( \pi \) is injective if and only if \( A \) is semisimple.

Proof. It is clear that \( \pi \) is linear. Furthermore, \( \hat{1} = \varphi(1) = 1 \) and

\[
\hat{a}(\varphi_1 \varphi_2) = (\varphi_1 \varphi_2)(a) = \varphi_1(a) \varphi_2(a) = \hat{a}(\varphi_1) \hat{a}(\varphi_2),
\]

so that \( \pi \) is a unital homomorphism. We have

\[
\| \hat{a} \| = \sup_{\varphi \in \Omega(A)} |\hat{a}(\varphi)| = \sup_{\varphi \in \Omega(A)} |\varphi(a)| = r(a) \leq \| a \|,
\]

where we used (11.7) and Proposition 11.11. ■

The Gelfand homomorphism can fail to be surjective or injective or both. See Section 18.2 for an important example for the failure of surjectivity and Exercise 18.6 for a non-trivial unital Banach algebra with very large radical.

18.5 Proposition Let \( A \) be a commutative unital Banach algebra and \( a \in A \) such that \( A \) is generated by \( \{1, a\} \). Then the map \( \hat{a} : \Omega(A) \rightarrow \sigma(a) \) is a homeomorphism. The same conclusion holds if \( a \in \text{Inv} A \) and \( A \) is generated by \( \{1, a, a^{-1}\} \).

Proof. We know from (11.7) that \( \hat{a}(\Omega(A)) = \sigma(a) \), thus \( \hat{a} \) is surjective. Assume \( \hat{a}(\varphi_1) = \hat{a}(\varphi_2) \), thus \( \varphi_1(a) = \varphi_2(a) \). Since the \( \varphi_i \) are unital homomorphisms, this implies \( \varphi_1(a^n) = \varphi_2(a^n) \) for all \( n \in \mathbb{N}_0 \), so that \( \varphi_1, \varphi_2 \) agree on the polynomials in \( a \). Since the latter are dense in \( A \) by assumption and the \( \varphi_i \) are continuous, this implies \( \varphi_1 = \varphi_2 \). Thus \( \hat{a} : \Omega(A) \rightarrow \sigma(a) \) is injective, thus a continuous bijection. Since \( \Omega(A) \) is compact and \( \sigma(a) \subseteq \mathbb{C} \) Hausdorff, \( \hat{a} \) is a homeomorphism. This proves the first claim.

For the second claim, note that \( \varphi(a) \varphi(a^{-1}) = \varphi(aa^{-1}) = \varphi(1) = 1, \) thus \( \varphi(a^{-1}) = \varphi(a)^{-1}, \) for each \( \varphi \in \Omega(A) \). This implies that \( \varphi_1(a^n) = \varphi_2(a^n) \) also holds for negative \( n \in \mathbb{Z} \). Now \( \varphi_1, \varphi_2 \) agree on all Laurent polynomials in \( a \), thus on \( A \) by density and continuity. The rest of the proof is the same. ■
18.6 Exercise Let \( \alpha : \mathbb{N}_0 \to (0, \infty) \) be a map satisfying \( \alpha(0) = 1 \) and \( \alpha_{n+m} \leq \alpha_n \alpha_m \forall n, m. \) For \( f : \mathbb{N}_0 \to \mathbb{C}, \) define \( \|f\| = \sum_{n \in \mathbb{N}_0} \alpha_n |f(n)|, \) and \( \mathcal{A} = \{ f : \mathbb{N}_0 \to \mathbb{C} \mid \|f\| < \infty \}. \) For \( f, g \in \mathcal{A}, \) define \( f \cdot g \) by \( (f \cdot g)(n) = \sum_{n=m+r} f(u)g(v). \)

(i) Prove that \( (\mathcal{A}, \cdot, 1, \| \cdot \|) \) is a commutative Banach algebra with unit \( 1 = \delta_0. \)

(ii) Prove that \( \delta_1 \) generates \( \mathcal{A} \) and \( r(\delta_1) = \lim_{n \to \infty} \alpha_1/n. \)

(iii) Find a sequence \( \{\alpha_n\} \) satisfying the above requirements such that \( \delta_1 \) is quasi-nilpotent.

(iv) Conclude that \( \text{rad} \mathcal{A} = \{ f \in \mathcal{A} \mid f(0) = 0 \}. \)

18.7 Remark 1. Since every commutative unital Banach algebra has at least one non-zero character \( \varphi, \) the worst that can happen is \( \text{rad} \mathcal{A} = \varphi^{-1}(0), \) which has codimension one, as in the preceding exercise.

2. If \( \mathcal{A} \) is a non-unital Banach algebra and \( a \in \mathcal{A} \) one defines \( \sigma(a) = \sigma_{\mathcal{A}}(a), \) where \( \mathcal{A} \) is the unitization of \( \mathcal{A} \) considered in Exercise 11.26. Now one defines \( r(a) = \sup_{\lambda \in \sigma(a)} |\lambda| \) and \( \text{rad} \mathcal{A} = r^{-1}(0) \subseteq \mathcal{A} \) as before. Now for the non-unital subalgebra \( \mathcal{A}' = \{ f \in \mathcal{A} \mid f(0) = 0 \} \) of the \( \mathcal{A} \) from Exercise 18.6 one easily proves \( \mathcal{A}' \cong \mathcal{A} \), thus \( r(a) = 0 \forall a \in \mathcal{A}' \) and \( \text{rad} \mathcal{A}' = \mathcal{A}'. \)

18.8 Exercise Let \( \mathcal{A} \) be a commutative unital Banach algebra generated by \( \{a_1, \ldots, a_n\} \subseteq \mathcal{A}. \) Define \( s : \Omega(\mathcal{A}) \to \mathbb{C}^n, \varphi \mapsto (\varphi(a_1), \ldots, \varphi(a_n)). \) Prove that \( s \) is a homeomorphism of \( \Omega(\mathcal{A}) \) onto a closed subspace of \( \sigma(a_1) \times \cdots \times \sigma(a_n). \) (The latter is one way of defining the joint spectrum \( \sigma(a_1, \ldots, a_n). \)

18.2 Application: Absolutely convergent Fourier series

The following example is lengthy, but very instructive:

18.9 Example Let \( (\mathcal{A} = \ell^1(\mathbb{Z}, \mathbb{C}), \| \cdot \|, *, 1) \) be the unital Banach algebra from Example 11.20. In view of \( \delta_n \ast \delta_m = \delta_{n+m}, \) this algebra is generated by \( a = \delta_1 \in \text{Inv} \mathcal{A} \) and \( a^{-1} = \delta_{-1}. \) We have \( \|a\| = \|a^{-1}\| = 1 \) so that by Exercise 11.22 we have \( \sigma(a) \subseteq S^1. \) Now, for \( z \in S^1 \) define

\[
\varphi_z : f \mapsto \sum_{n \in \mathbb{Z}} f(n)z^n, \quad (18.2)
\]

which is absolutely and uniformly convergent since \( f \in \ell^1. \) It is clear that \( \varphi_z(\delta_n) = z^n, \) so that \( \varphi_z(\delta_n)\delta_m = \varphi_z(\delta_{n+m}) = z^{n+m} = \varphi_z(\delta_n)\varphi_z(\delta_m), \) proving \( \varphi_z \in \Omega(\mathcal{A}). \) In particular, \( \varphi_z(a) = z, \) so that \( \sigma(a) = S^1. \) Now Proposition 18.5 gives \( \Omega(\mathcal{A}) = \{ \varphi_z \mid z \in S^1 \}. \) By uniform convergence in (18.2), one finds that \( \hat{f}(z) = \varphi_z(f) \) is continuous in \( z \) and \( \hat{f}(n) = \int_0^1 \hat{f}(e^{2\pi it})e^{-2\pi int} dt = f(n) \forall n. \)

We have

\[
\|\pi(f)\| = r(f) = \sup_{z \in S^1} |\varphi_z(f)| = \sup_{z \in S^1} |\hat{f}(z)| = \|\hat{f}\|_{\infty},
\]

which vanishes only if \( f = 0 \) (by the fact that \( g \in C(S^1, \mathbb{C}) \) vanishes if and only if \( \hat{g}(n) = 0 \forall n \in \mathbb{Z}, \) cf. e.g. [74, Chapter 2, Theorem 2.1]). Thus \( \mathcal{A} \) is semisimple and \( \pi : \ell^1(\mathbb{Z}) \to C(S^1, \mathbb{C}) \) is injective. But \( \pi \) is not surjective: Its image consists precisely of

\[
\mathcal{B} = \{ g \in C(S^1, \mathbb{C}) \mid \sum_{n \in \mathbb{Z}} |\hat{g}(n)| < \infty \}.
\]

This is an algebra since \( \mathcal{A} \) is. In (superficially) more elementary terms this is just the observation that the pointwise product of two elements of \( C(S^1, \mathbb{C}) \) corresponds to convolution of their
Fourier coefficients and the fact that $\ell^1(\mathbb{Z})$ is closed under convolution. For the $g \in \mathcal{B}$ the Fourier series converges absolutely uniformly to $g$, but we have proven in Section 9.4 that $C(S^1, \mathbb{C})$ has a dense subset of functions whose the Fourier series does not even converge pointwise everywhere. (Our proof was non-constructive, but as we remarked, single examples can be produced constructively.)

And functions in $C(S^1, \mathbb{C}) \setminus \mathcal{B}$ can be written down even more concretely: With some effort (see [48] for an exposition) the series $\sum_{n=2}^{\infty} \frac{\sin nz}{n \log n}$ can be shown to be uniformly convergent to some $f \in C(S^1, \mathbb{C})$, and its Fourier coefficients are not absolutely summable since $\sum_{n=2}^{\infty} (n \log n)^{-1} = \infty$.

We now turn the non-surjectivity of $\pi : \ell^1(\mathbb{Z}) \rightarrow C(S^1)$ into a virtue! For $g \in C(S^1, \mathbb{C})$ define $\|g\|_B = \sum_{n \in \mathbb{Z}} |\hat{g}(n)|$. Thus $\mathcal{B} = \{g \in C(S^1, \mathbb{C}) \mid \|g\|_B < \infty\}$. We have seen that the Gelfand representation of $\ell^1(\mathbb{Z})$ is a unital, the Stone-Weierstrass theorem (Corollary A.32) gives $\pi(\mathcal{B})$ an isometric $*$-isomorphism $(\ell^1(\mathbb{Z}), \|\cdot\|_1) \rightarrow (\mathcal{B}, \|\cdot\|_B)$. Now we can give Gelfand’s slick proof of the following result proven by Wiener with much more effort:

18.10 Theorem If $g \in \mathcal{B}$ satisfies $g(z) \neq 0 \forall z \in S^1$ and $h \in C(S^1, \mathbb{C})$ is its multiplicative inverse $h(z) = 1/g(z)$ then $h \in \mathcal{B}$ (thus $h$ has absolutely convergent Fourier series).

Proof. Let $f = \pi^{-1}(g) \in \ell^1(\mathbb{Z})$. We have seen that $\Omega(\mathcal{A}) = S^1$ and $\varphi_z(f) = g(z)$ for all $z \in S^1$. Now the assumption $g(z) \neq 0 \forall z$ implies that $0 \notin \sigma(f) = \{\varphi_z(f) \mid z \in S^1\}$, so that $f$ is invertible in $\ell^1(\mathbb{Z})$. Thus $\pi(f) = g \in \mathcal{B}$ is invertible. Since the product on $\mathcal{B}$ is pointwise multiplication, this proves that $h = g^{-1} \in \mathcal{B}$, thus $h$ has absolutely convergent Fourier series. ■

18.3 $C^*$-algebras. Continuous functional calculus revisited

In discussing when the Gelfand homomorphism $\pi : \mathcal{A} \rightarrow C(\Omega(\mathcal{A}), \mathbb{F})$ is an isomorphism, we limit ourselves to the case where $\mathcal{A}$ is a $C^*$-algebra over $\mathbb{C}$.

18.11 Theorem (Gelfand Isomorphism) If $\mathcal{A}$ is a commutative unital $C^*$-algebra then the Gelfand homomorphism $\pi : \mathcal{A} \rightarrow C(\Omega(\mathcal{A}), \mathbb{C})$ is an isometric $*$-isomorphism.

Proof. For all $a \in \mathcal{A}$, $\varphi \in \Omega(\mathcal{A})$, using Lemma 13.18 we have

$$\pi(a^*)(\varphi) = \hat{a}^*(\varphi) = \overline{\varphi(a)} = \overline{\varphi^*(a)} = \overline{\hat{a}(\varphi)} = \pi(a)(\varphi^*).$$

Thus $\pi(a^*) = \pi(a)^*$, so that $\pi$ is a $*$-homomorphism, and $\pi(\mathcal{A}) \subseteq C(\Omega(\mathcal{A}), \mathbb{C})$ is self-adjoint.

Since $\mathcal{A}$ is commutative, all $a \in \mathcal{A}$ are normal, thus satisfy $r(a) = \|a\|$ by Proposition 12.23(i). Together with $\|\pi(a)\| = r(a)$ for all $a$ this implies that $\pi$ is an isometry, thus injective. Since $\mathcal{A}$ is complete, this implies that the image $\pi(\mathcal{A}) \subseteq C(\Omega(\mathcal{A}), \mathbb{C})$ is complete, thus closed.

If $\varphi_1 \neq \varphi_2$ then there is an $a \in \mathcal{A}$ such that $\varphi_1(a) \neq \varphi_2(a)$, thus $\pi(a)(\varphi_1) = \hat{a}(\varphi_1) \neq \hat{a}(\varphi_2) = \pi(a)(\varphi_2)$. This proves that $\pi(\mathcal{A}) \subseteq C(\Omega(\mathcal{A}), \mathbb{R})$ separates the points of $\Omega(\mathcal{A})$. Since $\pi$ is also unital, the Stone-Weierstrass theorem (Corollary A.32) gives $\pi(\mathcal{A}) = \pi(\mathcal{A}) = C(\Omega(\mathcal{A}), \mathbb{C})$. ■

18.12 Remark 1. Theorem 18.11 can be strengthened to a (contravariant) equivalence of categories between the categories of commutative unital $C^*$-algebras and unital $*$-homomorphisms and of compact Hausdorff spaces and continuous maps.

2. With some work, the assumption of $\mathcal{A}$ having a unit can be dropped, cf. e.g. [49]. One finds that every commutative $C^*$-algebra is isometrically $*$-isomorphic to $C_0(X, \mathbb{C})$ for a locally compact Hausdorff space $X$, unique up to homeomorphism. $X$ is compact if and only if $\mathcal{A}$ is unital. And the equivalence of categories mentioned above extends to a contravariant
equivalence between the category of locally compact Hausdorff spaces and proper maps and the category of commutative $*$-algebras and non-degenerate homomorphisms.

3. The preceding comments is a sense end the theory of commutative $C^*$-algebras since the latter is reduced it to general topology. But the theory of non-commutative $C^*$-algebras is vast, see [32, 49] for accessible introductions, and it turns out that commutative $C^*$-algebras are a very useful tool for studying them, as results like Propositions 13.12 and 13.13 just begin to illustrate.

4. Comparing Theorem 18.11 with the non-surjectivity of the Gelfand-Homomorphism for $(\ell^1(\mathbb{Z}, \mathbb{C}), \ast)$ shows that $\ell^1(\mathbb{Z}, \mathbb{C})$ does not admit a norm that would make it a $C^*$-algebra. But $\ell^1(\mathbb{Z}, \mathbb{C})$ admits a non-complete $C^*$-norm $\| \cdot \|'$, and completing $\ell^1(\mathbb{Z}, \mathbb{C})$ w.r.t. the latter yields a $C^*$-algebra $C^*(\mathbb{Z})$ that is isomorphic to $C^*(U) \subseteq B(\ell^2(\mathbb{Z}, \mathbb{C}))$, where $U \in B(\ell^2(\mathbb{Z}, \mathbb{C}))$ is the two-sided shift unitary. One also has $C^*(\mathbb{Z}) \cong C(S^1, \mathbb{C})$, thus the $C^*$-completion ‘adds’ the continuous functions with non-absolutely convergent Fourier series. 

The following is a $C^*$-version of Proposition 18.5:

18.13 PROPOSITION Let $\mathcal{B}$ be a commutative unital $C^*$-algebra and $b \in \mathcal{B}$ such that $\mathcal{B} = C^*(1,b)$. Then the map $\widehat{b} : \Omega(\mathcal{B}) \rightarrow \sigma(b)$ is a homeomorphism.

Proof. The proof is similar to that of Proposition 18.5, enriched by the following argument: If $\varphi_1(b) = \varphi_2(b)$ then by Lemma 13.18 we have $\varphi_1(b^*) = \varphi_1(b) = \varphi_2(b) = \varphi_2(b^*)$. Thus $\varphi_1$ and $\varphi_2$ coincide on all polynomials in $b$ and $b^*$, and therefore on $\mathcal{B}$. 

Now we have another proof of the continuous functional calculus for normal elements of a $C^*$-algebra:

18.14 THEOREM Let $\mathcal{A}$ be a unital $C^*$-algebra and $a \in \mathcal{A}$ normal. Then

(i) There is a unique unital $*$-homomorphism $\alpha_a : C(\sigma(a), \mathbb{C}) \rightarrow \mathcal{A}$ such that $\alpha_a(z) = a$, where $z$ is the inclusion map $\sigma(a) \hookrightarrow \mathbb{C}$. As in Section 13.2, we interpret $\alpha_a(f)$ as $f(a)$.

(ii) If $f \in C(\sigma(a), \mathbb{C})$ then $\sigma(f(a)) = f(\sigma(a))$.

(iii) If $f \in C(\sigma(a), \mathbb{C})$ and $g \in C(\sigma(a), \mathbb{C})$ then $(g \circ f)(a) = g(f(a))$.

Proof. (i) Let $\mathcal{B} = C^*(1,a) \subseteq \mathcal{A}$ be the closed $*$-subalgebra generated by $\{1,a\}$. Since $a$ is normal, $\mathcal{B}$ is a commutative unital $C^*$-algebra, thus by Theorem 18.11, there is an isometric $*$-isomorphism $\pi : \mathcal{B} \rightarrow C(\Omega(\mathcal{B}), \mathbb{C})$. And by Proposition 18.13 we have a homeomorphism $\widehat{a} : \Omega(\mathcal{B}) \rightarrow \sigma(a)$. Now we define $\alpha_a$ to be the composite of the maps

$$C(\sigma(a), \mathbb{C}) \xrightarrow{\alpha_a} C(\Omega(\mathcal{B}), \mathbb{C}) \xrightarrow{\pi^{-1}} \mathcal{B} \hookrightarrow \mathcal{A},$$

where the first map is $\alpha_a : f \mapsto f \circ \widehat{a}$. It is clear that $\alpha_a$ is a unital homomorphism. If $z : \sigma(a) \hookrightarrow \mathbb{C}$ is the inclusion, then $\alpha_a(z) = \pi^{-1}(z \circ \widehat{a}) = \pi^{-1}(\widehat{a}) = a$. Any continuous unital homomorphism $\alpha : C(\sigma(a)) \rightarrow \mathcal{B}$ sending 1 to $1_{\mathcal{A}}$ and $z$ to $a$ coincides with $\alpha_a$ on the polynomials $\mathbb{C}[x]$. Since the latter are dense in $C(\sigma(a), \mathbb{C})$ by Stone-Weierstrass, we have $\alpha = \alpha_a$.

(ii) As used above, $C^*$-subalgebra $\mathcal{B} = C^*(1,a)$ is abelian and there is an isometric $*$-isomorphism $\pi : \mathcal{B} \rightarrow C(\sigma(a), \mathbb{C})$. By construction of the functional calculus we have $\pi(f(a)) = f \circ \iota$, where $\iota$ is the inclusion map $\sigma(a) \hookrightarrow \mathbb{C}$. Now, with Theorem 12.26 and Exercise 11.6 we have

$$\sigma_\mathcal{A}(f(a)) = \sigma_\mathcal{B}(f(a)) = \sigma_{C(\sigma(a))}(\pi(f(a))) = \sigma_{C(\sigma(a))}(f \circ \iota) = f(\sigma(a)).$$

(iii) This is essentially obvious, since applying $f$ to $a$ and $g$ to $f(a)$ is just composition of maps on the right hand side of the Gelfand isomorphism.
18.15 Remark 1. It should be clear that Theorem 18.11 is conceptually of fundamental importance, but it is not easy to find applications that are not just applications of the continuous functional calculus for normal operators. The proof of the latter that we gave in Section 13.3 was a good deal more elementary than the one above: It did not involve the weak-$^\ast$ topology and Alaoglu’s theorem, and it only needed Weierstrass’ classical theorem (in two dimensions) rather than the more general result of Stone. We invoked Theorem 18.11 in proving the uniqueness part of Proposition 13.10, but there it can be avoided by using the results proven at the end of Section 13.2.

2. Enriching Theorem 18.11 by some considerations on von Neumann algebras (which we don’t define here) one can prove representation theorems for commutative von Neumann algebras, cf. e.g. [41, Chapter 6] or [49, Section 4.4], which add additional perspective to the spectral theorem for normal operators.

18.16 Exercise Let $\mathcal{A}$ be a $C^*$-algebra and $a, b \in \mathcal{A}$ commuting self-adjoint elements. Put $c = a + ib$.

(i) Prove $\sigma(a) = \text{Re}(\sigma(c))$ and $\sigma(b) = \text{Im}(\sigma(c))$.

(ii) Prove that the joint spectrum $\sigma(a, b)$ defined in Exercise 18.8 coincides with

$$\sigma(a, b) = \{(\text{Re} \lambda, \text{Im} \lambda) \mid \lambda \in \sigma(a + ib)\}.$$ 

A Some topics from topology and measure theory

A.1 Unordered sums

If $S$ is a finite set, $A$ an abelian group and $f : S \to A$ a function, it is not hard to define $\sum_{s \in S} f(s)$ (even though few textbook authors bother to do so explicitly). One chooses a bijection $\alpha : \{1, 2, \ldots, \#S\} \to S$ and defines $\sum_{s \in S} f(s) = \sum_{i=1}^{\#S} f(\alpha(i))$. The only slight difficulty is proving that the result does not depend on the choice of $\alpha$.

In order to define infinite sums, we need a topology on $A$, and we restrict to the case of functions $f : S \to V$, where $(V, \| \cdot \|)$ is a normed space. Many authors of introductory texts (for a nice exception see [77, Vol. I, Section 8.2]) consider only those countable sums known as series, but for our purposes this is inadequate.

A.1 Definition Let $S$ be a set, $(V, \| \cdot \|)$ a normed space and $f : S \to V$ a function. We say that $\sum_{s \in S} f(s)$ exists (or converges) and equals $x$ if there is $x \in V$ such that for every $\varepsilon > 0$ there is a finite subset $T \subseteq S$ such that $\|x - \sum_{s \in U} f(s)\| < \varepsilon$ holds for every finite $U \subseteq S$ containing $T$.

In many cases, the above will be applied to $V = \mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$ and $\| \cdot \| = | \cdot |$.

This notion of summation has some useful properties:

A.2 Proposition (i) If $\sum_{s \in S} f(s)$ exists then the sum $x \in V$ is uniquely determined.

(ii) If $\sum_{s \in S} f(s) = x$ and $\sum_{s \in S} g(s) = y$ then $\sum_{s \in S} (cf(s) + dg(s)) = cx + dy$ for all $c, d \in \mathbb{F}$.

(iii) If $f(s) \geq 0 \forall s \in S$ then $\sum_{s \in S} f(s)$ exists if and only if $\sup\{\sum_{t \in T} f(t) \mid T \subseteq S \text{ finite}\} < \infty$, in which case the two expressions coincide. These equivalent conditions imply that the set $\{s \in S \mid f(s) \neq 0\}$ is at most countable.
(iv) If \((V, \| \cdot \|)\) is complete and \(\sum_{s \in S} \| f(s) \| < \infty\) then \(\sum_{s \in S} f(s)\) exists, and \(\| \sum_{s \in S} f(s) \| \leq \sum_{s \in S} \| f(s) \|\).

(v) If \(f : S \to \mathbb{F}\) is such that \(\sum_{s \in S} f(s)\) exists then \(\sum_{s \in S} |f(s)|\) exists.

The proofs of (i) and (ii) are straightforward and similar to those for the analogous statements about series. The equivalence in (iii) follows from monotonicity of the map \(P_{\text{fin}}(S) \to [0, \infty),\ T \mapsto \sum_{t \in T} f(t).\) If \(\sum_{s \in S} f(s)\) \(\to \infty\) then it follows that for every \(\varepsilon > 0\) there are at most finitely many \(s \in S\) such that \(|f(s)| \geq \varepsilon\). In particular, for every \(n \in \mathbb{N}\) the set \(S_n = \{ s \in S \mid |f(s)| \geq 1/n \}\) is finite. Since a countable union of finite sets is countable, we have countability of \(\{ s \in S \mid f(s) \neq 0 \} = \bigcup_{n=1}^{\infty} S_n\). The proof of (iv) is analogous to that of the implication \(\Rightarrow\) in Proposition 4.2.

The statement (v) probably is surprising since the analogous statement for series is false. Roughly, the reason is that our definition of \(\sum_{s \in S} f(s)\) imposes no ordering on \(S\), while the sum of a series \(\sum_{n=1}^{\infty} f(n)\) is invariant under reordering of the terms only if the series converges absolutely. I strongly suggest that you make an effort to understand this! A rigorous proof of (v) can be found, e.g., in [46].

In discussing the spaces \(\ell^p(S, \mathbb{F})\), the following (which just is an easy special case of Lebesgue’s dominated convergence theorem) is useful:

A.3 Proposition (Discrete Case of Dominated Convergence Theorem) Let \(S\) be a set and \(\{ f_n \}_{n \in \mathbb{N}}\) functions \(S \to \mathbb{C}\). Assume that

1. For each \(s \in S\), the limit \(\lim_{n \to \infty} f_n(s)\) exists. Define \(h : S \to \mathbb{C}, s \mapsto \lim_{n \to \infty} f_n(s)\).
2. There exists a function \(g : S \to [0, \infty)\) such that \(\sum_{s \in S} g(s) < \infty\) and \(\sum_{s \in S} |f_n(s)| \leq g(s)\) \(\forall s \in S\).

Then

(i) \(\sum_{s \in S} f_n(s)\) converges for each \(n \in \mathbb{N}\). So does \(\sum_{s \in S} h(s)\).
(ii) \(\lim_{n \to \infty} \sum_{s \in S} f_n(s) = \sum_{s \in S} h(s)\). (Thus limit and summation can be interchanged.)

Proof. (i) Assumption 1. gives \(\| h(s) \| \leq g(s)\) \(\forall s\). Now assumption 2. implies convergence of \(\sum_s h(s)\) and of \(\sum_s f_n(s)\) for all \(n\).
(ii) Let \(\varepsilon > 0\). Since \(\sum_s g(s) < \infty\), there is a finite subset \(T \subseteq S\) such that \(\sum_{s \in S \setminus T} g(s) < \frac{\varepsilon}{4}\). For each \(t \in T\) there is an \(n_t \in \mathbb{N}\) such that \(n \geq n_t \Rightarrow |f_n(t) - h(t)| < \frac{\varepsilon}{2\#T}\). Put \(n_0 = \max_{t \in T} n_t\). If \(n \geq n_0\) then

\[
\left| \sum_{s \in S} f_n(s) - \sum_{s \in S} h(s) \right| \leq \sum_{s \in T} \left| f_n(s) - h(s) \right| + \sum_{s \in S \setminus T} \left| f_n(s) - h(s) \right|.
\]

The first term on the r.h.s. is bounded by

\[
\sum_{s \in T} |f_n(s) - h(s)| \leq \#T \cdot \frac{\varepsilon}{2\#T} = \frac{\varepsilon}{2}
\]

due to the definition of \(n_0\) and \(n \geq n_0 \geq n_t\). And the second is bounded by

\[
\sum_{S \setminus T} (|f_n(s)| + |h(s)|) \leq 2 \sum_{S \setminus T} g(s) \leq \frac{\varepsilon}{2},
\]

where we used that \(g\) dominates \(|f_n|\) and \(|h|\), as well as the choice of \(T\). Putting the two estimates together gives \(n \geq n_0 \Rightarrow \left| \sum_{s \in S} f_n(s) - \sum_{s \in S} h(s) \right| \leq \varepsilon\), completing the proof. \(\blacksquare\)
A.2 Nets

The Definition A.1 of unordered sums is an instance of a much more general notion, the convergence of nets.

A.4 Definition A directed set is a set $I$ equipped with a binary relation $\leq$ on $I$ satisfying
1. $a \leq a$ for each $a \in I$ (reflexivity).
2. If $a \leq b$ and $b \leq c$ for $a, b, c \in I$ then $a \leq c$ (transitivity).
3. For any $a, b \in I$ there exists a $c \in I$ such that $a \leq c$ and $b \leq c$ (directedness).

A.5 Remark If only 1. and 2. hold, $(I, \leq)$ is called a pre-ordered set. Some authors, as e.g. [41], require in addition that $a \leq b$ and $b \leq a$ together imply $a = b$ (antisymmetry). Recall that a pre-ordered set with this property is called partially ordered. But the antisymmetry is an unnatural assumption in this context and is never used.

A.6 Example 1. Every totally ordered set $(X, \leq)$ is a directed set. Only the directedness needs to be shown, and it follows by taking $c = \max(a, b)$. In particular $\mathbb{N}$ is a directed set with its natural total ordering.

2. If $S$ is a set then the power set $I = \mathcal{P}(S)$ with its natural partial ordering is directed: For the directedness, put $c = a \cup b$. The same works for the set $\mathcal{P}_{\text{fin}}(S)$ of finite subsets of $S$, which appeared in the definition of unordered sums.

3. If $(X, \tau)$ is a topological space and $x \in X$, let $U_x$ be the set of open neighborhoods of $x$. Now for $U, V \in U_x$, define $U \leq V \iff U \subseteq V$, thus we take the reversed ordering. Then $(U_x, \leq)$ is directed with $c = a \cap b$.

A.7 Definition If $X$ is a set, a net in $X$ is a map $I \to X$, $\iota \mapsto x_\iota$, where $(I, \leq)$ is a directed set.

If $(X, \tau)$ is a topological space, a net $\{x_\iota\}_{\iota \in I}$ in $X$ converges to $z \in X$ if for every open neighborhood $U$ of $z$ there is a $\iota_0 \in I$ such that $\iota \geq \iota_0 \Rightarrow x_\iota \in U$.

When this holds, we write $x_\iota \to z$ or $\lim_{\iota \in I} x_\iota = z$. (The second notation should only be used if $X$ is Hausdorff, since this property is equivalent to uniqueness of limits of nets.)

A.8 Remark 1. With $I = \mathbb{N}$ and $\leq$ the natural total ordering, a net indexed by $I$ just is a sequence, and this net converges if and only if the sequence does.

2. Unordered summation is a special case of a net limit: If $S$ is any set, let $I$ be the set of finite subsets of $S$ and let $\leq$ be the ordinary (partial) ordering of subsets of $S$. If $T, U \in I$ let $V = T \cup U$. Clearly $T \leq V, U \leq V$, showing that $(I, \leq)$ is a directed set. (This is the same as Example A.6.2, except that now we only look at finite subsets of $S$.) Now given $f : S \to \mathbb{F}$, for every $T \in I$, thus every finite $T \subseteq S$, we can clearly define $\sum_{t \in T} f(t)$. Now

$$\sum_{s \in S} f(s) = \lim_{T \in I} \sum_{t \in T} f(t),$$

where the sum exists if and only if the limit exists. □

Why nets? The reason is that sequences are totally inadequate for the study of topological spaces that do not satisfy the first countability axiom.\textsuperscript{53} Given a metric space $X$ and a subset

\textsuperscript{53}Unfortunately many introductory books and courses sweep this under the rug and don’t even mention nets.
Y \subseteq X$, one proves that $x \in Y$ if and only if there is a sequence $\{y_n\}$ in $Y$ converging to $x$, but for general topological spaces this is false. Similarly, the statement that a function $f : X \to Y$ is continuous at $x \in X$ if and only if $f(x_n) \to f(x)$ for every sequence $\{x_n\}$ converging to $x$ is true for metric spaces, but false in general! (It is instructive to work out counterexamples.)

On the other hand:

A.9 **Proposition 1.** Let $X$ be a topological space and $Y \subseteq X$. If $\{y_i\}$ is a net in $Y$ that converges to $x \in X$ then $x \in \overline{Y}$.

2. Let $X$ be a topological space and $Y \subseteq X$. Then for every $x \in \overline{Y}$ there exists a net $\{y_i\}$ in $Y$ such that $y_i \to x$.

3. A topological space $X$ is Hausdorff if and only if there exists no net $\{x_i\}$ in $X$ that converges to two different points of $X$.

4. If $X,Y$ are topological spaces, $f : X \to Y$ a function, and $x \in X$, then $f$ is continuous at $x$ if and only if $f(x_n) \to f(x)$ for every net $\{x_n\}$ in $X$ converging to $x$.

For proofs see [46] or any decent book on topology or [41, Section 5.5].

If $(X,d)$ is a metric space, the problems with sequences mentioned above do not arise. Nevertheless, there are situations where the use of nets in Theorem 6.37 and 6.40 where we considered nets indexed by the finite subsets of an ONB $E$. In this case one wants:

A.10 **Definition** A net $\{x_n\}$, indexed by a directed set $(I,\leq)$, in a metric space $(X,d)$ is a Cauchy net if for every $\varepsilon > 0$ there is an $i_0 \in I$ such that $i,i' \geq i_0 \Rightarrow d(x_i,x_{i'}) < \varepsilon$.

(In a normed space, this definition is consistent with the one in Remark 3.11.)

A.11 **Lemma** In a complete metric space every Cauchy net converges.

**Proof.** Let $\{x_i\}_{i \in I}$ be Cauchy. Then for every $n \in \mathbb{N}$ there is an $i_n \in I$ such that $i,i' \geq i_n \Rightarrow d(x_i,x_{i'}) < 1/n$. We can also arrange that $i_1 \leq i_2 \leq \cdots$ (using directedness to replace $i_2$ by some $i'_2$ larger than $i_1$ and $i_2$ etc.). Now it is quite clear that $\{x_{i_n}\}_{n \in \mathbb{N}}$ is a Cauchy sequence in $X$. By completeness of $(X,d)$ it converges to some $x \in X$. Let now $\varepsilon > 0$. Pick $n_0 \in \mathbb{N}$ such that $n \geq n_0 \Rightarrow d(x_{i_n},x) < \varepsilon/2$. Put $i_0 = i_{2n}$. If $i \geq i_0$ then $i \geq i_{2n} \geq i_n$, so that $d(x_i,x) \leq d(x_i,x_{i_{2n}}) + d(x_{i_{2n}},x) < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon$, thus $x_i \to x$. ■

A.3 **The Stone-Čech compactification**

If $X$ is a topological space, a compactification of $X$ is a space $\tilde{X}$ together with a continuous map $\iota : X \to \tilde{X}$ such that $\iota(X) \subseteq \tilde{X}$ is a dense subset and $\iota : X \to \iota(X)$ is a homeomorphism.

You certainly know the one-point or Alexandrov compactification of a topological space. (Usually it is considered only for locally compact spaces.) It is the smallest possible compactification in that it just adds one point.

But for many purposes, another compactification is more important, the Stone-Čech compactification. It is defined for spaces that have the following property:

A.12 **Definition** A topological space $X$ is completely regular if for every closed $C \subseteq X$ and $y \in X \setminus C$ there exists a continuous function $f : X \to [0,1]$ such that $f \upharpoonright C = 0$ and $f(x) = 1$.

All subspaces of a completely regular space are completely regular. By Urysohn’s lemma, every normal space is completely regular, in particular every metrizable and every compact Hausdorff space. This implies that complete regularity is a necessary condition for a space $X$ to have a compactification $\tilde{X}$ that is Hausdorff. In fact, it also is sufficient:
A.13 **Theorem** Let \( X \) be a topological space. Then the following are equivalent:

(i) \( X \) is completely regular.

(ii) There exists a compact Hausdorff space \( \beta X \) together with a dense embedding \( X \hookrightarrow \beta X \) such that for every continuous function \( f : X \to Y \), where \( Y \) is compact Hausdorff, there exists a continuous \( \hat{f} : \beta X \to Y \) such that \( \hat{f} \upharpoonright X = f \). (This \( \hat{f} \) is automatically unique by density of \( X \subseteq \beta X \).

The universal property (ii) implies that \( \beta X \) is unique up to homeomorphism. 'It' is called the Stone-Čech compactification of \( X \).

Let \( X \) be completely regular and \( \mathbb{F} \in \{ \mathbb{R}, \mathbb{C} \} \). Then the restriction map \( \mathcal{C}(\beta X, \mathbb{F}) \to \mathcal{C}_b(X, \mathbb{F}) \) given by \( f \mapsto f \upharpoonright X \) is a bijection and an isomorphism of commutative unital \( \mathbb{F} \)-algebras.

There are many ways to prove the non-trivial implication (i)\( \Rightarrow \) (ii), the most common one using Tychonov's theorem, cf. [46]. But we can also use Gelfand duality for commutative \( C^\ast \)-algebras, cf. Section 18. Here is a sketch: If \( X \) is completely regular, \( \mathcal{A} = C_b(X, \mathbb{C}) \) with norm \( \| f \| = \sup_x | f(x) | \) is a commutative unital \( C^\ast \)-algebra. As such it has a spectrum \( \Omega(\mathcal{A}) \), which is compact Hausdorff. We define \( \beta X = \Omega(\mathcal{A}) \). There is a map \( \iota : X \to \Omega(\mathcal{A}), x \mapsto \varphi_x \), where \( \varphi_x(f) = f(x) \). This map is continuous by definition of the topology on \( \Omega(\mathcal{A}) \). Using the complete regularity of \( X \) one proves that \( \iota \) is an embedding, i.e. a homeomorphism of \( X \) onto \( \iota(X) \subseteq \Omega(\mathcal{A}) \). Now \( \iota(X) = \Omega(\mathcal{A}) \) is seen as follows: \( \iota(X) \neq \Omega(\mathcal{A}) \) would imply (using Urysohn or Tietze) that there are \( f \in \mathcal{A} \setminus \{0\} \) such that \( \iota(x)(f) = 0 \) for all \( x \in X \). This is a contradiction, since the elements of \( \mathcal{A} \) are functions on \( X \), so that \( \iota(x)(f) = 0 \ \forall x \) implies \( f = 0 \).

### A.4 Reminder of the choice axioms and Zorn’s lemma

A.14 **Definition** The Axiom of Choice (AC) is any of the following statements, which are easily shown to be equivalent:

- If \( f : X \to Y \) is a surjective function then there exists a function \( g : Y \to X \) such that \( f \circ g = \text{id}_Y \).
- If \( X \) is a set, there exists a function \( s : P(X)\setminus\{\emptyset\} \to X \) such that \( s(Y) \in Y \) for each \( Y \in P(X)\setminus\{\emptyset\} \), i.e. \( \emptyset \neq Y \subseteq X \).
- If \( \{X_i\}_{i \in I} \) is a family of non-empty sets then \( \prod_{i \in I} X_i \neq \emptyset \). Concretely, there exists a map \( f : I \to \bigcup_{j \in I} X_j \) such that \( f(i) \in X_i \ \forall i \in I \).

A.15 **Definition** Let \( (X, \leq) \) be a partially ordered set. Then

- \( m \in X \) is called a maximal element if \( y \in X, y \geq m \) implies \( y = m \).
- \( u \in X \) is called an upper bound for \( Y \subseteq X \) if \( x \leq u \) holds for each \( u \in y \). If this \( u \) is in \( Y \) it is called largest element of \( Y \) (which is unique).

A.16 **Theorem** Given the Zermelo-Frenkel axioms of set theory, the Axiom of Choice is equivalent to Zorn’s lemma, which says: If \( (X, \leq) \) is a non-empty partially ordered set such that every totally ordered subset \( Y \subseteq X \) has an upper bound then \( X \) has a maximal element.

A.17 **Definition** The Axiom of Countable Choice (AC\(_\omega \)) is the first (or third) of the above versions of AC with the restriction that \( Y \) (respectively \( I \)) be at most countable.

---


A.18 Definition The Axiom of Countable Dependent Choice (DCω) is the following: If X is a set and \( R \subseteq X \times X \) is such that for every \( x \in X \) there is a \( y \in X \) such that \( (x, y) \in R \) then for each \( x_1 \in X \) there is a sequence \( \{x_n\} \) in X such that \( (x_n, x_{n+1}) \in R \) for all \( n \in \mathbb{N} \).

It is easy to prove \( AC \Rightarrow DC_\omega \Rightarrow AC_\omega \). The converse implications are false.

A.5 Baire’s theorem

You should have seen the following. (If not, cf. e.g. [46].)

A.19 Lemma Let \((X, \tau)\) be a topological space with \( X \neq \emptyset \). Then \( Y \subseteq X \) is dense if and only if \( Y \cap W \neq \emptyset \) whenever \( \emptyset \neq W \in \tau \). (Equivalently, \( X \setminus Y \) has empty interior.)

A.20 Theorem \(^{56}\) Let \((X, d)\) be a complete metric space and \( \{U_n\}_{n \in \mathbb{N}} \) a countable family of dense open subsets. Then \( \bigcap_{n=1}^{\infty} U_n \) is dense.

Proof. Let \( \emptyset \neq W \in \tau \). Since \( U_1 \) is dense, \( W \cap U_1 \neq \emptyset \) by Lemma A.19, so we can pick \( x_1 \in W \cap U_1 \). Since \( W \cap U_1 \) is open, we can choose \( \varepsilon_1 > 0 \) such that \( B(x_1, \varepsilon_1) \subseteq W \cap U_1 \). We may also assume \( \varepsilon_1 < 1 \). Since \( U_2 \) is dense, \( U_2 \cap B(x_1, \varepsilon_1) \neq \emptyset \) and we pick \( x_2 \in U_2 \cap B(x_1, \varepsilon_1) \). By openness, we can pick \( \varepsilon_2 \in (0, 1/2) \) such that \( B(x_2, \varepsilon_2) \subseteq U_2 \cap B(x_1, \varepsilon_1) \). Continuing this iteratively, we find points \( x_n \) and \( \varepsilon_n \in (0, 1/n) \) such that \( B(x_n, \varepsilon_n) \subseteq U_n \cap B(x_{n-1}, \varepsilon_{n-1}) \) \( \forall n \). If \( i > n \) and \( j > n \) we have by construction that \( x_i, x_j \in B(x_n, \varepsilon_n) \) and thus \( d(x_i, x_j) \leq 2\varepsilon_n < 2/n \). Thus \( \{x_n\} \) is a Cauchy sequence, and by completeness it converges to some \( z \in X \). Since \( n > k \Rightarrow x_n \in B(x_k, \varepsilon_k) \), the limit \( z \) is contained in \( B(x_k, \varepsilon_k) \) for each \( k \), thus

\[
z \in \bigcap_n B(x_n, \varepsilon_n) \subseteq W \cap \bigcap_n U_n,
\]

thus \( W \cap \bigcap_n U_n \) is non-empty. Since \( W \) was an arbitrary non-empty open set, Lemma A.19 gives that \( \bigcap_n U_n \) is dense. \( \square \)

The following (equivalent) reformulation is useful:

A.21 Corollary Let \((X, d)\) be a complete metric space and \( \{C_n\}_{n \in \mathbb{N}} \) a countable family of closed subsets with empty interior. Then \( \bigcup_{n=1}^{\infty} C_n \) has empty interior.

Proof. Immediate by de Morgan and the preceding material. \( \square \)

A.22 Remark 1. There are many other ways of stating Baire’s theorem, but most of the alternative versions introduce additional terminology (nowhere dense sets, meager sets, sets of first or second category, etc.) and tend to obscure the matter unnecessarily.

2. An intersection \( \bigcap_n U_n \) of a countable family \( \{U_n\}_{n \in \mathbb{N}} \) of open sets is called a \( G_\delta \)-set.

3. The proof implicitly used the axiom DCω of countable dependent choice. (Making this explicit would be a tedious exercise.) Remarkably, one can prove that the (Zermelo-Frenkel) axioms of set theory (without any choice axiom) combined with Baire’s theorem imply DCω.

4. Some results usually proven using Baire’s theorem can alternatively be proven without it. But in most cases, such alternative proofs will also use DCω and therefore not be better from a foundational point of view. The proof of Theorem 9.2 is a rare exception. \( \square \)

---

\(^{56}\) René-Louis Baire (1874-1932). French mathematician, proved this for \( \mathbb{R}^n \) in his 1899 doctoral thesis. The generalization is due to Hausdorff (1914).
A typical application of Baire’s theorem is the following (for a proof see, e.g., [46]):

A.23 Theorem There is a $\| \cdot \|_\infty$-dense $G_d$-set $F \subseteq C([0,1], \mathbb{R})$ such that every $f \in F$ is nowhere differentiable.

Note that a single function $f \in C([0,1], \mathbb{R})$ that is nowhere differentiable can be written down quite explicitly and constructively, for example $f(x) = \sum_{n=1}^{\infty} 2^{-n} \cos(2^n x)$. But for proving that such functions are dense one needs Baire’s theorem.

A.6 Tietze’s extension theorem

A.24 Theorem (Tietze-Urysohn extension theorem) \(^{57}\) Let $(X, \tau)$ be normal, $Y \subseteq X$ closed and $f \in C_0(Y, \mathbb{R})$. Then there exists $\hat{f} \in C_0(X, \mathbb{R})$ such that $\hat{f}|_Y = f$ and $\|\hat{f}\| = \|f\|$.

(In other words, the restriction map $T : C_0(X, \mathbb{R}) \to C_0(Y, \mathbb{R})$, $f \mapsto f|_Y$ is surjective.)

Proof. Let $f \in C_0(Y, \mathbb{R})$, where we may assume $\|f\| = 1$, i.e. $f(Y) \subseteq [-1,1]$. Let $A = f^{-1}([-1, -1/3])$ and $B = f^{-1}([1/3, 1])$. Then $A, B$ are disjoint closed subsets of $Y$, which are also closed in $X$ since $Y$ is closed. Thus by Urysohn’s Lemma, there is a $g \in C(X, [-1/3, 1/3])$ such that $g|_A = -1/3$ and $g|_B = 1/3$. Thus $\|g\|_X = 1/3$ and $\|Tg - f\|_Y \leq 2/3$. (You should check this!) Now Lemma 10.2 is applicable with $m = 1/3$ and $r = 2/3$ and gives the existence of $\hat{f} \in C(X, \mathbb{R})$ with $T \hat{f} = f$ and $\|\hat{f}\| = \|f\|$ (since $m/(1 - r) = 1$). \(\square\)

The theorem is easily extended to $C$-valued functions.

A.7 The Stone-Weierstrass theorem

A.7.1 Weierstrass’ theorem

The following fundamental theorem of Weierstrass\(^{58}\) (1885) has been proven in many ways. A fairly standard proof due to E. Landau (1908) involves convolution of $f$ with a sequence $\{g_n\}$ of functions that is a polynomial approximate unit, cf. e.g. [77, Vol. II, Section 3.8]. The following proof, given in 1913 by Sergei Bernstein\(^{59}\), has the advantage of using no integration.

A.25 Theorem Let $f \in C([a,b], \mathbb{F})$ and $\varepsilon > 0$. Then there exists a polynomial $P \in \mathbb{F}[x]$ such that $|f(x) - P(x)| \leq \varepsilon$ for all $x \in [a,b]$. (As always, $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$.)

Proof. It clearly suffices to prove this for the interval $[0,1]$. For $n \in \mathbb{N}$ and $x \in [0,1]$, define

$$P_n(x) = \sum_{k=0}^{n} f\left(\frac{k}{n}\right) \binom{n}{k} x^k (1-x)^{n-k}.$$ 

Clearly $P_n$ is a polynomial of degree at most $n$, called Bernstein polynomial. In view of

$$1 = 1^n = (x + (1-x))^n = \sum_{k=0}^{n} \binom{n}{k} x^k (1-x)^{n-k} \quad \text{(A.1)}$$

\(^{57}\)H. F. F. Tietze (1880-1964), Austrian mathematician. He proved this for metric spaces (for which Urysohn’s lemma is a triviality). The generalization to normal spaces is due to Urysohn.

\(^{58}\)Karl Theodor Wilhelm Weierstrass (1815-1897). German mathematician and one of the fathers of rigorous analysis.

\(^{59}\)Sergei Natanovich Bernstein (1880-1968). Russian/Soviet mathematician. Important contributions to approximation theory, probability, PDEs.
we have
\[ f(x) - P_n(x) = \sum_{k=0}^{n} \left( f(x) - f \left( \frac{k}{n} \right) \right) \binom{n}{k} x^k (1-x)^{n-k}, \]
thus
\[ |f(x) - P_n(x)| \leq \sum_{k=0}^{n} \left| f(x) - f \left( \frac{k}{n} \right) \right| \binom{n}{k} x^k (1-x)^{n-k}. \]  \hspace{1cm} (A.2)

Since \([0,1]\) is compact and \(f : [0,1] \to \mathbb{F}\) is continuous, it is bounded and uniformly continuous. Thus there is \(M\) such that \(|f(x)| \leq M\) for all \(x\), and for each \(\varepsilon > 0\) there is \(\delta > 0\) such that \(|x-y| < \delta \Rightarrow |f(x) - f(y)| < \varepsilon\).

Let \(\varepsilon > 0\) be given, and chose a corresponding \(\delta > 0\) as above. Let \(x \in [0,1]\). Define
\[ A = \left\{ k \in \{0,1,\ldots,n\} \mid \left| \frac{k}{n} - x \right| < \delta \right\}. \]
For all \(k\) we have \(|f(x) - f(k/n)| \leq 2M\), and for \(k \in A\) we have \(|f(x) - f(k/n)| < \varepsilon\). Thus with (A.2) we have
\[ |f(x) - P_n(x)| \leq \varepsilon \sum_{k \in A} \binom{n}{k} x^k (1-x)^{n-k} + 2M \sum_{k \in A^c} \binom{n}{k} x^k (1-x)^{n-k} \]
\[ \leq \varepsilon + 2M \sum_{k \in A^c} \binom{n}{k} x^k (1-x)^{n-k}, \]  \hspace{1cm} (A.3)
where we used (A.1) again. In an exercise, we will prove the purely algebraic identity
\[ \sum_{k=0}^{n} \binom{n}{k} x^k (1-x)^{n-k}(k-nx)^2 = nx(1-x) \]  \hspace{1cm} (A.4)
for all \(n \in \mathbb{N}_0\) and \(x \in [0,1]\) (in fact all \(x \in \mathbb{R}\)). Now, \(k \in A^c\) is equivalent to \(\left| \frac{k}{n} - x \right| \geq \delta\) and to \((k-nx)^2 \geq n^2\delta^2\). Multiplying both sides of the latter inequality by \(\binom{n}{k} x^k (1-x)^{n-k}\) and summing over \(k \in A^c\), we have
\[ n^2\delta^2 \sum_{k \in A^c} \binom{n}{k} x^k (1-x)^{n-k} \leq \sum_{k \in A^c} \binom{n}{k} x^k (1-x)^{n-k}(k-nx)^2 \]
\[ \leq \sum_{k=0}^{n} \binom{n}{k} x^k (1-x)^{n-k}(k-nx)^2 = nx(1-x), \]  \hspace{1cm} (A.5)
where the last equality comes from (A.4). This implies
\[ \sum_{k \in A^c} \binom{n}{k} x^k (1-x)^{n-k} \leq \frac{nx(1-x)}{n^2\delta^2} \leq \frac{1}{n\delta^2}, \]  \hspace{1cm} (A.6)
where we used the obvious inequality \(x(1-x) \leq 1\) for \(x \in [0,1]\). Plugging (A.6) into (A.3) we have \(|f(x) - P_n(x)| \leq \varepsilon + \frac{2M}{n}\delta^2\). This holds for all \(x \in [0,1]\) since, by uniform continuity, \(\delta\) depends only on \(\varepsilon\), not on \(x\). Thus for \(n > \frac{2M}{\varepsilon\delta^2}\) we have \(|f(x) - P_n(x)| \leq 2\varepsilon\ \forall \ x \in [0,1]\) and are done.

**A.26 Exercise** Prove (A.4). Hint: Use basic properties of the binomial coefficients or differentiate \((x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k}\) twice with respect to \(x\) and then put \(y = 1 - x\).
An immediate consequence of Theorem A.25 is the following:

A.27 COROLLARY There exists a sequence \( \{p_n\}_{n \in \mathbb{N}} \subseteq \mathbb{R}[x] \) of real polynomials that converges uniformly on \([0, 1]\) to the function \( x \mapsto \sqrt{x} \).

The above corollary can also be proven directly:

A.28 EXERCISE Define a sequence \( \{p_n\}_{n \in \mathbb{N}_0} \subseteq \mathbb{R}[x] \) of polynomials by \( p_0 = 0 \) and

\[
p_{n+1}(x) = p_n(x) + \frac{x - p_n(x)^2}{2}.
\]

Prove by induction that the following holds:

(i) \( p_n(x) \leq \sqrt{x} \) for all \( n \in \mathbb{N}_0, x \in [0, 1] \).

(ii) The sequence \( \{p_n(x)\} \) increases monotonously for each \( x \in [0, 1] \) and converges uniformly to \( \sqrt{x} \).

A.7.2 The Stone-Weierstrass theorem in the simplest case

Theorem A.25 says that the polynomials, restricted to \([0, 1]\) are uniformly dense in \( C([0, 1]) \). Our aim is to generalize this, replacing replacing \([0, 1]\) by (locally) compact Hausdorff spaces. In order to see what should take the place of polynomials, notice that a polynomial on \( \mathbb{R} \) is a linear combination of powers \( x^n \), and the latter can be seen as powers \( f^n \) (under pointwise multiplication) of the identity function \( f = \text{id}_{\mathbb{R}} \). Thus the polynomials are the unital subalgebra \( P \subseteq C(\mathbb{R}, \mathbb{R}) \) generated by the single element \( \text{id}_{\mathbb{R}} \). Now, if \( X \) is a topological space and \( F \in \{\mathbb{R}, \mathbb{C}\} \) then \( C(X, F) \) is a unital algebra, and we will consider subalgebras (not necessarily singly generated) \( A \subseteq C(X, F) \). Since the functions on a (locally) compact Hausdorff space separate points, we clearly need to impose the following if we want to prove \( \overline{A} = C(X, F) \):

A.29 DEFINITION A subalgebra \( A \subseteq C(X, F) \) separates points if for any \( x, y \in X, x \neq y \) there is a \( f \in A \) such that \( f(x) \neq f(y) \).

A.30 THEOREM (M. H. STONE 1937) If \( X \) is compact Hausdorff and \( A \subseteq C(X, \mathbb{R}) \) is a unital subalgebra separating points then \( \overline{A} = C(X, \mathbb{R}) \).

Proof. Replacing \( A \) by \( \overline{A} \), the claim is equivalent to showing that \( A = C(X, \mathbb{R}) \). We proceed in several steps. We claim that \( f \in A \) implies \( |f| \in A \). Since \( f \) is bounded due to compactness, it clearly is enough to prove this under the assumption \( |f| \leq 1 \). With the \( p_n \) of Corollary A.27, we have \( x \mapsto p_n(f^2(x)) \in A \) since \( A \) is a unital algebra. Since \( p_n \circ f^2 \) converges uniformly to \( \sqrt{f^2} = |f| \), closedness of \( A \) implies \( |f| \in A \). In view of

\[
\max(f, g) = \frac{f + g + |f - g|}{2}, \quad \min(f, g) = \frac{f + g - |f - g|}{2},
\]

and the preceding result, we see that \( f, g \in A \) implies \( \min(f, g), \max(f, g) \in A \). By induction, this extends to pointwise minima/maxima of finite families of elements of \( A \).

Now let \( f \in C(X, \mathbb{R}) \). Our goal is to find \( f_\varepsilon \in A \) satisfying \( \|f - f_\varepsilon\| < \varepsilon \) for each \( \varepsilon > 0 \). Since \( A \) is closed, this will give \( A = C(X, \mathbb{R}) \).

If \( a \neq b \), the fact that \( A \) separates points gives us an \( h \in A \) such that \( h(a) \neq h(b) \). Thus the function \( h_{a,b}(x) = \frac{h(x) - h(a)}{h(b) - h(a)} \) is in \( A \), continuous and satisfies \( h(a) = 0, h(b) = 1 \). Thus also
\[ f_{a,b}(x) = f(a) + (f(b) - f(a))h_{a,b}(x) \] 

is in \( A \), and it satisfies \( f_{a,b}(a) = f(a) \) and \( f_{a,b}(b) = f(b) \).

This implies that the sets

\[ U_{a,b,\varepsilon} = \{ x \in X \mid f_{a,b}(x) < f(x) + \varepsilon \}, \quad V_{a,b,\varepsilon} = \{ x \in X \mid f_{a,b}(x) > f(x) - \varepsilon \} \]

are open neighborhoods of \( a \) and \( b \), respectively, for every \( \varepsilon > 0 \). Thus keeping \( b, \varepsilon \) fixed, \( \{U_{a,b,\varepsilon}\}_{a \in X} \) is an open cover of \( X \), and by compactness we find a finite subcover \( \{U_{a_i,b,\varepsilon}\}_{i=1}^n \).

By the above preparation, the function \( f_{b,\varepsilon} = \min(f_{a_1,b,\varepsilon}, \ldots, f_{a_n,b,\varepsilon}) \) is in \( A \). If \( x \in U_{a_i,b,\varepsilon} \) then \( f_{b,\varepsilon}(x) \leq f_{a_i,b,\varepsilon}(x) < f(x) + \varepsilon \) for all \( x \in X \), and since \( \{U_{a_i,b,\varepsilon}\}_{i=1}^n \) covers \( X \), we have \( f_{b,\varepsilon}(x) < f(x) + \varepsilon \) for all \( x \in X \).

For all \( x \in V_{b,\varepsilon} = \bigcap_{i=1}^n V_{a_i,b,\varepsilon} \), we have \( f_{a_i,b,\varepsilon}(x) > f(x) - \varepsilon \), and therefore \( f_{b}(x) = \min_i(f_{a_i,b,\varepsilon}) > f(x) - \varepsilon \). Now \( \{V_{b,\varepsilon}\}_{b \in X} \) is an open cover of \( X \), and we find a finite subcover \( \{V_{b_j,\varepsilon}\}_{j=1}^m \). Then \( f_{\varepsilon} = \max(f_{b_1,\varepsilon}, \ldots, f_{b_m}) \) is in \( A \). Now \( f_{\varepsilon}(x) = \max_j(f_{b_j,\varepsilon}) \leq f(x) + \varepsilon \)

holds everywhere, and for \( x \in V_{b_j,\varepsilon} \) we have \( f_{\varepsilon}(x) \geq f_{b_j,\varepsilon} > f(x) - \varepsilon \). Since \( \{V_{b_j,\varepsilon}\}_j \) covers \( X \), we conclude that \( f_{\varepsilon}(a) \in (f(x) - \varepsilon, f(x) + \varepsilon) \) for all \( x \), to wit \( \|f - f_{\varepsilon}\| < \varepsilon \). \( \blacksquare \)

Since the polynomial ring \( \mathbb{R}[x] \) is an algebra, and the polynomials clearly separate the points of \( \mathbb{R} \), Theorem A.30 recovers Theorem A.25. (This is not circular if one has used Exercise A.28 to prove Corollary A.27.) But we immediately have the higher dimensional generalization (which can also be proved by more classical methods, like approximate units):

A.31 Theorem Let \( X \subseteq \mathbb{R}^n \) be compact. Then the restrictions to \( X \) of the \( P \in \mathbb{R}[x_1, \ldots, x_n] \) (considered as functions) are uniformly dense in \( C(X, \mathbb{R}) \).

A.7.3 Generalizations

Having proven Theorem A.30, it is easy to generalize it to locally compact spaces or/and subalgebras of \( C(0, X, \mathbb{C}) \).\(^{60}\) Recall that a subset \( S \) of a \( * \)-algebra \( A \) is called self-adjoint if \( S = S^* := \{ s^* \mid s \in S \} \).

A.32 Corollary If \( X \) is compact Hausdorff and \( A \subseteq C(X, \mathbb{C}) \) is a self-adjoint unital subalgebra separating points then \( \overline{A} = C(X, \mathbb{C}) \).

Proof. Define \( B = A \cap C(X, \mathbb{R}) \). Let \( f \in A \). Since \( f^* \in A \), we also have \( \text{Re}(f) = \frac{f + f^*}{2} \in B \)

and \( \text{Im}(f) = \frac{f - f^*}{2i} = -\text{Re}(if) \in B \). Thus \( A = B + iB \). It is obvious that \( \text{Re}(A) \subseteq C(X, \mathbb{R}) \) is a unital subalgebra. If \( x \neq y \) then there is \( f \in C(X, \mathbb{C}) \) such that \( f(x) \neq f(y) \).

Thus \( \text{Re}(f)(x) \neq \text{Re}(f)(y) \) or \( \text{Re}(f)(x) = \text{Re}(f)(y) \) (or both). Since \( \text{Re}(f), \text{Re}(if) \in B \), we see that \( B \) separates points. Thus \( \overline{B} = C(X, \mathbb{R}) \) by Theorem A.30, implying \( \overline{A} = B + iB = B + iB = C(X, \mathbb{R}) + iC(X, \mathbb{R}) = C(X, \mathbb{C}) \). \( \blacksquare \)

A.33 Definition A subalgebra \( A \subseteq C_0(X, \mathbb{F}) \) vanishes at no point if for every \( x \in X \) there is an \( f \in A \) such that \( f(x) \neq 0 \).

A.34 Corollary If \( X \) is locally compact Hausdorff and \( A \subseteq C_0(X, \mathbb{R}) \) is a subalgebra separating points and vanishing at no point then \( \overline{A} = C_0(X, \mathbb{R}) \).

\(^{60}\)Some authors, mostly operator algebraists, write \( C(X) \) for \( C(X, \mathbb{C}) \), whereas topologists put \( C(X) = C(X, \mathbb{R}) \). I don’t use \( C(X) \).
Proof. Let $X_\infty = X \cup \{\infty\}$ be the one-point compactification of $X$. Recall that every $f \in C_0(X,\mathbb{R})$ extends to $\hat{f} \in C(X_\infty,\mathbb{R})$ with $\hat{f}(\infty) = 0$. Then $B = \{\hat{f} \mid f \in A\} + \mathbb{R}1$ clearly is a unital subalgebra of $C(X_\infty,\mathbb{R})$. We claim that $B$ separates the points of $X_\infty$. This is obvious for $x, y \in X$, $x \neq y$ since already $A$ does that. Now let $x \in X$. Since $A$ vanishes at no point, there is $f \in A$ such that $f(x) \neq 0$. Let $\hat{f} \in C(X,\mathbb{R})$ be the extension to $X_\infty$ with $\hat{f}(\infty) = 0$. In view of $\hat{f}(x) = f(x) \neq 0$, we see that $B$ also separates $\infty$ from the points of $X$, so that Theorem A.30 gives $\overline{B} = C(X,\mathbb{R})$. In view of $\overline{B} = \overline{A} + \mathbb{R}1$ and $C(X,\mathbb{R}) \mid X = C_0(X,\mathbb{R})$, we have $\overline{A} = \overline{B} \mid X = C_0(X,\mathbb{R})$. $\blacksquare$

A.35 Corollary If $X$ is locally compact Hausdorff and $A \subseteq C_0(X,\mathbb{C})$ is a self-adjoint subalgebra separating points and vanishing at no point then $\overline{A} = C_0(X,\mathbb{C})$.

Proof. The proof just combines the ideas of the proofs of Corollaries A.32 and A.34. $\blacksquare$

A.8 Totally bounded sets in metric spaces

Recall that a metric space $(X,d)$ is totally bounded if for every $\varepsilon > 0$ there are $x_1,\ldots,x_n \in X$ such that $X = B(x_1,\varepsilon) \cup \cdots \cup B(x_n,\varepsilon)$. And: A metric space is compact if and only if it is complete and totally bounded, cf. e.g. [46]. We will need the following:

A.36 Exercise Let $(X,d)$ be a metric space. Prove:

(i) If $(X,d)$ is totally bounded and $Y \subseteq X$ then $(Y,d)$ is totally bounded.
(ii) If $(Y,d)$ is totally bounded and $Y \subseteq X$ is dense then $(X,d)$ is totally bounded.
(iii) If $(X,d)$ is complete and $Y \subseteq X$ then $(Y,d)$ is totally bounded if and only if $Y$ is precompact.

A.9 The Arzelà-Ascoli theorem

If $(X,\tau)$ is a topological space and $(Y,d)$ metric, the set $C_b(X,Y)$ is topologized by the metric

$$D(f,g) = \sup_{x \in X} d(f(x),g(x)).$$

It is therefore natural to ask whether the (relative) compactness of a set $\mathcal{F} \subseteq C_b(X,Y)$ can be characterized in terms of the elements of $\mathcal{F}$, which after all are functions $f : X \to Y$. This will be the subject of this section, but we will restrict ourselves to compact $X$, for which $C(X,Y) = C_b(X,Y)$.

A.37 Definition Let $(X,\tau)$ be a topological space and $(Y,d)$ a metric space. A family $\mathcal{F}$ of functions $X \to Y$ is called equicontinuous if for every $x \in X$ and $\varepsilon > 0$ there is an open neighborhood $U \ni x$ such that $f \in \mathcal{F}$, $x' \in U \Rightarrow d(f(x),f(x')) < \varepsilon$. Then $\mathcal{F} \subseteq C(X,Y)$.

The point of course is that the choice of $U$ depends only on $x$ and $\varepsilon$, but not on $f \in \mathcal{F}$.

A.38 Theorem (Arzelà-Ascoli) ²¹ Let $(X,\tau)$ be a compact topological space and $(Y,d)$ a complete metric space. Then $\mathcal{F} \subseteq C(X,Y)$ is (pre)compact (w.r.t. the uniform topology $\tau_D$) if and only if

²¹Giulio Ascoli (1843-1896), Cesare Arzelà (1847-1912), Italian mathematicians. They proved special cases of this result, of which there also exist more general versions than the one above.

126
\begin{itemize}
\item \(\{f(x) \mid f \in \mathcal{F}\} \subseteq Y\) is (pre)compact for every \(x \in X\),
\item \(\mathcal{F}\) is equicontinuous.
\end{itemize}

\textbf{Proof.}\ \Rightarrow \text{If } f,g \in C(X,Y) \text{ then } d(f(x),g(x)) \leq D(f,g) \text{ for every } x \in X. \text{ This implies that the evaluation map } e_x : C(X,Y) \to Y, f \mapsto f(x) \text{ is continuous for every } x. \text{ Thus if } \mathcal{F} \text{ is compact, so is } e_X(\mathcal{F}). \text{ And compactness of } \mathcal{F} \text{ implies that } e_X(\mathcal{F}) = \{f(x) \mid f \in \mathcal{F}\} \text{ is compact, thus closed.}

Since \(e_X(\mathcal{F})\) contains \(e_X(\mathcal{F})\), also \(e_x(\mathcal{F}) \subseteq e_X(\mathcal{F})\) is compact.

To prove equicontinuity, let \(x \in X\) and \(\varepsilon > 0\). Since \(\mathcal{F}\) is compact, \(\mathcal{F}\) is totally bounded, thus there are \(g_1, \ldots, g_n \in \mathcal{F}\) such that \(\mathcal{F} \subseteq \bigcup_i B(g_i, \varepsilon)\). By continuity of the \(g_i\), there are open \(U_i \ni x, i = 1, \ldots, n\), such that \(x' \in U_i \Rightarrow d(g_i(x), g_i(x')) < \varepsilon\). Put \(U = \bigcap_i U_i\). If now \(f \in \mathcal{F}\), there is an \(i\) such that \(f \in B(g_i, \varepsilon)\), to wit \(D(f, g_i) < \varepsilon\). Now for \(x' \in U \subseteq U_i\) we have

\[
d((f(x), f(x')) \leq d((f(x), g_i(x))) + d(g_i(x), g_i(x')) + d(g_i(x'), f(x')) < 3\varepsilon,
\]

proving equicontinuity of \(\mathcal{F}\) (at \(x\), but \(x\) was arbitrary).

\(\Leftarrow\) We first prove a lemma:

\textbf{A.39 Lemma} Let \((X, \mathcal{d})\) be a metric space. Assume that for each \(\varepsilon > 0\) there are a \(\delta > 0\), a metric space \((Y, \mathcal{d}')\) and a continuous map \(h : X \to Y\) such that \((h(X), \mathcal{d}')\) is totally bounded and such that \(\mathcal{d}'(h(x), h(x')) < \delta\) implies \(\mathcal{d}(x, x') < \varepsilon\). Then \((X, \mathcal{d})\) is totally bounded.

\textbf{Proof.}\ For \(\varepsilon > 0\), pick \(\delta, (Y, \mathcal{d}'), h\) as asserted. Since \(h(X)\) is totally bounded, there are \(y_1, \ldots, y_n \in h(X)\) such that \(h(X) \subseteq \bigcup_i B(y_i, \delta) \subseteq Y\). Then \(X = \bigcup_i h^{-1}(B(y_i, \delta))\). For each \(i\) choose \(x_i \in X\) such that \(h(x_i) = y_i\). Now \(x \in h^{-1}(B(y_i, \delta)) \Rightarrow d(h(x), y_i) < \delta \Rightarrow d(x, x_i) < \varepsilon\), so that \(h^{-1}(B(y_i, \delta)) \subseteq B(x_i, \varepsilon)\). Thus \(X = \bigcup_{i=1}^n B(x_i, \varepsilon)\), and \((X, \mathcal{d})\) is totally bounded. \(\blacksquare\)

Let \(\varepsilon > 0\). Since \(\mathcal{F}\) is equicontinuous, for every \(x \in X\) there is an open neighborhood \(U_x\) such that \(f \in \mathcal{F}, x' \in U_x \Rightarrow d(f(x), f(x')) < \varepsilon\). Since \(X\) is compact, there are \(x_1, \ldots, x_n \in X\) such that \(X = \bigcup_{i=1}^n U_{x_i}\). Now define \(h : \mathcal{F} \to Y^{\times n} : f \mapsto (f(x_1), \ldots, f(x_n))\). Now \(d((y_1, \ldots, y_n), (y'_1, \ldots, y'_n)) = \sum_i d(y_i, y'_i)\) is a product metric on \(Y^{\times n}\) making \(h\) continuous. By assumption \(\{f(x) \mid f \in \mathcal{F}\}\) is compact for each \(x \in X\), thus \(h(\mathcal{F}) \subseteq \prod_i \{f(x_i) \mid f \in \mathcal{F}\} \subseteq Y^{\times n}\) is compact, thus \((h(\mathcal{F}), \tilde{d})\) is totally bounded. If now \(f, g \in \mathcal{F}\) satisfy \(\tilde{d}(h(f), h(g)) < \varepsilon\) then \(d(f(x_i), g(x_i)) < \varepsilon \forall i\) by definition of \(\tilde{d}\). For every \(x \in X\) there is \(i\) such that \(x \in U_{x_i}\), thus

\[
d((f(x), g(x)) \leq d((f(x), f(x_i))) + d(f(x_i), g(x_i)) + d(g(x_i), g(x)) < 3\varepsilon.
\]

Since this holds for all \(x \in X\), we have \(D(f, g) \leq 3\varepsilon\). Thus the assumptions of Lemma A.39 are satisfied, and we obtain total boundedness, thus precompactness, of \(\mathcal{F}\). \(\blacksquare\)

\textbf{A.40 Remark}\ 1. If \(Y = \mathbb{R}^n\), as in most statements of the theorem, then in view of the Heine-Borel theorem the requirement of precompactness of \(\{f(x) \mid f \in \mathcal{F}\}\) for each \(x\) reduces to that of boundedness, i.e. pointwise boundedness of \(\mathcal{F}\). One can also formulate the theorem in terms of existence of uniformly convergent (or Cauchy) subsequences of bounded equicontinuous sequences in \(C(X, \mathbb{R}^n)\).

2. We intentionally stated a more general version of the theorem than needed in order to argue that the result belongs to general topology rather than functional analysis. For \(Y = \mathbb{R}^n\) this is less clear, also since there are many alternative proofs of the theorem using various methods from topology and functional analytis, cf. e.g. [50]. (This is no surprise since, as explained in [46], the theorems of Alaoglu, the Stone-Čech compactification and Tychonov’s theorem for Hausdorff spaces are all equivalent, i.e. easily deducible from each other.) \(\Box\)
A.10 Some notions from measure and integration theory

A.41 Definition If $X$ is a set, a σ-algebra on $X$ is a family $\mathcal{A} \subseteq \mathcal{P}(X)$ of subsets such that

1. $\emptyset \in \mathcal{A}$.
2. If $A \in \mathcal{A}$ then $X \setminus A \in \mathcal{A}$.
3. If $\{A_n\}_{n \in \mathbb{N}} \subseteq \mathcal{A}$ then $\bigcup_{n=1}^{\infty} A_n \in \mathcal{A}$.

A measurable space is a pair $(X, \mathcal{A})$ consisting of a set and a σ-algebra on it.

The closedness of $\mathcal{A}$ under complements implies that a σ-algebra also contains $X$ and is closed under countable intersections. Obviously $\mathcal{P}(X)$ is a σ-algebra.

It is very easy to see that the intersection of any number of σ-algebras on $X$ is a σ-algebra on $X$. Thus if $\mathcal{F} \subseteq \mathcal{P}(X)$ is any family of subsets of $X$, we can define the σ-algebra generated by $\mathcal{F}$ as the intersection of all σ-algebras on $X$ that contain $\mathcal{F}$.

If $(X, \tau)$ is a topological space, the σ-algebra on $X$ generated by $\tau$ is called the Borel\textsuperscript{62} σ-algebra $\mathcal{B}(X)$ of $X$. (We should of course write $\mathcal{B}(X, \tau)$.) Apart from the open sets, it contains the closed sets, the $G_\delta$ sets and many more. A function $f : X \to \mathbb{C}$ is called Borel measurable if $f^{-1}(U) \in \mathcal{B}(X)$ for every open $U \subseteq \mathbb{C}$. (This is equivalent to $f^{-1}(B) \in \mathcal{B}(X)$ for every $B \in \mathcal{B}(\mathbb{C})$.) If $(X, \mathcal{A})$ is a measurable space, $\mathcal{B}^\infty(X, \mathbb{C})$ denotes the set of functions $f : X \to \mathbb{C}$ that are Borel-measurable and bounded, i.e. $\sup_{x \in X} |f(x)| < \infty$. It is not hard to check that this is an algebra (with the pointwise product).

A.42 Definition A positive measure on a measurable space $(X, \mathcal{A})$ is a map $\mu : \mathcal{A} \to [0, \infty]$ such that $\mu(\emptyset) = 0$ and $\mu(\bigcup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} \mu(A_n)$ whenever $\{A_n\} \subseteq \mathcal{A}$ is a countable family of mutually disjoint sets. If $\mu(X) < \infty$ then $\mu$ is called finite (then $\mu(A) \leq \mu(X) \forall A \in \mathcal{A}$).

A Borel measure on a topological space $(X, \tau)$ is a positive measure on $(X, \mathcal{B}(X))$.

There is a notion of regularity of a measure. Since we will only consider measures on compact subsets of $\mathbb{C}$, which are second countable, regularity of all finite Borel measures is automatic. (This follows e.g. from [61, Theorem 2.18].)

For the definition of integration of real or complex valued functions w.r.t. a measure see any book on measure theory or the appendix of [41].

The counting measure on $(X, \mathcal{P}(X))$ is defined by $\mu_c(A) = \#(A)$. It is easy to show that $f : X \to \mathbb{C}$ (obviously measurable) is $\mu_c$-integrable if and only if $\sum_{x \in X} f(x)$ exists, in which case $\int f(x) \, d\mu_c(x) = \sum_{x \in X} f(x)$.

A.43 Definition A complex measure on a measurable space $(X, \mathcal{A})$ is a map $\mu : \mathcal{A} \to \mathbb{C}$ such that $\mu(\emptyset) = 0$ and $\mu(\bigcup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} \mu(A_n)$ whenever $\{A_n\} \subseteq \mathcal{A}$ is a countable family of mutually disjoint sets.

Note that complex measures are by definition bounded. Furthermore, if $\{A_n\}$ is a countable family of mutually disjoint sets then automatically $\sum_{n} |\mu(A_n)| < \infty$ since $\mu(\bigcup_{n} A_n)$ is invariant under permutations of the $A_n$.

\textsuperscript{62}Emile Borel (1871-1956). French mathematician. One of the pioneers of measure theory.
B Supplements for the curious. NOT part of the course

B.1 Functional analysis over fields other than \( \mathbb{R} \) and \( \mathbb{C} \)?

The most general meaningful definition of (linear) functional analysis is as the theory of topological vector spaces over a topological field \( \mathbb{F} \) and continuous linear maps between them. If (the topology of) \( \mathbb{F} \) is discrete, we are effectively doing topological abelian group theory, and this would not be considered functional analysis. Thus we are restrict ourselves to non-discrete topological fields. The general theory of topological fields is a thorny subject, almost unknown to non-specialists. (For reviews see [82, 80].) There would be no point in going into this here since in this course we considered general topological vector spaces only as a step towards spaces that are at least metrizable.

But there is a complete (in a sense) classification of the non-discrete locally compact fields. In characteristic zero, these are precisely \( \mathbb{R} \), the \( p \)-adic fields \( \mathbb{Q}_p \), where \( p \) runs through the prime numbers, and all their finite (thus algebraic) extensions. (And in characteristic \( p \neq 0 \) one has the finite extensions of \( \mathbb{F}_p((x)) \), the field of formal Laurent series over the finite field \( \mathbb{F}_p \).) For a proof see e.g. [57]. While \( \mathbb{R} \) has only one algebraic extension (namely \( \mathbb{C} \)), \( \mathbb{Q}_p \) has infinitely many finite extensions, so that the algebraic closure of \( \mathbb{Q}_p \) (which is not complete!) is infinite dimensional over \( \mathbb{Q}_p \). Like \( \mathbb{R} \) and \( \mathbb{C} \), the \( p \)-adic fields and their finite extensions all have a norm, usually called ‘valuation’ or ‘absolute value’, i.e. a map \( \mathbb{F} \to [0, \infty) \) satisfying \( |x| = 0 \iff x = 0 \), \( |x + y| \leq |x| + |y| \) and \( |xy| = |x||y| \). Note that the norm is strictly multiplicative, not just submultiplicative. The locally compact fields are complete w.r.t. their absolute value \(|·|\).

Books entitled ‘Functional analysis’ or ‘Topological vector spaces’ tend to work entirely over \( \mathbb{R} \) and \( \mathbb{C} \) unless the title contains ‘\( p \)-adic’, ‘non-archimedian’ or ‘ultrametric’ (but there are exceptions like [5, 51]). Nevertheless, functional analysis over \( p \)-adic fields is a well-studied subject, cf. e.g. [60, 56], but a somewhat exotic one since it only seems to have applications to number theory, algebraic geometry and related fields.

In the remainder of this short section we briefly comment on the extent to which the theory covered in these notes remains valid over \( p \)-adic fields. As a rule of thumb, one must be very careful with theorems on normed/Banach spaces that involve \( \mathbb{R} \) or \( \mathbb{C} \) either in their statement or in the proof. Since then either the orderedness of \( \mathbb{R} \) or the algebraic completeness of \( \mathbb{C} \) tend to be used, and the \( p \)-adic fields are neither algebraically closed nor orderable! The Hahn-Banach theorem is a case in point since we first proved it for \( \mathbb{R} \), making essential use of the orderedness of the base field \( \mathbb{F} = \mathbb{R} \), thus not just of the set \([0, \infty)\) in which the norms take values, and then extended it to \( \mathbb{C} \). (There nevertheless is a \( p \)-adic Hahn-Banach theorem, but with slightly different hypotheses and a different proof.)

Theorems not explicitly referring to \( \mathbb{R} \) or \( \mathbb{C} \) have a better chance of carrying over to \( p \)-adic functional analysis. For example, the open mapping theorem and both versions of the uniform boundedness theorem generalize without change. However, one has to be careful with the above rule since there are properties, like connectedness, shared by \( \mathbb{R} \) and \( \mathbb{C} \), but not enjoyed by the \( p \)-adic fields! There are other problems: There is no a priori relationship between the subsets \( S_1 = \{|c| \mid c \in \mathbb{F}\} \) and \( S_2 = \{|\|x\| \mid x \in V\} \) of \([0, \infty)\). Thus given \( x \in V \setminus \{0\} \) there may not be a \( c \in \mathbb{F} \) such that \( \|cx\| = 1 \).

We also have to be very careful with results on Hilbert spaces, since scalars in \( \mathbb{F} \) can be

---

\(^{63}\) I am sceptical about claims of relevance of \( p \)-adic/ultrametric (functional) analysis to fundamental theoretical/mathematical physics (but statistical/condensed matter physics is another discussion).
The discussion in this section strongly borrows from [16]. Indeed this leads to problems adapting the proof of Theorem 6.24. The same holds for the polarization identities.

We leave the discussion here and refer to the literature on p-adic (functional) analysis for more information. See e.g. [23, 59, 60, 56].

**B.2 The dual space of \( \ell^\infty(S, \mathbb{F}) \)**

We have seen in Theorem 5.16(v) that there are bounded linear functionals \( \varphi \in \ell^\infty(S, \mathbb{F})^* \) that vanish on \( c_0(S, \mathbb{F}) \). Those clearly cannot be captured by the function \( g(s) = \varphi(\delta_s) \) widely used in the proof of Theorem 5.16. This suggests to consider \( \mu_\varphi(A) = \varphi(\chi_A) \) for arbitrary \( A \subseteq S \) instead. If \( A_1, \ldots, A_K \) are mutually disjoint, and \( A = \bigcup_{k=1}^K A_k \) then \( \chi_A = \sum_{k=1}^K \chi_{A_k} \), thus \( \mu_\varphi(A) = \sum_{k=1}^K \mu_\varphi(A_k) \), so that \( \mu_\varphi \) is finitely additive.\(^{64}\)

**B.1 Definition** If \( S \) is a set, a finitely additive finite \( \mathbb{F} \)-valued measure on \( S \) is a map \( \mu : P(S) \to \mathbb{F} \) satisfying \( \mu(\emptyset) = 0 \) and \( \mu(A_1 \cup \cdots \cup A_K) = \mu(A_1) + \cdots + \mu(A_K) \) whenever \( A_1, \ldots, A_K \) are mutually disjoint subsets of \( S \). The set of such \( \mu \), which we denote \( f\alpha(S, \mathbb{F}) \), is a vector space via \( (c_1 \mu_1 + c_2 \mu_2)(A) = c_1 \mu_1(A) + c_2 \mu_2(A) \). For \( \mu \in f\alpha(S, \mathbb{F}) \) we define

\[
\| \mu \| = \sup \left\{ \sum_{k=1}^K |\mu(A_k)| \mid K \in \mathbb{N}, A_1, \ldots, A_K \subseteq S, i \neq j \Rightarrow A_i \cap A_j = \emptyset \right\},
\]

\[
\| \mu \|' = \sup_{A \subseteq S} |\mu(A)|.
\]

**B.2 Theorem** (i) \( \| \cdot \| \) and \( \| \cdot \|' \) are equivalent norms on \( f\alpha(S, \mathbb{F}) \). We write

\[ \text{ba}(S, \mathbb{F}) = \{ \mu \in f\alpha(S, \mathbb{F}) \mid \| \mu \|' < \infty \iff \| \mu \| < \infty \} \].

(ii) \( \text{ba}(S, \mathbb{F}), \| \cdot \| \) is a Banach space.

(iii) If \( \varphi \in \ell^\infty(S, \mathbb{F})^* \) then \( \| \mu_\varphi \| \leq \| \mu \| \), thus we have a norm-decreasing linear map \( \ell^\infty(S, \mathbb{F})^* \to \text{ba}(S, \mathbb{F}), \varphi \mapsto \mu_\varphi \).

**Proof.** (i) It is immediate from the definition \( \| c \mu \| = |c|\| \mu \| \) and \( \| c \mu \|' = |c|\| \mu \|' \) for all \( c \in \mathbb{F}, \mu \in f\alpha(S, \mathbb{F}) \) and that \( \| \mu \| = 0 \iff \mu = 0 \iff \| \mu \|' = 0 \). Also \( \| \mu_1 + \mu_2 \|' \leq \| \mu_1 \|' + \| \mu_2 \|' \) is quite obvious. Now

\[
\| \mu_1 + \mu_2 \| = \sup \left\{ \sum_{k=1}^K |\mu_1(A_k) + \mu_2(A_k)| \mid \cdots \right\} \leq \sup \left\{ \sum_{k=1}^K |\mu_1(A_k)| + |\mu_2(A_k)| \mid \cdots \right\}
\]

\[
\leq \sup \left\{ \sum_{k=1}^K |\mu_1(A_k)| \mid \cdots \right\} + \sup \left\{ \sum_{k=1}^K |\mu_2(A_k)| \mid \cdots \right\} = \| \mu_1 \| + \| \mu_2 \|.
\]

Thus \( \| \cdot \|, \| \cdot \|' \) are norms on \( f\alpha(S, \mathbb{F}) \). The definition of \( \| \cdot \| \) clearly implies \( |\mu(A)| \leq \| \mu \| \) for each \( A \subseteq S \), whence \( \| \mu \|' \leq \| \mu \| \).

Assume \( \mu \in f\alpha(S, \mathbb{R}) \) and \( \| \mu \|' < \infty \). If \( A_1, \ldots, A_K \subseteq S \) are mutually disjoint, put

\[
A_+ = \bigcup \{ A_k \mid \mu(A_k) \geq 0 \}, \quad A_- = \bigcup \{ A_k \mid \mu(A_k) < 0 \}.
\]

\[^{64}\text{The discussion in this section strongly borrows from [16].}\]
Now by finite additivity, $\sum_k |\mu(A_k)| = \mu(A_+) + \mu(A_-) \leq 2\|\mu\|$ since $|\mu(A_\pm)| \leq \|\mu\|$'. Taking the supremum over the families $\{A_k\}$ gives $\|\mu\| \leq 2\|\mu\|$.

If $\mu \in f(a(S, \mathbb{F})$, writing $\mu = \Re \mu + i \Im \mu$ we find $\|\mu\| \leq 4\|\mu\|$. Thus $\|\mu\| \leq \|\mu\| \leq 4\|\mu\|$ for all $\mu$, and the two norms are equivalent.

(ii) Here it is more convenient to work with the simpler norm $||\cdot||$. Now let $\{\mu_n\}$ be a Cauchy sequence in $ba(S, \mathbb{F})$. Then $|\mu_n(A) - \mu_m(A)| \leq ||\mu_n - \mu_m||$, so that $\{\mu_n(A)\}$ is Cauchy, thus convergent. Define $\mu(n) = \lim_n \mu_n(A)$. It is clear that $\mu(0) = 0$. If $A_1, \ldots, A_K$ are mutually disjoint then

$$
\mu(A_1 \cup \cdots \cup A_K) = \lim_{n \to \infty} \mu_n(A_1 \cup \cdots \cup A_K) = \lim_{n \to \infty} (\mu_n(A_1) + \cdots + \mu_n(A_K)) = \mu(A_1) + \cdots + \mu(A_K),
$$

so that $\mu$ is finitely additive. Since $\{\mu_n\}$ is Cauchy, for every $\varepsilon > 0$ there is $n_0$ such that $n, m \geq n_0$ implies $||\mu_n - \mu_m|| < \varepsilon$. In particular there is $n_0$ such that $||\mu_m|| \leq ||\mu_n|| + 1$ for $m \geq n_0$. This implies boundedness of $\mu$. And taking $m \to \infty$ in $|\mu_n(A) - \mu_m(A)| \leq ||\mu_n - \mu_m|| < \varepsilon$ gives $||\mu_n - \mu|| < \varepsilon$, so that $||\mu_n - \mu|| \to 0$. Thus $ba(S, \mathbb{F})$ is complete (w.r.t. $||\cdot||$, thus also w.r.t. $||\cdot||$).

(iii) It is clear that $\ell^\infty(S, \mathbb{F})^* \to f(a(S, \mathbb{F})$, $\varphi \mapsto \mu_\varphi$ is linear. Now let $A_1, \ldots, A_K \subseteq S$ be mutually disjoint. Then

$$
\sum_{k=1}^K |\mu_\varphi(A_k)| = \sum_{k=1}^K sgn(\mu_\varphi(A_k))\mu_\varphi(A_k) = \sum_{k=1}^K sgn(\mu_\varphi(A_k))\varphi(x_{A_k}) = \varphi \left( \sum_{k=1}^K sgn(\mu_\varphi(A_k))x_{A_k} \right).
$$

Since the $A_k$ are mutually disjoint and $|sgn(z)| \leq 1$, we have $||\sum_{k=1}^K sgn(\mu_\varphi(A_k))x_{A_k}||_\infty \leq 1$, so that $\sum_{k=1}^K |\mu_\varphi(A_k)| \leq ||\varphi||$. Taking the supremum over the finite families $\{A_k\}$ gives $||\mu_\varphi|| \leq ||\varphi||$.

B.3 Theorem  (i) For each $\mu \in ba(S, \mathbb{F})$ there is a unique linear functional $\int_\mu \in \ell^\infty(S, \mathbb{F})^*$ such that $\int_\mu(x_A) = \mu(A)$ for all $A \subseteq S$. It satisfies $||\int_\mu|| \leq ||\mu||$.

(ii) The maps $\alpha : \ell^\infty(S, \mathbb{F})^* \to ba(S, \mathbb{F})$, $\varphi \mapsto \mu_\varphi$ and $\int : ba(S, \mathbb{F}) \to \ell^\infty(S, \mathbb{F})^*$, $\mu \mapsto \int_\mu$ are mutually inverse and isometric, thus $\ell^\infty(S, \mathbb{F})^* \cong ba(S, \mathbb{F})$.

Proof. (i) If $f \in \ell^1(S, \mathbb{F})$ has finite image, write $f = \sum_{k=1}^K c_k x_{A_k}$, where the $A_k$ are mutually disjoint, and define

$$
\int f d\mu = \sum_{k=1}^K c_k \mu(A_k).
$$

(We write $\int f d\mu$ or $\int f d\mu$ according to convenience.) If $f = \sum_{i=1}^L c_i x_{A_i}$ is another representation of $f$, then using finite additivity of $\mu$ it is straightforward to check, using the finite additivity of $\mu$, that $\sum_{k=1}^K c_k \mu(A_k) = \sum_{i=1}^L c_i \mu(A_i)$, so that $\int f d\mu$ is well-defined. Now $\int c f d\mu = c \int f d\mu$ for $c \in \mathbb{F}$ is obvious, and $\int (f + g) d\mu = \int f d\mu + \int g d\mu$ for all finite-image functions follows from the fact that $f + g$ again is a finite-image function and the representation independence of $f$. Thus $\int f : f \mapsto \int f d\mu$ is a linear functional on the bounded finite image functions. It is clear that this is the unique linear functional sending $x_A$ to $\mu(A)$ for each $A \subseteq S$. Now

$$
||\int f d\mu|| \leq \sum_{k=1}^K |c_k| |\mu(A_k)| \leq ||f|| \sum_{k=1}^K |\mu(A_k)| \leq ||f|| ||\mu||.
$$

Thus $\int_\mu$ is a bounded functional, and since the bounded finite-image functions are dense in $\ell^\infty(S, \mathbb{F})$ by Lemma 5.13, $\int_\mu$ has a unique extension to a linear functional $\int_\mu \in \ell^\infty(S, \mathbb{F})^*$ with $||\int_\mu|| \leq ||\mu||$.  

131
(ii) If $\mu \in ba(S, \mathbb{F})$ then by definition of $\int_\mu$, we have $\int \chi_A \, d\mu = \mu(A)$ for all $A \subseteq S$. Thus $\alpha \circ \int = \text{id}_{ba(S, \mathbb{F})}$.

If $\varphi \in \ell^\infty(S, \mathbb{F})$ then in view of the definition of $\int$ we have $\int \chi_A \, d\mu_\varphi = \mu_\varphi(A) = \varphi(\chi_A)$ for all $A \subseteq S$. Thus $\varphi$ and $\int_\mu$ coincide on all characteristic functions, thus on all of $\ell^\infty(S, \mathbb{F})$ by linearity, density of the finite-image functions and the $\| \cdot \|_\infty$ continuity of $\varphi$ and $\int_\mu$. Thus $\int \circ \alpha = \text{id}_{\ell^\infty(S, \mathbb{F})^*}$.

Since the maps $\alpha$ and $\int$ are mutually inverse and both norm-decreasing, they actually both are isometries.

This completes the determination of $\ell^\infty(S, \mathbb{F})^*$. (Note that we did not use the completeness of $ba(S, \mathbb{F})$ proven in Theorem B.2(ii). Thus it would also follow from the isometric bijection $ba(S, \mathbb{F}) \cong \ell^\infty(S, \mathbb{F})^*$ just established.)

B.4 Exercise Given $\mu \in ba(S, \mathbb{F})$, prove that $\mu$ is $\{0, 1\}$-valued if and only if $\int_\mu \in \ell^\infty(S, \mathbb{F})^*$ is a character, i.e. $\int_\mu(fg) = \int_\mu(f) \int_\mu(g)$ for all $f, g \in \ell^\infty(S, \mathbb{F})$.

Since $\ell^\infty(S, \mathbb{F})^*$ has a closed subspace $\ell(\ell^1(S, \mathbb{F}))$, it is interesting to identify the corresponding subspace of $ba(S, \mathbb{F})$.

B.5 Definition A finitely additive measure $\mu \in ba(S, \mathbb{F})$ is called countably additive if for every countable family $A \subseteq P(S)$ of mutually disjoint sets we have

$$\mu \left( \bigcup A \right) = \sum_{A \in A} \mu(A)$$

and totally additive if the same holds for any family of mutually disjoint sets. The set of countably and totally additive measures on $S$ are denoted $ca(S, \mathbb{F})$ and $ta(S, \mathbb{F})$, respectively.

B.6 Proposition For $\mu \in ba(S, \mathbb{F})$, consider the following statements:

(i) There is $g \in \ell^1(S, \mathbb{F})$ such that $\mu(A) = \sum_{s \in A} g(s)$ for all $A \subseteq S$.

(ii) $\int_\mu \in \ell^\infty(S, \mathbb{F})^*$ is normal, thus $\int f \, d\mu = \lim_i \int f_i \, d\mu$ for every net $\{f_i\} \in \mathbb{F}^S$ that is pointwise convergent and uniformly bounded.

(iii) $\mu$ is totally additive.

(iv) $\mu$ is countably additive.

Then (i)$\iff$(ii)$\iff$(iii)$\implies$(iv). If $S$ is countable then also (iv)$\implies$(iii).

Proof. (i)$\implies$(ii) If $\mu$ is of the given form then clearly $\int_\mu \chi_A \, d\mu = \mu(A) = \sum_{s \in A} g(s)$ for each $A \subseteq S$. By the way $\int_\mu$ is constructed from $\mu$, it is clear that $\int f \, d\mu = \sum_{s \in S} f(s)g(s)$ for all $f \in \ell^\infty(S, \mathbb{F})$. Thus $\int_\mu = \varphi g$, and normality of $\int_\mu$ follows from Proposition 14.6.

(ii)$\implies$(iii) We know that we can recover $\mu$ from $\int_\mu$ as $\mu(A) = \int \chi_A \, d\mu$. Let $A$ be a family of mutually disjoint subsets of $S$. Then the net $\{f_F = \chi_{\bigcup F}\}$, indexed by the finite subsets $F \subseteq A$, is uniformly bounded and converges pointwise to $\chi_B$, where $B = \bigcup A$. Now normality of $\int_\mu$ implies that $\mu(B) = \int_\mu \chi_B \, d\mu = \lim_F \int f_F = \lim_F \sum_{A \in F} \mu(A) = \sum_{A \in A} \mu(A)$, which is additivity of $\mu$.

(iii)$\implies$(i) If we put $g(s) = \mu(\{s\})$ then additivity of $\mu$ means that $\mu(A) = \sum_{s \in A} g(s)$ for all $A \subseteq S$, convergence being absolute. Now the finiteness of $\mu(S)$ gives $\|g\|_1 < \infty$.  

132
(iii)⇒(iv) is trivial. If \( S \) is countable then a family of mutually disjoint non-empty subsets of \( S \) is at most countable, so that (iii) and (iv) are equivalent. \( \square \)

Thus we have the situation of the following diagram:

\[
\begin{array}{ccc}
\ell^1(S,\mathcal{F}) & \cong & \mathbb{R} \\
| & \searrow & | \\
(\ell^\infty(S,\mathcal{F})^*)_n & \cong & ta(S,\mathcal{F}) \\
| & | & | \\
\ell^\infty(S,\mathcal{F})^* & \cong & ba(S,\mathcal{F})
\end{array}
\]

where \( ta(S,\mathcal{F}) \) can be replaced by \( ca(S,\mathcal{F}) \) if \( S \) is countable.

**B.3 \( c_0(\mathbb{N},\mathcal{F}) \subseteq \ell^\infty(\mathbb{N},\mathcal{F}) \) is not complemented**

**B.7 Definition** We say that a Banach space \( V \) has property \( S \) if there is a countable subset \( C \subseteq V^* \) separating the points of \( V \). I.e., if \( x \in V \) and \( \varphi(x) = 0 \) for all \( \varphi \in C \) then \( x = 0 \).

If \( V \) has property \( S \) then every closed subspace \( W \subseteq V \) has property \( S \). (And so would non-closed subspaces, but they are not Banach.)

It is easy to see that \( V \) has property \( S \) whenever \( V^* \) is separable. But this is not a necessary condition: \( V = \ell^\infty(\mathbb{N},\mathbb{C}) \) has property \( S \), as we see by taking \( C = \{\varphi_n\}_{n \in \mathbb{N}} \), where \( \varphi_n(f) = f(n) \). But \( V^* (\cong ba(\mathbb{N},\mathbb{C})) \) is not separable, since by Exercise 8.18 this would imply separability of \( V \), which is false by Exercise 5.15(i).

**B.8 Theorem** \( c_0(\mathbb{N},\mathbb{R}) \subseteq \ell^\infty(\mathbb{N},\mathbb{R}) \) is not complemented.

**Proof.** From now on we abbreviate \( \ell^\infty(\mathbb{N},\mathcal{F}) \) and \( c_0(\mathbb{N},\mathcal{F}) \) as \( \ell^\infty, c_0 \). Our strategy for proving that \( c_0 \subseteq \ell^\infty \) is not complemented is the following: If \( c_0 \subseteq \ell^\infty \) had a complementary closed subspace \( W \), Exercise 10.8 would give \( \ell^\infty \cong c_0 \oplus W \), thus \( \ell^\infty/c_0 \cong W \). Since \( W \) would have property \( S \), it would follow that \( Q = \ell^\infty/c_0 \) has property \( S \), but we will prove that it doesn’t!

The idea for doing so is to produce an uncountable subset \( \mathcal{F} \subseteq Q \) such that each functional \( \varphi \in Q^* \) is non-zero only on countably many elements of \( \mathcal{F} \). Then for any countable \( C \subseteq Q^* \) the set \( \mathcal{F}' = \bigcup_{\varphi \in C} \{ q \in Q \mid \varphi(q) \neq 0 \} \) is countable, so that the family \( C \subseteq Q^* \) vanishes identically on the uncountable set \( \mathcal{F} \setminus \mathcal{F}' \). It therefore cannot separate the elements of \( \mathcal{F} \), let alone those of \( Q \). Thus \( Q \) does not have property \( S \) and we are done. For the construction of such an \( \mathcal{F} \) we use the following lemma:

**B.9 Lemma** Every countably infinite set \( X \) admits a family \( \{X_\lambda\}_{\lambda \in \Lambda} \) of subsets of \( X \) such that

(i) \( \Lambda \) is uncountable.

(ii) \( X_\lambda \) is infinite for each \( \lambda \in \Lambda \).

(iii) \( X_\lambda \cap X_{\lambda'} \) is finite for all \( \lambda, \lambda' \in \Lambda, \lambda \neq \lambda' \).
Proof. Take $Y = (0,1) \cap \mathbb{Q}$ and $\Lambda = (0,1) \setminus \mathbb{Q}$. Clearly $Y$ is countable and $\Lambda$ is uncountable. For each $\lambda \in \Lambda$ pick a sequence $\{a_n\} \subseteq Y$ converging to $\lambda$ (for example $a_n = \lfloor n\lambda \rfloor / n$) and put $Y_\lambda = \{a_n \mid n \in \mathbb{N}\}$. That each $Y_\lambda$ is infinite follows from the irrationality of $\lambda$ and the rationality of the $a_n$. If $\lambda \neq \lambda'$ and $a_n \to \lambda, a'_n \to \lambda'$ then there exists $n_0$ such that $n, n' \geq n_0 \Rightarrow \max(|a_n - \lambda|, |a'_n - \lambda'|) < |\lambda - \lambda'|/2$, so that $a_n \neq a'_n$. This implies $\#(Y_\lambda \cap Y_{\lambda'}) < \infty$. We thus have a family of subsets of $Y$ with all desired properties. For an arbitrary countably infinite set $X$ the claim now follows using a bijection $X \cong Y$. \hfill \blacksquare

Let $\{X_\lambda\}_{\lambda \in \Lambda}$ be a family of subsets of $\mathbb{N}$ as provided by the lemma. For $\lambda \in \Lambda$, the characteristic function $\chi_{X_\lambda} : \mathbb{N} \to \{0,1\} \subseteq \mathbb{C}$ clearly is in $\ell^\infty$. Let $p : \ell^\infty \to Q = \ell^\infty/c_0$ be the quotient map. Now let $q_\lambda = p(X_{\lambda})$ and $\mathcal{F} = \{q_\lambda \mid \lambda \in \Lambda\}$. If $\lambda, \lambda' \in \Lambda, \lambda \neq \lambda'$ the symmetric difference $X_\lambda \Delta X_{\lambda'} = (X_\lambda \cup X_{\lambda'}) \setminus (X_\lambda \cap X_{\lambda'})$ is infinite by (ii) and (iii). Thus $\chi_{X_\lambda} \neq \chi_{X_{\lambda'}} \neq c_0 = \ker p$, so that $\lambda \mapsto q_\lambda$ is injective, thus with (i) we see that $\mathcal{F}$ is uncountable.

Let now $\varphi \in Q^*, m, n \in \mathbb{N}$ and let $\lambda_1, \ldots, \lambda_m \in \Lambda$ be mutually distinct and such that $|\varphi(q_{\lambda_i})| \geq 1/n$ for $i = 1, \ldots, m$. For each $i$ pick $t_i$ with $|t_i| = 1$ such that $t_i \varphi(q_{\lambda_i}) = |\varphi(q_{\lambda_i})|$. Put $f = \sum_{i=1}^m t_i \chi_{X_{\lambda_i}} \in \ell^\infty$. Since the sets $X_{\lambda_i}$ have pairwise finite intersections, the function $f$ has absolute value larger than one only on a subset of the finite set $\bigcup_{j,k=1}^m X_{\lambda_j} \cap X_{\lambda_k}$ and absolute value one on the infinite set $(\bigcup_i X_{\lambda_i}) \setminus (\bigcup_{j,k} X_{\lambda_j} \cap X_{\lambda_k})$. This implies that $\|p(f)\| = \inf_{g \in c_0} \|f - g\| = 1$. Thus

$$\|\varphi\| \geq |\varphi(p(f))| = \sum_{i=1}^m t_i |\varphi(p(\chi_{X_{\lambda_i}}))| = \sum_{i=1}^m t_i |\varphi(q_{\lambda_i})| = \sum_{i=1}^m |\varphi(q_{\lambda_i})| \geq m/n.$$ 

Thus $m \leq n\|\varphi\| < \infty$, so that for each $\varphi \in Q^*$ and $n \in \mathbb{N}$ there cannot be more than $m$ distinct $\lambda \in \Lambda$ with $|\varphi(q_{\lambda_i})| \geq 1/n$. If there was an uncountable $\mathcal{F}' \subseteq \mathcal{F}$ with $\varphi(q) \neq 0 \forall q \in \mathcal{F}'$, there would have to be an $n \in \mathbb{N}$ such that $|\varphi(q)| \geq 1/n$ for infinitely (in fact uncountably) many $q \in \mathcal{F}'$, contradicting what we just proved. This completes the proof. \hfill \blacksquare

### B.4 Banach spaces with no or multiple pre-duals

Recall that we write $\cong$ for isometric isomorphism and $\simeq$ for isomorphism of Banach spaces.

**B.10 Lemma** Let $V$ be a Banach space. Then:

(i) $P = \iota_{V^*} \circ (\iota_V)^* \in B(V^{***})$ is an idempotent and $PV^{***} = \iota_{V^*}(V^*)$.

(ii) $\iota_{V^*}(V^*) \subseteq V^{***}$ is a complemented subspace.

(iii) If a Banach space $W$ is isomorphic to $V^*$ with $V$ Banach then $\iota_W(W) \subseteq W^{**}$ is complemented.

(iv) $V^{***}/V^* \simeq (V^{**}/V^*)$. (We omitted the $\iota$’s for simplicity.)

**Proof.** (i) Since $\iota_V, \iota_{V^*}$ are bounded, with Lemma 12.1 we have boundedness of $P$. Let $\varphi \in V^*$ and $x \in X$. Then

$$(P\iota_{V^*}(\varphi))(\iota_V(x)) = \iota_{V^*}((\iota_V)^*(\iota_{V^*}(\varphi)))(\iota_V(x)) = [(\iota_V)^*(\iota_{V^*}(\varphi))](x) = \varphi(x) = \iota_{V^*}(\varphi)(\iota_V(x)),$$

where we used Exercise 8.10 several times, proves $P\iota_{V^*}(\varphi) = \iota_{V^*}(\varphi)$. Thus $P \uparrow \iota_{V^*}(V^*) = \text{id}$. On the other hand, it follows directly from the definition of $P$ that $PV^{***} \subseteq \iota_{V^*}(V^*)$. Combining these two facts gives $P^2 = P$ and $PV^{***} = \iota_{V^*}(V^*)$. 

134
(ii) This is an immediate consequence of (ii) and Exercise 7.12.

(iii) If \( T : W \rightarrow V^* \) is an isomorphism then we have isomorphisms \( T^* : V^{**} \rightarrow W^* \) and \( T^{**} : W^{**} \rightarrow V^{***} \). Using this it is straightforward to deduce the claim from (ii).

(iv) By Exercise 7.6 we have \( (V^{**}/V)^* \cong V^\perp \subseteq V^{***} \). And by (ii), \( V^{***} \cong V^* \oplus W \), where \( W \cong V^{**}/V^* \), the isomorphism being given by \( x^{**} \mapsto (Px^{***}, (1-P)x^{**}) \) with \( P \) as in (i). Thus \( PV^{***} \cong V^* \) and \( V^{***}/V^* \cong (1-P)V^{***} \). Thus the claimed isomorphism follows if we prove that the subspaces \( V^\perp \) and \( (1-P)V^{***} \) of \( V^{***} \) are equal.

Now, \( x^{**} \in (1-P)V^{***} \) means \( (1-P)x^{***} = x^{**} \), thus \( Px^{***} = 0 \). Since \( P = \iota_V^* \circ (\iota_V)^* \), where \( \iota_V^* \) is injective, this is equivalent to \( (\iota_V)^*(x^{**}) = 0 \). By the definition of the transpose, this means that \( x^{**} \circ \iota_V = 0 \). Since this is the same as \( x^{**} \in \iota_V(V)^\perp \), we are done. \( \blacksquare \)

B.11 Corollary \( c_0(\mathbb{N}, \mathbb{F}) \) is not isomorphic to the dual space of any Banach space.

*Proof.* We again abbreviate \( c_0(\mathbb{N}, \mathbb{F}) \) as \( c_0 \) etc. We know that \( c_0^* \cong \ell^1 \) and \( c_0^{**} \cong \ell^\infty \), the canonical map \( \iota_{c_0} : c_0 \rightarrow c_0^{**} \) just being the inclusion map \( c_0 \hookrightarrow \ell^\infty \). By Theorem B.8, \( c_0 \subseteq \ell^\infty \) is not complemented. Combining this with Lemma B.10(iii), the claim follows. \( \blacksquare \)

B.12 Corollary Let \( X = c_0 \oplus (\ell^\infty/c_0) \). Then \( X \not\cong \ell^\infty \), but \( X^* \cong (\ell^\infty)^* \).

*Proof.* \( X \cong \ell^\infty \) would imply that \( c_0 \subseteq \ell^\infty \) is complemented, which it is not by Theorem B.8. Thus \( X \not\cong \ell^\infty \). With \( c_0^* \cong \ell^1 \) we have \( X^* \cong c_0^* \oplus (\ell^\infty/c_0)^* \cong \ell^1 \oplus (\ell^\infty/c_0)^* \).

On the other hand, since \( \ell^1 \cong c_0^* \) is a dual space, we see that \( \ell^1 \cong (\ell^1)^* \cong (\ell^\infty)^* \) is complemented by Lemma B.10(iii). Thus \( (\ell^\infty)^* \cong \ell^1 \oplus (\ell^\infty)^*/\ell^1 \) by Exercise 10.8(i). Now Lemma B.10(iv) with \( V = c_0 \) gives \( (\ell^\infty)^*/\ell^1 \cong (\ell^\infty/c_0)^* \), so that \( X^* \cong (\ell^\infty)^* \). \( \blacksquare \)

One can also find Banach spaces \( V \) with \( V^* \cong \ell^1 \), while \( V \not\cong c_0 \). But this is a bit more involved.

### B.5 Normability. Separation theorems. Goldstine’s theorem

In this section we will prove the normability criterion for topological vector spaces stated in Remark 3.9 and Goldstine’s Theorem 17.19, which we used in Exercise 17.18 and will use again in Section B.6. The proofs require some preparations which, however, are quite fundamental, also for the study of locally convex spaces.

#### B.5.1 Minkowski functionals. Criteria for normability and local convexity

B.13 Proposition Let \( V \) be a topological vector space and \( U \) a convex open neighborhood of 0. Define the ‘Minkowski functional’ \( \mu_U : V \rightarrow [0, \infty) \) of \( U \) by

\[
\mu_U(x) = \inf\{t \geq 0 \mid x \in tU\}.
\]

Then \( \mu_U \) is sublinear and continuous, and \( U = \{x \in X \mid \mu_U(x) < 1\} \).

*Proof.* As \( t \rightarrow \infty \) we have \( t^{-1}x \rightarrow 0 \). Since \( U \) is an open neighborhood of 0, we have \( t^{-1}x \in U \) for \( t \) large enough. Thus \( \mu_U(x) < \infty \) for each \( x \in V \). It is quite obvious from the definition that \( \mu_U(cx) = c\mu_U(x) \) for \( c > 0 \). Thus \( \mu_U \) is positive-homogeneous. We have \( \mu_U(x) < 1 \) if and only if there exists \( t \in (0, 1) \) such that \( x \in tU \). Thus \( \mu_U(x) < 1 \Rightarrow x \in U \). And if \( x \in U \) then openness of \( U \) implies that \( (1-\varepsilon)x \in U \) for some \( \varepsilon > 0 \). Thus \( \mu_U(x) < 1 \), so that we have \( U = \{x \in X \mid \mu_U(x) < 1\} \).
Let \( x, y \in V \), and let \( s, t > 0 \) such that \( x \in sU, y \in tU \). i.e. there are \( a, b \in U \) such that \( x = sa, y = tb \). Thus \( x + y = sa + tb = (s + t)\frac{sa + tb}{s + t} \). Since \( \frac{s}{s+t}a + \frac{t}{s+t}b \in U \) due to convexity of \( U \), we have \( x + y \in (s + t)U \). Thus \( \mu_U(x + y) \leq s + t \), and since we have \( x \in sU, y \in tU \) for all \( s < \mu_U(x) + \epsilon, t < \mu_U(y) + \epsilon \) with \( \epsilon > 0 \), the conclusion is \( \mu_U(x + y) \leq \mu_U(x) + \mu_U(y) \), thus subadditivity. Being subadditive and positive homogeneous, \( \mu_U \) is sublinear.

Let \( \{x_i\}_{i \in I} \subseteq V \) be a net converging to zero. For each \( n \in \mathbb{N} \), \( n^{-1}U \) is an open neighborhood of zero. Thus there exists a \( t_n \in I \) such that \( t_n \in n^{-1}U \) and therefore, with the definition of \( \mu_U \), that \( \mu_U(x_i) \leq n^{-1} \). Thus \( \mu_U(x_i) \to 0 \), which is continuity of \( \mu_U \) at \( 0 \in V \).

If now \( x_i \to x \) then the subadditivity of \( \mu_U \) gives

\[
\mu_U(x) - \mu_U(x - x_i) \leq \mu_U(x_i) \leq \mu_U(x) + \mu_U(x_i - x),
\]

and since \( \mu_U(x_i - x) \to 0 \), we have \( \mu_U(x_i) \to \mu_U(x) \), thus continuity of \( \mu_U \) at all \( x \in V \). \( \blacksquare \)

### B.14 Definition
Let \( V \) be a topological vector space and \( 0 \in U \subseteq V \). Then \( U \) is called

- balanced if \( x \in U, |x| \leq 1 \Rightarrow \lambda x \in U \),
- bounded if for every open \( W \ni 0 \) there exists \( \lambda > 0 \) such that \( \lambda U \subseteq W \).

Note that if \( U \) is convex and contains zero, multiplication by \( t \in [0, 1] \) sends \( U \) into itself. Thus for checking balancedness it suffices to consider \( |\lambda| = 1 \).

### B.15 Proposition
Let \( (V, \tau) \) be a topological vector space and \( U \) a convex open neighborhood of zero. Then

(i) The Minkowski functional \( \mu_U \) is a seminorm if and only if \( U \) is balanced.

(ii) If \( U \) is bounded then \( \mu_U(x) = 0 \) implies \( x = 0 \).

(iii) If \( U \) is balanced and bounded then \( \|x\| = \mu_U(x) \) is a norm inducing the topology \( \tau \).

**Proof.** (i) Since \( \mu_U \) is subadditive and positive-homogeneous, it is a seminorm if and only if \( \mu_U(\lambda x) = \mu_U(x) \) for all \( x \in V \) and \( \lambda \in \mathbb{F} \) with \( |\lambda| = 1 \). If \( U \) is balanced then this is evidently satisfied. Now assume \( \mu_U(\lambda x) = \mu_U(x) \). The openness of \( U \) implies that \( \{ t > 0 \mid x \in tU \} = (\mu_U(x), \infty) \). Thus if \( |\lambda| = 1 \) then the assumption \( \mu_U(\lambda x) = \mu_U(x) \) implies that \( x \in U \) if and only if \( x \in \lambda U \). Thus \( U \) is balanced.

(ii) Assume that \( U \) is bounded and that \( x \neq 0 \). Since \( \tau \) is \( T_1 \), there is an open \( W \ni \lambda x \) such that \( 0 \in W \not\ni x \). Since \( U \) is bounded, there is \( \lambda > 0 \) such that \( \lambda U \subseteq W \), which clearly implies \( x \notin \lambda U \). Now the definition of \( \mu_U \) implies \( \mu_U(x) > \lambda > 0 \).

(iii) Proposition B.13 and the above (i) and (ii) show that \( \| \cdot \| = \mu_U \) is a continuous norm on \( V \). Thus \( x_n \to 0 \) implies \( \| x_n \| \to 0 \). If we prove the converse implication then \( \tau = \tau_{\| \cdot \|} \) follows since \( V \) is a topological vector space. Let \( \{x_n\} \) be a sequence such that \( \| x_n \| \to 0 \), and let \( W \) be an open neighborhood of 0. Since \( U \) is bounded, there is \( \lambda > 0 \) such that \( \lambda U \subseteq W \). Now, \( \| x_n \| \to 0 \) means that there is \( n_0 \in \mathbb{N} \) such that \( n \geq n_0 \Rightarrow \| x_n \| < \lambda / 2 \). With the definition of \( \mu_U \) this implies \( x_n \in \lambda U \), thus \( x_n \in \lambda U \subseteq W \) for all \( n \geq n_0 \). This proves \( x_n \to 0 \). \( \blacksquare \)

### B.16 Exercise
Let \( V \) be a topological vector space and \( A \subseteq V \) convex. Prove that the interior \( A^0 \) and the closure \( \overline{A} \) are convex.

We now know that a topological vector space is normable if the zero element has a balanced convex bounded open neighborhood. (The converse is easy.) But this can be improved:

### B.17 Lemma
Let \( V \) be a topological vector space and \( U \) a convex open neighborhood of 0. Then there exists a balanced convex open neighborhood \( U' \subseteq U \) of 0.
Proof. Since multiplication by scalars is continuous, there exists an \( \varepsilon > 0 \) such that \( \lambda U \subseteq U \) whenever \( |\lambda| \leq \varepsilon \). Thus with \( W = |\varepsilon|U \) we have \( tW \subseteq U \) whenever \( |t| \leq 1 \). Put \( Y = \bigcup_{|t| \leq 1} tW \subseteq U \). By construction, \( Y \) is a balanced open neighborhood of \( 0 \).

For every \( \lambda \in \mathbb{F} \) with \( |\lambda| = 1 \) it is clear that \( \lambda U \) is a convex open neighborhood of \( 0 \). Putting \( Z = \bigcap_{|\lambda| = 1} \lambda U \), it is manifestly clear that \( Z \) is balanced and \( 0 \in Z \). Furthermore, \( U' \) is convex (as an intersection of convex sets). Since \( tW \subseteq U \) for all \( |t| = 1 \), we have \( Y \subseteq Z \), so that \( Z \) has non-empty interior \( Z^0 \). Now we put \( U' = Z^0 \) and claim that \( U' \) has the desired properties. Clearly \( U' \) is an open neighborhood of \( 0 \), as the interior of a convex set it is convex (Exercise B.16). If \( |t| = 1 \) then the map \( Z \to Z, x \mapsto tx \) is a homeomorphism. Thus if \( x \in Z^0 = U' \) then \( tx \in Z^0 = U' \), showing that \( U' = Z^0 \) is balanced. ■

Now we are in a position to prove geometric criteria for normability and local convexity of topological vector spaces:

**B.18 Theorem** Let \( V \) be a topological vector space. Then \( V \) is normable if and only if there exists a bounded convex open neighborhood of \( 0 \).

**Proof.** If \( V \) is normable by the norm \( \| \cdot \| \) then \( B_{\| \cdot \|}(0,1) = \{ x \in V \mid \|x\| < 1 \} \) is clearly open, convex (and balanced). To show boundedness, let \( W \ni 0 \) be open. Then there is \( \varepsilon > 0 \) such that \( B(0,\varepsilon) \subseteq W \). Now clearly \( \varepsilon B(0,1) = B(0,\varepsilon) \subseteq W \), thus \( B(0,1) \) is bounded.

If there exists a bounded convex open neighborhood \( U \) of \( 0 \) then by Lemma B.17 we can assume \( U \) in addition to be balanced. (The \( U' \) provided by the lemma is a subset of \( U \), thus bounded if \( U \) is bounded.) Now by Proposition B.15(iii), \( \mu_U \) is a norm inducing the given topology on \( V \). ■

**B.19 Theorem** A topological vector space \( (V,\tau) \) is locally convex in the sense of Definition 3.18 (i.e. the topology \( \tau \) comes from a separating family \( \mathcal{F} \) of seminorms) if and only if it is Hausdorff and the zero element has an open neighborhood base consisting of convex sets.

**Proof.** Given a separating family \( \mathcal{F} \) of seminorms and putting \( \tau = \tau_{\mathcal{F}} \), a basis of open neighborhoods of \( 0 \) is given by the finite intersections of sets \( U_{p,\varepsilon} = \{ x \in V \mid p(x) < \varepsilon \} \), where \( p \in \mathcal{F}, \varepsilon > 0 \). Each of the \( U_{p,\varepsilon} \) is convex and open, thus also the finite intersections.

And if \( \tau \) has the stated property, Lemma B.17 gives that \( 0 \) has a neighborhood base consisting of balanced convex open sets. Defining \( \mathcal{F} = \{ \mu_U \mid U \text{ balanced convex open neighborhood of } 0 \} \), each of the \( \mu_U \) is a continuous seminorm by Propositions B.13 and B.15. Thus if \( x_i \to 0 \) then \( \|x_i\|_U := \mu_U(x_i) \to 0 \). And \( \|x_i\|_U \to 0 \) for all balanced convex open \( U \) implies that \( x_i \) ultimately is in every open neighborhood of \( 0 \), thus \( x_i \to 0 \). Thus \( \tau = \tau_{\mathcal{F}} \), and 3.17 gives that \( \mathcal{F} \) is separating. ■

**B.20 Exercise** Let \( 0 < p < 1 \).

(i) Prove that \( (\ell^p(S,\mathbb{F}),\tau_{d_p}) \) is normable if \( S \) is finite.

(ii) Prove that the open unit ball of \( (\ell^p(S,\mathbb{F}),\tau_{d_p}) \) does not contain any convex open neighborhood of \( 0 \) if \( S \) is infinite.

(iii) Prove that \( (\ell^p(S,\mathbb{F}),\tau_{d_p}) \) is neither normable nor locally convex if \( S \) is infinite.

**B.5.2 Hahn-Banach separation theorem. Goldstine’s theorem**

The Hahn-Banach theorem in the sublinear functional version (Theorem 8.2) has an important geometric application, namely the fact that disjoint convex sets can be separated by hyperplanes,
i.e. sets \( \{ x \in V \mid \text{Re}\varphi(x) = t \} \) for some \( \varphi \in V^* \) and \( t \in \mathbb{R} \). The following is just one of many versions, sufficient for our purposes.

B.21 **Theorem** Let \( V \) be a topological vector space and \( A, B \subseteq V \) disjoint non-empty convex subsets, \( A \) being open. Then there is a continuous linear functional \( \varphi : V \to \mathbb{F} \) such that \( \text{Re}\varphi(a) < \inf_{b \in B} \text{Re}\varphi(b) \) \( \forall a \in A \). (If \( \mathbb{F} = \mathbb{R} \), drop the ‘Re’.)

**Proof.** Pick \( a_0 \in A, b_0 \in B \) and put \( z = b_0 - a_0 \) and \( U = (A - a_0) - (B - b_0) = A - B + z, \) which is a convex (as pointwise sum of two convex sets) open (since \( U = \bigcup_{x \in B - a_0 + b_0}(A + x) \)) neighborhood of 0. Let \( p = \mu_U \) be the associated Minkowski functional. As a consequence of \( A \cap B = \emptyset \) we have \( 0 \notin A - B \), thus \( z \notin U \), and therefore \( p(z) \geq 1 \).

Put \( W = \mathbb{R}z \) and define \( \psi : W \to \mathbb{R}, cz \mapsto c. \) For \( c \geq 0 \) we have \( \psi(cz) = c \leq cp(z) = p(cz). \) Thus by sublinearity of \( p \) and Theorem 8.2 there exists a linear functional \( \varphi : V \to \mathbb{R} \) satisfying \( \varphi \mid W = \psi, \) thus \( \varphi(cz) = c, \) and \( \varphi(x) \leq p(x) \forall x \in V. \) Thus also \( -p(-x) \leq -\varphi(-x) = \varphi(x) \), and since \( x \to 0 \) implies \( p(x) \to 0, \) \( \varphi \) is continuous at zero, thus everywhere.

If now \( a \in A, b \in B \) then \( a - b + z \in U, \) so that \( p(a - b + z) < 1. \) Thus

\[
\varphi(a - b) + 1 = \varphi(a - b + z) \leq p(a - b + z) < 1,
\]

thus \( \varphi(a) < \varphi(b) \) for all \( a \in A \) and \( b \in B. \) Thus the subsets \( \varphi(A), \varphi(B) \) of \( \mathbb{R} \) are disjoint. Since \( A, B \) are convex, they are connected. Consequently, \( \varphi(A), \varphi(B) \) are connected, thus intervals. Since \( A \) is open, so is \( \varphi(A) \) (open mapping theorem). If we put \( s = \sup \varphi(a), \) we have \( \varphi(a) < s \leq \varphi(b) \) for all \( a \in A, b \in B, \) and this is equivalent to \( \varphi(a) \leq \inf_{b \in B} \varphi(b) \) for all \( a \in A. \)

\( \mathbb{F} = \mathbb{C}: \) Considering \( V \) as \( \mathbb{R} \)-vector space, apply the above to obtain a continuous \( \mathbb{R} \)-linear functional \( \varphi : V \to \mathbb{R} \) such that \( \varphi(a) < \inf_{b \in B} \varphi(b) \forall a \in A. \) Now define \( \varphi : V \to \mathbb{C}, x \mapsto \varphi_0(x) - i\varphi_0(ix). \) This clearly is continuous and satisfies \( \text{Re}\varphi = \varphi_0, \) so that the desired inequality holds. That \( \varphi \) is \( \mathbb{C} \)-linear follows from the same argument as in the proof of Theorem 8.5.

B.22 **Theorem** (Goldstine’s theorem) If \( V \) is a Banach space then \( V_{\leq 1} \) is \( \sigma(V^{**}, V^*) \)-dense in \( (V^{**})_{\leq 1}. \)

**Proof.** We abbreviate \( \tau = \sigma(V^{**}, V^*). \) The unit ball \( (V^{**})_{\leq 1} \) is \( \tau \)-compact by Alaoglu’s theorem, thus \( \tau \)-closed, so that \( B = V_{\leq 1}^{**} \tau, \) which is convex by Exercise B.16, is contained in \( (V^{**})_{\leq 1}. \) If this inclusion is strict, pick \( x^{**} \in (V^{**})_{\leq 1} \setminus V_{\leq 1}^{**} \tau. \) Then \( x^{**} \) has a \( \tau \)-open neighborhood \( U \) disjoint from \( B, \) and by Theorem B.19 there is a convex open \( A \subseteq U. \) Now Theorem B.21 applied to \( (V^{**}, \tau) \) and \( A, B \subseteq V^{**} \) gives a \( \tau = \sigma(V^{**}, V^*) \)-continuous linear functional \( \varphi \in (V^{**})^* \) such that \( \text{Re}\varphi(a) < \inf_{b \in B} \text{Re}\varphi(b) \forall a \in A. \) Now Exercise 17.13 gives \( \varphi \in V^* \subseteq V^{***}. \)

Putting \( \psi = -\varphi \) we have \( \sup_{b \in B} \text{Re}\psi(b) < \text{Re}\psi(a) \forall a \in A, \) which is more convenient. Since \( \psi \in V^* \) and \( B \supseteq V_{\leq 1}, \) we have \( \|\psi\| \leq \sup_{b \in B} \text{Re}\psi(b). \) On the other hand, with \( x^{**} \in A \) and \( \|x^{**}\| \leq 1, \) we have \( \text{Re}\psi(x^{**}) \leq |\psi(x^{**})| \leq \|x^{**}\||\psi|| \leq \|\psi\|. \) Combining these findings, we have \( \|\psi\| \leq \sup_{b \in B} \text{Re}\psi(b) < \text{Re}\psi(x^{**}) \leq \|\psi\|, \) which is absurd. This contradiction proves \( V_{\leq 1}^{**} \tau = (V^{**})_{\leq 1}. \)

B.6 **Strictly convex and uniformly convex Banach spaces**

B.6.1 **Strict convexity and uniqueness in the Hahn-Banach theorem**

B.23 **Definition** A Banach space \( V \) is called strictly convex if \( x, y \in V, \|x\| = \|y\| = 1, x \neq y \) implies \( \|x + y\| < 2. \)

B.24 **Exercise** (i) Prove that \( \ell^p(S, \mathbb{F}) \) is not strictly convex if \( \#S \geq 2 \) and \( p \in \{1, \infty\}. \)
(ii) Prove that $\ell^p(S, F)$ is strictly convex for every $S$ and $1 < p < \infty$.

(iii) Prove that all Hilbert spaces are strictly convex.

B.25 Proposition  Let $V$ be a Banach space. Then  

(i) The following are equivalent:  

(a) $V$ is strictly convex.  

(b) If $x, y \in V$ satisfy $\|x + y\| = \|x\| + \|y\|$ then $y = 0$ or $x = cy$ with $c \geq 0$.  

(ii) [Taylor-Foguel] The following are equivalent:  

(a) $V^*$ is strictly convex.  

(b) For every closed subspace $W \subseteq V$ and $\varphi \in W^*$ there is a unique $\hat{\varphi} \in V^*$ with $\hat{\varphi}|_W = \varphi$ and $\|\hat{\varphi}\| = \|\varphi\|$.  

B.26 Exercise  (i) Prove (b)$\Rightarrow$(a) in Proposition B.25(i).

(ii) Prove (a)$\Rightarrow$(b) in Proposition B.25(ii).

(These are the easier directions.)  

Proof of the remaining implications in Proposition B.25.  

(i) (a)$\Rightarrow$(b) Assume that $V$ is strictly convex and $x, y \in V$ satisfy $\|x + y\| = \|x\| + \|y\|$. If $x = 0$ or $y = 0$ then we are done. By rescaling and/or exchanging if necessary we may assume $1 = \|x\| \leq \|y\|$. Put $z = y/\|y\|$.

$$2 \geq \|x + z\| = \|x + y - (1 - \|y\|^{-1})y\| \geq \|x + y\| - (1 - \|y\|^{-1})\|y\| = \|x + y\| - \|y\| + 1$$

where we used $\|a - b\| \geq \|a\| - \|b\|$ and the assumptions $\|x + y\| = \|x\| + \|y\|$ and $\|x\| = 1$. This implies $\|x + z\| = 2$. Since $\|x\| = 1 = \|z\|$ (by assumption and by construction of $z$), the strict convexity implies $x = z = y/\|y\|$. Thus $y = \|y\| x$, and we have proven (b).

(ii) (b)$\Rightarrow$(a) Assume $V^*$ is not strictly convex. Then there are $\varphi_1, \varphi_2 \in V^*$ with $\varphi_1 \neq \varphi_2$ and $\|\varphi_1\| = \|\varphi_2\| = 1$ and $\varphi_1 + \varphi_2 = 2$. Then $W = \{x \in V \mid \varphi_1(x) = \varphi_2(x)\} \subseteq V$ is a closed linear subspace and proper (since $\varphi_1 \neq \varphi_2$). Put $\psi = \varphi_1|_W = \varphi_2|_W \in W^*$. We will prove $\|\psi\| = 1$. Then $\varphi_1, \varphi_2$ are distinct norm-preserving extensions of $\psi \in W^*$ to $V$, providing a counterexample for uniqueness of norm-preserving extensions.

Since $\varphi_1 - \varphi_2 \neq 0$, there exists $z \in V$ with $\varphi_1(z) - \varphi_2(z) = 1$. Now every $x \in V$ can be written uniquely as $x = y + cz$, where $y \in W$, $c \in \mathbb{C}$: Put $c = \varphi_1(x) - \varphi_2(x)$ and then $y = x - cz$. Now it is obvious that $y \in W$. Uniqueness of such a representation follows from $z \not\in W$.

Since $\|\varphi_1 + \varphi_2\| = 2$, we can find a sequence $\{x_n\} \subseteq V$ with $\|x_n\| = 1 \forall n$ such that $\varphi_1(x_n) + \varphi_2(x_n) \to 2$. Since $|\varphi_i(x_n)| \leq 1$ for $i = 1, 2$ and all $n$, it follows that $\varphi_i(x_n) \to 1$ for $i = 1, 2$. Now write $x_n = y_n + c_n z$, where $\{y_n\} \subseteq W$ and $\{c_n\} \subseteq \mathbb{C}$. Then $c_n = \varphi_1(x_n) - \varphi_2(x_n) \to 0$. Thus $\|x_n - y_n\| = |c_n| \|z\| \to 0$, so that $\|y_n\| \to 1$. And with $c_n \to 0$ we have

$$\lim_{n \to \infty} \varphi_1(y_n) = \lim_{n \to \infty} \varphi_1(y_n + c_n z) = \lim_{n \to \infty} \varphi_1(x_n) = 1.$$  

In view of $\{y_n\} \subseteq W$ and $\varphi_1|_W = \psi$, we have $\psi(y_n) = \varphi_1(y_n) \to 1$. Together with $\|y_n\| \to 1$ this implies $\|\psi\| \geq 1$. Since the converse inequality is obvious, we have $\|\psi\| = 1$, as claimed. 

In addition to the above we remark that $V$ is strictly convex if and only if $V^*$ is ‘smooth’, and conversely. (Cf. e.g. [42] for definition and proof.) Thus uniqueness of Hahn-Banach extensions for subspaces $W \subseteq V$ holds if and only if $V$ is smooth.
B.6.2 Uniform convexity and reflexivity. Duality of $L^p$-spaces reconsidered

B.27 Definition A Banach space $V$ is called uniformly convex if for every $\varepsilon > 0$ there exists $\delta(\varepsilon) > 0$ such that $x, y \in V$, $\|x\| = \|y\| = 1$ and $\|x - y\| \geq \varepsilon$ imply $\|x + y\| \leq 2 - \delta(\varepsilon)$.

It is obvious that uniform convexity implies strict convexity.

B.28 Theorem (Milman-Pettis 1938/9) Every uniformly convex Banach space is reflexive.

Proof. (Following Ringrose 1958) Assume $V$ is uniformly convex, but not reflexive. Let $S \subseteq V$ and $S^{**} \subseteq V^{**}$ be the unit spheres (sets of elements of norm one). Since $S = S^{**}$ easily implies $V = V^{**}$, we have $S \subseteq S^{**}$. If $x^* \in S^{**}\setminus S$ then by the obvious norm-closedness of $S \subseteq S^{**}$ there is $\varepsilon > 0$ such that $B(x^*, \varepsilon) \cap S = \emptyset$. Since $\|x^*\| = 1$, we can find $\varphi \in V^*$ with $\|\varphi\| = 1$ and $|\varphi^\prime(\varphi) - 1| > 1 - \delta(\varepsilon)/2$. Now $U = \{y^* \in V^* \mid |y^\prime(\varphi) - 1| > 1 - \delta(\varepsilon)/2\} \subseteq V^{**}$ is a $\tau := \sigma(V^{**}, V^*)$-open neighborhood of $x^*$. By Goldstine’s Theorem B.22, $V_{\leq 1} \subseteq (V^{**})_{\leq 1}$ is $\tau$-dense. If $\{x_\alpha\} \subseteq V_{\leq 1}$ is a net $\tau$-converging to $x \in S^{**}$ then $\|x_\alpha\| \to 1$ and $\frac{x_\alpha}{\|x_\alpha\|} \xrightarrow{\tau} x$. Thus $S \subseteq S^{**}$ is $\tau$-dense, thus $S \cap U \neq \emptyset$. If now $y_1, y_2 \in S \cap U$ then $|\varphi(y_1) + \varphi(y_2)| > 2 - \delta(\varepsilon)$. With $|\varphi| = 1$ this implies $\|y_1 + y_2\| > 2 - \delta(\varepsilon)$. Thus by uniform convexity we have $\|y_1 - y_2\| < \varepsilon$. Since every net in $S$ that $\tau$-converges to $x^*$ ultimately lives in $U$, picking any $y_1 \in S \cap U$ we have $\|x^* - y_1\| \leq \varepsilon$. But this contradicts the choice of $\varepsilon$.

The converse of the theorem is not true, but the construction of counterexamples is laborious. Note also that the dual of a uniformly convex space need not be uniformly convex!

B.29 Theorem For every measure space $(X, A, \mu)$ and $1 < p < \infty$, the space $L^p(X, A, \mu; \mathbb{F})$ is uniformly convex and therefore reflexive.

Proof. Reflexivity now follows from Theorem B.28.}

B.30 Remark The uniform convexity of $L^p$ for $1 < p < \infty$ was first proven by Clarkson in 1936 with a fairly complicated proof. (Reflexivity was known earlier thanks to F. Riesz’ proof of $(L^p)^* \cong L^q$.) A simpler proof, still giving optimal bounds, can be found in [31]. The above short proof is due to [30].

Now we are in a position to complete the determination of $L^p(X, A, \mu)^*$ for arbitrary measure space $(X, A, \mu)$ and $1 < p < \infty$ without invocation of the Radon-Nikodym theorem:

B.31 Corollary Let $1 < p < \infty$ and $(X, A, \mu)$ any measure space. Then the canonical map $L^q(X, A, \mu; \mathbb{F}) \to L^p(X, A, \mu; \mathbb{F})^*$ is an isometric bijection.

Proof. Let $(X, A, \mu)$ be any measure space, $1 < p < \infty$ and $q$ the conjugate exponent. We abbreviate $L^p(X, A, \mu)$ to $L^p$. As discussed (without complete, but hopefully sufficient detail) in Section 5.6, the map $\varphi : L^q \to (L^p)^*$, $g \mapsto \varphi_g$ is an isometry, so that only surjectivity remains to be proven. Assume $\varphi(L^q) \subseteq \subset (L^p)^*$. The subspace being closed (since $L^q$ is complete and $\varphi$ is an isometry), by Hahn-Banach there is a $0 \neq \psi \in (L^p)^*$ such that $\psi \upharpoonright \varphi(L^q) = 0$. By reflexivity of $L^p$ (Theorem B.29), there is an $f \in L^p$ such that $\psi = i_L^p(f)$. This implies $\varphi_g(f) = \psi(\varphi_g) = 0$ for all $g \in L^q$. With $\varphi_g(f) = \int f g d\mu = \varphi_f(g)$, where $\varphi : L^p \to (L^q)^*$ is the canonical map, this implies $\varphi_f = 0$. Since $\varphi$ is an isometry, we have $\hat{f} = 0$ and therefore $\psi = 0$, which is a contradiction. Thus $\varphi : L^q \to (L^p)^*$ is surjective.
B.7 Schur’s theorem

As on earlier occasions, we abbreviate $\ell^1 = \ell^1(\mathbb{N}, \mathbb{F})$.

B.32 Theorem (I. Schur) If $g$, $\{f_n\}_{n \in \mathbb{N}} \subseteq \ell^1(\mathbb{N}, \mathbb{F})$ and $f_n \overset{w}{\to} g$ then $\|f_n - g\|_1 \to 0$.

Proof. It clearly suffices to prove this for $g = 0$, thus $\ell^1 \ni f_n \overset{w}{\to} 0$ $\implies \|f_n\|_1 \to 0$. We will follow the gliding hump argument in [3] very closely.

Assume that $f_n \overset{w}{\to} 0$, but $\|f_n\| \neq 0$. Since $\delta_m \in \ell^\infty \cong (\ell^1)^*$, the first fact clearly implies $f_n(m) = \varphi_\delta_m(f_n) \overset{n \to \infty}{\to} 0$ for all $m$. And by the second assumption there exists $\varepsilon > 0$ such that $\|f_n\|_1 \geq \varepsilon > 0$ for infinitely many $n$. Using this, we inductively define $\{n_k\}, \{r_k\} \subseteq \mathbb{N}$ as follows:

(a) Let $n_1$ be the smallest number for which $\|f_{n_1}\|_1 \geq \varepsilon$.
(b) Let $r_1$ be the smallest number for which $\sum_{i=1}^{r_1} |f_{n_1}(i)| \geq \frac{\varepsilon}{5}$ and $\sum_{i=r_1+1}^{\infty} |f_{n_1}(i)| \leq \frac{\varepsilon}{5}$.

For $k \geq 2$:

(c) Let $n_k$ be the smallest number such that $n_k > n_{k-1}$ and $\|f_{n_k}\|_1 \geq \varepsilon$ and $\sum_{i=1}^{r_{k-1}} |f_{n_k}(i)| \leq \frac{\varepsilon}{5}$.
(d) Let $r_k$ be the smallest number such that $r_k > r_{k-1}$ and $\sum_{i=r_{k-1}+1}^{r_k} |f_{n_k}(i)| \geq \frac{\varepsilon}{2}$ and $\sum_{i=r_k+1}^{\infty} |f_{n_k}(i)| \leq \frac{\varepsilon}{5}$.

The reader should convince herself that the existence of such $n_k, r_k$ follows from our assumptions!

Now define $\{c_i\}_{i \in \mathbb{N}}$ by $c_i = \text{sgn}(f_{n_k}(i))$ where $k$ is uniquely determined by $r_{k-1} < i \leq r_k$ with $r_0 = 0$. Now clearly $c = \{c_i\} \in \ell^\infty$, and for all $k$ we have, using the lower bound in (b),(d),

$$\sum_{i=r_{k-1}+1}^{r_k} c_i f_{n_k}(i) = \sum_{i=r_{k-1}+1}^{r_k} |f_{n_k}(i)| \geq \frac{\varepsilon}{2},$$

while using $|c_i| \leq 1$ and the upper bounds in (b),(c),(d) we have

$$\sum_{i=1}^{r_{k-1}} |c_i f_{n_k}(i)| \leq \sum_{i=1}^{r_{k-1}} |f_{n_k}(i)| \leq \frac{\varepsilon}{5}, \quad \sum_{i=r_{k-1}+1}^{\infty} |c_i f_{n_k}(i)| \leq \sum_{i=r_{k-1}+1}^{\infty} |f_{n_k}(i)| \leq \frac{\varepsilon}{5}.$$

Thus $|\varphi_c(f_{n_k})| \geq \frac{\varepsilon}{2} - \frac{\varepsilon}{5} - \frac{\varepsilon}{5} = \frac{\varepsilon}{10}$ for all $k$, so that $\varphi_c(f_n) \neq 0$. Since this contradicts the assumption $f_n \overset{w}{\to} 0$, we must have $\|f_n\|_1 \to 0$.

\[\square\]

B.33 Remark In the above proof, the gliding hump philosophy is much more clearly visible than in the proof of Theorem 9.2: The gliding hump is precisely the dominant contribution to $\varphi_c(f_{n_k})$ coming from the $i$ in the interval $\{r_k + 1, \ldots, r_{k+1}\}$, which moves to infinity as $k \to \infty$.

Note also that the determination of the $n_k, r_k$ in the above proof was deterministic, using no choice axiom at all. In this sense the proof is better than the alternative one using Baire’s theorem, thus countable dependent choice, cf. e.g. [11, Proposition V.5.2], which nevertheless is instructive. But of course also the above proof is non-constructive in the somewhat extremist sense of intuitionism since the necessary $\varepsilon > 0$ cannot be found algorithmically.

For a high-brow interpretation of Schur’s theorem in terms of Banach space bases see [1, Section 2.3]. But also this discussion uses gliding humps.\[\square\]
B.8 The Fuglede-Putnam theorem

B.34 Theorem Let $\mathcal{A}$ be a unital $C^*$-algebra over $\mathbb{C}$.

(i) Let $a, c \in \mathcal{A}$. If $a$ is normal and $ac = ca$ then $a^*c = ca^*$ (and $ac^* = c^*a$).

(ii) Let $a, b, c \in \mathcal{A}$. If $a, b$ are normal and $ac = cb$ then $a^*c = cb^*$.

The theorem is quite remarkable, and asked for a proof one probably wouldn’t know where to begin. For matrices this is quite easy to prove, as we do for (i): Normality of $a$ implies the existence of an ONB $\{e_i\}$ such that $a = \sum_i \lambda_i P_i$, where $P_i(\cdot) = e_i(\cdot, e_i)$. Now $ac = ca$ implies $P_i c = c P_i$ for all $i$, from which $a^*c = ca^*$ is immediate. This argument can be extended to operators on infinite dimensional spaces, cf. [75]. But the following is quite different:

Proof. Obviously (i) is just the special case $a = b$ of (ii).

(ii) We define $f : \mathbb{C} \to \mathcal{A}, z \mapsto e^{z a^*} e^{-z b^*}$, where $e^a = \exp(a)$ is defined in terms of the power series as in Section 13.1. Expanding the two power series in the definition of $f$ we have

$$f(z) = e^{za^*} e^{-zb^*} = \left(\sum_{k=0}^{\infty} \frac{z^k (a^*)^k}{k!}\right) c \left(\sum_{l=0}^{\infty} \frac{(-z)^l (b^*)^l}{l!}\right) = \sum_{n=0}^{\infty} z^n d_n \ \forall z \in \mathbb{C}$$

for certain $d_n \in \mathcal{A}$. (The reshuffling is justified by the uniform convergence of the series.) We only need $d_1 = a^*c - cb^*$, which is quite obvious. Thus the theorem follows if we prove $d_1 = 0$.

By induction, the assumption $ab = cb$ is seen to imply $a^n c = cb^n$. Multiplying by $z^n/n!$ and summing over $n \in \mathbb{N}_0$ gives $e^{za^*} = e^{zb^*}$ for all $z \in \mathbb{C}$, thus also $e^{za^*} e^{-zb^*} = c$. Thus

$$f(z) = e^{za^*} e^{-zb^*} = e^{za^*} (e^{-za^*} e^{zb^*}) e^{-zb^*} = e^{za^* - za} e^{zb - zb^*} = e^{2 \text{Im}(za)} e^{-i 2 \text{Im}(zb)},$$

where $e^{za^*} e^{-za} = e^{za^* - za}$ is true due to the normality $aa^* = a^*a$ of $a$, and similarly for $b$. Now $2 \text{Im}(za), 2 \text{Im}(zb)$ are self-adjoint so that $e^{2 \text{Im}(za)}$ and $e^{2 \text{Im}(zb)}$ are unitary for all $z \in \mathbb{C}$, cf. Remark 12.24(ii), thus bounded. This proves that $f : \mathbb{C} \to \mathcal{A}$ is bounded.

Thus for every $\varphi \in \mathcal{A}^*$, the function $z \mapsto \varphi(f(z))$ is entire and bounded, thus constant by Liouville’s theorem. Thus for all $z, z' \in \mathbb{C}$, $\varphi \in \mathcal{A}^*$ we have $\varphi(f(z) - f(z')) = 0$. Hahn-Banach now implies $f(z) - f(z') = 0 \ \forall z, z'$, thus $f$ is constant. In particular, $d_1 = f'(0) = 0$. $\blacksquare$

B.35 Remark (i) was proven by Fuglede in 1950, (ii) by Putnam in 1951. The above elegant proof is due to Rosenblum (1958). Nevertheless, the appeal to complex analysis is redundant and somewhat misleading since for a bounded function given in terms of a power series of infinite convergence radius, as is the case here, Liouville’s theorem has a proof that involves neither the notion of holomorphy nor the general path independence of contour integrals: $\square$

B.36 Lemma Let $f(z) = \sum_{n=0}^{\infty} c_n z^n$ have infinite convergence radius (i.e. $\lim \sup_n |c_n|^{1/n} = 0$) and assume that $f$ is bounded. Then $c_n = 0 \ \forall n \geq 1$.

Proof. Let $r > 0$, $m \in \mathbb{N}$. Then

$$\int_0^{2\pi} e^{-imt} f(re^{it}) dt = \int_0^{2\pi} e^{-imt} \left(\sum_{n=0}^{\infty} r^n e^{int} c_n\right) dt = \sum_{n=0}^{\infty} r^n c_n \int_0^{2\pi} e^{i(n-m)t} dt = 2\pi r^m c_m,$$

where the interchange of integration and summation was justified by the uniform convergence of the series, and we used $\int_0^{2\pi} e^{i(n-m)t} dt = 2\pi \delta_{n,m}$. With $M = \sup_{z \in \mathbb{C}} |f(z)| < \infty$ we have

$$|c_m| = \frac{1}{2\pi r^m} \left|\int_0^{2\pi} e^{-imt} f(re^{it}) dt \right| \leq \frac{M}{r^m} \ \forall m \in \mathbb{N}, \ r > 0.$$
Taking the limit \( r \to +\infty \), we have \( c_m = 0 \) for all \( m \geq 1 \).  

For another instance where the standard invocation of Liouville’s theorem can be replaced by harmonic analysis see Rickart’s proof of Theorem 11.14, where only finite cyclic groups appear!

**B.9 Glimpse of non-linear FA: Schauder’s fixed point theorem**

In this final section we give a glimpse of non-linear functional analysis by proving Schauder’s fixed point theorem, which is a generalization of Brouwer’s fixed point theorem to Banach spaces.

**B.37 Definition** A topological space \( X \) has the fixed-point property if for every continuous \( f : X \to X \) there is \( x \in X \) such that \( f(x) = x \), i.e. a fixed-point.

**B.38 Theorem (Brouwer, Hadamard, 1910)** \(^{65} \) \([0,1]^n \) has the fixed point property. The same holds for every non-empty compact convex subset of \( \mathbb{R}^n \).

The second result follows from the first since such an \( X \) is homeomorphic to some \([0,1]^m\). There are many proofs of the first result. For what probably is the simplest proof (due to Kulpa) of the first statement, using only some easy combinatorics, see \(^{46} \). (Proofs using algebraic topology or analysis involve inessential elements and don’t reduce the combinatorics.)

**B.39 Theorem (Schauder 1930)** \(^{66} \) Every non-empty compact convex subset \( K \) of a normed vector space has the fixed point property.

**Proof.** Let \((V, \| \cdot \|)\) be a normed vector space, \( K \subseteq V \) a non-empty compact convex subset and \( f : K \to K \) continuous. Let \( \varepsilon > 0 \). Since \( K \) is compact, thus totally bounded, there are \( x_1, \ldots, x_n \in K \) such that \( K \subseteq \bigcup_{i=1}^{n} B(x_i, \varepsilon) \). Thus if we define \( \alpha_i(x) \geq 0 \) by

\[
\alpha_i(x) = \begin{cases} 
0 & \text{if } \|x - x_i\| \geq \varepsilon \vspace{2mm} \forall i = 1, \ldots, n, \\
\varepsilon - \|x - x_i\| & \text{if } \|x - x_i\| < \varepsilon 
\end{cases}
\]

we see that for each \( x \in K \) there is at least one \( i \) such that \( \alpha_i(x) > 0 \). The functions \( \alpha_i \) clearly are continuous. Thus also the map

\[
P_\varepsilon : K \to K, \quad x \mapsto \frac{\sum_{i=1}^{n} \alpha_i(x)x_i}{\sum_{i=1}^{n} \alpha_i(x)}
\]

is continuous. Since \( P_\varepsilon(x) \) is a convex combination of those \( x_i \) for which \( \|x - x_i\| < \varepsilon \), we have \( \|P_\varepsilon(x) - x\| < \varepsilon \) for all \( x \in K \). The finite dimensional subspace \( V_n = \text{span}(x_1, \ldots, x_n) \subseteq V \) is isomorphic to some \( \mathbb{R}^m \), and by Proposition 4.19 the restriction of the norm \( \| \cdot \| \) to \( V_n \) is equivalent to the Euclidean norm on \( \mathbb{R}^m \). Thus the convex hull \( \text{conv}(x_1, \ldots, x_n) \subseteq V_n \) into which \( P_\varepsilon \) maps is homeomorphic to a compact convex subset of \( \mathbb{R}^m \) and thus has the fixed point property by Theorem B.38. Thus if we define \( f_\varepsilon = P_\varepsilon \circ f \) then \( f_\varepsilon \) maps \( \text{conv}(x_1, \ldots, x_n) \) into itself and thus has a fixed point \( x' = f_\varepsilon(x') \). Now,

\[
\|x' - f(x')\| \leq \|x' - f_\varepsilon(x')\| + \|f_\varepsilon(x') - f(x')\| = \|f_\varepsilon(x') - f(x')\| = \|P_\varepsilon(f(x')) - f(x')\| < \varepsilon.
\]


\(^{66}\)Juliusz Schauder (1899-1943). Born in Lwow/Lviv (now Ukraine, then Lemberg in the Austrian empire) and killed by the Nazis during WW2.
Since $\varepsilon > 0$ was arbitrary, we find $\inf \{\|x - f(x)\| \mid x \in K\} = 0$. Since $K$ is compact and $x \mapsto \|x - f(x)\|$ continuous, the infimum is assumed, thus $f$ has a fixed point in $K$. ■

The use of methods/results from algebraic topology is quite typical for non-linear functional analysis. (But also linear functional analysis connects to algebraic topology, for example via K-theory, cf. e.g. [49, Chapter 7].)

B.40 Definition Let $V$ be a Banach space and $W \subseteq V$. A map $f : W \to V$ is called compact if it is continuous and $f(S) \subseteq V$ is precompact for every bounded $S \subseteq V$.

B.41 Corollary Let $V$ be a Banach space and $C \subseteq V$ closed, bounded and convex. If $f : C \to V$ is compact and $f(C) \subseteq C$ then $f$ has a fixed point in $C$.

Proof. If $W \subseteq V$ is compact then its convex hull $\text{co}(W) = \{tx + (1-t)y \mid x, y \in W, t \in [0,1]\}$ is the image of the compact space $W \times W \times [0,1]$ under the continuous map $(x, y, t) \mapsto tx + (1-t)y$ and therefore compact.

Now $\overline{f(C)} \subseteq V$ is compact by boundedness of $C$ and compactness of $f$, thus by the above also $K = \text{co}(f(C))$ is compact and, of course, convex. Thus $K$ has the fixed point property by Schauder’s theorem. Since $C$ is closed and convex, we have $K \subseteq C$, thus $f$ is defined on $K$ and maps it into $f(K) \subseteq f(C) \subseteq C$. Thus $f$ has a fixed point $x \in K \subseteq C$. ■
C Tentative schedule (14 lectures à 90 minutes)

1. Introduction, motivation. TVS. Normed spaces, bounded linear maps
2. Continuation of basic material (Sections 1-4). Sequence spaces $\ell^p(S)$: proof of Hölder and Minkowski inequalities.
4. From Riesz lemma to Sect. 6.3.
7. Strong convergence, Banach-Steinhaus. Many continuous functions with divergent Fourier series. Open mapping, bounded inverse and closed graph theorems. Invertibility of Banach space operators: (i)$\iff$(ii) in Proposition 10.24.
10. Sections 12.3-4 and 13.1-2 (incl. brief discussion of Weierstrass and Tietze theorems).
13. Section 16: Spectral theorems for normal operators.
All papers appearing in the bibliography are cited somewhere, but not all books. Still, all are worth looking at.

References


