Spring 2013: Functional Analysis[1]

Mark H. Kim
Based on Lectures by C. Sinan Güntürk

[1] Last updated: 5/13/2013

This is an extended version of the spring 2013 lectures on functional analysis by Professor Sinan Güntürk at the Courant Institute of Mathematical Sciences of New York University. The course assumes a modest background in linear algebra, advanced calculus, and measure theory, and it was aimed primarily at first-year doctoral students and advanced master's students.

The first two chapters contain edited transcrptions of the lectures. The results and their proofs recorded therein are largely from the lectures, though I have made modifications, rearrangements, and additions as I saw fit. For the most part, I have consulted [**Bre11**], [**Lax02**], [**Rud91**], [**SS11**], [**Yos80**], and [**Dou98**] for filling in the details. Other references I have used can be found in the bibliography.

Appendix A collects a number of notes accompanying the lectures. These are at times informal and are certainly not meant to be self-contained introductions to the topics covered. Appendix B contains the exercises in the main text and the homework problems given in class, as well as a selection of old oral qualifying exam problems from the Courant Institute. The problems are numbered sequentially throughout the chapter and the references to "Problem n" next to each exercise in the main text refers to this numbering.

Corrections and general comments about these notes are gratefully received at `markhkim@math.nyu.edu`. The current version of this document is available at

`http://markhkim.com/writing`

Mark Kim
Courant Institute, NYU

Thanks to Igor Balla, Giulia DeSalvo, Vitaly Kuznetsov, Hoyeon Lee, Cafer Yıldırım, and Scott Yang for corrections, comments, and clarifying remarks!

Contents

CHAPTER 1

Topological Vector Spaces

1. Introduction

The topic of this course is *functional analysis*, which is a lovely blend of linear algebra, topology, and analysis. Unlike linear algebra, however, our focus will be on infinite-dimensional vector spaces with a topology, movitated by the study of function spaces and operators on them. What difference does this make? Recall, for example, that the Bolzano-Weierstrass theorem guarantees the existence of a convergence subsequence for each bounded sequence in $\mathbb{R}^n$. In comparison, the sequence $e_n(k) = \delta_{nk}$ in the space $l^2(\mathbb{N})$ of sequences $(x_n)_{n=1}^{\infty}$ of complex numbers with finite l^2-*norm*

$$\|(x_n)\|_2 = (\sum_{n=1}^{\infty} |x_n|^2)^{1/2},$$

admits no convergent subsequence in the norm topology.

The sequence $(e_n)_{n=1}^{\infty}$ is nevertheless quite nice, and we seek a way to remedy the situation by considering a different notion of convergence. Recall that $l^2(\mathbb{N})$ is equipped with an inner product, given by

$$\langle (x_n), (y_n) \rangle = \sum_{n=1}^{\infty} x_n \overline{y_n}.$$

We now declare by fiat that a sequence $(x^k) = ((x_n^k)_{n=1}^{\infty})_{k=1}^{\infty}$ converges to a point $x = (x_n)$ in $l^2(\mathbb{N})$ *weakly* if $\langle x^k, y \rangle \to \langle x, y \rangle$ in the norm topology for all $y \in l^2(\mathbb{N})$. Parseval's theorem then yields the representation

$$\|y\|_2 = \sum |\langle e_n, y \rangle|^2$$

for each $y \in l^2(\mathbb{N})$, from which it is easy to see that $(e_n)_{n=1}^{\infty}$ converges to 0 weakly.

What have we done here? We have endowed $l^2(\mathbb{N})$ with a new topology, in which the usual notion of convergence is precisely the weak convergence defined above. This sort of reasoning—changing the topology on a given vector space to "force" convergence—will appear time and again. The spaces we shall encounter will often be normed linear spaces or inner-product spaces, but we will also see examples of *topological vector spaces*: vector spaces with a topology that turns vector addition and scalar multiplication into continuous maps.

Continuous linear operators between two topological vector spaces—in particular, continuous linear functionals—will play a crucial role in the study of these spaces. Of great importance in the study of these operators are the Hahn-Banach theorem, the Open Mapping Theorem, the Closed Graph Theorem, and the Uniform Boundedness Principle. While many examples of spaces that we study in functional analysis are complete inner-product spaces (*Hilbert spaces*) or complete normed linear spaces (*Banach spaces*), we shall see that these theorems find their natural context in *convexity* and *completeness*.

2. Normed Linear Spaces

Most of the spaces we study in this course are *normed linear spaces*:

DEFINITION 2.1. A *norm* on a vector space X over $\mathbb{F}$[1] is the function $\|\cdot\| : X \to [0,\infty)$ satisfying the following properties:

(i) $\|x\| = 0$ if and only if $x = 0$;
(ii) $\|x+y\| \leq \|x\| + \|y\|$ for all $x, y \in X$;
(iii) $\|\lambda x\| = |\lambda|\|x\|$ for each $\lambda \in \mathbb{F}$ and every $x \in X$.

X is a *normed linear space* if X is equipped with a norm.

With every normed linear space is a metric topology, given by the induced metric

$$d(x,y) = \|x-y\|.$$

EXERCISE 2.2. Check that this topology turns vector addition $(x,y) \mapsto x+y$ and scalar multiplication $(\lambda, x) \mapsto \lambda x$ into continous maps.

The induced metric satisfies the following properties:

(i) **Translation invariance**. $d(x+z, y+z) = d(x,y)$ for all $x, y, z \in X$.
(ii) **Homogeneous scaling**. $d(\lambda x, \lambda y) = |\lambda| d(x,y)$ for each $\lambda \in \mathbb{F}$ and every $x, y \in X$.

EXERCISE 2.3. Every vector space X with a metric that satisfies (i) and (ii) is a normed vector space with the norm given by $\|x\| = d(x,0)$.

Translation invariance and homogeneous scaling of the induced metric has the following consequences:

PROPOSITION 2.4. *Let X be a normed linear space over $\mathbb{F}$.*

(i) $\overline{A+y} = \overline{A} + y$ *and* $(A+y)^\circ = A^\circ + y$ *for each* $A \subseteq X$ *and every* $y \in X$.
(ii) $\overline{\lambda A} = \lambda \overline{A}$ *and* $(\lambda A)^\circ = \lambda A^\circ$ *for each* $\lambda \in \mathbb{F}$ *and every* $A \subseteq X$.
(iii) If U is open subset of X and A an arbitrary subset of X, then $U + A$ is open.
(iv) If A and B are compact subsets of X, then $A + B$ is compact.

We remark that the sum of two closed subsets is not necessarily closed. Indeed, $\mathbb{Z} + \sqrt{2}\mathbb{Z}$ is a proper dense subset of $\mathbb{R}$, which cannot be closed.

Another basic property of the norm topology is that *the norm is continuous*. In fact, the map $x \mapsto \|x\|$ is Lipschitz, for the triangle inequality yields the estimate

$$|\|x\| - \|y\|| \leq \|x-y\|.$$

As a consequence, the *open ball*

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

is open, and the *closed ball*

$$B_r[x] = \{y : \|x-y\| \leq r\}$$

is closed. Translation and scaling yields

$$B_r(x) = x + rB_1(0) \qquad \text{and} \qquad B_r[x] = x + rB_1[0].$$

[1] Throughout the course, $\mathbb{F}$, shall denote an "arbitrary field": $\mathbb{R}$ or $\mathbb{C}$.

EXERCISE 2.5. A *metric linear space* is a vector space with a translation-invariant metric. Show that every ball in a normed linear space is convex, but $\mathbb{R}^2$ with the metric

$$d(x,y) = \sqrt{|x_1 - y_1|} + \sqrt{|x_2 - y_2|}$$

is a metric linear space whose unit ball $B_1(0)$ fails to be convex.

A basic result in the theory of metric spaces is that $\overline{B_r(x)} \subseteq B_r[x]$ and $B_r[x]^\circ \supseteq B_r(x)$. In a general metric space, the equalities need not hold, for any set of cardinality at least 2 with the discrete metric

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

has the following "unit balls":

$$B_1(x) = \{x\}, \quad \overline{B_1(x)} = \{x\}, \quad B_1[x] = X.$$

The equality holds in normed linear spaces:

PROPOSITION 2.6. *In a normed linear space,* $\overline{B_1(0)} = B_1[0]$.

PROOF. It suffices to show that $B_1[0] \subseteq \overline{B_1(0)}$. We pick $x \in B_1[0]$ and construct a sequence

$$x_n = \left(1 - \frac{1}{n}\right) x.$$

Then $x_n \in B_1(0)$ for each $n \in \mathbb{N}$ and $x_n \to x$ in the norm topology, whence $x \in \overline{B_1(0)}$. □

Let us now introduce some fundamental examples of normed linear spaces.

EXAMPLE. The *p-norms* on $\mathbb{F}^n$, given by

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

is a norm. More generally, if (X, μ) is a measure space, then the *Lebesgue space* $L^p(X, \mu)$ of μ-measurable complex-valued functions on X with finite L^p-norm

$$\|f\|_p = \begin{cases} \left(\int_X |f|^p \, d\mu \right)^{1/p} & \text{if } 1 \leq p < \infty; \\ \operatorname{ess\,sup} |f| & \text{if } p = \infty; \end{cases}$$

is a normed linear space.

EXAMPLE. Let E be an arbitrary set, and consider the collection $\mathscr{B}(E)$ of bounded real-valued functions on E. The L^∞-norm is a norm on $\mathscr{B}(E)$. In particular, if E is a compact metric space, then the space $\mathscr{C}(E)$ of continuous real-valued functions on E with the L^∞-norm is a normed linear space. Similarly, $C^1([0,1])$ of continuously differentiable functions on $[0,1]$ with a "modified" L^∞-norm

$$\|f\| = \|f\|_\infty + \|f'\|_\infty$$

is a normed linear space.

EXAMPLE. If A is an invertible n-by-n matrix over $\mathbb{F}$, then

$$\|x\|_{A,p} = \|Ax\|_p$$

is a norm on $\mathbb{F}^n$.

The Lebesgue spaces are more than just normed linear spaces— they are complete metric spaces. Since these examples are of paramount importance in analysis, we abstract this property and give it the name of one of the founders of functional analysis, Stefan Banach:

DEFINITION 2.7. A *Banach space* is a normed linear space whose induced metric is complete.

Not every space is a Banach space, of course. $\mathscr{C}([0,1])$, as a subset of $L^2([0,1])$, is a dense proper subspace, hence it cannot be closed. Nevertheless, there is a Banach space that contains it: namely, $L^2([0,1])$. As it turns out, we can always find a Banach space that contains the given normed linear space.

For a precise formulation of this result, we will need some definitions. Recall that an *isomorphism* of two vector spaces is a bijective linear transformation. An isomorphism of vector spaces preserves linear structures in a way that renders the two spaces in question "essentially the same". An appropriate analogue for normed linear spaces is as follows:

DEFINITION 2.8. An *isometric isomorphism* between normed linear spaces X and Y is an isomorphism $T : X \to Y$ of vector spaces that is *isometric*, viz., $\|x\| = \|Tx\|$ for all $x \in X$. X and Y are said to be *isometrically isomorphic* if there exists an isometric isomorphism between them.

The adjective *isometric* is important, because there is another notion of isomorphism for normed linear spaces: namely, bijective linear transformations that are homeomorphisms with respect to the metric topology generated by the norm. See Appendix A, Section 1 for a more detailed discussion.

Isometrically isomorphic normed linear spaces are considered to be the "same", and so the result we wish to formalize must be understood in the isomorphic sense as well:

DEFINITION 2.9. An *embedding* of a vector space X into another vector space Y is an injective linear transformation $T : X \to Y$. An embedding $T : X \to Y$ is said to be an *isometric embedding* if X and Y are normed linear spaces and T is an isometry.

It is easy to see that T is an isometric isomorphism between X and $\operatorname{im} T$. We are now in a position to state the following theorem:

THEOREM 2.10. *Every normed linear space can be isometrically embedded into a Banach space. The Banach space can be chosen in a way that the embedding is dense. This choice is unique up to an isometric isomorphism and is referred to as the* completion *of the normed linear space in question.*

We will prove this result in §5 of this chapter. For now, we consider the following simple exercise that highlights the basic concepts related to the above result.

EXERCISE 2.11. Let $(X, \|\cdot\|)$ be a normed linear space and M a linear subspace of X. Show that $\overline{M}$ is also a linear subspace of X. Moreover, if $(x_n)_{n=1}^\infty$ is a sequence in X, then show that $x_n \to x$ implies $\|x_n\| \to \|x\|$ in $\mathbb{R}$. Show that if $(x_n)_{n=1}^\infty$ is a Cauchy sequence, then $(\|x_n\|)_{n=1}^\infty$ is a Cauchy sequence.

Before we proceed, we remark that not every metric on a vector space is induced by a norm.

EXERCISE 2.12. Check that

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

is a translation-invariant metric on $\mathbb{R}^{\mathbb{N}}$ that does not scale homogeneously. This is an example of a *Fréchet space* (Definition 9.7), a generalization of a Banach space with multiple, rather than one, "complete norms".

We also remark that not every norm generates a distinct norm topology.

DEFINITION 2.13. Two norms $\|\cdot\|_a$ and $\|\cdot\|_b$ on X are *equivalent* if there exist two positive constants K_1 and K_2 such that

$$K_1\|x\|_a \leq \|x\|_b \leq K_2\|x\|_a$$

for all $x \in X$.

EXERCISE 2.14. Show that equivalent norms induce the same metric topology, viz., a subset E is open with respect to $\|\cdot\|_a$ if and only if E is open with respect to $\|\cdot\|_b$.

It turns out that there is only one norm topology on finite-dimensional vector spaces of a fixed dimension.

THEOREM 2.15. *All norms on a finite-dimensional vector space are equivalent.*

PROOF. Let X be a finite-dimensional normed linear space over $\mathbb{F}$ with the norm $\|\cdot\|$. We fix a basis $\{v_1, \ldots, v_n\}$ of X and consider the canonical isomorphism $T : \mathbb{F}^n \to X$, given by

$$T(\lambda_1, \ldots, \lambda_n) = \sum_{i=1}^n \lambda_i v_i$$

for each $\lambda = (\lambda_1, \ldots, \lambda_n)$ in $\mathbb{F}^n$. The Cauchy-Schwarz inequality implies that

$$\|T\lambda\| \leq \sum_{i=1}^n |\lambda_i|\|v_i\| \leq \left(\sum_{i=1}^n \|v_i\|^2\right)^{1/2} \left(\sum_{i=1}^n |\lambda_i|^2\right)^{1/2} = K_2\|\lambda\|_2$$

for each $\lambda \in \mathbb{F}^n$. From this, we see that $\lambda \mapsto \|T\lambda\|$ is continuous, for

$$|\|T\lambda_1\| - \|T\lambda_2\|| \leq \|T(\lambda_1 - \lambda_2)\| \leq K_2\|\lambda_1 - \lambda_2\|_2.$$

We now take the unit sphere $S = \{x \in \mathbb{F}^n : \|x\|_2 = 1\}$ in $\mathbb{F}^n$, which is compact. The map $\lambda \mapsto \|T\lambda\|$ therefore attains its minimum K_1 on S. The minimum must be strictly postiive, for the norm is only zero at 0. We then have

$$\|T\lambda\| = \|\lambda\|\|T(\|\lambda\|^{-1}\lambda)\| \geq K_1\|\lambda\|_2$$

for all nonzero $\lambda \in \mathbb{F}^n$, and it follows that

$$K_1\|\lambda\|_2 \leq \|T\lambda\| \leq K_2\|\lambda\|_2. \tag{2.16}$$

We now take another norm $\|\cdot\|'$ on X. The same argument as above shows that we have positive constants K_1' and K_2' such that

$$K_1'\|\lambda\|_2 \leq \|T\lambda\|' \leq K_2'\|\lambda\|_2. \tag{2.17}$$

Let $x = T\lambda$. We apply (2.17) and then (2.16) to obtain the estimate

$$\|x\|' \leq K_2'\|\lambda\|_2 \leq \frac{K_2'}{K_1}\|x\|.$$

Similarly, we apply (2.16) and then (2.17) to obtain the estimate

$$\|x\| \leq K_2\|\lambda\|_2 \leq \frac{K_2}{K_1'}\|x\|'.$$

Combining the two estimates, we have

$$\frac{K_1'}{K_2}\|x\|' \leq \|x\| \leq \frac{K_2'}{K_1}\|x\|'$$

for all $x \in \operatorname{im} T$. Since T is an isomorphism, this accounts for all vectors in X. □

The equivalence of norms implies that all finite-dimensional normed linear spaces of the same dimension are homeomorphic. Since finite-dimensional vector spaces of the same dimension are isomorphic, we see that they are isomorphic as topological vector spaces. They are not necessarily isomorphic as normed linear spaces, however.

Example. Consider $(\mathbb{R}^2, \|\cdot\|_1)$ and $(\mathbb{R}^2, \|\cdot\|_2)$. If $T : (\mathbb{R}^2, \|\cdot\|_1) \to (\mathbb{R}^2, \|\cdot\|_2)$ is an isometric isomorphism, then

$$\|(1-\lambda)x + \lambda y\|_1 = \|(1-\lambda)Tx + \lambda Ty\|_2 \tag{2.18}$$

for all $\lambda \in \mathbb{R}$. Letting $x = (1,0)$ and $y = (0,1)$, we see that

$$\|(1-\lambda)x + \lambda y\|_1 = 1$$

for each $0 \leq \lambda \leq 1$. But Tx and Ty are points on the ordinary unit circle in the coordinate plane $\mathbb{R}^2$, whence the segment connecting the two points cannot be on the unit circle. It follows that

$$\|(1-\lambda)x + \lambda y\|_2 < 1$$

if $0 < \lambda < 1$, contradicting (2.18). It follows that there is no isometric isomorphism from $(\mathbb{R}^2, \|\cdot\|_1)$ to $(\mathbb{R}^2, \|\cdot\|_2)$.

Nevertheless, all finite-dimensional normed linear spaces of the same dimension are similar enough for the most part. For one, all of them are complete:

Corollary 2.19. *All finite-dimensional normed linear spaces are Banach spaces.*

Proof. Let $(X, \|\cdot\|)$ be an n-dimensional normed linear space and fix a Cauchy sequence $(x_n)_{n=1}^\infty$ in V. The equivalence of norms (Theorem 2.15) implies that the Euclidean norm (l^2-norm) on $\mathbb{R}^n$ and $\|\cdot\|$ are equivalent. More precisely, if we fix a basis $\{v_1, \ldots, v_n\}$ of X and take the canonical isomorphism $T : X \to \mathbb{R}^n$, given by

$$T\left(\sum_{i=1}^n \lambda_i v_i\right) = (\lambda_1, \ldots, \lambda_n),$$

then we can find positive constants K_1 and K_2 such that

$$K_1\|Tx\|_2 \leq \|x\| \leq K_2\|Tx\|_2. \tag{2.20}$$

We now note that $(Tx_n)_{n=1}^{\infty}$ is Cauchy in $\mathbb{R}^n$, and so we can find its limit y. We set $x = T^{-1}y$, so that $\|Tx_n - Tx\|_2 \to 0$ as $n \to \infty$. (2.15) now implies that

$$\lim_{n\to\infty} \|x_n - x\| \leq \lim_{n\to\infty} K_2\|Tx_n - Tx\|_2 = 0,$$

whence X is a Banach space. □

This, in particular, shows that all finite-dimensional subspaces of a normed linear space are complete. Since completeness implies closedness, we have the following corollary:

COROLLARY 2.21. *All finite-dimensional subspaces of a normed linear space are closed.* □

We remark that the finiteness assumption is essential. Indeed, there are many examples of infinite-dimensional subspaces that are not closed.

EXERCISE 2.22. Fix $1 \leq p < \infty$ and recall that the space $\mathcal{C}^\infty([0,1])$ of smooth functions is a dense subspace of $L^p([0,1])$. Conclude that the space $\mathcal{C}^n([0,1])$ of n-times continuously differentiable functions on $[0,1]$ is not closed in the norm topology of $L^p([0,1])$.

EXERCISE 2.23. Nevertheless, there are concrete examples of nontrivial closed infinite-dimensional subspaces of infinite-dimensional spaces that are not merely closures of infinite-dimensional subspaces. As an exercise, determine which of the following subspaces of the Banach space $l^\infty(\mathbb{N})$ are closed.

(a) c, the space of convergent sequences.
(b) c_0, the space of sequences convering to zero.
(c) $l^1(\mathbb{N})$, the space of absolutely summable sequences.
(d) c_c, the set of finitely supported sequences.

Hint: For (c) and (d), use the *cutoff operator* $\mathscr{C}_N$ that sends $(a_n)_{n=1}^{\infty}$ to the sequence

$$a_1, a_2, \ldots, a_k, 0, 0, \ldots.$$

EXERCISE 2.24. The space of bounded and continuous functions on $\mathbb{R}$ is a proper closed subspace of $L^\infty(\mathbb{R})$.

We also note that finiteness is used in a crucial manner in the proof of the equivalence of norms: first, to compute the constant $(\sum_{i=1}^n \|v_i\|^2)^{1/2}$, and second, to obtain the minimum of the map $\lambda \mapsto \|T\lambda\|$ via the compactness of the unit sphere in $\mathbb{F}^n$.

EXAMPLE. On an infinite-dimensional vector space, we can easily construct non-equivalent norms. Consider, for example, the vector space X of sequences $(x_n)_{n=1}^{\infty}$ that vanish after finitely many terms. X can then be considered as linear subspaces of $l^1(\mathbb{N})$ and $l^2(\mathbb{N})$, respectively. We now let

$$x_n^N = \begin{cases} 1 & \text{if } n \leq N; \\ 0 & \text{if } n > N; \end{cases}$$

and consider the family of sequences $x^N = (x_n^N)_{n=1}^{\infty}$. We have

$$\|x^N\|_1 = N \quad \text{and} \quad \|x^N\|_2 = \sqrt{N},$$

and, by sending $N \to \infty$, we see that no norm equivalence is possible.

More generally, every infinite-dimensional vector space admits non-equivalent norms.

EXERCISE 2.25. Let V be an infinite-dimensional vector space and let $\mathcal{B} = \{v_\alpha\}$ be a Hamel basis of V. Check that each function $f : \mathcal{B} \to [0, \infty)$ defines a norm $\|\cdot\|_f$, given by

$$\left\| \sum_\alpha \lambda_\alpha v_\alpha \right\| = \sum_\alpha |\lambda_\alpha| f(v_\alpha).$$

Conclude that V admits non-equivalent norms $\|\cdot\|_f$ and $\|\cdot\|_g$.

This suggests that the unit sphere in an infinite-dimensional normed linear space is not compact, for otherwise the proof of Theorem 2.15 would have gone through.

THEOREM 2.26. *A normed linear space is finite-dimensional if and only if the closure of each bounded subset is compact.*

EXERCISE 2.27 ([Tao10], Example 1.9.8). As a concrete example, we consider $l^p(\mathbb{N})$ with $1 \leq p < \infty$. The set $\{e_n\}_{n \in \mathbb{N}}$ of sequences

$$e_n(m) = \begin{cases} 1 & \text{if } n = m; \\ 0 & \text{if } n \neq m; \end{cases}$$

is closed and bounded in $l^p(\mathbb{N})$, but is not compact. Nevertheless, a closed and bounded set $K \subseteq l^p(\mathbb{N})$ such that every $\varepsilon > 0$ furnishes an index n with the estimate

$$\left(\sum_{m > n} |f(m)|^p \right)^{1/p} \leq \varepsilon$$

for all $f \in K$ is compact.

PROOF OF THEOREM. The "only if" part follows at once from the equivalence of norms. To show the "if" part, we assume that X is infinite-dimensional. We shall need the following lemma, due to F. Riesz:

LEMMA 2.28 (Riesz). *Let $(X, \|\cdot\|)$ be a normed linear space and Y a closed, proper subspace of X. For each $\varepsilon > 0$, we can find a vector $z \in X$ such that $\|z\| = 1$ and $\|z - y\| > 1 - \varepsilon$ for all $y \in Y$.*

PROOF OF LEMMA. $X \smallsetminus Y$ is nontrivial, hence we can pick a vector x_0 in it. The quantity

$$d = \inf_{y \in Y} \|x_0 - y\|$$

is positive, for otherwise x_0 is in the closure of Y. Since Y is closed, this is obviously false.

Now, for each $\eta > 0$, we can find $y_\eta \in Y$ such that

$$d \leq \|x_0 - y_\eta\| < d + \eta.$$

We let

$$z_\eta = \|x_0 - y_\eta\|^{-1}(x_0 - y_\eta).$$

For each $y \in Y$, we have the estimate

$$\begin{aligned} \|z_\eta - y\| &= \left\| y - \|x_0 - y_\eta\|^{-1}(x_0 - y_\eta) \right\| \\ &= \frac{1}{\|x_0 - y_\eta\|} \left\| \|x_0 - y_\eta\| y - (x_0 - y_\eta) \right\| \\ &= \frac{1}{\|x_0 - y_\eta\|} \|x_0 - (y_\eta + \|x_0 - y_\eta\| y)\| \\ &> \frac{1}{d+\eta} \cdot d, \end{aligned}$$

because $y_\eta + \|x_0 - y_\eta\| y \in Y$.

We now fix a sufficiently small $\varepsilon > 0$ and pick η such that $\frac{d}{d+\eta} = 1 - \varepsilon$. Letting $z = z_\eta$, we see that the above estimate translates to $\|z\| = 1$ and

$$\|z - y\| > \frac{d}{d+\eta} = 1 - \varepsilon,$$

which was the conclusion of the lemma. □

We now suppose for a contradiction that the closure of each bounded subset of X is compact. Fix a unit vector x_1 in X and let Y_1 be the span of x_1. The lemma above furnishes a unit vector x_2 in X such that $\|x_2 - y\| \geq \frac{1}{2}$ for all $y \in Y_1$. In particular, $\|x_2 - x_1\| \geq \frac{1}{2}$, and so the span Y_2 of x_1 and x_2 is distinct from Y_1. Since X is infinite-dimensional, we continue this process to obtain a strictly increasing sequence

$$Y_1 \subset Y_2 \subset \cdots \subset Y_n \subset \cdots .$$

of subspaces of X, and the corresponding sequence $(x_n)_{n=1}^\infty$ of vectors such that $\|x_m - x_n\| \geq \frac{1}{2}$ whenever $m \neq n$. We can see at once that the sequence $(x_n)_{n=1}^\infty$ has no convergent subsequence, despite its boundedness. It follows that $\{x_n\}_{n \in \mathbb{N}}$ is a closed, bounded subset of X that is not compact, which is absurd. □

3. Bounded Linear Operators

The argument for the continuity of $\lambda \mapsto \|T\lambda\|$ given in proof of the equivalence of norms in finite-dimensional vector spaces shows that $\lambda \mapsto T\lambda$ is Lipschitz continuous. Indeed, we have

$$\|T\lambda_1 - T\lambda_2\| \leq C\|\lambda_1 - \lambda_2\|_2$$

with the positive constant $C = (\sum_{i=1}^n \|v_i\|^2)^{1/2}$, and the Lipschitz continuity follows at once. Again, finiteness plays an essential role here, as not every linear transformation between infinite-dimensional normed linear spaces is continuous.

Example. For example, we consider the differential operator $D : \mathscr{C}^1([0,1]) \to \mathscr{C}([0,1])$ given by

$$Df = f'.$$

We let $f_n(x) = n^{-1} \sin(nx)$ for each $n \in \mathbb{N}$ and observe that $\|f_n\|_\infty \to 0$ in $\mathcal{C}^1[0,1]$. We see, however, that $Df_n = \cos nx$, and so $\|Df_n\|_\infty = 1$ for all $n \in \mathbb{N}$. Therefore, $Df_n \not\to 0$ in $\mathcal{C}[0,1]$, whence D cannot be continuous.

Example. As another example, we consider the space X of sequences that vanish after finitely many terms. The map

$$l(x) = \sum_{n=1}^{\infty} x_n$$

is a linear transformation from X to $\mathbb{R}$. We now let

$$x_n^N = \begin{cases} \frac{1}{N} & \text{if } n \leq N; \\ 0 & \text{if } n > N; \end{cases}$$

for each $N \in \mathbb{N}$ and observe that $\|x^N\|_\infty \to 0$ in X. Nevertheless, $l(x^N) = 1$ for all $N \in \mathbb{N}$, so that $l(x^N) \not\to 0$ in $\mathbb{R}$. It follows that l cannot be continuous.

If, however, we can find a positive constant C such that

$$\|Tx\| \leq C\|x\|,$$

as we have in the proof of the equivalence of norms, then it is straightforward to show that T is, in fact, continuous. We give this condition a name.

Definition 3.1. A linear transformation $T : X \to Y$ between $(X, \|\cdot\|_X)$ and $(Y, \|\cdot\|_Y)$ is *bounded* if there exists a positive constant C such that

$$\|Tx\| \leq C\|x\|$$

for all $x \in X$.

As it turns out, boundendess is also necessary for a linear transformation to be continuous, rendering boundedness an equivalent condition for continuity.

Proposition 3.2. *Let $T : X \to Y$ be a linear transformation between $(X, \|\cdot\|_X)$ and $(Y, \|\cdot\|_Y)$. The following are equivalent:*

(a) T is uniformly continuous.
(b) T is continuous.
(c) T is continuous at 0.
(d) T is bounded.

Proof. (a) $\Rightarrow$ (b) and (b) $\Rightarrow$ (c) are trivial. To see (c) $\Rightarrow$ (a), we fix $\varepsilon > 0$ and find a $\delta > 0$ such that $\|x\|_X < \delta$ implies $\|Tx\|_Y < \varepsilon$. Let $x_1, x_2 \in X$. If $\|x_1 - x_2\|_X < \delta$, then $\|T(x_1 - x_2)\|_Y = \|Tx_1 - Tx_2\|_Y < \varepsilon$, and so T is uniformly continuous.

(d) $\Rightarrow$ (c) is straightforward, hence it remains to show that (c) $\Rightarrow$ (d). To this end, we find a $\delta > 0$ such that $\|x\| < \delta$ implies $\|Tx\| = \|Tx - T(0)\| < 1$. Given an arbitrary vector $x \in X$, we now see that

$$\|Tx\| = \frac{2\|x\|}{\delta} \left\| T\left(\frac{\delta}{2\|x\|} x\right) \right\| < \frac{2}{\delta}\|x\|,$$

and so T is bounded. □

We now introduce a quantitive measurement of the boundedness of a continuous linear transformation.

Definition 3.3. Let $T : X \to Y$ be a bounded linear transformation. The infimum of all constants C such that

$$\|Tx\| \leq C\|x\|$$

for all $x \in X$ is the *operator norm* of T, denoted by $\|T\|_{X \to Y}$.

The name *operator* norm comes from classical analysis, in which an object that takes a function and transforms it was called an operator of functions. Since the spaces we consider in functional analysis are, by and large, abstractions of function spaces, much of the classical terminology remains. In fact, we shall often refer to linear transformations as *linear operators* from now on.

There are a number of different ways of computing the operator norm.

EXERCISE 3.4. If $T : X \to Y$ is a bounded linear operator, then

$$\|T\|_{X\to Y} = \sup_{x\neq 0} \frac{\|Tx\|_Y}{\|x\|_X} = \sup_{\|x\|_X=1} \|Tx\|_Y.$$

The name operator *norm* suggests that the collection of bounded linear operators should form a normed linear space, and this is indeed the case.

EXERCISE 3.5. The space $\mathscr{B}(X, Y)$ of bounded linear operators from X to Y is a normed linear space with the operator norm.

EXERCISE 3.6. If $T_n \to T$ in $\mathscr{B}(X, Y)$ and $x_n \to x$ in X, then $T_n x_n \to Tx$ in Y.

We remark that the notion of convergence in $\mathscr{B}(X, Y)$ is precisely the notion of convergence on the unit sphere in X. Specifically, we have the following result:

EXERCISE 3.7. $\|T_n \to T\|_{X\to Y} \to 0$ if and only if

$$\sup_{\|x\|_X=1} \|T_n x - Tx\|_Y \to 0.$$

THEOREM 3.8. *If Y is a Banach space, then so is $\mathscr{B}(X, Y)$.*

PROOF. Let $(T_n)_{n=1}^\infty$ be a Cauchy sequence in $\mathscr{B}(X, Y)$. For each $x \in X$, the sequence $(T_n x)_{n=1}^\infty$ is Cauchy in Y, as

$$\|T_n x - T_m x\| \leq \|T_n - T_m\|\|x\|.$$

By the completeness of Y, we can find a limit Tx of the sequence $(T_n x)_{n=1}^\infty$. T is clearly linear. We also observe that

$$\|Tx\| \leq \|Tx - T_n x\| + \|T_n x\| \leq \|Tx - T_n x\| + \|T_n\|\|x\|.$$

Since $(T_n)_{n=1}^\infty$ is Cauchy, $M = \sup_n \|T_n\|$ is finite, and so

$$\|Tx\| \leq \|Tx - T_n x\| + M\|x\|.$$

We now send $n \to \infty$ to conclude that

$$\|Tx\| \leq M\|x\|,$$

and so T is bounded. Therefore, $T \in \mathscr{B}(X, Y)$.

It remains to show that $\|T_n - T\| \to 0$. To this end, we fix $\varepsilon > 0$ and find $N \in \mathbb{N}$ such that $\|T_n - T_m\| < \varepsilon$ for all $m > n > N$. Then $\|T_n x - T_m x\| < \varepsilon\|x\|$ for all $x \in X$, and sending $m \to \infty$ yields

$$\|T_n x - Tx\| < \varepsilon\|x\|.$$

This implies that $\|T_n - T\| < \varepsilon$ for all $n > N$, and so $T_n \to T$ in $\mathscr{B}(X, Y)$, as was to be shown. □

Let us now consider a few important special cases of the space of bounded linear operators. If the target space Y is the ground field $\mathbb{F}$, we obtain the space of *linear functionals* that are continuous:

Definition 3.9. The *(topological) dual space* X^* of the normed linear space X is the space $\mathscr{B}(X, \mathbb{F})$ of bounded linear functionals.

The name *topological* dual space contrasts X^* with a related construct, the *algebraic dual space* X'. X' is defined to be the collection of all linear functionals on X, continuous or not. We remark that X' always has a plenty of elements: just take a basis of X' (which requires the Axiom of Choice) and define a linear functional by assigning a scalar to each basis element. The topological dual, however, is not always interesting: for some spaces, the topogical duals contain nothing but the zero functional, which sends every vector to zero. We shall have an occasion to say more about this later, when we discuss the Hahn-Banach theorem in §5 of this chapter.

In any case, we can apply Theorem 3.8 to $X^* = \mathscr{B}(X, \mathbb{F})$ even if X^* is trivial. And we do so at once:

Corollary 3.10. *X^* is a Banach space.* □

4. Completeness: Applications of the Baire Category Theorem

With the foresight that X^* is nontrivial in plenty of cases, we now exploit the *completness* of Banach spaces to establish several powerful results. These results will be derived as corollaries of the *Baire category theorem*, which, in turn, deals with a characterization of the notion of *smallness* in topological spaces.

4.1. Nowhere Dense Sets and the Baire Category Theorem. We begin by singling out an "obviously negligible" class of sets in topological spaces.

Definition 4.1. A subset A of a topological space X is *nowhere dense* in X if $(\overline{A})^\circ = \varnothing$.

Note that a finite union of nowhere dense sets is nowhere dense. Extending this idea, we can define a notion of smallness in topological spaces as follows:

Definition 4.2. A subset A of a topological space X is *of the first category*, or *meager*, if it is a countable union of nowhere dense sets. A is *of the second category* if A is not meager.

It is instructive to compare the notion of sets of category with the sets of measure zero. The next theorem shows that complete metric spaces cannot be too small.

Theorem 4.3 (Baire category theorem). *A complete metric space is of the second category.*

Proof. We shall need the following lemma, which shows that open, dense subsets of a complete space cannot be too small:

Lemma 4.4. *If $\{U_n\}_{n=1}^\infty$ is a sequence of open, dense subsets of a complete metric space X, then $\bigcap_n U_n$ is also dense.*

Proof of lemma. Let V be an open subset of X. We shall show that the intersection of V and $\bigcap_n U_n$ is nontrivial. To this end, we first note that $U_1 \cap V$ contains a ball $B_{r_1}(x_1)$ centered at a point $x_1 \in U_1 \cap V$. We now define $B_{r_n}(x_n)$ inductively for $n > 1$ as follows: noting that $U_n \cap B_{r_{n-1}}(x_{n-1})$ has nonempty interior, we can find a ball $B_{r_n}(x_n)$ such that $B_{r_n}[x_n] \subseteq U_n \cap B_{r_{n-1}}(x_{n-1})$.

Let us relabel $(B_{r_n}(x_n))_{n=1}^{\infty}$ if necessary to have r_n tend to zero as $n \to \infty$. Then $(x_n)_{n=1}^{\infty}$ is Cauchy in the complete metric space X, whence $x_n \to x$ for some $x \in X$. By the construction,

$$x \in U_n \cap B_{r_1}(x_1) \subseteq U_n \cap V$$

for all n, which implies the desired conclusion. □

We now return to the task at hand. Suppose that $(E_n)_{n=1}^{\infty}$ is a sequence of sets such that $(\overline{E_n})^{\circ} = \varnothing$. For each $n \in \mathbb{N}$, the complement of $\overline{E_n}$ is open. In fact, it is dense, as

$$\overline{(\overline{E_n})^c} = [(\overline{E_n})^{\circ}]^c = X.$$

By the above lemma, we have

$$\bigcap_{n=1}^{\infty} (\overline{E_n})^c \neq \varnothing,$$

which implies that

$$\bigcup_{n=1}^{\infty} \overline{E_n} \neq X.$$

Therefore,

$$\bigcup_{n=1}^{\infty} E_n \neq X,$$

and the proof is complete. □

4.2. The Open Mapping Theorem. We now derive three powerful corollaries of the Baire category theorem. The first in line is the *open mapping theorem*, which provides a sufficient condition for a linear operator to be open. We first recall the definition of openness:

DEFINITION 4.5. A function $f : X \to Y$ between topological spaces X and Y is *open* if the image of each open subset of X is open in Y.

The open mapping theorem asserts that surjectivity is enough if we are dealing with bounded linear operators between normed linear spaces.

THEOREM 4.6 (Banach-Schauder, the open mapping theorem). *If $T : X \to Y$ is a surjective bounded linear operator between Banach spaces X and Y, then T is open.*

PROOF. We simplify our task by making the following observation:

LEMMA 4.7. *Let $T : X \to Y$ be a linear operator between normed linear spaces X and Y. Then T is open if and only if $T(B_1(0))$ has nonempty interior.*

PROOF OF LEMMA. If T is open, then $T(B_1(0))$ is a nonempty open set, whence $T(B_1(0))$ must have a nonempty interior. Conversely, if $T(B_1(0))$ has a nonempty interior, then Proposition 2.4 implies that $T(B_r(x))$ must be open for each $r > 0$ and every $x \in X$. We now let E be an arbitrary open set in X. For each $x \in E$, we can find a real number $r > 0$ such that $B_r(x) \subseteq E$, whence $T(B_r(x))$ is an open neighborhood of Tx contained in $T(E)$. It follows that $T(E)$ is open. □

In light of this, we shall show that $\overline{T(B_1(0))}$ contains a ball, and generalize this result to show that $\overline{T(B_{1/2^n}(0))}$ contains a ball regardless of $n \in \mathbb{N}$. A strengthening of this argument will allow us to remove the closure, and the desired result will follow from Lemma 4.7.

We first prove that $\overline{T(B_1(0))}$ has nonempty interior[2]. Since $X = \bigcup_n B_n(0)$, we can write $Y = \bigcup T(B_n(0))$ by the surjectivity of T. The linearity of T allows us to rewrite the latter identity as

$$Y = \bigcup_{n=1}^{\infty} nT(B_1(0)),$$

whence the Baire category theorem (Theorem 4.3) applied to Y implies that at least one of the $nT(B_1(0))$ must fail to be nowhere dense. Since $nT(B_1(0))$ is homeomorphic to $T(B_1(0))$, it follows that $T(B_1(0))$ fails to be nowhere dense. In other words, $\overline{T(B_1(0))}$ has nonempty interior. We can now pick a point y_0 in the interior of $\overline{T(B_1(0))}$ and find a ball $B_r(y_0)$ contained in the interior of $\overline{T(B_1(0))}$.

Let us now show that $\overline{T(B_1(0))}$ contains a ball *centered at the origin*. Since $y_0 \in \overline{T(B_1(0))}$, we can find a point $y_1 \in T(B_1(0))$ and $x_1 \in B_1(0)$ such that $y_1 = Tx_1$ and $\|y_1 - y_0\| < r/2$. This, in particular, implies that the ball $B_{r/2}(y_1)$ is contained in $B_r(y_0)$, which, in turn, is contained in $\overline{T(B_1(0))}$.

Now, if $\|y\| < r/2$, then $y + y_1 \in B_{r/2}(y_1) \subseteq \overline{T(B_1(0))}$, whence

$$y = (y + y_1) - y_1 = (y + y_1) - Tx_1$$

is in $\overline{T(B_2(0))}$. Therefore, $\overline{T(B_2(0))}$ contains the ball $B_{r/2}(0)$, and a simple scaling argument shows that

$$B_{r/2^{n+2}}(0) \subseteq \overline{T(B_{1/2^n}(0))} \tag{4.8}$$

for each $n \in \mathbb{N}$.

Finally, we show that

$$T(B_1(0)) \supseteq B_{r/8}(0). \tag{4.9}$$

To this end, we fix a point $z \in B_{r/8}(0)$. (4.8) with $n = 1$ allows us to pick a point $p_1 \in B_{1/2}(0)$ such that $z - Tp_1 \in B_{r/16}(0)$. Similarly, we pick $p_k \in B_{1/2^k}(0)$ inductively such that

$$\left(z - \sum_{i=1}^{k} Tp_i\right) \in B_{r/2^{k+3}}(0),$$

applying (4.8) with $n = k$. The end result is a sequence $(p_k)_{k=1}^{\infty}$ in the Banach space X with

$$\sum_{k=1}^{\infty} \|p_k\| < \sum_{k=1}^{\infty} \frac{1}{2^k} = 1,$$

whence the following lemma implies that $\sum p_k$ must converge to a point $p \in B_1(0)$:

LEMMA 4.10. *If $\sum \|x_n\| < \infty$, then $\sum x_n$ convergences.*

[2]This is *not* the same thing as $T(B_1(0))$ having nonempty interior, which is the desired result.

Proof of lemma. If $\sum_{n=1}^{\infty} \|x_n\| < \infty$, then the tail $\sum_{n=N}^{\infty} \|x_n\|$ can be made as small as desired, whence the partial sums $\left(\sum_{n=1}^{N} x_n\right)_{N=1}^{\infty}$ form a Cauchy sequence:

$$\left\|\sum_{n=1}^{M} x_n - \sum_{n=1}^{N} x_n\right\| = \left\|\sum_{n=N}^{M} x_n\right\| \leq \sum_{n=N}^{M} \|x_n\| \to 0 \text{ as } N, M \to \infty.$$

It follows that the series $\sum x_n$ converges. □

By continuity of T, we have

$$\|z - Tp\| = \lim_{k\to\infty} \left\|\left(z - \sum_{i=1}^{k} Tp_i\right)\right\| = 0,$$

and so $z \in T(B_1(0))$, thus establishing (4.9). This completes the proof. □

Exercise 4.11. The converse of Lemma 4.10 also holds. Indeed, a normed linear space X is complete if and only if $\sum \|x_n\| < \infty$ implies the convergence of $\sum x_n$.

An immediate corollary of the open mapping theorem is that it is not very difficult to upgrade a linear isomorphism between two Banach spaces to an isomorphism of topological vector spaces.

Corollary 4.12. *A bijective bounded linear operator between Banach spaces is an isomorphism of topological vector spaces.*

Proof. It suffices to show that the inverse is continuous, which is equivalent to the openness of the operator in question. □

Since there is always a *bounded* linear isomorphism on a vector space onto itself, it follows that all complete norms on a single vector space must produce the same topology.

Corollary 4.13. *Let $\|\cdot\|$ and $\|\cdot\|'$ be two complete norms on a vector space X. If there exists a constant K_2 such that $\|\cdot\|' \leq K_2\|\cdot\|$, then there exists a constant K_1 such that $K_1\|\cdot\| \leq \|\cdot\|'$, rendering two norms equivalent.*

Proof. The identity operator I from $(X, \|\cdot\|)$ to $(X, \|\cdot\|')$ is a bijective bounded linear operator, hence an isomorphism of topological vector spaces by the previous corollary. In particular, it is a homeomorphism, which is to say that the two norms are equivalent. □

4.3. Closed Graph Theorem. Let us move onto the second result, the *closed graph theorem*, which concerns the *graph*

$$\Gamma(T) = \{(x, Tx) \in X \times Y : x \in X\}$$

of a linear operator $T : X \to Y$. We note that $\Gamma(T)$ is a linear subspace of the vector space $X \times Y$.

Exercise 4.14. $X \times Y$ with componentwise scalar multiplication and vector addition is a vector space, isomorphic to the (external) direct sum $X \oplus Y$.

In particular, if X and Y are normed linear spaces, then $\Gamma(T)$ inherits a norm from the direct sum $X \times Y$, as per the following exercise.

EXERCISE 4.15. If X and Y are normed linear spaces, then $\|\cdot\|_{X\times Y}$ given by

$$\|(x,y)\|_{X\times Y} = \|x\|_X + \|y\|_Y$$

is a norm on $X \times Y$ that generates the product topology. We refer the reader to Section 8 for a definition of the product topology.

We also observe that $\Gamma(T)$ is closed in $X \times Y$ if and only if $x_n \to x$ in X and $T(x_n) \to y$ in Y implies that $y = Tx$. While this is a consequence of the continuity of T, it is *a priori* a weaker condition than continuity.

EXERCISE 4.16 ([Tao10], Example 1.7.18). The operator on the space c_0 of sequences vanishing at infinity that maps (a_n) to (na_n) is unbounded, but is graph is closed.

The next theorem shows that it is equivalent to continuity, provided that the spaces in question are complete.

THEOREM 4.17 (Closed graph theorem). *A linear operator $T : X \to Y$ between Banach spaces X and Y is bounded if and only if $\Gamma(T)$ is closed.*

PROOF. We have already established the "only if" part of the theorem. To show the "if" part, we assume that $\Gamma(T)$ is closed. Since X and Y are complete, $X\times Y$ is also complete, whence the closed linear subspace $\Gamma(T)$ of $X\times Y$ is complete as well.

We now consider the projection map $\pi : \Gamma(T) \to X$ given by $\pi(x, Tx) = x$ for each $x \in X$. π is a bijective bounded linear operator, and so Corollary 4.12 implies that π is an isomorphism of topological vector spaces. In particular, π^{-1} is a bounded linear operator, hence we can find a positive constant C such that

$$\|x\| + \|Tx\| \leq C\|x\| \tag{4.18}$$

for all $x \in X$.

Now, excluding the trivial case that T is the zero operator, we see that there is at least one x such that $\|Tx\| > 0$. Therefore, the constant C in (4.18) is strictly larger than 1, so we can rewrite (4.18) as

$$\|Tx\| \leq (C-1)\|x\|.$$

It follows that T is bounded, as was to be shown. □

EXERCISE 4.19 ([Tao10], Theorem 1.7.19). A linear operator $T : X \to Y$ between Banach spaces X and Y is bounded if and only if there exists a Hausdorff topology $\mathscr{T}$ on Y that contains strictly fewer open sets than the norm topology on Y such that $T : X \to (Y, \mathscr{T})$ is still continuous. We refer the reader to Section 8 for a detailed discussion on *weak topologies*.

A nice application of the closed graph theorem is the following theorem of Alexandre Grothendieck on closed subspaces of Lebesgue spaces. See Chapter 4, Theorem 4.2 in [SS11] for a proof, which makes use of the theory of Hilbert spaces we shall develop in Chapter 2, Section 1.

THEOREM 4.20 (Grothendieck). *If (X, μ) is a finite measure space, $1 \leq p < \infty$, E a closed subspace of $L^p(X, \mu)$, and $E \subseteq L^\infty(X, \mu)$, then E is finite-dimensional.*

4.4. Uniform Boundedness Principle. The third and final theorem provides a convenient method of upgrading pointwise bounds to uniform bounds.

THEOREM 4.21 (Banach-Steinhaus, the uniform boundedness principle). *Let Λ be a nonempty index set, and, for each $\alpha \in \Lambda$, we let $T_\alpha : X \to Y$ be a bouned linear operator between two normed linear spaces X and Y. We set*

$$E = \left\{ x \in X : \sup_{\alpha \in \Lambda} \|T_\alpha x\| < \infty \right\}.$$

(1) If E is of the second category in X, then

$$\sup_{\alpha \in \Lambda} \|T_\alpha\| < \infty.$$

(2) If $E = X$ and X is a Banach space, then

$$\sup_{\alpha \in \Lambda} \|T_\alpha\| < \infty.$$

PROOF. (2) is a trivial consequence of the Baire category theorem (4.3), and so it suffices to establish (1). To this end, we let

$$E_n = \left\{ x \in X : \sup_{\alpha \in \Lambda} \|T_\alpha x\| \leq n \right\},$$

so that $E = \bigcup_n E_n$. Note that each E_n is closed, for

$$E_n = \bigcap_{\alpha \in \Lambda} \{x \in X : \|T_\alpha x\| \leq n\}.$$

E is of the second category, and so we can find some $n_0 \in \mathbb{N}$ such that E_{n_0} has nonempty interior. Therefore, we can find $x_0 \in X$ and $r > 0$ such that $B_r(x_0) \subseteq E_{n_0}$, whence $\|T_\alpha x\| \leq n_0$ for all $\alpha \in \Lambda$ whenever $\|x - x_0\| < r$. We now see that each $\|y\| < r$ satisfies the estimate

$$\|T_\alpha y\| \leq \|T_\alpha(y + x_0)\| + \|T(-x_0)\| = \|T_\alpha(y + x_0)\| + \|T(x_0)\| \leq 2n_0$$

for all $\alpha \in \Lambda$. Therefore, if $\|x\| \leq 1$, then we have

$$\|T_\alpha x\| = \|r^{-1} T_\alpha(rx)\| \leq \frac{2n_0}{r}$$

for all $\alpha \in \Lambda$, and so $\|T_\alpha\|$ is uniformly bounded. ☐

As an application of the uniform boundedness principle, we discuss the Fourier series. Recall that the *Fourier series* of a complex-valued L^1-function f on the circle group $\mathbb{T} \cong \mathbb{R}/\mathbb{Z}$ is given by

$$\sum_{n \in \mathbb{Z}} \hat{f}(n) e^{2\pi i n x}, \tag{4.22}$$

where

$$\hat{f}(n) = \int_0^1 f(t) e^{-2\pi i n t} \, dt$$

for each $n \in \mathbb{Z}$. Here we are considering f as a 1-periodic function on $\mathbb{R}$.

It is a classical result that (4.22) converges to f in the $L^2(\mathbb{T})$-norm if $f \in L^2(\mathbb{T})$. In other words,

$$\lim_{N \to \infty} \left\| f(x) - \sum_{n=-N}^{N} \hat{f}(n) e^{2\pi i n x} \right\|_2 = 0.$$

This, of course, does not imply that the series converges pointwise to f. Investigating pointwise convergence of functions in a large space like L^1 or L^2 is difficult, so we opt to study a smaller space: the space $\mathscr{C}(\mathbb{T})$ of continuous complex-valued functions on $\mathbb{T}$. As it turns out, however, we can find, for each $x_0 \in \mathbb{T}$, a function $f \in \mathscr{C}(\mathbb{T})$, whose Fourier series diverges at x_0.

To see this, we assume without loss of generality that $x_0 = 0$. Observe that

$$\begin{aligned}(S_N f)(x) &= \sum_{n=-N}^{N} \hat{f}(n) e^{2\pi i n x} \\ &= \int_{\mathbb{T}} f(t) \left(\sum_{n=-N}^{N} e^{2\pi i n (x-t)} \right) dt \\ &= \int_{\mathbb{T}} f(t) D_N(x-t)\, dt,\end{aligned}$$

where D_N is the *nth Dirichlet kernel*

$$D_N(x) = \sum_{n=-N}^{N} e^{2\pi i n x}.$$

Exercise 4.23. It is not very hard to show that

$$D_N(x) = \frac{\sin \pi(2N+1)x}{\sin \pi x}.$$

Show also that there is a fixed constant $c > 0$ such that $\|D_N\|_1 \geq c \log N$ for all $N \in \mathbb{N}$.

We define a linear functional $l_{N,x} : \mathscr{C}(\mathbb{T}) \to \mathbb{C}$ by setting

$$l_{N,x}(f) = (S_N f)(x).$$

Observe that

$$|l_{N,0}(f)| = \left| \int f(t) D_N(t)\, dt \right| \leq \int |f(t)||D_N(t)|\, dt \leq \|f\|_\infty \|D_N\|_1.$$

This is, in fact, an equality:

Exercise 4.24. Show that $\|l_{N,x}\| = \|D_N\|_1$ for all $N \in \mathbb{N}$ and $x \in \mathbb{T}$.

Combining the two exercises above, we can deduce the lower bonud $\|l_{N,0}\| \geq c \log N$, whence $\sup_N \|l_{N,0}\| = \infty$. Since $\mathscr{C}(\mathbb{T})$ is complete, it follows from the uniform boundedness principle (Theorem 4.21) that $(l_{N,0})_{N=1}^\infty$ is not bounded pointwise. In other words, we can find a function $f \in \mathscr{C}(\mathbb{T})$ such that $l_{N,0}(f) = (S_N f)(0)$ does not converge to $f(0)$.

Exercise 4.25. The end result can be strengthened as follows: if A is a countable subset of $\mathbb{T}$, then there exists an $f \in \mathscr{C}(\mathbb{T})$ such that $(S_N f(x))_{N=1}^\infty$ fails to converge for all $x \in A$.

Hint: Show first aht if X is a Banach space, Y a normed linear space, and $(T_\alpha)_\alpha$ a collection of bounded linear maps from X to Y, then

$$M = \{x \in X : \sup_\alpha \|T_\alpha x\| = \infty\}$$

is either empty or a dense G_δ set in X.

Is it possible to find a larger set on which the Fourier series of a continuous function diverges? A result of Lennart Carleson states that the Fourier series of an L^2-function must converge almost everywhere to the function. This, in particular, implies that the Fourier series of a continuous function cannot diverge on too large of a set. See Volume 2, Chapter 7 of [**MS13**], Chapter 11 of [**Gra10b**], or Chapter 7 of [**Thi06**] for modern expositions on the theorem and the techniques used in the proof.

5. Duality: The Hahn-Banach theorems

Let us now recall from Corollary 3.10 that the (topological) dual space X^* of continuous linear functionals on a normed linear space X over $\mathbb{F}$ is a Banach space. The discussion preceding the corollary establishes that the algebraic dual space X' always has many elements, but it is not clear whether there exists a nontrivial element in X^*. The goal of this section is to show that X^* has *plenty* of elements in the following sense: for each nonzero element $x \in X$, we can find $l \in X^*$ such that $\|l\| = 1$ and $l(x) = \|x\|_X$. It then follows immediately that X^* *separates the points of* X, viz., every pair of distinct points $x_1, x_2 \in X$ admit $l \in X^*$ such that $l(x_1) \neq l(x_2)$.

Before we prove the result for the general case, we discuss some examples.

Example. If X is finite-dimensional, then $X^* = X'$, and so X^* is nontrivial.

Example. If $X = L^p(\Omega, \mu)$ with $1 \leq p \leq \infty$, then the linear functional $l_g : L^p(\Omega, \mu) \to \mathbb{C}$

$$l_g(f) = \int fg \, d\mu$$

is shown to be bounded via Hölder's inequality, provided that $g \in L^q(\Omega, \mu)$ with $1/p + 1/q = 1$:

$$\|l_g(f)\| \leq \|g\|_q \|f\|_p.$$

Therefore, $L^q(\Omega, \mu)$ can be isometrically embedded into $(L^p(\Omega, \mu))^*$, and $(L^p)^*$ is nontrivial. In fact, the Riesz representation theorem for Lebesgue spaces shows that the isometric embedding is an isomorphism if $1 < p < \infty$, but we will not prove this fact in this course. This result extends to $p = 1$ if (Ω, μ) is, for example, σ-finite.

Example. In general, $(L^\infty)^*$ is *not* L^1, even if the measure space is σ-finite. For example, if (Ω, μ) is $\mathbb{N}$ with the counting measure, then the space $\mathscr{C}_0(\mathbb{N})$ of sequences that converge to zero is a proper subspace of the space $l^\infty(\mathbb{N})$ of bounded sequences. We shall show that $(\mathscr{C}_0(\mathbb{N}))^* \cong l^1(\mathbb{N})$.

The above example shows that every $a = (a_n) \in l^1$ can be identified with a bounded linear functional on l^∞. Therefore, a can also be identified with a bounded linear functional on $\mathscr{C}_0$: indeed, if $a \in l^1$ and $x \in \mathscr{C}_0$, then

$$\left| \sum_n a_n x_n \right| \leq \|a\|_1 \|x\|_\infty$$

by Hölder's inequality. Therefore, there is an isometric embedding of l^1 into $\mathscr{C}_0$.

Conversely, we fix $l \in (\mathscr{C}_0)^*$. We shall show that l can be represented by an element of l^1. To this end, we define, for each $k \in \mathbb{N}$, a sequence $e^k = (e^k_n)_{n=1}^\infty$ by setting

$$e_n^k = \begin{cases} 1 & \text{if } k = n; \\ 0 & \text{if } k \neq n. \end{cases}$$

Then, for each $x = (x_k)_{k=1}^\infty \in \mathscr{C}_0$, we see that

$$l(x) = l\left(\sum_{k=1}^\infty x_k e^k\right) = \sum_{k=1}^\infty x_k l(e^k).$$

But then

$$|x_k l(e^k)| \leq |x_k||l(e^k)| \leq \|x\|_\infty |l(e^k)|,$$

and so

$$|l(x)| \leq \|x\|_\infty \|l(e^k)\|_1.$$

It follows that l^1 is isometrically isomorphic to $(\mathscr{C}_0)^*$. □

Let us now turn to the business of stating the theorem precisely. Here we work with a generalization of a norm.

DEFINITION 5.1. A *seminorm* on a vector space X over $\mathbb{F}$ is a map $\rho : X \to [0, \infty)$ such that $\rho(\lambda x) = |\lambda|\rho(x)$ and $\rho(x + y) \leq \rho(x) + \rho(y)$ for all $\lambda \in \mathbb{F}$ and $x, y \in X$.

We remark that the L^p-norm is *a priori* a seminorm and only becomes a norm once the almost-everywhere equality formalism is introduced. See Section 9 for other examples of seminorms that arise naturally in analysis.

THEOREM 5.2 (Complex Hahn-Banach). *Let X be a vector space over $\mathbb{C}$. Let Y be a linear subspace of X, l a $\mathbb{C}$-linear functional on Y, and ρ a seminorm on X that dominates l on Y, viz., $|l(y)| \leq \rho(y)$ for all $y \in Y$. We can then find a $\mathbb{C}$-linear functional L on X that extends l and is dominated by ρ on X.*

The complex Hahn-Banach theorem will be derived from an analogous version of the theorem over $\mathbb{R}$. We now that we can relax the hypothesis in the real case to allow for sublinear functionals.

DEFINITION 5.3. A *sublinear functional* on a vector space X is a map $\rho : X \to [0, \infty)$ such that $\rho(\lambda x) = \lambda\rho(x)$ and $\rho(x + y) \leq \rho(x) + \rho(y)$ for all $\lambda \geq 0$ and $x, y \in X$.

THEOREM 5.4 (Real Hahn-Banach). *Let X be a vector space over $\mathbb{R}$. Let Y be a linear subspace of X, l an $\mathbb{R}$-linear functional on Y, and ρ a sublinear functional on X that dominates l on Y, viz., $l(y) \leq \rho(y)$ for all $y \in Y$. We can then find an $\mathbb{R}$-linear functional L on X that extends l and is dominated by ρ on X.*

PROOF OF $\mathbb{R}$-HB $\Rightarrow$ $\mathbb{C}$-HB. We first recall that a $\mathbb{C}$-linear functional l on X can be written as the sum of two $\mathbb{R}$-linear functionals $l_1 = \operatorname{Re} l$ and $il_2 = i \operatorname{Im} l$. Simple computations show that $l_2(x) = -l_1(ix)$.

We now suppose that the $\mathbb{C}$-linear functional l on a subspace M of X is dominated by a seminorm ρ on X. This, in particular, implies that $l_1(x) \leq \rho(x)$ for all $x \in M$, whence the real Hahn-Banach theorem implies that we can extend l_1 to an $\mathbb{R}$-linear functional L_1 on X such that $L_1(x) \leq \rho(x)$ for all $x \in X$. Note that

$$-L_1(x) = L_1(-x) \leq \rho(-x) = \rho(x),$$

so that $|L_1(x)| \leq \rho(x)$ for all $x \in X$.

We now set $L(x) = L_1(x) - iL_1(ix)$, which is an extension of l onto X. Pick $\theta(x)$ such that
$$L(x) = e^{i\theta(x)}|L(x)|,$$
and observe that
$$|L(x)| = e^{-i\theta(x)}e^{i\theta(x)}|L(x)| = e^{-i\theta(x)}L(x) = L(e^{-i\theta(x)}x).$$
As $|L(x)| \geq 0$, we see that $|L(x)| = L_1(e^{-i\theta(x)}x) = |L_1(e^{-i\theta(x)}x)|$. It follows that
$$|L(x)| = |L_1(e^{-i\theta(x)}x)| \leq \rho(e^{-i\theta(x)}x) = |e^{-i\theta(x)}|\rho(x) = \rho(x),$$
as was to be shown. □

Proof of $\mathbb{R}$-HB. We shall first demonstrate how to extend a linear functional by one dimension and then use Zorn's lemma. We assume for now that M is a proper subspace of an $\mathbb{R}$-vector space X, and l a linear functional on M dominated by a sublinear functional ρ on X. We fix $x_0 \in X \smallsetminus M$ and extend l onto $M \oplus \mathbb{R}x_0$.

Let us construction an extension $L : M \oplus \mathbb{R}x_0 \to \mathbb{R}$ of l that is still dominated by ρ. We shall define $L(x_0)$ such that $L(x_0) \leq \rho(x_0)$, and extend onto $M \oplus \mathbb{R}x_0$ by linearity:
$$L(x + \lambda x_0) = l(x) + \lambda L(x_0).$$
L is clearly linear. To check that $L(x + \lambda x_0) \leq \rho(x + \lambda x_0)$, it suffices to show that
$$L(x + x_0) \leq \rho(x + x_0) \quad \text{and} \quad L(x - x_0) \leq \rho(x - x_0). \tag{5.5}$$
Indeed, (5.5) implies that
$$L(x + \lambda x_0) = \lambda \left[L(\lambda^{-1}x + x_0)\right] \leq \lambda\rho(\lambda^{-1}x + x_0) = \rho(x + \lambda x_0)$$
if $\lambda > 0$, and
$$L(x + \lambda x_0) = -\lambda \left[L(-\lambda^{-1}x - x_0)\right] \leq -\lambda\rho(-\lambda^{-1}x - x_0) = \rho(x + \lambda x_0)$$
if $\lambda < 0$.

We now note that (5.5) is equivalent to
$$L(x_1) - \rho(x_1 - x_0) \leq L(x_0) \leq \rho(x_2 + x_0) - L(x_2)$$
for all $x_1, x_2 \in M$. It thus suffices to define $L(x_0)$ to safisfy this inequality, which is possible if and only if
$$\sup_{x_1 \in M} \left[l(x_1) - \rho(x_1 - x_0)\right] \leq \inf_{x_2 \in M} \left[\rho(x_2 + x_0) - l(x_2)\right]. \tag{5.6}$$
To see this, we observe that
$$l(x_1) + l(x_2) = l(x_1 - x_0) + l(x_2 + x_0) \leq \rho(x_2 + x_0) + \rho(x_1 - x_0),$$
for all $x_1, x_2 \in M$, so that
$$l(x_1) - \rho(x_1 - x_0) \leq \rho(x_2 + x_0) - l(x_2).$$
(5.6) now follows by taking the supremum on the left-hand side and the infimum on the right-hand side. We have therefore successfully extended our linear functional by one dimension.

We now turn to the general case. We let $\mathcal{P}$ be a collection of pairs $(\tilde{M}, \tilde{l})$ consisting of a subspace $\tilde{M}$ of X that contains M and a linear functional $\tilde{l}$ on $\tilde{M}$ that extends l such that $\tilde{l}$ dominates l on M. We define a partial order on $\mathcal{P}$ by setting $(\tilde{M}_1, \tilde{l}_1) \prec (\tilde{M}_2, \tilde{l}_2)$ if and only if $\tilde{M}_1 \subseteq \tilde{M}_2$ and $\tilde{l}_2|_{\tilde{M}_1} = \tilde{l}_1$. If $\{(\tilde{M}_\alpha, \tilde{l}_\alpha)\}_\alpha$ is a chain in $\mathcal{P}$, then the upper bound of the chain is given by $(\tilde{M}, \tilde{l})$, where $\tilde{M} = \bigcup_\alpha \tilde{M}_\alpha$ and

$\tilde{l}$ is defined to agree with $\tilde{l}_\alpha$ on $\tilde{M}_\alpha$. Zorn's lemma therefore furnishes a maximal element (L, X) of $\mathcal{P}$. This L is unique and is the desired extension of l onto X. □

COROLLARY 5.7. *Let X be a normed linear space over $\mathbb{F}$. For each $x_0 \in X \smallsetminus \{0\}$, we can find $l \in X^*$ such that $\|l\| = 1$ and $l(x) = \|x_0\|$.*

PROOF. We fix $x_0 \in X \smallsetminus \{0\}$ and let $Y = \mathbb{F}x_0$. Define a linear functional $l : Y \to \mathbb{F}$ by $l(\lambda x_0) = \lambda\|x_0\|$ for each $\lambda \in \mathbb{F}$, so that $l(x_0) = \|x_0\|$. Furthermore, each $y \in Y$ admits a scalar $\lambda \in \mathbb{F}$ such that $y = \lambda x_0$, and so

$$|l(y)| = |l(\lambda x_0)| = |\lambda||l(x_0)| = |\lambda|\|x_0\| = \|\lambda x_0\| = \|y\|.$$

We now define a seminorm ρ on X by setting $\rho(x) = \|x\|$, so that $|l(y)| = \rho(y)$ for each $y \in Y$. By the complex Hahn-Banach theorem (5.2), we can extend l to a linear functional $L : X \to \mathbb{C}$ that is dominated by ρ on X, whence $L \in X^*$. In particular, we have

$$|Lx| \leq \rho(x) = \|x\|$$

for all $X \in X$, so we have $\|L\| \leq 1$. It now suffices to note that $|L(x_0)| = |l(x_0)| = \|x_0\|$, whence $\|L\| = 1$. □

By the definition of boundedness, we have

$$|l(x)| \leq \|l\|_{X^*}\|x\|_X$$

for each $l \in X^*$ and every $x \in X$. Instead of considering l as a function that takes x, we now consider x as a function taking l: more precisely, we define a linear functional $\hat{x}$ on X^* by setting $\hat{x}(l) = l(x)$. Then we can rewrite the above inequality to see that

$$|\hat{x}(l)| \leq \|x\|_X\|l\|_{X^*},$$

whence $\|\hat{x}\|_{X^{**}} \leq \|x\|_X$. It follows that $\hat{x}$ is an element of the *double dual* $X^{**} = (X^*)^*$ of X. Furthermore, if $x \neq 0$, then Corollary 5.7 furnishes $l \in X^*$ such that $\|l\| = 1$ and

$$\hat{x}(l) = l(x) = \|x\|,$$

and so $\|\hat{x}\|_{X^{**}} \geq \|x\|_X$. We conclude that the map $x \mapsto \hat{x}$ is an isometric embedding from X into X^{**}.

Since Corollary 3.10 implies that X^{**} is always complete regardless of the completeness of X, the isometric embedding given above is not an isomorphism in general. Even when X is a Banach space, X^{**} is not always isometrically isomorphic to X: see the $\mathscr{C}_0$ example above. The cases in which X^{**} *is* isometrically isomorphic form an important subclass of the class of Banach spaces, and we give them a name.

DEFINITION 5.8. A Banach space X is *reflexive* if the canonical embedding $x \mapsto \hat{x}$ into X^{**} is an isometric isomorphism[3].

We remark that the canonical embedding $x \mapsto \hat{x}$ provides a way to construct the *completion* of an incomplete normed linear space that we alluded to in Theroem 2.10. Indeed, if we let Y be the image of the canonical embedding $x \mapsto \hat{x}$, then $\overline{Y}$ is a complete subspace of X that includes an isometrically isomorphic copy Y of X.

We now establish two restatements of the Hahn-Banach theorem, specialized for particular purposes. We shall consider some applications of each restatement as well.

[3]This isomorphism is *natural*, in the sense that the construction of the isomorphism does not require a choice of basis: see Appendix A, Section 1 for a discussion.

THEOREM 5.9 (Analytic Hahn-Banach). *Let X be a normed linear space, M a subspace of X, and l a bounded linear functional on M. Then there exists a bounded extension L of l onto the whole space X such that $\|l\|_{M^*} = \|L\|_{X^*}$.*

PROOF. Let $\rho(x) = \|l\|_{M^*}\|x\|$, so that ρ is a norm on X. Since $|l(x)| \leq \rho(x)$ for all $x \in M$, the complex Hahn-Banach theorem (Theorem 5.2) furnishes an extension L of l onto the whole space X that is still dominated by $\rho(x)$. Therefore, $\|L\|_{X^*} \leq \|l\|_{M^*}$. The reverse inequality is trivially established. ☐

COROLLARY 5.10 (Existence of nontrivial annihilators). *Let X be a normed linear space over $\mathbb{F}$, M a subspace of X, and $x_0 \in X$ such that $d(x_0, M) > 0$. Then there exists $L \in X^*$ such that $L|_M = 0$, $L(x_0) = 1$, and $\|L\|_{X^*} = 1/d(x_0, M)$.*

See Appendix A, Section 7 for a discussion on the annihilators of subsets of a topological vector space.

PROOF OF COROLLARY. Let $M_1 = M \oplus \mathbb{F}x_0$ and define a linear functional l on M_1 by setting $l(x + \lambda x_0) = \lambda$. Then $l = 0$ on M, and

$$\begin{aligned}
\|l\|_{M_1^*} &= \sup_{\substack{\lambda \neq 0 \\ x \in M}} \frac{|l(x+\lambda x_0)|}{\|x + \lambda x_0\|} \\
&= \sup_{\substack{\lambda \neq 0 \\ x \in M}} \frac{|\lambda|}{|\lambda|\|\lambda^{-1}x + x_0\|} \\
&= \frac{1}{\inf_{x,\lambda} \|\lambda^{-1}x + x_0\|} \\
&= \frac{1}{\inf_{u \in M} \|u + x_0\|} \\
&= \frac{1}{d(x_0, M)}.
\end{aligned}$$

The desired bounded linear functional L is now obtained as an extension of l via the analytic Hahn-Banach theorem (Theorem 5.9). ☐

COROLLARY 5.11. *If the dual space X^* of a Banach space X is separable, then so is X.*

EXERCISE 5.12. $l^p(\mathbb{N})$ is a separable Banach space if and only if $1 \leq p < \infty$. In tandem with the above corollary, we conclude that $(l^\infty(\mathbb{N}))^* \neq l^1(\mathbb{N})$.

PROOF OF COROLLARY. Let $\{l_n\}_{n \in \mathbb{N}}$ be a countable dense subset of X^*. We pick a sequence $(x_n)_{n=1}^\infty$ in X such that $\|x_n\| = 1$ and $|l_n(x_n)| \geq \frac{1}{2}\|l_n\|_{X^*}$ for all $n \in \mathbb{N}$. If we let M to be the span of $\{x_n\}_n$ and $\mathcal{L}$ the space of finite linear combinations of the elements of $\{x_n\}_n$ with rational coefficients, then $\mathcal{L}$ is a countable dense subset of M.

It therefore suffices to show that M is dense in X. We suppose for a contradiction that M is not dense in X, and pick $x \in X$ such that $d(x, M) > 0$. By Corollary 5.10, we can find a nonzero $L \in X^*$ that vanishes on M. Now, there exists a subsequence $(l_{n_k})_{k=1}^\infty$ of $(l_n)_{n=1}^\infty$ that converges to L in X^*. But

$$\|L - l_{n_k}\|_{X^*} \geq |(L - l_{n_k})(x_{n_k})| = |l_{n_k}(x_{n_k})| \geq \frac{1}{2}\|l_{n_k}\|_{X^*},$$

so that $\|l_{n_k}\|_{X*} \to 0$ as $n_k \to \infty$. This evidently contradicts the fact that L is nonzero, whence we must conclude that M is a dense subset of X. It follows that L is a dense subset of X, and X is separable. □

We conclude the section by introducing the notion of *adjoints*, which, in a sense, is a generalization of the complex conjugate operation on matrices.

Definition 5.13. The *(algebraic) adjoint* of a linear operator $T : X \to Y$ between two vector spaces X and Y is a linear operator $T' : Y' \to X'$ defined to be

$$T'y' = y'T$$

for each $y' \in Y'$.

If l is a linear functional on a vector space X, we often denote $l(x)$ by (l, x). Then the above definition states that $(T'y', x) = (y', Tx)$ for all $x \in X$ and $y' \in Y'$.

Definition 5.14. The *(topological) adjoint* of a bounded linear operator $T : X \to Y$ between two normed linear spaces X and Y is a linear operator $T^* : Y^* \to X^*$ defined to be

$$T^*y^* = y^*T$$

for all $y^* \in Y^*$.

We remark that T^*y^* is indeed an element of X^*, for it is a composition of two continuous maps. Furthermore, the adjoint operation preserves the operator norm.

Proposition 5.15. $\|T^*\| = \|T\|$

Proof. Observe that

$$\|T^*\| = \sup_{\|y^*\|\le 1} \|T^*y^*\|_{X^*} = \sup_{\|y^*\|\le 1} \sup_{\|x\|\le 1} |T^*y^*(x)| = \sup_{\|y^*\|\le 1} \sup_{\|x\|\le 1} |y^*(Tx)|.$$

By the existence of nontrivial annihilators (Corollary 5.10), there exists a bounded linear functional y^* such that $\|y^*\| = 1$ and $y^*(Tx) = \|Tx\|$. Therefore,

$$\|T^*\| = \sup_{\|y^*\|\le 1} \sup_{\|x\|\le 1} |y^*(Tx)| \ge \sup_{\|x\|\le 1} \|Tx\| = \|T\|.$$

Conversely,

$$\|T^*l\| = \|lT\| = \sup_{\|x\|\le 1} \|lTx\| \le \sup_{\|x\|\le 1} \|l\|\|Tx\| = \|l\|\|T\| = \|T\|\|l\|$$

for all $l \in Y^*$, and so $\|T^*\| \le \|T\|$. □

Also, bounded operators and their adjoints possess an injectivity-surjectivity duality relation.

Proposition 5.16. *Let X and Y be normed linear spaces. The adjoint T^* of $T \in \mathscr{B}(X, Y)$ is injective if and only if $\operatorname{im} T$ is dense in Y.*

Proof. Before we proceed, we observe that the injectivity of T^* is equivalent to the statement that $l_1 \circ T = l_2 \circ T$ implies $l_1 = l_2$ for all $l_1, l_2 \in Y^*$. This is a direct consequence of the definition of T^*.

($\Leftarrow$) We suppose that $\operatorname{im} T$ is dense in Y. If $l_1, l_2 \in Y^*$ satisfy $l_1 \circ T = l_2 \circ T$, then $l_1|_{\operatorname{im} T} = l_2|_{\operatorname{im} T}$, and so l_1 and l_2 are continuous functions that agree on a dense set. It follows that $l_1 = l_2$, and so T^* is injective.

($\Rightarrow$) Conversely, suppose that T^* is injective and assume for a contradiction that $\operatorname{im} T$ is not dense. Then $\overline{\operatorname{im} T}$ is a closed proper linear subspace of Y, whence the existence of nontrivial annihilators (Corollary 5.10) implies that there are $l_1, l_2 \in Y^*$ with $l_1|_{\operatorname{im} T} = l_2|_{\operatorname{im} T} = 0$ but $l_1 \neq l_2$. This contardicts the injectivity of T^*. $\square$

The adjoint operation is a direct generalization of the conjugate transpose of square matrices: see Chapter 2, Definition 2.6.

6. Convexity, Part 1: the Geometric Hahn-Banach Theorems

The next restatement of the Hahn-Banach theorem tells us why the proper functional-analytic context for the Hahn-Banach theorem is *convexity*. We begin with a few definitions.

Definition 6.1. Let X be a vector space over $\mathbb{R}$, and S a subset of X. A point $x_0 \in X$ is an *internal point* of S if, for each $u \in X$, there exists an $\varepsilon > 0$ such that $|t| < \varepsilon$ implies $x_0 + tu \in S$.

We remark that all internal points of S are points in S. Furthermore, if X is a normed linear space, then every interior point of S is an internal point of S. The converse is true in $\mathbb{R}$, but even in $\mathbb{R}^2$ there are readily available counterexamples:

Example. Let

$$S = \{(r\cos\theta, r\sin\theta) : \theta - \pi \leq r \leq \pi - \theta,\ 0 \leq \theta < \pi\},$$

so that 0 is an internal point, but $0 \notin S^\circ$.

In general, if 0 is an internal point of a subset S of a real vector space X, then, for all $x \in X$, there exists a $\lambda_0 > 0$ such that $\lambda > \lambda_0$ implies $x \in \lambda S$. This motivates us to define the following:

Definition 6.2. The *Minkowski functional*, or the *gauge*, of a subset S of a real vector space X is given by the formula

$$\rho_S(x) = \inf\{\lambda > 0 : x \in \lambda S\}.$$

Let us note some basic properties of ρ_S:

Proposition 6.3. *The Minkowski functional ρ_S satisfies the following properties:*

(a) ρ_S is positive homogeneous, i.e., $\rho_S(\lambda x) = \lambda \rho_S(x)$ for all $\lambda \geq 0$ and $x \in S$.
(b) $\rho_S(x) \leq 1$ for all $x \in S$.
(c) If x is an internal point of S, then $\rho_S(x) < 1$.

Proof. (a) is trivial. To show (b), it suffices to note that $x \in 1S$ whenever $x \in S$. (c) follows from the fact that there exists a $\delta > 0$ such that $x + \delta x \in S$, so that $x \in (1+\delta)^{-1}S$. $\square$

We now prove that the implication in (c) can be reversed if the underlying set is convex.

Proposition 6.4. *If K is a convex subset of a real vector space X containing 0 as an internal point, then x is an internal point of K if and only if $\rho_K(x) < 1$.*

Proof. It is enough to prove the "if" part. Suppose that $\rho_K(x) < 1$, and find $\delta > 0$ such that $x \in (1-\delta)K$. Since 0 is an internal point of K, each $u \in X$ admits an $\varepsilon > 0$ such that $|t| < \varepsilon$ implies $tu \in K$. We let $\varepsilon' = \varepsilon\delta$ and $t' = \delta t$. If $|t'| < \varepsilon'$, then $|t| < \varepsilon$, so that $tu \in K$. In particular, $t'u \in \delta K$. It now suffices to observe that $x \in (1-\delta)K$ and $t'u \in \delta K$ implies

$$x + t'u \in (1-\delta)K + \delta K = K,$$

as was to be shown. □

But what do Minkowski functionals have to do with the Hahn-Banach theorem? As it turns out, a Minkowski functional is a sublinear functional if the underlying set is convex.

Proposition 6.5. *If K is a convex subset of a real vector space X containing 0 as an internal point, then ρ_K is a sublinear functional on X.*

Proof. It suffices to check subadditivity of ρ_K. Let $x \in \lambda K$ and $y \in \mu K$, so that $\rho_K(x) \leq \lambda$ and $\rho_K(y) \leq \mu$. We now consider the representation

$$\frac{x+y}{\lambda+\mu} = \frac{\lambda}{\lambda+\mu}\frac{x}{\lambda} + \frac{\mu}{\lambda+\mu}\frac{y}{\mu}.$$

Since x/λ and y/μ are in K, we see that $(x+y)/(\lambda+\mu)$ is in K, and so $x+y \in (\lambda+\mu)K$. It follows that $\rho_K(x+y) \leq \lambda+\mu$. It now follows that

$$\rho_K(x+y) \leq \inf_\lambda \lambda + \inf_\mu \mu = \rho_K(x) + \rho_K(y),$$

where the infima are taken over all values of λ and μ such that $\rho_K(x) \leq \lambda$ and $\rho_K(y) \leq \mu$. □

We are now in a position to state the first version of the geometric Hahn-Banach theorem.

Theorem 6.6 (Geometric Hahn-Banach for the algebraic dual). *Let X be a real vector space and K_1 and K_2 two nonempty disjoint convex sets in X. If K_1 has an internal point, then there exists a nonzero $L \in X'$ such that $L(x) \leq L(y)$ for all $x \in K_1$ and $y \in K_2$.*

In other words, l separates K_1 and K_2: indeed, there exists a real number α such that

$$\sup_{x \in K_1} l(x) \leq \alpha \leq \inf_{y \in K_2} l(y).$$

Proof. Let x_1 be an internal point of K_1 and x_2 a point of K_2. We note that $K_1 - K_2$ is convex, and that $x_1 - x_2$ is an internal point of $K_1 - K_2$. We let $K = K_1 - K_2 - x_0$, so that 0 is an internal point of K. Since K_1 and K_2 are disjoint, $0 \notin K_1 - K_2$, whence $-x_0 \notin K$. Proposition 6.4 now implies that $\rho_K(-x_0) \geq 1$.

Let $M = x_0\mathbb{R}$ and set $l(\lambda x_0) = -\lambda$ for each $\lambda \in \mathbb{R}$. l is a linear functional on M with $l(-x_0) = 1 \leq \rho_K(-x_0)$. For each $\lambda \geq 0$, linearity of l and positive homogeneity of ρ_k imply that

$$l(-\lambda x_0) \leq \rho_K(-\lambda x_0),$$

which, combined with the non-negativity of ρ_K, yields the estimate

$$l(\lambda x_0) = -\lambda \leq 0 \leq \rho_K(\lambda x_0).$$

Therefore, l is dominated by ρ_K, which, by Proposition 6.5, is sublinear. The real Hahn-Banach theorem (Theorem 5.4) now furnishes an extension $L \in X'$ of l dominated by ρ_K on X.

In particular, Proposition 6.3 (b) implies that the extension L is bounded above by 1 on K. We now fix arbitrary $y_1 \in K_1$ and $y_2 \in K_2$, respectively, and note that $y_1 - y_2 - x_0 \in K$. Therefore,

$$1 \geq L(y_1 - y_2 - x_0) = L(y_1) - L(y_2) + L(-x_0) = L(y_1) - L(y_2) + 1,$$

whence we have

$$L(y_1) \leq L(y_2).$$

We conclude that L is the desired linear functional. □

The second version of the geometric Hahn-Banach theorem can now be deduced as an easy consequence of the first one.

Corollary 6.7 (Geometric Hahn-Banach for the topological dual). *Let X be a real normed linear space and K_1 and K_2 two nonempty disjoint convex sets in X. If K_1 has nonempty interior, then there exists a nonzero $L \in X^*$ such that $L(x) \leq L(y)$ for all $x \in K_1$ and $y \in K_2$.*

Proof. As in the above proof, we construct L on X. Note that $L(x) \leq \rho_K(x)$ and $-L(x) = L(-x) \leq \rho_K(-x)$, so that

$$-\rho_K(-x) \leq L(x) \leq \rho_K(x).$$

Note also that $K_1^\circ \neq \varnothing$ implies $0 \in K^\circ$, whence we can find a ball $B_{\varepsilon_0}(0) \subseteq K$.

We claim that ρ_K is continuous at 0. Indeed, for each $\varepsilon > 0$, we know that $B_\varepsilon(0) \subseteq \frac{\varepsilon}{\varepsilon_0} K$. It follows that $x_n \to 0$ implies $\rho_K(x_n) \to 0$, and the claim is established. In particular, L is continous at 0, hence continuous everywhere. Therefore, $L \in X^*$ □

For both cases, we remark that if all points of K_1 are internal, then all points of K are internal as well. Indeed, $K = \{x : \rho_K(x) < 1\}$, and so $L(x) < 1$ for all $x \in K$. It then follows that $L(y_1) < L(y_2)$ for all $y_1 \in K_1$ and $y_2 \in K_2$. We also remark that the same results hold for $\mathrm{Re}(L)$ if the underlying field is $\mathbb{C}$.

Exercise 6.8. The nonempty interior assumption in Corollary 6.7 is there to guarantee the existence of internal points, which is crucial. Consider, for example, the subsets

$$E_\alpha = \{f \in \mathscr{C}([-1,1]) : f(0) = \alpha\}$$

of space $L^2([-1,1])$, which are dense and convex in L^2. Moreover, $E_\alpha \cap E_\beta = \varnothing$ whenever $\alpha \neq \beta$. Even so, E_α and E_β do not have any interior point, and there is no nonzero continuous linear functional l on L^2 for which

$$\sup_{f \in E_\alpha} l(f) \leq \inf_{f \in E_\beta} l(f).$$

Utilizing the additional geometric structure in Corollary 6.7, we now establish the following geometric result:

Theorem 6.9 (Hyperplane separation theorem). *Let X be a real normed linear space and A and B nonempty, convex, disjoint subsets of X. If A is compact and B is closed, then there exists an $L \in X^*$ and an $\alpha \in \mathbb{R}$ such that*

$$\sup_{x \in A} L(x) < \alpha < \inf_{y \in B} L(y).$$

In other words, the hyperplane $\{z : L(z) = \alpha\}$ separates A and B.

PROOF. Consider the *Hausdorff metric*

$$d_H(A,B) = \inf_{\substack{x \in A \\ y \in B}} \|x - y\|.$$

Since $d_H(A,B) = \inf_{x \in A} d(x,B)$, the minimum is attained on the compact set A. Therefore, $d_H(A,B) > 0$.

Let $\varepsilon = d_H(A,B)/4$. We "beef up" A and B by setting

$$\tilde{A} = A + B_\varepsilon(0) \quad \text{and} \quad \tilde{B} = B + B_\varepsilon(0);$$

note that $\tilde{A}$ and $\tilde{B}$ are disjoint convex subsets of X with nonempty interior.The geometric Hahn-Banach theorem for the topological dual (Corollary 6.7) furnishes an $L \in X^*$ such that

$$\sup_{x \in \tilde{A}} L(x) \leq \alpha \leq \inf_{y \in \tilde{B}} L(y).$$

For each $x \in A$ and $u \in B_1(0)$, we see that $x + \varepsilon u \in \tilde{A}$. This, in turn, implies that

$$\sup_{u \in B_1(0)} L(x + \varepsilon u) = L(x) + \sup_{u \in B_1(0)} L(\varepsilon u) = L(x) + \varepsilon\|L\| \leq \alpha + \varepsilon\|L\|,$$

whence $L(x) < \alpha$ for each $x \in A$. Similarly, $L(y) > \alpha$ for all $y \in B$, and L is the desired linear functional. □

As a corollary, we deduce an alternate characterization of closed convex subsets of normed linear spaces. We shall need the a definition.

DEFINITION 6.10. The *closed half-space* of a real vector space X with respect to $l \in X'$ at $\alpha \in \mathbb{R}$ is

$$H(l, \alpha) = \{x \in X : l(x) \leq \alpha\}.$$

We remark that $H(l, \alpha)$ is a closed subset of X if X is a normed linear space and l is bounded.

COROLLARY 6.11. *Every closed convex set K in a real normed linear space X can be written as the intersection of all closed half-spaces of X that contain K.*

PROOF. Evidently, the intersection C of all closed half-spaces that contain K is a closed superset of K. To establish the reverse inclusion, we suppose for a contradiction that there exists $x_0 \in C \smallsetminus K$ and note that the singleton set $\{x_0\}$ is a compact, convex subset of X. Since K is a closed, convex subset of X disjoint from $\{x_0\}$, the hyperplane separation theorem (Theorem 6.9) applied to $\{x_0\}$ and K furnishes an $L \in X^*$ and an $\alpha \in \mathbb{R}$ such that

$$L(x_0) < \alpha < \inf_{x \in K} L(x).$$

In particular, $K \subseteq H(-L, -\alpha)$, and so $C \subseteq H(-L, -\alpha)$ by the minimality of C. Since $x_0 \in C$, it follows that $-L(x_0) \leq -\alpha$, or $L(x_0) \geq \alpha$, which contradicts the above inequality. □

7. Convexity, Part 2: the Krein-Milman Theorem

With the tools at hand, let us now carry out a detailed study of convex sets. We are, in particular, interested in the "generating sets" of convex sets. To make this notion precise, we make the following definition:

DEFINITION 7.1. The *convex hull* of a subset E of a real vector space X is the smallest convex set $\operatorname{co}(E)$ that contains E, viz., the intersection of all convex sets containing E. Concretely, we can write $\operatorname{co}(E)$ as the collection

$$\left\{\sum a_i v_i \,:\, a_i \in [0,1],\, \sum a_i = 1,\, v_i \in E\right\},$$

of convex combinations of elements of E

We can think of E as a generating set for the convex set $A = \operatorname{co}(E)$. To illustrate the utility of generating sets, we investigate the follwoing problem of maximizing convex functionals.

PROPOSITION 7.2. *Let E be a nonempty, bounded subset of a real vector space X and $A = \overline{\operatorname{co}(E)}$. If f is a continuous convex functional on X, then*

$$\sup_{x \in A} f(x) = \sup_{x \in E} f(x).$$

PROOF. Since f is continuous, the supremum of f over $A = \overline{\operatorname{co}(E)}$ agrees with the supremum over $\operatorname{co}(E)$. Observe that

$$\sup_{x \in \operatorname{co}(E)} f(x) \leq \sup_{\substack{x=\sum a_i v_i \\ \sum a_i = 1 \\ v_i \in E}} \sum a_i f(v_i) \leq \sum a_i \sup_{x \in E} f(x) = \sup_{x \in E} f(x),$$

and so $\sup_{x \in A} f(x) \leq \sup_{x \in E} f(x)$. The reverse inequality is trivial. □

Since $A = \operatorname{co}(E \cup F)$ for any subset $F \subseteq C$, it is quite possible that the generating set at hand contains many extraneous points. It is thus desirable to find a generating set E of a given convex set A that is "minimal" in the following sense: if D is a proper subset of E, then $\operatorname{co}(D) \neq A$. The set of vertices of a triangle is such a set—if any one of them is omitted, then the convex hull of the remaining two points is merely a line.

To make the notion of minimal generating sets precise, we fix two points x and y in a vector space X and consider the map $t \mapsto tx + (1-t)y$. The *closed line segment* $[x, y]$ and the *open line segment* $]x, y[$ from x to y are the images of $[0, 1]$ and $(0, 1)$ under this map, respectively.

DEFINITION 7.3. A subset E of a subset A of a real vector space X is *extreme* if, for each $x, y \in A$, the nontriviality of the intersection $]x, y[\cap E$ implies that $x, y \in E$. A point $x \in A$ is an *extreme point* if $\{x\}$ is an extreme subset of A, and the collection of all such points of A is denoted by $\operatorname{ext}(A)$.

Note that extreme points are precisely the points that cannot be written as a convex combination of distinct points. As a conseqeunce, if E is a proper subset of $\operatorname{ext}(A)$, then $\operatorname{co}(E)$ cannot be A, for none of the points in $\operatorname{ext}(A)$ can be generated by taking a convex combination of points in $A \smallsetminus \operatorname{ext}(A)$.

EXAMPLE. The extreme points of a convex n-gon in $\mathbb{R}^2$ are precisely the vertices.

Example. Let $X = l^p(\mathbb{Z})$ and set $B = \{(a_n) \in X : \|(a_n)\|_p \leq 1\}$. If $p = 1$, then the extreme points of B are the sequence $(e_n^N)_{n=1}^{\infty}$ such that

$$|e_n^N| = \begin{cases} 1 & \text{if } n = N \text{ and} \\ 0 & \text{otherwise.} \end{cases}$$

If $1 < p < \infty$, then $\operatorname{ext}(B) = \{(a_n) \in X : \|(a_n)\|_1 = 1\}$, where as $\operatorname{ext}(B) = \{(a_n) \in X : |a_n| = 1 \text{ for all } n \in \mathbb{N}\}$ if $p = \infty$. This seems like a lot, but the unit ball in the subspace $_0$ of l^∞ has *no* extreme points.

Example. Let $X = L^p([0,1])$ and set $B = \{f \in X : \|f\|_p \leq 1\}$. If $p = 1$, then B has no extreme points, whereas $\operatorname{ext}(B) = \{f \in X : \|f\|_p = 1\}$ if $1 < p < \infty$. If $p = \infty$, then

$$\operatorname{ext}(B) = \{f \in X : |f(x)| = 1 \text{ for almost all } x \in [0,1]\}.$$

In particular, the unit ball in $\mathscr{C}([0,1])$ has only two extreme points: 1 and -1.

Example. Let Y be a compact metric space with the Borel σ-algebra. We define $\mathcal{M}(Y)$ to be the collection of all complex Borel measures on Y, and $\mathcal{A}(Y)$ to be the collection of all Borel probability measures on Y, viz., positive measures whose total measures are 1.

Given a continuous map $T : Y \to Y$, we say that $\mu \in \mathcal{A}(Y)$ is *T-invariant* if $\mu(T^{-1}B) = \mu(B)$ for every Borel set B in Y. The *Krylov-Bogoliubov theorem* guarantees that the collection $\mathcal{M}_T(Y)$ of T-invariant Borel probability measures is always nonempty. Recall now that μ is an *ergodic measure with respect to T* if $\mu(B \,\Delta\, T^{-1}(B)) = 0$ implies $\mu(B) = 0$ and $\mu(Y \smallsetminus B) = 0$. A theorem of Oxtoby characterizes $\operatorname{ext}_T(\mathcal{M}_T(Y))$ in terms of ergodic measures with respect to T.

While the set of extreme points is evidently minimal, they often do not serve as generating sets for the corresponding convex sets. Indeed, some very obviously convex sets, such as the L^1 unit ball, have no extreme points! The following theorem spells out useful criteria for extreme points serving as a generating set.

Theorem 7.4 (Krein-Milman). *If A is a nonempty, compact, convex subset of a normed linear space X, then $A = \overline{\operatorname{co}(\operatorname{ext}(A))}$.*

The rest of the section will be devoted to proving this theorem. To this end, we first estbalish a few basic properties of extreme sets and extreme points.

Proposition 7.5. *If F is an extreme subset of E and E an extreme subset of A, then F is an extreme subset of A.*

Proof. Let $x, y \in A$ and suppose that $y = tx + (1-t)y$ is in F for some $t \in (0,1)$. Since u is in the extreme subset E of A, the points x and y must be in E. F is an extreme subset of E, and so $u \in F$ implies that $x, y \in F$. $\square$

Proposition 7.6. *Let f be a convex functional on a real vector space X. For each $A \subseteq X$, the preimage $E = (f|_A)^{-1}(\alpha)$ is an extreme subset of A, where $\alpha = \sup_{x \in A} f(x)$.*

Proof. If $x, y \in A$ and $tx + (1-t)y \in E$ for some $t \in (0,1)$, then

$$\alpha = f(tx + (1-t)y) \leq tf(x) + (1-t)f(y) \leq t\alpha + (1-t)\alpha = \alpha,$$

whence $tf(x) + (1-t)f(y) = \alpha$. Therefore, $f(x) = f(y) = \alpha$, and so $x, y \in E$. $\square$

Let us now return to the task of proving the Krein-Milman theorem. To show that the extreme points serve as a generating set, we need to know that extreme points exist.

LEMMA 7.7. *Every nonempty compact set K of a normed linear space X has an extreme point.*

PROOF. We shall exhibit a minimal extreme subset of A and check that it is a singleton set. Let $\mathcal{E}_K$ be the collection of nonempty, compact, extreme subsets of K, ordered by reverse inclusion. $\mathcal{E}_K$ is nonempty, as K is in $\mathcal{E}_K$. If $(E_\alpha)_\alpha$ is a chain in $\mathcal{E}_K$, then we claim that $E = \bigcap_\alpha E_\alpha$ is the upper bound of the chain. It is clear that E is nonempty and compact. If $x, y \in K$ and $u = tx + (1-t)y$ is in E for some $t \in (0,1)$, then $u \in E_\alpha$, and so $x, y \in E_\alpha$, for each α. Therefore, E is an extreme subset of K, and thus an upper bound of $(E_\alpha)_\alpha$.

We apply Zorn's lemma on $\mathcal{E}_K$ to obtain the minimal extreme subset E_0 of K. If E_0 is not a singleton, then we can find two distinct points x_1 and x_2 in E_0. By the hyperplane separation theorem (Theorem 6.9), there exists an $l \in X^*$ such that $l(x_1) < l(x_2)$. We let $\beta = \sup_{x \in E_0} l(x)$ and set $E_0' = (l_{E_0})^{-1}(\beta)$. By the compactness of K, we can find $x_0 \in E_0'$, whence, in particular, E_0' is nonempty. E_0' is an extreme subset of E_0 as per Proposition 7.6, and thus an extreme subset of K by Proposition 7.5. Since $l(x_1) < l(x_2)$, it follows that $x_1 \in E_0 \setminus E_0'$, which is now a nonempty extreme subset of K smaller than E_0, contradicting the minimality of E_0. We conclude that E_0 must be a singleton. □

The rest of the proof of the Krein-Milman theorem proceeds as follows:

PROOF OF THE KREIN-MILMAN THEOREM. Let A be a nonempty, compact, convex subset of a normed linear space X, and set $K = \overline{\operatorname{co}(\operatorname{ext}(A))}$. Since A is convex, $\operatorname{co}(\operatorname{ext}(A)) \subseteq A$, and the inclusion $K \subseteq A$ follows from the closedness of A.

To establish the reverse inclusion, we suppose for a contradiction that we can find an element $a \in A \setminus K$. Applying the hyperplane separation theorem (Theorem 6.9) on K and $\{a\}$, we obtain $l \in X^*$ such that

$$\sup_{x \in K} l(x) < l(a). \tag{7.8}$$

Let $\alpha = \sup_{x \in A} l(x)$ and set $E = (l_A)^{-1}(\alpha)$. The compactness of A guarantees that E is nonempty. E is an extreme subset of K as per Proposition 7.6, and thus an extreme subset of A by Proposition 7.5. Furthermore, E is a closed subset of the compact set A, hence compact.

By Lemma 7.7, E has an extreme point e. Since $e \in E$, we see that $l(a) \leq l(e)$. On the other hand, $E \subseteq K$, and so (7.8) implies that $l(e) < l(a)$. This is evidently absurd, and we conclude that $A \setminus K$ is empty, or that $A = K$. □

As a corollary, we now derive a useful strengthening of Proposition 7.2:

COROLLARY 7.9. *Let A be a nonempty, compact, convex subset of a normed linear space X and l a continuous convex functional on X. Then*

$$\sup_{x \in A} l(x) = \sup_{x \in \operatorname{ext}(A)} l(x) \tag{7.10}$$

and there exists $a \in A$ such that $l(a) = \sup_{x \in A} l(x)$.

PROOF. The Krein-Milman theorem (Theorem 7.4) shows that $A = \overline{\text{co}(\text{ext}(E))}$, and so Proposition 7.2 implies (7.10). Let $\alpha = \sup_{x \in A} l(x)$ and set $E = (l_A)^{-1}(\alpha)$. The compactness of A guarantees that E is a nonempty compact set, and Proposition 7.6 implies that E is an extreme subset of A. Lemma 7.7 furnishes an extreme point a of E, whence by Proposition 7.5 a is an extreme point of A. Since $a \in E$, we see that

$$l(a) = \sup_{x \in A} l(x),$$

as was to be shown. □

8. Weak and Weak-$*$ Topologies

By now, the power and utility of convexity and compactness surely needs no further proof: we have seen, for example, the Krein-Milman theorem (Theorem 7.4), the geometric Hahn-Banach theorem (Theorem 6.7), and the hyperplane separation theorem (Theorem 6.9). Unfortunately, compact sets are quite hard to come by in a normed linear space. We have shown in Section 2 that the Heine-Borel property fails in normed linear spaces (Lemma 2.26), which, in particular, implies that closed balls are never compact. One conseqeunce is that a compact set in a normed linear space must have empty interior, for otherwise it contains a suitably small closed ball, implying that closed balls must be compact.

At the core of the problem is the over-abundance of open sets. The definition of compactness requires us to check every open cover, and a single failure to produce a finite subcover results in lack of compactness. A natural way to circumvent this problem is to reduce the number of open sets in a given space, thereby reducing the severity of this obstruction.

This method comes with a catch, however. We would like our functions and operators to be continuous whenever possible, and this property hinges on the abundance of open sets. Indeed, the inverse images of open sets would not be open if not too many sets were open to begin with.

8.1. The Weak Topology Paradigm and the Product Topology. To find a happy medium, we consider the following paradigm: let X be a set, $\{Y_\alpha\}$ a collection of topological spaces, and, for each index α, a function $f_\alpha : X \to Y_\alpha$. Let us construct the *smallest* topology on X that turns all f_α continuous. Requiring the continuity of f_α amounts to defining $f_\alpha^{-1}(U_\alpha)$ be an open set in X whenever U_α is open in Y_α. This forces us to declare any union or finite intersection of these preimage. Collecting these open sets, we obtain a topology on X generated by the collection

$$\bigcup_\alpha \{f_\alpha^{-1}(U_\alpha) : U_\alpha \text{ open in } Y_\alpha\}_\alpha,$$

which we call the *weak topology with respect to the maps f_α*.

Let us consider an example. Recall that the *Cartesian product* of a collection $\{X_\alpha\}_{\alpha \in \mathcal{I}}$ of sets indexed by an index set $\mathcal{I}$ is defined to be the set $\prod_{\alpha \in \mathcal{I}} X_\alpha$ of functions $x : \mathcal{I} \to \bigcup_\alpha X_\alpha$ such that $x^\alpha = x(\alpha) \in X_\alpha$. Note that there is the *canonical projection map* $\pi_\alpha(f) = f(\alpha)$ for each $\alpha \in \mathcal{I}$. If each X_α is a topological space, then we can endow the *product topology* on the Cartesian product $\prod_\alpha X_\alpha$, which is the weak topology on $\prod_\alpha X_\alpha$ with respect to the canonical projection maps. Here we see that our aim of creating as many compact sets as possible is accomplished quite well:

THEOREM 8.1 (Tychonoff). *If $\{X_\alpha\}$ is a family of compact topological spaces, then the product topology on $\prod_\alpha X_\alpha$ turns $\prod_\alpha X_\alpha$ into a compact topological space.*

We sketch the proof, if only to introduce the notion of generalized sequences:

SKETCH OF PROOF. A *net*[4] in a topological space X is an X-valued function $\Phi : D \to X$ on a directed set D, viz., a partially ordered set in which every pair of elements α and β has a common maximum γ. A net Φ in X is *eventually in* a subset U of X if there exists an index $\alpha_0 \in D$ such that $\Phi(\alpha) \in U$ for all $\alpha \geq \alpha_0$. Φ *converges to* $x \in X$ if, for each open neighborhood U of x, the net Φ is eventually in U. Φ is *universal* if, for each subset A of X, Φ is eventually in either A or $X \setminus A$.

Now, a topological space is compact if and only if every universal net therein converges. The image $\pi_\alpha \circ \Phi$ of a universal net Φ in $\prod_\alpha X_\alpha$ is universal in the compact space X_α and is thus convergent. Since a net Φ in $\prod_\alpha X_\alpha$ converges if and only if the components $\pi_\alpha \circ \Phi$ converge, the desired result follows. □

The component-wise convergence criterion for convergence of nets in a product space is worth reiterating. If $X_\alpha = X$ for all $\alpha \in \mathcal{I}$, then the Cartesian product $X^{\mathcal{I}} = \prod_{\alpha \in \mathcal{I}} X$ consists of X-valued functions $\varphi : \mathcal{I} \to X$ on $\mathcal{I}$. In this case, the component-wise convergence criterion states that a net (φ_β) of X-valued functions on $\mathcal{I}$ converges to $\varphi : \mathcal{I} \to X$ in the product topology of $X^{\mathcal{I}}$ if and only if

$$\pi_\alpha(\varphi_\beta) = \varphi_\beta(\alpha) \to \varphi(\alpha) = \pi_\alpha(\varphi)$$

for each $\alpha \in \mathcal{I}$.

EXAMPLE. we consider $X = \mathbb{F}$ and $\mathcal{I} = [a, b]$. Here a net (φ_α) of $\mathbb{F}$-valued functions $\varphi_\alpha : [a, b] \to \mathbb{F}$ converges to φ in the product topology on $\mathbb{F}^{[a,b]}$ if and only if (f_α) converges pointwise to φ in the standard Euclidean topology on $\mathbb{F}$.

For this reason, the product topology is sometimes referred to as the *topology of pointwise convergence*. Using this idea, we can define a notion of boundedness on $X^{\mathcal{I}}$, provided that boundedness at each component is well-defined:

DEFINITION 8.2. Let X be a metric space. A subset E in the product space $X^{\mathcal{I}}$ is *pointwise bounded* if

$$\pi_\alpha(E) = \{\varphi(\alpha) : \varphi \in E\}$$

is bounded for each $\alpha \in \mathcal{I}$.

It turns out that this notion of boundedness yields the Heine-Borel property for spaces of scalar-valued functions. We derive this as a quick corollary of Tychonoff's theorem.

COROLLARY 8.3 (Pointwise Arzelà-Ascoli theorem). *A subset K of the product space $\mathbb{F}^{\mathcal{I}}$ with the product topology is compact if and only if K is closed and pointwise bounded.*

[4]The details can be found in Appendix A, Section 3. For the remainder of this section, it suffices to think of nets as sequences with "uncountable index sets" as opposed to $\mathbb{N}$, retaining much of the properties that sequences in a topological space have. In analogy with sequences, we write $(x_\beta)_{\beta \in D}$ to denote a net with indices β in a directed set D.

PROOF. ($\Leftarrow$) Since the product of Hausdorff topological spaces is Hausdorff, $\mathbb{F}^{\mathcal{I}}$ is Hausdorff. Compact sets in a Hausdorff space are closed, and so K is closed. Moreover, the continuous image $\pi_\alpha(K)$ of the compact set K is compact for each $\alpha \in \alpha \in \mathcal{I}$, hence bounded. It follows that K is pointwise bounded.

($\Rightarrow$) Let K be closed and pointwise bounded. $K_\alpha = \overline{\pi_\alpha(K)}$ is closed and bounded in $\mathbb{F}$, hence compact by the Heine-Borel theorem. By Tychonoff's theorem (Theorem 8.1), $\prod_{\alpha \in \mathcal{I}} K_\alpha$ is a compact subset of $\mathbb{F}^{\mathcal{I}}$. In particular, $\prod_\alpha K_\alpha$ contains the closed set K, whence K must be compact. □

8.2. Weak-$*$ Topology. Given a vector space X over $\mathbb{F}$, let us now consider the X-fold product

$$\mathbb{F}^X = \prod_{x \in X} \mathbb{F}$$

of the base field $\mathbb{F}$ with the product topology. $\mathbb{F}^X$ is precisely the set of $\mathbb{F}$-valued functions on X and includes the set X' of algebraic duals as its subset. If X is a topological vector space, then the topological dual X^* is also a subset of $\mathbb{F}^X$.

DEFINITION 8.4. Let X be a topological vector space over $\mathbb{F}$. The *weak-$*$ topology* on X^* is the subspace topology with respect to the product topology on $\mathbb{F}^X$ and is denoted by $\sigma(X^*, X)$. We write $l_\beta \overset{*}{\rightharpoonup} l$ to denote the convergence of a net (l_β) to l in the weak-$*$ topology.

Once again, we see that our aim of obtaining many compact sets has been met:

THEOREM 8.5 (Banach-Alaoglu). *Let X be a normed linear space. The norm unit ball*

$$B = \{l \in X^* : \|l\| \leq 1\}$$

is weak-$*$ compact. Consequently, every closed norm ball in X^* is weak-$*$ compact.

PROOF. By the pointwise Arzelà-Ascoli theorem (Corollary 8.3), it suffices to check that B is weak-$*$ closed and pointwise bounded. If $l \in B$, then

$$|l(x)| \leq \|l\|_{X^*}\|x\|_X \leq \|x\|_X$$

for each $x \in X$, whence B is pointwise bounded.

To show that B is weak-$*$ closed, we take a net (l_β) in B that converges to some element $l \in \mathbb{F}^X$. l is evidently linear, and so it suffices to check that $\|l\|_{X^*} \leq 1$. For each $x \in X$, we have $l_\beta(x) \to l(x)$ by the component-wise convergence criterion, whence

$$|l(x)| = \lim_\beta |l_\beta(x)| \leq \lim_\beta \|x\| = \|x\|.$$

We conclude that $l \in B$. □

Let us now check that the weak topology paradigm that we have introduced in the beginning of the section has been adpoted in the construction of the weak-$*$ topology.

PROPOSITION 8.6. *Let X be a topological vector space over $\mathbb{F}$. For each $x \in X$, the evaluation functional $\hat{x} : X^* \to \mathbb{F}$ by setting $\hat{x}(l) = l(x)$ for each $l \in X^*$ is a continuous linear functional on X^* with respect to $\sigma(X^*, X)$. Furthermore, $\sigma(X^*, X)$ is the weak topology generated by the collection of evaluation functionals.*

PROOF. It suffices to mull over the fact that the evaluation functional $\hat{x}$ is precisely the canonical projection map $\pi_x : \mathbb{F}^X \to \mathbb{F}$ restricted to X^*. □

Using this characterization, we can give a simple description of the open sets in $\sigma(X^*, X)$.

PROPOSITION 8.7. *Let X be a topological vector space. The sets*

$$V(x_1, \ldots, x_n; \varepsilon) = \{l \in X^* : |l(x_i) - l_0(x_i)| < \varepsilon \text{ for all } 1 \leq i \leq n\},$$

where $l_0 \in X^*$, $x_1, \ldots, x_n \in X$, and $\varepsilon > 0$, form a basis of neighborhoods of l_0 in the weak-$*$ topology $\sigma(X^*, X)$.

PROOF. Let $y_i = l_0(x_i)$ and observe that

$$V(x_1, \ldots, x_n; \varepsilon) = \bigcap_{i=1}^{n} \widehat{x_i}^{-1}\left(B_\varepsilon(y_i)\right), \tag{8.8}$$

which is an open neighborhood of l_0 with respect to $\sigma(X^*, X)$ by Proposition 8.6. Conversely, the second half of Proposition 8.6 states that every open set in $\sigma(X^*, X)$ is generated by the preimages of open sets in $\mathbb{F}$ under the evaluation functionals, whence every neighborhood of l_0 can be written as a union of open sets of the form (8.8). The desired result now follows. □

Proposition 8.6 tells us that the evaluation functionals are elements of the dual space[5] $(X^\star)^*$ of the weak-$*$ dual $X^\star$ of X. If X is, for example, a normed linear space, then the Hahn-Banach theorem guarantees that these maps are distinct, so that $(X^\star)^*$ is nontrivial. We now show that every element of $(X^\star)^*$ is an evaluation functional, thereby proving that every normed linear space is reflexive[6] in the weak-$*$ sense.

THEOREM 8.9. *Every normed linear space X is linearly isomorphic to the dual $(X^\star)^*$ of the weak-$*$ dual $X^\star$.*

We shall require two technical lemmas. The first provides a characterization of continuity of linear functionals using boundedness on a neighborhood.

LEMMA 8.10. *A linear functional $l : Y \to \mathbb{F}$ on a topological vector sapce Y is continuous if and only if l is bounded on a neighborhood of 0.*

PROOF OF LEMMA. ($\Leftarrow$) If l is continuous, then $l^{-1}((-1,1))$ is an open neighborhood of 0, on which l is bounded by 1.

($\Rightarrow$) It evidently suffices to prove that l is continuous at 0: indeed, a minor modification of the argument given in Proposition 3.2 would work. Suppose without loss of generality that l is bounded by 1 on a neighborhood U of 0. If (y_β) is a net in Y that converges to 0, then, for each $\varepsilon > 0$, we can find an index β_ε such that $\beta \geq \beta_\varepsilon$ implies $y_\beta \in \varepsilon U$, or $(1/\varepsilon)y_\beta \in U$. It now follows that $|l((1/\varepsilon)y_\beta)| \leq 1$, whence $|l(y_\beta)| \leq \varepsilon$ for all $\beta \geq \beta_\varepsilon$. Therefore, $l(y_\beta) \to 0$, and we conclude that l is continuous at 0. □

The second linear-algebraic lemma furnishes a necessary and sufficient condition for a linear functional to be a linear combination of other linear functionals.

---

[5]The use of $\star$ to denote the weak-$*$ dual is not standard and used only in this section for notational convenience.

[6]See Definition 5.8

LEMMA 8.11. *If V is a vector space and $l_1, \dots, l_n, l$ are elements of the algebraic dual V' of V, then l is a linear combination of $l_1, \dots, l_n$ if and only if*

$$\bigcap_{i=1}^{n} \ker(l_i) \subseteq \ker l. \tag{8.12}$$

PROOF OF LEMMA. ($\Rightarrow$) If l is a linear combination of $l_1, \dots, l_n$, then $l_i(x) = 0$ for all i evidently implies $l(x) = 0$, which is precisely the statement of (8.12).

($\Leftarrow$) We define $\pi : V \to \mathbb{F}^n$ by setting

$$\pi(x) = (l_1(x), \dots, l_n(x))$$

and consider the mapping $\pi(x) \overset{\Phi}{\mapsto} l(x)$ We first check that Φ is a well-defined linear functional on $\pi(V) \subseteq \mathbb{F}^n$. Indeed, if $\pi(x_1) = \pi(x_2)$, then $l_i(x_1) = l_i(x_2)$ for all $1 \leq i \leq n$, whence $x_1 - x_2 \in \bigcap \ker(l_i)$. (8.12) now implies that $x_1 - x_2 \in \ker l$, which, in particular, shows that

$$\Phi(\pi(x_1)) = l(x_1) = l(x_2) = \Phi(\pi(x_2)).$$

Now, as a linear functional on the finite-dimensional vector space $\pi(V)$, Φ admits a concrete representation:

$$\Phi(u_1, \dots, u_n) = \sum_{i=1}^{n} \lambda_i u_i.$$

It now follows that

$$l(x) = \Phi(\pi(x)) = \sum_{i=1}^{n} \lambda_i l_i(x),$$

as was to be shown. □

PROOF OF THEOREM 8.9. The mapping $x \mapsto \hat{x}$ is evidently linear. If $\hat{x} = \hat{y}$, then $l(x - y) = 0$ for all $l \in X^*$. By the existence of nontrivial annihilators on a normed linear space (Proposition 5.10), it follows that $x - y = 0$, whence the mapping $x \mapsto \hat{x}$ is injective.

To show that the mapping $x \mapsto \hat{x}$ is surjective, we fix $\varphi \in (X^\star)^*$ and invoke Lemma 8.10 to find a neighborhood V of 0 on which φ is bounded by 1. By Proposition 8.7, we may assume without loss of generality that $V = V_{x_1,\dots,x_n;\varepsilon}$. Our goal is to write φ as a linear combination of $\widehat{x_1}, \dots, \widehat{x_n}$, which establishes the surjectivity of the mapping $x \mapsto \hat{x}$.

By Lemma 8.11, it suffices to check that

$$\bigcap_{i=1}^{n} \ker(\widehat{x_i}) \subseteq \ker \varphi.$$

Observe that if l is an element of the linear subspace $\bigcap \ker(\widehat{x_i})$ of $X^\star$, then $l(x_i) = 0$ for all $1 \leq i \leq n$, whence $\delta l \in V_{x_1,\dots,x_n;\varepsilon}$ for all $\delta > 0$. We see that $|\varphi(\delta l)| \leq 1$, which implies that $|\varphi(l)| \leq \delta^{-1}$ for all $\delta > 0$. It follows that $\varphi(l) = 0$, whence $l \in \ker \varphi$, as was to be shown. □

8.3. Weak Topology. We now employ the weak topology paradigm on the topological vector space X itself, rather than its dual.

DEFINITION 8.13. The *weak topology* $\sigma(X, X^*)$ on a topological vector space X is the weak topology with respect to the elements of X^*, the continuous linear functionals. We write $x_\beta \rightharpoonup x$ to denote the convergence of a net (x_β) to x in the weak topology.

Recall that if X is a normed linear space, then the mapping $x \mapsto \hat{x}$ embeds X isometrically into X^{**}; let us write $\widehat{X}$ to denote the image. Certainly X^{**} is a subset of the product space $\mathbb{F}^{X^*}$, and so we may consider $\widehat{X}$ with the subspace topology inherited from the product topology on $\mathbb{F}^{X^*}$. Here the action of a typical canonical projection map looks like this:

$$\pi_l(\hat{x}) = \hat{x}(l) = l(x).$$

Therefore, the continuity of the projection maps π_l is equivalent to the continuity of the coressponding linear functionals, whence the subspace topology on $\widehat{X} \cong X$ inherited from the product topology on $\mathbb{F}^{X^*}$ is precisey the weak topology on X.

We now exhibit a simple collection of open sets that serves as a neighborhood basis in the weak topology, similar to that producted in Proposition 8.7.

PROPOSITION 8.14. *Let X be a topological vector space. The sets*

$$V(l_1, \ldots, l_n; \varepsilon) = \{x \in X : |l_i(x) - l_i(x_0)| < \varepsilon \text{ for all } 1 \leq i \leq n\},$$

where $l_1, \ldots, l_n \in X^*$, $x_0 \in X$, and $\varepsilon > 0$, form a basis of neighborhoods of x_0 in the weak topology $\sigma(X, X^*)$.

PROOF. Let $y_i = l_i(x_0)$ and observe that

$$V(l_1, \ldots, l_n; \varepsilon) = \bigcap_{i=1}^{n} l_i^{-1}\left(B_\varepsilon(y_i)\right), \tag{8.15}$$

which is an open neighborhood of x_0 with respect to $\sigma(X, X^*)$. Conversely, the definition of the weak topology on X states that every open set in $\sigma(X, X^*)$ is generated by the preimages of open sets in $\mathbb{F}$ under the bounded linear functionals, whence every neighborhood of x_0 can be written as a union of open sets of the form (8.15). The desired result now follows. ☐

EXERCISE 8.16. Using the above characterization, we can show that a bounded set in an infinite-dimensional normed linear space can never be weak open. In particular, the weak topology on an infinite-dimensional space is always strictly weaker than the norm topology.

EXERCISE 8.17. Nevertheless, it is possible for *strong convergence* and *weak convergence* to coincide on an infinite-dimensional space—for example, on $l^1(\mathbb{N})$. If $f_n \rightharpoonup f$ but $f_n \not\to f$, then we may assume without loss of generality that $f = 0$ and $\|f_n\| > 1$ and consider a subsequence $(f_{n_k})_{k=1}^\infty$ and an increasing sequence $(M_k)_{k=1}^\infty$ of indices such that

$$\|f_{n_k}\|_1 - \varepsilon < \sum_{n=M_{k-1}}^{M_k} |f_{n_k}(m)|$$

for all $k \in \mathbb{N}$. Letting

$$g(m) = \begin{cases} |f_{n_k}(m)|/f_{n_k}(m) & \text{if } M_{k-1}+1 \leq m \leq M_k \text{ for some } k; \\ 0 & \text{otherwise;} \end{cases}$$

we see that $|\sum_k f_{n_k}(m)g(m)| > 1 - 2\varepsilon$, which contradicts $f_n \rightharpoonup 0$.

Our next task is to characterize the dual of a normed linear space X with the weak topology. We say that a linear functional on X is *weakly continuous* if l is continuous with respect to the weak topology $\sigma(X, X^*)$. In contrast, a bounded linear functional on X is often said to be *strongly continuous.* The following proposition shows that this distinction is, in fact, unnecessary.

PROPOSITION 8.18. *A linear functional l on X is weakly continuous if and only if it is strongly continuous.*

PROOF. Let U be an arbitrary open set in $\mathbb{F}$. If l is weakly continuous, then $l^{-1}(U)$ is open in $\sigma(X, X^*)$, which is coarser than the norm topology on X. Therefore, $l^{-1}(U)$ is open in the norm topology of X, and l is strongly continuous. Conversely, if l is strongly continuous, then $l^{-1}(U)$ is an element of the generating set of $\sigma(X, X^*)$ and is thus open in $\sigma(X, X^*)$. It follows that l is weakly continuous. □

We remark that strong continuity always implies weak continuity, while weakly continuous *nonlinear* maps are often not strongly continuous.

EXERCISE 8.19. [[**Bre11**], Exercise 4.20] We fix $p, q \in [1, \infty)$ and let $a : \mathbb{R} \to \mathbb{R}$ be a continuous function satisfying the estimate

$$|a(t)| \leq C(|t|^{p/q} + 1)$$

for all $t \in \mathbb{R}$. Show that the operator $A : L^p(0,1) \to L^q(0,1)$ defined by the formula

$$(Af)(x) = a(f(x))$$

is strongly continuous but is never weakly continuous unless a is affine.

See Appendix A, Theorem 4.1 for a weak compactness result analogous to the Banach-Alaoglu theorem (Theorem 8.5).

8.4. Concluding Remarks. For the sake of concreteness, we let X be a normed linear space. X can be endowed with either of the following two topologies.

(1) The *norm topology*, which gives rise to the strong convergence $x_\beta \to x$ of nets, i.e., $\lim_\beta \|x_\beta - x\| = 0$. This is equivalent to the **uniform** convergence of the corresponding nets $(\widehat{x_\beta})$ of evaluation functionals on the norm unit ball $B_1[0]$ in X^{**}, which can also be written as

$$\lim_\beta \sup_{\|l\| \leq 1} |\widehat{x_\beta}(l) - \hat{x}(l)| = \lim_\beta \sup_{\|l\| \leq 1} |l(x_\beta) - l(x)| = 0.$$

(2) The *weak topology* $\sigma(X, X^*)$, which gives rise to the weak convergence $x_\beta \rightharpoonup x$ of nets, i.e., $l(x_\beta) \rightharpoonup l(x)$ for all $l \in X^*$. This is equivalent to the **pointwise** convergence of the corresponding nets $(\widehat{x_\beta})$ of evaluation functionals in the double dual X^{**}.

X^*, on the other hand, allows for three possible topologies.

(1) The *norm topology*, which gives rise to the strong convergence $l_\beta \to l$ of nets of linear functionals. This is equivalent to the **uniform** convergence of the nets on the norm unit ball $B_1[0]$ in X.

(2) The *weak-$*$ topology* $\sigma(X^*, X)$, which gives rise to the weak-$*$ convergence $l_\beta \overset{*}{\rightharpoonup} l$ of nets of linear functionals. This is equivalent to the **pointwise** convergence of the nets (l_β) on X.
(3) The *weak topology* $\sigma(X^*, X^{**})$, which gives rise to the weak convergence $l_\beta \rightharpoonup l$ of nets of linear functionals, i.e., $\varphi(l_\beta) \to \varphi(l)$ for all $\varphi \in X^{**}$.

Let us now discuss two cases in which some of these topologies might be the same.

PROPOSITION 8.20. *If X is finite-dimensional, then all of the topologies are the same.*

PROOF. For notational convenience, we consider $\mathbb{R}^n$ with the Euclidean norm topology. We can do this without loss of generality, as all norms on a finite-dimensional vector space are equivalent (Theorem 2.15). The projection maps onto the kth coordinate $\pi_k(x) = x^k$ are continuous linear functionals on $\mathbb{R}^n$. Since the weak topology with respect to the projection maps $\pi_1, \ldots, \pi_n$ is already the Euclidean norm topology, the weak and norm topologies on $\mathbb{R}^n$ are the same. Similarly, the weak and norm topologies on $(\mathbb{R}^n)^* \cong \mathbb{R}^n$ are the same. The product-space definition of the weak-$*$ topology implies that the weak-$*$ topology on $(\mathbb{R}^n)^*$ is precisely the weak topology generated by the projection maps $\pi_1, \ldots, \pi_n$, which is the same as the weak and norm topologies on $(\mathbb{R}^n)^*$. $\square$

PROPOSITION 8.21. *If X is reflexive, then the weak-$*$ topology $\sigma(X^*, X)$ and the weak topology $\sigma(X^*, X^{**})$ are the same.*

PROOF. If X is reflexive, then every element of X^{**} is an evaluation functional. Therefore, the weak convergence $l_\beta \rightharpoonup l$ of nets of linear functionals in this case is equivalent to $\hat{x}(l_\beta) \to \hat{x}(l)$ for all $x \in X$. We can rewrite this as $l_\beta(x) \to l(x)$, which is precisely the pointwise-convergence criterion for the weak-$*$ convergence $l_\beta \overset{*}{\rightharpoonup} l$ of nets of linear functionals. $\square$

Much of our discussion in this section relied on net characterizations of continuity and compactness. These can be reduced to the usual sequential characterizations with the added assumption of metrizability. Unfortunately, metrizability turns out to be too strong of an assumption in general. See Appendix A, Section 5 for a discussion.

9. Locally Convex Spaces

We now turn to the other recurring theme of the chapter: convexity. The goal of this is to convince ourselves that the natural setting for many of the results developed so far is a topological vector space with sufficiently many open convex sets, which we define precisely as follows:

DEFINITION 9.1. A topological vector space is said to be *locally convex* if it has a neighborhood basis consisting of convex sets.

We see at once that normed linear spaces are locally convex; this is a consequence of the triangle inequality. The following proposition provides a typical way of creating a locally convex space.

Proposition 9.2. *Let X be a vector space, $\{\rho_\alpha\}_{\alpha \in \mathcal{I}}$ a family of seminorms on X, x_0 a fixed point in X, and $\mathscr{B}(x_0)$ a collection of sets of the form*

$$V(\alpha_1, \ldots, \alpha_n; \varepsilon) = \{x \in X : \rho_{\alpha_i}(x - x_0) < \varepsilon \ \text{ for all } 1 \leq i \leq n\}.$$

Then $\mathscr{B}(x_0)$ is a basis of neighborhoods of x_0 consisting of convex sets, and the topology generated by

$$\bigcup_{x_0 \in X} \mathscr{B}(x_0) \tag{9.3}$$

turns X into a locally convex space.

In particular, if X is a normed linear space, then the weak topology $\sigma(X, X^*)$ turns X into a locally convex space (Proposition 8.14), and the weak-$*$ topology $\sigma(X^*, X)$ turns X^* into a locally convex space (Proposition 8.7).

Proof. Let X be endowed with the topology generated by (9.3). We observe that a net (x_β) in X converges to $x_0 \in X$ if and only if $\rho_\alpha(x_\beta - x_0) \to 0$ in for all α. Indeed, the seminorm convergence criterion shows that that each single-index neighborhood $V(\alpha; \varepsilon)$ admits an index β_α such that $\beta \geq \alpha$ implies $x_\beta \in V(\alpha; \varepsilon)$. We then see that x_β is in

$$V(\alpha_1, \ldots, \alpha_n; \varepsilon) = \bigcap_{i=1}^{n} V(\alpha_i; \varepsilon)$$

for all $\beta \geq \max\{\beta_1, \ldots, \beta_n\}$. Since $\mathscr{B}(x_0)$ is a basis of neighborhoods of x_0, it follows that $x_\beta \to x_0$. The converse is established analogously.

We now show that X is a topological vector space. If $((\lambda_\beta, x_\beta))_\beta$ be a net in $\mathbb{F} \times X$ that converges to (λ, x), then

$$\rho_\alpha(\lambda_\beta x_\beta - \lambda x) \leq \rho_\alpha(\lambda_\beta(x_\beta - x)) + \rho_\alpha((\lambda_\beta - \lambda)x) \to 0$$

for all α, whence the scalar multiplication operation is continuous. To show that vector addition is continuous, we fix a net (x_β, y_β) in $X \times X$ that converges to (x, y). We observe that

$$\rho_\alpha((x_\beta + y_\beta) - (x + y)) \leq \rho_\alpha(x_\beta - x) + \rho_\alpha(y_\beta - y) \to 0$$

for each α, whence $x_\beta + y_\beta \to x + y$. It follows that the vector addition operation is continuous, and X is a topological vector space.

It remains to show that X is a locally convex space. To this end, we observe that $V(\alpha; \varepsilon)$ based at x_0 is convex, for if $x, y \in V(\alpha; \varepsilon)$, then

$$\begin{aligned} \rho_\alpha([(1-\lambda)x + \lambda y] - x_0) &= \rho_\alpha((1-\lambda)(x - x_0) + \lambda(y - x_0)) \\ &\leq (1-\lambda)\rho_\alpha(x - x_0) + \lambda\rho_\alpha(y - x_0) \\ &< \varepsilon, \end{aligned}$$

whence $(1-\lambda)x + \lambda y \in V(\alpha; \varepsilon)$ for all $0 \leq \lambda \leq 1$. Since intersections of convex sets are convex, $\mathscr{B}(x_0)$ consists of convex sets, and the desired result follows. □

We now generalize the notion of bounded operators (Definition 3.1) via seminorms.

PROPOSITION 9.4. *Let X and Y be locally convex spaces whose topologies are generated by $\{\rho_\alpha\}_\alpha$ and $\{\sigma_\beta\}_\beta$, respectively, à la Proposition 9.2. A linear operator $T : X \to Y$ is continuous if and only if each Y-index β admits a finite collection of X-indices $\{\alpha_1, \ldots, \alpha_n\}$ and a positive constant A such that*

$$\sigma_\beta(Tx) \leq A \sum_{i=1}^{n} \rho_{\alpha_i}(x) \tag{9.5}$$

for all $x \in X$.

PROOF. ($\Leftarrow$) Let (x_γ) be a net in X that converges to $x \in X$. By Proposition 9.2, $\rho_\alpha(x_\gamma - x) \to 0$ for all α, whence $\sigma_\beta(Tx_\gamma - Tx) \to 0$ by (9.5). It follows that T is continuous.

($\Rightarrow$) Suppose that T is continuous, so that the preimage $T^{-1}(V(\beta; 1))$ of

$$V(\beta; 1) = \{y \in Y : \sigma_\beta(y) < 1\}$$

is an open neighborhood of 0. By Proposition 9.2, we can find a neighborhood $V(\alpha_1, \ldots, \alpha_n; \varepsilon)$ of 0 that is entirely contained in $T^{-1}(V(\beta; 1))$.

We fix $x \in X$. If $\rho_{\alpha_i}(x) \neq 0$ for some $1 \leq i \leq n$, then we can define the vector

$$z = \left[\frac{\varepsilon}{2} \left(\sum_{i=1}^{n} \rho_{\alpha_i}(x) \right)^{-1} \right] x,$$

so that

$$\rho_{\alpha_i}(z) = \frac{\varepsilon}{2} \cdot \frac{\rho_{\alpha_i}(x)}{\sum \rho_{\alpha_i}(x)} < \varepsilon$$

for all $1 \leq i \leq n$. Therefore, $z \in V(\alpha_1, \ldots, \alpha_n; \varepsilon) \subseteq T^{-1}(V(\beta; 1))$, and we have

$$\frac{\varepsilon}{2} \frac{\sigma_\beta(Tx)}{\sum \rho_{\alpha_i}(x)} = \sigma_\beta(Tz) < 1,$$

whence it follows that

$$\sigma_\beta(Tx) \leq \frac{2}{\varepsilon} \sum_{i=1}^{n} \rho_{\alpha_i}(x),$$

as was to be shown.

On the other hand, if $\rho_{\alpha_i}(x) = 0$ for all $1 \leq i \leq n$, then $\rho_{\alpha_i}(Mx) = 0$ for all $M > 0$. This implies that $Mx \in V(\alpha_1, \ldots, \alpha_n; \varepsilon) \subseteq T^{-1}(V(\beta; 1))$ for all $M > 0$, which, in turn, shows that $\sigma_\beta(Tx) < 1/M$ for all $M > 0$. It follows that $\sigma_\beta(Tx) = 0$, and (9.5) is obtained trivially. □

As a corollary, we obtain a generalization of the notion of equivalence of norms (Definition 2.13).

COROLLARY 9.6. *Two collections of seminorms $\{\rho_\alpha\}_\alpha$ and $\{\sigma_\beta\}_\beta$ generate the same topology on a vector space X if and only if the following two conditions hold:*

(1) For each β, there exist indices $\alpha_1, \ldots, \alpha_n$ and a positive constant A such that

$$\sigma_\beta(Tx) \leq A \sum_{i=1}^{n} \rho_{\alpha_i}(x)$$

for all $x \in X$.

(2) *For each α, there exist indices $\beta_1, \ldots, \beta_n$ and a positive constant B such that*

$$\rho_\alpha(Tx) \leq B \sum_{i=1}^{n} \sigma_{\beta_i}(x)$$

for all $x \in X$.

PROOF. For notational convenience, we write X_α and X_β to denote X with topologies generated by $\{\rho_\alpha\}$ and $\{\sigma_\beta\}$, respectively. If (1) and (2) hold, then Proposition 9.4 implies that the identity map $\mathrm{id} : X_\alpha \to X_\beta$ is a homeomorphism. Conversely, if $X_\alpha \cong X_\beta$, then $\mathrm{id} : X_\alpha \to X_\beta$ is a homeomorphism, whence (1) and (2) follow from Proposition 9.4. □

The generalizations we have carried out above suggest that much of the Banach-space theory could be carried over to an appropriate class of locally convex spaces. We now introduce a metrizable topological vector space that is complete with respect to the seminorms that generate the topology.

DEFINITION 9.7. A *Fréchet space* is a topological vector space X whose topology is generated by a complete translation-invariant metric.

In what sense is a Fréchet space complete with respect to seminorms? The next three propositions detail a typical method of creating a Fréchet space.

PROPOSITION 9.8. *The topology on a topological vector space X generated by a countable family of seminorms $\{\rho_n\}_{n \in \mathbb{N}}$ is metrizable by the following translation-invariant metric:*

$$d(x, y) = \sum_{n=1}^{\infty} \frac{1}{2^n} \frac{\rho_n(x-y)}{1 + \rho_n(x-y)}.$$

PROOF. For notational convenience, we write X_ρ and X_d to denote X with topologies generated by $\{\rho_n\}$ and d, respectively. Our goal is to show that X_ρ is homeomorphic to X_d. To do so, it suffices to show that an open set in X_ρ is open in X_d and vice versa.

($\Rightarrow$) By Proposition 9.2 and the translation invariance of the two topologies at hand, it suffices to show that

$$V(n_1, \ldots, n_k; \varepsilon) = \{x \in X : \rho_{n_i}(x) < \varepsilon \text{ for all } 1 \leq i \leq k\}$$

is open in X_d. We fix $x \in V(n_1, \ldots, n_k; \varepsilon)$. If $y \in X$ satisfies the lower bound $\rho_{n_i}(y) \geq \varepsilon$ for some $1 \leq i \leq n$, then

$$\begin{aligned}
d(x,y) &\geq \frac{1}{2^{n_i}} \frac{\rho_{n_i}(x-y)}{1+\rho_{n_i}(x-y)} \\
\text{(triangle inequality for } \rho_{n_i}) &\geq \frac{1}{2^{n_i}} \frac{|\rho_{n_i}(x) - \rho_{n_i}(y)|}{1+\rho_{n_i}(x)+\rho_{n_i}(y)} \\
\text{(upper bound of } \rho_{n_i}(x)) &\geq \frac{1}{2^{n_i}} \frac{\rho_{n_i}(y) - \rho_{n_i}(x)}{(1+\varepsilon)+\rho_{n_i}(y)} \\
\text{(removing dependence on } n_i) &\geq \frac{1}{2^{n_i}} \frac{\rho_{n_i}(y) - \max_{1\leq i \leq n} \rho_{n_i}(x)}{(1+\varepsilon)+\rho_{n_i}(y)} \\
\text{(lower bound of } \rho_{n_i}(y)) &\geq \frac{1}{2^{n_i}} \cdot \frac{\varepsilon - \max_{1\leq i \leq n} \rho_{n_i}(x)}{1+2\varepsilon} \\
\text{(removing dependence on } n_i) &\geq 2^{-\max\{n_1,\ldots,n_k\}} \cdot \frac{\varepsilon - \max_{1\leq i \leq n} \rho_{n_i}(x)}{1+2\varepsilon}
\end{aligned}$$

It then follows that

$$d(x,z) < 2^{-\max\{n_1,\ldots,n_k\}} \cdot \frac{\varepsilon - \max_{1\leq i \leq n} \rho_{n_i}(x)}{1+2\varepsilon}$$

implies $\rho_{n_i}(z) < \varepsilon$ for all $1 \leq i \leq n$, hence $z \in V(n_1, \ldots, n_k; \varepsilon)$. Therefore, $V(n_1, \ldots, n_k; \varepsilon)$ is open in X_d.

($\Leftarrow$) By the translation invariance of the two topologies at hand, it suffices to show that $B_\varepsilon(0)$ is open in X_ρ. For each $x \in B_\varepsilon(0)$, we set

$$r = \frac{1}{2} \min\{d(x,0), |\varepsilon - d(x,0)|\}$$

and note that $B_r(x) \subseteq B_\varepsilon(0)$. We find a natural number k such that $2^{-k} < r/2$ and observe that $\rho_i(x-y) < r/k$ implies

$$d(x,y) < \sum_{i=1}^{k} \frac{1}{2^n} \frac{\rho_i(x-y)}{1+\rho_i(x-y)} + \frac{1}{2^k} < \frac{1}{2} \sum_{i=1}^{k} \rho_i(x-y) + \frac{r}{2} < \frac{1}{2} \cdot k \cdot \frac{r}{k} + \frac{r}{2} = r,$$

whence $V(1, \ldots, k; r/k)$ is an X_ρ-open neighborhood of x in $B_r(x)$, hence in $B_\varepsilon(0)$. It follows that $B_\varepsilon(0)$ is open in X_ρ. □

Proposition 9.9. *A sequence $(x_n)_{n=1}^\infty$ in a topological vector space X generated by a countable family of seminorms $\{\rho_m\}_{m \in \mathbb{N}}$ converges to $x \in X$ if and only if $\rho_m(x_n - x) \to 0$ for each fixed $m \in \mathbb{N}$.*

Proof. This is a trivial consequence of Proposition 9.8. □

Proposition 9.10. *Let $\{\rho_m\}_{m\in\mathbb{N}}$ be a countable family of seminorms that generates the topology on a topological vector space X. If every sequence $(x_n)_{n=1}^\infty$ such that $\rho_m(x_{n_1} - x_{n_2}) \to 0$ as $n_1, n_2 \to \infty$ for each fixed $m \in \mathbb{N}$ converges to a limit in X, then X is a Fréchet space.*

Proof. By Proposition 9.9, the hypothesis translates to the metric

$$d(x,y) = \sum_{n=1}^{\infty} \frac{1}{2^n} \frac{\rho_n(x-y)}{1+\rho_n(x-y)}$$

being complete, whence the desired result follows from Proposition 9.8. ☐

We now consider two of the most important examples of Fréchet spaces.

EXAMPLE. The space $\mathscr{C}^\infty(K)$ of smooth functions on a compact subset K of $\mathbb{R}$ comes with the natural seminorms

$$\rho_n(f) = \sup_{x \in K} |f^{(n)}(x)|$$

which are complete in the sense of Proposition 9.10. Similarly, the *Schwartz space* $\mathscr{S}(\mathbb{R})$ of smooth functions $f : \mathbb{R} \to \mathbb{C}$ such that the quantity

$$\rho_{n,m}(f) = \sup_{x \in \mathbb{R}} |x^m f^{(n)}(x)|$$

is finite for each $m, n \in \mathbb{N}$ is a Fréchet space with the seminorms $\rho_{n,m}$. These are examples of spaces of *test functions*, which forms the basis of distribution theory. See Appendix A, Section 8 for a discussion.

We conclude this section by placing some of the major theorems established in this chapter in their proper contexts. The omitted proofs can be found in many of the standard functional analysis textbooks.

THEOREM 9.11 (Banach-Schauder, open mapping theorem). *If $T : X \to Y$ is a surjective continuous linear operator between Fréchet spaces X and Y, then T is open.*

THEOREM 9.12 (Closed graph theorem). *A linear operator $T : X \to Y$ between Fréchet spaces X and Y is bounded if and only if $\Gamma(T)$ is closed.*

THEOREM 9.13 (Banach-Steinhaus-Dieudonné, uniform boundedness principle). *Let X be a Fréchet space, Y a normed linear space. Let Γ be a nonempty index set, and, for each $\alpha \in \Lambda$, we let $T_\alpha : X \to Y$ be a continuous linear operator. If*

$$\sup_{\alpha \in \Lambda} \|T_\alpha x\| < \infty$$

for all $x \in X$, then $\{T_\alpha\}_{\alpha \in \Lambda}$ is equicontinuous, viz., for each $\varepsilon > 0$, there exists a $\delta > 0$ such that $d(x, y) < \delta$ implies $\|T_\alpha x - T_\alpha y\| < \varepsilon$ for all $x, y \in X$ and $\alpha \in \Lambda$.

THEOREM 9.14 (Analytic Hahn-Banach). *Let X be a locally convex space, M a subspace of X, and l a continuous linear functional on M. Then there exists a continuous extension L of l onto the whole space X.*

THEOREM 9.15 (Existence of nontrivial annihilators). *If X is a locally convex space and M a closed proper subspace of X, then there exists a nonzero continuous linear functional L on X such that $L|_M = 0$.*

THEOREM 9.16 (Hyperplane separation theorem). *Let X be a real locally convex space and A and B nonempty, convex, disjoint subsets of X. If A is compact and B is closed, then there exists an $L \in X^*$ and an $\alpha \in \mathbb{R}$ such that*

$$\sup_{x \in A} L(x) < \alpha < \inf_{y \in B} L(y).$$

In other words, the hyperplane $\{z : L(z) = \alpha\}$ separates A and B.

THEOREM 9.17 (Krein-Milman). *If X is a locally convex space and K a nonempty, compact, convex subset of X, then $K = \overline{\operatorname{co}(\operatorname{ext}(A))}$.*

THEOREM 9.18 (Banach-Alaoglu). *Let X be a topological vector space. If K is the collection of continuous linear functionals on X that are bounded by 1 on a fixed neighborhood of 0 in X, then K is weak-$*$ compact.*

CHAPTER 2

Spectral Theory of Operators on Hilbert Spaces

As we have seen in our study of convexity, the geometry of topological vector spaces—even those of Banach spaces—can be significantly different from that of the Euclidean space. What is missing here is the *orthogonality properties* of the Euclidean space, with which we can recover much of the Euclidean geometry in the abstract setting. The added structure lends itself to a detailed analysis of the operators between these spaces, giving rise to *spectral theory*.

1. The Geometry of Hilbert Spaces

1.1. Inner Products. Recall that two vectors v and w in $\mathbb{R}^n$ are *orthogonal* if their dot product $v \cdot w = \sum v_i w_i$ is zero. If we are to speak of orthogonality in a broader context, it is first necessary to generalize the dot product.

Definition 1.1. An *inner-product space* is a complex vector space X with a map $\langle \cdot, \cdot \rangle : X \times X \to \mathbb{C}$ such that

(i) $\langle \alpha x + \beta y, z \rangle = \alpha \langle x, z \rangle + \beta \langle y, z \rangle$,
(ii) $\langle y, x \rangle = \overline{\langle x, y \rangle}$,
(iii) $\langle x, x \rangle \geq 0$, and
(iv) $\langle x, x \rangle = 0$ if and only if $x = 0$

for all $\alpha, \beta \in \mathbb{C}$ and $x, y, z \in V$.

(i) and (ii) imply that $\langle \cdot, \cdot \rangle$ is a *Hermitian form*, and (iii) implies that $\langle \cdot, \cdot \rangle$ is a *positive-definite form*; a positive-definite Hermitian form is a *semi-inner product.* If X is a real vector space, then Hermiticity reduces to *bilinearity*, i.e.,

$$\langle \alpha x + \beta y, z \rangle = \alpha \langle x, z \rangle + \beta \langle y, z \rangle \text{ and } \langle x, \alpha y + \beta z \rangle = \alpha \langle x, y \rangle + \beta \langle x, z \rangle,$$

and *symmetry*

$$\langle x, y \rangle = \langle y, x \rangle.$$

Therefore, a semi-inner product on a real vector space is a positive-definite, symmetric bilinear form, and (iv) turns it into a *real inner product.* We prefer to work with complex inner products, as they possess certain advantages over real inner products that we shall see in later sections.

By the positive definiteness of the inner product, we see that

$$\|x\| = \sqrt{\langle x, x \rangle}$$

is nonnegative, and hermiticity implies that it is homogeneous just like a norm. We shall show that $\| \cdot \|$ is indeed a norm in due course, but, for now, we harness the basic properties of the inner product $\langle \cdot, \cdot \rangle$ and the associated "norm" $\| \cdot \|$. The first is a simple computational principle that directly generalizes the law of cosines in $\mathbb{R}^2$.

PROPOSITION 1.2 (Generalized law of cosines). *For all* $x, y \in X$,

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

PROOF. Hermiticity. □

In classical Euclidean geometry, the Pythagorean theorem could be deduced as a special case of the law of cosines in which the two sides are perpendicular to each other. We now generalize the notion of perpendicularity and use it to establish a Pythagorean identity for inner-product spaces.

DEFINITION 1.3. Two vectors $x, y \in X$ are *orthogonal* if $\langle x, y\rangle = 0$. We write $x \perp y$ to denote that x and y are orthogonal.

PROPOSITION 1.4 (Generalized Pythagorean theorem). *For all* $x_1, \ldots, x_n \in X$ *such that* $x_i \perp x_j$ *for all* $1 \leq i < j \leq n$,

$$\left\|\sum_{i=1}^n x_i\right\|^2 = \sum_{i=1}^n \|x_i\|^2.$$

PROOF. For $n = 2$, this is a special case of the generalized law of cosines. For $n > 2$, it suffices to apply the $n = 2$ case repeatedly. □

The next result is the foundational inequality of Schwarz, which generalizes Cauchy's inequality in $\mathbb{R}^n$.

PROPOSITION 1.5 (Schwarz's inequality). $|\langle x, y\rangle| \leq \|x\|\|y\|$ *for all* $x, y \in X$, *and the equality holds if and only if* x *is a scalar multiple of* y.

The key idea is to decompose x into two orthogonal vectors and to obtain an estimate via the Pythagorean theorem. This decomposition is (almost) always possible, and we take this up in Theorem 1.18.

PROOF. We assume without loss of generality that $y \neq 0$ and set $\hat{y} = \|y\|^{-1}y$, so that $\|\hat{y}\| = 1$. We let

$$u = \langle x, \hat{y}\rangle\hat{y} \quad \text{and} \quad v = x - \langle x, \hat{y}\rangle\hat{y},$$

so that $x = u + v$ and $u \perp v$. By the Pythagorean theorem, we have

$$\|x\|^2 = \|u\|^2 + \|v\|^2 = \|\langle x, \hat{y}\rangle\hat{y}\|^2 + \|x - \langle x, \hat{y}\rangle\hat{y}\|^2.$$

Combined with the positive-definiteness of the inner product, the above identity implies that

$$\|x\|^2 \geq \|\langle x, \hat{y}\rangle\hat{y}\|^2 = |\langle x, \hat{y}\rangle|^2\|\hat{y}\|^2 = |\langle x, \hat{y}\rangle|^2,$$

where the equality holds if and only if $\|x - \langle x, \hat{y}\rangle\hat{y}\|^2$. The inequality is precisely Schwarz's inequality; the equality condition is equivalent to the attainment of the identity

$$x = \langle x, \hat{y}\rangle\hat{y} = \left(\frac{\langle x, \|y\|^{-1}y\rangle}{\|y\|}\right) y = \|y\|^{-2}\langle x, y\rangle y. \tag{1.6}$$

Therefore, if the equality in Schwarz's inequality holds, then (1.6) holds, whence, in particular, x is a scalar multiple of y. Conversely, if $x = \lambda y$, for some scalar λ, then

$$\|y\|^{-2}\langle x, y\rangle y = \|y\|^{-2}\langle \lambda y, y\rangle y = \lambda y = x,$$

which is precisely (1.6). □

With Schwarz's inequality at hand, we can now deduce that $\|\cdot\|$ is indeed a norm.

PROPOSITION 1.7. *$\|x\| = \sqrt{\langle x, x\rangle}$ is a norm on X.*

PROOF. Homogeneity follows hermiticity; positive-definiteness is inherited. To verify the triangle inequality, we observe that

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

by the law of cosines. Applying Schwarz's inequality, we see that

$$2\operatorname{Re}\langle x, y\rangle = \langle x, y\rangle + \overline{\langle x, y\rangle} \leq \|x\|\|y\| + \overline{\|x\|\|y\|} = 2\|x\|\|y\|,$$

whence we have

$$\|x+y\|^2 \leq (\|x\| + \|y\|)^2,$$

as desired. □

We now define the single most important object in functional analysis.

DEFINITION 1.8. A *Hilbert space* is a complete inner-product space, in the sense that the norm associated with the inner product induces a complete metric.

EXAMPLE. The Euclidean spaces $\mathbb{R}^n$ and $\mathbb{C}^n$ with the usual dot products are Hilbert spaces.

EXAMPLE. The space $L^2(X, \mu)$ of square-integrable functions on a measure space (X, μ) with the inner product

$$\langle f, g\rangle_2 = \int_X f\overline{g}\, d\mu$$

is a Hilbert space.

EXAMPLE. Let $\mathcal{T}$ be the set of all *trigonometric polynomials*

$$p(x) = \sum_{n=1}^{N} p_n e^{2\pi i n x}$$

on $[0, 1]$, with the inner product

$$\langle p, q\rangle = \sum_{n=1}^{N} p_n \overline{q_n}.$$

is an inner-product subspace of $L^2([0, 1])$. The space $\mathcal{T}_N$ of trigonometric polynomials of degree at most N is a closed subspace of $\mathcal{T}$ and hence a Hilbert subspace of $L^2([0, 1])$.

EXAMPLE. Let $K \in \mathscr{C}_c(\mathbb{R}^2)$ be nonnegative and *symmetric*, i.e., $K(x, y) = K(y, x)$. The *weighted L^2 inner product*

$$\langle f, g\rangle_K = \iint f(x)\overline{g(y)} K(x, y)\, dx\, dy$$

is a complete inner product on $L^2(\mathbb{R}^2)$. Here we only show that

$$\|f\|_K = \left(\iint |f(x)|^2 K(x, y)\, dx\, dy\right)^{1/2}$$

is a complete norm. To this end, we let $(f_n)_{n=1}^\infty$ be a Cauchy sequence with respect to $\|\cdot\|_K$ and extract a subsequence $(f_{n_k})_{k=1}^\infty$ such that $\|f_{n_k} - f_{n_{k+1}}\|_K < 2^{-k}$. This subsequence converges pointwise almost everywhere to

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

which is in $L^2(\mathbb{R}^2)$. Now, $|f_{n_k} - f|^2 K \in L^1(\mathbb{R}^2)$, and so

$$\begin{aligned}\lim_{k\to\infty} \|f_{n_k} - f\|_K &= \lim_{k\to\infty} \iint |f_{n_k}(x,y) - f(x,y)|^2 K(x,y)\,dx\,dy \\ &= \iint \lim_{k\to\infty} |f_{n_k}(x,y) - f(x,y)|^2 K(x,y)\,dx\,dy \\ &= 0\end{aligned}$$

by the dominated convergence theorem. The desired result now follows.

1.2. Orthogonality. Having reviewed the definition and basic examples of a Hilbert space, we now turn to the promised discussion on orthogonality. The next result encodes the orthogonality property of inner product spaces in a norm identity, attributed to Pascual Jordan and John von Neumann in [**Tes09**].

THEOREM 1.9 (Parallelogram law / Polarization identity). *If X is an inner-product space, then the* **parallelogram law**

$$\|x+y\|^2 + \|x-y\|^2 = 2(\|x\|^2 + \|y\|^2)$$

and the **polarization identity**

$$\langle x, y\rangle = \frac{1}{4}\left(\|x+y\|^2 - \|x-y\|^2 + i\|x+iy\|^2 - i\|x-iy\|^2\right)$$

hold. Conversely, if X is a normed linear space that satisfies the parallelogram law, then the quadratic form given by the polarization identity is an inner product on X.

PROOF. If X is an inner-product space, then the parallelogram law follows from the law of cosines (Proposition 1.2) and the polarization identity from the hermiticity of the inner product. Conversely, we assume that X is a normed linear space satisfying the parallelogram law and consider the quadratic form $\langle\cdot,\cdot\rangle$ given by the polarization identity. Observe that

$$\begin{aligned}4\langle x,y\rangle &= \|x+y\|^2 - \|x-y\|^2 + i\|x+iy\|^2 - i\|x-iy\|^2 \\ &= \|y+x\|^2 - \|y-x\|^2 + i\|y-ix\|^2 - i\|y+ix\|^2 \\ &= \overline{(\|y+x\|^2 - \|y-x\|^2 - i\|y-ix\|^2 + i\|y+ix\|^2)} \\ &= 4\overline{\langle y,x\rangle}.\end{aligned}$$

In particular, $\langle x,x\rangle = \overline{\langle x,x\rangle}$, and so

$$4\langle x,x\rangle = 4\,\mathrm{Re}\langle x,x\rangle = \|x+x\|^2 + \|x-x\|^2 = 4\|x\|^2,$$

whence $\langle\cdot,\cdot\rangle$ is positive-definite. In particular, $\langle x,x\rangle = 0$ if and only if $x = 0$.

It remains to show that $\langle\cdot,\cdot\rangle$ is linear in the first variable. By the parallelogram law, we have

$$\begin{aligned} & 4(\langle x,z\rangle + \langle y,z\rangle) \\ = \; & \|(x+y)+2z\|^2 - \|(x+y)-2z\|^2 + i\|(x+y)+2iz\|^2 - \|(x+y)-2iz\|^2 \\ = \; & 8(\langle (x+y)/2, z\rangle). \end{aligned}$$

In particular, $\langle x,z\rangle = \frac{1}{2}\langle 2x,z\rangle$, and so

$$\langle x,z\rangle + \langle y,z\rangle = 2\langle (x+y)/2, z\rangle = \langle x+y,z\rangle. \tag{1.10}$$

Iterating these two identities imply that

$$\sum_{n=-N}^{N} a_n 2^n \langle x,z\rangle = \left\langle \left(\sum_{n=-N}^{N} a_n 2^n\right) x, z \right\rangle \tag{1.11}$$

for all $N \in \mathbb{N}$ and $a_n \in \mathbb{N}$.

Now, the norm is continuous, and so $\langle\cdot,\cdot\rangle$ is continuous in the first variable. If this claim holds, then (1.11) implies that

$$r\langle x,z\rangle = \langle rx,z\rangle \tag{1.12}$$

for all $r \geq 0$. Simple computations show that $-\langle x,z\rangle = \langle -x,z\rangle$ and $i\langle x,z\rangle = \langle ix,z\rangle$, whence (1.12) now implies that

$$\lambda\langle x,z\rangle = \langle \lambda x,z\rangle.$$

This, combined with (1.10), establishes the hermiticity of $\langle\cdot,\cdot\rangle$, thereby showing that $\langle\cdot,\cdot\rangle$ is an inner product. □

The continuity property used in the above proof holds in general:

PROPOSITION 1.13. *The inner product is continuous as a map from $\langle\cdot,\cdot\rangle$ into $\mathbb{C}$.*

PROOF. If $x_n \to x$ and $y_n \to y$, then

$$\lim_{n\to\infty} |\langle x,y\rangle - \langle x_n,y_n\rangle| \leq \lim_{n\to\infty} \|x-x_n\|\|y-y_n\| = 0$$

by Schwarz's inequality (Proposition 1.5). □

With the parallelogram law and the polarization identity at hand, we can now construct the *Hilbert completion* of an inner-product space. Given an inner-product space X, we take its Banach completion (Chapter 1, Theorem 2.10) $\tilde{X}$. Since the norm is continuous, the parallelogram law on X evidently carries over to $\tilde{X}$, whence the inner product given by the polarization identity turns $\tilde{X}$ into a Hilbert space. In light of this, analysts often refer to inner-product spaces as *pre-Hilbert spaces*.

EXAMPLE. $\mathscr{C}^1[0,1]$ with the inner product

$$\langle f,g\rangle = \int_0^1 f(x)\overline{g(x)}\,dx + \int_0^1 f'(x)\overline{g'(x)}\,dx$$

is a pre-Hilbert space. Its Hilbert completion is the *Sobolev space of order 1*, denoted by H^1.

EXAMPLE. While all norms on $\mathbb{R}^n$ are equivalent (Chapter 1, Theorem 2.15), not too many of them admit an inner product. For example, the maximum norm

$$\|x\|_\infty = \max_{1 \leq i \leq n} |x_i|$$

does not satisfy the parallelogram law. Indeed,

$$\|(2,0)+(0,1)\|_\infty^2 + \|(2,0)-(0,1)\|_\infty^2 = 8 \neq 10 = 2(\|(2,0)\|_\infty^2 + \|(0,1)\|_\infty^2),$$

and this counterexample can easily be extended to handle the $\mathbb{R}^n$ case.

We now exploit the orthogonality property to study the geometry of Hilbert spaces. First, we generalize the notion of *vectors perpendicular to a plane* used in the study of three-dimensional Euclidean geometry.

DEFINITION 1.14. Given a subset E of an inner product space V, we define the *orthogonal complement of* E to be the set

$$E^\perp = \{x \in V : x \perp y \text{ for all } y \in E\}.$$

Recall that a *projection* on a vector space V is an idempotent linear endomorphism, i.e., a linear transformation $P : V \to V$ such that $P^2 = P$. Note that $P|_{\operatorname{im} P}$ is the identity operator I, and that we have the direct-sum decomposition

$$V = P(V) \oplus (I-P)(V),$$

viz., each $x \in V$ can be written as the sum $x = y + z$ with $y \in P(V)$ and $z \in (I-P)(V)$ in precisely one way: namely, $y = Px$ and $z = (I-P)x = x - Px$. In view of the direct sum decomposition, we often refer to P as *the projection of* V *onto* $P(V)$ *along* $(I-P)(V)$.

One of the most important results in Hilbert-space theory is the existence of *orthogonal projections*:

DEFINITION 1.15. Let M be a closed linear subspace of $\mathcal{H}$. The *orthogonal projection* of $\mathcal{H}$ onto M is the projection P of $\mathcal{H}$ onto M along $M^\perp$.

EXERCISE 1.16 (Problem 18). P is an orthogonal projection if and only if $(\operatorname{im} P)^\perp = \ker P$.

LEMMA 1.17. *If E is a subset of an inner-product space X, then $E^\perp$ is a closed linear subspace of X.*

PROOF OF LEMMA. It is clear that $E^\perp$ is a linear subspace of X. To show that $E^\perp$ is closed, it suffices to note that

$$E^\perp = \bigcap_{y \in E} \{y\}^\perp,$$

and that each $\{y\}^\perp$ is the preimage of $\{0\}$ of the continuous mapping $x \mapsto \langle x, y \rangle$ (Proposition 1.13). □

THEOREM 1.18 (Existence of orthogonal projections). *Let $\mathcal{H}$ be a Hilbert space. If M is a closed, convex subset of $\mathcal{H}$, then there exists a linear operator each $x \in \mathcal{H}$ admits precisely one $x_M \in M$ such that*

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

If M is, in addition, a linear subspace of $\mathcal{H}$, then there exists a projection P of $\mathcal{H}$ along $M^\perp$ onto M, called the orthogonal projection *of $\mathcal{H}$ onto M. In particular,*

$$\mathcal{H} = M \oplus M^\perp.$$

We remark that the hypothesis that M is a closed *convex* linear subspace is superfluous, as every linear subspace is convex.

Proof of theorem. Assume for now that M is a closed, convex subset of $\mathcal{H}$. Let $(y_n)_{n=1}^\infty$ be a sequence in M such that

$$\lim_{n\to\infty} \|x - y_n\| = d(x, M) = \inf_{y\in M} \|x - y\|.$$

Since M is convex, $(y_n + y_m)/2 \in M$, and so $\|x - (y_n + y_m)/2\| \geq d(x, M)$. The parallelogram law (Theorem 1.9) implies that

$$\begin{aligned}
4d(x, M)^2 &\leq 4\|x - (y_n + y_m)/2\|^2 \\
&\leq 4\|x - (y_n + y_m)/2\|^2 + \|y_n - y_m\|^2 \\
&= \|(x - y_m) + (x - y_n)\|^2 + \|(x - y_m) - (x - y_n)\|^2 \\
&= 2(\|x - y_m\|^2 + \|x - y_n\|^2),
\end{aligned}$$

whence

$$\begin{aligned}
\lim_{n,m\to\infty} \|y_n - y_m\|^2 &\leq \lim_{n,m\to\infty} 2(\|x - y_m\|^2 + \|x - y_n\|^2) - 4d(x, M)^2 \\
&= 2(d(x, M)^2 + d(x, M)^2)) - 4d(x, M)^2 \\
&= 0.
\end{aligned}$$

It follows that $(y_n)_{n=1}^\infty$ is Cauchy.

We invoke the completeness of $\mathcal{H}$ to find $x_M = \lim_n y_n$. By construction, $\|x - x_M\| = d(x, M)$. To show that the minimizer is unique, we let $y \in M$ be another minimizer, i.e., $\|x - y\| = d(x, M)$. This, in particular, implies that $\|x - (x_M + y)/2\| = d(x, M)$, and so

$$\begin{aligned}
4d(x, M)^2 &= 2(d(x, M)^2 + d(x, M)^2) \\
&= 2(\|x - x_M\|^2 + \|x - y\|^2) \\
&= \|(x - x_M) + (x - y)\|^2 + \|(x - x_M) - (x - y)\|^2 \\
&= 4\|x - (x_M + y)/2\|^2 + \|y - x_M\|^2 \\
&= 4d(x, M)^2 + \|y - x_M\|^2
\end{aligned}$$

by the parallelogram law (Theorem 1.9). It follows that $\|y - x_M\|^2 = 0$, whence $y = x_M$.

We now assume that M is a closed linear subspace. Define an operator $P : \mathcal{H} \to \mathcal{H}$ by setting $Px = x_M$ for each $x \in \mathcal{H}$. We claim that P is a projection of $\mathcal{H}$ along $M^\perp$ onto M. To this end, we first note that $P^2x = (x_M)_M$ is evidently x_M, for the unique distance minimizer for any element of M is the element itself. Therefore, $P^2 = P$, and the argument also shows that $M \subseteq P(M)$. and P is a projection. Secondly, we have shown above that the distance minimizer is always in M, and so $P(M) \subseteq M$. It follows that P is a projection onto M.

Lastly, we must check that $(I - P)x = x - x_M$ is an element of $M^\perp$, which is a closed linear subspace by Lemma 1.17. This amounts to showing that $\langle x - x_M, y\rangle$

vanishes for all $y \in M^\perp$. To this end, we fix $y \in M^\perp$ and consider $f : \mathbb{R} \to [0, \infty)$ given by the formula

$$f(t) = \|x - x_M + ty\|^2.$$

By the law of cosines (Proposition 1.2), we can write

$$f(t) = \|x - x_M\|^2 + t^2\|y\|^2 + 2t\,\mathrm{Re}\langle x - x_M, y\rangle.$$

Now, we observe that

$$f(t) = \|x - (x_M - ty)\|^2$$

and $x_M - ty \in M$, whence the distance function f is minimized when $x_M - ty = x_M$, or $t = 0$. Therefore,

$$\|x - x_M\|^2 = f(0) \leq f(t) = \|x - x_M\|^2 + t^2\|y\|^2 + 2t\,\mathrm{Re}\langle x - x_M, y\rangle$$

for all $t \in \mathbb{R}$. This is only possible if $\langle x - x_M, y\rangle = 0$. It now follows that P is a projection along $M^\perp$, thus completing the proof. □

An immediate corollary of the direct-sum decomposition is that proper closed subspaces must have a nontrivial orthogonal complement.

Corollary 1.19 (Existence of nontrivial orthogonal complements). *If M is a closed proper linear subspace of $\mathcal{H}$, then there exists a nonzero vector $y \in \mathcal{H}$ such that $\langle x, y\rangle = 0$ for all $x \in M$.* □

For each $u \in M^\perp$, we obtain a linear functional $l_u : \mathcal{H} \to \mathbb{C}$ given by $l_u(x) = \langle x, u\rangle$. By the continuity of the inner product (Proposition 1.13), l_u is bounded. Furthermore, l_y vanishes on M, whence l_u is an an element of the *annihilator* of M, viz., the collection of bounded linear functionals on that vanish on M. In this sense, Corollary 1.19 can be thought of as the existence of nontrivial annihilators that we have established in Chapter 1 (Corollary 5.10). Direct comparisons bewteen orthogonal complements and annihilators are made in Appendix A, Section 7.

The construction of the bounded linear functional l_u evidently applies to any $u \in \mathcal{H}$, and this shows that there are plenty of bounded linear functionals on $\mathcal{H}$. Shockingly, they fill up $\mathcal{H}^*$, as we shall see in the following result.

Theorem 1.20 (Riesz representation). *Every bounded linear functional $l : \mathcal{H} \to \mathbb{C}$ can be written as an action of the inner product against a vector $u \in \mathcal{H}$, i.e.,*

$$lx = \langle x, u\rangle$$

for all $x \in \mathcal{H}$. Furthermore, this u is unique, and $\|l\| = \|u\|$, whence $\mathcal{H}$ is isometrically isomorphic to $\mathcal{H}^*$ and is thus reflexive.

We remark that the natural notion of isomorphisms between Hilbert spaces is that of a *unitary isomorphism*:

Definition 1.21. A linear operator $T : X \to Y$ between two inner-product spaces is *unitary* if $\langle Tx, Ty\rangle = \langle x, y\rangle$ for all $x, y \in X$. T is a *unitary isomorphism* if T is a unitary operator and a linear isomorphism.

Nevertheless, the polarization identity (Theorem 1.9) implies that a linear isomorphism between pre-Hilbert spaces is unitary if and only if it is isometric. In practice, therefore, we never have to check whether the inner product is preserved.

Proof of theorem. If l is the zero functional, then 0 is clearly the only vector that does the job. If l is nontrivial, then $\ker l$ is a closed proper linear subspace of $\mathcal{H}$, whence $(\ker l)^\perp$ is nontrivial by Corollary 1.19. We pick a unit vector $z \in (\ker l)^\perp$ and set $u = (\overline{lz})z$. Note that $(lx)z - (lz)x \in \ker l$ for all $x \in \mathcal{H}$, and so

$$0 = \langle (lx)z - (lz)x, z\rangle = (lx)\langle z, z\rangle - \langle x, (\overline{lz})z\rangle = lx - \langle x, u\rangle,$$

or $lx = \langle x, u\rangle$. Finally, we observe that

$$\|u\| = |\overline{(lz)}| = |lz| \leq \|l\|\|z\| = \|l\|$$

and

$$|lz| = |\overline{lz}|\|z\| = \|u\|\|z\|;$$

the latter implies that $\|u\| \geq \|l\|$, whence $\|u\| = \|l\|$. □

We pause to reflect on a particularly useful consequence of the orthogonality property of the Euclidean space: $\mathbb{R}^n$ is equipped with n mutually orthogonal axes that admit a particularly convenient coordinate representation of n-vectors. Our next goal is to recover this coordinate representation in the Hilbert-space setting. The proper way to set up "mutually orthogonal axes" is via an orthonormal set vectors.

Definition 1.22. A set of vectors $\{u_\alpha\}_{\alpha\in\mathcal{I}}$ in an inner-product space X is *orthonormal* if $\|u_\alpha\| = 1$ for all $\alpha \in \mathcal{I}$ and $u_\alpha \perp u_\beta$ for all $\alpha \neq \beta$.

The next result shows that a finite orthonormal set is, in general, not sufficient for a full-blwon coordinate representation.

Proposition 1.23 (Finite Bessel's inequality). *If $\{u_1, \ldots, u_N\}$ is an orthonormal set in an inner-product space X, then*

$$\sum_{n=1}^N |\langle x, u_n\rangle|^2 \leq \|x\|^2$$

for all $x \in X$, and the equality holds if and only if

$$x = \sum_{n=1}^N \langle x, u_n\rangle u_n.$$

Proof. Since $(\sum\langle x, u_n\rangle u_n)$ is orthogonal to $x - \sum\langle x, u_n\rangle$, the Pythagorean theorem (Proposition 1.4) implies that

$$\|x\|^2 = \left\|\sum_{n=1}^N \langle x, u_n\rangle u_n\right\|^2 + \left\|x - \sum_{n=1}^N \langle x, u_n\rangle u_n\right\|^2.$$

Now, the vectors $\langle x, u_1\rangle u_1, \ldots, \langle x, u_n\rangle u_n$ are mutually orthogonal, and so another application of the Pythagorean theorem yields

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

The desired result now follows. □

We now recall that the sum of an infinite set $\{r_\alpha\}_{\alpha \in \mathcal{I}}$ of nonnegative real numbers is defined as the supremum of all the finite partial sums. If $\sum_\alpha r_\alpha$ is finite, then all but countably many elements in $\{r_\alpha\}$ must be zero, for otherwise we can extract a monotonically increasing sequence $(r_n)_{n=1}^\infty$ in $\{r_\alpha\}$. With this observation, we can extend finite Bessel's inequality to infinite orthonormal sets.

COROLLARY 1.24 (Bessel's inequality). *If $\{u_\alpha\}_{\alpha \in \mathcal{I}}$ is an orthonormal set in an inner-product space, then*

$$\sum_{\alpha \in \mathcal{I}} |\langle x, u_\alpha \rangle|^2 \leq \|x\|^2$$

for all $x \in X$. In particular, $\{\alpha : \langle x, u_\alpha \rangle \neq 0\}$ is a countable set. □

In order to recover the equality condition we have established for finite Bessel's inequality, we must deal with infinite sums of vectors. The appropriate notion of convergence for us is that of *unconditional convergence*: a countable sum

$$\sum_{n=1}^\infty x_n$$

of vectors in a topological vector space *converges uncounditionally* if there exists a vector x such that the sequence of partial sums converges to x for all possible reordering of indices. Note that unconditional convergence of vectors in a normed linear space is *not* equivalent to the absolute convergence

$$\sum_{n=1}^\infty \|x_n\| < \infty;$$

in fact, a theorem of Dvoretzky and Rogers ([DR50], Theorem 1) shows that unconditional convergence implies absolute convergence in a Banach space if and only if the space is finite-dimensional.

With this in mind, we deal with the task of recovering the equality condition:

PROPOSITION 1.25 (Definition of orthonormal basis). *Let $\mathscr{U} = \{u_\alpha\}_{\alpha \in \mathcal{I}}$ be an orthonormal set in a Hilbert space $\mathcal{H}$. The following are equivalent:*

(a) $\mathscr{U}$ is an **orthonormal basis**, *viz.,*

$$x = \sum_{\alpha \in \mathcal{I}} \langle x, u_\alpha \rangle u_\alpha$$

for all $x \in \mathcal{H}$, where the sum converges unconditionally;

(b) **Parseval's identity** *holds for $\mathscr{U}$, viz.,*

$$\|x\|^2 = \sum_{\alpha \in \mathcal{I}} |\langle x, u_\alpha \rangle|^2$$

for all $x \in \mathcal{H}$.

(c) $\mathscr{U}$ is **complete**, *i.e. $\mathscr{U}^\perp = \{0\}$.*

PROOF. $(a) \Rightarrow (b)$. By Bessel's inequality, the set $\mathcal{I}_x = \{\alpha \in \mathcal{I} : \langle x, u_\alpha \rangle \neq 0\}$ is countable. We index the elements of $\mathcal{I}_x$ by natural numbers: $\mathcal{I}_x = \{\alpha_n\}_{n \in \mathbb{N}}$. Parseval's identity is now obtained by considering the limiting case of the identity

$$\|x\|^2 = \sum_{n=1}^N |\langle x, u_n \rangle|^2 + \left\| x - \sum_{n=1}^N \langle x, u_n \rangle u_n \right\|^2$$

from the proof of finite Bessel's inequality.

$(b) \Rightarrow (c)$. If $\mathscr{U}^\perp$ contains a nonzero vector x, then

$$\|x\|^2 > 0 = \sum_{\alpha \in \mathcal{I}} |\langle x, u_\alpha \rangle|^2.$$

$(c) \Rightarrow (a)$. By Bessel's inequality, the set $\mathcal{I}_x = \{\alpha \in \mathcal{I} : \langle x, u_\alpha \rangle \neq 0\}$ is countable. We index the elements of $\mathcal{I}_x$ by natural numbers: $\mathcal{I}_x = \{\alpha_n\}_{n \in \mathbb{N}}$. Bessel's inequality implies that $\sum |\langle x, u_\alpha \rangle|^2 < \infty$, whence it follows from the Pythagorean theorem (Proposition 1.4) that

$$\lim_{N \to \infty} \left\| \sum_{n=N}^{N+m} \langle x, u_n \rangle u_n \right\|^2 = \lim_{N \to \infty} \sum_{n=N}^{N+m} |\langle x, u_n \rangle|^2 = 0$$

for all $m \in \mathbb{N}$. By completeness of $\mathcal{H}$, the series $\sum \langle x, u_n \rangle u_n$ converges. Now,

$$\left\langle x - \left(\sum_{n=1}^{\infty} \langle x, u_n \rangle u_n \right), u_n \right\rangle = \langle x, u_n \rangle - \langle x, u_n \rangle = 0$$

for all $n \in \mathbb{N}$, whence by the completness of $\mathscr{U}$ we conclude that $x - \sum \langle x, u_n \rangle u_n = 0$, or $x = \sum \langle x, u_n \rangle u_n$. □

With the new tools available, we can now establish swiftly the coordinate representation we have advertised.

THEOREM 1.26. *Every Hilbert space has an orthonormal basis.*

SKETCH OF PROOF. Apply Zorn's lemma on the collection of orthonormal sets ordered by inclusion: the maximal element is complete. □

The existence of an orthonormal basis, combined with Parseval's identity, implies the following characterization of Hilbert spaces.

COROLLARY 1.27. *Every Hilbert space is unitarily isomorphic to $l^2(\mathcal{I})$ for some set $\mathcal{I}$ with the counting measure.*

SKETCH OF PROOF. Let $\mathcal{H}$ be a Hilbert space. By Theorem 1.26, $\mathcal{H}$ has an orthonormal basis $\{u_\alpha\}_{\alpha \in \mathcal{I}}$. For each $x \in \mathcal{H}$, we define $f_x \in l^2(\mathcal{I})$ by setting

$$f_x(\alpha) = \langle x, u_\alpha \rangle.$$

The mapping $x \mapsto f_x$ is then an isometric isomorphism. □

1.3. Separability. While theoretically satisfying, the coordinate presentation give by Theorem 1.26 is unwieldy in many cases. In practice, it is useful to have a countable orthonormal basis, so as to dispense with the infinite-sums-as-suprema business. To this end, we recall that a topological space is *separable* if it contains a countable dense subset. An orthonormal set $\mathscr{U}$ is an orthonormal basis if and only if the rational linear combination of $\mathscr{U}$ is dense, and so we expect to see a strong connection between separability and the existence of a countable orthonormal basis.

THEOREM 1.28. *A Hilbert space $\mathcal{H}$ has a countable orthonormal basis if and only if $\mathcal{H}$ is separable.*

Proof. The equivalence is trivial if $\mathcal{H}$ is finite-dimensional, hence we assume that $\mathcal{H}$ is infinite-dimensional.

($\Leftarrow$) Let $\mathscr{U}$ be a countable orthonormal basis of $\mathcal{H}$ and take the collection $\mathscr{B}$ of rational linear combinations of $\mathscr{U}$. $\mathscr{B}$ is dense in span $\mathscr{U}$, and $\mathscr{U}$ is dense in $\mathcal{H}$ by the completeness characterizaiton of orthonormal bases (Proposition 1.25). Since $\mathscr{B}$ is countable, it follows that $\mathcal{H}$ is separable.

($\Rightarrow$) Let $\{x_m\}_{m\in\mathbb{N}}$ be a dense subset of $\mathcal{H}$. We build a countable linearly independent subset $\{e_n\}_{n\in\mathbb{N}}$ of $\{x_m\}_{m\in\mathbb{N}}$ as follows: let $e_1 = x_1$ and $m_1 = 1$; for each $n > 1$, we let m_n be the smallest natural number bigger than m_{n-1} such that $\{e_1, \ldots, e_{n-1}, x_{m_{n-1}+1}, \ldots, x_{m_n}\}$ is linearly independent. This process must continue indefinitely, as $\operatorname{span}\{x_m\} = \operatorname{span}\{ e_n \}$. In particular, $\overline{\operatorname{span}\{ e_n \}} = \mathcal{H}$.

The bulk of the remaining work is to establish the infinite-dimensional generalization of a familiar tool from linear algebra:

Lemma 1.29 (Gram-Schmidt process). *If $\{e_n\}_{n\in\mathbb{N}}$ is a linearly independent subset of $\mathcal{H}$, then there exists an orthonormal set $\{u_N\}_{N\in\mathbb{N}}$ such that* $\operatorname{span}\{e_n\} = \operatorname{span}\{u_N\}$ *and that each u_N is a linear combination of $e_1, \ldots, e_N$.*

Proof of lemma. We first note that $\|e_1\| > 0$ by the linear independence of $\{e_n\}$. Set

$$u_1 = \frac{e_1}{\|e_1\|}$$

and, for each $N > 1$, define

$$u_{N+1} = \frac{e_{N+1} - \sum_{n=1}^{N} \langle e_{N+1}, e_n\rangle e_n}{\left\| e_{N+1} - \sum_{n=1}^{N} \langle e_{N+1}, e_n\rangle e_n \right\|}.$$

At each step, the linear independence of $\{e_n\}$ implies that u_{N+1} is well-defined and $\{u_1, \ldots, u_{N+1}\}$ is an orthonormal set whose linear span agrees with the linear span of $\{e_1, \ldots, e_{N+1}\}$. The lemma now follows. □

We now observe that Proposition 1.17 implies

$$\{u_N\}^\perp = (\operatorname{span}\{u_N\})^\perp = (\operatorname{span}\{e_n\})^\perp = (\overline{\operatorname{span}\{e_n\}})^\perp = \mathcal{H}^\perp = \{0\},$$

whence by the completeness characterization of orthonormal bases (Proposition 1.25) $\{u_N\}$ is an orthonormal basis. □

Corollary 1.30. *All infinite-dimensional separable Hilbert spaces are unitarily isomorphic to one another.*

Proof. This follows at once from Theorem 1.28 and Corollary 1.27. □

Note that we do not need Zorn's lemma to develop the theory of separable Hilbert spaces, which, in turn, makes the theory more *constructive*. The reliance of the above collary on Theorem 1.26 (from which Corollary 1.27 follows) is merely for convenience. Duplicating the proof of Corollary 1.27 with Theorem 1.28 instead of 1.26 allows us to avoid Zorn's lemma.

We conclude the section by presenting examples of commonplace separable Hilbert spaces and their orthonormal bases.

Example (Kronecker delta basis of l^2). Let

$$e_n(m) = \begin{cases} 1 & \text{if } n = m; \\ 0 & \text{otherwise} \end{cases}$$

for all $n, m \in \mathbb{Z}$. Then $\{e_n\}_{n\in\mathbb{Z}}$ is an orthonormal basis in l^2.

Example (L^2 theory of Fourier series). The *Fourier basis* $\{e_n = e^{inx}\}_{n\in\mathbb{Z}}$ on $[-\pi, \pi]$ is an orthonormal basis on $L^2([-\pi,\pi])$. The basis property yields the *Fourier inversion formula*

$$f(x) = \sum_{n\in\mathbb{Z}} \langle f, e_n\rangle e_n = \sum_{n\in\mathbb{Z}} \left(\frac{1}{2\pi} \int_{-\pi}^{\pi} f(t) e^{-int}\, dt \right) e^{inx},$$

and Parseval's identity yields the *Plancherel identity*

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} |f(x)|^2\, dx = \sum_{n\in\mathbb{Z}} \left| \frac{1}{2\pi} \int_{-\pi}^{\pi} f(t) e^{-int}\, dt \right|^2 .$$

Example (One-dimensional dyadic harmonic analysis). A *dyadic interval* in $\mathbb{R}$ is an interval of the form

$$[m2^{-k}, (m+1)2^{-k})$$

for some $m, k \in \mathbb{Z}$. Given a dyadic interval $I = [m2^{-k}, (m+1)2^{-k})$, we let

$$I_L = [m2^{-k}, (m+1/2)2^{-k}) \quad \text{and} \quad I_R = [(m+1/2)2^{-k}, (m+1)2^{-k})$$

and define the *Haar function associated with the interval I*

$$h_I = m(I)^{-1/2} \left(\chi_{I_L} - \chi_{I_R}\right).$$

The collection

$$\mathscr{H} = \{h_I : I \text{ is a dyadic interval in } \mathbb{R}\}$$

is a countable orthonormal basis in $L(\mathbb{R})$, called the *Haar wavelet basis* of $L^2(\mathbb{R})$.

In general, $\varphi \in L^2(\mathbb{R})$ is called a *wavelet* if the set $\{\varphi_{m,k} : m, k \in \mathbb{Z}\}$ consisting of the functions

$$\varphi_{m,k}(x) = 2^{-m/2} \varphi(2^m x - n)$$

is an orthonormal basis of $L^2(\mathbb{R})$. Observe that the Haar function $h_{[0,1)}$ is a wavelet. It can be shown that a smooth wavelet exists.

Dyadic harmonic analysis is deeply intertwined with Littlewood-Paley theory of frequency localization: see Chapter 8 of [**MS13**] or Chapter 4 of [**Gra10a**] for details.

Example.

2. The C^*-Algebra $\mathscr{B}(\mathcal{H})$

We now turn to the study of operators between Hilbert spaces. We shall focus on *bounded endomorphisms*, viz., continuous linear operators on a Hilbert space into itself, for they are direct generalizations of square matrices. Our aim is to develop a theory of bounded endomorphisms analogous to that of eigenvalues, eigenvectors, and diagonalizable matrices in linear algebra.

We have already seen that the space $\mathscr{B}(\mathcal{H}) = \mathscr{B}(\mathcal{H}, \mathcal{H})$ of all bounded endomorphisms on a Hilbert space $\mathcal{H}$ is a normed linear space (Chapter 1, Exercise **??**) whose norm is complete (Chapter 1, Theorem 3.8). The space $\mathscr{M}_n(\mathbb{F})$ n-by-n

matrices with $\mathbb{F}$-valued entries is more than a normed linear space, however. For example, $\mathscr{M}_n$ is equipped with a binary *matrix multiplication* operation and a unary *conjugate transpose* operation. The goal of this section is to introduce the appropriate generalizations of the algebraic structures that $\mathscr{M}_n$ possesses. We shall see at the end of the section that these additional structures turn $\mathscr{B}(\mathcal{H})$ into a C^**-algebra*.

For the remainder of this section, we fix a Hilbert space $\mathcal{H}$ over $\mathbb{C}$.

DEFINITION 2.1. The *product* TU of two operators $T, U \in \mathscr{B}(\mathcal{H})$ is the composite function $T \circ U$.

We observe at once that $TU \in \mathscr{B}(\mathcal{H})$. The following basic properties of the multiplication operators are trivially verified.

PROPOSITION 2.2. *Let $T, U, V \in \mathscr{B}(\mathcal{H})$ and $\lambda \in \mathbb{C}$. The multiplication operation satisfies the following properties:*
(a) **Associativity**. $(TU)V = T(UV)$*;*
(b) **Distributivity**. $T(U + V) = TU + TV$ *and* $(T + U)V = TV + UV$*;*
(c) **Bilinearity**. $\lambda(TU) = (\lambda T)U = T(\lambda U)$*;*
(d) **Existence of Unit**. $IU = UI = I$*, where I is the identity operator.*

The above properties turns $\mathscr{B}(\mathcal{H})$ into a *unital associative* $\mathbb{C}$*-algebra*:

DEFINITION 2.3. A vector space $\mathscr{A}$ over $\mathbb{F}$[1] is an *associative algebra over* $\mathbb{F}$, or an *associative* $\mathbb{F}$*-algebra*, if there exists a multiplication map $\mathfrak{m} : \mathscr{A} \times \mathscr{A} \to \mathscr{A}$ that is

- *associative*, i.e., $x(yz) = (xy)z$,
- *distributive*, i.e., $x(y + z) = xy + xz$ and $(x + y)z = xz + yz$, and
- *bilinear*, i.e., $\lambda(xy) = (\lambda x)y = x(\lambda y)$

for all $x, y, z \in \mathscr{A}$ and $\lambda \in \mathbb{F}$, where we have written xy to denote $\mathfrak{m}(x, y)$. An associative $\mathbb{F}$-algebra $\mathscr{A}$ is *unital* if there exists an element $1 \in \mathscr{A}$ such that

$$1x = x1 = 1$$

for all $x \in \mathscr{A}$, and *commutative* if

$$xy = yx$$

for all $x, y \in \mathscr{A}$.

Recall that every normed linear space is a topological vector space. Indeed, the vector addition operation and the scalar multiplication operation are continuous. It is then natural to expect that newly-defined multiplication operation is continuous as well—and it indeed is.

PROPOSITION 2.4. $\|TU\| \leq \|T\|\|U\|$ *for all* $T, U \in \mathscr{B}(\mathcal{H})$.

In fact, we have already used this fact once in the proof of Chapter 1, Proposition 5.15. This turns $\mathscr{B}(\mathcal{H})$ into a *unital Banach algebra*:

DEFINITION 2.5. An associative $\mathbb{F}$-algebra $\mathscr{A}$ is a *Banach algebra* if there exists a norm $\|\cdot\|$ on $\mathscr{A}$ that satisfies the inequality

$$\|xy\| \leq \|x\|\|y\|$$

[1]Of course, associative algebras over a ring R can be defined by considering R-moudles with an appropriate multiplication operation, but this level of generality is unnecessary for this course.

for all $x, y \in \mathscr{A}$. A Banach algebra $\mathscr{A}$ is *unital* or *commutative* if $\mathscr{A}$ as an associative algebra is unital or commutative, respectively.

We now generalize the conjugate transpose operation on $\mathscr{M}_n(\mathbb{F})$. Recall the definition of the *adjoint* $T^* : Y^* \to X^*$ of a bounded linear operator $T : X \to Y$ between normed linear space from Chapter 1, Definition 5.14. If $X = Y = \mathcal{H}$, then X^* and Y^* are isometrically isomorphic to $\mathcal{H}$ by the Riesz representation theorem (Theorem 1.20), and so we could consider T^* as an element of $\mathscr{B}(\mathcal{H})$. Indeed, for each $l \in \mathcal{H}^*$, the Riesz representation theorem yields a unique $y \in \mathcal{H}$ such that $lx = \langle x, y\rangle$, and so

$$(T^*l)(x) = (lT)(x) = l(Tx) = \langle Tx, y\rangle.$$

By the Riesz representation theorem, there exists a unique $y^* \in \mathcal{H}$ such that $T^*l(x) = \langle x, y^*\rangle$, and so

$$\langle x, y^*\rangle = (T^*l)(x) = \langle Tx, y\rangle.$$

Considering T^* as the operator that sends y to y^*, we are led to the following definition:

Definition 2.6. The *Hilbert adjoint*, or *Hermitian adjoint*, of $T \in \mathscr{B}(\mathcal{H})$ is the operator $T^* : \mathcal{H} \to \mathcal{H}$ that satisfies the identity

$$\langle Tx, y\rangle = \langle x, T^*y\rangle$$

for all $x, y \in \mathcal{H}$.

By the above discussion, we see that the Hilbert adjoint of an operator is always well-defined. Since the conjugate transpose A^* of an n-by-n matrix A with complex entries satisfies the dot-product identity

$$(Ax) \cdot y = x \cdot (A^*y),$$

the Hilbert adjoint can be considered as a direct generalization of the conjugate transpose operation.

The following basic properties of the adjoint operation are easily verified.

Proposition 2.7. *Let $T, U \in \mathscr{B}(\mathcal{H})$ and $\lambda \in \mathbb{C}$.*

(a) $*$ is an **involution**: $(T^*)^* = T$.
(b) $*$ is **conjugate-linear**: $(T + U)^* = T^* + U^*$ *and* $(\lambda T)^* = \bar{\lambda}T^*$.
(c) $*$ is **contravariant**: $(TU)^* = U^*T^*$.

The above properties turn $\mathscr{B}(\mathcal{H})$ into a *unital Banach $*$-algebra*:

Definition 2.8. An *involution* on a Banach algebra $\mathscr{A}$ is a map $* : \mathscr{A} \to \mathscr{A}$ such that $(x^*)^* = x$ for all $x \in \mathscr{A}$. A *Banach $*$-algebra*[2] is a Banach algebra $\mathscr{A}$ with an involution operation $*$ that is

- *conjugate-linear*, i.e., $(x + y)^* = x^* + y^*$ and $(\lambda x)^* = \bar{\lambda}x^*$,
- *contravariant*, i.e., $(xy)^* = y^*x^*$, and

for all $x, y \in \mathscr{A}$ and $\lambda \in \mathbb{F}$. A Banach $*$-algebra $\mathscr{A}$ is *commutative* if $\mathscr{A}$ as a Banach algebra is commutative, and *unital* if $\mathscr{A}$ as a Banach algebra is unital and satisfies the additioanl identity

- $1^* = 1$.

[2] Again, we could define a *$*$-algebra* by defining a $*$-involution operation on an associative $\mathbb{F}$-algebra, thereby dropping the "Banach" portion of the definition.

To illustrate that the $*$-involution is indeed a generalization of the complex conjugate, we now recover the standard fact that $(A^{-1})^* = (A^*)^{-1}$ for all $A \in \mathscr{M}_n(\mathbb{F})$ in the Banach $*$-algebra setting.

Definition 2.9. An element x of a unital associative $\mathbb{F}$-algebra is *invertible* if there exists an element y such that $xy = yx = 1$, called an *inverse* of x. There is at most one inverse of x, and we write x^{-1} to denote the inverse if it exists.

Proposition 2.10. *An element x of a unital Banach $*$-algebra is invertible if and only if x^* is; if x is invertible, then $(x^*)^{-1} = (x^{-1})^*$.*

Proof. $1 = 1^* = (xx^{-1})^* = (x^{-1})^*x^*$ by contravariance. □

We now extend the notions of self-adjoint, normal, and unitary matrices to Banach $*$-algebras. As a particular instance of the following generalization, we obtain the definitions of self-adjoint, normal, and unitar operators.

Definition 2.11. An element of a Banach $*$-algebra $\mathscr{A}$ is *self-adjoint* if $x^* = x$ and *normal* if $x^*x = xx^*$. If $\mathscr{A}$ is unital, then $x \in \mathscr{A}$ is *unitary* if $x^*x = xx^* = 1$, or, equivalently, if $x^* = x^{-1}$.

We have seen in Chapter 1, Proposition 5.15 that the adjoint operation preserves the operator norm. Since matrices of the form A^*A were of pivotal importance in the study of diagonalizable matrices in linear algebra, we seek to know more about operators of the form T^*T. An important basic property of such operators is the following proposition:

Proposition 2.12. $\|T^*T\| = \|T^*\|\|T\| = \|T\|^2$ *for all* $T \in \mathscr{B}(\mathcal{H})$

Proof. The adjoint operation is an isometry (Chapter 1, Proposition 5.15), and so we have the identity $\|T^*\|\|T\| = \|T\|^2$. Since $\mathscr{B}(\mathcal{H})$ is a Banach algebra (Proposition 2.4), we have the norm inequality $\|T^*T\| \leq \|T^*\|\|T\|$, whence it suffices to establish the converse inequality. To this end, we fix a sequence $(x_n)_{n=1}^{\infty}$ in $\mathcal{H}$ such that $\|x_n\| = 1$ and $\lim_n \|Tx_n\| = \|T\|$ and observe that

$$\begin{aligned}
\|T^*T\| &= \sup_{\|x\|\leq 1} \|T^*Tx\| \\
\text{(def. of operator norm)} \quad &\geq \lim_{n\to\infty} \|T^*Tx_n\| \\
(\|x_n\| = 1) \quad &= \lim_{n\to\infty} \|T^*Tx_n\|\|x_n\| \\
\text{(Schwarz)} \quad &\geq \lim_{n\to\infty} |\langle T^*Tx_n, x_n\rangle| \\
\text{(def. of Hilbert adjoint)} \quad &= \lim_{n\to\infty} |\langle Tx_n, Tx_n\rangle| \\
\text{(def. of inner product norm)} \quad &= \lim_{n\to\infty} \|Tx_n\|^2 \\
\text{(def. of } (x_n)) \quad &= \|T\|^2 \\
\text{(isometry)} \quad &= \|T^*\|\|T\|,
\end{aligned}$$

as was to be shown. □

The above property turns $\mathscr{B}(\mathcal{H})$ into a *unital C^*-algebra*, as advertised in the beginning of this section:

DEFINITION 2.13. A Banach $*$-algebra $\mathscr{A}$ is a *C^*-algebra* if every element x of $\mathscr{A}$ satisfies the *C^*-identity*

$$\|x^*x\| = \|x\|^2.$$

A C^*-algebra $\mathscr{A}$ is *unital* or *commutative* if $\mathscr{A}$ as a Banach $*$-algebra is unital or commutative, respectively.

We observe that the identity akin to that in Proposition 2.12 holds for all C^*-algebras.

PROPOSITION 2.14. *The $*$-involution on a C^*-algebra $\mathscr{A}$ is an isometry, whence*

$$\|x^*x\| = \|x^*\|\|x\| \tag{2.15}$$

for all $x \in \mathscr{A}$. Conversely, if the $*$-involution is an isometry, then (2.15) implies the C^*-identity.

PROOF. If the C^* identity holds, then

$$\|x\|^2 = \|x^*x\| \leq \|x^*\|\|x\|,$$

and so $\|x^*\| \leq \|x\|$. Similarly,

$$\|x^*\|^2 = \|(x^*)^*x^*\| = \|xx^*\| \leq \|x\|\|x^*\|,$$

and so $\|x^*\| \geq \|x\|$. It follows that the $*$-involution is an isometry, whence (2.14) follows. The converse is trivially established. □

In fact, (2.15) implies (2.14) without the isometry condition on the $*$-involution

The theory of C^*-algebras provide powerful tools for studying operators. In fact, the abstract definition presented in Definition 2.13 grew out of the study of *operator algebras*, which we define below.

DEFINITION 2.16. The *C^*-algebra of an operator $T \in \mathscr{B}(\mathcal{H})$*, denoted by $C^*(T)$,is the norm closure of the space of polynomials in two variables T and T^*. In other words, $C^*(T)$ is the norm closure of the span of the set

$$\{T^n\}_{n\in\mathbb{N}} \cup \{(T^*)^n\}_{n\in\mathbb{Z}}.$$

In general, a *C^*-algebra of operators* in $\mathscr{B}(\mathcal{H})$ is a closed linear subspace $\mathscr{A}$ of $\mathscr{B}(\mathcal{H})$ such that $T, U \in \mathscr{A}$ implies $TU \in \mathscr{A}$ and $T^* \in \mathscr{A}$.

Indeed, the name *C^*-algebra* comes from "Norm-**C**losed $*$-sub**algebra** of $\mathscr{B}(\mathcal{H})$." A fundamental theorm of Gelfand and Naimark shows that the study of C^*-algebras is no more general than the study of operator algebras.

THEOREM 2.17 (Gelfand–Naimark theorem). *Every C^*-algebra is isometrically $*$-isomorphic to a C^*-algebra of operators. In other words, for every C^*-algebra $\mathscr{A}$, there exists a Hilbert space $\mathcal{H}$, a norm-closed Banach $*$-subalgebra $\mathscr{B}$ of $\mathscr{B}(\mathcal{H})$, and an isometric isomorphism $\pi : \mathscr{A} \to \mathscr{B}$ such that $\pi(x^*) = \pi(x)^*$ for all $x \in \mathscr{A}$.*

We will not have an occasionl to use this theorem, so we do not provide a proof. Interested readers should refer to Chapter 4 of [**Dou98**], or Chapter 1 of [**Arv76**]. The former reference covers the topics of this chapter in much greater detail; the latter is a standard textbook in the basic theory of C^*-algebras, assuming some exposure to the theory of commutative Banach algebras.

We conclude this section with a remark on the terminology. The term *C^*-algebra* was coined by Irving Segal in [**Seg47**] and was used in the context of concrete operator algebras a là Definition 2.16. The abstract definition appeared as Definition 2.13 in this section was first introduced by Gelfand and Naimark [**GN43**] as a technical hypothesis for proving a weaker version of Theorem 2.17; here Banach algebras as in Definition 2.5 are referred to as *normed rings*, and Banach $*$-algebras (Definition 2.8) with isometric $*$-involutions as *$*$-rings*. Abstract C^*-algebras were first studied systematically by Charles Rickart in [**Ric45**], where he employs the term *B^*-algebras* for what we have named C^*-algebras in this section. But the full version of the Gelfand–Naimark theorem—which is the one we stated as Theorem 2.17—implies that C^*-algebras and B^*-algebras are precisely the same, and the division of concrete algebras as C^* and abstract algebras as B^* is no longer necessary. In fact, performing searches on widely used mathematical databases such as MathSciNet, MathOverflow, and Math StackExchange indicates that the term *B^*-algebras* is no longer in use, even though fairly modern textbooks such as [**Lax02**] and [**Rud91**] continue to use the term to refer to abstract C^*-algebras.

3. Spectral Theory of Banach Algebras

Having laid down the linguistic foundations, we proceed with the task of recovering a suitable theory of eigenvalues, eigenvectors, and diagonalizable matrices in the context of bounded endomorphisms on a Hilbert space. We shall develop the basic notions in their natural setting, the Banach algebras over $\mathbb{C}$. Key examples include $\mathscr{B}(\mathcal{H})$, the space $\mathscr{B}(X)$ of bounded endomorphisms on a Banach space X, Lebesgue space $L^1(\mathbb{R}^d)$ with convolution as the multiplication operation, the space $\mathscr{C}(K)$ of continuous complex-valued functions on a compact Hausdorff space K, and the space $\mathscr{C}_0(L)$ of continuous complex-valued functions on a locally compact Hausdorff space L that vanish at infinity.

Recall that an *eigenvalue* of a square matrix A is a scalar λ such that

$$\det(\lambda I - A) = 0. \tag{3.1}$$

This is equivalent to the statement that the matrix $\lambda I - A$ is *not* invertible. Since it is not very clear how we should define the determinant of a bounded endomorphism on an infinite-dimensional Hilbert space, we shall take the invertibility criterion as our definition.

Definition 3.2. The *resolvent set* of an element x of a unital Banach algebra $\mathscr{A}$ is the set $\rho_{\mathscr{A}}(x)$ of complex numbers λ such that $\lambda 1 - x$ is invertible in $\mathscr{A}$. The *spectrum* of x is defined to be the set

$$\sigma_{\mathscr{A}}(x) = \mathbb{C} \smallsetminus \rho_{\mathscr{A}}(x).$$

In particular, the operator $\lambda I - T$ must be invertible *in* $\mathscr{B}(X)$ in order for λ to be in the resolvent $\rho(T)$ of $T \in \mathscr{B}(\mathcal{H})$. But this is automatic, as every bijective bounded operator on a Hilbert space has a bounded inverse (Chapter 1, Corollary 4.12).

Since (3.1) is a polynomial equation over $\mathbb{C}$, the fundamental theorem of calculus guarantees the existence of a root—an eigenvalue. In order for spectral theory, the "theory of the spectrum", to be of substance, we need to be able to recover this existence result in the generalized setting. Therefore, our immediate goal is to establish the following:

Theorem 3.3. *$\sigma_{\mathscr{A}}(x)$ is nonempty.*

Proof of theorem. We break down the proof into several lemmas. The key idea is to study the map $\lambda \mapsto (\lambda 1 - x)^{-1}$ via complex-analytic techniques and deduce a contradiction in the $\rho_{\mathscr{A}}(x) = \mathbb{C}$ case. To do so, we first need to establish a proper domain for the function $\lambda \mapsto (\lambda 1 - x)^{-1}$. Since complex analysis operates on open domains, we shall prove that $\rho_{\mathscr{A}}(x)$ is open.

Lemma 3.4. *If $x_1 \in \mathscr{A}$ is invertible and $x_2 \in \mathscr{A}$ satisfies the norm estimate $\|x_2\| < \|x_1^{-1}\|^{-1}$, then $x_1 - x_2$ is invertible. In particular, $\rho_{\mathscr{A}}(x_1)$ is open.*

Proof of lemma. We assume for now that $x_1 = 1$. In this case, the norm estimate $\|x_2\| < 1$ implies that the sequence of partial sums $(\sum_{n=0}^{N} x_2^n)_{N=1}^{\infty}$ is Cauchy in $\mathscr{A}$, whence

$$y = \sum_{n=0}^{\infty} x_2^n = \lim_{N\to\infty} \sum_{n=0}^{N} x_2^n \tag{3.5}$$

exists. It follows that once that $y = x_2^{-1}$.

We now lift the assumption that $x_1 = 1$. We factor $x_1 - x_2$ as follows:

$$x_1 - x_2 = x_1(1 - x_1^{-1}x_2)$$

Observe that

$$\|x_1^{-1}x_2\| \leq \|x_1^{-1}\|\|x_2\| < \|x_1^{-1}\|\|x_1^{-1}\|^{-1} = 1,$$

whence the above argument implies that $1 - x_1^{-1}x_2$ is invertible. The product of two invertible elements is invertible, and so $x_1 - x_2$ is invertible.

Finally, if $\lambda \in \rho_{\mathscr{A}}(x_1)$, then

$$\|\varepsilon 1\| = \varepsilon < \|(\lambda 1 - x_1)^{-1}\|^{-1}$$

for all sufficiently small $\varepsilon \in \mathbb{C}$, and so

$$(\lambda - \varepsilon)1 - x_1 = (\lambda 1 - x_1) - \varepsilon 1$$

is invertible for all such ε. It follows that $\rho_{\mathscr{A}}(x_1)$ is open. □

We now define a function $f : \rho_{\mathscr{A}}(x) \to \mathscr{A}$ by setting $f(\lambda) = (\lambda 1 - x)^{-1}$. In order to apply the "complex-analytic techniques" as we advertised, we must have a notion of holomorphicity for Banach-valued functions.

Definition 3.6. A function $f : O \to \mathscr{A}$ from an open subset of $\mathbb{C}$ into $\mathscr{A}$ is *holomorphic* if $lf : O \to \mathbb{C}$ is holomorphic for each bounded linear functional l on $\mathscr{A}$ as a Banach space.

We claim that f is holomorphic on $\rho_{\mathscr{A}}(x)$. To see this, we make use of the following computational result.

Lemma 3.7. *If $x_1, x_2 \in \mathscr{A}$ are invertible, then*

$$x_1^{-1} - x_2^{-1} = x_1^{-1}(x_2 - x_1)x_2^{-1}.$$

Proof of lemma. $x_1^{-1} - x_2^{-1} = x_1^{-1}x_2x_2^{-1} - x_1^{-1}x_1x_2^{-1} = x_1^{-1}(x_2 - x_1)x_2^{-1}$. □

For each fixed $\mu \in \rho_{\mathscr{A}}(T)$, Lemma 3.7 implies that

$$\begin{aligned} \frac{(\lambda 1 - x)^{-1} - (\mu 1 - x)^{-1}}{\lambda - \mu} &= \frac{(\lambda 1 - x)^{-1}(\mu 1 - \lambda 1)^{-1}(\mu 1 - x)^{-1}}{\lambda - \mu} \\ &= -(\lambda 1 - x)^{-1}(\mu 1 - x)^{-1} \end{aligned}$$

for all $\lambda \neq \mu$. Therefore,

$$\lim_{\mu \to \lambda} \frac{(\lambda 1 - x)^{-1} - (\mu 1 - x)^{-1}}{\lambda - \mu} = -(\lambda 1 - x)^{-2},$$

and so

$$\lim_{\mu \to \lambda} \frac{l(\lambda 1 - x)^{-1} - l(\mu 1 - x)^{-1}}{\lambda - \mu} = l(-(\lambda 1 - x)^{-2}) < \infty,$$

for each bounded linear functional l on $\mathscr{A}$, thereby establishing the claim of holomorphicity.

We now suppose for a contradiction that $\rho_{\mathscr{A}}(x) = \mathbb{C}$, which, in particular, implies that lf is an entire function for each bounded linear functional l on $\mathscr{A}$. Recall that we have derived a power-series formula (3.5) of an inverse in the proof of Lemma 3.4. We apply this formula to $(\lambda 1 - x)^{-1}$ to obtain

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

By summing the geometric series, we see that

$$\|l(\lambda 1 - x)^{-1}\| \leq \|l\| \|\lambda 1 - x)^{-1}\| \leq \frac{\|l\|}{|\lambda| - \|x\|},$$

whence $\|l(\lambda 1 - x)^{-1}\| \to 0$ as $|\lambda| \to \infty$. Applying the following Banach-valued extension of Liouville's theorem to f, we conclude that f must be a constant function. In particular, f must be the zero function, as the above estimate shows that $(\lambda 1 - x)^{-1} \to 0$ as $|\lambda| \to \infty$.

Lemma 3.8 (Banach-valued Liouville's theorem). *If a Banach-valued holomorphic function $f : O \to \mathscr{A}$ is weakly bounded, viz., lf is bounded in $\mathbb{C}$ for each bounded linear functional l on $\mathscr{A}$, then f is a constant function.*

Proof of lemma. Evidently, each lf is a constant function. If $f(\lambda_1) \neq f(\lambda_2)$, then the existence of nontrivial annihilators (Chapter 1, Corollary 5.10) guarantees the existence of a bounded linear functional l on $\mathscr{A}$ such that $l(f(\lambda_1) - f(\lambda_2))$ is nonzero. In particular, $lf(\lambda_1) \neq lf(\lambda_2)$, contardicting the fact that lf must be a constant function. □

But now, $f(\lambda) = (\lambda 1 - x)^{-1}$ is invertible, whence it cannot be the zero element. This is evidently absurd, and we conclude that $\rho_{\mathscr{A}}(x) \neq \mathbb{C}$. In particular, $\sigma_{\mathscr{A}}(x)$ is nonempty. □

Recall that another definition of an eigenvalue λ of an n-by-n matrix is that the identity

$$Ax = \lambda x \tag{3.9}$$

must hold for all $x \in \mathbb{F}^n$. While the determinant definition (3.1) and the scaling definition (3.9) agree on finite-dimensional operator algebras, the nontriviality of a spectrum of an element of a Banach algebra does not necessarily imply the existence of a constant satisfying (3.9). The next example illustrates this point.

Example. We define the *right-shift operator* $R \in \mathscr{B}(l^2(\mathbb{Z}))$ by setting

$$l((a_n)_{n \in \mathbb{Z}}) = (a_{n+1})_{n \in \mathbb{Z}}.$$

By Theorem 3.3, the spectrum of R is nonempty, but the scaling identity (3.9) is never satisfied for all $(a_n) \in l^2(\mathbb{Z})$.

In particular, the spectra of operators on an infinite-dimensional space can be decomposed into a number of different parts, which we shall investigate in due course. In this section, we content ourselves with facts about the spectrum that we could gather at the level of Banach algebras. We first note that the nontriviality of spectrum (Theorem 3.3) implies the following classification result.

COROLLARY 3.10 (Gelfand–Mazur theorem). *If $\mathscr{A}$ is a Banach algebra in which every nonzero element is invertible, then there exists a unique isometric isomorphism from $\mathscr{A}$ to $\mathbb{C}$.*

PROOF. For each $x \in \mathscr{A}$, Theorem 3.3 implies that the spectrum $\sigma_{\mathscr{A}}(x)$ is nonempty. If $\lambda \in \sigma_{\mathscr{A}}(x)$, then $\lambda 1 - x$ is not invertible, whence by the hypothesis $\lambda 1 - x = 0$. Now, if $\mu \in \mathscr{A}$, then

$$\mu 1 - x = \mu 1 - \lambda 1 + (\lambda 1 - x) = (\mu - \lambda)1$$

must not be invertible, but this is true if and only if $\mu = \lambda$. It follows that $\sigma_{\mathscr{A}}(x)$ contains precisely one element, say λ_x, and the mapping $x \mapsto \lambda_x$ is the desired isometric isomorphism. □

(fill in the details)

DEFINITION 3.11. The *spectral radius* of an element x of a unital Banach algebra $\mathscr{A}$ is

$$r(x) = \sup_{\lambda \in \sigma_{\mathscr{A}}(x)} |\lambda|.$$

PROPOSITION 3.12. $r(x) \leq \|x\|$. *In particular, the spectral radius is always finite and the spectrum is compact.*

PROOF. □

THEOREM 3.13 (Gelfand's formula). $r(x) = \lim_n \|x^n\|^{1/n}$.

LEMMA 3.14 (Spectral mapping theorem for polynomials). *If x is an element of a unital Banach algebra $\mathscr{A}$ and p a non-constant polynomial, then*

$$\sigma_{\mathscr{A}}(p(x)) = p(\sigma_{\mathscr{A}}(x)),$$

where the right-hand side is defined to be the set $\{p(\lambda) : \lambda \in \sigma_{\mathscr{A}}(x)\}$.

PROOF OF LEMMA. [**Car12**], Proposition 1.12 □

PROOF OF THEOREM. [**Car12**], Theorem 1.21 □

THEOREM 3.15. *If x is a self-adjoint element of a unital C^*-algebra $\mathscr{A}$, then $\sigma_{\mathscr{A}}(x) \subseteq \mathbb{R}$.*

PROOF. [**Car12**], Theorem 3.5 □

(some technical remarks about adjoining the identity)
Recall:

THEOREM 3.16 (Finite-dimensional spectral theorem). *An n-by-n matrix A with complex entries is normal if and only if A is unitarily diagonalizable, viz., there exists a unitary matrix U such that $A = UDU^*$.*

THEOREM 3.17 (Abstract spectral theorem). *If $T \in \mathscr{B}(\mathcal{H})$ is a normal operator, then the operator algebra $C^*(T)$ is isometrically isomorphic to the C^*-algebra $\mathscr{C}(\sigma(T))$ of continuous functions on the spectrum $\sigma(T)$ of T.*

The next two sections will be devoted to developing two concrete instantiations of the above theorem.

4. Spectral Theory of Compact Operators

In order to generalize Theorem 3.16 to various subalgebras of $\mathscr{B}(\mathcal{H})$, it makes sense to examine the case that models the finite-dimensional theory as closely as possible. While the domain of an operator is irrevocably infinite-dimensional on an infinite-dimensional space, it is well within our reach to restrict the dimension of the range.

Definition 4.1. An operator $T \in \mathscr{B}(X, Y)$ between two Banach spaces X and Y is *of finite rank* if $\operatorname{im} T$ is finite-dimensional. The *rank* of a finite-rank operator T is the dimension of its image. We write $\mathscr{F}(X, Y)$ to denote the space of all finite-rank operators in $\mathscr{B}(X, Y)$.

(fill in the details)

Proposition 4.2. *Let $\mathscr{K}(\mathcal{H})$ be the norm closure of the collection of all finite-rank operators in $\mathscr{B}(\mathcal{H})$. If $T \in \mathscr{K}(\mathcal{H})$, then $\overline{T(B)}$ is compact in $\mathcal{H}$ whenever $B \subseteq \mathcal{H}$ is norm-bounded.*

Definition 4.3. Let X and Y be Banach spaces. $T \in \mathscr{B}(X, Y)$ is *compact* if $\overline{T(B)}$ is compact in Y whenever $B \subseteq X$ is norm-bounded. The collection of all compact operators in $\mathscr{B}(X, Y)$ is denoted by $\mathscr{K}(X, Y)$.

Example. Let $K(x, y)$ be a continuous complex-valued funciton on $[0, 1]^2$ and consider the operator $T_K \in \mathscr{B}(\mathscr{C}([0, 1]))$ defined by

$$(T_K f)(x) = \int_0^1 f(y) K(x, y) \, dy.$$

Whenever B is a bounded set in $\mathscr{C}([0, 1])$, the image set $T_K(B)$ is pointwise bounded and equicontinuous. It follows from the Arzelà-Ascoli theorem that $\overline{T_K(B)}$ is compact.

Proposition 4.4. *If $T \in \mathscr{K}(X, Y)$ and $U \in \mathscr{K}(Y, Z)$, then $UT \in \mathscr{K}(X, Z)$.*

Theorem 4.5 (Schauder). *If $T \in \mathscr{K}(X, Y)$, then $T^* \in \mathscr{K}(Y^*, X^*)$.*

Corollary 4.6. *$\mathscr{K}(\mathcal{H})$ is the C^*-algebra of compact operators on $\mathcal{H}$.*

Example. Not true in general. [**Bre11**], p.158

Proposition 4.7. *If $T \in \mathscr{K}(X, Y)$ and $x_n \rightharpoonup x$ in X, then $Tx_n \to Tx$ in Y.*

We now recall from the last section that the spectrum of an element of a Banach algebra need not contain any eigenvalue in the linear-algebraic sense. Since we have set out to model the finite-dimensional case as closely as possible, we expect the spectrum and the collection of

Definition 4.8. The *point spectrum* of $T \in \mathscr{B}(X)$ is the set $\sigma_p(T)$ of all complex numbers λ such that $Tx = \lambda x$ for each $x \in X$. An element λ of the point spectrum is called an *eigenvalue* of T, and the corresponding vector x an *eigenvector* of T with respect to λ.

Theorem 4.9 (F. Riesz–Schauder). *If $T \in \mathscr{K}(X)$, then $\sigma(T) \setminus \{0\} = \sigma_p(T) \setminus \{0\}$, and one of the following is true:*

- *$\sigma(T) = \{0\}$,*
- *$\sigma(T)$ is finite, or*
- *$\sigma(T)$ is a countable set whose only limit point is 0.*

Proof of theorem. We note that the problem of characterizing the point spectrum of T is equivalent to the study of the nonzero solutions of the equation

$$Tx = \lambda x \tag{4.10}$$

for a fixed λ. Indeed, if (4.10) has a nonzero solution x, then x is an eigenvector, and $\lambda \in \sigma_p(T)$. On the other hand, if (4.10) has no nonzero solutions, then the kernel of $\lambda I - T$ is trivial. In the finite-dimensional case, the rank-nullity theorem would now imply that $\lambda I - T$ is invertible, whence $\lambda \in \rho(T)$. Our immediate goal is to recover this connection between injectivity and surjectivity of linear operator in the setting of compact operators.

Lemma 4.11 (Fredholm alternative). *The equation $Ux = \lambda x$ satisfies the following properties:*

(a) If $U \in \mathscr{K}(X, Y)$ and $\operatorname{im} U$ is closed in Y, then $U \in \mathscr{F}(X, Y)$.
(b) If $U \in \mathscr{K}(X)$, then $\dim \ker(\lambda I - U) < \infty$ for every non-zero scalar λ.
(c) If $U \in \mathscr{K}(X)$ and $\lambda \neq 0$, then $\lambda I - U$ is injective if and only if $\lambda I - U$ is surjective.

Proof of lemma. (a) Since $\operatorname{im} U$ is a Banach space, we resrict the codomain of U to $\operatorname{im} U$ and apply the open mapping theorem (Chapter1, Theorem 4.6) to conclude that $B = U(B_1(0))$ is open. Therefore, B contains a closed ball centered at $0 = U(0)$, which is compact by the compactness of U. By translation invariance, it follows that all closed balls in $\operatorname{im} T$ are compact, whence by Theorem 2.26 in Chapter 1 that $\operatorname{im} U$ is finite-dimensional.

(b) Let $Z = \ker(\lambda I - U)$, which is evidently a Banach space. If $x \in Z$, then $Ux = \lambda x$, and so $U(Z) \subseteq Z$. Moreover, $U(\lambda^{-1}x) = x$, whence $U(Z) \supseteq Z$. It follows from (a) that Z is finite-dimensional.

(c) We assume without loss of generality that $\lambda = 1$.

($\Rightarrow$) Suppose for a contradiction that $I - U$ is injective but not surjective. By (a), $X_1 = (I - U)(X)$ is a proper closed subspace of X. Let $X_0 = X$. Since the restriction of a compact operator onto a closed subspace is compact, we may apply (a) repeatedly to define compact operators $U_n = (I - U)|_{X_n}$ on X_n and closed proper subspaces $X_{n+1} = U_n(X_n)$ of X_n. Riesz's Lemma (Chapter 1, Lemma 2.28) now furnishes a sequence $(x_n)_{n=1}^{\infty}$ in X such that $\|x_n\| = 1$, $x_n \in X_n$, and $\|x_n - x_{n+1}\| > \frac{1}{2}$ for all $n \in \mathbb{N}$. Since $\{x_n\}_{n\in\mathbb{N}}$ is a bounded set, the compactness of T implies that the image set $\{Ux_n\}_{n\in\mathbb{N}}$ must have a convergent subsequence. But this is absurd, as

$$Tx_n - Tx_m =$$

□

Lemma 4.12. *If $\dim X = \infty$, and if $T \in \mathscr{K}(X)$, then $0 \in \sigma(T)$.*

Proof of lemma. □

Lemma 4.13. *$\sigma(T) \setminus \{0\}$ is discrete.*

□

Example. The Fredholm alternative is false in general. $X = L^2([0,2])$. $(Tf)(x) = xf(x)$. $Tf = f$ has no solution on X, but $1 \notin \rho(T)$.

Example. Solve for ψ: $\lambda\psi = T\psi + \varphi$. T compact, $\lambda \neq 0$, φ. If uniqueness is guaranteed, then λ is not in $\sigma(T)$, and so $\lambda I - T$ is invertible. Therefore, $\psi = (\lambda I - T)^{-1}\varphi$. Uniqueness for compact maps imply existence of solutions.

Theorem 4.14 (Spectral theorem for compact operators). *If $\mathcal{H}$ is a separable Hilbert space and T a compact self-adjoint operator on $\mathcal{H}$, then there exists an orthonormal basis of $\mathcal{H}$ consisting of the eigenvectors of T.*

Definition 4.15. Let T be a compact self-adjoint operator on a separable Hilbert space $\mathcal{H}$. The *Rayleigh quotient* of T is given by

$$R_T(x) = \frac{\langle Tx, x\rangle}{\langle x, x\rangle}.$$

Lemma 4.16. *If $m = \inf_x R_T(x)$ and $M = \sup_x R_T(x)$, then following holds:*

(1) $\sigma(T) \subseteq [m, M]$
(2) $m, M \in \sigma(T)$.
(3) $r(T) = \max(|m|, |M|)$.

$\sigma(T) \subseteq [m, M]$, $m, M \in \sigma(T)$

CHAPTER 3

Appendix A - Additional Remarks and Further Results

Collected in this chapter are supplementary notes for the main lectures. For the most part, each section can be read independently of the others. Dependencies to other sections are explicitly labeled, and the sections are ordered so that they can be read linearly if the reader chooses to peruse the entire chapter.

1. Categories and Natural Transformations

Here is a brief introductory language course, following [**Mac98**], [**Wei95**], and [**nLa**]. There are no propositions, lemmas, or theorems in this section, and the sporadic references to the language of category theory made in the remainder of this chapter can be "understood intuitively" without reading this section.

A *category*[1] $\mathcal{C}$ consists of a class $\operatorname{Ob}\mathcal{C}$ of *objects*, a set $\operatorname{Hom}_{\mathcal{C}}(A, B)$ of *morphisms* for each ordered pair (A, B) of objects, an *identity morphism* $\operatorname{id}_A \in \operatorname{Hom}_{\mathcal{C}}(A, A)$ for each object A, and a *composition rule* $\operatorname{Hom}_{\mathcal{C}}(A, B) \times \operatorname{Hom}_{\mathcal{C}}(B, C) \to \operatorname{Hom}_{\mathcal{C}}(A, C)$ for every ordered triple (A, B, C) of objects that satisfies the following two properties:

- $(h \circ g) \circ f = h \circ (g \circ f)$ whenever $f \in \operatorname{Hom}_{\mathcal{C}}(A, B)$, $g \in \operatorname{Hom}_{\mathcal{C}}(B, C)$, and $\operatorname{Hom}_{\mathcal{C}}(C, D)$;
- $\operatorname{id}_B \circ f = f = f \circ \operatorname{id}_A$ whenever $f \in \operatorname{Hom}_{\mathcal{C}}(A, B)$.

A morphism $f \in \operatorname{Hom}_{\mathcal{C}}(A, B)$ is an *isomorphism* in $\mathcal{C}$ if there exists a morphism $g \in \operatorname{Hom}_{\mathcal{C}}(B, A)$ such that $gf = \operatorname{id}_A$ and $fg = \operatorname{id}_B$.

Example. The category **Set** of sets with functions as morphisms is a category. The isomorphisms are bijective functions.

Example. The category **Top** of topological spaces with continuous functions as morphisms is a category. The isomorphisms are homeomorphisms.

Example. The category $\mathbb{F}-$**Vect** of vector spaces over $\mathbb{F}$ with linear transformations as morphisms is a category. The isomorphisms are linear isomorphisms.

Example. Let us fix a base field $\mathbb{F}$. The category **TopVect** of topological vector spaces consists of topological vector spaces with continuous linear transformations as morphisms. f is an isomorphism in **TopVect** if and only if f is an isomorphism in **Top** and $\mathbb{F}-$**Vect**.

A category $\mathcal{D}$ is a *subcategory* of $\mathcal{C}$ if every object of $\mathcal{D}$ is an object of $\mathcal{C}$ and $\operatorname{Hom}_{\mathcal{D}}(A, B) \subseteq \operatorname{Hom}_{\mathcal{C}}(A, B)$ for each $A, B \in \operatorname{Ob}\mathcal{D}$.

[1] Technically, we are defining *locally small* categories, as we require collections of morphisms to be sets—but we shall not dwell on this point.

Example. The category **NLS** of normed linear spaces consists of normed linear spaces with bounded linear transformations as morphisms. The category **Ban** of Banach spaces is then a subcategory of **NLS**.

For technical reasons, category theorists often restrict the morphisms in **NLS** and **Ban** to be *short linear maps*, viz., linear maps $T : X \to Y$ whose norm $\|T\|$ is at most 1. One reason for this is that **NLS** and **Ban** as we have defined above have too many isomorphisms. Indeed, any bounded, bijective linear transformation whose inverse is bounded is an isomorphism in these categories, whether it is isometric (see Chapter 1, Definition 2.8) or not. If we only consider short linear maps to be morphisms, then the categorical isomorphisms are surjective linear isometries.

To highlight this distinction, **NLS** and **Ban** with bounded linear transformations as morphisms are often referred to as *isomorphic categories*, and those with short linear maps as morphisms *isometric categories*.

A *covariant functor* $F : \mathcal{C} \to \mathcal{D}$ from a category $\mathcal{C}$ to another category $\mathcal{D}$ is a rule that sends each object $A \in \operatorname{Ob}\mathcal{C}$ to $F(A) \in \operatorname{Ob}\mathcal{D}$ and every morphism $f \in \operatorname{Hom}_{\mathcal{C}}(A, B)$ to $F(f) \in \operatorname{Hom}_{\mathcal{D}}(F(A), F(B))$. A *contravariant functor* $G : \mathcal{C} \to \mathcal{D}$ from $\mathcal{C}$ to $\mathcal{D}$ is a rule that sends each object $A \in \operatorname{Ob}\mathcal{C}$ to $G(A) \in \operatorname{Ob}\mathcal{D}$ and every morphism $f \in \operatorname{Hom}_{\mathcal{C}}(A, B)$ to $G(f) \in \operatorname{Hom}_{\mathcal{D}}(G(B), G(A))$.

Example. Define the *dual functor* $D : \mathbf{NLS} \to \mathbf{NLS}$ by setting $D(V) = V^*$ and $D(T) = T^*$, the topological adjoint of T (Chapter 1, Definition 5.14). This is a contravariant functor, as the "arrow" $V \xrightarrow{T} W$ reverses its direction when the adjoint is taken: $W^* \xrightarrow{T^*} V^*$.

The *double dual functor* D^2 on **NLS**. D^2 is given by the composition of the dual functor D D^2 is a covariant functor, as the arrow $V \xrightarrow{T} W$ gets flipped twice. For illustrative purposes, we write down the action of $T^{**} : V^{**} \to W^{**}$ explicitly: for each $\varphi \in V^{**}$ and every $l \in W^*$, we have

$$(T^{**}\varphi)l = (\varphi T^*)l = \varphi(T^*l) = \varphi(lT).$$

Example. The dual functor is a special case of *Hom functors*. Given a category $\mathcal{C}$ and an object A in $\mathcal{C}$, we define the *covariant Hom functor* $\operatorname{Hom}(A, -)$ from $\mathcal{C}$ to **Set** by sending each $B \in \operatorname{Ob}\mathcal{C}$ to $\operatorname{Hom}_{\mathcal{C}}(A, B)$ and $f \in \operatorname{Hom}_{\mathcal{C}}(B, C)$ to $f^* = \operatorname{Hom}(A, f) \in \operatorname{Hom}_{\mathbf{Set}}(\operatorname{Hom}_{\mathcal{C}}(A, B), \operatorname{Hom}_{\mathcal{C}}(A, C))$, given by the formula

$$f^*(g) = f \circ g.$$

Similarly, we can define the *contravariant Hom functor* $\operatorname{Hom}(-, A)$ from $\mathcal{C}$ to **Set** by sending each $B \in \operatorname{Ob}\mathcal{C}$ to $\operatorname{Hom}_{\mathcal{C}}(B, A)$ and $f \in \operatorname{Hom}_{\mathcal{C}}(B, C)$ to $f_* = \operatorname{Hom}(f, A) \in \operatorname{Hom}\mathbf{Set}(\operatorname{Hom}_{\mathcal{C}}(C, A), \operatorname{Hom}_{\mathcal{C}}(B, A))$, given by the formula

$$f_*(g) = g \circ f.$$

In this sense, the dual functor D can be thought of as the contravariant Hom functor $\operatorname{Hom}_{\mathbf{NLS}}(-, \mathbb{F})$. Here the base field $\mathbb{F}$ is considered as a one-dimensional normed linear space with some norm. Note that the Hom-set $\operatorname{Hom}_{\mathbf{NLS}}(X, \mathbb{F})$ is precisely the dual space X^*, which, in fact, is given the structure of a normed linear space. Therefore, the contravariant Hom functor $\operatorname{Hom}_{\mathbf{NLS}}(-, \mathbb{F})$ also maps **NLS** into **NLS**.

Given two covariant functors F and G from $\mathcal{C}$ to $\mathcal{D}$, we can define a *natural transformation (of covariant functors)* $\alpha : F \to G$ from F to G as a rule that associates a morphism $\alpha_A \in \operatorname{Hom}_{\mathcal{D}}(F(A), G(A))$ to each object $A \in \operatorname{Ob}\mathcal{C}$ such that the diagram

$$\begin{array}{ccc} F(A) & \xrightarrow{Ff} & F(B) \\ {\scriptstyle \alpha_A}\downarrow & & \downarrow{\scriptstyle \alpha_B} \\ G(A) & \xrightarrow[Gf]{} & G(B) \end{array}$$

commutes, i.e., $(Gf)\alpha_A = \alpha_B(Ff)$, for all $A, B \in \operatorname{Ob}\mathcal{C}$. If each α_A is an isomorphism, then α is said to be a *natural isomorphism*.

EXAMPLE. As an example, we establish a natural transformation from the identity functor I that does nothing to the objects and the morphisms to the double dual functor D^2. Recall from Chapter 1, Theorem 2.10 that there is a canonical embedding $\alpha_V : V \to V^{**}$ of a vector space V into its double dual V^{**}, which associates each $x \in V$ to $\hat{x} \in V^{**}$. $\hat{x}$ is given by the formula

$$\hat{x}l = lx$$

for each $l \in V^*$. We claim that $\alpha : I \to D^2$ that associates V to α_V is a natural transformation from the identity functor I to the double dual functor D^2. In other words, the diagram

$$\begin{array}{ccc} V & \xrightarrow{T} & W \\ {\scriptstyle \alpha_V}\downarrow & & \downarrow{\scriptstyle \alpha_W} \\ V^{**} & \xrightarrow[T^{**}]{} & W^{**} \end{array}$$

commutes for all normed linear spaces V and W. To prove this assertion, we first note that $T^{**}\alpha_V x$ for a fixed $x \in V$ is an element of W^{**}, hence it takes elements of W^* and produces a scalar. We now fix $l \in W^*$ and observe that

$$\begin{aligned} (T^{**}\alpha_V x)l &= (T^{**}\hat{x})l = (\hat{x}T^*)l = \hat{x}(T^*l) = \hat{x}(lT) \\ &= (lT)x = l(Tx) = (\widehat{Tx})l = (\alpha_W Tx)l, \end{aligned}$$

which is the desired result.

EXAMPLE. We remark that α_V in the above example is an isomorphism if V is finite-dimensional, hence α is a natural isomorphism if we restrict to the category of finite-dimensional normed linear spaces. The same argument shows that α is a natural transformation on the category of Banach spaces, and a natural isomorphism on the category of reflexive Banach spaces.

2. The First Uncountable Ordinal ω_1

The following set-theoretic material, which is used in the construction of a sequentially compact but not compact topological space in Section 3 of this appendix, is taken from [**Jec06**] and [**Kun09**].

A *relation* on a nonempty set A is a subset R of the Cartesian product $A \times A$; we often write xRy to denote the fact that (x, y) is an element of R. A *partial order* on A is a relation $\leq$ that is

- **reflexive**: $x \leq x$ for all $x \in A$;
- **antisymmetric**: $x \leq y$ and $y \leq x$ imply $x = y$;
- **transitive**: $x \leq y$ and $y \leq z$ imply $x \leq z$.

A *total order* or a *linear order* on A is a partial order $\leq$ that satisfies the **trichotomy property**, viz., for every pair of elements x and y of A, one of the three must hold:

$$x \leq y, \quad x = y, \quad \text{or} \quad y \leq x.$$

A *well-order* on A is a total order $\leq$ on A such that every non-empty subset E of A has a **least element**, viz., an element $x \in E$ such that $x \leq y$ for all $y \in E$. If $\leq$ is any kind of order on A, we write $(A, \leq)$ to denote the fact that A is equipped with the order $\leq$.

The following theorem is equivalent to the axiom of choice:

Theorem 2.1 (Zermelo). *Every set admits a well-order.*

We now exhibit the canonical examples of well-ordered sets: the ordinal numbers. Not only are ordinals the canonical examples, they are, in a sense, the only examples.

The von Neumann definition of the natural numbers goes as follows:

- $0 = \varnothing$;
- $1 = \{0\} = \{\varnothing\}$;
- $2 = \{0, 1\} = \{\varnothing, \{\varnothing\}\}$;
- $n = \{0, 1, \ldots, n-1\}$.

We define the *first infinite ordinal* ω to be the set of all natural numbers, and continue with the construction above:

- $\omega = \{0, 1, \ldots, n, \ldots\}$;
- $\omega + 1 = \{0, 1, \ldots, n, \ldots, \omega\}$;
- $\omega + 2 = \{0, 1, \ldots, n, \ldots, \omega, \omega + 1\}$;
- $2\omega = \{0, 1, \ldots, n, \ldots, \omega, \omega + 1, \ldots, \omega + n, \ldots\}$;
- $k\omega = \{0, 1, \ldots, n, \ldots, \omega, \ldots, (k-1)\omega, \ldots, (k-1)\omega + n, \ldots\}$;
- $\omega^2 = \{0, \ldots, \omega, \ldots, n\omega, \ldots\}$;
- $\omega^2 + 1 = \{0, \ldots, \omega, \ldots, n\omega, \ldots, \omega^2\}$;
- $\omega^\omega = \{0, \ldots, \omega, \ldots, n\omega, \ldots, \omega^2, \ldots, \omega^n, \ldots\}$;
- $\varepsilon_0 = \omega^{\omega^{\omega^{\cdot^{\cdot^{\cdot}}}}}$.

Axiomatically, ordinals are defined as follows:

Definition 2.2. A set κ is *transitive* if every element of κ is a subset of κ. A transitive set κ is an *ordinal* if the relation "is either an element of or the same as" is a well-order on κ.

Note that the sets constructed above are indeed examples of ordinals. A valid (and important!) question to ask at this stage is to ask whether these sets even exist, as "…" is a super-vague notation. We won't discuss this here: see Chapter 2 of [**Jec06**].

If we let $\leq$ to denote the relation "is either an element of or the same as", then we see that ordinals are well-ordered:

$$1 \leq 2 \leq \cdots \leq n \leq \cdots \leq \omega \leq \omega + 1 \leq \cdots \leq n\omega \leq \cdots \leq \omega^2 \leq \cdots \leq \omega^\omega \leq \cdots .$$

This allows us to do two things. Firstly, we can now talk about *upper bounds* of a set of ordinals, hence the *supremum* of it, which exists by the well-ordering of ordinals. Note that we can write the supremum of a set of ordinals $\{\kappa_\alpha\}_\alpha$ as

$$\sup_\alpha \{\kappa_\alpha\}_\alpha = \bigcup_\alpha \kappa_\alpha,$$

which, in particular, shows that the supremum is also an ordinal. Secondly, any set O of ordinals (and each ordinal by itself as well!) can be turned into a topological space by giving it the order topology—the topology generated by the "open intervals"

$$(\kappa, \mu) = \{\nu \in O : \kappa < \nu < \mu\}$$

and the "half-open intervals"

$$[0, \kappa) = \{\nu \in O : \nu < \kappa\}.$$

In particular, if κ is an ordinal, then κ is precisely the half-open interval $[0, \kappa)$ containing every ordinal strictly smaller than κ.

We now state precisely why ordinals are the only examples of well-ordered sets. Two sets with order $(A, \leq)$ and $(B, \lesssim)$ are *order-isomorphic* if there exists a bijection $f : A \to B$ such that $x \leq y$ implies $f(x) \lesssim f(y)$. The following is a standard fact from set theory:

THEOREM 2.3. *Every well-ordered set is order-isomorphic to a unique ordinal.*

Combining the above theorem with Zermelo's well-ordering theorem (Theorem 2.1), we see that every set can be made order-isomorphic to an ordinal.

Recall that two sets A and B are said to be *equinumerous* if there exists a bijection $A \to B$. We write $|A| = |B|$ to denote the equinumerosity of A and B. A set A is *countable* if A is equinumerous to a natural number or ω. It is important to note that *every single example of ordinals we have produced above is countable.* Nevertheless, Cantor's diagonal argument shows that uncountable sets exist.

THEOREM 2.4 (Cantor). *A is never equinumerous to its power set $\mathcal{P}(A)$.*

In particular, uncountable ordinals exist.

At this stage, we exhibit the minimal example of an uncountable ordinal. To this end, we define ω_1 to be the supremum of all countable ordinals, which is an ordinal by the well-ordering of ordinals. Furthermore, ω_1 must be uncountable, for otherwise $\omega_1 + 1$ is countable.

A quick remark: we *cannot* say that ω_1 is equinumerous to, say, $\mathcal{P}(\omega)$, even though we'd be hard-pressed to find an example of an uncountable set whose "cardinality sits in bewteen those of ω and $\mathcal{P}(\omega)$". The existence of such a set is *independent* of the Zermelo-Fraenkel set theoretic axioms with the axiom of choice (ZFC), meaning we can neither prove nor disprove it within the realm of ordinary axiomatic set theory. This is the content of the famous *continuuum hypothesis.*

3. Moore–Smith Theory of Nets

Following [**Bre93**] (for the most part), we develope the Moore–Smith theory of nets in this section. In particular, we focus on generalizing the sequential characterization of compactness. Recall that sequential compactness and compactness are equivalent on metric spaces.

Theorem 3.1 (Sequential criterion for compactness in metric spaces). *A set K in a metric space is compact if and only if every sequence in K admits a convergent subsequence.*

Unfortunately, this criterion does not hold in the category of topological spaces.

Proposition 3.2. *The space $[0,1]^{[0,1]}$ of $[0,1]$-fold product of $[0,1]$ with the product topology is compact but not sequentially compact.*

Proof. By Tychonoff's theorem (Theorem 3.15), $[0,1]^{[0,1]}$ is compact. To see that $[0,1]^{[0,1]}$ is not sequentially compact, we define the nth binary expansion function $B_n : [0,1] \to [0,1]$ by setting $B_n(x)$ to be the nth digit of the binary expansion of x, so that $(B_n)_{n=1}^\infty$ is a sequence in $[0,1]^{[0,1]}$. Given a subsequence $(B_{n_k})_{k=1}^\infty$, we fix $x \in [0,1]$ such that

$$B_{n_k}(x) = \begin{cases} 0 & \text{if } k \text{ is odd;} \\ 1 & \text{if } k \text{ is even.} \end{cases}$$

Then $(B_{n_k}(x))_{k=1}^\infty$ clearly does not converge, whence neither does $(B_{n_k})_{k=1}^\infty$ in $[0,1]^{[0,1]}$. □

Proposition 3.3. *The first uncountable ordinal ω_1 with the order topology is sequentially compact but not compact.*

Proof. We refer readers unfamiliar with the construction of $\omega_1 = [0,\omega_1)$ with the order topology to Section 2 of this appendix. We observe first that ω_1 is noncompact, for the open cover

$$\{[0,\kappa) : \kappa \in [0,\omega_1)\}$$

clearly doesn't admit a finite subcover.

On the other hand, ω_1 is sequentially compact. To see this, we pick an arbitrary sequence $(\kappa_n)_{n=1}^\infty$ of ordinals in ω_1. We define the *limit superior* of the sequence by setting

$$\limsup_n \kappa_m = \min_{n \geq 0} \sup_{m \geq n} \kappa_m,$$

which exists by the well-ordering of ordinals. Note that

$$\limsup_n \kappa_m = \min_{n \geq 0} \bigcup_{m \geq n} \kappa_m.$$

Since a countable union of countable ordinals is a countable ordinal, we see that $\kappa = \limsup_n \kappa_m$ is a countable ordinal.

It therefore suffices to produce a subsequence subsequence of (κ_n) that converges to κ, which we could if κ were a limit point. We suppose for a contradiction that there exists an open interval (μ,ν) containing κ that does not contain any term of the sequence (κ_n). This, in particular, implies that κ cannot be attained as the supremum of a subsequence. By the well-ordering of ordinals, however, $\kappa = \sup_{m \geq N} \kappa_m$ for some $N \in \mathbb{N}$, which is absurd. □

Similarly, it is not necesarily true that $f(x_n) \to f(x)$ for all convergent sequences (x_n) implies the continuity of f. Sequential characterizations provide a convenient way to think about topological properties concretely, so it is desirable to introduce the notion of generalized sequences that will allow us to recover them. Let us first introduce generalized index sets:

Definition 3.4. A *directed set* is a partially ordered set $(D, \leq)$ such that every pair of elements α and β of D has a common maximum γ in D, i.e., $\gamma \geq \alpha$ and $\gamma \geq \beta$.

Note that the set of natural numbers with the usual ordering is a directed set. Observing that sequences are functions on the directed set $\mathbb{N}$, we make the following generalization:

Definition 3.5. A *net* in a topological space X is a function $\Phi : D \to X$ on a directed set D.

For convenience, we typically write $(x_\beta)_{\beta \in D}$ or (x_β) to detnote a net. The convergence of a net is defined analogously to that of a sequence:

Definition 3.6. A net (x_β) in X is *eventually in* $U \subseteq X$ if there exists an index β_0 such that $x_\beta \in U$ for all $\beta \geq \beta_0$. (x_β) *converges to* $x \in X$ if, for each neighborhood U of x, the net (x_β) is eventually in U.

Immediately, we recover the sequential characterization of continuity.

Proposition 3.7. *A function $f : X \to Y$ between topological spaces is continuous if and only if, for each $x \in X$ and every net (x_β) in X converging to x, the net $(f(x_\beta))$ converges to $f(x)$.*

Defining subnets requires some care. We would like to recover, for example, the Bolzano-Weierstrass property, but the notion of subnets given by truncating the initial terms of a net is too restrictive for this.

Definition 3.8. A function $h : D' \to D$ between directed sets is *final* if, for each $\beta_0 \in D$, there exists $\beta_0' \in D'$ such that $\beta' \geq \beta_0'$ implies $h(\beta') \geq \beta_0$. A *subnet* of a net $\Phi : D \to X$ in X is the composite map $\Phi \circ h$ with a final function h.

At this point, we present the required machinery for the net-theoretic proof of Tychonoff's theorem. The key concept, whose clever definition absorbs the bulk of the work, is as follows:

Definition 3.9. A net (x_β) in X is *universal* if, for each subset A of X, the net (x_β) is eventually in either A or $X \smallsetminus A$.

The next lemma, due to Kelley, shows that universal nets exist. In fact, there are lots of them. For the sake of brevity, we introduce the following definition.

Definition 3.10. A net (x_β) is *frequently in* $A \subseteq X$ if, for each index β_0, there exists an index $\beta \geq \beta_0$ such that $x_\beta \in A$.

Note that (x_β) is frequently in A if and only if (x_β) is *not* eventually in $X \smallsetminus A$.

Lemma 3.11 (Kelley's theorem). *Every net has a universal subnet.*

Proof. Let $(x_\beta)_{\beta \in D}$ be a net in a topological space X. We consider a collection $\mathscr{D}$ of all subsets $\mathcal{C}$ of $\mathcal{P}(X)$ such that

(1) if $A \in \mathcal{C}$, then (x_β) is frequently in A, and
(2) if $A, B \in \mathcal{C}$, then $A \cap B \in \mathcal{C}$.

$\mathscr{D}$ is nonempty, as $\mathcal{C} = \{X\}$ is such a set. The inclusion relation is a partial order on $\mathscr{D}$, and any chain on $\mathscr{D}$ has a maximum—namely, the union of the sets in the chain. It follows from Zorn's lemma that $(\mathscr{D}, \subseteq)$ has a maximal element $\mathcal{C}_0$.

We set $D' = \{(A, \beta) \in \mathcal{C}_0 \times D : x_\beta \in A\}$ and define a partial order $\leq$ on D' by setting

$$(A_1, \beta_1) \leq (A_2, \beta_2) \quad \text{if and only if} \quad A_1 \supseteq A_2 \text{ and } \beta_1 \leq \beta_2.$$

If $(A_1, \beta_1), (A_2, \beta_2) \in D'$, then $A = A_1 \cap A_2$ is in $\mathcal{C}_0$. Moreover, β_1 and β_2 have a common maximum β', and we can find $\beta \geq \beta'$ such that $x_\beta \in A$. It follows that (A, β) is a common maximum of (A_1, β_1) and (A_2, β_2), whence D' is a directed set. The function $(A, \beta) \mapsto \beta$ is a final map from D' to D, hence we obtain a subnet of (x_β), which we denote by $(x_{(A,\beta)})_{(A,\beta) \in D'}$.

We shall show that $(x_{(A,\beta)})$ is universal. We fix a subset E of X and suppose that $(x_{(A,\beta)})$ is frequently in E. We shall show that $(x_{(A,\beta)})$ is eventually in E, which establishes the desired result. For each index (A_0, β_0), we can find another index $(A, \beta) \geq (A_0, \beta_0)$ such that $x_{(A,\beta)} \in E$. By definition, $x_\beta = x_{(A,\beta)}$ is an element of A, whence $x_\beta \in E \cap A$. Furthermore, $A_0 \supseteq A$, and so $E \cap A_0 \supseteq E \cap A$. It follows that $x_\beta \in E \cap A_0$. Since β_0 was arbitrary, we conclude that (x_β) is frequently in $E \cap A_0$. By maximality of $\mathcal{C}_0$, the collection $\mathcal{C}_0$ must include E and $E \cap A_0$ for all $A_0 \in \mathcal{C}_0$. In particular, $(x_{(A,\beta)})$ cannot be frequently in $X \smallsetminus E$, for otherwise $E \cap (X \smallsetminus E) = \varnothing$ is in $\mathcal{C}_0$, which is absurd. We conclude that $(x_{(A,\beta)})$ is not frequently in $X \smallsetminus E$, hence eventually in E. □

We now establish three useful characterizations of compactness.

LEMMA 3.12. *Let X be a topological space. The following are equivalent.*

(1) X is compact.
(2) ***Universal net criterion****. Every universal net in X converges.*
(3) ***Bolzano-Weierstrass property****. Every net in X has a convergent subnet.*
(4) ***Finite intersection property criterion****. If a collection of closed sets in X has the* finite intersection property*—the intersection of any finite subcollection is nonempty—then the intersection of the entire collection is nonempty.*

PROOF. (1) ⇒ (2). Let X be a compact space and suppose for a contradiction that there exists a universal net (x_β) that does not converge. For each $x \in X$, we pick a neighborhood U_x such that (x_β) is *not* eventually in U_x. By universality, (x_β) is eventually in $X \smallsetminus U_x$, whence we can find an index β_x such that $\beta \geq \beta_x$ implies $x_\beta \notin U_x$. We now extract a finite subcover $\{U_{x_1}, \ldots, U_{x_n}\}$ of the open cover $\{U_x\}_{x \in x}$ and set $\beta_0 = \max\{\beta_{x_1}, \ldots, \beta_{x_n}\}$. Then, for each $\beta \geq \beta_0$, we see that $x_\beta \notin U_{x_i}$ for all $1 \leq i \leq n$, whence $x \notin X$. This is evidently absurd, and we conclude that every universal net converges.

(2) ⇒ (3). This is a trivial consequence of Kelley's theorem (Lemma 3.11).

(3) ⇒ (4). We let $\mathscr{C}$ be a collection of closed subsets of X with the finite intersection property. Adding in finite intersections of elements of $\mathscr{C}$ does not break the finite intersection property, so we assume without loss of generality that $\mathscr{C}$ is closed under finite intersection. This turns $\mathscr{C}$ into a directed set with $\supseteq$ as the partial order. We define a net on $\mathscr{C}$ by fixing an element x_C of each index set $C \in \mathscr{C}$.

By the Bolzano-Weiestrass property, we can find a directed set D and a final map $f : D \to \mathscr{C}$ such that the subnet $(x_{f(\beta)})_{\beta \in X}$ converges, say, to x. For each $C \in \mathscr{C}$, we can find $\beta_C \in D$ such that $\beta \geq \beta_C$ implies $f(\beta) \geq C$. Unraveling the definitions, we have the inclusion relation

$$x_{f(\beta)} \in f(\beta) \subseteq C$$

for all $\beta \geq \beta_C$, whence by the closedness of C the point x must be in C. Since C was arbitrary, it follows that $x \in \bigcap_{C \in \mathscr{C}} C$, as was to be shown.

(4) $\Rightarrow$ (1) Let $\{U_\alpha\}$ be an open cover of X. Let $C_\alpha = X \smallsetminus U_\alpha$ for each index α and consider the collection $\mathscr{C} = \{C_\alpha\}$. Since $\bigcup U_\alpha = X$, the intersection $\bigcap C_\alpha$ is empty, whence the contrapositive of (4) implies that there exists at least one finite subcollection $\{C_1, \ldots, C_n\}$ of $\mathscr{C}$ whose intersection is empty. Setting $U_n = C_n$, this implies that the union of the subcollection $\{U_1, \ldots, U_n\}$ of $\{U_\alpha\}$ is X, whence $\{U_1, \ldots, U_n\}$ is the desired finite subcover. □

We are almost ready to present a proof of Tychonoff's theorem. We require two more relatively simple lemmas.

Lemma 3.13. *If $\Phi : D \to X$ is a universal net and $f : X \to Y$ an arbitrary function between topological spaces, then $f \circ \Phi$ is a universal net in Y.*

Proof. We let $x_\beta = \Phi(\beta)$ for notational convenience. Given $A \subseteq Y$, the net (x_β) is eventually in either $f^{-1}(A)$ or $X \smallsetminus f^{-1}(A) = f^{-1}(Y \smallsetminus A)$. It follows that $f(x_\beta)$ is in either A or $Y \smallsetminus A$. □

Lemma 3.14. *A net $(x_\beta)_\beta$ in $\prod_\alpha X_\alpha$ converges to $x = (x^\alpha)$ if and only if each component net $(x^\alpha_\beta) = (\pi_\alpha(x_\beta))_\beta$ converges to $x^\alpha = \pi_\alpha(x)$.*

Proof. Suppose that $x_\beta \to x$. If U_α is a neighborhood of x^α, then we can find an index β_0 such that $x_\beta \in \pi^{-1}(U_\alpha)$ for all $\beta \geq \beta_0$, whence $x^\alpha_\beta \in U_\alpha$ for all $\beta \geq \beta_0$. It follows that $x^\alpha_\beta \to x^\alpha$.

Conversely, we suppose that $x^\alpha_\beta \to x^\alpha$ for all α. If U is a neighborhood of x, then we can find neighborhoods U_α of x^α such that $\prod_\alpha U_\alpha \subseteq U$ and $U_\alpha = X$ for all but finitely many α. If $U_\alpha = X_\alpha$, then x^α_β is trivially in U_α. We let $\alpha_1, \ldots, \alpha_n$ be the indices such that $U_{\alpha_i} \neq X_\alpha$, and find an index β_i such that $\beta \geq \beta_i$ implies $x^{\alpha_i}_\beta \in U_{\alpha_i}$. Setting $\beta_0 = \max\{\beta_1, \ldots, \beta_n\}$, we see that $\beta \geq \beta_0$ implies $x^\alpha_\beta \in U_\alpha$ for all α, hence in U. It follows that $x_\beta \to x$. □

Finally, the theorem and the proof:

Theorem 3.15 (Tychonoff). *If $\{X_\alpha\}$ is a family of compact topological spaces, then the product topology on $\prod_\alpha X_\alpha$ turns $\prod_\alpha X_\alpha$ into a compact topological space.*

Proof. Let (x_β) be a universal net in $\prod_\alpha X_\alpha$. By Lemma 3.13, $(x^\alpha_\beta)_\beta$ is a universal net in X_α for each α, whence by Lemma 3.12 it is a convergent net in X_α. We invoke the Axiom of Choice to fix a limit x^α of $(x^\beta_\alpha)_\beta$ for each α. By Lemma 3.14, the net (x_β) converges to $x = (x^\alpha)$ in $\prod_\alpha X_\alpha$. Therefore, every universal net in $\prod_\alpha X_\alpha$ converges, and it follows from Lemma 3.12 that $\prod_\alpha X_\alpha$ is compact. □

4. Reflexive Spaces

THEOREM 4.1 (Kakutani). *A Banach space is reflexive if and only if the norm unit ball is weak compact.*

PROOF. [**Bre11**], Theorem 3.17 □

THEOREM 4.2. *Every closed subspace of a reflexive Banach space is reflexive.*

PROOF WITH KAKUTANI'S THEOREM. [**Bre11**], Proposition 3.20 □

PROOF WITHOUT KAKUTANI'S THEOREM. [**Lax02**], Chapter 8, Theorem 15 □

THEOREM 4.3. *Let X be a reflexive Banach space, A a nonempty, closed, convex subset of X, and $\varphi : A \to (-\infty, \infty]$ a convex, lower semicontinuous function that is not $+\infty$ everywhere on A. If either A is bounded or $\varphi(x) \to +\infty$ as $\|x\| \to \infty$ in A, then there exists an $x_0 \in A$ such that*

$$\varphi(x_0) = \min_{x \in A} \varphi(x).$$

PROOF. [**Bre11**], Corollary 3.23 □

THEOREM 4.4 (Milman-Pettis). *Every uniformly convex Banach space is reflexive.*

PROOF. [**Bre11**], Theorem 3.31 □

5. Metrization Theorems on Weak and Weak-$*$ Topologies

In this section, we discuss the metrization theorems alluded to at the end of Chapter 1, Section 8.

THEOREM 5.1. *The weak topology on X is metrizable if and only if X is finite-dimensional.*

We shall require the following technical lemma.

LEMMA 5.2. *Every Hamel basis in an infinite-dimensional Banach space is unctounable.*

PROOF OF LEMMA. Let X be an infinite-dimensional Banach space and suppose for a contradiction that $\{v_n\}_{n\in\mathbb{N}}$ is a Hamel basis of X. For each $n \in \mathbb{N}$, we set $X_n = \operatorname{span}\{v_1, \ldots, v_n\}$. By Chapter 1, Corollary 2.19, each X_n is complete, hence closed in X. Furthermore, X_n evidently cannot contain a norm ball, hence X_n has empty interior. It now follows that $X = \bigcup_n X_n$ is of the first category, which contradicts the Baire category theorem (Chapter 1, Theorem 4.3). □

PROOF OF THEOREM, TAKEN FROM [Lec]. If X is finite-dimensional, then the weak topology is the same as the norm topology, which is evidently metrizable. Conversely, we suppose that the weak topology on X is induced by a metric d. We shall produce a countable spanning set of X^*. By Chapter 1, Corollary 3.10, X^* is a Banach space, and so Lemma 5.2 forces X^* to be finite-dimensional. This, in particular, implies that X^* is linearly isomorphic to X^{**}. Since X embeds into X^{**}, it follows that X must be finite-dimensional as well.

For each $n \in \mathbb{N}$, we take the metric ball $B_{1/n}(0)$ of radius $1/n$ centered at 0. By Chapter 1, Proposition 8.14, each n furnishes a finite collection $\mathcal{L}_n = \{l_{1,n}, \ldots, l_{k_n,n}\} \subseteq X^*$ and a positive scalar ε_n such that $\mathcal{V}_n = V_{l_{1,n},\ldots,l_{k_n,n};\varepsilon_n}$ is contained in $B_{1/n}(0)$. We shall show that $\mathcal{L} = \bigcup_n \mathcal{L}_n$ spans X^*.

To this end, we fix $l \in X^*$ and consider the weak neighborhood $V_{l;1}$ based at 0. There exists an $N \in \mathbb{N}$ such that $B_{1/N}(0) \subseteq V_{l;1}$, so that $\mathcal{V}_N \subseteq V_{l;1}$. We claim that l is a linear combination of the elements of $\mathcal{L}_N$. By Chapter 1, Lemma 8.11, it suffices to show that

$$\bigcap_{i=1}^{k_N} \ker(l_{i,N}) \subseteq \ker l.$$

If $l_{i,N}(x) = 0$ for all $1 \leq i \leq k_N$, then $l_{i,N}(Mx) = 0$ for all $M > 0$, whence $Mx \in \mathcal{V}_n \subseteq V_{l;1}$. This, in particular, implies that $|l(Mx)| < 1$, or $|l(x)| < 1/M$ for all $M > 0$, whence $l(x) = 0$. The proof is now complete. □

THEOREM 5.3. *The weak-$*$ topology on X^* is metrizable if and only if X is finite-dimensional.*

PROOF. (fill in the details) □

The following theorem shows that the norm-bounded subsets of the weak-$*$ dual is metrizable if and only if the underlying normed linear space is separable.

THEOREM 5.4. *A normed linear space X is separable if and only if every weak-$*$ compact subset of X^* is metrizable.*

PROOF OF THEOREM. (fill in the details)

□

Since metrizability guarantees the equivalence of compactness and sequential compactness, we have the following result.

THEOREM 5.5 (Helly). *If X is separable, then the norm unit ball of X^* is weak-$*$ sequentially compact.*

PROOF. The Banach-Alaoglu theorem (Theorem 8.5) implies that the norm unit ball is weak-$*$ compact. Since compactness is equivalent to sequential compactness in metrizable spaces, it follows from Theorem 5.4 that the norm unit ball is weak-$*$ sequentially compact. $\square$

6. Quotient Spaces

In this section, we take a more algebraic turn and discuss the quotient space formalism in the context of linear algebra and functional analysis.

DEFINITION 6.1. Let V be a vector space and W a linear subspace of W. The *quotient vector space* V/W is defined to be the set of equivalence classes induced by the equivalence relation

$$x \sim y \text{ if and only if } x - y \text{ is in } W.$$

The vector space structure on V/W is defined as follows:

- $[x] + [y] = [x + y]$;
- $\lambda[x] = [\lambda x]$.

For each quotient space V/W, there is the *canonical projection* $\pi : V \to V/W$ given by the formula $\pi(x) = [x]$. We observe at once that π is a surjective linear transformation. Combined with the *canonical embedding* $\iota : W \to V$ given by the formula $\iota(x) = x$, we obtain a *short exact sequence*

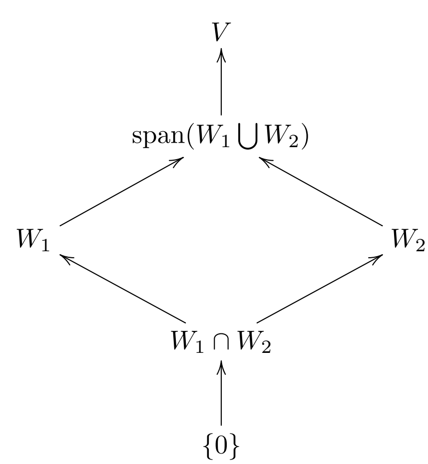
$$0 \to W \xrightarrow{\iota} V \xrightarrow{\pi} V/W \to 0.$$

We shall neither define nor discuss short exact sequences in this section: for now, it suffices to note that the above is the model example of a short exact sequence.

Note that quotienting V out by W collapses every element of W onto the zero vector. Indeed, the equivalence class $[0]$ consists precisely of the vectors in W. It is then natural to expect that the quotient space V/W is smaller if W is larger. More precisely, if we define a partial order on the set of all linear subspaces of V by setting

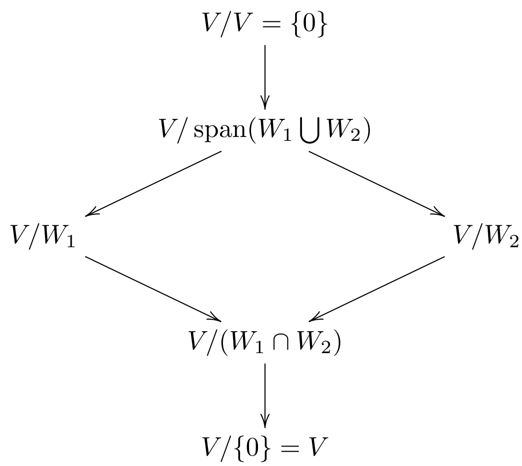
$$W_1 \leq W_2 \text{ if and only if } W_1 \text{ is a linear subspace of } W_2, \tag{6.2}$$

then, given any two linear subspaces W_1 and W_2 of V, we can draw a *lattice of linear subspaces*



where the arrows indicate an increase in size with respect to the partial order $\leq$.

We contend that the act of quotienting out reverses the arrows in the lattice, i.e.,



where the arrows now indicate the "embeds to" relation. The precise statement is as follows:

PROPOSITION 6.3. *The relation*

$$V/W_1 \lesssim V/W_2 \Leftrightarrow \text{ there exists an embedding } V/W_1 \to V/W_2 \tag{6.4}$$

is a reflexive, transitive relation on the set quotient spaces of V such that $W_1 \geq W_2$ in the sense of (6.2) implies $V/W_1 \lesssim V/W_2$. □

We omit the proof of the above proposition. Note that $\lesssim$ is not quite a partial order, for it is possible for two non-equal quotient spaces to be isomorphic, thereby admitting embeddings in both directions. The correct set-up would be to consider *isomorphism classes* of subspaces of V and quotient spaces of V, respectively, and consider the partial orders on these sets given by the embedding relation (6.4). The antisymmetry property that failed in the setting of Proposition 6.3 now holds, and this is a consequence of the following theorem:

THEOREM 6.5 (Schröder-Bernstein for vector spaces). *If two vector spaces X and Y admit embeddings $X \to Y$ and $Y \to X$, then X is linearly isomorphic to Y.*

See [**MO1**] for a discussion on the Schröder-Bernstein property on different categories. We do not pursue this matter further in these notes.

Returning to the topic at hand, we observe that shrinking the vector space reduces the size of the corresponding quotient space. Once again, we omit the proof of the following proposition.

PROPOSITION 6.6. *Let V be a vector space and W a linear subspace of V. Every linear subspace of the quotient space V/W is of the form X/W, where $W \leq X \leq V$ in the sense of (6.2). Conversely, if X is a linear subspace of V containing W, then X/W is a linear subspace of V/W.* □

With this we now leave behind the order properties of quotient spaces and establish the basic tools of the trade.

THEOREM 6.7 (Splitting lemma for vector spaces). *If V is a vector space and W a linear subspace of W, then*

$$V \cong W \oplus V/W$$

in the category of vector spaces.

PROOF. We take a Hamel basis $\{w_\alpha\}_\alpha$ of W and extend it to a Hamel basis $\{w_\alpha, v_\beta\}_{\alpha,\beta}$ of V. It is clear that $V \cong W \oplus \operatorname{span}\{v_\beta\}_\beta$. whence it suffices to show that $V/W \cong \operatorname{span}\{v_\beta\}$. To this end, we let $\pi : V \to V/W$ be the canonical projection and take For each $[x] \in V/W$, we take the basis expansion

$$x = \sum_{\alpha,\beta} \lambda_\alpha w_\alpha + \lambda_\beta v_\beta$$

and apply π to obtain the linear combination

$$[x] = \pi(x) = \sum_{\alpha,\beta} \lambda_\alpha \pi(w_\alpha) + \lambda_\beta \pi(v_\beta) = \sum_\beta \lambda_\beta \pi(v_\beta),$$

whence $\mathscr{V}$ is a spanning set in V/W. Furthermore, $\mathscr{V}$ is linearly independent, as

$$0 = \sum_\beta \lambda_\beta \pi(v_\beta)$$

implies that $\sum \lambda_\beta v_\beta \in \ker \pi = W$, which is only possible if $\lambda_\beta = 0$ for all β. It follows that $\mathscr{V}$ is a basis of V/W, and the desired result follows. $\square$

We now recall that the *kernel* of a linear transformation $T : X \to Y$ between vector spaces is the linear subspace

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

A fundamental result that connects kernels and images of a linear transformation is as follows:

THEOREM 6.8 (First isomorphism theorem for vector spaces). *If $T : X \to Y$ is a linear transformation, then*

$$X/\ker T \cong \operatorname{im} T$$

in the category of vector spaces.

PROOF. We define $\tilde{T} : X/\ker T \to \operatorname{im} T$ by setting $\tilde{T}[x] = Tx$. This map is evidently surjective. Moreover, $\tilde{T}[x] = 0$ if and only if $x \in \ker T = [0]$, whence $\ker \tilde{T}$ is trivial. It follows that $\tilde{T}$ is injective, and the desired result follows. $\square$

Combining the splitting lemma and the first isomorphism theorem, we obtain the following generalization of the usual rank-nullity theorem:

COROLLARY 6.9 (Rank-nullity theorem). *If $T : X \to Y$ is a linear transformation, then*

$$X \cong \operatorname{im} T \oplus \ker T.$$

In particular, if X and Y are finite-dimensional, then

$$\dim X = \operatorname{rank} T + \operatorname{nullity} T.$$

We now consider quotient spaces of a normed linear space.

DEFINITION 6.10. Let $(X, \|\cdot\|)$ be a normed linear space and Y a subspace of X. The *quotient norm* on the quotient vector space X/Y is given by

$$\|[x]\| = \inf_{z\in[x]} \|z\| = \inf_{y\in Y} \|x+y\|. \tag{6.11}$$

Note that the canonical projection $\pi : X \to X/Y$ is a bounded linear operator. Indeed,

$$\|\pi(x)\| = \inf_{z\in[x]} \|z\| \leq \|x\|.$$

The next result shows that act of taking quotients preserves completeness in many cases.

PROPOSITION 6.12. *If X is a Banach space and Y is a closed linear subspace of X, then X/Y is a Banach space.*

PROOF. Let $([x_n])_{n=1}^{\infty}$ be a Cauchy sequence in X/Y and extract a subsequence $([x_{n_k}])_{k=1}^{\infty}$ such that

$$\|[x_{n_k} - x_{n_{k+1}}]\| = \|[x_{n_k}] - [x_{n_{k+1}}]\| < 2^{-k}$$

for each $k \in \mathbb{N}$. We now pick $z_0 \in [x_1]$ and fix, for each $k \in \mathbb{N}$, an element $z_k \in [x_{n_k} - x_{n_{k+1}}]$ such that $\|z_k\| < 2^{-k}$. Setting $a_K = z_0 + \cdots + z_K$, we see that $(a_k)_{k=1}^{\infty}$ is a Cauchy sequence in X, whence it converges to a limit $a \in X$. Now,

$$a - a_K = \sum_{k=K+1}^{\infty} z_k$$

is an element of $[a] - [a_K] = [a] - [x_{n_K}] = [a - x_{n_K}]$, and so

$$\|[a] - [x_{n_K}]\| \leq \left\| \sum_{k=K+1}^{\infty} z_k \right\| \leq \sum_{k=K+1}^{\infty} \|z_k\| < 2^{-K},$$

It follows that $[x_{n_K}] \to [a]$ in X/Y, and the desired result follows. □

The first isomorphism theorem generalizes readily to the category of Banach spaces, taken as a subcategory of the category of topological vector spaces. We remark that the isomorphism we get is *not* an isometric isomophism in general.

THEOREM 6.13 (First isomorphism theorem for Banach spaces). *If $T : X \to Y$ is a bouned linear transformation, then*

$$X/\ker T \cong \operatorname{im} T$$

in the category of topological vector spaces.

PROOF. We take the linear isomorphism $\tilde{T} : X/\ker T \to T$ constructed in the proof of Theorem 6.8. We observe that

$$\|\tilde{T}[x]\| = \|Tz\| \leq \|T\|\|z\|$$

for each $z \in [x]$, whence taking the infimum over $[x]$ yields

$$\|\tilde{T}[x]\| \leq \|T\|\|[x]\|.$$

Therefore, $\tilde{T}$ is bounded. It now follows from the open mapping theorem (Chapter 1, Corollary 4.6) that $\tilde{T}^{-1}$ is bounded. □

7. Annihilators

In this section, we generalize the notion of orthogonal complements to Banach spaces. Here we make use of the quotient space formalism and the isomorphism theorems that come with it, as developed in Section 6 of this appendix. We also expect the reader to have read Section 4.

DEFINITION 7.1. Let X be a topological vector space. The *annihilator* of a subset Y of X is the collection of all continuous linear functionals $l : X \to \mathbb{F}$ that vanish on Y. The annihilator of Y is denoted by $Y^\perp$.

If X is a Hilbert space, then the Riesz representation theorem (Chapter 2, Theorem 1.20) implies that $Y^\perp$ is precisely the orthogonal complement (Chapter 2, Definition 1.14) of Y.

If X is a normed linear space and Y a closed subspace of X, then Chapter 1, Corollary 5.10 guarantees that $Y^\perp$ is nontrivial as long as Y is a proper subspace. The following proposition quantifies this phenomenon precisely.

PROPOSITION 7.2. *If X is a Banach space and Y a closed linear subspace of X, then $Y^\perp \cong (X/Y)^*$ and $Y^* \cong X^*/Y^\perp$ in the category of topological vector spaces.*

PROOF. To establish the first isomorphism, we let $\pi : X \to X/Y$ be the canonical projection onto the quotient space X/Y. Observe that the adjoint $\pi^* : (X/Y)^* \to X^*$ is injective. Indeed, if $\pi^*(l_1) = \pi^*(l_2)$, then

$$l_1 = l_1\pi = \pi^*(l_1) = \pi^*(l_2) = l_2\pi = l_2$$

by the surjectivity of π. In particular, $\ker \pi^*$ is trivial, and the first isomorphism theorem (Theorem 6.13) implies that $(X/Y)^* \cong Y^\perp$, provided that $\operatorname{im} \pi^* = Y^\perp$. But this is trivial, as $\pi^*(l) = l\pi$ are precisely the linera functionals that vanish on Y.

To establish the second isomorphism, we let $\iota : Y \to X$ be the canonical embedding. Observe that the adjoint $\iota^* : X^* \to Y^*$ is surjective. Indeed, if $l \in Y^*$, then the analytic Hahn-Banach theorem (Chapter 1, Theorem 5.9) furnishes an $L \in X^*$ such that

$$\iota^*(L) = L\iota = L|_Y = l.$$

It now follows from the first isomorphism theorem (Theorem 6.13) that implies that $X^*/Y^\perp \cong Y^*$ in the category of topological vector spaces. provided that $\ker \iota^* = Y^\perp$. But this is trivial, as $\iota^*(l) = l\iota = l|_Y$. □

We remark that the isomorphism $Y^\perp \cong (X/Y)^*$ by itself does not guarantee the nontriviality of $Y^\perp$. The Hahn-Banach theorem, however, guarantees that X^* is always "bigger" than Y^*. Therefore, if Y is a proper subspace, then $Y^\perp$ must be nontrivial to make up for this difference, as quantified by the isomorphism $Y^* \cong X^*/Y^\perp$.

Tacitly used in the proof of the above proposition is an assumption that annihilators are subspaces. In fact, the quotient construction requires $Y^\perp$ to be a *closed* subspace of X^*. We address this in the following proposition, which does not require completeness.

PROPOSITION 7.3. *If Y is a subset of a normed linear space X, then $Y^\perp$ is a closed linear subspace of X^*. Furthermore,*

$$Y^\perp = [\overline{\operatorname{span}(Y)}]^\perp. \tag{7.4}$$

PROOF. Whenever $l_1, l_2 \in Y^\perp$ and $\lambda \in \mathbb{F}$

$$(\lambda l_1 + l_2)(y) = \lambda l_1(y) + l_2(y) = 0 + 0 = 0$$

for all $y \in Y$. Therefore, $Y^\perp$ is a linear subspace of X^*. To show that $Y^\perp$ is closed, we pick a sequence $(l_n)_{n=1}^\infty$ in $Y^\perp$ converging to $l \in X^*$ and observe that

$$|l(y)| = |l(y) - l_n(y)| = \lim_{n\to\infty} |l(y) - l_n(y)| \leq \lim_{n\to\infty} \|l - l_n\|\|y\| = 0$$

for all $y \in Y$.

It remains to verify the formula (7.4). $Y \subseteq \overline{\operatorname{span}(Y)}$ clearly implies that $Y^\perp \supseteq [\overline{\operatorname{span}(Y)}]^\perp$, and so it suffices to show the reverse inclusion. But for each $l \in Y^\perp$, $l|_{\operatorname{span}(Y)} = 0$ holds trivially and the desired result now follows form the continuity of l. □

Experiences with finite-dimensional inner-product spaces suggest that the orthogonal complement of the orthogonal complement of a subspace should be the subspace itself. This is not true in general, but we can prove an analogous result for reflexive Banach spaces.

PROPOSITION 7.5. *If Y is a subset of a reflexive Banach space X, then*

$$(Y^\perp)^\perp \cong \overline{\operatorname{span}(Y)}.$$

In particular, if Y is a closed linear subspace, then $(Y^\perp)^\perp \cong Y$.

PROOF. We prove the "in particular" part of the proposition, from which we can derive the proposition in its full generality via Proposition 7.3. We therefore assume that Y is a closed linear subspace of X. By Proposition 7.2, we see that

$$(Y^\perp)^\perp \cong (X^*/Y^\perp)^* \cong (Y^*)^*.$$

It now follows from Theorem 4.2 that $(Y^\perp)^\perp \cong Y$. □

Annihilators provide a useful way of thinking about dense subspaces of a space whose dual space admits a concrete characterizations.

EXAMPLE. Let $1 \leq p < \infty$. By the L^p Riesz representation theorem, the dual of $L^p(X,\mu)$ is isometrically isomorphic to $L^q(X,\mu)$, where $1/p + 1/q = 1$. By Propositions 7.2 and 7.3, a subspace $\mathcal{D}$ of $L^p(X,\mu)$ is dense in $L^p(X,\mu)$ if and only if each $u \in L^q(X,\mu)$ satisfying

$$\int_X fu \, d\mu = 0$$

for all $f \in \mathcal{D}$ must be the zero function. Standard examples of dense subspaces of $L^p(\mathbb{R}^d)$ include the space of simple functions on cubes, the space $\mathscr{C}(\mathbb{R}^d)$ of continuous functions, the Schwartz space $\mathscr{S}(\mathbb{R}^d)$, and the space $\mathscr{C}_c^\infty(\mathbb{R}^d)$ of continuous functions with compact support.

Familiar examples of denses subspaces abound if we restrict our attention to compact subsets of $\mathbb{R}^d$. For example, the Weierstrass approximation theorem asserts that polynomials on $[a,b]$ are dense in $\mathscr{C}([a,b])$, which is then dense in $L^p([a,b])$. It now follows from (7.4) in Proposition 7.3 that any $u \in L^q$ satisfying $\int_a^b f(x)x^n \, dx = 0$ for all $n \in \mathbb{N}$ must be the zero function. Similarly, the L^2 Fourier inversion theorem shows that trigonometric polynomials on $[-\pi,\pi]$ are dense in $L^2([-\pi,\pi])$, whence any $f \in L^2$ satisfying $\hat{f}(n) = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(x)e^{-inx} \, dx$ for all $n \in \mathbb{N}$ must be the zero function. In particular, the Fourier series of an L^2-function is unique.

8. Distributions

CHAPTER 4

Appendix B - Solutions to Exercises and Miscellaneous Problems

1. Exercises from the Main Text

1.1. Problems and Solutions for Chapter 1.

Problem 1 (Exercise **??**). Check that this topology turns vector addition $(x, y) \mapsto x + y$ and scalar multiplication $(\lambda, x) \mapsto \lambda x$ into continous maps.

Proof. Let X be a normed linear space. Before we check the continuity of the operations, we must discuss the topologies on $X \times X$ and $\mathbb{F} \times X$.

Lemma 1.1. *Let Y be another normed linear space. $\|\cdot\|_{X\times Y}$, given by*

$$\|(x, y)\|_{X\times Y} = \|x\|_X + \|y\|_Y$$

for each $(x, y) \in X \times Y$, is a norm on $X \times Y$ that generates the product topology.

Proof of lemma. The properties of the norm are trivially verified. For example, we observe that

$$\begin{aligned}\|(x_1 + x_2, y_1 + y_2)\|_{X\times Y} &= \|x_1 + x_2\|_X + \|y_1 + y_2\|_Y \\ &\leq (\|x_1\|_X + \|y_1\|_Y) + (\|x_2\|_X + \|y_2\|_Y) \\ &= \|(x_1, y_1)\|_{X\times Y} + \|(x_2, y_2)\|_{X\times Y}.\end{aligned}$$

for all $(x_1, y_1), (x_2, y_2) \in X \times Y$, and so the triangle inequality holds.

We now fix a point (x_0, y_0) in an arbitrary product ball $B_r(x) \times B_s(y)$, and find two open balls $B_{r_0}(x_0)$ and $B_{s_0}(y_0)$ contained in $B_r(x)$ and $B_s(y)$, respectively. Then the open ball

$$\tilde{B}_{\min(r_0,s_0)}((x_0, y_0)) = \{(v, w) \in X \times Y : \|(v, w) - (x_0, y_0)\|_{X\times Y} < \min(r_0, s_0)\}$$

in the norm topology of $X \times Y$ is a norm-open neighborhood of (x_0, y_0) contained in $B_r(x) \times B_s(y)$, for each (v, w) in the ball satisfies the estimates

$$\begin{aligned}\|v - x_0\|_X &\leq \|(v, w) - (x_0, y_0)\|_{X\times Y} < r_0 \\ \|w - y_0\|_Y &\leq \|(v, w) - (x_0, y_0)\|_{X\times Y} < s_0.\end{aligned}$$

It follows that $B_r(x) \times B_s(y)$ is open in the norm topology of $X \times Y$. Since the product topology on $X \times Y$ is generated by the product balls, we conclude that every product-open set in $X \times Y$ is norm-open.

Conversely, we take an arbitrary norm ball

$$\tilde{B}_t(x, y) = \{(v, w) \in X \times Y : \|(v, w) - (x, y)\|_{X\times Y} < t\}$$

in $X \times Y$ and fix an element $(x_0, y_0) \in \tilde{B}_t(x, y)$. Set $r = \|x_0 - x\|_X$ and $s = \|y_0 - y\|$. The product ball $B_{t/2-r}(x_0) \times B_{t/2-s}(y_0)$ is a product-open neighborhood of (x_0, y_0)

contained in the norm ball $\tilde{B}_t(x,y)$, for each (v,w) int he product ball satisfies the estimate

$$\begin{aligned}\|(v,w)-(x,y)\|_{X\times Y} &= \|v-x\|_X + \|w-y\|_Y \\ &\leq \|v-x_0\|_X + \|x_0-x\|_X + \|w-y_0\|_Y + \|y_0-y\|_Y \\ &< \left(\frac{t}{2}-r\right) + r + \left(\frac{t}{2}-s\right) + s \\ &= t.\end{aligned}$$

It follows that $\tilde{B}_t(x,y)$ is open in the product topology of $X \times Y$. Since the norm topology on $X \times Y$ is generated by the norm balls, we conclude that every norm-open set in $X \times Y$ is product-open.

We have thus shown that the norm topology and the product topology agree with each other. □

The (Lipschitz) continuity of vector addition follows trivially from the lemma, for

$$\|x+y\|_X \leq \|(x,y)\|_{X\times X}$$

by the triangle inequality. As for scalar multiplication, we observe that

$$\|(\lambda,x)\|_{\mathbb{F}\times X} < \sqrt{\varepsilon}$$

yields the estimate

$$\varepsilon > \|(\lambda,x)\|^2_{\mathbb{F}\times X} = |\lambda|^2 + \|x\|^2_X + 2|\lambda|\|x\|_X \geq 2|\lambda|\|x\|_X > |\lambda|\|x\|_X = \|\lambda x\|_X,$$

which establishes continuity. □

Problem 2 (Exercise **??**). Show that every vector space X with a translation-invariant, homogeneously-scaling metric is a normed vector space, with the norm given by $\|x\| = d(x,0)$.

Proof. Clearly, $\|x\| = d(x,0) = 0$ if and only if $x = 0$, and

$$\|\lambda x\| = d(\lambda x, 0) = |\lambda| d(x,0) = |\lambda|\|x\|.$$

As for the triangle inequality, we observe that

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

as was to be shown. □

Problem 3 (Exercise **??**). Verify that the examples introduced above are indeed normed linear spaces.

Proof. For all of the examples, verifying triangle inequality is the only non-trivial part. We begin by establishing the triangle inequality for the L^p-norm, a result more commonly known as *Minkowski's inequality.* If $p = \infty$, the inequality is a direct consequence of the ordinary triangle inequality in $\mathbb{C}$:

$$\|f+g\|_\infty = \operatorname{ess\,sup} |f+g| \leq \text{e.s.}(|f|+|g|) \leq \text{e.s.}\,|f| + \text{e.s.}\,|g| = \|f\|_\infty + \|g\|_\infty.$$

Incidentally, the triangle inequality for the "Sobolev norm" $\|f\| = \|f\|_\infty + \|f'\|_\infty$ on $\mathcal{C}^1[0,1]$ follows directly from this observation as well. If $1 \leq p < \infty$, we assume without loss of generality that $\|f\|_p + \|g\|_p = 1$, dividing f and g by $\|f\|_p + \|g\|_p$ if necessary. We set $\theta = \|g\|_p$, so that $\|f\|_p = 1-\theta$. If θ is either 0 or 1, then the desired result reduces to a triviality, hence we assume that $0 < \theta < 1$.

Letting $F = (1-\theta)^{-1}f$ and $G = \theta^{-1}g$, we see that the inequality to be established becomes

$$\|(1-\theta)F + \theta G\|_p = \|f+g\|_p \leq \|f\|_p + \|g\|_p = 1, \tag{1.2}$$

with $\|F\|_p = |1-\theta|^{-1}(1-\theta) = 1$ and $\|G\|_p = |\theta|^{-1}\theta = 1$. Now, we observe that the convexity of the map $z \mapsto |z|^p$ yields the pointwise estimate

$$|(1-\theta)F(x) + \theta G(x)|^p \leq (1-\theta)|F(x)|^p + \theta|G(x)|^p,$$

and integrating this gives us

$$\|(1-\theta)F(x) + \theta G(x)\|_p^p \leq (1-\theta)\|F\|_p^p + \theta\|G\|_p^p = (1-\theta) + \theta = 1.$$

We can now obtain (1.2) by taking the pth root, completing the proof.

Finally, we check that the triangle inequality holds for the (A,p)-norm

$$\|x\|_{A,p} = \|Ax\|_p$$

on $\mathbb{F}^n$. This is easy to see, as

$$\|x+y\|_{A,p} = \|A(x+y)\|_p = \|Ax + Ay\|_p \leq \|Ax\|_p + \|Ay\|_p \leq \|x\|_{A,p} + \|y\|_{A,p}$$

by Minkowski's inequality. We remark that the invertibility of A is necessary, for otherwise the (A,p)-norm of a nonzero vector can be zero. Indeed, if A is singular, then $Ax = 0$ for any vector x in the nullspace of A, which is nontrivial.

In general, the (A,p)-norm

$$\|f\|_{A,p} = \|Af\|_p$$

is a norm on the Lebesgue space $L^p(X,\mu)$ whenever A is a linear automorphism on $L^p(X,\mu)$. The $\mathbb{F}^n$ case can be recovered by considering the L^p-space of $\mathbb{F}$-valued functions on a base set X of cardinality n. □

Problem 4 (Exercise **??**). Let $(X, \|\cdot\|)$ be a normed linear space and M a linear subspace of X. Show that $\overline{M}$ is also a linear subspace of X. Moreover, if $(x_n)_{n=1}^\infty$ is a sequence in X, then show that $x_n \to x$ implies $\|x_n\| \to \|x\|$ in $\mathbb{R}$. Show that if $(x_n)_{n=1}^\infty$ is a Cauchy sequence, then $(\|x_n\|)_{n=1}^\infty$ is a Cauchy sequence.

Proof. Let $x, y \in \overline{M}$ and find sequences $(x_n)_{n=1}^\infty$ and $(y_n)_{n=1}^\infty$ in M such that $x_n \to x$ and $y_n \to y$ in the norm topology. For each $\lambda \in \mathbb{F}$, we have the estimate

$$\|(\lambda x_n + y_n) - (\lambda x + y)\| \leq \|\lambda(x_n - x) + (y_n - y)\| \leq |\lambda|\|x_n - x\| + \|y_n - y\|,$$

so that $\lambda x_n + y_n \to \lambda x + y$. It follows that $\lambda x + y \in \overline{M}$, and so $\overline{M}$ is a linear subspace of X. Furthermore, we have the bound

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

whence $\|x_n\| \to \|x\|$. A similar bound, i.e.,

$$|\|x_n\| - \|x_m\|| \leq \|x_n - x_m\|$$

implies that $(\|x_n\|)_{n=1}^\infty$ is Cauchy if $(x_n)_{n=1}^\infty$ is. □

Problem 5 (Exercise **??**). Check that

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

is a translation-invariant metric on $\mathbb{R}^{\mathbb{N}}$ that does not scale homogenously. This is an example of a *Frechét space* (Chapter 1, Definition 9.7), a generalization of a Banach space with multiple, rather than one, "complete norms".

PROOF. We first show that d is indeed a metric on $\mathbb{R}^{\mathbb{N}}$. If $x = y$, then $x_n = y_n$ for each $n \in \mathbb{N}$, and so

$$d(x, y) = \sum_{n=1}^{\infty} 2^{-n} \cdot 0 = 0.$$

If $x \neq y$, then there exists at least one $N \in \mathbb{N}$ such that $x_N \neq y_N$. Since $\min(1, |x_n - y_n|) \geq 0$ for all $n \in \mathbb{N}$, we see that

$$d(x, y) = \sum_{n=1}^{\infty} 2^{-n} \min(1, |x_n - y_n|) \geq 2^{-N} \min(1, |x_N - y_N|) > 0,$$

for neither 1 nor $|x_N - y_N|$ is zero. Symmetry $d(x, y) = d(y, x)$ is trivial, and the triangle inequality is easily established as well:

$$\begin{aligned}
d(x, y) &= \sum_{n=1}^{\infty} 2^{-n} \min(1, |x_n - y_n|) \\
&= \sum_{n=1}^{\infty} 2^{-n} \min(1, |(x_n - z_n) + (z_n - y_n)|) \\
&\leq \sum_{n=1}^{\infty} 2^{-n} \min(1, |(x_n - z_n)| + |(z_n - y_n)|) \\
&\leq \sum_{n=1}^{\infty} 2^{-n} [\min(1, |(x_n - z_n)|) + \min(1, |(z_n - y_n)|)] \\
&= d(x, z) + d(z, y).
\end{aligned}$$

Translation invariance also follows from routine computations:

$$\begin{aligned}
d(x + z, y + z) &= \sum_{n=1}^{\infty} 2^{-n} \min(1, |(x_n + z_n) - (y_n + z_n)|) \\
&= \sum_{n=1}^{\infty} 2^{-n} \min(1, |x_n - y_n + z_n|) \\
&= d(x, y).
\end{aligned}$$

We now produce an example to show that d does not scale homogeneously. We let $x_n = 1$ and $y_n = 0$ for each $n \in \mathbb{N}$ and fix a scalar λ such that $|\lambda| > 1$. We then see that

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

as was to be shown. □

PROBLEM 6 (Exercise **??**). Show that equivalent norms induce the same metric topology, viz., a subset E is open with respect to $\| \cdot \|_a$ if and only if E is open with respect to $\| \cdot \|_b$.

Proof. We suppose that there exist two positive constants K_1 and K_2 such that

$$K_1\|x\|_a \leq \|x\|_b \leq K_2\|x\|_a$$

for all $x \in X$. If E is open with respect to $\|\cdot\|_a$, then, for each $x \in E$, then we can find a number $r_x > 0$ such that $y \in E$ whenever $\|x-y\|_a < r_x$. By equivalence of norms, $\|x-y\|_b < K_1^{-1}r_x$ implies $\|x-y\|_a < r_x$, so that the b-ball of radius $K_1^{-1}r_x$ is contained in E. If follows that E is open with respect to $\|\cdot\|_b$.

Similarly, if E is open with respect to $\|\cdot\|_b$, then, for each $x \in E$, then we can find a number $R_x > 0$ such that $y \in E$ whenever $\|x-y\|_b < R_x$. By equivalence of norms, $\|x-y\|_a < K_2R_x$ implies $\|x-y\|_b < R_x$, so that the a-ball of radius K_2R_x is contained in E. If follows that E is open with respect to $\|\cdot\|_a$. ☐

Problem 7 (Exercise **??**). If $T : X \to Y$ is a bounded linear operator, then

$$\|T\|_{X\to Y} = \sup_{x\neq 0}\frac{\|Tx\|_Y}{\|x\|_X} = \sup_{\|x\|_X=1}\|Tx\|_Y.$$

Proof. If x is a unit vector, then

$$\frac{\|Tx\|_Y}{\|x\|_X} = \|Tx\|_Y,$$

and so

$$\sup_{x\neq 0}\frac{\|Tx\|_Y}{\|x\|_X} \geq \sup_{\|x\|_X=1}\|Tx\|_Y.$$

Conversely, if x is a nonzero vector, then

$$\frac{\|Tx\|_Y}{\|x\|_X} = \|T(\|x\|_X^{-1}x)\|_Y,$$

where $\|x\|_X^{-1}x$ is a unit vector. Therefore,

$$\sup_{x\neq 0}\frac{\|Tx\|_Y}{\|x\|_X} \leq \sup_{\|x\|_X=1}\|Tx\|_Y,$$

and the equality holds.

It now suffices to show that

$$\|T\|_{X\to Y} = \sup_{\|x\|_X=1}\|Tx\|_Y.$$

To see this, we let y be a nonzero vector and observe that

$$\|Ty\|_Y = \|T(\|y\|_X^{-1}y)\|\|y\| \leq \left(\sup_{\|x\|_X=1}\|Tx\|_Y\right)\|y\|,$$

whence

$$\sup_{\|x\|_X=1}\|Tx\|_Y \geq \|T\|_{X\to Y}.$$

Conversely, we have

$$\|Tx\| \leq \|T\|_{X\to Y}\|x\| = \|T\|_{X\to Y}$$

for every unit vector x, and so

$$\sup_{\|x\|_X=1}\|Tx\|_Y \leq \|T\|_{X\to Y},$$

and the equality holds. ☐

PROBLEM 8 (Exercise **??**). The space $\mathscr{B}(X,Y)$ of bounded linear operators from X to Y is a normed linear space with the operator norm.

PROOF. Continuity of $\mathbb{F}$-valued functions is preserved under addition and multiplication by scalar, hence $\mathscr{B}(X,Y)$ is a vector space. The zero operator clearly has the norm of 0. If T is a nontrivial bounded operator from X to Y, then there exists a vector $x \in X$ such that $\|Tx\| > 0$, so that $\|T\| > 0$. For each $\lambda \in \mathbb{F}$ and every $T \in \mathscr{B}(X,Y)$, we see that

$$\|\lambda T\| = \sup_{\|x\|=1} \|\lambda Tx\| = |\lambda| \sup_{\|x\|=1} \|Tx\| = |\lambda|\|T\|.$$

Finally, if $T_1, T_2 \in \mathscr{B}(X,Y)$, then

$$\begin{aligned} \|T_1 + T_2\| &= \sup_{\|x\|=1} \|(T_1 + T_2)x\| \\ &= \sup_{\|x\|=1} \|T_1 x + T_2 x\| \\ &\leq \sup_{\|x\|=1} \|T_1 x\| + \sup_{\|x\|=1} \|T_2 x\| \\ &= \|T_1\| + \|T_2\|. \end{aligned}$$

We conclude that $\mathscr{B}(X,Y)$ is a normed linear space. □

PROBLEM 9 (Exercise **??**). $\|T_n \to T\|_{X\to Y} \to 0$ if and only if

$$\sup_{\|x\|_X=1} \|T_n x - Tx\|_Y \to 0.$$

PROOF. Obvious, as $\|T_n - T\| = \sup_{\|x\|_X=1} \|T_n x - Tx\|_Y$ for each $n \in \mathbb{N}$. □

PROBLEM 10 (Exercise **??**). Let $T : X \to Y$ be a linear operator between normed linear spaces X and Y. Then T is open if and only if $T(B_1(0))$ has a nonempty interior.

PROOF. If T is open, then $T(B_1(0))$ is a nonempty open set, whence $T(B_1(0))$ must have a nonempty interior. Conversely, if $T(B_1(0))$ has a nonempty interior, then Proposition 2.4 implies that $T(B_r(x))$ must be open for each $r > 0$ and every $x \in X$. We now let E be an arbitrary open set in X. For each $x \in E$, we can find a real number $r > 0$ such that $B_r(x) \subseteq E$, whence $T(B_r(x))$ is an open neighborhood of Tx contained in $T(E)$. It follows that $T(E)$ is open. □

PROBLEM 11 (Exercise **??**). A normed linear space X is complete if and only if $\sum \|x_n\| < \infty$ implies the convergence of $\sum x_n$.

PROOF. ($\Rightarrow$) We assume that X is complete. If $\sum_{n=1}^{\infty} \|x_n\| < \infty$, then the tail $\sum_{n=N}^{\infty} \|x_n\|$ can be made as small as desired, whence the partial sums $\left(\sum_{n=1}^{N} x_n\right)_{N=1}^{\infty}$ form a Cauchy sequence:

$$\left\|\sum_{n=1}^{M} x_n - \sum_{n=1}^{N} x_n\right\| = \left\|\sum_{n=N}^{M} x_n\right\| \leq \sum_{n=N}^{M} \|x_n\| \to 0 \text{ as } N, M \to \infty.$$

It follows that the series $\sum x_n$ converges.

($\Leftarrow$) We assume that every absolutely convergent series in X is convergent. Let $(x_n)_{n=1}^{\infty}$ be a Cauchy sequence in X. Set $N_0 = 0$. For each $k \in \mathbb{N}$, find $N_k \in \mathbb{N}$ such that $N_k > N_{k-1}$ and $\|x_n - x_m\| < 2^{-k}$ for all $n, m \geq N_k$. Then

$$\sum_{k=1}^{\infty} \|x_{N_{k+1}} - x_{N_k}\| < \sum_{k=1}^{\infty} 2^{-k} = 1,$$

whence the series $\sum x_{N_k+1} - x_{N_k}$ is convergent by the hypothesis. Since

$$\sum_{k=1}^{K} x_{N_{k+1}} - x_{N_k} = x_{N_{K+1}} - x_{N_1},$$

sending $K \to \infty$, we see that

$$\sum_{k=1}^{\infty} x_{N_{k+1}} - x_{N_k} = \Big(\lim_{K\to\infty} x_{N_{K+1}}\Big) - x_{N_1}.$$

We let $x = \lim x_{N_{K+1}}$ and claim that $x_n \to x$. To see this, we observe that

$$\|x - x_n\| \leq \|x - x_{N_K}\| + \|x_{N_k} - x_n\|,$$

where both terms in the right-hand side can be made as small as desired by first picking a large k and then picking an n larger than N_k. This completes the proof. □

Problem 12 (Exercise **??**). It is not very hard to show that

$$D_N(x) = \frac{\sin \pi(2N+1)}{\sin \pi x}.$$

Show also that there is a fixed constant $c > 0$ such that $\|D_N\|_1 \geq c \log N$ for all $N \in \mathbb{N}$.

Proof. Set $\omega = e^{2\pi i x}$, so that

$$D_N(x) = \sum_{n=-N}^{N} \omega^n.$$

Then

$$\begin{aligned} \sum_{n=0}^{N} &= \omega^n = \frac{\omega^{N+1} - 1}{\omega - 1} \quad \text{and} \\ \sum_{n=-N}^{-1} \omega^n &= \omega^{-N} \sum_{n=0}^{N-1} \omega^n = \omega^{-N} \cdot \frac{\omega^N - 1}{\omega - 1} = \frac{1 - \omega^{-N}}{\omega - 1}, \end{aligned}$$

and so

$$D_N(x) = \frac{\omega^{N+1} - \omega^{-N}}{\omega - 1} = \frac{\omega^{1/2}(\omega^{N+1/2} - \omega^{-N-1/2})}{\omega^{1/2}(\omega^{1/2} - \omega^{-1/2})} = \frac{\sin(\pi(2N+1)x)}{\sin(\pi x)},$$

as was to be shown.

We now show that

$$\|D_N\|_1 \geq c \log N$$

for some constant $c > 0$. Note first that

$$\|D_N\|_1 = \int_{-\frac{1}{2}}^{\frac{1}{2}} |D_N(y)|\, dy = \int_{-\frac{1}{2}}^{\frac{1}{2}} \left| \frac{\sin(\pi(2N+1)y)}{\sin(\pi y)} \right| = 2 \int_0^{\frac{1}{2}} \left| \frac{\sin(\pi(2N+1)y)}{\sin(\pi y)} \right|.$$

The first equality follows from the translation invariance of the Haar measure on $\mathbb{T}$, the scond from the result above, and the third from the oddness of the sine function. Since $|\sin y| \leq |y|$ for all $y \in \mathbb{T}$, we have

$$\begin{aligned}
\|D_N\|_1 &\geq 2\int_0^{\frac{1}{2}} \frac{|\sin(\pi(2N+1)y)|}{|\pi y|}\,dy \\
&\geq 2\int_0^{N+\frac{1}{2}} \frac{|\sin \pi y|}{|\pi y|}\,dy \\
&= 2\sum_{n=1}^{N}\int_{n-1}^{n} \frac{|\sin \pi y|}{\pi y}\,dy \\
&\geq 2\sum_{n=1}^{N}\frac{1}{\pi n}\int_{n-1}^{n} |\sin \pi y|\,dy \\
&= \left(\frac{2}{\pi}\int_0^1 |\sin \pi y|\right)\sum_{n=1}^{N}\frac{1}{n}\,dy \\
&\geq c\log N,
\end{aligned}$$

as was to be shown. □

PROBLEM 13 (Exercise **??**). Show that $\|l_{N,x}\| = \|D_N\|_1$ for all $N \in \mathbb{N}$ and $x \in \mathbb{T}$.

PROOF. We shall produce a sequence $(f_n)_{n=1}^{\infty}$ of functions such that $\|f_n\|_\infty \leq 1$ and $l_{N,x}(f_n) \to \|D_N\|_1$ as $n \to \infty$. The existence of such a sequence shows that

$$\|D_N\|_1 = \lim_{n\to\infty} \|l_{N,x}(f_n)\| \leq \lim_{n\to\infty} \|l_{N,x}\|\|f_n\|_\infty \leq \|l_{N,x}\|.$$

Combined with the reverse inequality that we have already established in class, we will have obtained the desired equality.

Let $f(y) = \operatorname{sgn}(D_N)(y)$, so that f is a measurable function bounded by 1. Let $g(y) = f(-(x-y))$ and observe that

$$\begin{aligned}
&\int_0^1 f(-(x-y))D_N(x-y)\,dy \\
&\qquad = \int_0^1 |D_N(x-y)|\,dy \\
&\qquad = \int_0^1 |D_N(y)|\,dy \\
&\qquad = \|D_N\|_1.
\end{aligned}$$

The domain of $l_{N,x}$ is $\mathscr{C}(\mathbb{T})$, however, and g is not continuous, whence we cannot plug g into $l_{N,x}$ and derive the desired conclusion.

Indeed, we need to approximate g by continuous functions. This can be done in L^1-sense: there exists a sequence $(f_n)_{n=1}^{\infty}$ bounded by 1 in $\mathscr{C}(\mathbb{T})$ that converges to g in $L^1(\mathbb{T})$. Then we apply the dominated convergence theorem to conclude that

$$\begin{aligned}\lim_{n\to\infty} |l_{N,x}(f_n) - \|D_N\|_1| &\leq \lim_{n\to\infty} \int_0^1 |f_n(y) - g(y)||D_N(x-y)|\,dy \\ &\leq \lim_{n\to\infty} \|f_n - g\|_1 \|D_N\|_{L^\infty(\mathbb{T})} \\ &= 0\end{aligned}$$

for each $N \in \mathbb{N}$. □

PROBLEM 14 (Exercise **??**). The end result can be strengthened as follows: if A is a countable subset of $\mathbb{T}$, then there exists an $f \in \mathscr{C}(\mathbb{T})$ such that $(S_N f(x))_{N=1}^\infty$ fails to converge for all $x \in A$.

PROOF. The previous problem implies that $\sup_N \|l_{N,x}\| = \infty$ regardless of $x \in \mathbb{T}$, whence, for each point $x \in \mathbb{T}$, we can find a continuous function whose Fourier series diverges at x. To upgrade this statement to a statement about divergence on countable sets, we establish the following lemma:

LEMMA 1.3. *Let X be a Banach space, Y be a normed vector space, and $(T_\alpha)_\alpha$ be a collection of bounded linear maps from X to Y. Define*

$$M = \{x \in X : \sup_a \|T_\alpha x\|_Y = \infty\}.$$

Show that M is either empty or a dense G_δ set (a countable intersection of open sets) in X.

PROOF OF LEMMA. Let $U_n = \{x \in X : \sup_\alpha \|T_\alpha x\|_Y > n\}$ for each $n \in \mathbb{N}$ and note that $M = \bigcap_n U_n$. If each U_n is dense, then M is a countable intersection of open, dense subsets of a complete space X, whence by Lemma 4.4 in Chapter 1 M must be dense as well.

We now assume that U_{n_0} is not dense for some $n_0 \in \mathbb{N}$. We can then find a point $x_0 \in X \smallsetminus U_{n_0}$ and a real number $r > 0$ such that $B_r(x_0) \subseteq X \smallsetminus U_{n_0}$. Then, for all α, we have $\|T_\alpha x\| \leq n_0$ whenever $\|x - x_0\| < r$. We now see that each $\|y\| < r$ satisfies the estimate

$$\|T_\alpha y\| \leq \|T_\alpha(y + x_0)\| + \|T(-x_0)\| = \|T_\alpha(y + x_0)\| + \|T(x_0)\| \leq 2n_0$$

for all α. Therefore, if $\|x\| \leq 1$, then we have

$$\|T_\alpha x\| = \|r^{-1} T_\alpha(rx)\| \leq \frac{2n_0}{r}$$

for all α, and so $\|T_\alpha\|$ is uniformly bounded. It follows that M is empty. □

For each $a \in \mathbb{T}$, we set

$$M_a = \left\{ f \in \mathscr{C}(\mathbb{T}) : \sup_N |S_N f(a)| = \infty \right\}.$$

Our goal is to show that

$$M = \bigcap_{a \in A} M_a$$

is nonempty. Since each M_a is nonempty, the above lemma shows that M_a can be written as a countable intersection of dense open sets. Therefore, M is also a countable intersection of dense open sets, whence by Lemma 4.4 in Chapter 1 M must be dense as well. In particular, M is nonempty. □

PROBLEM 15. Recall that a topological space X is *Hausdorff* if, for each distinct pair of points x and y in X, there exist disjoint open sets U and V that contain x and y, respectively. Show that the product of a collection of Hausdorff topological spaces with the product topology is a hausdorff topological space.

PROOF. □

PROBLEM 16. Show that closed subsets of compact sets are compact. Conversely, show that compact subsets of a Hausdorff topological space is closed.

PROBLEM 17 (Exercise 8.19). We fix $p, q \in [1, \infty)$ and let $a : \mathbb{R} \to \mathbb{R}$ be a continuous function satisfying the estimate

$$|a(t)| \leq C(|t|^{p/q} + 1)$$

for all $t \in \mathbb{R}$. Show that the operator $A : L^p(0,1) \to L^q(0,1)$ defined by the formula

$$(Af)(x) = a(f(x))$$

is strongly continuous but is never weakly continuous unless a is affine.

PROOF. □

2. Problems and Solutions for Chapter 2

PROBLEM 18 (Exercise 1.16). A projection P of $\mathcal{H}$ is orthogonal if and only if $(\operatorname{im} P)^{\perp} = \ker P$.

PROOF. ($\Rightarrow$) We suppose that P is an orthogonal projection, so that

$$\mathcal{H} = \operatorname{im} P \oplus \operatorname{im}(I - P) = \operatorname{im} P \oplus (\operatorname{im} P)^{\perp}.$$

For each $x \in (\operatorname{im} P)^{\perp}$, □

PROBLEM 19. If $T \in \mathscr{B}(\mathcal{H})$, then $(\operatorname{im} T)^{\perp} = \ker T^{*}$.

PROOF. Observe that

$$\begin{aligned} y \in \ker T^{*} \quad &\Leftrightarrow \quad T^{*}y = 0 \\ &\Leftrightarrow \quad \langle x, T^{*}y \rangle = 0 \text{ for all } x \in \mathcal{H} \\ &\Leftrightarrow \quad \langle Tx, y \rangle = 0 \text{ for all } x \in \mathcal{H} \\ &\Leftrightarrow \quad \langle z, y \rangle = 0 \text{ for all } z \in \operatorname{im} T \\ &\Leftrightarrow \quad y \in (\operatorname{im} T)^{\perp}, \end{aligned}$$

as was to be shown. □

3. Homework Problems from Class

PROBLEM 20 (Homework set 1, Problem 4). Let X and Y be two normed vector spaces, and (T_n) be a convergent sequence of bounded linear operators from X to Y. Show that if (x_n) is a convergent sequence in X, then $T_n(x_n)$ is convergent in Y.

PROOF. Let $T = \lim T_n$, $x = \lim x_n$, and observe that

$$\left\| \lim_{n\to\infty} T_n(x_n) - Tx \right\| = \lim_{n\to\infty} \|T_n(x_n) - Tx\| \leq \lim_{n\to\infty} \|T_n - T\| \|x_n - x\| = 0.$$

□

PROBLEM 21 (Homework set 1, Problem 6). Let X be a normed vector space. Show that every ball in X is a convex set. Is this true if X is only a metric linear space? If not, give an example.

PROOF. By scaling and translation, we need only to establish convexity for the unit ball. If x and y are in the closed unit ball, then $\|x\| \leq 1$ and $\|y\| \leq 1$, and so

$$\|\lambda x + (1-\lambda)y\| \leq \lambda\|x\| + (1-\lambda)\|y\| \leq \lambda + (1-\lambda) = 1$$

for each $\lambda \in [0,1]$. If x and y are in the open unit ball, then $\|x\| < 1$ and $\|y\| < 1$, and so

$$\|\lambda x + (1-\lambda)y\| \leq \lambda\|x\| + (1-\lambda)\|y\| < \lambda + (1-\lambda) = 1$$

for each $\lambda \in [0,1]$. Therefore, both the open unit ball and the closed unit ball are convex.

Not all metric linear spaces have convex unit balls. To see this, we define a metric d by setting

$$d(x,y) = \sqrt{|x_1 - y_1|} + \sqrt{|x_2 - y_2|}$$

for each pair of points x and y in $\mathbb{R}^2$. We claim that the open unit ball B with respect to d is not convex. To see this, we examine the line segment connecting $(1/2, 0)$ and $(0, 1/2)$. Since

$$d((0,0),(1/2,0)) = ((0,0),(0,1/2)) = \frac{1}{\sqrt{2}} < 1,$$

both points are in B. We observe, however, that

$$d((0,0),(1/4,1/4)) = \frac{2}{\sqrt{2}} > 1,$$

whence the midpoint $(1/4, 1/4)$ of the two points is not in B. It follows that B is not convex. □

PROBLEM 22 (Homework set 1, Problem 7). Show that on any infinite dimensional vector space, one can find nonequivalent norms.

PROOF. Invoking the Axiom of Choice, we take a Hamel basis $B = \{v_\alpha\}$ of an infinite-dimensional vector space X. Each function $f : B \to [0,\infty)$ defines a norm $\|\cdot\|_f$ given by

$$\left\| \sum_\alpha \lambda_\alpha v_\alpha \right\|_f = \sum_\alpha |\lambda_\alpha| f(v_\alpha).$$

Clearly, $\|v\|_\alpha = 0$ if and only if $v = 0$, and

$$\left\|\rho \sum_\alpha \lambda_\alpha v_\alpha\right\|_f = \left\|\rho \sum_\alpha (\rho\lambda_\alpha) v_\alpha\right\|_f = \sum_\alpha |\rho\lambda_\alpha| f(v_\alpha) = |\rho| \sum_\alpha |\lambda_\alpha| f(v_\alpha).$$

Furthermore,

$$\left\|\sum_\alpha \lambda_\alpha v_\alpha + \sum_\beta \rho_\beta v_\beta\right\| = \sum_{\alpha,\beta} |\lambda_\alpha + \rho_\beta| v_{\alpha,\beta} \leq \sum_\alpha |\lambda_\alpha| v_\alpha + \sum_\beta |\rho_\beta| v_\beta,$$

and so $\|\cdot\|_f$ is a norm.

We now note that if f and g are two nonnegative functions on B such that

$$K_1 f(v_\alpha) \leq g(v_\alpha) \leq K_2 f(v_\alpha)$$

for all $v_\alpha \in B$, then $\|\cdot\|_f$ and $\|\cdot\|_g$ are equivalent. It therefore suffices to pick f and g that violate this condition. For example, if f is chosen to be a constant function and g to take arbitrarily large and arbitrarily small values, then this condition cannot hold. □

Problem 23 (Homework set 1, Problem 8). Recall the definition that a Banach algebra is a Banach space that is also algebra such that $\|xy\| \leq \|x\|\|y\|$ for all x, y. Show that the following are Banach algebras.

(a) The space of bounded linear operators on a Banach space with respect to composition (product) of operators.
(b) $\mathscr{C}(K)$ with respect to pointwise multiplication. Here K is a compact space.
(c) $L^1(\mathbb{R})$ or $l^1(\mathbb{Z})$ with respect to convolution product.

Proof. We already know that (a) - (c) are Banach spaces, so it suffices to check the norm inequality $\|xy\| \leq \|x\|\|y\|$.

(a) Pick $T, U \in \mathscr{B}(X)$ and observe that

$$\|TUx\| \leq \|T\|\|Ux\| \leq \|T\|\|U\|\|x\|$$

for all $x \in X$, whence $\|TU\| \leq \|T\|\|U\|$.

(b) Pick $f, g \in \mathscr{C}(K)$ and observe that

$$\|fg\| = \sup_{x\in K} |f(x)g(x)| \leq \sup_{x\in K} |f(x)| \sup_{x\in K} |g(x)| = \|f\|\|g\|.$$

(c) We prove Young's inequality for general measure spaces. Given f, g in $L^1(X, \mu)$, we observe that

$$\begin{aligned}
\|f * g\|_1 &= \int \left| \int f(x-y)g(y)\,dy \right| dx \\
&\leq \iint |f(x-y)||g(y)|\,dy\,dx \\
&= \int |g(y)| \int |f(x-y)|\,dx\,dy \\
&= \int |g(y)|\|f\|_1 \\
&= \|f\|_1\|g\|_1,
\end{aligned}$$

which is the desired result. □

PROBLEM 24 (Homework set 1, Problem 9). Recall that a metric space is separable if it has a countable dense subset. For which p is $l^p(\mathbb{N})$ separable?

PROOF. $l^\infty(\mathbb{N})$ is not separable, for the set B of sequences whose terms are either 0 or 1 is an uncountable, discrete set in $l^\infty(\mathbb{N})$. Indeed, if (a_n) and (b_n) are distinct elements of B, then $\|(a_n) - (b_n)\|_\infty = 1$.

Why does this show that $l^\infty(\mathbb{N})$ is not separable? If D is any dense subset of $l^\infty(\mathbb{N})$, then, for each $(a_n) \in B$, the open ball of radius $1/2$ intersects nontrivially with D. Since there are uncountably many such balls that are pairwise disjoint, we are forced to conclude that D must be uncountable.

We now assume that $1 \leq p < \infty$. For each sequence $(a_n) \in l^p(\mathbb{Z})$, we define the *kth cutoff operator* $\mathcal{C}_k : l^p(\mathbb{Z}) \to c_c$ by setting $\mathscr{C}_k[(a_n)]$ to be the sequence

$$a_1, a_2, \ldots, a_k, 0, 0, \ldots.$$

Note that

$$\|(a_n) - \mathcal{C}_k[(a_n)]\|_p^p = \sum_{n=k+1}^{\infty} |a_n|^p$$

converges to zero as $k \to \infty$, whence the space c_c of finitely supported sequences is a dense subspace of $l^p(\mathbb{N})$.

We claim that the space R of finitely supported sequences of rational numbers is dense in c_c. To see this, we let $(a_n) \in c_c$ and find $N \in \mathbb{N}$ such that $a_n = 0$ for all $n > N$. By the density of $\mathbb{Q}$ in $\mathbb{R}$, we can construct a sequence $((a_n^k)_{n=1}^\infty)_{k=1}^\infty$ of sequences $(a_n^k)_{n=1}^\infty$ in R by setting a_n^k to be a rational number such that $|a_n - a_n^k| \leq 1/k$ if $n \leq N$, and $a_n^k = 0$ if $n > N$. Then

$$\|(a_n) - (a_n^k)\|_p = \sum_{n=1}^{N} |a_n - a_n^k| \leq \frac{N}{k},$$

which converges to 0 as $k \to \infty$, as was to be shown.

We now show that the countable set R is dense in $l^p(\mathbb{N})$. For each $(a_n) \in l^p(\mathbb{N})$ and every $\varepsilon > 0$, we can find sequences $(b_n) \in c_c$ and $(c_n) \in R$ such that $\|(a_n) - (b_n)\|_p < \varepsilon/2$ and $\|(b_n) - (c_n)\|_p < \varepsilon/2$. It follows that $\|(a_n) - (c_n)\|_p < \varepsilon$, as desired. □

PROBLEM 25 (Homework set 1, Problem 10). Which of the following subspaces of the Banach space $l^\infty(\mathbb{N})$ are closed?

(a) c, the set of convergent sequences.
(b) c_0, the set of sequences convergent to zero.
(c) $l^1(\mathbb{N})$.
(d) c_c, the set of finitely supported sequences.

PROOF. (a) c is a closed subspace of $l^\infty(\mathbb{N})$. To see this, we let $(a_n)_{n=1}^\infty$ be a limit point of c. For each $k \in \mathbb{N}$, we find $(a_n^k)_{n=1}^\infty \in c$ such that $\|(a_n^k) - (a_n)\|_\infty < \frac{1}{k}$. Let

$$L^k = \lim_{n \to \infty} a_n^k$$

for each $k \in \mathbb{N}$, and observe that

$$|L^{k_1} - L^{k_2}| \leq |L^{k_1} - a_{n_1}^{k_1}| + |L^{k_2} - a_{n_2}^{k_2}|.$$

Both terms in the right-hand side of the inequality can be made as small as desired, and so $(L^k)_{k=1}^\infty$ is a Cauchy sequence in the Banach space $l^\infty(\mathbb{N})$, converging to, say, L. Observe that

$$\begin{aligned} |a_n - L| &\leq |a_n - a_n^k| + |a_n^k - L^k| + |L^k - L| \\ &\leq \|(a_n) - (a_n^k)\|_\infty + |a_n^k - L^k| + |L^k - L|, \end{aligned}$$

and note that all three terms in the final upper bound can be made as small as desired. It follows that $a_n \to L$, whence $(a_n) \in c$.

(b) c_0 is a closed subspace of $l^\infty(\mathbb{N})$. To see this, we let $(a_n)_{n=1}^\infty$ be a limit point of c_0. For each $k \in \mathbb{N}$, we find $(a_n^k)_{n=1}^\infty \in c_0$ such that $\|(a_n^k) - (a_n)\| < \frac{1}{k}$. Observe that

$$|a_n| \leq |a_n - a_n^k| + |a_n^k| \leq \|a_n - a_n^k\|_\infty + |a_n^k|,$$

and note that both terms of the final upper bound can be made as small as desired. It follows that $a_n \to 0$, whence $(a_n) \in c_0$.

(c) $l^1(\mathbb{N})$ is *not* a closed subspace of $l^\infty(\mathbb{N})$. To see this, we let (a_n) be an element of $c_0 \smallsetminus l^1(\mathbb{N})$, which is evidently nonempty. For each $\varepsilon > 0$, we find $N_\varepsilon \in \mathbb{N}$ such that $n > N_\varepsilon$ implies $|a_n| < \varepsilon$, so that

$$\|(a_n) - \mathscr{C}_{N_\varepsilon}[(a_n)]\|_\infty < \varepsilon.$$

Here $\mathscr{C}_{N_k}$ is the N_kth cutoff operator defined in Problem 24, which maps c_0 to c_c in this case. Since $c_c \subseteq l^1(\mathbb{N})$, we see that (a_n) is a limit point of $l^1(\mathbb{N})$ that is not an element of $l^1(\mathbb{N})$.

(d) c_c is *not* a closed subspace of $l^\infty(\mathbb{N})$. To see this, we let (a_n) be an element of $c_0 \smallsetminus c_c$, which is evidently nonempty. For each $\varepsilon > 0$, we find $N_\varepsilon \in \mathbb{N}$ such that $n > N_\varepsilon$ implies $|a_n| < \varepsilon$, so that

$$\|(a_n) - \mathscr{C}_{N_\varepsilon}[(a_n)]\|_\infty < \varepsilon.$$

It follows that (a_n) is a limit point of $c_c(\mathbb{N})$ that is not an element of $c_c(\mathbb{N})$. □

Problem 26 (Homework set 2, Problem 4). Show that there is a nonzero bounded linear functional on $L^\infty(\mathbb{R})$ which vanishes on the space $\mathcal{BC}(\mathbb{R})$ of bounded and continuous functions on $\mathbb{R}$.

Proof. Since the uniform limit of continuous functions is cotninuous, $\mathcal{BC}(\mathbb{R})$ is a closed subspace of $L^\infty(\mathbb{R})$. The existence of nontrivial annihilators on closed subspaces is an immediate corollary (Chapter 1, Corollary 5.10) of the Hahn-Banach theorem, and so there exist nonzero bounded linear functionals on $L^\infty(\mathbb{R})$ that vanish on $\mathcal{BC}(\mathbb{R})$. □

Problem 27 (Homework set 2, Problem 5). Let X be a normed vector space and A be a subset of X. Show that $\sup_{x \in A} |l(x)| < \infty$ for every $l \in X^*$ if and only if A is bounded.

Proof. ($\Leftarrow$) Let $A \subseteq X$ be bounded, so that $\|x\| \leq M$ for all $x \in A$. Then

$$\sup_{x \in A} |l(x)| \leq \sup_{x \in A} \|l\|\|x\| \leq M\|l\| < \infty$$

for each $l \in X^*$.

($\Rightarrow$) Suppose that $\sup_{x \in A} |l(x)| < \infty$. For each $x \in X$, we define $\hat{x} : X^* \to \mathbb{F}$ by setting $\hat{x}(l) = l(x)$, and observe that

$$\sup_{x \in A} |\hat{x}(l)| = \sup_{x \in A} |l(x)| < \infty$$

for each $l \in X^*$. Since X^* is complete, the uniform boundednes principle (Chapter 1, Theorem 4.21) applies to $\{\hat{x}\}_{x\in A}$, and we conclude that

$$\sup_{x\in A} \|\hat{x}\|_{X^{**}} < \infty.$$

But then $\|\hat{x}\|_{X^{**}} = \|x\|_X$, and so we have

$$\sup_{x\in A} \|x\|_X < \infty,$$

as was to be shown. □

Problem 28 (Homework set 2, Problem 6). Let X and Y be normed vector spaces and $T \in \mathscr{B}(X, Y)$. Show that T^* is 1-1 if and only if $T(X)$ is dense in Y.

Proof. Before we proceed, we observe that the injectivity of T^* is equivalent to the statement that $l_1 \circ T = l_2 \circ T$ implies $l_1 = l_2$ for all $l_1, l_2 \in Y^*$. This is a direct consequence of the definition of T^*.

($\Leftarrow$) We suppose that $T(X)$ is dense in Y. If $l_1, l_2 \in Y^*$ satisfy $l_1 \circ T = l_2 \circ T$, then $l_1|_{T(X)} = l_2|_{T(X)}$, and so l_1 and l_2 are continuous functions that agree on a dense set. It follows that $l_1 = l_2$, and so T^* is injective.

($\Rightarrow$) Conversely, suppose that T^* is injective and assume for a contradiction that $T(X)$ is not dense. Then $\overline{T(X)} \neq 0$, and so we can find two points $y_1, y_2 \in Y \smallsetminus \overline{T(X)}$ such that $d(y_1, \overline{T(X)}) \neq d(y_2, \overline{T(X)})$. By the Hahn-Banach theorem, we can then construct $l_1, l_2 \in Y^*$ such that $l_1|_{\overline{T(X)}} = l_2|_{\overline{T(X)}} = 0$, $l_1(y_1) = 1/d(y_1, \overline{T(X)})$, and $l_2(y_2) = 1/d(y_2, \overline{T(X)})$. Note, in particular, that $l_1|_{T(X)} = l_2|_{T(X)}$ but $l_1 \neq l_2$, which contradicts the assumption that T^* is injective. It follows that $T(X)$ must be dense. □

Problem 29 (Homework set 3, Problem 1). Consider $X = L^2([-1,1])$ with its native norm. For each scalar α, let

$$E_\alpha = \{f \in \mathscr{C}([-1,1]) : f(0) = \alpha\}.$$

(a) Show that each E_α is convex and dense in X.
(b) Show that for $\alpha \neq \beta$, E_α and E_β are disjoint but there is no nonzero continuous linear functional l on X for which

$$\sup_{f\in E_\alpha} l(f) \leq \inf_{f \in E_\beta}.$$

Explain why the geometric Hahn-Banach theorem could not be employed. (Does E_α have any internal point?)

Proof. (a) If $f, g \in E_\alpha$, then $(1-\lambda)f + \lambda g$ is continuous and $(1-\lambda)f(0) + \lambda g(0) = (1-\lambda)\alpha + \lambda\alpha = \alpha$, and so $(1-\lambda)f + \lambda g \in E_\alpha$. It follows that E_α is convex.

As for the density of E_α, it evidently suffices to show that E_α is dense in $\mathscr{C}([-1,1])$. To this end, we fix $f \in \mathscr{C}([-1,1])$. We assume without loss of generality that $\|f\|_2 > 0$. Let $\varepsilon > 0$ be small enough that $\|f\|_2 > 2\varepsilon$. The mapping

$$t \mapsto \left(\int_{-t}^{t} |f(x)|^2 \, dx \right)^{1/2}$$

is continuous, and so the intermediate value theorem furnishes a point $t_1 \in (0,1)$ such that

$$\left(\int_{-t}^{t} |f(x)|^2 \, dx \right)^{1/2} < \varepsilon.$$

for all $t \in (-t_1, t_1)$.

Now, for each $t \in (-t_1, t_1)$, we define

$$f_t(x) = \begin{cases} \frac{f(-t)-\alpha}{-t}x + \alpha & \text{if } -t < x \leq 0; \\ \frac{f(t)-\alpha}{t}x + \alpha & \text{if } 0 \leq x < t; \\ f(x) & \text{if } t \leq |x| \leq 1; \end{cases}$$

so that $f_t \in E_\alpha$. Elementary computations show that

$$F(t) = \left(\int_{-t}^{t} |f_t(x)|^2 \right)^{1/2} dx$$

is a continuous function that converges to 0 as $t \to 0$, we can find a point $t_2 \in (-t_1, t_1)$ such that $|F(t)| < \varepsilon$. It now follows that

$$\|f - f_{t_2}\|_2 = \|(f - f_{t_2})\chi_{[-t_2,t_2]}\|_2 \leq \|f\chi_{[-t_2,t_2]}\|_2 + \|f_{t_2}\chi_{[-t_2,t_2]}\|_2 < 2\varepsilon,$$

and we conclude that E_α is dense in $\mathscr{C}([-1,1])$.

(b) We recall the following standard result:

Lemma 3.1. *Let $f, g \in \mathscr{C}([-1,1])$. If $f(x_0) \neq g(x_0)$ for some $x_0 \in [-1,1]$, then there exists an $\varepsilon > 0$ such that $f(x) \neq g(x)$ on $(x_0 - \varepsilon, x_0 + \varepsilon) \cap [-1,1]$. In particular, f and g differ on a set of positive measure.*

If $\alpha \neq \beta$, then $f \in E_\alpha$ and $g \in E_\beta$ differ at a point, whence the above result implies that $f \neq g$ in L^2. It follows that $E_\alpha \cap E_\beta = \varnothing$.

To show the second part, we prove the following lemma:

Lemma 3.2. *If l is a nonzero continuous functional on a normed linear space Y, then $\sup l(x) = \infty$ and $\inf l(x) = \infty$ on D, whenever D is a dense subset of Y.*

Proof of lemma. Fix $x_0 \in Y$ such that $l(x_0) \neq 0$, and set $x_1 = x_0/l(x_0)$, so that $l(\lambda x_1) = \lambda$ for all $\lambda \in \mathbb{R}$. Fix $\varepsilon > 0$. For each $\lambda \in \mathbb{R}$, we can find $y_\lambda \in D$ such that $\|\lambda x_1 - y_\lambda\| < \varepsilon$. It then follows that

$$|l(\lambda x_1) - l(y_\lambda)| \leq \|l\|\|x_1 - y_\lambda\| < \varepsilon\|l\|,$$

whence $l(y_\lambda) > l(\lambda x_1) - \varepsilon$. It follows that

$$\sup_{x \in D} l(x) = \infty.$$

The other half of the lemma is established analogously. □

It now follows from the above lemma that

$$\sup_{f \in E_\alpha} l(f) = \infty \quad \text{and} \quad \inf_{f \in E_\beta} l(f) = -\infty,$$

whence

$$\sup_{f \in E_\alpha} l(f) \leq \inf_{f \in E_\beta} l(f)$$

can never be true.

Lastly, we show that the geometric Hahn-Banach theorem cannot be applied in this setting. Let $f \in E_\alpha$. If $g \in X$ is nonzero at 0, then $f(0) + \lambda g(0) \neq \alpha$ for all $\lambda \neq 0$, whence $f + \lambda g \notin E_\alpha$. It follows that f cannot be an internal point of E_α, and thus E_α has none. □

Problem 30 (Homework set 3, Problem 2). For each of the following normed vector spaces find the extreme points of its unit ball B.

(a) $X = l^1(\mathbb{N})$
(b) $X = l^p(\mathbb{N})$, $1 < p < \infty$
(c) $X = l^\infty(\mathbb{N})$
(d) $X = L^1([0,1])$
(e) $X = c_0$ (supremum norm)
(f) $X = C([0,1])$ (supremum norm)

Proof. (a) Let

$$e_n^N = \begin{cases} 1 & \text{if } N = n; \\ 0 & \text{if } N \neq n. \end{cases}$$

For each $N \in \mathbb{N}$, $(e_n^N)_{n=1}^\infty$ is an extreme point of B, for if $(1-\lambda)(a_n)+\lambda(b_n) = (e_n^N)$, then either $a_N > 1$ or $b_N > 1$, putting the sequence outside of B. Similarly, $(-e_n^N)_{n=1}^\infty$ is an extreme point of B for all $N \in \mathbb{N}$. If (c_n) is any other sequence in B, then $|c_n| < 1$ for all $n \in \mathbb{N}$. We find $\varepsilon > 0$ such that $2\varepsilon + \sum |c_n| < 1$ and set

$$a_n = \begin{cases} c_1 + \varepsilon & \text{if } n = 1; \\ c_n & \text{if } n \geq 2. \end{cases} \quad \text{and} \quad b_n = \begin{cases} c_1 - \varepsilon & \text{if } n = 1; \\ c_n & \text{if } n \geq 2. \end{cases}$$

Then $(a_n), (b_n) \in B$ and $\frac{1}{2}(a_n) + \frac{1}{2}(b_n) = (c_n)$, whence (c_n) is not an extreme point of B.

(b) We recall the following standard result:

Lemma 3.3 (Minkowski's inequality). *Let (X, μ) be a measure space. If $1 < p < \infty$ and $f, g \in L^p(X, \mu)$, then $\|f + g\|_p \leq \|f\|_p + \|g\|_p$, and the inequality is strict unless $f = 0$, $g = 0$, or $f = \lambda g$ for some $\lambda > 0$.*

Let $(c_n) \in B$ be of p-norm 1. If $\|(a_n)\|_p < 1$ or $\|(b_n)\|_p < 1$, then

$$\|(1-\lambda)(a_n) + \lambda(b_n)\|_p \leq (1-\lambda)\|(a_n)\|_p + \lambda\|(b_n)\|_p < 1,$$

and so (c_n) cannot be a convex combination of (a_n) and (b_n). If $\|(a_n)\|_1 = 1$ and $\|(b_n)\|_1 = 1$, then

$$\|(1-\lambda)(a_n) + \lambda(b_n)\|_p \leq 1,$$

and the equality holds if and only if (a_n) is a positive scalar multiple of (b_n). But $\|(a_n)\|_p = \|(b_n)\|_p = 1$, and so we are forced to conclude that $(a_n) = (b_n)$. Therefore, (c_n) cannot be a convex combination of two distinct elements of B, and so (c_n) is an extreme point.

If $\|(c_n)\|_p < 1$, then we can find an $\varepsilon > 0$ such that

$$\left(|c_1 + \varepsilon|^p + \sum_{n=2}^\infty |c_n|^p\right)^{1/p} < 1 \quad \text{and}$$
$$\left(|c_1 - \varepsilon|^p + \sum_{n=2}^\infty |c_n|^p\right)^{1/p} < 1.$$

We set

$$a_n = \begin{cases} c_1 + \varepsilon & \text{if } n = 1; \\ c_n & \text{if } n \geq 2. \end{cases} \quad \text{and} \quad b_n = \begin{cases} c_1 - \varepsilon & \text{if } n = 1; \\ c_n & \text{if } n \geq 2. \end{cases}$$

Then $(a_n), (b_n) \in B$ and $\frac{1}{2}(a_n) + \frac{1}{2}(b_n) = (c_n)$, whence (c_n) is not an extreme point of B.

(c) If $|c_n| = 1$ for all $n \in \mathbb{N}$, then (c_n) is an extreme point of B. Indeed, if $(1-\lambda)(a_n) + \lambda(b_n) = (c_n)$, then either $|a_n| > 1$ or $|b_n| > 1$ for some $n \in \mathbb{N}$, putting the sequence outside of B. If $(c_n)_{n=1}^{\infty}$ is any other sequence in B, then there exists at least one $N \in \mathbb{N}$ such that $|c_N| < 1$. We find an $\varepsilon > 0$ such that $|c_N| + \varepsilon < 1$, and set

$$a_n = \begin{cases} c_1 + \varepsilon & \text{if } n = 1; \\ c_n & \text{if } n \geq 2. \end{cases} \quad \text{and} \quad b_n = \begin{cases} c_1 - \varepsilon & \text{if } n = 1; \\ c_n & \text{if } n \geq 2. \end{cases}$$

Then $(a_n), (b_n) \in B$ and $\frac{1}{2}(a_n) + \frac{1}{2}(b_n) = (c_n)$, whence (c_n) is not an extreme point of B.

(d) Let $f \in B$. If $\|f\|_1 = 0$, then $f = 0$ almost everywhere. Setting $g = 1/2$ and $h = -1/2$, we see that $g, h \in B$ and $(1/2)g + (1/2)h = f$ almost everywhere, whence f is not an extreme point of B.

We now suppose that $\|f\|_1 > 0$. Observe that the maps $t \mapsto \|f\chi_{[0,t]}\|_1$ and $t \mapsto \|f\chi_{[t,1]}\|_1$ are continuous on the compact set $[0,1]$, and their values range from 0 to $\|f\|_1$. Therefore, we can find points $t_m, t_M \in [0,1]$ such that

$$\begin{cases} \|f\chi_{[0,t]}\|_1 = 0 & \text{if } t \leq t_m; \\ \|f\chi_{[0,t]}\|_1 > 0 & \text{if } t > t_m; \end{cases} \quad \text{and} \quad \begin{cases} \|f\chi_{[t,1]}\|_1 = 0 & \text{if } t \geq t_M; \\ \|f\chi_{[t,1]}\|_1 > 0 & \text{if } t < t_M. \end{cases}$$

Necessarily, we have $t_m < t_M$, for otherwise $\|f\|_1 = 0$.

Let $l = t_M - t_m$. We set

$$\varepsilon = \min\left\{ \|f\chi_{[t_m, t_m + l/4]}\|_1,\ \|f\chi_{[t_M - l/4, t_M]}\|_1,\ \frac{1}{4}\|f\|_1 \right\}.$$

Using the continuity of the map $s \mapsto \|f\chi_{[t_m,s]}\|_1$, we can find $s_m \in [t_m, t_m + l/4]$ such that $\|f\chi_{[t_m,s_m]}\|_1 = \varepsilon$. This, in particular, implies that $\|\frac{1}{2}f\chi_{[t_m,s_m]}\|_1 = \varepsilon/2$ and $\|\frac{3}{2}f\chi_{[t_m,s_m]}\|_1 = \frac{3}{2}\varepsilon$. We now use the continuity of the map $s \mapsto \|\frac{1}{2}f\chi_{[s,t_M]}\|_1$ to find $s_M \in [t_M - l/4, t_M]$ such that $\|\frac{1}{2}f\chi_{[s_M,t_M]}\|_1 = \varepsilon/2$. Note once again that we have $\|\frac{3}{2}f\chi_{[s_M,t_M]}\|_1 = \frac{3}{2}\varepsilon$.

Let us set

$$\begin{aligned} g &= \frac{1}{2}f\chi_{[t_m,s_m]} + f\chi_{[s_m,s_M]} + \frac{3}{2}f\chi_{[s_M,t_M]} + f\chi_{[0,1]\smallsetminus[t_m,t_M]}; \\ h &= \frac{3}{2}f\chi_{[t_m,s_m]} + f\chi_{[s_m,s_M]} + \frac{1}{2}f\chi_{[s_M,t_M]} + f\chi_{[0,1]\smallsetminus[t_m,t_M]}. \end{aligned}$$

Then $(1/2)g + (1/2)h = f$ and

$$\begin{aligned} \|g\|_1 &= \left\|\frac{1}{2}f\chi_{[t_m,s_m]}\right\|_1 + \|f\chi_{[s_m,s_M]}\|_1 + \left\|\frac{3}{2}f\chi_{[s_M,t_M]}\right\|_1 + \|f\chi_{[0,1]\smallsetminus[t_m,t_M]}\|_1 \\ &= \frac{1}{2}\varepsilon + \|f\chi_{[s_m,s_M]}\|_1 + \frac{3}{2}\varepsilon + \|f\chi_{[0,1]\smallsetminus[t_m,t_M]}\|_1 \\ &= \varepsilon + \|f\chi_{[s_m,s_M]}\|_1 + \varepsilon + \|f\chi_{[0,1]\smallsetminus[t_m,t_M]}\|_1 \\ &= \|f\chi_{[t_m,s_m]}\|_1 + \|f\chi_{[s_m,s_M]}\|_1 + \|f\chi_{[s_M,t_M]}\|_1 + \|f\chi_{[0,1]\smallsetminus[t_m,t_M]}\|_1 \\ &= \|f\|_1. \end{aligned}$$

Similarly, $\|h\|_1 = \|f\|_1$. It follows that $g, h \in B$, whence f is not an extreme point of B.

(e) Let $(a_n) \in B$, and find $N \in \mathbb{N}$ such that $a_n = 0$ for all $n \geq N$. We let

$$b_n = \begin{cases} a_n & \text{if } n < N; \\ 1 & \text{if } n = N; \\ 0 & \text{if } n > N; \end{cases} \quad \text{and} \quad c_n = \begin{cases} a_n & \text{if } n < N; \\ -1 & \text{if } n = N; \\ 0 & \text{if } n > N, \end{cases}$$

so that $(b_n), (c_n) \in B$ and $\frac{1}{2}(b_n) + \frac{1}{2}(c_n) = (a_n)$. It follows that B has no extreme points.

(f) Let $f \in B$. If $|f| = 1$ everywhere, then f is an extreme point of B. Indeed, if $(1-\lambda)g + \lambda h = f$, then either $|g(x)| > 1$ or $|h(x)| > 1$ for some $x \in [0,1]$, putting the function outside of B.

We now suppose that $|f(x_0)| < 1$ for some $x_0 \in [0,1]$. The continuity of f furnishes an $\varepsilon > 0$ and a $\delta > 0$ such that $|f(x)| < 1 - \delta$ for all $x \in (x_0 - \varepsilon, x_0 + \varepsilon) \cap [0,1]$. We now set

$$g(x) = \begin{cases} f(x) + \delta & \text{if } x \in (x_0 - \varepsilon, x_0) \cap [0,1]; \\ f(x) & \text{otherwise}; \end{cases} \quad \text{and}$$

$$h(x) = \begin{cases} f(x) - \delta & \text{if } x \in (x_0 - \varepsilon, x_0) \cap [0,1]; \\ f(x) & \text{otherwise}; \end{cases}$$

so that $g, h \in B$ and $(1/2)g + (1/2)h = f$. It follows that f is not an extreme point of B. $\square$

PROBLEM 31 (Homework set 3, Problem 3). (a) Show that an internal point of a set cannot be an extreme point.

(b) Show that if K is a convex, compact set in $\mathbb{R}^2$, then $\operatorname{ext}(K)$ is compact.

(c) Let $K \subseteq \mathbb{R}^3$ be the convex hull of $\{(1,0,1), (1,0,-1), (\cos\theta, \sin\theta, 0) : 0 \leq \theta \leq 2\pi\}$. Show that K is compact, but $\operatorname{ext}(K)$ is not.

PROOF. (a) If x is an internal point of A, then we can find a scalar $\varepsilon > 0$ and a vector u such that $x + \varepsilon u$ and $x - \varepsilon u$ are in A. x is evidently a convex combination of these two vectors.

(b) Let $(a_n)_{n=1}^{\infty}$ be a sequence in $\operatorname{ext}(K)$ with a limit $a \in K$. We suppose for a contradiction that $a \notin \operatorname{ext}(K)$. We find $b', c' \in K$ such that $(1/2)b' + (1/2)c' = a$ and let $b = (1/2)b' + (1/2)a$ and $c = (1/2)c' + (1/2)a$, so that $b, c \notin \operatorname{ext}(K)$ and that $a = (1/2)b + (1/2)c$.

For each $n \in \mathbb{N}$, we let H_n be the convex hull of $\{a_n, b, c\}$. Since K is convex, each H_n is a subset of K. Furthermore, every element of H_n can be written as a convex combination of distinct elements of K.

We assume without loss of generality that $a_1 \neq a_2$. If either a_1 or a_2 is contained in the line segment that connects b, a, and c, then it is not an extreme point, a contradiction. We also rule out the case in which a_2 is an element of H_1, for then a_2 is a convex combination of elements of H_1. We then see that $H_1 \cup H_2$, the convex combination of $\{a_1, a_2, b, c\}$, is a quadrilateral whose sides are positive distances away from a. Therefore, $H_1 \cup H_2$ contains a disk centered at a, which must contain all but finitely many terms of (a_n)—but then none of them is an extreme point of K. We conclude that a must be in $\operatorname{ext}(K)$.

(c) We shall make use of the following lemma:

LEMMA 3.4. *If X be a topological vector space, and if $C_1, \ldots, C_N$ are nonempty, convex, compact subsets of X, then* $\operatorname{co}\left(\bigcup C_n\right)$ *is compact.*

Proof of lemma. We know that

$$\operatorname{co}\left(\bigcup_{n=1}^{N} C_n\right) = \left\{\sum_{n=1}^{N} \lambda_n x_n : x_n \in C_n,\ \lambda_n \geq 0,\ \text{and}\ \sum_{n=1}^{N} \lambda_n = 1\right\}.$$

If we set

$$A = \left\{(\lambda_1, \ldots, \lambda_N) \in \mathbb{R}^N : \lambda_n \geq 0 \text{ and } \sum_{n=1}^{N} \lambda_n = 1\right\},$$

then $A \times C_1 \times \cdots \times C_N$ is a compact set in $\mathbb{R}^N \times X \times \cdots \times X$ by Tychonoff's theorem. We now define a continuous map $f : \mathbb{R}^N \times X \times \cdots \times X \to X$ by setting

$$f(\lambda, x_1, \ldots, x_N) = \sum_{n=1}^{N} \lambda_n x_n.$$

Then $\operatorname{co}(\bigcup C_n)$ is the image of the compact set $A \times C_1 \times \cdots \times C_N$ under the continuous map f, whence it is compact. □

K is the convex hull of the union of $\{(1,0,1)\}$, $\{(1,0,-1)\}$, and

$$\operatorname{co}(\{(\cos\theta, \sin\theta, 0) : 0 \leq \theta \leq 2\pi\}),$$

and so the above lemma implies that K is compact. We now observe that every point on $\{(\cos\theta, \sin\theta, 0) : 0 \leq \theta \leq 2\pi\}$ *except* $(1,0,0)$ is an extreme point of K, as $(1,0,0)$ is a convex combination of $(1,0,1)$ and $(1,0,-1)$. It follows that $\operatorname{ext}(K)$ is not closed, hence noncompact. □

Problem 32 (Homework set 3, Problem 4). Let $0 < p < 1$, and consider $X = L^p([0,1])$ with the topology induced by the metric

$$d_p(f,g) = \int_0^1 |f-g|^p \, dx.$$

(a) Show that d_p is indeed a metric.
(b) Show that there exists $c = c_p < 1$ such that any nonzero $f \in X$ can be written as $(f_0 + f_1)/2$, where

$$d_p(f_0, 0) = d_p(f_1, 0) = c d_p(f, 0).$$

Hence, no ball in X is convex.
(c) Let $A = \{f, f_0, f_1, f_{00}, f_{01}, f_{10}, f_{11}, \ldots\}$. Argue that A has no extreme point.
(d) The only limit point of A is 0. Using A, construct a compact set K which has no extreme point.

Proof. (a) We shall make use of the following lemma:

Lemma 3.5. *If $a \geq 1$ and $0 < p \leq s$, then*

$$(a+1)^p - a^p \leq (a+1)^s - a^s.$$

Proof of lemma. If $t > 0$, then $x \mapsto x^t$ is a strictly increasing function on $[0, \infty)$. Therefore, $(a+1)^p > a^p$. We also observe that $(a+1)^{s-p} - 1 > a^{s-p} - 1$, provided that $s - p > 0$. If $s - p = 0$, then we have the identity $(a+1)^{s-p} - 1 = a^{s-p} - 1$. Combining the two estimates, we obtain

$$1 \leq \frac{(a+1)^p}{a^p} \frac{(a+1)^{s-p} - 1}{a^{s-p} - 1}.$$

ultiplying both sides of the inequality by $a^p(a^{s-p}-1) = a^s - a^p$, we obtain

$$a^s - a^p \le (a+1)^s - (1+a)^p,$$

or

$$(a+1)^p - a^p \le (a+1)^s - a^s.$$

□

Setting $s = 1$ and replacing a by a/b, we have $b^p(a+b)^p \le b^p(a^p + b^p)$, or

$$(a+b)^p \le a^p + b^p$$

for all $a \ge 0$ and $b \ge 0$. We now set $a = |f-g|$ and $b = |g-h|$ to conclude that d_p is a metric.

(b) $t \mapsto \int_0^t |f|^p$ is a continuous mapping, and so we can invoke the intermediate value theorem to find t_0 such that $\int_0^{t_0} |f|^p = \frac{1}{2}\int_0^1 |f|^p$. If we let $f_0 = 2f\chi_{[0,t_0]}$ and $f_1 = 2f\chi_{(t_0,1]}$, then $(f_0+f_1)/2 = f$ and

$$d_p(f_0, 0) = d_p(f_1, 0) = 2^{p-1} d_p(f, 0).$$

(c) $f_{nm} = (1/2)f_{nm0} + (1/2)f_{nm1}$, and so no element of A is an extreme point.

(d) Let $K = A \cup \{0\} \cup -A$. Since $f, f_0, f_1, f_{00}, f_{01}, f_{10}, f_{11}, \ldots$ is a sequence that converges to 0, every sequence in A converges to a point in K. Similarly, every sequence in $-A$ converges to a point in K. It follows that K is sequentially compact, hence compact. □

Problem 33 (Homework set 3, Problem 5). Now consider $X = l^p(\mathbb{N})$, $0 < p < 1$. Using the reverse strategy of the above problem, find a compact set in X whose convex hull is not bounded.

Proof. Let $(a_n)_{n=1}^\infty$ be a positive, strictly decreasing sequence such that $a_n \to 0$ and $a_n < 1$. We set

$$a_n^N = \begin{cases} a_n & \text{if } n = N \text{ and} \\ 0 & \text{otherwise,} \end{cases}$$

so that $\{(a_n^N)_{n=1}^\infty : n \in \mathbb{N}\}$ is a countable subset of $l^p(\mathbb{N})$. Observe that

$$d_p((a_n^N), 0) == |a_N|^p,$$

which decreases strictly to 0. It follows that

$$\{0\} \cup \{(a_n^N)_{n=1}^\infty : n \in \mathbb{N}\}$$

is sequentially compact, hence compact.

We now let $(b_n^{1,1})_n = (a_n^1)_n$, $(b_n^{2,k})_n = \frac{1}{2}(a_n^{2k-1})_n + \frac{1}{2}(a_n^{2k})_n$, and

$$(b_n^{l,k})_n = \frac{1}{2}(b_n^{l-1,2k-1})_n + \frac{1}{2}(b_n^{l-1,2k})_n$$

for all $l, k \in \mathbb{N}$. Then $(b_n^{l,k})_n \in \operatorname{co}(K)$ for all $l, k \in \mathbb{N}$. We see that

$$d_p((b_n^{l,k})_n, 0) = \frac{1}{2^{lp}} \sum_{n=1}^{2^{lp}} |a_n|^p.$$

This implies that $\operatorname{co}(K)$ is unbounded if the sequence $(a_n)_{n=1}^\infty$ also satisfies the identity

$$\lim_{N\to\infty} \frac{1}{N^p} \sum_{n=1}^{N} |a_n|^p = \infty.$$

But there are many such sequences. Take, for example, $a_n = 1/\log(n+1)$. $\square$

Bibliography

[Arv76] William Arveson, *An invitation to c^*-algebras*, Springer, 1976.

[Bre93] Glen E. Bredon, *Topology and geometry*, Springer, 1993.

[Bre11] Haïm Brezis, *Functional analysis, sobolev spaces and partial differential equations*, Springer, 2011.

[Car12] Eric A. Carlen, *Math 508 - operator algebras*, unpublished lecture notes, 2012.

[Dou98] Ronald G. Douglas, *Banach algebra techniques in operator theory*, Springer, 1998.

[DR50] A. Dvoretzky and C. A. Rogers, *Absolute and unconditional convergence in normed linear spaces*, Proceedings of the National Academy of Sciences **36** (1950), no. 3, 192–197.

[Fol99] Gerald B. Folland, *Real analysis: Modern techniques and their applications*, Wiley-Interscience, 1999.

[GN43] Israel Gelfand and Mark Naimark, *On the embedding of normed rings into the ring of operators in hilbert space*, Mathematicheskii Sbornik **12** (1943), no. 2, 197–217.

[Gra10a] Loukas Grafakos, *Classical fourier analysis*, Springer, 2010.

[Gra10b] ———, *Modern fourier analysis*, Springer, 2010.

[Jec06] Thomas Jech, *Set theory*, 3rd ed., Springer, 2006.

[Kim12] Mark H. Kim, *Interpolation theorems in harmonic analysis*, http://arxiv.org/abs/1206.2690, 2012, unpublished senior thesis.

[Kun09] Kenneth Kunen, *The foundation of mathematics*, College Publications, 2009.

[Lax02] Peter D. Lax, *Functional analysis*, Wiley-Interscience, 2002.

[Lec] *David Lecomte's lecture notes on weak topologies*, http://perso.crans.org/lecomte/Math/WeakTopologies.pdf.

[Mac98] Saunders Mac Lane, *Categories for the working mathematician*, 2nd ed., Springer, 1998.

[MO1] *When does Cantor-Bernstein hold?*, http://mathoverflow.net/questions/1058/.

[MS13] Camil Muscalu and Wilhelm Schlag, *Classical and multilinear harmonic analysis*, Cambridge University Press, 2013.

[nLa] *Banach Space in nLab*, http://ncatlab.org/nlab/show/Banach+space.

[Ric45] Charles E. Rickart, *Banach algebras with an adjoint operation*, Annals of Mathematics **47** (1945), no. 3, 528–550.

[Rud91] Walter Rudin, *Functional analysis*, second ed., McGraw-Hill, 1991.

[Seg47] Irving E. Segal, *Irreducible representations of operator algebras*, Bulletin of the American Mathematical Society **53** (1947), 73–88.

[SS05] Elias M. Stein and Rami Shakarchi, *Real analysis: measure theory, integration, and hilbert spaces*, Princeton University Press, 2005.

[SS11] ———, *Functional analysis: Introduction to further topics in analysis*, Princeton University Press, 2011.

[Tao10] Terence Tao, *An epsilon of room, i: Real analysis*, American Mathematical Society, 2010.

[Tes09] Gerald Teschel, *Mathematical methods in quantum mechanics: With applications to schrödinger operators*, AMerican Mathematical Society, 2009.

[Thi06] Christoph Thiele, *Wave packet analysis*, American Mathematical Society, 2006.

[Wei95] Charles A. Weibel, *An introduction to homological algebra*, Cambridge University Press, 1995.

[Yos80] Kôsaku Yosida, *Functional analysis*, sixth ed., Springer, 1980.