FUNCTIONAL ANALYSIS

Eric A. Carlen
Professor of Mathematics
Rutgers University

March, 2017

Contents

Chapter 1

Essential Results form Topology

1.1 Introduction

The basic strategy in real analysis is *approximation*. In particular, one often tries to approximate general elements of some infinite dimensional vector space of functions by elements of a subspace consisting of well-behaved functions, or one tries to construct solutions to equations as limits of solutions to "approximate" equations. The basic framework for making mathematical sense of "approximation" is provided by the theory of *metric spaces*, and more generally the theory of *topological spaces*. Methods of approximation are especially effective in a metric space that is *complete* and rich in *compact sets*, as we explain in this introductory chapter.

In the next chapter, we study the Lebesgue theory of integration. A fundamental advantage of it over previous integration theories is that it permits the construction of many complete metric spaces in which compact sets can be concretely described. This provides an extremely useful framework for solving a wide variety of equations. First, we introduce the fundamental topological theory, and illustrate it with examples that do not require the Lebesgue theory of integration.

1.2 Metric Spaces

1.2.1 DEFINITION (Metric Space)**.** A *metric space* (X, d) consists of a set X and a function $d : X \times X \to [0, \infty)$ satisfying:

(1) $d(x, y) \geq 0$ for all $x, y \in X$ and $d(x, y) = 0 \iff x = y$.

(2) $d(x, y) = d(y, x)$ for all $x, y \in X$.

(3) $d(x, z) \leq d(x, y) + d(y, z)$ for all $x, y, z \in X$.

The inequality in *(3)* is called the triangle inequality, and a function d satisfying *(1)*, *(2)* and *(3)* is called a *metric* on X.

1.2.2 EXAMPLE. *For real numbers $a < b$, let $\mathcal{C}([a, b], \mathbb{R})$ denote the set of of continuous real-valued* *functions on* $[a,b]$*. Given any two* $f,g \in \mathcal{C}([a,b],\mathbb{R})$*. Define*

$$d_\infty(f,g) = \sup_{t\in[a,b]} |f(t) - g(t)| \ .$$

It is easy to check that d_∞ *is a metric on* $\mathcal{C}([a,b],\mathbb{R})$*, called the* uniform metric*. Since* $\mathcal{C}([a,b],\mathbb{R})$ *has an obvious vector space structure, it is our first example of a metric space that is also a vector space of functions.*

1.2.1 Continuity in metric spaces

A function f from X to Y is continuous if a sufficiently small change in the input results in a small change in the output. In other words, $f(x)$ will be a close approximation of $f(x_0)$, if x is a sufficiently close approximation of x_0. Here is the precise version of this in the metric space setting:

1.2.3 DEFINITION (Continuous functions from one metric space to another)**.** Let (X,d_X) and (Y,d_Y) be two metric spaces. Let f be a function from X to Y. Then f *is continuous at* $x_0 \in X$ in case for every $\epsilon > 0$, there is a $\delta_\epsilon > 0$ such that

$$d_X(x,x_0) < \delta_\epsilon \quad \Rightarrow \quad d_Y(f(x),f(x_0)) < \epsilon \ . \tag{1.2.1}$$

The function f is *continuous* in case it is continuous at each $x_0 \in X$.

1.2.4 THEOREM (Continuity and sequences)**.** *Let* (X,d_X) *and* (Y,d_Y) *be two metric spaces. Let* f *be a function from* X *to* Y*. Then* f *is continuous at* $x_0 \in X$ *if and only if for every sequence* $\{x_k\}$ *in* X

$$\lim_{k\to\infty} x_k = x_0 \quad \Rightarrow \quad \lim_{k\to\infty} f(x_k) = f(x_0) \ . \tag{1.2.2}$$

Proof. Suppose that f is continuous, and that $\lim_{k\to\infty} x_k = x_0$. Pick any $\epsilon > 0$, and let δ_ϵ be as in (1.2.1). Choose N so large that for all $k > N$, $d(x_k,x_0) < \delta_\epsilon$. Then, for all $k > N$, $d(f(x),f(x_0)) < \epsilon$. Since ϵ is arbitrary, this shows that $\lim_{k\to\infty} f(x_k) = f(x_0)$.

Next suppose that f is not continuous at x_0. Then there exists some $\epsilon > 0$ so that for every $\delta > 0$, there is at least one point x satisfying $d_X(x,x_0) < \delta$ such that $d_Y(f(x),f(x_0)) > \epsilon$. Define a sequence $\{x_k\}$ as follows: For each k, choose x_k so that $d_X(x_k,x_0) < 1/k$ such that $d_Y(f(x_k),f(x_0)) > \epsilon$. Then $\lim_{k\to\infty} d_X(x_k,x_0) = 0$, but it is not the case that $\lim_{k\to\infty} f(x_k) = f(x_0)$. Thus, when f is not continuous at x_0, (2.1.5) does not hold true. Hence, whenever (2.1.5) does hold true, it must be the case that f is continuous at x_0. □

There is another characterization of continuity involving the notion of *open sets*, which we now define:

1.2.5 DEFINITION (Open sets in metric spaces)**.** Let X be a metric space with metric d. Given a number $r > 0$, and a point $x \in X$, let $B_r(x)$ be defined by

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

This set is called the *open ball of radius* r *about* x.

A subset U of X is *open* in case *either*:

(1) It is the empty set $\emptyset$, *or else*

(2) For each $x \in U$, there is an $r > 0$, depending on x, such that

$$B_r(x) \subset U \ .$$

It is possible, and as we shall see, useful to characterize the continuity of functions between two metric spaces simply in terms of open sets, without explicit reference to the the specific metrics themselves.

1.2.6 THEOREM (Continuity and open sets)**.** *Let X and Y be metric spaces with metrics d_X and d_Y respectively. Let f be a function from X to Y. Then f is continuous if and only if for every open set U in Y, $f^{-1}(U)$ is open in X.*

Proof. Suppose that f is continuous, and let U be an open set in Y. If $f^{-1}(U) = \emptyset$, then $f^{-1}(U)$ is open by *(1)*. Otherwise, if $f^{-1}(U) \neq \emptyset$, consider any $x \in f^{-1}(U)$. Then $f(x) \in U$, and since U is open, there exists a $\epsilon > 0$ such that $B_\epsilon(f(x)) \subset U$. Then, since f is continuous at x_0, there is a $\delta > 0$ so that

$$d_X(\tilde{x}, x) < \delta \Rightarrow d_Y(f(\tilde{x}), f(x)) < \epsilon \ .$$

Hence

$$f(B_\delta(x)) \subset B_\epsilon(f(x)) \subset U \ .$$

But this means that

$$B_\delta(x) \subset f^{-1}(U) \ .$$

Since x was any point in $f^{-1}(U)$, we have shown that $f^{-1}(U)$ contains an open ball about each of its members, and hence is open.

Conversely, suppose that f has the property that whenever U is open in Y, $f^{-1}(U)$ is open in X. Fix any $x \in X$ and any $\epsilon > 0$. $f^{-1}(B_\epsilon(f(x)))$ is open and contains x. Therefore, there is some $\delta_\epsilon > 0$ such that

$$B_{\delta_\epsilon}(x) \subset f^{-1}(B_\epsilon(f(x))) \ .$$

But then

$$f(B_{\delta_\epsilon}(x)) \subset B_\epsilon(f(x)) \ ,$$

which is just another way to write (1.2.1). Since ϵ is arbitrary, f is continuous at x. Since x is arbitrary, f is continuous. □

1.2.2 Complete metric spaces

The metric space $(\mathcal{C}([a,b],\mathbb{R}), d_\infty)$ has another feature that is desirable for analysis: It is *complete.*

1.2.7 DEFINITION (Complete metric space)**.** A metric space (X, d) is complete in case whenever $\{x_k\}_{k\in\mathbb{N}}$ is a Cauchy sequence in X, there exists an $x \in X$ such that

$$\lim_{k\to\infty} d(x_k, x) = 0 \ .$$

1.2.8 THEOREM. $(\mathcal{C}([a,b],\mathbb{R}), d_\infty)$ *is complete.*

Proof. Suppose that $\{f_k\}_{k\in\mathbb{N}}$ is a Cauchy sequence in $(\mathcal{C}([a,b],\mathbb{R}),d_\infty)$. Then for any $\epsilon > 0$, there is an N_ϵ so that

$$k,\ell \ge N_\epsilon \Rightarrow d_\infty(f_k,f_\ell) \le \epsilon \ .$$

Fix any $t \in [a,b]$, Since $|f_k(t) - f_\ell(t)| \le d_\infty(f_k,f_\ell)$,

$$k,\ell \ge N_\epsilon \to |f_k(t) - f_\ell(t)| \le \epsilon \ , \tag{1.2.3}$$

and hence $\{f_k(t)\}$ is a Cauchy sequence in $\mathbb{R}$. By the completeness of the real numbers, it has a limit which we denote by $f(t)$. This defines a real valued function f on $[a,b]$, and it remains to show that $f \in \mathcal{C}([a,b],\mathbb{R})$, and that $\lim_{k\to\infty} d_\infty(f_k,f) = 0$.

By the definition of $f(t)$, $|f_k(t) - f(t)| = \lim_{\ell\to\infty}|f_k(t) - f_\ell(t)|$. For any $s,t \in [0,1]$,

$$|f(t) - f(s)| \le |f(t) - f_k(t)| + |f_k(t) - f_k(s)| + |f_k(s) - f(s)| \ ,$$

and so

$$k \ge N_\epsilon \to |f_k(t) - f(t)| \le \epsilon \ . \tag{1.2.4}$$

For $k = N_\epsilon$, this becomes

$$|f(t) - f(s)| \le |f_{N_\epsilon}(t) - f_{N_\epsilon}(s)| + 2\epsilon \ .$$

Since f_{N_ϵ} is continuous, there is a $\delta > 0$ so that $|s-t| \le \delta \Rightarrow |f_{N_\epsilon}(s) - f_{N_\epsilon}(t)| \le \epsilon$. Therefore,

$$|s-t| \le \delta \Rightarrow |f(s) - f(t)| \le 3\epsilon \ .$$

Since $\epsilon > 0$ is arbitrary, this proves that $f \in \mathcal{C}([a,b],\mathbb{R})$. Finally, since (1.2.4) is valid uniformly in t,

$$\lim_{k\to\infty} d_k(f_k,f) = 0 \ .$$

□

The next theorem provides an example of the importance of completeness.

1.2.9 THEOREM (Banach Contraction Mapping Theorem)**.** *Let (X,d) be a complete metric space. Let $\Phi : X \to X$ have the property that for some $\lambda < 1$, $d(\Phi(x),\Phi(y)) \le \lambda d(x,y)$ for all $x,y \in X$. Then there is a unique $x_0 \in X$ such that*

$$x_0 = \Phi(x_0) \ . \tag{1.2.5}$$

Moreover, for all $x \in X$, let the sequence $\{x_k\}_{k\in\mathbb{N}}$ be defined by $x_1 = x$ and for $x_{k+1} = \Phi(x_k)$. Then

$$\lim_{k\to\infty} d(x_0,x_k) = 0 \ . \tag{1.2.6}$$

Proof. Pick any $x \in X$, and construct the sequence $\{x_k\}_{k\in\mathbb{N}}$ as described in the theorem. Define $R := d(x_2,x_1) = d(\Phi(x),x))$. By the hypothesis

$$d(x_3,x_2) = d(\Phi(x_2),\Phi(x_1)) \le \lambda d(x_2,x_1) = \lambda R \ .$$

Then by a simple induction,

$$d(x_{k+2},x_{k+1}) \le \lambda^k R$$

for all $k \in \mathbb{N}$. By the triangle inequality, for all $\ell > k$,

$$d(x_\ell, x_k) \leq \sum_{j=0}^{\ell-k-1} d(x_{k+j+1}, x_{k+j} \leq \sum_{j=0}^{\infty} \lambda^{k+j-1} R = \frac{\lambda^{k-1}}{1-\lambda} R \ .$$

Since $\lim_{k\to\infty} \lambda^k = 0$, this proves that $\{x_k\}_{k\in\mathbb{N}}$ is a Cauchy sequence. Since (X, d) is complete, this sequence has a limit x_0. Clearly

$$\Phi(x_0) = \lim_{k\to\infty} \Phi(x_k) = \lim_{k\to\infty} x_{k+1} = x_0$$

so that x_0 is a fixed point of Φ. Finally if y_0 is any fixed point of Φ, then

$$d(x_0, y_0) = d(\Phi(x_0), \Phi(y_0)) \leq \lambda d(x_0, y_0) \ .$$

The only $t \in [0, \infty)$ satisfying $t \leq \lambda t$ is $t = 0$. By property *(1)* in the definition of a metric, this implies that $y_0 = x_0$ which proves the uniqueness. □

1.2.10 EXAMPLE. *The previous theorem is the basis of the basic existence and uniqueness theorem for ordinary differential equations; we now sketch the main points in the simplest case. Let $v(x,t)$ be a continuous function of $\mathbb{R} \times \mathbb{R}$ such that for some $L < \infty$,*

$$|v(x,t) - v(y,t)| \leq L|x - y|$$

for all x, y, t. For $x(\cdot) \in \mathcal{C}([0,(2L)^{-1}], \mathbb{R})$ and $x_0 \in \mathbb{R}$ define

$$\Phi(x(\cdot))(t) = x_0 + \int_0^t v(x(s), s)\mathrm{d}s \ .$$

Note that $x(\cdot)$ is a fixed point of Φ is and only if for all t,

$$x(t) = x_0 + \int_0^t v(x(s), s)\mathrm{d}s \ .$$

Suppose that such a fixed point exists. Then, since the right hand side is continuously differentiable, $x(t)$ is continuously differentiable, and differentiating both sides,

$$x'(t) = v(x(t), t) \tag{1.2.7}$$

for all t, and moreover, $x(0) = x_0$. Conversely, any solution of (1.2.7) with $x(0) = x_0$ is a fixed point of Φ. Hence, proving existence and uniqueness of fixed points of Φ is tantamount to proving the existence and uniqueness of solutions of the ordinary differential equation (1.2.7), at least on this time interval. To apply the Banach Contraction Mapping Theorem, consider the complete metric space $\mathcal{C}([0,(2L)^{-1}], \mathbb{R})$ equipped with the d_∞ metric.

Let $x(\cdot), y(\cdot) \in \mathcal{C}([0,(2L)^{-1}], \mathbb{R})$. Then for all $0 \leq t \leq (2L)^{-1}$,

$$\begin{aligned}
|\Phi(x(\cdot))(t) - \Phi(y(\cdot))(t)| &= \left| \int_0^t [v(x(t), t) - v(y(t), t)]\mathrm{d}t \right| \\
&\leq \int_0^t |v(x(t), t) - v(y(t), t)]|\mathrm{d}t \\
&\leq L \int_0^t |x(t) - y(t)|\mathrm{d}t \\
&\leq Lt \sup_{0\leq s\leq t} |x(t) - y(t)| \leq \frac{1}{2} d_\infty(x(\cdot), y(\cdot))
\end{aligned}$$

Thus,

$$d_\infty(\Phi(x(\cdot)), \Phi(y(\cdot))) \leq \frac{1}{2} d_\infty(x(\cdot), y(\cdot)) \ ,$$

and the Banach Contraction Mapping Theorem yields the existence of a unique fixed point of Φ*. This yields the existence and uniqueness of a solution to (1.2.7) on the interval* $[0, (2L)^{-1}]$*. The basic idea explained here can be used to prove much more.*

We have just explained how *completeness* can be applied to solve equations. We shall do the same for compactness after developing more of the theory.

1.2.3 Compactness in metric spaces

Alongside completeness, the other fundamental concept pertaining to approximation in analysis is that of compactness:

1.2.11 DEFINITION (Sequentially compact subset of metric space)**.** Let (X, d) be a metric space, and $A \subset X$. Then A is *sequentially compact* in case every sequence $\{x_k\}_{k\in\mathbb{N}}$ in A contains a subsequence converging to an element of A. A metric space (X, d) is *sequentially compact* in case X itself is sequentially compact.

Note that if (X, d) is a metric space, and $A \subset X$, and d_A denotes the restriction of d to $A \times A$, the (A, d_A) is itself a metric space, and A is sequentially compact subset of X if an only if (A, d_A) is a compact metric space. Thus, characterizing compact subsets of a metric space reduces to a question about whether or not a metric spaces is compact.

Later in this chapter we shall prove the *Arzela-Ascoli Theorem* which characterizes compact sets in $(\mathcal{C}([a, b]), d_\infty)$, and we shall give an example of the application of this to solving equations. There is an equivalent formulation of sequential compactness in a metric space that will be useful in proving the Arzela-Ascoli Theorem and other theorems characterizing compact sets.

1.2.12 DEFINITION (Totally bounded)**.** A metric space (X, d) is totally bounded if and only if for every $\epsilon > 0$, there is a finite set $\mathcal{U}_\epsilon$ consisting of finitely many open balls of radius ϵ; i.e.,

$$\mathcal{U}_\epsilon = \{B_\epsilon(x_1), \ldots, B_\epsilon(x_{n_\epsilon})\} \ ,$$

that covers X; i.e, $X = \bigcup_{j=1}^{n_\epsilon} B_\epsilon(x_j)$.

1.2.13 THEOREM (Sequential compactness and total boundedness)**.** *A metric space* (X, d) *is sequentially compact if and only if it is complete and totally bounded.*

Proof. Suppose that (X, d) is not totally bounded. We shall show that then it is not sequentially compact. To do this, we construct a sequence that has no convergent subsequence. By hypothesis, for some $\epsilon > 0$, there does not exist any finite cover of X by open balls of radius ϵ. Thus given any set $\{x_1, \ldots, x_n\}$ of X, $\bigcup_{j=1}^{n_\epsilon} B_\epsilon(x_j) \neq X$, and so we can select x_{n+1} so that

$$d(x_{n+1}, x_j) \geq \epsilon \ , \quad \text{for all } j = 1, \ldots, n \ .$$

Thus, starting from an arbitrary choice of x_1, using a simple induction we can construct an infinite sequence $\{x_k\}_{k\in\mathbb{N}}$ such that

$$d(x_k, x_\ell) \geq \epsilon$$

for all $k \neq \ell$. Clearly, such a sequence has no convergent subsequence.

Next, suppose that (X, d) is totally bounded and complete. Let $\{x_k\}_{k\in\mathbb{N}}$ be any infinite sequence in X. For each $m \in \mathbb{N}$, there exists a set $\{B_{1/m}(x_1), \dots B_{1/m}(x_{n_m})\}$ of open balls of radius $1/m$ such that

$$X = \bigcup_{j=1}^{n_{1/m}} B_\epsilon(x_j) \ .$$

By the pigeon-hole principle, at least one of these balls contains infinitely many elements of any infinite sequence in X.

By what we have explained above, there exists an infinite subsequence $\{x_k^{(1)}\}_{k\in\mathbb{N}}$ of $\{x_k\}_{k\in\mathbb{N}}$ such that all elements of this subsequence lie in some open ball of radius 1. Next, for the same reason, exists an infinite subsequence $\{x_k^{(2)}\}_{k\in\mathbb{N}}$ of $\{x_k^{(1)}\}_{k\in\mathbb{N}}$ such that all elements of this subsequence lie in some open ball of radius 1/2. Continuing the obvious induction, we obtain a sequence $\{x_k^{(j)}\}_{k\in\mathbb{N}}$ of sequences such that each $\{x_k^{(j+1)}\}_{k\in\mathbb{N}}$ is a subsequence of $\{x_k^{(j)}\}_{k\in\mathbb{N}}$, and all elements in $\{x_k^{(j)}\}_{k\in\mathbb{N}}$ lie in some open ball of radius $1/j$.

Now we use Cantor's "diagonal sequence" construction: define $y_j = x_j^{(j)}$. Then $\{y_j\}_{j\in\mathbb{N}}$ is a subsequence of $\{x_k\}_{k\in\mathbb{N}}$, and for all m,

$$j, k \geq m \Rightarrow d(y_j, y_k) \leq 1/m \ .$$

Thus, $\{y_k\}_{k\in\mathbb{N}}$ is a Cauchy sequence. Since (X, d) is complete, it is also convergent. We have thus shown the existence of a convergent subsequence of an arbitrary sequence of X. □

Many important theorems can be proved by using the fact that real valued continuous functions on a sequentially compact metric space have maxima and minima. That is, if f is a real valued continuous function on a sequentially compact metric space, the there exist x_0 and x_1 such that

$$f(x_0) \leq f(x) \leq f(x_1)$$

for all $x \in X$.

While most often in the chapters that follow, we shall be working with the notions of continuity and compactness in a metric space setting, this is not always possible, and it is not always convenient even when it is possible. It is advantageous to develop these notions in a more general setting, that of *topological spaces*. We now introduce this more general setting.

1.3 Topological spaces

Since by Theorem 1.2.6 we can characterize continuous functions from one metric space to another in terms of open sets, without explicitly mentioning either metric at all, it is sometimes useful to "strip away" the metric structure, and only refer to the open sets.

1.3.1 DEFINITION (Topological Spaces, Hausdorff Topological Spaces)**.** Let X be any set, and let $\mathcal{O}$ be any collection of sets in X satisfying:

(1) The empty set $\emptyset$ belongs to $\mathcal{O}$, as does X itself.

(2) The union of any *arbitrary* set of sets in $\mathcal{O}$ belongs to $\mathcal{O}$.

(3) The intersection of any *finite* set of sets in $\mathcal{O}$ belongs to $\mathcal{O}$.

In this case, $\mathcal{O}$ is said to be a *topology* on X, and the sets belonging to $\mathcal{O}$ are called *open* sets in X (for the topology in question). A subset A of X is *closed* in case its complement, A^c is open.

The pair $(X, \mathcal{O})$ is said to be a *topological space*. It is a *Hausdorff* if whenever x, y are any two distinct elements in X, there exists disjoint open sets sets U and V with $x \in U$ and $y \in V$.

Note that by De Morgan's laws, the intersection of any *arbitrary* set of closed sets in X is itself closed.

1.3.2 EXAMPLE. *It is left as an easy exercise to show that if X is any metric space, and $\mathcal{O}$ is the collection of all open sets in X, as defined above in terms of open balls, $\mathcal{O}$ does indeed constitute a topology on X.*

If (X, d) is any metric space and x, y are distinct in X, then $r := d(x, y) > 0$. If $w \in B_{r/3}(x)$ and $z \in B_{r/3}(y)$, then

$$r = d(x, y) \leq d(x, w) + d(w, z) + d(z, y) < d(w, z) + 2r/3 \ ,$$

so that $d(w, z) > r/3$, In particular, $B_{r/3}(x) \cap B_{r/3}(y) = \emptyset$. Since $B_{r/3}(x)$ and $B_{r/3}(y)$ are open sets containing x and y respectively, this proves that every metric space is Hausdorff.

Not every topology is Hausdorff. If X is an set, the trivial topology *on X is given by $\mathcal{O} = \{\emptyset, X\}$. This is evidently a topology, and when X contains more than a single element, it cannot be Hausdorff.*

1.3.3 EXAMPLE (Relative topology on a subset)**.** *Let $(X, \mathcal{O})$ be a topological space. Let A be any subset of X. The* relative topology *on A induced by $\mathcal{O}$ is denoted by $\mathcal{O}_A$ and is given by*

$$\mathcal{O}_A = \{\ U \cap A \ :\ U \in \mathcal{O}\ \}\ .$$

It is readily checked that this is a topology.

1.3.4 DEFINITION (Metrizable Topology)**.** Let $(X, \mathcal{O})$ be a topological space. The topology $\mathcal{O}$ is *metrizable* if there exists some metric d on $X \times X$ such that the set of open sets for this metric is exactly $\mathcal{O}$.

The remarks in Example 1.3.2 show that non-Hausdorff topologies are never metrizable. We shall be (almost) exclusively concerned with Hausdorff topologies, but as we shall see, there are useful Hausdorff topologies that are not metrizable.

The next definitions introduces some more useful terminology

1.3.5 DEFINITION (Interior, closure and neighborhoods)**.** Let $(X, \mathcal{O})$ be a topological space, and A a subset of X.The *interior* of A, A^o, is the union of all of the open sets contained in A. The *closure of* A, $\overline{A}$, is the intersection of all of the closed sets containing A. Finally for any $x \in X$, the set $\mathcal{N}_x$ of *neighborhoods of* x consists of all sets B such that $x \in B^o$.

1.3.1 Continuity in topological spaces

1.3.6 DEFINITION (Continuous functions between topological spaces)**.** Let $(X, \mathcal{O}_X)$ and $(Y, \mathcal{O}_Y)$ be two topological spaces. A function f from X to Y is *continuous at* $x \in X$ in case for every neighborhood V of $f(x)$, there is a neighborhood U of x such that $f(U) \subset V$.

A function f from X to Y is *continuous* whenever U is open in Y, $f^{-1}(U)$ is open in X.

It is easy to see that $f : X \to Y$ is continuous if and only if it is continuous at each $x \in X$.

1.3.7 Remark. By Example 1.3.2 and by Theorem 1.2.6, the definition of continuity that we make next is consistent with our existing notion of continuity in the metric space setting.

We now turn to the notion of *approximation* in topological spaces. The following definition is the starting point.

1.3.8 DEFINITION (Limit points in a topological space)**.** Let $(X, \mathcal{O}_X)$ be a topological space. If A is any set in X, a point $x \in X$ is a *limit point* of X in case every for every open set U that contains x,

$$A \cap (U \backslash \{x\}) \neq \emptyset \ .$$

That is, every open set U that contains x also contains some point in A *other than* x. Note that x itself may or may not be in A.

Note that if $(X, \mathcal{O}_X)$ is Hausdorff, and x is a limit point of $A \subset X$, with $x \notin A$, then not only is $A \cap U$ non-empty for every neighborhood U of x: $A \cap U$ *must contain infinitely many points.* To see this, suppose $y \in A \cap U$. Let V_x and V_y be disjoint open sets containing x and y respectively. Then $V_x \cap U$ is another neighborhood of x, contained in U, so $A \cap (V_x \cap U) \neq \emptyset$. Since $A \cap (V_x \cap U) \subset A \cap U$, and is missing at least y. Repeating this procedure, it is clear that we can repeatedly remove elements from $A \cap U$ without ever emptying it, and so it must contain infinitely many points.

1.3.9 DEFINITION (convergent sequence)**.** Let $(X, \mathcal{O})$ be a topological space. Let $\{x_k\}$ be a sequence of elements of X. Then $\{x_k\}_{k\in\mathbb{N}}$ *is convergent to* x in case every open set U containing x also contains all but finitely many terms in the sequence $\{x_k\}_{k\in\mathbb{N}}$.

Note the differences between the notions of limit point the notion of the limit of a sequence. One difference is that a sequence $\{x_k\}_{k\in\mathbb{N}}$ is a *function* from $\mathbb{N}$ to X, though it is common practice to identify the sequence with its range, which is a subset of X. Apart from this, there is the essential difference between “infinitely many” and “all but finitely many”.

In particular, in a Hausdorff space, a sequence can have at most one limit, but (identified with its range, and considered as a set), but it may have more than one limit point, since x is a limit point of $\{x_n\}_{n\in\mathbb{N}}$ if and only if every open set U containing x also infinitely many terms in the sequence $\{x_k\}$, while $\lim_{n\to\infty} x_n = x$ if and only if every neighborhood U of x contains all but finitely many terms in the sequence.

In a metric space, there is of course a characterization of limit points in terms of sequences; x is a limit point of A if and only if there is a sequence $\{x_n\}_{n\in\mathbb{N}}$ of elements in A such that $\lim_{n\to\infty} x_{k_n} = x$.

1.3.10 EXAMPLE (The right order topology on $\mathbb{R}$)**.** *Let $\mathcal{O}_r$ be the set of all subsets of $\mathbb{R}$ of the form (a, ∞), $a \in \mathbb{R}$, together with $\emptyset$ and $\mathbb{R}$. It is readily checked that this is a topology, called the* right order topology. *Conisder the function from $\mathbb{R}$ to $\mathbb{R}$ defined by*

$$f(x) := \begin{cases} 1 & x > 0 \\ 0 & x \leq 0 \ . \end{cases}$$

Then

$$f^{-1}((a,\infty)) = \begin{cases} \emptyset & a \geq 1 \\ (0,\infty) & 0 \leq a < 1 \\ \mathbb{R} & a < 0\ . \end{cases}$$

Therefore, if we equip the range with the right order topology and the domain with the usual topology, f is conitnuous, though it is not continuous if both the domain and range are equipped with the usual topology.

Let $(X.\mathcal{O})$ be a topological space. A functions that is continuous from $(X.\mathcal{O})$ to $(\mathbb{R},\mathcal{O}_r)$ is called lower semicontinuous*. The left order topology and upper semicontinuous functions are defined in the analogous way using the sets $(-\infty,b)$.*

We are now ready for the theorem that justifies the terminology "closed":

1.3.11 THEOREM (Closed sets and limit points)**.** *Let $(X,\mathcal{O})$ be any topological space. A subset A of X is closed if and only if A contains all of its limit points.*

Proof. Suppose that A is closed, and $x \in A^c$. Since A^c is open, there is an open set U containing x that has an empty intersection with A. Thus, x is not a limit point of A. Since x was an arbitrary point outside A, A must contain all of its limit points.

On the other hand, suppose that A contains all of its limit points. We must show that A is closed, or, what is the same thing, that A^c is open. Consider any point $x \in A^c$. Since it is not a limit point of x, there is an open set U_x containing x that has empty intersection with A. For each $x \in A^c$, chose such a U_x. But then, since U_x contains x,

$$A^c \subset \bigcup_{x\in A^c} U_x$$

On the other hand, since each $U_x \subset A^c$,

$$\bigcup_{x\in A^c} U_x \subset A^c$$

Thus, $A = \bigcup_{x\in A^c} U_x$, and by *(2)* in the definition of topological spaces, $\bigcup_{x\in A^c} U_x$ is open. □

We close this subsection with one more definition:

1.3.12 DEFINITION (Density)**.** Let $(X,\mathcal{O})$ be a topological space. Let $A \subset B \subset X$. Then A *is dense in* B in case the closure of A contains B.

By Theorem 1.3.11, A is dense in B if and only if every point in B is a limit point in A; i.e, if every point in B can be approximated arbitrarily well by points in A.

1.3.2 Compactness in topological spaces

1.3.13 DEFINITION (Compact Sets)**.** Let $(X,\mathcal{O}_X)$ be a topological space. A subset K is called *compact* in X in case for *open cover* $\mathcal{U}$ of K; that is, for every collection $\mathcal{U}$ of open sets such that

$$K \subset \bigcup_{u\in\mathcal{U}} U\ ,$$

there is a finite subset $\mathcal{G}$ of $\mathcal{U}$ that also covers K:

$$K \subset \bigcup_{u \in \mathcal{G}} U \ .$$

$\mathcal{G}$ is called a *finite subcover* of A. The topological space $(X, \mathcal{O})$ is called compact in case X itself is compact.

1.3.14 THEOREM. *Let $(X, \mathcal{O})$ be any compact topological space. Then every closed subset K of X is compact. Conversely, if X is Hausdorff, every compact subset of X is closed.*

Proof. Let $\mathcal{U}$ be any open cover of K. Then $\mathcal{U} \cup \{K^c\}$ is an open cover of X. Thus, there exists a finite set of the sets in sets in $\mathcal{U} \cup \{K^c\}$ that covers all of X, and K^c is not needed to cover K. Hence K is compact.

Conversely, suppose K is compact and $(X, \mathcal{O})$ is Hausdorff. Suppose that $y \notin K$, but that y is a limit point of K. For each $x \in K$, there exist open sets U_x and V_x such that $U_x \cap V_x = \emptyset$, $x \in U_x$ and $y \in V_x$. Then $\{U_x \ : \ x \in K\}$ is an open cover of K, so there exists a finite subcover $\{U_{x_1}, \ldots, U_{x_n}\}$ that covers K. However,

$$y \in V := \bigcap_{j=1}^{n} V_{x_j}$$

and V is open. Since y is a limit point of K, $K \cap V \neq \emptyset$. But this is impossible since $V \cap U_{x_j} = \emptyset$ for each j, and $K \subset \cup_{j=1}^{n} U_{x_j}$. □

Our first example of a convergence theorem involving compactness is the classical result known as *Dini's Theorem.* In proving this, we shall make use of the following fact: If $\mathcal{U}$ is any set of open subsets of X, then by De Morgan's laws,

$$\left(\bigcup_{U \in \mathcal{U}} U \right)^c = \bigcap_{U \in \mathcal{U}} U^c \ ,$$

Thus $\mathcal{U}$ is an open cover if and only if $\{ \ U^c \ : \ U \in \mathcal{U} \ \}$ is a set of closed subsets of X with empty intersection.

Therefore, X is compact if and only if whenever $\mathcal{K}$ is a set of closed subsets of X such that

$$\bigcap_{K \in \mathcal{K}} K = \emptyset \ ,$$

there is a finite subset $\{K_1, \ldots, K_n\} \subset \mathcal{K}$ such that

$$\bigcap_{j=1}^{n} K_j = \emptyset \ .$$

This analysis is often summarized by saying that X *is compact if and only if* X *has the "finite intersection property"*.

1.3.15 THEOREM (Dini's Theorem)**.** *Let $(X, \mathcal{O})$ be a compact topological space, and let $\{f_n\}_{n \in \mathbb{N}}$ be a sequence of real valued continuous functions on X, and suppose that there is a continuous real valued function f on X such that for each $x \in X$, the sequence $\{f_n(x)\}_{n \in \mathbb{N}}$ is monotone non-decreasing, and*

$$\lim_{n \to \infty} f_n(x) = f(x) \ .$$

Then

$$\lim_{n\to\infty} f_n = f$$

uniformly.

In other words, pointwise convergence, together with compactness and monotonicity, imply uniform convergence. Also note that replacing each f_n by $-f_n$, one converts a monotone non-decreasing sequence into a monotone non-increasing sequence, and so the theorem remains true if one replaces "monotone non-decreasing" by "monotone non-increasing".

Proof. Fix $\epsilon > 0$. Define the sets K_ℓ, $\ell \in \mathbb{N}$, by

$$K_\ell := \{\ x \in X\ :\ f(x) - f_\ell(x) \geq \epsilon\ \}\ .$$

Since f and f_ℓ are continuous, K_ℓ is closed. Since $\{f_n(x)\}_{n\in\mathbb{N}}$ is monotone non-decreasing,

$$\ell \geq k \Rightarrow K_\ell \subset K_k\ . \tag{1.3.1}$$

Also, since for each x, $\lim_{\ell\to\infty} f_\ell(x) = f(x)$, $\displaystyle\bigcap_{k=1}^{\infty} K_\ell = \emptyset$ Then, by the compactness of X, there is some $n \in \mathbb{N}$ such that

$$\bigcap_{\ell=1}^{n} K_\ell = \emptyset. \tag{1.3.2}$$

Combining (1.3.1) and (1.3.2), we see that $K_\ell = \emptyset$ for all $\ell \geq n$. Hence, for all $\ell > n$, and all x, $|f_\ell(x) - f(x)| < \epsilon$. which proves the uniform convergence. $\square$

Next, we turn to one of the main theorems on compactness.

1.3.16 THEOREM (Compactness, Continuity, and Minima)**.** *Let $(X, \mathcal{O})$ be any topological space, and let K be a compact subset of X. Let f be a function from X to $\mathbb{R}$ that is continuous when $\mathbb{R}$ is equipped with its usual metric topology. Then there exists and $x \in K$ so that*

$$f(x) \leq f(y) \qquad \text{for all} \qquad y \in K\ . \tag{1.3.3}$$

Proof. Consider the open sets $(-n, \infty)$ in $\mathbb{R}$, since f is continuous,

$$\mathcal{U} = \{\ f^{-1}((-n,\infty))\ :n \in \mathbb{N}\ \}$$

is an open cover of X, and hence K. Since K is compact, there exists an open subcover. But for $n > m$, $f^{-1}((-m,\infty)) \subset f^{-1}((-n,\infty))$, so there is an n with $K \subset f^{-1}((-n,\infty))$. In particular, f is bounded from below on K.

Now let a be the greatest lower bound of the numbers $f(y)$ for $y \in K$. We claim that there exists an $x \in K$ with $f(x) = a$. If so, then plainly (1.3.3) is true.

To prove this, let us suppose that there is no such x. Then

$$\mathcal{U} = \{\ f^{-1}((a + 1/n,\infty))\ :n \in \mathbb{N}\ \}$$

is an open cover of K. This means that there is a finite subcover, and again, since the sets in the open cover are nested, a single one of them, say $f^{-1}((a + 1/n, \infty))$, covers K. But this would mean that $f(y) \geq a + 1/n$ for each y in k, which is not possible since a is the greatest lower bound. $\square$

Any point x for which (1.3.3) is true is called a *minimizer of f on X*. Likewise, any point x for which

$$f(x) \geq f(y) \qquad \text{for all} \qquad y \in K\ . \tag{1.3.4}$$

is called a *maximizer of f on X*.

There are several important conclusions to be drawn from this proof. First, if f is continuous, so is $-f$, and a minimizer of $-f$ is a maximizer of f. Hence the theorem implies the existence of both minimizers and maximizers for continuous functions on compact sets.

Now suppose we have a real valued function f defined on a sets K, and we want to know if f has a minimizer in K. If we can find a topology on K that makes f continuous, and makes K compact, then we can apply the previous theorem.

However, the demands of continuity and compactness pull in opposite directions when we look for our topology: The topology has to have sufficiently many open sets in it for f to be continuous, since we need $f^{-1}(U)$ to be open for every open set U in $\mathbb{R}$. On the other hand, the more open sets we include in our topology, the more open covers we have to worry about when showing that every open cover has a finite subcover.

Very often, one is stuck between a rock and a hard place, and there is *no* topology that both makes f continuous, and K compact. Indeed, there are many very nice functions f – such as the exponential function of $\mathbb{R}$ – that simply do not have minimizers or maximizers. While $\mathbb{R}$ is compact under the trivial topology $\mathcal{O} = \{\emptyset, \mathbb{R}\}$, and while the exponential function is continuous under the usual metric topology on $\mathbb{R}$, the fact that the exponential function does not have either a maximizer or a minimizer shows that there is *no* topology on $\mathbb{R}$ under which $\mathbb{R}$ is compact and the exponential function is continuous.

A situation that is frequently encountered in applications is that a function f on X does have, say, a minimizer, but not a maximizer. Also in this situation, it is impossible to find a topology for which f is continuous and X is compact, since them both minima and maxima would exist.

However, if we are just looking for minima, it is worth noticing that in our proof of Theorem 1.3.16, we did not use the full strength of the continuity hypothesis. The same proof yields the same conclusion if we assume only the property that $f^{-1}((t,\infty))$ is open for each t in $\mathbb{R}$.

1.3.17 DEFINITION (Upper and lower semicontinuous function)**.** Let $(X, \mathcal{O}_X)$ be a topological space. A function f from X to $\mathbb{R}$ is called *lower semicontinuous* in case for all t in $\mathbb{R}$, $f^{-1}((t,\infty))$ is open. It is called *upper semicontinuous* in case for all t in $\mathbb{R}$, $f^{-1}((-\infty,t))$ is open. As has been explained in Example 1.3.10, lower semicontinuity of f is the same as conitnuity of f from $(X, \mathcal{O}_X)$ to $(\mathbb{R}, \mathcal{O}_r)$, where $\mathcal{O}_r$ is the right order topology on $\mathbb{R}$.

Summarizing the discussion above, we have the following variant of Theorem 1.3.16:

1.3.18 THEOREM. *Let $(X, \mathcal{O})$ be a topological space, and let $K \subset X$ be either compact or sequentially compact. Let f be a lower semicontinous real valued function on $(X, \mathcal{O})$. Then there exists $x_0 \in K$ such that $f(x_0) \leq f(x)$ for all $x \in K$.*

Thus, we can prove existence of minimizers for f on X by finding a topology that makes f lower semicontinuous, and K compact. This turns out to be a very useful strategy, as we shall see.

Still, to use either Theorem 1.3.1 or Theorem 1.3.16, we need criteria for compactness. How can we tell if a set X is compact? In metric spaces, we can reduce this to a question about sequences.

1.3.19 DEFINITION (Sequential compactness)**.** A topological space $(X, \mathcal{O})$ is *sequentially compact* in case every sequence $\{x_n\}_{n\in\mathbb{N}}$ has a convergent subsequence $\{x_{n_k}\}_{k\in\mathbb{N}}$.

1.3.20 THEOREM (Compactness and subsequences in a Metric space)**.** *Let (X, d) be any metric space, and let K be any subset of X. Then K is compact if and only if every infinite sequence $\{x_k\}$ of elements of K has an infinite subsequence $\{x_{k_n}\}$ that converges to some x in K.*

In other words, a metric space is compact if and only if it is sequentially compact. In the broader setting of topological spaces, there is no relation between compactness and sequential compactness. There are topological spaces that are compact, but not sequentially compact, and there are sequentially compact spaces that are not compact.

The notion of compactness as we have defined it in terms of open covers is a 20th century notion. In the 19th century, mathematicians thought about compactness issues in terms of sequential compactness.

It is important to note that this theorem is *not true* in the general setting of topological spaces; it is important that the topology be a metric topology. Likewise, in the general topological setting, it is not true that a function f is continuous if and only if it takes convergent sequences to convergent sequences.

The close connection between sequences and continuity and compactness that one has in metric spaces does not carry over to the more general topological setting at all. Fortunately, almost all of the topologies that we shall encounter are metric topologies.

Proof of Theorem 1.3.20. The fact that compactness implies sequential compactness in a metric space is relatively easy. Suppose there exists a sequence $\{x_n\}_{n\in\mathbb{N}}$ that has no convergent subseqeunce. Then there is no $y \in K$ such that for all $r > 0$, $B_r(y)$ contains x_k for infinitely many k, since otherwise there would be a subsequence converging to y. (Consider the balls $B_{1/n}(y)$, and apply Cantor's diagonal sequences argument to the sequence of subsequences coinained in these balls.) Hence, for each $y \in K$ there exists an open set U_y that contains y, but which cointains x_k for at most finitely many k. Then $\{U_y \ : \ y \in K\}$ is an open cover of K. Since K is compact, there exists a finite subcover $\{U_{y_1}, \ldots, U_{y_n}\}$. Thus, every k, x_k belongs to U_{y_j} for some j, but each U_{y_j} contains x_k for only finitely many k, which is impossible. Hence a convergent subsequence exists.

Fot the other implication, we assume sequention compactness, and shall prove compactness in four steps.

Step 1: K is bounded: We first show that K is bonded, which means that

$$\sup_{x,y\in K} d(x, y) < \infty \ .$$

This supremum is called the *diameter* of K.

To see that the diameter is finite, suppose that it is not. Under this hypothesis, we construct a sequence $\{x_n\}_{n\in\mathbb{N}}$ as follows. First, fix any $x \in X$. Now for each $n \in N$, choose some $x_n \in K\backslash B_n(x)$. The set $K\backslash B_n(x)$ is not empty when the diameter of K is infinite.

Then, by hypothesis, there is a subsequence $\{x_{n_k}\}_{k\in\mathbb{N}}$ and some $y \in K$ such that

$$\lim_{k\to\infty} x_{n_k} = y \ . \tag{1.3.5}$$

Then by the triangle inequality, we would have

$$d(x, x_{n_k}) \le d(x, y) + d(y, x_{n_k}) \ .$$

But this cannot be: By construction, $d(x, x_{n_k}) > n_k$, while $d(x, y)$ is some fixed, finite number, and for all sufficiently large k, $d(y, x_{n_k}) \le 1$, by (1.3.5). This contradiction shows that K must be bounded.

Step 2: K contains a dense sequence: We next show that there is a sequence $\{x_n\}_{n\in\mathbb{N}}$ that is *dense* in K; i.e., that for every $\epsilon > 0$, and every $x \in K$, there is some n such that $d(x_n, x) < \epsilon$.

In other words, the sequence $\{x_n\}_{n\in\mathbb{N}}$ passes arbitrarily close to every point in K. Here is how to construct it:

Pick the first term x_1 arbitrarily. We then define the rest of the sequence recursively as follows: Suppose that $\{x_1, \dots, x_k\}$ have been chosen. For each $y \in K$, define

$$d_k(y) := \min_{1\le j\le k} \{d(y, x_j)\} \ .$$

This is, by definition, the distance from y to the set $\{x_1, \dots, x_k\} \subset K$, and of course, this is no greater than the diameter of K, which is finite by the first step.

Therefore, d_k, defined by

$$d_k := \sup_{y\in K} d_k(y)$$

is no greater than the diameter of K.

Armed with this knowledge, we are ready to choose x_{k+1}: We choose x_{k+1} to be any element of K with

$$d_k(x_{k+1}) \ge \frac{1}{2} d_k \ .$$

We now claim that $\lim_{k\to\infty} d_k = 0$. It should be clear that $\{x_n\}_{n\in\mathbb{N}}$ is dense if and only if this is the case. So, to complete Step 2, we need to prove that $\lim_{k\to\infty} d_k = 0$.

Towards this end, the first thing to observe is that $\{d_k\}_{k\in\mathbb{N}}$ is a monotone decreasing sequence, bounded below by zero: Indeed, for any sets $A \subset B \subset K$, the distance from y to B is no greater than the distance from y to A. Therefore, we only have to show that *some subsequence* of $\{d_k\}_{k\in\mathbb{N}}$ converges to zero.

To do this, let $\{x_{k_n}\}_{n\in\mathbb{N}}$ be a convergent subsequence of $\{x_k\}_{k\in\mathbb{N}}$, and let y be the limit; i.e.,

$$\lim_{n\to\infty} x_{k_n} = y \ .$$

Then of course since by the triangle inequality

$$d(x_{k_n}, x_{k_{n+1}}) \le d(x_{k_n}, y) + d(y, x_{k_{n+1}}) \ ,$$

and since $\lim_{n\to\infty} d(x_{k_n}, y) = \lim_{n\to\infty} d(y, x_{k_{n+1}}) = 0$,

$$\lim_{n\to\infty} d(x_{k_n}, x_{k_{n+1}}) = 0 \ .$$

But since

$$x_{k_n} \in \{x_1, \dots, x_{k_{n+1}-1}\} \ ,$$

$$d(x_{k_n}, x_{k_{n+1}}) \geq d_{k_{n+1}-1}(x_{k_{n+1}}) \geq \frac{1}{2} d_{k_{n+1}-1} \ .$$

Therefore,

$$\lim_{n\to\infty} d_{k_{n+1}-1} = 0 \ ,$$

and then, since the entire sequence is monotone decreasing, $\lim_{k\to\infty} d_k = 0$. Hence, the sequence we have constructed is dense.

Step 3: Given any open cover of K, there exists a countable subcover. To prove this, consider any open cover $\mathcal{G}$ of K. Consider the set of open balls $B_r(x_k)$ where $r > 0$ is rational, and $\{x_k\}_{k\in\mathbb{N}}$ is the dense sequence that we have constructed in Step 2. This set of balls is countable since a countable union of countable sets is countable.

The countable subcover is constructed as follows: For each rational $r > 0$ and each $k \in \mathbb{N}$, choose $U_{r,k}$ to be *some* open set in $\mathcal{G}$ that contains $B_r(x_k)$ if there is such a set, and otherwise, do not define $U_{r,k}$. Let $\mathcal{U}$ be the set of open sets defined in this way; clearly $\mathcal{U}$ is countable by construction.

We now claim that $\mathcal{U}$ is an open cover of K. Clearly the sets in $\mathcal{U}$ are open. To see that they cover, pick any $x \in K$. Since $\mathcal{G}$ is an open cover of K, $x \in V$ for some $V \in \mathcal{G}$. Then, since V is open, for some rational $r > 0$, $B_{2r}(x) \subset V$.

Then, since $\{x_k\}_{k\in\mathbb{N}}$ is dense, there is some k with $x_k \in B_r(x)$. But then $x \in B_r(x_k)$ and

$$B_r(x_k) \subset B_{2r}(x) \subset V \ ,$$

(where the first containment holds by the triangle inequality). This shows that for the pair (r, k), there is *some* $V \in \mathcal{G}$ containing $B_r(x_k)$. Therefore, by construction, $U_{r,k} \in \mathcal{U}$ contains $B_r(x_k)$, and hence $x \in U_{r,k}$. Since x is an arbitrary element of K, $\mathcal{U}$ covers K.

Step 4: Some finite subcover of the countable cover is a cover. Now order the sets in our countable cover $\mathcal{U}$ into a sequence of open sets $\{U_k\}_{k\in\mathbb{N}}$ that covers K.

Suppose that for each n, it is *not* the case that

$$K \subset \bigcup_{k=1}^{n} U_k \ . \tag{1.3.6}$$

Then we can construct a sequence $\{x_n\}_{n\in\mathbb{N}}$ be choosing $x_n \in K \backslash \left(\bigcup_{k=1}^{n} U_k\right)$.

Let $\{x_{n_j}\}_{j\in\mathbb{N}}$ be a subsequence with $\lim_{j\to\infty} x_{n_j} = y \in K$. Then, since $\mathcal{U}$ is an open cover of K, there is some U_k with $y \in U_k$. But then all but finitely many terms of the sequence $\{x_{n_j}\}_{j\in\mathbb{N}}$ lie in U_k, and so the whole sequence lies in some finite union of the sets in $\mathcal{U}$. This is a contradiction, and so (1.3.6) is true for some $n \in \mathbb{N}$. □

Part of the proof made use of a dense sequence in our sequentially compact metric space. The existence of a dense sequence is often useful, and so we make the following definition.

1.3.21 DEFINITION (Separable topological space)**.** A topological space $(X, \mathcal{O})$ is *separable* in case it contains a countable dense subset.

We have seen the a compact metric space is always separable, but also many non-compact spaces are separable. We shall se examples shortly.

We shall soon prove two powerful theorems on approximation and compactness in an infinite dimensional vector space, $\mathcal{C}(X,\mathbb{R})$, the space of real valued continuous functions on a compact Hausdorff space X, equipped with the uniform metric

$$d_\infty(f,g) = \sup_{x\in X}\{|f(x)-g(x)|\}\ .$$

Note that by Theorem 1.3.16 and the fact that X is compact, there exists an $x_0 \in X$ such that

$$|f(x_0)-g(x_0)| = \sup_{x\in X}\{|f(x)-g(x)|\}\ .$$

It is then very easy to see that d_∞ is indeed a metric on $\mathcal{C}(X,\mathbb{R})$, called the *uniform metric*. It is left as an exercise to generalize Theorem 1.2.8 and show that $(\mathcal{C}(X,\mathbb{R}), d_\infty)$ is complete.

1.3.3 Generated topologies

When working a class of functions $\mathcal{F}$ on a set X with values in a topological space $(Y,\mathcal{U})$, it is often useful to introduce a topology that makes every function $f \in \mathcal{F}$ continuous, but which contains the minimal number of open sets for this purpose. given two topologies $\mathcal{O}_1$ and $\mathcal{O}_2$ on a sets X, we say that $\mathcal{O}_1$ is weaker than $\mathcal{O}_2$ in case $\mathcal{O}_1 \subset \mathcal{O}_2$. In the case at hand, there is a unique *weakest possible* topology $\mathcal{O}_\mathcal{F}$ on X with respect to which each $f \in \mathcal{F}$ is continuous, and such that $\mathcal{O}_\mathcal{F}$ is weaker than any other topology with this property.

The weaker a topology is, the more compact sets there will be, and so such topologies are useful when we wish to apply theorems requiring both continuity and compactness.

1.3.22 THEOREM (Generated topologies)**.** *Given a class of functions $\mathcal{F}$ on a set X with values in a topological space $(Y,\mathcal{U})$, define*

$$\mathcal{E} = \{f^{-1}(U)\ :\ f\in\mathcal{F}\ ,\ U\in\mathcal{U}\ \},$$

and define $\widehat{\mathcal{E}}$ to be the set of all finite intersections of sets in $\mathcal{E}$. Finally define $\mathcal{O}_\mathcal{F}$ to be the set of arbitrary unions of sets in $\widehat{\mathcal{E}}$. Then $\mathcal{O}_\mathcal{F}$ is a topology, and it is the weakest topology with respect to which each $f \in \mathcal{F}$ is continuous.

Proof. Since each $x \in X$ lies in $f^{-1}(Y) \in \mathcal{E}$ for any $f \in \mathcal{F}$, it is clear that $X \in \mathcal{O}_\mathcal{F}$, and taking the empty union, $\emptyset \in \mathcal{O}_\mathcal{F}$. Evidently $\mathcal{O}_\mathcal{F}$ is closed under arbitrary unions. Thus, if $\mathcal{O}_\mathcal{F}$ is closed under finite intersections, is is a topology. Let $E,F \in \mathcal{O}_\mathcal{F}$. If $x \in E\cap F$, then, by definition, there exist E_x, F_x in $\widehat{\mathcal{E}}$ such that $E_x \subset E$ and $F_x \subset F$. Since $\widehat{\mathcal{E}}$ is closed under finite intersections, $x \in E_x \cap F_x \in \widehat{\mathcal{E}}$, and $E_x\cap F_x \subset E\cap F$. Making such a construction for each $x \in E\cap F$, we have

$$E\cap F = \bigcup_{x\in E\cap F} E_x\cap F_x\ ,$$

showing that $E\cap F \in \mathcal{O}_\mathcal{F}$. Thus, $\mathcal{O}_\mathcal{F}$ is a topology.

Now let $\mathcal{O}$ be any topology with respect to which each $f \in \mathcal{F}$ is continuous. Evidently, $\mathcal{O}$ must contain all of the sets

$$\mathcal{E} = \{f^{-1}(U)\ :\ f\in\mathcal{F}\ ,\ U\in\mathcal{U}\ \}.$$

Certainly also any such topology must contain, $\widehat{\mathcal{E}}$, the set of all finite intersections of sets in $\mathcal{E}$, and then it must contain all arbitrary unions of sets in $\widehat{\mathcal{E}}$. Hence $\mathcal{O}$ must contain $\mathcal{O}_\mathcal{F}$. □

1.3.23 DEFINITION (Weak topology)**.** Given a set X and a family $\mathcal{F}$ of functions from X to a topological space $(Y, \mathcal{O})$, the *weak topology on X generated by $\mathcal{F}$.*

1.4 Some frequently used theorems

1.4.1 Baire's Theorem

1.4.1 DEFINITION (Nowhere dense)**.** Let $(X, \mathcal{O})$ be a topological space. *A subset$X$* is *nowhere dense* in case the closure of A has empty interior; i.e., $(\overline{A})^\circ = \emptyset$.

Note that if A is closed, it is nowhere dense if and only if $A^\circ = \emptyset$, which is the case iff and only if $A^c \cap U \neq \emptyset$ for all $U \in \mathcal{O}$. Therefore, a closed set A is nowhere densre if and only if its complement is an open dense set in X.

1.4.2 LEMMA (Baire's Lemma)**.** *Let (X, d) be a complete metric space. Let $\{U_n\}_{n\in\mathbb{N}}$ be a sequence of open dense sets in X. Then $\bigcap_{n\in\mathbb{N}} U_n$ is dense in X.*

Proof. Let W be any open subset of X. It suffices to show that

$$\bigcap_{n\in N} U_n \cap W \neq \emptyset \ . \tag{1.4.1}$$

Now construct a Cauchy seequence as follows: Since U_1 is open and dense, and W is open, $U_1 \cap W$ is open and non-empty. Hence for some $r_1 > 0$ and some $x_1 \in X$, $\overline{B_{r_1}(x_1)} \subset U_1 \cap W$. (Note that if $B_{s_1}(x_1) \subset U_1 \cap W$, then for any $r_1 < s_1$, $\overline{B_{r_1}(x_1)} \subset U_1 \cap W$ because $\overline{B_{r_1}(x_1) \subset B_{s_1}(x_1)}$.) Without loss of generality, we may suppose that $r_1 < 1$. Since U_2 is open and dense and $B_{r_1}(x_0)$, is open, there exist $r_2 > 0$ and some $x_2 \in X$ so that $B_{r_2}(x_2) \subset B_{r_1}(x_1) \cap U_2$. Without loss of generality, we may suppose that $r_2 < r_1/2$.

Now supposing that $\{(r_1, x_1), \dots (r_n, x_n)\}$ are chosen, pick (r_{n+1}, x_{n+1}) so that

$$\overline{B_{r_{n+1}}(x_{n+1})} \subset B_{r_n}(x_n) \cap U_{n+1} \qquad \text{and} \qquad r_{n+1} < r_n/2 \ .$$

By contruction

$$\overline{B_{r_n}(x_n)} \subset B_{r_m}(x_m) \subset W \quad \text{for all} \quad n > m \tag{1.4.2}$$

and

$$\overline{B_{r_n}(x_n)} \subset \bigcap_{m=1}^{n} U_m \cap W \ . \tag{1.4.3}$$

The sequence $\{x_n\}_{n\in\mathbb{N}}$ is Cauchy since (1.4.2) says that for all $j, k \geq m$, $d(x_j, x_k) < r_m$ and $\lim_{m\to\infty} r_m = 0$. Since (X, d) is complete, there exists $x_0 \in X$ such that $\lim_{n\to\infty} x_n = x_0$. By (1.4.2) again, for all $n \geq m$, $x_n \in \overline{B_{b_m}(x_m)}$, and hence $x_0 \in \overline{B_{b_m}(x_m)}$ for all m. By (1.4.3), $x_0 \in \bigcap_{n=1}^{\infty} U_n \cap W$. ∎

1.4.3 THEOREM (Baire's Theorem)**.** *A complete metric space (X, d) is never the countable union of nowhere dense sets.*

Proof. Let $\{E_m\}_{n\in\mathbb{N}}$ be a sequence of nowhere dense sets in X. It suffice to show $\bigcup_{n=1}^{\infty} \overline{E_n} \neq X$. By Baire's Lemma,

$$\left(\bigcup_{n=1}^{\infty} \overline{E_n}\right)^c = \bigcap_{n=1}^{\infty} (\overline{E_n})^c \neq \emptyset$$

since each $(\overline{E_n})^c$ is open and dense. □

Notice that to prove Baire's Theorem, we only needed to know that the inetrsection of a sequence of open dense sets is non-empty, while Baire's lemma tells us that not only is $\cap_{n=1}^{\infty} U_n$ non-empty, it is even dense. This stronger information is sometimes useful in applications.

In fact, in a metric space (X,d) that has no isoolated points, $\cap_{n=1}^{\infty} U_n$ is uncountable. To see this, suppose on the contrary that $\cap_{n=1}^{\infty} U_n$ is countable. Let $\{x_n\}_{n\in\mathbb{N}}$ be a sequence such that $\{x_n \ : \ n \in \mathbb{N}\} = \cap_{n=1}^{\infty} U_n$. (There will be repeats in the sequence if the latter set is finite.) Define $E_n = U_n^c \cup \{x_n\}$. Then each E_n is closed and has empty interior (since X has no isolated points). Then

$$\bigcup_{n=1}^{\infty} E_n = \left(\bigcup_{n=1}^{\infty} U_n^c\right) \cup \left(\bigcap_{n=1}^{\infty} U_n\right) = \left(\bigcap_{n=1}^{\infty} U_n\right)^c \cup \left(\bigcap_{n=1}^{\infty} U_n\right) = X \ .$$

By Baire's Theorem, this is impossible.

1.4.2 The Arzelà-Ascoli Theorem

Let X be a compact topological space, and consider the metric space, and hence topological space, consisting of $\mathcal{C}(X,\mathbb{R})$ equipped with the uniform metric.

1.4.4 DEFINITION (Equicontinuous, pointwise bounded)**.** Let $\mathcal{F} \subset \mathcal{C}(X,\mathbb{R})$. Then $\mathcal{F}$ is *equicontinuous* in case for each $\epsilon > 0$ and each $x \in X$, there is a neighborhood U_ϵ of x such that for all $f \in \mathcal{F}$,

$$y \in U_\epsilon \Rightarrow |f(y) - f(x)| < \epsilon \ .$$

Also, $\mathcal{F}$ is *pointwise bounded* in case for each $x \in X$, $\{f(x) \ : f \in \mathcal{F} \ \}$ is a bounded subset of $\mathbb{R}$.

The first thing to observe is that if $\mathcal{F}$ is a compact subset of $\mathcal{C}(X,\mathbb{R})$, then $\mathcal{F}$ is both pointwise bounded and equicontinuous.

Indeed, suppose that $\mathcal{F}$ is compact. Then evidently $\mathcal{F} \subset \bigcup_{f\in\mathcal{F}} B_f(1)$, and hence there exists a finite set $\{f_1,\dots,f_n\}$ in $\mathcal{F}$ such that

$$\mathcal{F} \subset \bigcup_{j=1}^{n} B_{f_j}(1) \ .$$

Let M_j denote the maximum of $|f_j|$, which is continuous on X, and hence bounded. Let $M = \max_{j=1,\dots,n} M_j$. Then for any $f \in \mathcal{F}$, there is some j such that $d_\infty(f,f_j) < 1$, and hence for all x,

$$|f(x)| \leq |f_j(x)| + |f(x) - f_j(x)| \leq M_j + d_\infty(f,f_j) \leq M + 1 \ .$$

This even shows that $\mathcal{F}$ uniformly bounded.

To show that $\mathcal{F}$ is equicontinuous, fix $x \in X$ and $\epsilon > 0$. For each $f \in \mathcal{F}$, define V_f by

$$V_f = \{ \ g \in \mathcal{F} \ : \ d(g,f) < \epsilon/3 \ \} \ .$$

Clearly, each V_f is open, and $\displaystyle\bigcup_{f\in\mathcal{F}} V_f = \mathcal{F}$. Hence there exists a finite set $\{f_1,\ldots,f_n\} \subset \mathcal{F}$ such that

$$\bigcup_{j=1}^{n} V_{f_j} = \mathcal{F} \ .$$

Next, define $U_j \subset X$ by $U_j = \{\ y \in X\ :\ |f_j(y) - f_j(x)| < \epsilon/3\}$. Evidently $U = \cap_{j=1}^{n} U_j$ is a neighborhood of x.

For any $f \in \mathcal{F}$, $d_\infty(f, f_j) < \epsilon/3$ for some j, and so for $y \in U$,

$$|f(y) - f(x)| \le |f(y) - f_j(y)| + |f_j(y) - f_j(x)| + |f_j(x) - f(x)| \le |f_j(y) - f_j(x)| + 2d_\infty(f, f_j) < \epsilon \ .$$

Thus, U is a neighborhood of x such that $|f(y) - f(x)| < \epsilon$ for all $y \in U$ and all $f \in \mathcal{F}$.

Thus, a necessary condition for $\mathcal{F}$ to be compact in $\mathcal{C}(X,\mathbb{R})$ is that $\mathcal{F}$ be equicontinuous and pointwise bounded. The Arzelà-Ascoli Theorem says that these conditions are essentially sufficient as well:

1.4.5 THEOREM (Arzelà-Ascoli)**.** *Let X be a compact topological space space, and let $\mathcal{F}$ be an equicontinuous and pointwise bounded subset of $\mathcal{C}(X,\mathbb{R})$. Then the closure of $\mathcal{F}$ is compact.*

Proof. The first thing to observe is that if $\mathcal{F}$ is equicontinuous and pointwise bounded, then so is the closure of $\mathcal{F}$. Hence, let us assume that $\mathcal{F}$ is closed as well as equicontinuous and pointwise bounded. We shall then show that $\mathcal{F}$ is compact.

By Theorem 1.3.20, it suffices to show that for any infinite sequence $\{f_\ell\}_{\ell\in\mathbb{N}}$ in $\mathcal{F}$, there is a convergent subsequence, and then by the completeness of $\mathcal{C}(X,\mathbb{R})$ it suffices to show that for any infinite sequence $\{f_\ell\}_{\ell\in\mathbb{N}}$ in $\mathcal{F}$, there is a Cauchy subsequence. Therefore, fix any infinite sequence $\{f_\ell\}_{\ell\in\mathbb{N}}$ in $\mathcal{F}$. We must prove that there is a subsequence $\{f_{\ell_j}\}_{j\in\mathbb{N}}$ such that for all $\epsilon > 0$, there is an $N_\epsilon \in \mathbb{N}$ such that

$$j,k \ge N_\epsilon \Rightarrow |f_{\ell_j}(x) - f_{\ell_k}(x)| < \epsilon \quad \text{for all } x \in X. \tag{1.4.4}$$

Fix $\epsilon > 0$. Use the compactness of X and the equicontinuity of $\mathcal{F}$ to select a finite set of points $\{x_1,\ldots,x_m\}$ and neighborhoods $\{U_1,\ldots,U_m\}$ that cover $\mathcal{F}$ and are such that

$$x \in U_j \Rightarrow |f(x) - f(x_j)| < \frac{\epsilon}{3} \qquad \text{for all } f \in \mathcal{F} \ . \tag{1.4.5}$$

Since for each i, the set $\{f(x_i)\ :\ f \in \mathcal{F}\ \}$ is a bounded subset of $\mathbb{R}$, we can choose a subsequence of $\{f_\ell\}_{\ell\in\mathbb{N}}$ along which $f_{\ell_k}(x_i)$ converges for each i. Since convergent sequences are Cauchy, it follows that there exists $N_\epsilon \in \mathbb{N}$ such that

$$j,k \ge N_\epsilon \Rightarrow |f_{\ell_j}(x_i) - f_{\ell_k}(x_i)| \le \frac{\epsilon}{3} \ . \tag{1.4.6}$$

But then since each $x \in X$ belongs to U_i for some i, we have (for this i), and $j,k \ge N_\epsilon$,

$$\begin{aligned}
|f_{\ell_j}(x) - f_{\ell_k}(x)| &\le |f_{\ell_j}(x) - f_{\ell_k}(x_i)| + |f_{\ell_j}(x_i) - f_{\ell_k}(x_i)| + |f_{\ell_j}(x_i) - f_{\ell_k}(x)| \\
&\le \frac{\epsilon}{3} + |f_{\ell_j}(x_i) - f_{\ell_k}(x_i)| + \frac{\epsilon}{3} \\
&\le \epsilon \ .
\end{aligned}$$

where the first inequality is the triangle inequality, the second is (1.4.5) and the third is (1.4.6). This proves (1.4.4). $\square$

Another proof of the Arzela-Ascoli Theorem may be given using the fact that complete totally bounded subsets of a metric space are compact. This approach is useful in proving other compactness theorems, so we briefly explain it as well. The key is the following lemma of Hanche-Olsen and Holden:

1.4.6 LEMMA. *Let (X, d_X) be a metric space. Suppose that for all $\epsilon > 0$ there exists a $\delta > 0$ and a metric space (W, d_W) and a function $\Phi : X \to W$ such that:*

(1) $\Phi(X)$ is totally bounded in W.

(2) For all $x, y \in X$,

$$d_W(\Phi(x), \Phi(y)) < \delta \Rightarrow d_X(x, y) < \epsilon \ .$$

Then X is totally bounded.

1.4.7 Remark. If $\Phi : X \to W$ were invertible, then *(2)* could be rewritten as: For all $w, z \in W$,

$$d_W(w, z) < \delta \Rightarrow d_X(\Phi^{-1}(w), \Phi^{-1}(z)) < \epsilon \ ,$$

which would mean that Φ^{-1} is continuous. However, we do not assume that Φ is invertible. Nonetheless, even when Φ is not invertible, and A is any subset of W, we write $\Phi^{-1}(A)$ to denote the preimage of A under Φ, i.e., $\Phi^{-1}(A) = \{x \in X \ : \ \Phi(x) \in A\}$. Condition *(2)* then says that the diameter of the preimage of a set of diameter less than δ is less than ϵ.

Proof. Fix $\epsilon > 0$, Let δ, (W, d_W) and Φ be such that *(1)* and *(2)* are satisfied. Since $\Phi(X)$ is totally bounded, there is a finite cover $\{U_1, \dots, U_n\}$ of $\Phi(X)$ be balls of radius δ in W. It follows immediately that $\{\Phi^{-1}(U_1), \dots \Phi^{-1}(U_n)\}$ is a cover of X by sets of diameter 2ϵ. For each $i = 1, \dots, n$, pick $x_i \in \Phi^{-1}(U_i)$. Then $\Phi^{-1}(U_i) \subset B_{2\epsilon}(x_i)$, and so $\{B_{2\epsilon}(x_1), \dots, B_{2\epsilon}(x_n)\}$ is a finite cover of X by balls of radius 2ϵ . Since $\epsilon > 0$ is arbitrary, (X, d) is totally bounded. □

Second proof of the Arzela-Ascoli Theorem. We may suppose as before that $\mathcal{F}$ is closed and hence that $(\mathcal{F}, d_\infty)$ is a complete metric space. It therefore suffices to show that it is totally bounded. Starting as before, fix $\epsilon > 0$, and use the compactness of X and the equicontinuity of $\mathcal{F}$ to select a finite set of points $\{x_1, \dots, x_m\}$ and neighborhoods $\{U_1, \dots, U_m\}$ that cover $\mathcal{F}$ and are such that

$$x \in U_j \Rightarrow |f(x) - f(x_j)| < \frac{\epsilon}{3} \qquad \text{for all } f \in \mathcal{F} \ . \tag{1.4.7}$$

Define $\Phi : \mathcal{F} \to \mathbb{R}^m$ by

$$\Phi(f) = (f(x_1), \dots, f(x_m)) \ .$$

Since $\mathcal{F}$ is pointwise bounded, $\Phi(\mathcal{F})$ lies in some bounded rectangle in $\mathbb{R}^m$, and bounded sets in $\mathbb{R}^m$ are totally bounded. Thus, Φ satisfies condition *(1)* of Lemma 1.4.6. Next, consider any $f, g \in \mathcal{F}$. Denoting the Euclidean norm on $\mathbb{R}^m$ by $\|\cdot\|$,

$$\|\Phi(f) - \Phi(g)\| = \|(f(x_1), \dots, f(x_m)) - (g(x_1), \dots, g(x_m))\| \geq \max_{j=1}^{m} |f(x_j) - g(x_j)| \ . \tag{1.4.8}$$

fix any $x \in X$. Then for some j, $x \in U_j$, and for this j,

$$\begin{aligned} |f(x) - g(x)| &\leq |f(x) - f(x_j)| + |f(x_j) - g(x_j)| + |g(x_j) - g(x)| \\ &\leq \frac{\epsilon}{3} + |f(x_j) - g(x_j)| + \frac{\epsilon}{3} \\ &\leq \frac{\epsilon}{3} + \|\Phi(f) - \Phi(g)\| + \frac{\epsilon}{3} \end{aligned}$$

where the first inequality is the triangle inequality, the second is (1.4.7) and the third is (1.4.8). Thus, condition *(2)* of Lemma 1.4.6 is satisfied for $\delta = \epsilon/3$, and we have proved that $\mathcal{F}$ is totally bounded. $\square$

1.4.3 The Stone-Wierstrass Theorem

The Stone-Wierstrass Theorem is an approximation theorem that generalizes the classical Wierstrass Approximation Theorem that we discussed at the beginning of these notes.

We begin with two definitions. Let X be a compact topological space, and let $\mathcal{C}(X,\mathbb{R})$ be the space of continuous real valued functions on X equipped with the uniform metric d_∞. A subset $\mathcal{A}$ of $\mathcal{C}(X,\mathbb{R})$ is an *algebra* in case $\mathcal{A}$ is a vector subspace over $\mathbb{R}$ of $\mathcal{C}(X,\mathbb{R})$ equipped with its usual rules of addition and scalar multiplication, and if, moreover, for every f and g in $\mathcal{A}$, the pointwise product fg also belongs to $\mathcal{A}$.

A subset $\mathcal{A}$ of $\mathcal{C}(X,\mathbb{R})$ is *separating* in case for pair of distinct points x, y in X, there is an $f \in \mathcal{A}$ such that $f(x) \neq f(y)$.

Notice that if X is not Hausdorff, not even $\mathcal{C}(X,\mathbb{R})$ the space of *all* continuous real valued functions on X is separating. Indeed, if X is not Hausdorff, there exist two distinct points x and y in X such that every neighborhood U of x contains y. But then for any continuous function f, $f(x) = f(y)$. Indeed, if $|f(x)-f(y)| := r > 0$, then $f^{-1}((f(x)-r.2, f(x)+r/2))$ would be an open neighborhood of x that excluded y. Thus, for all continuous f, $f(x) = f(y)$, so not even $\mathcal{C}(X,\mathbb{R})$ separates, let alone any proper subset of $\mathcal{C}(X,\mathbb{R})$. Hence throughout this subsection, we shall only be concerned with Hausdorff topological spaces.

The primary example of a separating algebra to keep in mind is $X = [0,1]$, with $\mathcal{A}$ being the algebra of all polynomials in the real variable $x \in [0,1]$. To see that this algebra is separating, consider the polynomial $p(x) = x$. Then for $x_0 \neq x_1$ in X, $p(x_0) \neq p(x_1)$.

1.4.8 THEOREM (Stone-Wierstrass)**.** *Let X be a compact topological space, and let $\mathcal{A}$ be a subset of $\mathcal{C}(X,\mathbb{R})$ that is a separating algebra. Let $\mathcal{B}$ be the uniform closure of $\mathcal{A}$. Then either $\mathcal{B} = \mathcal{C}(X,\mathbb{R})$, or else $\mathcal{B}$ consists of all continuous functions on X that vanish at some fixed point x_0. In particular, if $\mathcal{A}$ contains the constant functions, $\mathcal{B} = \mathcal{C}(X,\mathbb{R})$.*

We will prove Theorem 1.4.8 as a consequence of two lemmas, and shall make use of the *partial order* in $\mathcal{C}(X,\mathbb{R})$: If $f, g \in \mathcal{C}(X,\mathbb{R})$, we write $f \leq g$ in case $f(x) \leq g(x)$ for all $x \in X$. With this partial order, $\mathcal{C}(X,\mathbb{R})$ is a *lattice*: Given any $f, g \in \mathcal{C}(X,\mathbb{R})$ there is a unique function $g \wedge f \in \mathcal{C}(X,\mathbb{R})$ such that $g \wedge f \leq f, g$, and such that $h \leq g \wedge f$ whenever $h \leq f, g$. Of course, $g \wedge f$ is given by

$$g \wedge f(x) = \min\{\ f(x)\ ,\ g(x)\ \}\ ,$$

which is continuous.

Likewise, given any $f, g \in \mathcal{C}(X,\mathbb{R})$ there is a unique function $g \vee f \in \mathcal{C}(X,\mathbb{R})$ such that $f, g \leq g \vee f$, and such that $g \vee f \leq h$ whenever $f, g \leq h$. Of course, $g \vee f$ is given by

$$g \vee f(x) = \max\{\ f(x)\ ,\ g(x)\ \}\ ,$$

which is continuous.

A subset $\mathcal{F}$ of $\mathcal{C}(X,\mathbb{R})$ is itself a lattice if and only if whenever $f, g \in \mathcal{F}$, then both $f \wedge g$ and $f \vee g$ belong to $\mathcal{F}$. Then observing that

$$f \wedge g = \frac{1}{2}(f + g - |f - g|) \qquad \text{and} \qquad f \vee g = \frac{1}{2}(f + g + |f - g|) \ , \tag{1.4.9}$$

we see that a subset $\mathcal{F}$ of $\mathcal{C}(X,\mathbb{R})$ that is a *vector space* is a lattice if and only if whenever $f \in \mathcal{F}$, then $|f| \in \mathcal{F}$.

1.4.9 LEMMA (Limit point criterion for lattices in $\mathcal{C}(X,\mathbb{R})$)**.** *Let X be a compact Hausdorff space. Let $\mathcal{F} \subset \mathcal{C}(X,\mathbb{R})$ be a lattice.*

If f is any element of $\mathcal{C}(X,\mathbb{R})$ with the property that for every $x, y \in X$, there exists a function $f_{x,y} \in \mathcal{F}$ for which

$$f_{x,y}(x) = f(x) \qquad \text{and} \qquad f_{x,y}(y) = f(y) \ . \tag{1.4.10}$$

Then f is a limit point of $\mathcal{F}$; i.e., it belongs to the closure of $\mathcal{F}$.

Proof. Fix any $f \in \mathcal{C}(X,\mathbb{R})$ with the property every $x, y \in X$, there exists a function $f_{x,y} \in \mathcal{F}$ such that (1.4.10) is satisfied. Fix any $\epsilon > 0$. We must show that there exists some $g \in \mathcal{F}$ with $|g(x) - f(x)| < \epsilon$ for all $x \in X$.

First, for each $(x, y) \in X \times X$, make some choice of $f_{x,y}$, and define the open set $U_{x,y} \subset X$ by

$$U_{x,y} = \{\ z \ : \ f_{x,y}(z) < f(z) + \epsilon\ \} \ .$$

Evidently, $x, y \in U_{x,y}$. Therefore

$$X = \bigcup_{x \in X} U_{x,y} \ ,$$

and then, since X is compact, there exists a finite set $\{x_1, \ldots, x_n\} \subset X$ such that

$$X = \bigcup_{j=1}^{n} U_{x_j,y} \ .$$

Now define the function f_y by

$$f_y = f_{x_1,y} \wedge f_{x_2,y} \wedge \cdots \wedge f_{x_n,y} \ .$$

Since $\mathcal{F}$ is a lattice, $f_y \in \mathcal{F}$, and

$$f_y \le f + \epsilon$$

in the lattice order; i.e., everywhere on X.

Furthermore, since $f_{x_j,y}(y) = f(y)$ for each j, $f_y(y) = f(y)$. Therefore, defining the open set V_y by

$$V_y := \{\ z \in X \ : \ f(z) - \epsilon < f_y(z)\ \} \ ,$$

we have $y \in V_y$, and hence

$$X = \bigcup_{y \in X} V_y \ ,$$

hen, since X is compact, there exists a finite set $\{y_1, \ldots, y_m\} \subset X$ such that

$$X = \bigcup_{k=1}^{m} V_{y_k} \ .$$

Now define g by

$$g = f_{y_1} \vee f_{y_2} \vee \ldots, \vee f_{y_m} \ .$$

Then since $\mathcal{F}$ is a lattice, $g \in \mathcal{F}$, and by construction,

$$f - \epsilon \leq g \leq f + \epsilon \ ,$$

which means that $|f(x) - g(x)| < \epsilon$ for all $x \in X$. □

1.4.10 LEMMA (A closed algebra in $\mathcal{C}(X, \mathbb{R})$ is a lattice)**.** *Let X be a compact Hausdorff space. Let $\mathcal{B}$ be a closed subset of $\mathcal{C}(X, \mathbb{R})$ that is also a subalgebra of $\mathcal{C}(X, \mathbb{R})$. Then $\mathcal{B}$ is a lattice.*

Proof. By the remarks we have made concerning (1.4.9), it suffices to show that for all $f \in \mathcal{B}$, $|f| \in \mathcal{B}$. Since X is compact and f is continuous, f is bounded above and below, and hence there is a finite positive number c such that $|cf| \leq 1$. Then since $|cf| = c|f|$, we may freely suppose that $f| \leq 1$.

Therefore, fix any $f \in \mathcal{B}$ with $|f| \leq 1$, We shall complete the proof by showing that there exists a sequence of polynomials $\{p_n\}_{n \in \mathbb{N}}$ so that

$$|f| = \lim_{n \to \infty} p_n(f^2) \tag{1.4.11}$$

in the uniform topology. Since $\mathcal{B}$ is an algebra, $p_n(f^2) \in \mathcal{B}$ for each n, and then since $\mathcal{B}$ is closed, $|f| \in \mathcal{B}$.

For any number $a \in [0, 1]$, we define a sequence $\{b_n\}_{n \in \mathbb{N}}$ recursively as follows: We set $b_1 = 0$ and then for all $n \in \mathbb{N}$,

$$b_{n+1} = b_n + \frac{a - b_n^2}{2} \ .$$

Notice that

$$b_1 = 0 \ , \quad b_2 = \frac{a}{2} \ , \quad b_3 = a - \frac{a^2}{8} \ ,$$

and so forth. It is easy to see by induction that for each n, there is a polynomial p_n, independent of the value of a, so that such that $b_n = p_n(a)$.

We claim that $\sqrt{a} = \lim_{n \to \infty} b_n$. This will give us a sequence of polynomials $\{p_n\}_{n \in \mathbb{N}}$ such that for each $a \in [0, 1]$,

$$\sqrt{a} = \lim_{n \to \infty} p_n(a) \ ,$$

and therefore, such that

$$|f(x)| = \lim_{n \to \infty} p_n(f^2(x))$$

for all x in X. Then, since X is compact, Dini's Theorem implies that (1.4.11) is true with uniform convergence.

Hence, we need only verify the claim that $\sqrt{a} = \lim_{n \to \infty} b_n$. To do this, note that

$$\sqrt{a} - b_{n+1} = \sqrt{a} - b_n - \frac{(\sqrt{a} - b_n)(\sqrt{a} + b_n)}{2} = (\sqrt{a} - b_n)\left(1 - \frac{\sqrt{a} + b_n}{2}\right) \ .$$

Since $a \leq 1$, as long as $b_n \leq \sqrt{a}$, the right hand side is non-negative, and therefore $b_{n+1} \leq \sqrt{a}$. Since $b_1 \leq \sqrt{a}$, it follows that $\sqrt{a}$ is an upper bound for the sequence $\{b_n\}_{n\in\mathbb{N}}$.

Now, knowing that $b_n^2 \leq a$ for all n, it is clear from the definition that $\{b_n\}_{n\in\mathbb{N}}$ is a monotone non-decreasing sequence. Therefore the limit $b = \lim_{n\to\infty} b_n$ exists and satisfies

$$b = b + \frac{a - b^2}{2} \ .$$

This means that $b^2 = a$, and since $b \geq 0$, $b = \sqrt{a}$. □

Proof of Theorem 1.4.8: Fix $x \neq y$ in X, and consider the linear transformation from $\mathcal{A}$ to $\mathbb{R}^2$ given by

$$f \mapsto (f(x), f(y)) \ .$$

The range of this linear transformation is a subspace S of $\mathbb{R}^2$.

Since $\mathcal{A}$ separates, there can be at most one point $x_0 \in X$ for which $g(x_0) = 0$ for all $g \in \mathcal{A}$.

Let us first assume first that neither x nor y is such a point. Since $\mathcal{A}$ is an algebra, and a vector space in particular, if g is in $\mathcal{A}$ so is very multiple of g. By assumption, there is some $g \in \mathcal{A}$ such that $g(x) \neq 0$, and by choosing an appropriate multiple, we may arrange that $g(x) = 1$.

Thus, S contains a vector of the form $(1, a)$. (Since $\mathcal{A}$ separates, we can choose $g \in \mathcal{A}$ so that $g(y) = a \neq 1$.)

Now there are two cases to consider. If also $a \neq 0$, then the two vectors $(1, a)$ and $(1, a^2)$ are linearly independent, and $(1, a^2)$ also belongs to S since $\mathcal{A}$ is an algebra (so that $g^2 \in \mathcal{A}$). On the other hand if $a = 0$ then S contains the vector $(1, 0)$, and, since there is some other g with $g(y) = 1$, there is some $b \in \mathbb{R}$ such that $(b, 1) \in S$. Hence in this case, S contains the two vectors $(1, 0)$ and $(b, 1)$ which are linearly independent. Either way, $S = \mathbb{R}^2$, and so we have proved that as long as $g(x) \neq 0$ and $h(y) \neq 0$ for *some* $g, h \in \mathcal{A}$, then S is all of $\mathbb{R}^2$.

This has the consequence that for *any* $f \in \mathcal{C}(X, \mathbb{R})$, we can find a function $f_{x,y} \in \mathcal{A}$ for which

$$(f(x), f(y)) = (f_{x,y}(x), f_{x,y}(y)) \ . \tag{1.4.12}$$

Now we have two cases once more: Suppose first that there is no point $x_0 \in X$ with $f(x_0) = 0$ for all $f \in \mathcal{A}$. Then the above argument applies for *all* x and y in X and all $f \in \mathcal{C}(X, \mathbb{R})$, we can find $f_{x,y} \in \mathcal{A}$ such that (1.4.12) is true. Moreover, by Lemma 1.4.10, $\mathcal{B}$ is a lattice. Therefore, by Lemma 1.4.9, f is a limit point of $\mathcal{B}$, and since $\mathcal{B}$ is closed, $f \in \mathcal{B}$. Since f is an arbitrary element of $\mathcal{C}(X, \mathbb{R})$, we see that in this case, $\mathcal{B} = \mathcal{C}(X, \mathbb{R})$.

The remaining case to consider is that in which there is one point x_0 such that $g(x_0) = 0$ for all $g \in \mathcal{A}$, and hence $\mathcal{B}$, so that $\mathcal{B}$ is certainly contained in the closed subset of $\mathcal{C}(X, \mathbb{R})$ consisting of continuous functions f on X such that $f(x_0) = 0$.

Let f be any such function. The argument made above show that as long as neither x nor y equals x_0, then there is some $g \in \mathcal{A}$, and hence $\mathcal{B}$, for which (1.4.12) is true. Now suppose that $x = x_0$, and $y \neq x_0$. Then we trivially have

$$f(x_0) = g(x_0) = 0$$

for all $g \in \mathcal{B}$. And since $\mathcal{A}$ separates, and is a vector space, we can choose g so that $f(y) = g(y)$. Therefore, for any $f \in \mathcal{C}(X, \mathbb{R})$ with $f(x_0) = 0$, no matter how x and y are chosen, we can can find $g_{x,y} \in \mathcal{B}$ so that (1.4.12) is true.

Then the argument made above shows that every $f \in \mathcal{C}(X, \mathbb{R})$ with $f(x_0) = 0$ is a limit point of $\mathcal{B}$, and hence belongs to $\mathcal{B}$. Therefore, in this second case, $\mathcal{B}$ is the subset of $\mathcal{C}(X, \mathbb{R})$ consisting of functions f with $f(x_0) = 0$. □

In our proof of Theorem 1.4.8, we made use of the fact that our functions f were real valued, and not complex valued: The real numbers are ordered, while the complex numbers are not, and the order on the real numbers played a crucial role in the proof through our use of Lemma 1.4.9.

This is not simply an artifact of the proof: If in the statement of the theorem we replace $\mathcal{C}(X, \mathbb{R})$ by, $\mathcal{C}(X, \mathbb{C})$, the space of continuous complex valued functions on X, the statement becomes false.

To see this, take X to be the closed unit disc in the complex plane $\mathbb{C}$. Take $\mathcal{A}$ to be the algebra of all complex polynomials in the complex variable z, which clearly separates. Polynomials in z are analytic, and uniform limits of analytic functions are analytic, and so the closure of $\mathcal{A}$ consists of functions that are analytic in the interior of the the unit disc. Obviously, not every continuous function of the closed unit disc is analytic in the interior of the disc; $f(z) = z^*$, the complex conjugate of z, is an example. Hence, the uniform closure of $\mathcal{A}$ is not the full set of continuous complex valued functions on the closed unit disc.

However, under one simple additional condition on the algebra $\mathcal{A}$, one can reduce the complex valued case to the real case.

A (complex) subalgebra $\mathcal{A}$ of the algebra of complex valued function on a compact Hausdorff space is called a *$*$-algebra* in case it is closed under complex conjugation. That is, whenever $f \in \mathcal{A}$, then $f^* \in \mathcal{A}$, where f^* is the function defined by $f^*(x) = (f(x))^*$ for all $x \in X$.

In this case, for every $f \in \mathcal{A}$, the real and imaginary parts of f both belong to $\mathcal{A}$. It is also easy to see that when $\mathcal{A}$ separates, so does the real algebra consisting of the real and imaginary parts of functions in $\mathcal{A}$. Applying the Stone-Wierstrass Theorem to this algebra, one can separately approximate, in the uniform metric, the real and imaginary parts of any continuous complex valued function on X by functions in $\mathcal{A}$.

In summary, we have:

1.4.11 THEOREM (Complex Stone-Wierstrass). *Let X be a compact topological space, and let $\mathcal{A}$ be a subset of $\mathcal{C}(X, \mathbb{C})$ that is a separating $*$-algebra. Let $\mathcal{B}$ be the uniform closure of $\mathcal{A}$. Then either $\mathcal{B} = \mathcal{C}(X, \mathbb{C})$, or else $\mathcal{B}$ consists of all continuous functions on X that vanish at some fixed point x_0. In particular, if $\mathcal{A}$ contains the constant functions, $\mathcal{B} = \mathcal{C}(X, \mathbb{R})$.*

Here is one important application of Theorem 1.4.11: Let X be the unit circle in $\mathbb{C}$, with its usual topology. Let $\mathcal{A} \subset \mathcal{C}(x, \mathbb{C})$ be the set consisting of functions f of the form

$$f(z) = \sum_{j=-n}^{n} a_j z^n$$

for some $n \in \mathbb{N}$, and some numbers $a_{-n}, \ldots, a_n$ in $\mathbb{C}$. (Each element of X is a complex number z, and z^n denotes the nth power of z.) The elements of $\mathcal{A}$ are called *complex trigonometric polynomials*

It is easy to see that $\mathcal{A}$ is a $*$-algebra, and that $\mathcal{A}$ separates. Hence, by Theorem 1.4.11, $\mathcal{A}$ is dense in $\mathcal{C}(X, \mathbb{C})$. This proves:

1.4.12 THEOREM (Density of Complex Trigonometric Polynomials). *Let X be the unit circle in $\mathbb{C}$, with its usual topology. Then the set of complex trigonometric polynomials is dense in $\mathcal{C}(X, \mathbb{C})$, with respect to the uniform metric.*

1.4.4 Tychonoff's Theorem

Let X be a set. The *Cartesian product of X with itself*, $X \times X$, is the set of all ordered pairs (x_1, x_2) of elements of X. Of course (x_1, x_2) is the graph of a unique function $f : \{1, 2\} \to X$, namely the one with $f(1) = x_1$ and $f(2) = x_2$. (One can accommodate Cartesian products of two different sets Y and Z in this framework by considering $X = Y \cup Z$ and restricting attention to functions f such that $f(1) \in Y$ and $f(2) \in Z$. No real generality is lost in taking the sets to be the same, and the notation is much simpler, so that is how we shall proceed.)

More generally, given any set non-empty S, the Cartesian product of X indexed by S, denotes X^S, is the set of all functions from S to X. For example, $X^{\mathbb{N}}$ is the set of all infinite sequences $\{x_n\}_{n\in\mathbb{N}}$ of elements of X.

On any Cartesian product, there is a natural family of functions with values in X, namely the *coordinate functions*: For each $s \in S$, define

$$\varphi_s : X^S \to X$$

by

$$\varphi(f) = f(s) \ .$$

That is, one simply evaluates the function $f \in X^S$ at s.

Note that when $S = \{1, 2\}$, $\varphi_j((x_1, x_2)) = x_j$, which is why the φ_s are called coordinate functions.

1.4.13 DEFINITION. Let $(X, \mathcal{O})$ be a topological space, and S and arbitrary set. The product topology on the Cartesian product X^S is the topology generated by the coordinate functions. That is, it is the weakest topology for which each of the coordinate functions is continuous.

Now suppose that $(X, \mathcal{O})$ is a compact topological space. When is $(X, \mathcal{O})$ compact in the product topology? The answer, given by Tychonov's theorem is: "Always."

1.4.14 THEOREM (Tychonoff's Theorem). *Let $(X, \mathcal{O})$ be a compact topological space, and S any non-empty set. Then X^S, equipped with the product topology, is compact.*

The special case of this theorem in which X is a compact metric space and S is countable (or finite) is fairly easy to prove using the theorems presented so far in these notes. This is developed in the exercises that follow. The general case involves either the theory of "nets" or the theory of "filters", and this would be a digression, since we shall not invoke the general case in this course, nor shall we have any other occasion to use the theory of nest of filters. Furthermore, the proof of the general case involves the axiom of choice in a much more subtle way than does the spacial case. This is not a problem, but discussion of these subtleties would take us far afield. (The axiom of choice enters the subject, even in the special case, in an essential way: It is the axion of choice which assures us that X^S is non-empty: one can always choose, for each $x \in s$, some $x(s) \in X$. Moreover, it is known that Tychanov's Theorem is logically equivalent to the Axiom of Choice.)

It is well worth knowing the general case nonetheless. It shows that advantage of the 20th century notion of *compactness*, as defined above, in terms of open covers, and the 19th century notion of sequential compactness. As shown in the exercises, if we take $X = [0,1]$ with its usual topology, and equip X^X, the set of all functions from $[0,1]$ to $[0,1]$, then X^X is not sequentially compact, but is compact *by Tychonov's Theorem*. Many theorems in which compactness is an hypothesis remain true if this hypothesis is replaced by sequential compactness (see the exercises). Tychonov's Theorem is an important example for which this is not the case.

1.5 Exercises

1. Let (X,d) be a separable metric space. Let Y be any subset of X, and define d_Y to be the restriction of d to $Y \times Y$. Show that (Y, d_Y) is separable.

2. Suppose that (X,d) is a complete metric space with a finite diameter; i.e., there exists $D < \infty$ such that $d(x,y) \leq D$ for all $x, y \in X$. Is it true that every continuous real valued function on X is bounded? Prove this assertion or give a counterexample.

3. Let (X,d) be a compact metric space.

(a) Show that if $f : X \to X$ is continuous but not onto, there is some $x_0 \in X$ and some $r > 0$ so that $d(f(x), x_0) \geq r$ for all $x \in X$.

(b) Let f be an isometry from X into itself; i.e., a function with the property that

$$d(f(x), f(y)) = d(x,y)$$

for all $x, y \in X$. Show that f is necessarily one to one and onto, and hence invertible.

4. Let (X,d) be a compact metric space, and let $f : X \to \mathbb{C}$ be continuous. Show that for all $\epsilon > 0$, there exists $L < \infty$ so that

$$|f(x) - f(y)| \leq Ld(x,y) + \epsilon$$

for all $x, y \in X$.

5. (a) Let (X,d) be a complete metric space in which bounded sets are totally bounded. Let $A \subset X$ be closed and $B \subset X$ be compact. Show that there exist $x_1 \in A$ and $x_2 \in B$ such that

$$d(x_1, x_2) \leq d(x,y) \qquad \text{for all } x \in A \ , y \in B \ .$$

(b) Show by example that this is false if we weaken the assumption to only suppose that B is closed.

6. Define ℓ_1 to be the set of complex valued sequences $\{x_j\}_{j\in\mathbb{N}}$ such that $\displaystyle\sum_{j=1}^{\infty} |x_j| < \infty$. Define a function on d_{ℓ_1} on ℓ_2 by

$$d_{\ell_1}\left(\{x_j\}\ ,\ \{y_j\}\right) = \sum_{j=1}^{\infty} |x_j - y_j| \ .$$

(a) Show that (ℓ_1, d_{ℓ_1}) is a metric space.

(b) Show that the metric space (ℓ_1, d_{ℓ_1}) is complete.

7. Let (ℓ_2, d_{ℓ_2}) Show that a bounded subset X of ℓ_2 is totally bounded if and only if for all $\epsilon > 0$, there exists $N_\epsilon \in \mathbb{N}$ such that

$$\sum_{k>N_\epsilon} |x_j|^2 < \epsilon^2$$

for all $\{x_j\} \in X$.

8. Let (ℓ_1, d_{ℓ_1}) be defined as in Exercise 6. Show that $X \subset \ell_1$ is totally bounded if and only if for all $\epsilon > 0$, there exists $N_\epsilon \in \mathbb{N}$ such that

$$\sum_{k>N_\epsilon} |x_j| < \epsilon$$

for all $\{x_j\} \in X$. Then show that $B_1(\{0\})$, the ball of radius 1 about the zero sequence, is not totally bounded, and hence that the closed ball of radius 1 about the zero sequence is not compact.

9. Let $X = [0,1]$. Each $x \in X$ has a binary expansion

$$x = \sum_{n=1}^{\infty} b_n(x) 2^{-n}$$

with each $b_n(x) \in \{0,1\}$. We stipulate that if x is a dyadic rational, only finitely many of the $b_n(x)$ are non-zero, and under this condition, the $b_n(x)$ are uniquely determined, so that $b_x : X \to \{0,1\} \subset X$ is a well-defined function for each n.

(a) Show that no subsequence of $\{b_n\}_{n\in\mathbb{N}}$ converges pointwise.

(b) Equip X^X with its product topology and note that each b_n is a function from X to X, and hence is an element of X^X. Show that no subsequence of $\{b_n\}_{n\in\mathbb{N}}$ converges in the product topology, and thus that the analog of Tychonov's Theorem for sequential compactness is false.

10. Let (X,d) be a compact metric space. Then $X^{\mathbb{N}}$ consists of all sequences $\{x_k\}_{k\in\mathbb{N}}$ in X. Define a function d on $X^{\mathbb{N}} \times X^{\mathbb{N}}$ by

$$d(\{x_k\}_{k\in\mathbb{N}}, \{y_k\}_{k\in\mathbb{N}}) = \sum_{k=1}^{\infty} 2^{-k} d(x_k, y_k) \ .$$

(a) Show that d is a metric on $X^{\mathbb{N}} \times X^{\mathbb{N}}$.

(b) Show that the metric topology in $X^{\mathbb{N}}$ induced by d is at least as strong as the product topology.

(c) Show that with the metric topology induced by d, $X^{\mathbb{N}}$ is sequentially compact.

(d) Show directly, without invoking Tychonov's Theorem that $X^{\mathbb{N}}$ compact in the product topology.

11. Let (X, d_X) and (Y, d_Y) be two compact metric spaces. Let $\mathcal{C}(X \times Y, \mathbb{R})$ be the set of all valued functions on $X \times Y$ continuous real that are continuous with respect to the product topology. Let $\mathcal{A}$ be the set of functions f on $X \times Y$ of the form

$$f(x,y) = \sum_{j=1}^{n} g_j(x) h_j(y)$$

for some $n \in \mathbb{N}$ and some $\{g_1, \ldots, g_n\} \subset \mathcal{C}(X, \mathbb{R})$ and some $\{h_1, \ldots, h_n\} \subset \mathcal{C}(Y, \mathbb{R})$. Show that $\mathcal{A}$ is dense in $\mathcal{C}(X \times Y, \mathbb{R})$ in the uniform topology. (Note: The usual notation for $\mathcal{A}$ is $\mathcal{C}(X, \mathbb{R}) \otimes \mathcal{C}(Y, \mathbb{R})$, and it is called the tensor product of $\mathcal{C}(X, \mathbb{R})$ and $\mathcal{C}(Y, \mathbb{R})$.)

12. A topological space is *locally compact* in case every point has a neighborhood whose closure is compact. Let (X, d_X) and (Y, d_Y) be locally compact metric spaces, and suppose that $f : X \to Y$ is continuous and bijective. Show that f^{-1} is continuous if and only if $f^{-1}(K)$ is compact for all compact $K \subset Y$.

13. Let $(X, \mathcal{O})$ be a compact topological space. Let A and B be non-empty closed and disjoint subsets of X. Suppose that for every $b \in B$, there exist a continuous function $f_b : X \to [0,1]$ such that $f_b(b) = 1$ and $f_b(a) = 0$ for all $a \in A$. Show that there exist open sets U and V such that $A \subset U$, $B \subset V$ and $U \cap V = \emptyset$.

14. Let $(X, \mathcal{O})$ be a compact topological space, and let $\mathcal{F}$ be a set of functions real valued on X that is equicontinuous and uniformly bounded. Define

$$g(x) = \sup_{f \in \mathcal{F}} f(x) \ .$$

Is $g(x)$ necessarily continuous? Prove that your answer is correct.

15. Let (X, d_X) and (Y, d_Y) be metric spaces with Y complete. For $L \in (0, \infty)$, a function $f : X \to Y$ is L-Lipschitz in case $d_Y(f(x), f(y)) \leq L d_X(x, y)$ for all $x, y \in X$.

Let S be a dense subset of X. Let $g : S \to Y$ satisfy

$$d_Y(g(x), g(y)) \leq L d_X(x, y) \quad \text{for all} \quad x, y \in S \ .$$

Show that there exists a unique L-Lipschitz function $f : X \to Y$ such that the restriction of f to S is g.

16. Let $\{f_n\}_{n \in \mathbb{N}}$ be a sequence of continuous real valued functions on $[0,1]$ that are continuously differentiable on $(0,1)$. Suppose that $f_n(0) = 0$ for all n and that there is a continuous function $g : [0,1] \to [0, \infty)$ such that $|f_n'(x)| \leq g(x)$ for all $n \in \mathbb{N}$ and all $x \in (0,1)$. Show that there exists a uniformly convergent subsequence $\{f_{n_k}\}_{k \in \mathbb{N}}$.

17. Let (X, d_X) and (Y, d_Y) be two metric spaces. Let $f : X \to Y$ be continuous uand surjective, and supose that

$$d_X(x_1, x_2) \leq d_Y(f(x_1), f(x_2))$$

for all $x_1, x_2 \in X$.

(a) If (X, d_X) complete, must (Y, d_Y) be complete? Prove this or give a counterexample.

(b) If (Y, d_Y) complete, must (X, d_X) be complete? Prove this or give a counterexample.

18. Let (X, d) be a metric space. If every real-vauled conttinuous function f on X has a maximum, does this mean that X is compact? Prove your answer is correct.

19. Let (X, d) be a compact metric space. Ptove that if $\{U_1, \ldots, U_k\}$ is an open cover of X, there there exists a closed cover $\{C_1, \ldots, C_k\}$ with $C_j \subset U_j$ for $j = 1, \ldots, k$.

20. Let A and B be compact subsets of a Hausdorff topological space. Prove that there exist open sets U and V such that $A \subset U$, $B \subset V$ and $U \cap V = \emptyset$.

21. Let (X, d_X) and (Y, d_Y) be metric spaces and let $F : X \to Y$ be a continuous functions such that f maps closed sets to closed sets and such that the inverse image of any point in Y is compact. Show that $f^{-‘}(K)$ is compact whenever K is compact.

22. Ler $\{f_n\}_{n \in \mathbb{N}}$ be a sequence of uniformly continuous functions from $\mathbb{R}$ to $\mathbb{R}$. Suppose that f_n converges uniformly to f. Must f be uniformly continuous?

Chapter 2

Topological Vector Spaces

2.1 Topological Vector Spaces

2.1.1 Neighborhood bases for topological vector spaces

2.1.1 DEFINITION (Topological Vector Space)**.** A *topological vector space* is a vector space X over $\mathbb{C}$ that is equipped with a topology of open sets $\mathscr{O}$ such that the maps $(\alpha, x) \mapsto \alpha x$ and $(x, y) \mapsto x + y$ are continuous on $\mathbb{C} \times X$ and $X \times X$ respectively. A *real topological vector space* is defined in the same way except that the field $\mathbb{C}$ replaced by $\mathbb{R}$.

Let X be a vector space. For all $x \in X$ and $A, B \subset X$, define

$$x + B := \{x + y \ :\ y \in B\} \quad \text{and} \quad A + B := \{x + y \ :\ x \in A, y \in B\} \ . \tag{2.1.1}$$

Then $A + B$ is said to be the *Minkowski sum of* A *and* B. For $x \in X$, let T_x denote the map from X to X given by $T_x(y) = x + y$. Then $T_x^{-1} = T_{-x}$, and T_x is vector space isomorphism on X. The maps T_x, $x \in X$, are called *translations*.

Likewise, for $\alpha \in \mathbb{C}\backslash\{0\}$ let S_α denote the map form X to X given by $x \mapsto \alpha x$. Since $S_\alpha^{-1} = S_{\alpha^{-1}}$, each S_α is a vector space isomorphism on X. The maps S_α, $\alpha \in \mathbb{C}\backslash\{0\}$ are called *scale transformations*. For all $\alpha \in \mathbb{C}$ and all $A \subset X$, define

$$\alpha A := \{\alpha x \ :\ x \in A\} \ . \tag{2.1.2}$$

Whenever $(X, \mathscr{O})$ is a topological vector space, both T_x and T_x^{-1} are continuous. Hence each $T_x, x \in X$, is a homeomorphism on X. Likewise, each S_α, $\alpha \in \mathbb{C}\backslash\{0\}$, is a homeomorphism on X. It follows that $U \subset X$ is open if and only if $x + U = T_x(U)$ is open for each $x \in X$. If U is any non-empty set, and $-x \in U$, then $0 \in x + U = T_x(U)$. Therefore, *the sets in* $\mathscr{O}$ *are precisely the translates of the sets in* $\mathscr{O}$ *that contain* 0. Likewise, $U \subset X$ is open if and only $S_\alpha(U) \in \mathscr{O}$ for all $\alpha \in \mathbb{C}\backslash\{0\}$.

2.1.2 DEFINITION (Neigborhood base at 0)**.** Let $(X, \mathscr{O})$ be a topological vector space. A *neigborhood base at* 0 for the topology $\mathscr{O}$ is a set $\mathscr{V} \subset \mathscr{O}$ such that $0 \in V$ for all $V \in \mathscr{V}$, and if $0 \in U \in \mathscr{O}$, there exists some $V \in \mathscr{V}$ such that $V \subset U$. A topologocal vector space is *locally convex* in case it has a neigborhood base at 0 consisting of convex sets.

2.1.3 LEMMA. *$(X,\mathscr{O})$ be a topological vector space, and let $\mathscr{V}$ be a neighborhood base at 0 for the topology $\mathscr{O}$. Then a non-empty $U \subset X$ is open if and only if for each $x \in U$, there is some $V \in \mathscr{V}$ such that $x + V \subset U$.*

Proof. Suppose that for each $x \in U$, there is some $V_x \in \mathscr{V}$ such that $x + V_x \subset U$. Then each $x + V_x$ is open and writing $U = \bigcup\{x + V_x \ : x \in U\}$ displays U as a union of open sets. Hence $U \in \mathscr{O}$.

Conversely, suppose that $U \in \mathscr{O}$. Then for each $x \in U$, $-x + U$ is an open set containing 0. Hence for some $V_x \in \mathscr{V}$, $V_x \subset -x + U$. But then $x + V_x \subset U$. □

It is a simple matter to construct topologies $\mathscr{O}$ on a vector space X such that each translation map T_x is a homeomphism on $(X,\mathscr{O})$

2.1.4 LEMMA. *Let X be a vector space, and let $\mathscr{W}$ be a family of subsets of X that is closed under finite intersections.*

Define $\mathscr{O}$ to be the set of subsets U of X given by

$$\mathscr{O} = \emptyset \cup \{U \subset X \ : \ \text{for all } x \in U \text{ there exists } W_x \in \mathscr{W} \text{ such that } x + W_x \subset U\} \ . \tag{2.1.3}$$

Then $\mathscr{O}$ is a topology on X, and under this topology, each of the maps T_x, $x \in X$, is continuous.

Proof. Evidently, $\emptyset \in \mathscr{O}$, and evidently an arbitrary union of sets U such that for each $x \in U$, there is some $W_x \in \mathscr{W}$ such that $x + W_x \subset U$ also has this same property. Hence $\mathscr{O}$ is closed under arbitrary unions.

To see that $\mathscr{O}$ is closed under finite intersections, let $\{U_1, \dots, U_n\} \subset \mathscr{O}$, and let $U = \cap_{j=1}^n U_j$. By definition, for each $x \in U$, $x \in U_j$, $j = 1, \dots, n$, and hence there is some $W_{x,j}$ such that $x + W_{x,j} \subset U_j$. But then

$$x + \bigcap_{j=1}^n W_{x,j} \subset U \ ,$$

and by the closure under finite intersections, $\cap_{j=1}^n W_{x,j} \in \mathscr{W}$, and hence $U \in \mathscr{O}$.

For the final statement, since $T_x^{-1} = T_{-x}$, it suffices to show that each T_x is open. By definition, the general open set U has the form $U = \cup\{y + W_y \ : y \in U\}$ for some set $\{W_y \ : \ y \in U\} \subset \mathscr{W}$. Bu then $T_x(U) = \cup\{x + y + W_y \ : y \in U\}$ which is open. □

The condition that $(x,y) \mapsto x+y$ is continuous on $X \times X$ in the product topology is *stronger* than the condition that each of the maps T_x are continuous. The latter condition amounts to *separate continuity* of the map $(x,y) \mapsto x+y$, and for $(X,\mathscr{O})$ to be a topological vector space, we require *joint continuity.* To achieve this, and to ensure that the topology $\mathscr{O}$ provided by Lemma 2.1.4 has the other properties that would make $(X,\mathscr{O})$ is a topological vector space, and moreover is is Hausdorff, we must impose further conditions on the sets in $\mathscr{W}$. We close this section with some useful results that make direct use of the joint continuity of the algebraic operations.

2.1.5 DEFINITION. Let X be a vector space over $\mathbb{C}$. a set $A \subset X$ is *absorbing* if for all $x \in X$, there is a $\delta_x > 0$ so that for $t\ in(-\delta_x, \delta_x)$, $tx \in A$, or, what is the same thing, that $x \in tA$ for all $t > 1/\delta_x$. A set $A \subset X$ is *balanced* in case for all $\alpha \in \mathbb{C}$ with $|\alpha| = 1$, $\alpha A \subset A$. (Hence every balanced set contains 0.)

2.1.6 LEMMA. *In any topological vector space, every neighborhood V of 0 is absorbing.*

Proof. Let $x \in X$, and let V be a neighborhood of 0. Since $0x = 0$, the continuity of scalar multiplication implies that there exists $r > 0$ so that $|\alpha| < r \Rightarrow \alpha x \in V$. (For this proof, even separate continuity would suffice.) $\square$

In a topological vector space, not every neighborhood of 0 need be balanced, but there will always be a nieighborhood base consisting of balanced sets.

2.1.7 LEMMA. *Let $(X, \mathscr{O})$ be a topological vector space. Every neighborhood U of 0 contains a balanced neighborhood V of 0, and thus that there exists a neighborhood base for $\mathscr{O}$ consisting of balanced sets.*

Proof. By the joint continuity of scalar multiplication, there is an open set W containing 0 and an $r > 0$ so that $\alpha w \in U$ for all $|\alpha| < r$ and all $w \in W$. That is, defining $D := \{\alpha \ : \ |\alpha| < r\}$, $DW \subset U$. Since $DW = \bigcup_{\alpha \in D} \alpha W$ is a union of open sets, $V := DW$ is open and evidently balanced. $\square$

2.1.8 LEMMA. *Let $(X, \mathscr{O})$ be a topological vector space. For every neighborhood U of 0, there is a balanced open set G (necessarily a neighborhood of 0) such that $G + G \subset U$.*

Proof. By the joint continuity of vector addition, for $x, y \in X$, and for every neighborhood of U of 0, there exist neighborhoods V, W of 0 such that $(x + V) + (y + W) \subset (x + y + U)$. Let G be a balanced neighborhood of 0 contained in $V \cap W$, which exists by Lemma 2.1.7. Then $(x+G)+(y+G) \subset (x+y+U)$. Specializing to $x = y = 0$, we have that $G + G \subset U$. $\square$

Lemmas 2.1.7 and 2.1.8 made use of the joint continuity of the algebraic operations in their proofs. The next two theorems are consequences of these lemmas.

2.1.9 THEOREM. *Let $(X, \mathscr{O})$ be a topological vector space. Then X is Hausdorff if and only if $\{0\}$ is closed in X.*

Proof. If X is Hausdorff, then for each $x \neq 0$, there is a open neighborhood V_x of x such that V_x^c contains a neighborhood of 0. Choosing such a set for each $x \neq 0$, $\{0\}^c = \bigcup_{x \neq 0} \{V_x\}$, which is open.

Conversely, suppose that $\{0\}$ is closed in X. Let $z \neq 0$, and let U be a neighborhood of 0 such that $0 \notin z + U$. Then there is a balanced neighborhood G of 0 such that $0 \notin z + G - G$. Now let $x, y \in X$ with $x \neq y$, and put $z = x - y$. Then for all $g_1, g_2 \in G$, so that $-g_2$ also belongs to G since G is balanced, $0 \notin z + G - G$ means that $0 \neq x - y + g_1 - g_2$, and hence $y + g_2 \neq x + g_1$, which means that $(y + G) \cap (x + G) = \emptyset$. $\square$

2.1.10 THEOREM. *Let $(X, \mathscr{O})$ be a topological vector space. Let $K \subset X$ be compact, and $F \subset X$ be closed, and suppose that $K \cap F = \emptyset$. Then there exists a neighborhood V of 0 such that $(K + V) \cap F = \emptyset$*

Proof. Let $x \in K$. Then $x \notin F$, and since F is closed, there is a neighborhood U of 0 such that $(x + U) \cap F = \emptyset$. By Lemma 2.1.8, for each $x \in K$, there exists a balanced neighborhood G_x of 0 such that $(x + G_x + G_x) \cap F = \emptyset$. Then $\{x + G_x\}_{x \in K}$ is an open cover of K. Since K is compact, there exists a finite set $\{x_1, \dots, x_n\}$ such that

$$K \subset \bigcup_{j=1}^{n} (x_j + G_{x_j}) \ .$$

Set $W = \bigcup_{j=1}^{n} G_{x_j}$, which is an open neighbrood of 0.

For all $x \in K$, there exists $j \in \{1, \dots, n\}$ such that $x = x_j + g$ for some $g \in G_{x_j}$. For all $w \in W$, $x + w = x_j + g + w \in x_j + G_{x_j} + G_{x_j}$. Since $(x_j + G_{x_j} + G_{x_j}) \cap F = \emptyset$ $(x + W) \cap F = \emptyset$. $\square$

2.1.2 Generating locally convex topologies

2.1.11 THEOREM. *Let X be a vector space, and let $\mathscr{W}$ be a family of subsets of X such that that is closed under finite intersections. Suppose further that: each $W \in \mathscr{W}$ is absorbing, convex and balanced, and that for all $W \in \mathscr{W}$, $\frac{1}{2}W$ belongs to $\mathscr{W}$. Let $\mathscr{O}$ be defined by*

$$\mathscr{O} = \emptyset \cup \{U \subset X \ : \ \text{for all } x \in U \text{ there exists } W_x \in \mathscr{W} \text{ such that } x + W_x \subset U\} \ .$$

Then $(X, \mathscr{O})$ is a topological vector space, and $\mathscr{W}$ is a neighborhood base at 0 for this topology. Moreover, if for all $x \neq 0$, there exists $W \in \mathscr{W}$ such that $x \notin W$, then $(X, \mathscr{O})$ is a Hausdorff topological vector space.

Proof. By Lemma 2.1.4, $\mathscr{O}$ is a topology on X under which translation is a homeomorphism.

We now show that the map $(\alpha, x) \mapsto \alpha x$ is jointly continuous on $(\mathbb{C}\backslash\{0\}) \times (X, \mathscr{O})$. Let $\alpha_0 \in \mathbb{C}\backslash\{0\}$ and $x \in X$. It suffices to show that if $U \in \mathscr{O}$ and $\alpha_0 x \in U$, then there is a $\delta > 0$ and a $\widetilde{W} \in \mathscr{W}$ such $D(x + \widetilde{W}) \subset U$ where $D := \{\alpha \in \mathbb{C} \ : \ |\alpha - \alpha_0| < \delta\}$. Since $\alpha_0 x \in U \in \mathscr{O}$, there is some $W \in \mathscr{W}$ such that $\alpha_0 x + W \subset U$. Then

$$\alpha x + \frac{1}{2}W = (\alpha - \alpha_0)x + (\alpha_0 x + \frac{1}{2}W) \ .$$

Since W is absorbing and balanced, for δ sufficiently small, if $|\alpha - \alpha_0| < \delta$, then $(\alpha - \alpha_0)x \in \frac{1}{2}W$. Then since W is convex, $\alpha x + \frac{1}{2}W \subset \frac{1}{2}W + (\alpha_0 x + \frac{1}{2}W) \subset \alpha_0 x + W \subset U$. Taking $\widetilde{W} = \frac{1}{2}W$, we have what we sought for the scalar multiplication.

We now show that map $(x, y) \mapsto x + y$ is jointly continuous on $X \times X$. Let $x, y \in X$, and let $U \in \mathscr{O}$ be such that $x + y \in U$. Then there exists $W \in \mathscr{W}$ such that $x + y + W \subset U$. But then $(x + \frac{1}{2}W) + (y + \frac{1}{2}W) \subset U$.

Finally, by Lemma 2.1.9, $(X, \mathscr{O})$ is Hausdorff in case $\{0\}$ is closed. For a balanced neighborhood of 0, $x \not\in W$ is the same as $0 \notin x + W$. For each $x \neq 0$, choose $W_x \in \mathscr{W}$ such that $0 \notin W_x$. Then $\{0\}^u = \bigcup_{x \neq 0} W_x$ is open. □

Let X be a vector space and let $\mathscr{V}$ be a set of of absorbing, convex and balanced subsets of X. Since any finite intersection of absorbing, convex and balanced sets is again absorbing, convex and balanced, if we define $\mathscr{W}$ to be the set of all finite intersections of sets in $\mathscr{V}$, then by Theorem 2.1.11, $\mathscr{W}$ is a neighborhood base at 0 of a uniquely determined topology $\mathscr{O}$ on X such that $(X, \mathscr{O})$ is a locally convex topological vector space.

The vector spaces topologies that we consider below are always generated this way, and we will often need to determine when two such topologies are comparable. The following lemma facilitates this.

2.1.12 LEMMA. *Let X be a vector space. Let $\mathscr{V}_1$ and $\mathscr{V}_2$ be two sets of absorbing, convex and balanced subsets of X, both also closed under multiplication by 2^{-k}, $k \in \mathbb{N}$. Let $\mathscr{W}_1$ and $\mathscr{W}_2$ be the sets of all finite intersections of sets in $\mathscr{V}_1$ and $\mathscr{V}_2$ respectively. Let $\mathscr{O}_1$ and $\mathscr{O}_2$ be the topological vector space topologies on X that have $\mathscr{W}_1$ and $\mathscr{W}_2$ respectively as neighborhood bases at 0. Then $\mathscr{O}_1 \subset \mathscr{O}_2$ if each $V_1 \in \mathscr{V}_1$ contains some $V_2 \in \mathscr{V}_2$.*

Proof. Let $U \in \mathscr{O}_1$. By definition, for each $x \in U$, there exists $W_x \in \mathscr{W}_1$ such that $x + W_x \subset U$. Suppose that each $W \in \mathscr{W}_1$ contains some $\widetilde{W} \in \mathscr{W}_2$. In particular, writing $U = \cup_{x \in U}\{x + W_x\}$, and letting $\widetilde{W}_x$ denote some element of $\mathscr{W}_2$ contained in W_x, we have $U = \cup_{x \in U}\{x + \widetilde{W}_x\}$. Thus $U \in \mathscr{O}_2$.

Therefore, it suffice to show that whenever each $V_1 \in \mathscr{V}_1$ contains some $V_2 \in \mathscr{V}_2$, then each $W_1 \in \mathscr{W}_1$ contains some $W_2 \in \mathscr{W}_2$.

Let $W_1 \in \mathscr{W}_1$. Then W_1 is a finite intersection of elements of $\mathscr{V}_1$. Each of these contains a set in $\mathscr{V}_2$. The intersection over the later sets belongs to $\mathscr{W}_2$, and is contained in $\mathscr{W}_1$. □

2.1.3 Finite dimensional subspaces

Let $(X, \mathscr{O})$ be a Hausdorff topological vector space. Let Y be a finite dimensional subspace of X. Let $\{y_1, \ldots, y_n\}$ be a basis for Y. Then associated to this basis, there is the natural linear map T from $\mathbb{C}^n$ into X and onto Y given by

$$T(\alpha_1, \ldots, \alpha_n) = \sum_{j=1}^{n} \alpha_j y_j \ , \tag{2.1.4}$$

which is evidently injective as well as surjective onto Y. By the joint continuity of scalar multiplication and vector additions, it is evident that T is continuous.

Let S be the unit sphere in $\mathbb{C}^n$. That is, S consists of all $(\alpha_1, \ldots, \alpha_n) \in \mathbb{C}^n$ such that $\sum_{j=1}^{n} |\alpha_j|^2 = 1$. Equip $\mathbb{C}^n$ with is usual topology. Then S is compact in $\mathbb{C}^n$, and hence $T(S)$ is compact in X. Since $(X, \mathscr{O})$ is Hausdorff, $T(S)$ is closed and since T is injective, $0 \notin T(S)$. Hence there is an open set V containing 0 that does not intersect $T(S)$, and by Lemma 2.1.7, we may take V to be balanced.

If $x \in T^{-1}(V)$, then $\|x\| \neq 1$ since $V \cap S = \emptyset$. Furthermore if $\|x\| > 1$, then for some $t \in (0, 1)$, $tx \in S$, and then $tT(X) = T(tx) \in T(S)$. this is impossible since V is balanced so that $tT(x) \in V$. Hence $T^{-1}(V)$ is contained in the open unit ball $B(1, 0)$ in $\mathbb{C}^n$. Therefore, $V \subset T(B(1, 0))$. By homogeneity, for all $r > 0$, $T(B(r, 0))$ contains a neighborhood of 0 in Y. This means that T is open, and hence T^{-1} is continuous.

This shows that every n dimensional subspaces Y of any Hausdorff topological vector space $(X, \mathscr{O})$ is naturally isomorphic and homeomorphic to $\mathbb{C}^n$ under the map T defined in (2.1.4) using any basis of Y. In particular, all Hausdorff topological vector spaces of the same finite dimension n are homomorphically isomorphic to one another.

Finally, let Y be any n-dimensional subspace of a Hausdorff topological vector space $(X, \mathscr{O})$, and let $x \in X$ but $x \notin Y$. Let $\{y_1, \ldots, y_n\}$ be a basis for Y, and note that $\{y_1, \ldots, y_n, x\}$ is linearly independent. Let $T : \mathbb{C}^{n+1} \to W$ be the homeomorphic isomorphism induced by this basis. Note that

$$Y = T^{-1}(\{(\alpha_1, \ldots, \alpha_n, 0) \ : \ (\alpha_1, \ldots, \alpha_n) \in \mathbb{C}^n\}) \ ,$$

and $x = T^{-1}(0, \ldots, 0, 1)$. Evidently, $T^{-1}(x) \notin T^{-1}(Y)$, which is a closed subspace of $\mathbb{C}^{n+1}$. Hence there is an neighborhood U of 0 in X such that $(x + U) \cap Y = ((x + U) \cap W) \cap Y = \emptyset$, showing that x is not in the closure of Y. Thus, Y is closed. We summarize the content of this section in a theorem:

2.1.13 THEOREM. *Let $(X, \mathscr{O})$ be a Hausdorff topological vector space. Let Y be a finite dimensional subspace of X. Then Y is closed, and if $\{y_1, \ldots, y_n\}$ is any basis for Y, the map T defined by (2.1.4) is a homeomorphic isomorphism of $\mathbb{C}^n$ onto Y.*

Now let $(X, \mathscr{O})$ be a topological vector space, not necessarily Hausdorff. Let Y be a subspace of X. Define an equivalence relation $\sim$ on X by $x_1 \sim x_2$ if and only if $x_1 - x_2 \in Y$. For $x \in X$, let $\{x\}_\sim$ denote the equivalence class of x. Let X/Y denote the set of equivalence classes in X. This is a vector space with the operations $\alpha\{x\}_\sim = \{\alpha x\}_\sim$ and $\{x\}_\sim + \{y\}_\sim = \{x+y\}_\sim$.

The *quotient topology* on X/Y is the topology $\mathscr{U}$ consisting of those subsets $U \subset X/Y$ such that the preimage of U under the map $\pi : x \mapsto \{x\}_\sim$ belongs to $\mathscr{O}$. Then π is continuous from $(X, \mathscr{O})$ to $(X/Y, \mathscr{U})$. It is easy to check that $(X/Y, \mathscr{U})$ is a topological vector space. Moreover, $\pi^{-1}(\{0\}_\sim) = Y$, and so the singleton consisting of $\{0\}_\sim$ alone is closed in X/Y if and only if Y is closed in X. Thus, by Lemma 2.1.9, X/Y us Hausdorff if and only if Y is closed.

Now let Z be a finite dimensional subspace of X, and Y a closed subspace of X. Then $\pi(Z)$ is a finite dimensional subspace of X/Y, which is Hausdorff, and then by Theorem 2.1.13, $\pi(Z)$ is a closed subspace of X/Y. Then $Y + Z = \pi^{-1}(\pi(Z))$ is closed in X. This proves:

2.1.14 THEOREM. *Let $(X, \mathscr{O})$ be a topological vector space, not necessarily Hausdorff. Let Y and Z be subspaces of X with Y closed and Z finite dimensional. Then $Y + Z$ is closed in X.*

2.1.4 Seminorms, norms and normed vector spaces

2.1.15 DEFINITION. Let X be a vector space over $\mathbb{C}$ or $\mathbb{R}$. A function $f : X \to \mathbb{R}$ is *convex* in case for all $\lambda \in (0,1)$ and all $x, y \in X$,

$$f((1-\lambda)x + \lambda y) \le (1-\lambda)f(x) + \lambda f(y) \ , \tag{2.1.5}$$

and is *subadditive* in case for all $x, y \in X$,

$$f(x+y) \le f(x) + f(y) \ , \tag{2.1.6}$$

and is *homogenous of degree one* in case for all $x \in X$ and $\alpha \in \mathbb{C}$

$$f(\alpha x) = |\alpha| f(x) \ . \tag{2.1.7}$$

A function $f : X \to \mathbb{R}$ is *positively homogenous of degree one* in case for all $t \in [0, \infty)$

$$f(tx) = tf(x) \ . \tag{2.1.8}$$

2.1.16 LEMMA. *Let X be a vector space over $\mathbb{C}$ or $\mathbb{R}$ and let $f : X \to \mathbb{R}$ be positively homogeneous of degree one. Then f is convex if and only if f is subadditive.*

Proof. Suppose that f is homogeneous of degree one and convex. Then

$$f(x+y) = 2f\left(\frac{x+y}{2}\right) \le 2\left(\frac{1}{2}f(x) + \frac{1}{2}f(y)\right) = f(x) + f(y) \ .$$

Conversely, Suppose that f is homogeneous of degree one and subadditive. Then for $\lambda \in (0,1)$

$$f((1-\lambda)x + \lambda y) \le f((1-\lambda)x) + f(\lambda y) = (1-\lambda)f(x) + \lambda f(x) \ .$$

□

2.1.17 DEFINITION (Seminorm and norm)**.** Let X be a vector space over $\mathbb{C}$ or $\mathbb{R}$. A *seminorm* on X is a function $p : X \to [0, \infty)$ such that p is homogeneous of degree one and convex, or, what is the same thing, homogeneous of degree one and subadditive. A seminorm p is a *norm* in case $p(x) = 0$ implies that $x = 0$. A different notation, namely $x \mapsto \|x\|$ is usually used for norm functions, also sometimes for seminorms.

2.1.18 LEMMA. *Let p be a seminorm on the vector space X. Define the set B_p by*

$$B_p := \{x \in X \ \ p(x) \le 1\} \ . \tag{2.1.9}$$

Then B_p is absorbing, balanced and convex. Moreover,

$$B_p = \cap_{r>1} r B_p \ . \tag{2.1.10}$$

Proof. For $x \in X$, $\lambda > 0$, $p(\lambda x) = \lambda p(x) < 1$, so that $\lambda x \in B_p$, for all $\lambda < 1/p(x)$. Therefore, B_p is absorbing. For $x \in B_p$, $\alpha \in \mathbb{C}$, $|\alpha| \le 1$, $p(\alpha x) = |\alpha| p(x) \le 1$, so that $\alpha x \in B_p$. Therefore, B_p is balanced. For all $x, y \in B_p$, and all $\lambda \in (0,1)$, $p((1-\lambda)x + \lambda y) \le (1-\lambda)p(x) + \lambda p(y) \le 1$. Therefore, B_p is convex.

To prove (2.1.10), note that $x \in rB_p$ if and only if $r^{-1}x \in B_p$ if and only if $p(r^{-1}x) \le 1$. By the homogeneity of p, $p(r^{-1}x) \le 1$ is the same as $p(x) \le r$. Hence $x \in \cap_{r>1} r B_p$ if and only if $p(x) \le r$ for all $r > 1$, and this means that $p(x) \le 1$. □

It turns out that there is a one-to-one correspondence between absorbing, balanced, convex sets V in X with the property that $V = \cap_{r>1} rV$, and seminorms p on X, as the lemmas show.

2.1.19 LEMMA. *Let V be any absorbing, convex set in X. Define a function $p_V : X \to [0,\infty)$ by*

$$p_V(x) = \inf\{t > 0 \ : \ x \in tV\} \ . \tag{2.1.11}$$

Then p_V is subadditive and positivel homogeneous of degree one. If morover V is balanced, then p_V is a seminorm on X.

Proof. Since V is absorbing, $p_V(x) < \infty$ for all $x \in X$. Next, let $x, y \in X$. Let t, s be such that $x \in tV$ and $y \in sV$. That is, $t^{-1}x, s^{-1}y \in V$. Since V is convex, $(1-\lambda)t^{-1}x + \lambda s^{-1} y \in V$ for all $\lambda \in (0,1)$. Define $\lambda = s/(t+s)$. Then we have $x + y \in (t+s)V$. Therefore, $p_V(x+y) \le p_V(x) + p_V(y)$.

For all $r, t > 0$,

$$t(rx) \in V \iff (tr)x \in V$$

and therefore $p_V(rX) = rp_V(x)$. This porves that p_V is positively homogeneous of degree one, and completes the proof of the first part of the lemma.

Now suppose that V is balanced. Then $p_V(\alpha x) = p_V(x)$ when $|\alpha| = 1$. Then for $r > 0$ and $\alpha \in \mathbb{C}$, $\alpha = 1$, for all $x \in X$,

$$p_V(r\alpha x) = rp_V(\alpha x) = rp_V(x) = |r\alpha| p_v(x) \ .$$

This shows that p_V is homogeneous of degree one. Altogether, we have shown that p_V is a seminorm. □

2.1.20 LEMMA. *Let V be any absorbing, balanced, convex set in X. Suppose also that $V = \cap_{r>1} rV$. Let p_V be the seminorm defined by (2.1.11), and then let B_{p_V} be the absorbing, balanced, convex set defined by (2.1.9) with p_V in place of V. Then $V = B_{p_V}$.*

Proof. If $x \in V$, then $p_V(x) \le 1$, and hence $x \in B_{p_V}$. If $x \in B_{p_V}$, $p_V(x) \le 1$, and then $x \in rV$ for all $r > 1$. Since $V = \cap_{r>1} rV$, $x \in V$. Thus, $V = B_{p_V}$. □

2.1.21 LEMMA. *Let $V \subset X$ be absorbing, balanced and convex, and let p_V be the seminorm defined in (2.1.11). Then p_V is a norm if and only if for all non-zero $x \in X$, there is some $t > 0$ such that $x \notin tV$*

Proof. If for all non-zero $x \in X$, there is some $t_0 > 0$ such that $x \notin tV$, $p_V(x) = \inf\{t > 0 \ : \ x \in tV\} \geq t_0$, and hence $p_V(x) > 0$ for all $x \neq 0$. Conversely, if $p_V(x) > 0$ for all $x \neq 0$, then for all x and all $0 < t < p_V(x)$, $x \notin tV$. □

Summarizing, we have the following reuslt:

2.1.22 THEOREM. *Let X be a vector space over $\mathbb{C}$. For each absorbing, balanced, convex set V in X, the function p_V defined by (2.1.11) is a seminorm. Conversely, for each seminorm p, the set is absorbing, balanced and convex. Moreover the map $p \mapsto B_p$, with B_p defined in (2.1.9), is a bijection between the set of seminorms on X and the set of absorbing, balanced, convex sets $V \subset X$ with the property that $V = \cap_{r>1} rV$. Finally, for any absorbing, balanced and convex $V \subset X$, the seminorm p_V defined in (2.1.11) is a norm if and only if for all non-zero $x \in X$, there is some $t > 0$ such that $x \notin tV$.*

As a consequence of Theorem 2.1.11 and Theorem 2.1.22, there is a topology $\mathscr{O}$ on a vector space X associated to any set $\mathscr{P}$ of seminorms p on X, and $(X, \mathscr{O})$ is a topological vector space such that for each $p \in \mathscr{P}$ and each $\epsilon > 0$, $\epsilon B_p \in \mathscr{O}$, and consequently, such that each $p \in \mathscr{P}$ is continuous.

Let p be any seminorm on the vector space X. Consider the nested family of absorbing, balanced, convex sets $\mathscr{W} = \{2^k B_p \ : \ k \in \mathbb{Z}\}$. Since this set is nested, it is closed under finite intersections. Then by Theorem 2.1.11, the set $\mathscr{O}$ defined by (2.1.3) is a topology on X such that $(X, \mathscr{O})$ is a topological vector space. This topology is Hausdorff if and only if p is a norm.

To get a Hausdorff topology out of seminorms, one must use a sufficiently large family of them. Let $\mathscr{P}$ be a set of seminorms on X. Define

$$\mathscr{W}_{\mathscr{P}} := \{2^k B_p \ : \ k \in \mathbb{Z} \ , \ p \in \mathscr{P}\} \ . \tag{2.1.12}$$

Then $\mathscr{W}$ is closed under finite intersections and each set $W \in \mathscr{W}$ is balanced, convex and absorbing and by Theorem 2.1.11, the set $\mathscr{O}$ defined by (2.1.3) is a topology on X such that $(X, \mathscr{O})$ is a topological vector space. This topology is Hausdorff if and only if for each $x \in$, there is some $p \in \mathscr{P}$ such that $p(x) \neq 0$.

2.1.23 DEFINITION. Let $\mathscr{P}$ be a family of seminorms on a vector space X. The weakest topology on X that contains all translates of all of the sets in $\mathscr{W}_{\mathscr{P}}$, as defined in (2.1.12) is called the topology on X generated by $\mathscr{P}$.

If $\mathscr{P}$ is a *countable* set of seminorms on X, then topology on X generated by $\mathscr{P}$ is metrizable. This is the main content of the next theorem.

2.1.24 THEOREM. *Let $\mathscr{P}$ be a countable family of seminorms on a vector space X, and suppose that for each $x \in X$, there is some $p \in \mathscr{P}$ such that $p(x) \neq 0$. Then there is a translation is a translation invariant metric ρ on X such that the topology induced on X by this metric coincides with the topology on X generated by $\mathscr{P}$.*

Proof. Order the elements of $\mathscr{P}$ in a sequence $\{p_n\}_{n \in \mathbb{N}}$. Define the function $\phi : [0, \infty) \to [0, 1)$ by $\phi(t) = t/(1+t)$. Then for $s, t \geq 0$,

$$\phi(s+t) = \frac{s}{1+s+t} + \frac{t}{1+s+t} \leq \phi(s) + \phi(t) \tag{2.1.13}$$

and ϕ is strictly monotone increasing. For $x, y \in X$ define

$$\rho(x,y) = \sum_{n=1}^{\infty} 2^{-n}\phi(p_n(x-y)) \ , \tag{2.1.14}$$

and note that the sum converges absolutely and in fact $\rho(x, y) < 1$ for all $x, y \in X$. It is evident that $\rho(x, y) = \rho(y, x)$, and since if $x \neq y$, there is some n such that $p_n(x - y) \neq 0$, then $\rho(x, y) \neq 0$. Finally, by the subadditivity of each p_n, for all $x, y, z \in X$,

$$p_n(x-z) = p_n((x-y)+(y-z)) \le p_n(x-y) + p_n(y-z) \ ,$$

and then by (2.1.13), $\phi(p_n(x - z)) \le \phi(p_n(x - y)) + \phi(p_n(x - z))$. It follows that ρ satisfies the triangle inequality, and therefore is a metric on X. Notice that by definition, for all $x, y, z \in X$,

$$\rho(T_x y, T_x z) = \rho(y, z) \ ,$$

so that ρ is translation invariant. For $r > 0$, let $B_\rho(r, 0) := \{x \in X \ : \ \rho(x, 0) < r\}$.

It remains to show that for each $k \in \mathbb{Z}$ and $n \in \mathbb{N}$, $2^k B_{p_n}$ contains $B_\rho(r, 0)$ for some $r > 0$, and that for each $r > 0$, $B_\rho(r, 0)$ contains a set that is open in the topology generated by $\mathscr{P}$.

Consider the set $2^k B_{p_n} = \{x \ : \ p_n(x) < 2^k\}$. Since for each n, and each x, $\rho(x, 0) \ge 2^{-n}\phi(p_n(x))$,

$$p_n(x) \le \frac{2^n \rho(x,0)}{1 - 2^n \rho(x,0)}$$

for all x such that $\rho(x, 0) < 2^{-n}$. Hence for each $n \in N$, and each $\epsilon > 0$, there is an $r_{n,\epsilon} > 0$ such that $p_n(x) < \epsilon$ whenever $x \in B(r_{n,\epsilon}, 0)$. Taking $\epsilon = 2^k$, this can be written as

$$B(r_{n,2^k}, 0) \subset 2^k B_{p_n} \ .$$

Next, fix $r > 0$ and $n \in \mathbb{N}$. Fix $N \in N$ such that $2^{-N} < r/2$. Then

$$\rho(x,0) = \sum_{n=1}^{N} 2^{-n}\phi(p_n(x)) + \sum_{n=N+1}^{\infty} 2^{-n} \le \sum_{n=1}^{N} 2^{-n}\phi(p_n(x)) + r/2$$

Since for all $t > 0$, $\phi(t) \le t$, if $x \in (r/2)B_{p_n}$, then $\phi(p_n(x)) \le (r/2)$. Now pick $\ell \in \mathbb{N}$ so that $2^{-\ell} < r/2$, and note that

$$W := \bigcap_{n=1}^{N} 2^{-\ell} B_{p_n}$$

is in the canonical neighborhood base at 0 of the topology generated by $\mathscr{P}$. By what we have noted above, for all $x \in \mathscr{P}$, $\rho(x, 0) < r$. That is $W \subset B(r, 0)$. $\square$

A single seminorm that is not a norm cannot generate a Hausdorff topology: If p is a seminorm but not a norm, there is some non-zero $x \in X$ such that $p(x) = 0$, and therefore $tx \in B_p$ for all $t > 0$, or what is the same, $x \in 2^k B_p$ or all $k \in \mathbb{Z}$. Hence x belongs to each open set containing 0. However, when the seminorm p is norm, this problem is eliminated. A very important class of topological vector space topologies arises this way.

2.1.5 Normed vector spaces

2.1.25 DEFINITION (Normed vector space)**.** A *normed vector space* $(X, \|\cdot\|)$ is a vector space X equipped with a norm function $\|\cdot\|$. For each $r > 0$ and $x_0 \in X$, define

$$B(r, x_0) = \{x \in X \ : \ \|x - x_0\| < r\} \ . \tag{2.1.15}$$

For each $r > 0$, $B(r, 0)$ is balanced, convex and absorbing. Define $\mathscr{W} = \{B(2^k, 0) \ : \ k \in \mathbb{Z}\}$, which is nested and therefore closed under finite intersections. Then by Theorem 2.1.11, $\mathscr{W}$ is a neigborhood base for a topology on X that makes it a topological vector space. This topology is called the *norm topology.* The set $B(r, x_0)$ is call the *open ball of radius r centered at x_0*.

Let $(X, \|\cdot\|)$ be a vector space, and define a function $\rho : X \times X \to [0, \infty)$ by $\rho(x, y) = \|x - y\|$. Then evidently $\rho(x, y) = \rho(y, x)$ for all x, y, and $\rho(x, y) = 0$ if and only if $x = y$. Moreover, by the subadditivity of the norm, for all $x, y, z \in X$,

$$\rho(x, z) = \|x - z\| = \|(x - y) + (y - z)\| \leq \|x - y\| + \|y - z\| = \rho(x, y) + \rho(y, z) \ .$$

Therefore, the triangle inequality is satisfied, and hence ρ is a metric on X called the *norm metric.*

For all $x_0 \in X$, the metric open ball of radius r about is precisely the set $B(r, x_0)$ defined in (2.1.15), and $U \subset X$ is open in the metric topology if and only if for each x_0 in U, there is some $r > 0$ such that $B(r, x_0) \subset U$. Decreasing r if need be, we may assume that $r = 2^k$ for some $k \in Z$, and then since $B(2^k, x_0) = x_0 + B(2^k, 0)$, the metric topology is precisely the topology determined by the neighborhood base $\mathscr{W} = \{B(2^k, 0) \ : \ k \in \mathbb{Z}\}$.

A normed vector space is therefore not only a Hausdorff topological vector space, but the norm topology is a metric topology, and it has a countable neighborhood base.

2.1.26 DEFINITION (Equivalent norms)**.** Let $\|\cdot\|_0$ and $\|\cdot\|_1$ be two norms on a vector space X. These norms are *equivalent* in case for some constant $0 < C < \infty$,

$$C\|x\|_1 \leq \|x\|_0 \leq \frac{1}{C}\|x\|_1 \qquad \text{for all } x \in X \ . \tag{2.1.16}$$

$\|\cdot\|_0$ and $\|\cdot\|_1$ be two norms on a vector space X such that (2.1.16) is satisfied. For $j = 0, 1$, let $B_j(r, 0)$ be the centered open ball of radius r for the norm $\|\cdot\|_j$. Then is equivalent to

$$B_1(Cr, 0) \subset B_0(r, 0) \subset B_1(r/C, 0) \tag{2.1.17}$$

for all $r > 0$. Therefore, two norms generate the same topology if and only if they are equivalent.

2.1.27 THEOREM. *Let $(X, \|\cdot\|)$ be a normed vector space, and let Y be a finite dimensional subspace of X. Then Y is closed. Moreover, all norms on any finite dimensional vector space are equivalent to one another.*

Proof. since normed vector sapces are Hausforff topological vector spaces, this is immediate from Theorem 2.1.13. □

2.1.28 THEOREM. *Let $(X, \|\cdot\|_X)$ and $(Y, \|\cdot\|_Y)$. Let $B_X(r, 0)$ and $B_Y(r, 0)$ denote the open balls of radius $r > 0$ in X and Y respectively. Let T be a linear transformation from X to Y. Then T is continuous if and only if*

$$\|T\| := \sup\{ \ \|Tx\|_Y \ : \ x \in B_X(1, 0) \ \} < \infty \ . \tag{2.1.18}$$

Proof. Suppose that T is continuous. Then $T^{-1}(B_Y(1,0))$ contains $B_X(r,0)$ for some $r > 0$, and hence for all $x \in B_X(1,0)$, $\|Tx\|_Y \leq 1/r$. Hence $\|T\| \leq 1/r$. Conversely, suppose that (3.1.1) is valid. Then for all $x_1, x_2 \in X$ and all $\lambda > \|x_1 - x_2\|_X$,

$$\|Tx_1 - Tx_2\|_Y = \lambda\|T((x_2 - x_2)/\lambda)\|_Y \leq \lambda\|T\| \ .$$

It follows that

$$\|Tx_1 - Tx_2\|_Y \leq \|T\|\|x_1 - x_2\|_X \ .$$

Thus, T is not only continuous, it is Lipschitz continuous. □

On account of this theorem, the term *bounded linear transformation* is often used as a synonym for the term continuous linear transformation when referring to transformations from one normed vector space to another. The bounded transformations themselves are often called *operators*.

2.1.29 DEFINITION. Let $(X, \|\cdot\|_X)$ and $(Y, \|\cdot\|_Y)$ be normed vector spaces. Let $\mathscr{B}(X,Y)$ denote the vector space of continuous (bounded) linear transformations from X to Y. The function $T \mapsto \|T\|$, with $\|T\|$ defined by (3.1.1) is called the *operator norm* on $\mathscr{B}(X,Y)$.

The terminology in the previous definition is appropriate; it is easy to see, and left to the reader, that the operator norm, is indeed a norm on the vector space $\mathscr{B}(X,Y)$

We shall be forced to consider topologies $\mathscr{O}$ in vector spaces X that are not topologies coming from norms on account a simple corollary of the following result of Riesz.

2.1.30 LEMMA (Riesz's Lemma)**.** *Let $(X, \|\cdot\|)$ be a normed vector space. Let Y be a proper, closed subspace of X. Then for all $\alpha \in (0,1)$, there exists $u \in X$, $\|u\| = 1$, such that*

$$\alpha \leq \inf\{\|u - y\| \ : \ y \in Y\} \ .$$

Proof. Since Y is proper, there exists some $x_0 \notin Y$, and then since Y is closed, there exists some $r > 0$ such that $B(r, x_0) \cap Y = \emptyset$. Therefore, $d := \inf\{\|u - y\| \ : \ y \in Y\} \geq r > 0$. By the definition of d, for all $\alpha \in (0,1)$, there exists $y_0 \in Y$ such that $\|x_0 - y_0\| < d/\alpha$. Then, by the definition of d,

$$\alpha\|x_0 - y_0\| \leq d \leq \|x_0 - y_0 - y\| \quad \text{for all} \quad y \in Y \ .$$

That is,

$$\alpha \leq \left\| \frac{x_0 - y_0}{\|x_0 - y_0\|} - \frac{y}{\|x_0 - y_0\|} \right\| \quad \text{for all} \quad y \in Y \ .$$

Let $u = \|x_0 - y_0\|^{-1}(x_0 - y_0)$. Notice that $\|x_0 - y_0\|^{-1}y$ ranges over all of Y as y ranges over Y. □

2.1.31 COROLLARY. *In every infinite dimensional normed vector space $(X, \|\cdot\|)$, there exists an infinite sequence $\{u_n\}_{n\in\mathbb{N}}$ such that $\|u_n\| = 1$ for all n, but from which no convergence subsequence can be extracted. In other words, the closed unit ball in an infinite dimensional normed vector space is never sequentially complete.*

Proof. The sequence $\{u_n\}_{n\in\mathbb{N}}$ is constructed inductively as follows: Pick some $u_1 \in X$ with $\|u_1\| = 1$. For the induction, suppose that we have found vectors $\{u_1, \dots, u_n\}$ such that $\|u_j\| = 1$ for each j and if $j \neq k$, then $\|u_j - u_k\| \geq 1/2$. Let $V_n := \text{span}(\{u_1, \dots, u_n\})$ which is is closed by Theorem 2.1.27 proper since X is infinite dimensional. By Riesz's Lemma, we may choose u_{n+1} with $\|u_{n+1}\| = 1$ snd $\|u_{n+1} - y\| \geq 1/2$ for all $y \in Y$. In particular we have $\|u_j - u_j\| \geq 1/2$ for all $1 \leq j < k \leq n+1$. □

2.1.6 Topologies induced by sets of linear transformations

The usual way of constructing seminorms on X is to consider linear transformations T from X to a normed vector space $(Y, \|\cdot\|)$. (A particularly important example is that in which $(Y, \|\cdot\|) = (\mathbb{C}, |\cdot|)$.) Then evidently the function p_T on X defined by

$$p_T(x) = \|Tx\| \tag{2.1.19}$$

is a seminorm, and is a norm if and only if T is injective.

Let $(Y, \|\cdot\|)$ be a normed space and let $\mathscr{T}$ be any set of linear maps from $X \to Y$. For each $T \in \mathscr{T}$, let p_T be defined as in (2.1.19). Let $\widetilde{\mathscr{O}}$ be any translation invariant topology on X. Then T is continuous from $(X, \widetilde{\mathscr{O}})$ to Y equipped with the norm topology if and only if for each $x_0 \in X$, and each $\epsilon > 0$, there is a set $U \in \widetilde{\mathscr{O}}$ such that for all $x - x_o \in U$,

$$\|Tx - Tx_0\| = \|T(x - x_0)\| < \epsilon \ .$$

Since $p_T(x - x_0) = \|T(x - x_0)\|$, for all $k \in Z$,

$$x \in x_0 + 2^{-k} B_{p_T} \iff x - x_0 \in 2^{-k} B_{p_T} \iff \|Tx - Tx_0\| < 2^{-k} \ .$$

Hence T is continuous on $(X, \widetilde{\mathscr{O}})$ if and only if if $\widetilde{\mathscr{O}}$ contains each of the sets $x_0 + 2^k B_{p_T}$, $k \in \mathbb{Z}$, $x_0 \in X$. This proves:

2.1.32 THEOREM. *Let $\mathscr{T}$ be any set of linear transformations from a vector space X to a normed space $(Y, \|\cdot\|)$. Let $\mathscr{P}$ be the set of seminorms on X given by $\mathscr{P} = \{p_T \ : T \in \mathscr{T}\}$ where $p_T(x) = \|T(x)\|$. Let $\mathscr{O}$ be the topology generated by $\mathscr{P}$, as in Definition 2.1.23. Then $(X, \mathscr{O})$ is a topological vector space such that each $T \in \mathscr{T}$ is continuous from $(X, \mathscr{O})$ to $(Y, \|\cdot\|)$, and $\mathscr{O}$ is the weakest topology of any sort on X such that each $T \in \mathscr{T}$ is continuous from $(X, \mathscr{O})$ to $(Y, \|\cdot\|)$.*

In what follows, the case in which $(X, \|\cdot\|_X)$ and $(Y, \|\cdot\|_Y)$ are two normed vector spaces, and $\mathscr{T} = \mathscr{B}(X, Y)$, the set of all bounded linear transformations from X to Y, is especially important. In many cases we shall be able to show that norm closed, norm bounded, convex sets $K \subset X$ are compact in the weak topology on X generated by $\mathscr{T}$. (To say that K is *norm bounded*, or simply *bounded*, means that there exists $C < \infty$ such that $\|x\| \leq C$ for all $x \in K$.) This will be more useful to us if we have not only compactness, but sequential compactness, which is the same thing if this weak topology, or at least the relative weak topology on K is metrizable.

The following theorem gives a useful sufficient condition for weak topology generated by $\mathscr{T}$ to be metrizable on bounded subsets of X.

2.1.33 THEOREM. *$(X, \|\cdot\|_X)$ and $(Y, \|\cdot\|_Y)$ are two normed vector spaces, and let $\mathscr{T}$ be any subset of $\mathscr{B}(X, Y)$. Suppose that $\mathscr{T}_0$ is a countable subset of $\mathscr{T}$ that is dense in $\mathscr{T}$ in the operator norm. Then restricted to bounded subsets the weak topologies generated by $\mathscr{T}_0$ and $\mathscr{T}$ coincide.*

Proof. Let K be a bounded subset of X. Let $\mathscr{O}_0$ and $\mathscr{O}$ be the relative weak topologies generated by $\mathscr{T}_0$ and $\mathscr{T}$ respectively. Evidently, $\mathscr{O}_0 \subset \mathscr{O}$.

Let $C < \infty$ be such that $\|x\|_X \le C$ for all $x \in K$. For any $T \in \mathscr{T}$ and any $\epsilon > 0$, there is $T_0 \in \mathscr{T}_0$ such that $\|T - T_0\| < \epsilon$, and then for all $x \in K$,

$$|\|Tx\|_Y - \|T_0x\|_Y| \le \|(T - T_0)x\|_Y \le \|T - T_0\|\|x\| \le \epsilon C \ .$$

It follows that $\|T_0x\|_Y \le \|Tx\|_Y + \epsilon C$ for all $x \in K$. Hence for all $k \in \mathbb{N}$, if we choose ϵ (and then T_0) so that $\epsilon < 2^{-k-1}/C$, $2^{-k-1}B_{p_{T_0}} \cap K \subset 2^{-k}B_{p_T} \cap K$. By Lemma 2.1.12, $\mathscr{O} \subset \mathscr{O}_0$. □

2.2 Banach spaces

2.2.1 Banach spaces of bounded linear transformations

2.2.1 DEFINITION (Banach space)**.** A *Banach space* is a normed vector space $(X, \|\cdot\|)$ that is complete in its norm topology.

2.2.2 THEOREM. *Let $(X, \|\cdot\|_X)$ be a normed space, and let $(Y, \|\cdot\|_Y)$ be a Banach space. Then $\mathscr{B}(X, Y)$, equipped with the operator norm is Banach space.*

Proof. Let $\{T_n\}_{n\in\mathbb{N}}$ be a Cauchy sequence in $\mathscr{B}(X, Y)$. Then for each $x \in X$, $\{T_nx\}_{n\in\mathbb{N}}$ is a Cauchy sequence in Y. Since Y is complete, there exists $y \in Y$ such that $y = \lim_{n\to\infty} T_nx$. Define a function T mapping X to Y by $Tx = y$.

To see that T is linear, fix $x_1, x_2 \in X$ and $\alpha_1, \alpha_2 \in \mathbb{C}$. Then

$$T(\alpha_1x_1 + \alpha_2x_2) = \lim_{n\to\infty} T_n(\alpha_1x_1 + \alpha_2x_2) = \alpha_1 \lim_{n\to\infty} T_nx_1 + \alpha_2 \lim_{n\to\infty} T_nx_2 = \alpha_1T_nx_1 + \alpha_2Tx_2 \ .$$

To see that T is bounded, first observe that since $|\|T_n\| - \|T_m\|| \le \|T_n - T_m\|$, $\{\|T_n\|\}_{n\in\mathbb{N}}$ is a Cauchy sequence in $\mathbb{R}$, and hence $a := \lim_{n\to\infty} \|T_n\|$ exists.

Now consider any $x \in X$ with $\|x\|_X < 1$. By the triangle inequality,

$$\|Tx\|_Y \le \|T_nx\|_Y + \|(T - T_n)x\|_Y \le \|T_n\| + \|(T - T_n)x\|_Y \ , \tag{2.2.1}$$

and since $\lim_{n\to\infty} \|(T - T_n)x\|_Y = 0$, $\|Tx\|_Y \le \lim_{n\to\infty} \|T_n\|$. This shows that T is bounded, and in fact, $\|T\| \le \lim_{n\to\infty} \|T_n\|$. Now the same reasoning that led to (2.2.1) shows that $\|T_nx\| \le \|T\| + \|(T - T_n)x\|_Y$ for all $x \in X$ with $\|x\|_X < 1$. Therefore,

$$\lim_{n\to\infty} \|T_nx\| \le \|T\| \ . \tag{2.2.2}$$

Altogether, $\|T\| = \lim_{n\to\infty} \|T_n\|$. Now considering the Cauchy sequence $\{T_n - T\}_{n\in\mathbb{N}}$, we conclude that $\lim_{n\to\infty} \|T_n - T\| = 0$. which shows that T is the operator norm limit of the Cauchy sequence $\{T_n\}_{n\in\mathbb{N}}$. □

An important spacial case of Theorem 2.2.2 is that in which $(Y, \|\cdot\|) = (\mathbb{C}, |\cdot|)$, which is of course the simplest example of a Banach space.

2.2.3 DEFINITION (Dual space)**.** Let $(X, \|\cdot\|)$ be a normed space. Let X^* denote the set of bounded linear transformations from $(X, \|\cdot\|)$ to $(\mathbb{C}, |\cdot|)$ and let $\|\cdot\|_*$ denote the operator norm on X^*. Then $(X^*, \|\cdot\|_*)$ is a Banach space called the *dual space* to $(X, \|\cdot\|)$. The norm $\|\cdot\|_*$ is called the *dual norm.* Elements of X_* are referred to as *bounded linear functionals on* X. Applying the same construction to $(X^*, \|\cdot\|_*)$ we obtain the *second dual* $(X^{**}, \|\cdot\|_{**})$, which is also a Banach space.

There is a natural embedding of X into X^{**}: For each $x \in X$, define the function ϕ_x on X^* by

$$\phi_x(L) = L(x) \quad \text{for all } L \in X^* \ . \tag{2.2.3}$$

It is evident that ϕ_x is a linear functional from X^* to $\mathbb{C}$, and for all $x \in X$, $L \in X^*$, $|\phi_x(L)| = |L(x)| \leq \|L\|_*\|x\|$. Therefore ϕ_x is bounded, so that $\phi_x \in X^{**}$, and in fact,

$$\|\phi_x\|_{**} \leq \|x\| \ .$$

Thus, the map $x \mapsto \phi_x$ is a contractive linear map from $(X, \|\cdot\|)$ into $(X^{**}, \|\cdot\|_{**})$, which is a Banach space. The results of the next section show that this contractive map is actually an isometry, so that every normed vector spaces is isometrically embedded into a Banach space in a canonical manner. We may identify the closure of the image as the *completion* of $(X, \|\cdot\|)$.

2.2.2 The real Hahn-Banach extension theorem

In this section we consider real vector spaces X. Every complex vector space X is also a vector space over $\mathbb{R}$, only now, for each non-zero $x \in X$, x and ix are linearly independent, and in the next section we shall extend the results of this section to complex vector spaces through this connection.

2.2.4 LEMMA (Helly's Lemma)**.** *Let X be a real vector space, and let V be a subspace of X. Let $x \in X$, $x \notin V$, and let $W = \mathrm{span}(\{x\}\cup V)$, so that V is a subspace of W of co-dimension 1. Let p be a function on X with values in $[0,\infty)$ that is sub-additive; i.e., $p(x+y) \leq p(x)+p(y)$ for all $x, y \in X$, and is postitively homogeneous of degree one; i.e., such that $p(tx) = tp(x)$ for all $t \geq 0$ and all $x \in X$.*

Let L be a linear functional on V such that

$$L(y) \leq p(y) \quad \text{for all } y \in V \ . \tag{2.2.4}$$

Then there exists an extension $\widetilde{L}$ of L as a linear functional to W such that $L(w) \leq p(w)$ for all $w \in W$.

2.2.5 Remark. In many, but not all, applicartions, $p(x)$ will be norm on X; i.e., $p(x) = \|x\|$. Then of course $\|\cdot\|$ is sub-additive, and it is homogenous of degree one; i.e., $p(\lambda x) = |\lambda| p(x)$ for all $\alpha in \mathbb{R}$, and not only $p(tx) = tp(x)$ for all $t \geq 0$ which is what we shall use in the proof. In particular, p need not even be a semi-norm. We also do not require that $p(x) > 0$ for $x \neq 0$. This is true when p is a norm, but it has no role in the proof.

Proof of Helly's Lemma. Let $y_1, y_2 \in V$. By the subadditivity of p, ,

$$L(y_1) - L(y_2) = L(y_1 - y_2) \leq p(y_1 - y_2) = p((y_1 + x) - (y_2 + x)) \leq p(y_1 + x) + p(-y_2 - x) \ .$$

Rearranging terms, to bring all y_2 terms to the left, and all y_1 terms to the right,

$$-L(y_2) - p(-y_2 - x) \leq -L(y_1) + p(y_1 + x) \ .$$

Since y_1, y_2 are arbitrary,

$$a := \sup_{y\in V}\{\ -L(y) - p(y + x)\} \leq \inf_{y\in V}\{\ -L(y) + p(-y - x)\} =: b \ .$$

Pick any $\tilde{\lambda} \in [a,b]$. Define $\widetilde{L}(y+tx) = L(y) + t\tilde{\lambda}$. Then for $t > 0$,

$$\widetilde{L}(y+tx) = t[L(y/t) + \tilde{\lambda}] \le t[L(y/t) - L(y/t) + p(y/t + x)] = p(y+tx) \ ,$$

while for $t < 0$,

$$\widetilde{L}(y+tx) = t[L(y/t) + \tilde{\lambda}] \ge t[L(y/t) - L(y/t) - p(-y/t - x)] = |t|p(-y/t - x) = p(y+tx) \ ,$$

Hence, for all $t \in \mathbb{R}$, and all $y \in V$, $|\widetilde{L}(y+tx)| \le p(y+tx)$. □

2.2.6 THEOREM (The Real Hahn-Banach Extension Theorem)**.** *Let X be a real vector space, and let Y be a subspace of X. Let p be a function on X with values in $[0,\infty)$ that is sub-additive and postitively homogeneous of degree one.*

Let L be a linear functional on Y such that for some $L(y) \le p(y)$ for all $y \in Y$. Then there exists a linear functional $\widetilde{L}$ on all of X such $\widetilde{L}(x) \le p(x)$ for all $x \in X$ and such that $\widetilde{L}y = Ly$ for all $Y \in Y$.

Proof. Consider the set of pairs (V, L_V) of subspaces V of X and L_V on V with $L_V(v) \le p(v)$ for all $v \in V$. Partially order this set of pairs so that $(V, L_V) \prec (W, L_W)$ in case $V \subset W$ and $L_W\big|_V = L_V$. By Zorn's Lemma, since every linearly ordered chain containing (Y, L) has a maximal element, the set of all such pairs has a maximal element $(\widehat{L}, \widehat{X})$ (V, L_V). By Lemma 2.2.4, $\widehat{X} = X$, or else $(\widehat{L}, \widehat{X})$ would not be maximal. □

We now give an important geometric application of the Hahn-Banach extension that uses its full generality.

2.2.7 THEOREM (Hahn-Banach Separation Theorem)**.** *Let $(X, \mathscr{O})$ be a real locally convex topological vector space.*

(1) Let A and B be convex, non-empty subsets of X with A open, and $A \cap B = \emptyset$. Then there exists a continuous linear functional L on X such that for all $x \in A$ and $y \in B$,

$$L(x) < 1 \le L(y) \ . \tag{2.2.5}$$

(2) Let K and F be convex, non-empty subsets of X with K compact, and F closed, and $K \cap F = \emptyset$. Then there exists a continuous linear functional L on X such that for all $x \in A$ and $y \in B$,

$$L(x) < 1 \le L(y) \ . \tag{2.2.6}$$

Proof. We first prove *(1)*. Choose $x_0 \in A$ and $y_0 \in B$, which is possible since A and B are non-empty. Define $z_0 := y_0 - x_0$ and $C = A - B + z_0$ (where $A - B$ denotes $A + (-1)B$). C is open and convex, and since $-z_0 \in A - B$, $0 \in C$. Since $A \cap B = \emptyset$, $z_0 \notin C$.

Since C is a neighborhood of 0, by Lemma 2.1.6, C is absorbing so that for all $x \in X$, there is some $t \in (0,\infty)$ such that $x \in tC$. Therefore, we may define a function p_C on X with values in $[0,\infty)$ by

$$p_C(x) = \inf\{\ r > 0 \ : \ x \in rC\ \} \ , \tag{2.2.7}$$

and since C is convex, Lemma 2.1.19 says that p_C is subadditive and positively homogeneous of degree one. Also since C is convex, for $r \in [0,1]$, $rC \subset C$, and hence $z_0 \notin rC$ for any $r < 1$. Therefore, $p_C(z_0) \ge 1$

Define a linear functional L_0 on span$(\{z_0\})$ by $L_0(tz_0) = 1$. Then for $t > 0$, $L_0(tz_0) = t \leq tp(z_0) = p(tz_0)$, whicle for $t \leq 0$, $L_0(tz_0) = t \leq p(tz_0)$ is trivially true. Hence $L_0 \leq p$ in span$(\{z_0\})$. By Theorem 2.2.6, there exists an extension L of L_0 to all of X such that $L \leq p_C$ on all of X.

Let $x \in A$ and $y \in B$. Then $x-y+z_0 \in C$, which is open, and hence for some $\epsilon > 0$, $(1+\epsilon)(x-y+z_0) \in C$. Therefore, $p_C(x-y+z_0) \leq (1+\epsilon)^{-1}$. Then we have $L(x-y+z_0) \leq (1+\epsilon)^{-1}$. Rearranging terms,

$$L(x) \leq L(y) + (1+\epsilon)^{-1} - L(z_0) = L(y) - \frac{\epsilon}{1+\epsilon} \ .$$

In particular, $L(x) < L(y)$. Define $a = \sup_{x\in A}\{L(x)\} > 0$. Then $L(x) \leq a \leq L(y)$ for all $x \in A$ and $y \in B$.

We now claim that $L(x) < a$ for all $x \in A$. Suppose on the contrary that $x \in A$ and $L(x) = a$. Since A is open, there is a neighborhood V of 0 such that $x + V \subset A$, and by Lemma 2.1.7, we may take V to be ballanced. It follows that $L(v) \leq 0$ for all $v \in V$. Since V is balanced, and $L \neq 0$, this is impossible. Hence for all $x \in A$, $L(x) < a$. Replacing L by $a^{-1}L$, we obtain the desired linear functional.

To complete the first part of the proof we must show that L is continuous. By Lemma 2.1.7, C contains a balanced neighborhood W of 0. Since $L \leq P_C \leq 1$ on C and hence on W, for all $w \in W$, $L(w) \leq 1$ and $-L(w) = L(-w) \leq 1$ so that $|L(w)| \leq 1$. Hence $\frac{1}{2}W \subset L^{-1}((-1,1))$, showing that L is continuous.

We next prove *(2)*. By Theorem 2.1.10, there is a balanced, convex neighborhood V of 0 such that $(K+V)\cap F = \emptyset$. $K+V$ is open and convex, and therefore by part *(1)*, there exists a continuous linear functional L such that for all $x \in K$ and all $y \in F$, $L(x) < 1 \leq L(y)$. Now define $a := \max_{x\in K}\{L(x)\} > 0$. Since K is compact and L is continuus, there is some $x_0 \in K$ such that $a = L(x_0) < 1$. Replacing L by $((1+a)/2)^{-1}L$, we obtain the desired linear functional. □

2.2.3 The complex Hahn-Banach extension theorem

In this section $(X, \mathscr{O})$ will denote a complex locally convex topological vector space, and $(X_{\mathbb{R}}, \mathscr{O})$ will denote the real normed vector space obtained by regarding X as a real vector space, and equipping it with the same topology. We write X^* denotes the set of continuous (complex) linear functionals on $(X, \mathscr{O})$, and $X^*_{\mathbb{R}}$ denotes the set of continuous (real) linear functional on $(X_{\mathbb{R}}, \mathscr{O})$.

2.2.8 LEMMA (Murray's Lemma)**.** *Let $(X, \mathscr{O})$ be a complex normed vector space. Let $R \in (X_{\mathbb{R}}, \mathscr{O})$ be a continuous real linear functional on $X_{\mathbb{R}}$. Define the functional L_R on X by*

$$L_R(x) = R(x) - iR(ix) \ . \tag{2.2.8}$$

Then L_R is complex linear and continuous on $(X, \mathscr{O})$, and for all $x \in X$, $R(x) = \Re(L_R(x))$, and for any balanced neighborhood V of 0 in X,

$$\sup\{|L_R(x)| \ : \ x \in V\ \} = \sup\{|R(x)| \ : \ x \in V\ \} \ . \tag{2.2.9}$$

In particular, if the topology $\mathscr{O}$ on X is generated by a norm, $\|L_R(x)\| = \|L(x)\|$ for all x inX. Finally, if $R = \Re\circ L$, $L \in X^*$, then $L = L_R$. Thus the map $L \mapsto \Re\circ L$ is a real linear homeomorphic isomorphism between X^* and $X^*_{\mathbb{R}}$ that is even isometric when the topology $\mathscr{O}$ is generated by a norm.

Proof. Evidently, for all $x_1, x_2 \in X$ and $t_1, t_2 \in \mathbb{R}$, $L_R(t_1x_1 + t_2x_2) = L_R(x_1) + t_2L_R(x_2)$. Therefore, it suffices to show that for all $x, y \in X$, $L_R(x + iy) = L_R(x) + iL_R(y)$, and this is a simple calculation. The fact that $R(x) = \Re(L_R(x))$ is evident. It follows that $|R(x)| \le |L_R(x)|$ for all $x \in X$. Therefore, for any neighborhood V of 0 in X, $\sup\{|R(x)| \ : \ x \in V\} \le \sup\{|L_R(x)| \ : \ x \in V\}$. When V is a balanced neighborhood of 0, and $x \in V$ there is some $\theta \in [0, 2\pi)$ such that $e^{i\theta}L_R(x) > 0$, and then $e^{i\theta}x \in V$. Therefore,

$$|L_R(x)| = e^{i\theta}L_R(x) = L_R(e^{i\theta}x) = \Re(L_R(e^{i\theta}x)) = R(e^{i\theta}x) \le \sup\{|R(x)| \ : \ x \in V\} \ .$$

Since $x \in V$ is arbitrary, this completes the proof of (2.2.9). If $\mathscr{O}$ is a norm topology, take V to be the open unit ball in X to conclude from (2.2.9) that $\|L_R\| \le \|R\|$. In the general setting, first set $W = R^{-1}((-1, 1))$ which is an open neighborhood of 0 in X. By Lemma 2.1.7, there exists a balanced neighborhood V of 0 contained in W, and then for this V, $\sup\{|R(x)| \ : \ x \in V\} \le 1$. Then (2.2.9) shows that $V \subset L_R^{-1}(\{\alpha \in \mathbb{C} \ : \ |\alpha| < 1\})$. Thus, L is continuous. Finally, if $R = \Re \circ L$, then $L(x)$ and $L_R(x)$ are complex linear functionals that have the same real parts for all $x \in X$. Hence $L = L_R$. □

2.2.9 THEOREM (The Complex Hahn-Banach Extension Theorem)**.** *Let $(X, \|\cdot\|)$ be a complex normed vector space, and let Y be a subspace of X. Let L be a linear functional on Y such that for some $C < \infty$, $|L(y)| \le C\|y\|$ for all $y \in Y$. Then there exists a linear functional $\widetilde{L}$ on all of X such that for this same C, $|\widetilde{L}(x)| \le C\|x\|$ for all $x \in X$ and such that $\widetilde{L}y = Ly$ for all $y \in Y$.*

Proof. Define $R = \Re \circ L$. By Lemma 2.2.8, $\|R\| = \|L\|$, where the norms are computed on Y. By Theorem 2.2.6, there exists a linear functional $\widetilde{R}$ on $X_\mathbb{R}$ such that $\widetilde{R}(y) = R(y)$ for all $y \in Y$, and such that $\|\widetilde{R}\| = \|R\|$. Then $L_{\widetilde{R}}$ is a complex linear functional on X such that for all $y \in Y$,

$$\Re(L_{\widetilde{R}}(y)) = \widetilde{R}(y) = R(y) = \Re(L(y)) \ .$$

Replacing y by iy, yields equality of the imaginary parts as well. Therefore, $L_{\widetilde{R}}(y) = L(y)$ for all $y \in Y$, and again by Lemma 2.2.8, $\|L_{\widetilde{R}}\| = \|\widetilde{R}\|$. Altogether, $\|L_{\widetilde{R}}\| = \|L\|$. □

2.2.10 THEOREM. *Let $(X, \|\cdot\|)$ be a non-trivial normed vector space. For all $x \in X$, there exists $L_x \in X^*$ such that $\|L_x\| = 1$ and $L_x(x) = \|x\|$.*

Proof. Let $x \in X$, $x \neq 0$, and let Y be the one-dimensional subspace X spanned by x. The general element of Y has the form αx, $\alpha \in \mathbb{C}$. Define a linear functional L on Y by $L(\alpha x) = \alpha\|x\|$. Evidently $\|L\| = 1$ and $L(x) = \|x\|$. Let L_x denote the norm-preserving extension of L to all of X that is provided by the Hahn-Banach Theorem. For $x = 0$, the claim is true, since by the first part, unit vectors L in X^* exist, and any such unit vector will do. □

2.2.11 THEOREM. *Let $(X, \|\cdot\|)$ be a normed vector space. For $x \in X$, let ϕ_x denote the element of X^{**} given by $\phi_x(L) = L(x)$ for all $L \in X^*$. The map $x \mapsto \phi_x$ is an isometric imbedding of X into X^{**}.*

Proof. We have already observed that $x \mapsto \phi_x$ is linear and that $\|\phi_x\| \le \|x\|$. Let L_x be an element of X^* such that $\|L_x\|_* = 1$ and $L_x(x) = \|x\|$. Then $\phi_x(L_x) = L_x(x) = \|x\|$, and hence $\|\phi_x\| \ge \|x\|$. Altogether, we have the isometry $\|\phi_x\| = \|x\|$. □

2.2.12 DEFINITION (Natural isometry, reflexive)**.** The map $x \mapsto \phi_x$ described in Theorem 2.2.11 is called the *natural isometry* embedding X in X^{**}. A Banach space $(X, \|\cdot\|)$ is *reflexive* in case every the natrual isometry is surjective.

Theorem 2.2.11 provides a natural way to *complete* a normed vector space that is not a Banach space: Identify it with its image in the Banach space X^{**} under the isometric map $x \mapsto \phi_x$. The closure of the image in X^{**} is a Banach space in which the isometric image of X is dense.

A number of applications of the Hahn-Banach Theorem will be made it what follows, but a simple alternative proof of Lemma 2.1.30, Riesz's Lemma, is worthwhile to give at this point. Recall that Lemma 2.1.30 says that if $(X, \|\cdot\|)$ be a normed vector space, and Y be a proper, closed subspace of X, then for all $\alpha \in (0,1)$, there exists $u \in X$, $\|u\| = 1$, such that

$$\alpha \le \inf\{\|u - y\| \ : \ y \in Y\} \ .$$

Second Proof of Lemma 2.1.30. Let $x \in X$, $x \notin Y$. Define L on $\mathrm{span}(Y \cup \{x\})$ by $L(y + \alpha x) = \alpha$ for all $y \in Y$, $\alpha \in \mathbb{C}$. Then $\ker(L) = Y$ which is closed in $\mathrm{span}(Y \cup \{x\})$, and hence is bounded on $\mathrm{span}(Y \cup \{x\})$. By the Hahn-Banach Theorem, there is a non-zero element $\widetilde{L}$ of X^* that is zero on all of Y.

By the definition of $\|\widetilde{L}\|_*$, for all $\alpha \in (0,1)$, there is a unit vector $u \in X$ such that $\widetilde{L}(u)$ is real and $\widetilde{L}(u) \ge \alpha \|\widetilde{L}\|_*$. Since for all $y \in Y$,

$$\alpha \|\widetilde{L}\|_* \le \widetilde{L}(u) = \widetilde{L}(u - y) \le \|\widetilde{L}\|_* \|u - y\| \ ,$$

$\inf\{\|u - y\| \ : \ y \in Y\} \ge \alpha$. □

2.2.13 THEOREM. *Let $(X, \|\cdot\|)$ be a reflexive Banach space space. For all $L \in X^*$, there exists a unit vector $x \in X$ such that $L(x) = \|x\|$.*

Proof. By Theorem 2.2.10, there exists $\phi \in X^{**}$ such that $\|\phi\|_{**} = 1$ and $\phi(L) = \|L\|$. Since X is refexive, for some $x \in X$ with $\|x\| = 1$, $\phi = phi_x$. Thus, $L(x) = \phi_x(L) = \|L\|$. □

It is a much deeper theorem of James that the converse of Theorem 2.2.13 is also true: If for every $L \in X^*$ there exists a unit vector $x \in X$ such that $L(x) = \|L\|$, then X is reflexive.

2.2.14 THEOREM. *Let $(X, \|\cdot\|)$ Banach space space. Then X is reflexive if and only if X^{**} is reflexive.*

Proof. Let $f \in X^{***}$. The natrual isometry of X into X^{**} $x \mapsto \phi_x$ is continuous (even isometric), and hence $x \mapsto f(\phi_x)$ is continuous, so that for some $L_f \in X^*$, for all $x \in X$,

$$f(\phi_x) = L_f(x) = \phi_x(L_f) \ . \tag{2.2.10}$$

Suppose that X is reflexive. Then every $\phi \in X^{**}$ has the form ϕ_x for some $x \in X$, and then (2.2.10) becomes

$$f(\phi) = \phi(L_f) \ . \tag{2.2.11}$$

which shows that every element f of X^{***} is an evaluation functional on X^{**}. Hence X^* is reflexive.

For the converse, suppose that X^* is reflexive. If X is not reflexive, the image of X under the isometric embedding $x \mapsto \phi_x$ is a proper closed subspace V of X^{**}, and then by the Hahn-Banach Theorem, there exists a non-zero $f \in X^{***}$ that vanishes on V. Since X^* is reflexive, there exists $L_f \in X^*$ such that (2.2.11) is true for all $\phi \in X^{**}$. But then

$$0 = f(\phi_x) = L_f(\phi_x) = L_f(x)$$

for all x, so that $L_f = 0$, and hence $f = 0$. This contradiction shows that V cannot be a proper subspace of X^{**}.

□

2.2.4 The weak and weak-$*$ topologies

2.2.15 DEFINITION. Let $(X, \|\cdot\|)$ be a normed space, and let $(X^*, \|\cdot\|_*)$ be its dual. For $x \in X$, let $\phi_x \in X^{**}$ be given by $\phi_x(L) = L(x)$ for all $L \in X^*$. The *weak topology on* X is the weakest topology on X under which each $L \in X^*$ is continuous. The *weak-$*$ topology on* X^* is the weakest topology on X under which each of the functionals ϕ_x, $x \in X$, is continuous

By Theorem 2.1.32, the weak topology makes X a topological vectors space, and the weak-$*$ topology makes X^* a topological vector space. By Theorem 2.1.33, if X^* is separable, then the relative weak topology is metrizable on bounded subsets of X, and if X is separable, then the relative weak-$*$ topology is metrizable on bounded subsets of X^*

The sets

$$V_{L_1,\dots,L_n,\epsilon} = \bigcap_{j=1}^{n} \{x \in X \ : \ |L_j(x)| < \epsilon\} \tag{2.2.12}$$

where $\{L_1. \dots, L_n\}$ is a (finite) subset of X^* and $\epsilon > 0$, constitute a neighborhood base at 0 for the weak topology on X. It is left as a simple exercise to show that in fact one may require that $\{L_1. \dots, L_n\}$ be linearly independent, and then this smaller set of neighborhoods is still a base at 0. Partly forr this reason, the following lemma will be useful:

2.2.16 LEMMA. *Let $(X, \|\cdot\|)$ be a normed space. Let $\{L_1, \dots, L_n\}$ be a linearly independent subset of X^*. Then there exists a set $\{x_1, \dots, x_n\}$ of X such that $L_i(x_j) = \delta_{i,j}$ for all $1 \le i, j \le n$ (and therefore linearly independent). Moreover, defining*

$$Z := \bigcap_{j=1}^{n} \ker(L_j) \qquad \text{and} \qquad W := \mathrm{span}(\{x_1, \dots, x_n\}) \ , \tag{2.2.13}$$

$X = Z \oplus W$.

Proof. If $n = 1$ the claim is trivial. Suppose the claim is true for all linearly independent sets of $n-1$ vectors in X^*, and let $\{y_1, \dots, y_{n-1}\}$ be such that $L_i(y_j) = \delta_{i,j}$ for all $1 \le i, j \le n-1$. If $\sum_{j=1}^{n-1} \alpha_j y_j = 0$, then for each $k = 1, \dots, n-1$, $\alpha_k = L_k(\sum_{j=1}^{n-1} \alpha_j y_j) = 0$, and so $\{y_1, \dots, y_{n-1}\}$ is linearly independent. For all $x \in X$,

$$L_n(x) = L_n\left(x - \sum_{j=1}^{n-1} L_j(x) y_j\right) + \sum_{j=1}^{n-1} L_n(y_j) L_j(x) \ ,$$

and since $x - \sum_{j=1}^{n-1} L_j(x) y_j$ belongs to $\bigcap_{k=1}^{n-1} \ker(L_k)$, if it were the case that

$$\bigcap_{k=1}^{n-1} \ker(L_k) \subset \ker(L_n) \ ,$$

then it would follow that $L_n - \sum_{j=1}^{n-1} L_n(y_j)L_j = 0$, and this would contradict the linear independence of $\{L_1, \ldots, L_n\}$. Hence there exists $x_n \in \bigcap_{k=1}^{n-1} \ker(L_k)$ such that $L_n(x_n) = 1$. For each $j = 1, \ldots, n-1$, define $x_j = y_j - L_n(y_j)x_n$. One readily checks that $L_i(x_j) = \delta_{i,j}$ for all $1 \le i, j \le n$.

Finally, for all $x \in X$, $x = \left(x - \sum_{j=1}^{n} L_j(x)x_j\right) + \left(\sum_{j=1}^{n} L_j(x)x_j\right)$, and the first term on the right belongs to Z, and the second to W. If $x \in Z \cap W$, $x = \sum_{j=1}^{n} \alpha_j$ and for each k, $\alpha_k = L_k(x) = 0$ and hence $x = 0$. Therefore, $X = Z \oplus W$. □

One reason it is often useful to consider the weak-$*$ topology on X^* is on account of the following Theorem:

2.2.17 THEOREM (Alaoglu's Theorem)**.** *Let $(X, \|\cdot\|)$ be a normed space, and let B denote the closed unit ball in X^*. Then B is compact in the weak-$*$ topology.*

Proof. For each $x \in X$, define $D_x = \{z \in \mathbb{C} \ \ |z| \le \|x\|\}$ which is a compact subset of $\mathbb{C}$. By Tychonov's Theorem, $\mathscr{D} := \prod_{x\in X} D_x$, which consists of all complex values functions ϕ on X such that for each $x \in X$, $\phi(x) \in D_x$, is compact in the product topology. Let $\mathscr{L}$ denote the subset of $\mathscr{D}$ consisting of linear functions. It is easy to that Ł is a closed subset of $\mathscr{D}$. (The same argument that we applied in the Hilbert space setting can be made here.) Hence $\mathscr{L}$ is compact, and the elements of $\mathscr{L}$ are precisely the elements of B. Moreover, the product topology on $\mathscr{D}$ is precisely the weakest topology that makes all of the evaluation maps $\phi \mapsto \phi(x)$ continuous, but this is precisely the weak-$*$ topology on B. □

2.2.18 COROLLARY (Corollary to Alaoglu's Theorem)**.** *Let $(X, \|\cdot\|)$ be a reflexive Banach space, B denote the closed unit ball in X. Then B is compact in the weak topology.*

Proof. Since X is isometric with X^{**} under the natural embedding, the weak topology on X is the same as the weak-$*$ topology on X^{**}. □

If $(X, \|\cdot\|)$ is reflexive, then the closed unit ball in X is weakly compact since we may then identify the weak topology on X with the weak-$*$ topology on X^{**}. But when $(X, \|\cdot\|)$ is reflexive, the unit ball in X need not be weakly compact.

While the weak topology on a Banach space $(X, \|\cdot\|)$ is weaker than the norm topology, and strictly so when X is infnite dimensional (since then the norm topology is not metrizable), every weakly continuous function is norm continuous, but not *vice-versa*. However, by the very definition of the weka topology, every norm continuous linear functional is weakly continuous.

There is an important reult of this type for subsets of X: While the class of weakly open sets is strictly smaller than the class of norm closed sets, every norm closed convex set is also weakly closed.

2.2.19 DEFINITION (Half-space)**.** Let $(X, \|\cdot\|)$ be a Banach space. A set $H \subset X$ is a *closed half-space of* X in case for some $a \in \mathbb{R}$ and $L \in X^*$,

$$H = \{\ x \in X \ :\ \Re(L(x)) \ge a\ \}\ . \tag{2.2.14}$$

By the definition of the weak topology, the functions $x \mapsto \Re(L(x))$ is weakly continuous, and hence H is weakly closed, and therefore norm closed.

2.2.20 THEOREM (Mazur's Theorem)**.** *Let $(X, \|\cdot\|)$ be a Banach space. Every norm closed convex set $K \subset X$ is the intersection of the half-spaces containing K. In particular, every norm closed convex set K is also weakly closed.*

Proof. Regard $(X, \|\cdot\|)$ as a real Banach space $(X_{\mathbb{R}}, \|\cdot\|)$. Let $z \in K^c$. Then $\{z\}$ is compact and convex in a trivial manner, while K is closed and convex, and $\{z\} \cap K = \emptyset$. By the Real Hahn-Banach Separation Theorem, there is a continuous linear functional R_z on $(X_{\mathbb{R}}, \|\cdot\|)$ such that and an $\epsilon > 0$ such that for all $y \in K$,

$$R_z(z) + \epsilon \leq 1 \leq R_z(y) \ .$$

Thus, $H_z := \{\ x \in X \ \ : \ \ R_z(x) \geq 1\ \}$ contains K, and does not contain z. By Murray's Lemma. Lemma 2.2.8, there exists a complex contiuous linear functional on $(X, \rangle)$ such that $R = \Re \circ L$, and hence $H_z\{\ x \in X \ : \ \Re(L_z(x)) \geq 1\ \}$, which is manifestly closed. Thus, $K = \bigcap_{z \in K^c} H_z$, and this displays K as the intersection over a family of weakly closed sets. $\square$

In particular, the norm closed unit ball in a Banach space X is also weakly closed. The unit ball in its dual X^* is closed in the weak-$*$ topology since the weak-$*$ topology is Hausdorff, and it is weak-$*$ compact by Alaglu's Theroem. However, one can give a more elementary proof of this fact as follows: Let B be the unit ball in X^*. If $L \in X^{**}$ and $L \notin B$, then there exists a unit vector $z \in L$ such that $\Re(L(z)) > 1$. But $|M(z)| \leq 1$ for all $M \in B$. Hence

$$H_L := \{M \in X^* \ : \ \Re(M(x)) \leq 1\} = \{M \in X^* \ : \ \Re(\phi_x(M)) \leq 1\}$$

is weak-$*$ closed, contains B, and does not contain L. Therefore,

$$B = \bigcap_{L \notin B} H_L$$

is weakly closed.

2.2.21 THEOREM. *Let $(X, \|\cdot\|_X)$ and $(Y, \|\cdot\|_Y)$ be Banach spaces, with X reflexive. Let $T \in \mathscr{B}(X, Y)$. Let $K \subset X$ be convex, bounded and norm closed. Then $T(K)$ is convex, bounded and norm closed in Y.*

Proof. By Mazur's Theorem, K is a weakly closed subset of rB, for some $r > \infty$ where B is the closed unit ball in X. By Corollary 2.2.18 to Alaoglu's Theorem, K is weakly compact. By Theorem 2.2.27, T is weak-weak contiuous, and hence $T(K)$ is weakly compact in Y. Since the weka topology is Hausdorff, $T(K)$ is weakly closed, and therefore, norm closed. A linear image of a convex set is evidently convex, so that $T(K)$ is convex, and $\|Tx\|_Y \leq \|T\|\|x\|_X$ for all $x \in K$, so that $T(K)$ is bounded. $\square$

2.2.5 The uniform boundedness principle and related theorems

2.2.22 THEOREM. *Let $(X, \|\cdot\|_X)$ be a Banach space, and let $(Y, \|\cdot\|_Y)$ be a normed space. Let $\mathscr{T} \subset \mathscr{B}(X, Y)$. For each T in $\mathscr{T}$, let $\|T\|$ denote the operator norm of T. Suppose that for all $x \in X$,*

$$\sup_{T \in \mathscr{T}} \{\|Tx\|_Y\} < \infty \ . \qquad (2.2.15)$$

Then

$$\sup_{T \in \mathscr{T}} \{\|T\|\} < \infty \ . \qquad (2.2.16)$$

Proof. For $n \in \mathbb{N}$, define $E_n = \bigcap_{T\in\mathscr{T}} \{x \in X \ : \|Tx\|_Y \le n\} = \{x \in X \ : \ \sup_{T\in\mathscr{T}}\{\|Tx\|_Y\} \le n\}$. Since for each T, $\{x \in X \ : \|Tx\|_Y \le n\}$ is closed, E_n is closed, and by (2.2.15), $\bigcup_{n\in\mathbb{N}} E_n = X$. By Baire's Theorem, for some n, the interior of E_n is non-empty.. Hence for some $n \in \mathbb{N}$, $r > 0$ and $x_0 \in X$, $B(r, x_0) \subset E_n$. That is, if $\|x\|_X \le 1$, and $0 \le s < r$, $x_0 + sx \in E_n$. Therefore, for all $T \in \mathscr{T}$,

$$s\|Tx\|_Y = \|T(sx - x_0) + Tx_0\|_Y \le n + \|Tx_0\| \ .$$

Define $C = \sup_{T\in\mathscr{T}}\{\|Tx_0\|_Y\}$ which is finite by (2.2.15). Then $\|Tx\|_Y \le (n + C)/s$ for all x with $\|x\|_X \le 1$, and this means that $\|T\| \le (n + C)/r$, showing that $\sup_{T\in\mathscr{T}}\{\|T\|\} \le (n + C)/r$. □

2.2.23 COROLLARY. *Let $(X, \|\cdot\|_X)$ be a Banach space, and let $(Y, \|\cdot\|_Y)$ be a normed space. Let T be a linear transformation from X to Y such that $L \circ T \in X^*$ for all $L \in Y^*$. Then $T \in \mathscr{B}(X, Y)$.*

Proof. Define $\mathcal{A} = \{L \circ T \ : \ L \in Y^* \ , \ \|L\|_{Y^*} = 1\}$, which, by hypothesis, is set of continuous linear transformations form $(X, \|\cdot\|_X)$ to $(\mathbb{C}, |\cdot|)$, By the definition of $\|\cdot\|_{Y^*}$,

$$\sup_{L\circ T\in\mathcal{A}}\{|L(T(x))|\} \le \|T(x)\|_Y\} < \infty \ .$$

By the Uniform Boundedness Principle, $C := \sup_{f\circ T\in\mathcal{A}}\{\|L \circ T\|_{Y^*}\} < \infty$. Thereofre for all unit vectors $x \in X$ and $L \in Y^*$, $|L(T(x))| \le C$, and this shows that $\|T\| \le C$. □

2.2.24 DEFINITION (Weakly bounded)**.** Let $(X, \|\cdot\|_X)$ be a Banach space. $A \subset X$ is *weakly bounded* in case for every weak neighborhood U od 0, there is an $r < \infty$ such that $A \subset rU$.

2.2.25 COROLLARY. *$(X, \|\cdot\|_X)$ be a Banach space. $A \subset X$ is wekaly bounded if and only if it is norm bounded.*

Proof. Suppose A is norm bounded: For some $C < \infty$, $\|x\| \le C$ for all $x \in A$. Let U be any weakly open set containing 0. Then U contain $V_{L_1,\dots,L_n,\epsilon}$ for some $\{L_1, \dots, L_n\} \subset X^*$ and $\epsilon > 0$. Let $m := \max\{\|L_1\|_*, \dots, \|L_1\|_*\}$. Then for all $x \in A$, $j = 1, \dots, n$, $|L_j(x)| \le \|L_j\|_*\|x\| \le mC$. Therefore, if $r > mC/\epsilon$, $x \in V_{L_1,\dots,L_n,\epsilon} \subset U$. Thus, A is weakly bounded.

Conversely, let $x \mapsto \phi_x$ be the natural embedding of X into X^{**}. Suppose A is weakly bounded. Then for each $L \in X^*$, there is some $r_L < \infty$ so thet $A \subset r_L V_{L,1}$, which is the same as $|\phi_x(L)| = |L(x)| \le r_L$ for all $x \in A$. That is,

$$\sup_{x\in A} |\{|\phi_x(L)|\} = \sup_{x\in A} |\{|L(x)|\} < \infty \ .$$

By the Uniform Boundedness Principle, $\sup_{x\in A} |\{\|\phi_x\|\} < \infty$, but, as a consequence of the Hahn-Banach Theorem, for each x, $\|\phi_x\| = \|x\|$. Thus, weak boundedness implies strong boundedness. □

Corollary 2.2.25 says that in a Banach space, where one might expect two distinct notions of boundedness – norm boundedness and weak boundedness – there is onty one. There are also fewer dicrinct classes of linear transformations than one might expect.

2.2.26 DEFINITION (Mixed continuity)**.** Let $(X, \|\cdot\|_X)$ and $(Y, \|\cdot\|_Y)$ be Banach spaces. Let T be a linear transformation from X to Y

(1) T is *weak-weak* continuous in case T is continuous when X and Y are equipped with their weak topologies.

(2) T is *strong-strong* continuous in case T is continuous when X and Y are equipped with their norm topologies.

(3) T is *strong-weak* continuous in case T is continuous when X is equipped with its norm topology and Y is equipped with its weak topology.

(4) T is *weak-strong* continuous in case T is continuous when X is equipped with its weak topology and Y is equipped with its norm topology.

The next theorem says that the first three types of continuity are the same, and that the fourth type of continuity is extremely resrictive in an infinite dimensional setting. The key step is to prove that strong-weak continuity implies strong-strong continuity, but this follows easily from Corollaray 2.2.23, and hence this thoerem is another fairly direct conseqeunce of the Uniform Boundedness Principle.

2.2.27 THEOREM. *$(X, \|\cdot\|_X)$ be a Banach space, and let $(Y, \|\cdot\|_Y)$ be a normed space. Let T be a linear transformation from X to Y. Then the following are equivalent:*

(1) T is strong-strong *continuous.*

(2) T is weak-weak *continuous.*

(3) T is strong-weak *continuous.*

Moreover, T is weak-strong *continuous if and only if for some $n \in \mathbb{N}$, there exists $\{L_1, \ldots, L_n\} \subset X^*$ and $\{y_1, \ldots, y_n\} \subset X$ such that for all $x \in X$,*

$$Tx = \sum_{j=1}^{n} (L_j(x))y_j \ . \tag{2.2.17}$$

Proof. Suppose that T is strong-strong continuous, and therefore bounded. Let $\{L_1, \ldots, L_n\}$ be any finite subset of Y^*, and let $\epsilon > 0$. We must show that $T^{-1}(V_{L_1,\ldots,L_n,\epsilon})$ is weakly open. Note that since $|L_j(T(x))| = |T^*L_j(x)|$, $\max_{1\le j\le n}\{|L_j(T(x)|\} < \epsilon \iff \max_{1\le j\le n}\{|T^*L_j(x)|\} < \epsilon$. Therefore,

$$T^{-1}(V_{L_1,\ldots,L_n,\epsilon}) = V_{T^*L_1,\ldots,T^*L_n,\epsilon} \ ,$$

which is open. Thus whenwever T is strong-strong continuous, T is weak-weak continuous.

Since the norm topology is stronger than the weak topology, weak-weak continuity implies strong weak continuity.

Suupose that T is strong-weak continuous. Since every $L \in X^*$ is weakly continuous by the definition of the weak topology, $L \circ T$ is continuous on X equipped with is norm topology. That is, $L \circ T \in X^*$. Then by Corollary 2.2.23, T is bounded, and hence strong-strong continuous. This completes the proof of the equivalence of *(1)*, *(2)* and *(3)*.

Finally, suppose that T is weak-strong continuous. Then for some $\{L_1, \ldots, L_n\} \subset Y^*$ and some $\epsilon > 0$,

$$V_{L_1,\ldots,L_n,\epsilon} \subset T^{-1}(B_Y(1,0)) \ .$$

Let $Z = \bigcap_{j=1}^n \ker(L_j)$, and note that $Z \subset V_{L_1,\ldots,L_n,\epsilon}$. If $z \in Z$, then $tz \in Z$ for all $t \in \mathbb{R}$. Then $|t|\|Tz\|_Y = \|T(tz)\|_Y < 1$ for all t, and this means that $Tz = 0$. We may assume without loss of generality that $\{L_1, \ldots, L_n\}$ is linearly independent. By Lemma 2.2.16, there exists a set $\{x_1, \ldots, x_n\}$ such that $L_j(x_j) = \delta_{i,j}$ for $1 \leq i,j \leq n$. Then for all $x \in X$, $x - \sum_{j=1}^n L_j(x)x_j \in Z$, and hence $T(x) = T\left(\sum_{j=1}^n L_j(x)x_j\right) = \sum_{j=1}^n (L_j(x))Tx_j$. This proves (2.2.17) with $y_j = Tx_j$ for $j = 1, \ldots, n$.

Conversely, any operator of the form (2.2.17) is weak-strong continuous since each L_j is weakly continuous, and scalar multiplication continuous and vector addition are norm continuous. □

2.2.28 THEOREM. *Let $(X, \|\cdot\|_X)$ and $(Y, \|\cdot\|_Y)$ be Banach spaces, with X reflexive. Let $T \in \mathscr{B}(X,Y)$. Let $K \subset X$ be convex, bounded and norm closed. Then $T(K)$ is convex, bounded and norm closed in Y.*

Proof. By Mazur's Theorem, K is a weakly closed subset of rB, for some $r > \infty$ where B is the closed unit ball in X. By Corollary 2.2.18 to Alaoglu's Theorem, K is weakly compact. By Theorem 2.2.27, T is weak-weak contiuous, and hence $T(K)$ is weakly compact in Y. Since the weka topology is Hausdorff, $T(K)$ is weakly closed, and therefore, norm closed. A linear image of a convex set is evidently convex, so that $T(K)$ is convex, and $\|Tx\|_Y \leq \|T\|\|x\|_X$ for all $x \in K$, so that $T(K)$ is bounded. □

2.2.29 THEOREM. *Let $(X, \|\cdot\|)$ be a Banach space. Let $\{x_n\}_{n\in\mathbb{N}}$ be a sequence in X. Then $\{x_n\}_{n\in\mathbb{N}}$ converges weakly to $x \in X$ if and only if for all $L \in X^*$,*

$$\lim_{n\to\infty} L(x_n) = L(x) \ . \tag{2.2.18}$$

Moreover, if $\{x_n\}_{n\in\mathbb{N}}$ is wekly convergent, then

$$\sup_{n\in\mathbb{N}}\{\|x_n\|\} < \infty \ . \tag{2.2.19}$$

Proof. By definition, $\{x_n\}_{n\in\mathbb{N}}$ converges weakly to $x \in X$ if and only if for every weak neighborhood V of 0, $x_n \in x + V$ for all but finitely many $n \in \mathbb{N}$. In particular, for all $L \in X^*$ and all $\epsilon > 0$, $x_n - x \in V_{L,\epsilon}$, with is the same thing as $|L(x_n) - L(x)| < \epsilon$ for all but finitely many n, so that (2.2.18) is valid.

Conversely if (2.2.18) is valid for all L, then for any $\{L_1, \ldots, L_n\} \subset X^*$ and any $\epsilon > 0$, $|L_j(x_n - x)| < \epsilon$ for all but finitely many n, and thus

$$x_n - x \in \bigcap_{j=1}^n V_{L_j,\epsilon} = V_{L_1,\ldots,L_n,\epsilon}$$

for all but finitely many n. Since the sets $V_{L_1,\ldots,L_n,\epsilon}$ are a neighborhood basis at 0 for the weak topology, this means that $\{x_n\}_{n\in\mathbb{N}}$ converges weakly to x.

Finally, by what we just proved, if $\{x_n\}_{n\in\mathbb{N}}$ converges weakly to x, then $\{L(x_n)\}_{n\in\mathbb{N}}$ is a Cauchy sequence in $\mathbb{C}$, and hence is bounded, so that for all $L \in X^*$, there exists $C_L < \infty$ such that $\sup n \in \mathbb{N}\{L(x_n)\} = \sup n \in \mathbb{N}\{\phi_{x_n}(L)\} \leq C_L$. By the Unifrom Boundednes Principle, $\sup n \in \mathbb{N}\{\|\phi_{x_n}\|\} < \infty$, and as a consequence of the Hahn-Banach Theorem, $\|\phi_{x_n}\| = \|x_n\|$, thus proving (2.2.19). □

2.2.30 THEOREM (Open Mapping Theorem)**.** *Let $(X, \|\cdot\|_X)$ and $(Y, \|\cdot\|_Y)$ be Banach spaces. Let $T \in \mathscr{B}(X,Y)$ be surjective. Then T is an open mapping. In particular, when $T \in \mathscr{B}(X,Y)$ is injective as well as surjective, its inverse belongs to $\mathscr{B}(Y,X)$.*

Proof. Let $B_X(r,0)$ denote the open unit ball of radius $r > 0$ in X, and let $B_Y(r,0)$ denote the open unit ball of radius $r > 0$ in Y, Since $\{B_X(r,0)\ \ r > 0\}$ is a neighborhood base at the origin for the topology on X, and $\{B_Y(r,0)\ \ r > 0\}$ is a neighborhood base at the origin for the topology on Y, if for each $r > 0$, there is an $s > 0$ so that $B_Y(s,0) \subset T(B_X(r,0))$, then T maps open sets in X to open sets in Y. moreover, by homogeneity, to show that T has this property, it suffices to show that there is some $s > 0$ so that $B_Y(s,0) \subset T(B_X(1,0))$.

Since $X = \cup_{n\in\mathbb{N}} B_X(n,0)$, and since T is surjective,

$$Y = \bigcup_{n\in\mathbb{N}} \overline{T(B_X(n,0))} = \bigcup_{n\in\mathbb{N}} n\overline{T(B_X(1,0))}\ .$$

By Baire's Theorem, for some n, the interior of $n\overline{T(B_X(1,0))}$ is non-empty. Since these sets are a homeomorphic to one another, each of them has a non-empty interior, and hence $\overline{T(B_X(1,0))} \neq \emptyset$.

Thus, for some $y_0 \in Y$, and some $t > 0$, $B_Y(t,y_0) \subset \overline{T(B_X(1,0))}$. In other words, of $\|y - y_0\|_Y < t$, then $y - y_0 \in \overline{T(B_X(1,0))}$. In particular, $y_0 \in \overline{T(B_X(1,0))}$ Then since $\overline{T(B_X(1,0))}$ is convex,

$$\frac{1}{2}y = \frac{1}{2}(y - y_0) + \frac{1}{2}y_0 \in \overline{T(B_X(1,0))}\ .$$

This shows that $B_Y(t/2,0) \subset \overline{T(B_X(1,0))}$.

What we have proved so far shows that for all $y \in B_Y(t/2,0)$ and all $\epsilon > 0$, there is an $x \in B_X(1,0)$ such that $\|y - Tx\|_Y < \epsilon$. By homogeneity, we have that for all $r > 0$,

$$\text{for all } \epsilon > 0\ ,\ \|y\|_Y < r \quad \Rightarrow \quad \text{there exists } x \in B_X(2r/t,0) \text{ such that } \|y - Tx\|_Y < \epsilon\ . \tag{2.2.20}$$

Pick $y \in B_Y(t/4,0)$, and then pick $x_1 \in B_X(1/2,0)$ so that $\|y - Tx_1\|_Y \le t/8$. Now apply (2.2.20) with $y - Tx_1$ in place of y, and choose $x_2 \in B_X(1/4,0)$ such that $\|(y_1 - Tx_1) - Tx_2\| \le t/16$. Proceeding inductively, we construct an infinite sequence $\{x_n\}_{n\in\mathbb{N}}$ such that for all n, $x_n \in B_X(2^{-n-1},0)$ and $\|y - T(\sum_{j=1}^n x_j)\|_Y \le t2^{-n-2}$. Note the $\sum_{j=1}^\infty x_j$ coverages to an element $x \in B_X(1,0)$, and then $\|y - Tx\|_Y = \lim_{n\to\infty} \|y - T(\sum_{j=1}^n x_j)\|_Y = 0$. Thus $y \in T(B_X(1,0))$ whenever $y \in B_Y(t/4,0)$ which means that $B_Y(t/4,0) \subset T(B_X(1,0))$.

The final statement is clear since T^{-1} is continuous if and only if whenever U is open in X, $(T^{-1})^{-1}(U) = T(U)$ is open in Y. □

2.2.31 DEFINITION. Let $(X, \|\cdot\|_X)$ and $(Y, \|\cdot\|_Y)$ be normed vector spaces, and let T be a linear transformation from X to Y, not necessarily bounded. The graph of T, $\Gamma(T)$, is the subset of $X \times Y$ given by

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

The norm $\|(x,y)\| := \|x\|_X + \|y\|_Y$ makes $X \times Y$ into a normed vector space with the obvious rules for vector addition and scalar multiplication. A linear operator T from X to Y is said to be *closed* in case $\Gamma(T)$ is norm closed in $X \times Y$.

Beware the terminology, which is standard: while a map T is open if and only if the image under T of every open set is open, to say that T is closed does not mean that the image under T of every closed set is closed.

2.2.32 THEOREM (Closed Graph Theorem)**.** *Let $(X, \|\cdot\|_X)$ and $(Y, \|\cdot\|_Y)$ be Banach spaces. Let T be a linear transformation from X to Y. If T is closed, then $T \in \mathscr{B}(X,Y)$.*

Proof. It is easy to see that $X \times Y$ is a Banach space in the norm $\|(x,y)\| = \|x\| + \|y\|$. Then since $\Gamma(T)$ is a norm closed subspace of $X \times Y$, it is a Banach space in this norm. The map $(x, Tx) \mapsto x$ is a linear bijection of X with $\Gamma(X)$, and since $\|x\|_X \leq \|x\|_X + \|Tx\|_Y$, it is bounded. (Its operator norm is no greater than 1.) By the Open Mapping Theorem, its inverse, which is the map $x \mapsto (x, Tx)$ is bounded. That is, there is a finite constant C such that $\|x\|_X + \|Tx\|_Y \leq C\|x\|_X$, which certainly implies that the operator norm of T is no greater than C. □

The following example of the use of the Closed Graph Theorem is due to Terry Tao.

2.2.33 THEOREM. *Let $(X, \mathscr{O})$ be a Hausdorff topological vector space, and let $\|\cdot\|_1$ and $\|\cdot\|_2$ be two norms on X such that these norm topologies are both at least as strong as $\mathscr{O}$, and such that X is complete in both norms. Then $\|\cdot\|_1$ and $\|\cdot\|_2$ are equivalent norms.*

Proof. To say that $\|\cdot\|_1$ and $\|\cdot\|_2$ are equivalent norms means that the identity transformation I is bounded from $(X, \|\cdot\|_1)$ to $(X, \|\cdot\|_2)$ and *vice-versa*. Note that $\Gamma(I) = \{(x,x) \ : \ x \in X\}$. Let (x,y) belong to the closure of $\Gamma(I)$. Then there is a sequence $\{x_n\}_{n \in N}$ such that $\lim_{n\to\infty} \|x_n - x\|_1 = 0$ and $\lim_{n\to\infty} \|x_n - y\|_2 = 0$. But then every open set $U \in \mathscr{O}$ that contains x contains x_n for all but finitely n, and every open set $V \in \mathscr{O}$ that contains y contains x_n for all but finitely n. If $U \cap V = \emptyset$, this is impossible, and since $\mathscr{O}$ is Hausdorff, it must be that $x = y$. Hence $(x,y) \in \Gamma(I)$. This $\Gamma(I)$ is closed, and hence I is bounded from $(X, \|\cdot\|_1)$ to $(X, \|\cdot\|_2)$. By symmetry, it is also bounded from $(X, \|\cdot\|_2)$ to $(X, \|\cdot\|_1)$. □

The utility of the closed graph theorem lies in the fact that if we seek to prove continuity of a linear transformation T between two normed space $(X, \|\cdot\|_X)$ and $(Y, \|\cdot\|_Y)$, we must show that whenever $\lim_{n\to\infty} x_n = x$ in X, then $\lim_{n\to\infty} Tx_n = Tx$ in Y. When the normed spaces are Banach spaces, the closed graph theorem reduces our burden to show that if $\lim_{n\to\infty} x_n = x$ in X and $\lim_{n\to\infty} Tx_n = y$, then $y = Tx$. That is we may assume that both sequences converge and need only identify the limit, as in the proof of Theorem 2.2.33. We do not need to prove that $\{Tx_n\}_{n\in\mathbb{N}}$ is convergent.

2.3 Exercises

1. Let $\mathscr{O}$ be the topology on $\mathbb{R}$ generates by the half-open intervals $(x, y]$. (This is clalled the *lower-limit topology*, and $(\mathbb{R}, \mathscr{O})$ is called the *Sorgenfrey line*. Show that $(\mathbb{R}, \mathscr{O})$ is not a topological vector space over $\mathbb{R}$.

2. Let $(X, \mathscr{O})$ be a topological vector space. Show that for all subsets A, B of X, $\overline{A} + \overline{B} \subset \overline{A+B}$. Show that if $C \subset X$ is convex, then $\overline{C}$ abd C° are convex.

3. Let $(X, \mathscr{O})$ be a topological vector space. Let A and B be compact subsets of X. Show that $A+B$ is compact. Let A be closed and B compact. Show that $A+B$ is closed.

4. Let $(X, \mathscr{O})$ be a topological vector space. Let L be a linear map from X to $\mathbb{C}$. Show that T is continuous if and only if $\ker(T)$ is closed.

5. Let $(X, \|\cdot\|)$ be a normed vector space. Show that if X^* is sepearable, then so is X. Show that if $(X, \|\cdot\|)$ is a separable, reflexive Banach space, then so is X^*. Show that in this case, the weak topology on bounded subset of X is metrizable.

6. Let $X = \ell^\infty$, the space of bounded sequences $x = \{x_n\}_{:n\in n}$ which, equipped with the norm $\|x\| := \sup_{n\in\mathbb{N}}\{|x_n|\}$ is a Banach space. Find a sequence $\{L_n\}_{n\in\mathbb{N}}$ in the unit ball of X^* that has no weak-$*$ convergent subsequence, showing that the unit ball B in X^*, which is weak-$*$ compact by Alaoglu's Theorem is not sequentially weak-$*$ compact.

7. Show that the norm function $x \mapsto \|x\|$ is lower semicontinuous on any Banach space $(X, \|\cdot\|)$.

8. Let $(X, \|\cdot\|_X)$ be a normed vector spaces, and let $(Y, \|\cdot\|_Y)$ be a Banach space. Suppose $\{T_n\}_{n\in\mathbb{N}}$ is a uniformly bounded sequence in $\mathscr{B}(X,Y)$; i.e.; $\sup_{n\in\mathbb{N}}\{\|T_n\|\} < \infty$. Suppose there is a dense set $E \subset X$ such that $\{T_n x\}_{n\in\mathbb{N}}$ converges for all $x \in E$. Show that $\{T_n x\}_{n\in\mathbb{N}}$ converges for all $x \in X$.

9. Let $(X, \|\cdot\|)$ be a Banach space. Let X^* be its dual and X^{**} be the dual of X^*. Let B be the closed unit ball in X, andl let B^** be the closed unit ball in X^{**}. The natural isometrix embedding of X into X^{**} idetifies B with a subset of B^{**}. Show that this subset is weak-$*$ dense in B^{**}. Then show that X is reflexive if and only if B, identified with a subset of X^{**} via the natural isometric embedding, is weak-$*$ compact.

10. Let $(X, \|\cdot\|)$ be a reflexive Banach space. Show that for every $L \in X^*$, there exists a unit vector $u \in X$ such that $L(u) = \|L\|$.

11. Let $(X, \|\cdot\|)$ be a reflexive Banach space. Let K be a closed convex subset of X. For any $x_0 \in X$, define the function $F : X \to [0, \infty)$ by $F(x) = \|x - x_0\|$. Show that F is weakly lower semicontinuous on X, and then show that there exists an $x \in K$ such that $F(x) \le F(y)$ for all $y \in K$.

12. Let $X = \mathbb{C}([0,])$ the Banach space of continuous functions f in $[0,1]$ with the norm $\|f\| = \max\{|f(x)| : x \in [0,1]\}$. For $a \in [0,\infty)$, define

$$K_a := \left\{ f \in X \ : \ \int_0^1 f(x)\mathrm{d}x = 1 \ , f(0) = a \right\} \ .$$

(a) Show that for each $a \in \mathbb{R}$, K_a is closed and convex.

(b) For $f \in X$, define $d(f, K_a) = \inf\{\|f - g\| \ : \ g \in K_a\}$. Show that there is no $g \in K_0$ such that $\|g\| = d(0, K_0)$, while there is a unique $g \in K_1$ such that $\|g\| = d(0, K_1)$, and there are ininiftely many $g \in K_2$ such that $\|g\| = d(0, K_2)$.

13. Let $(X, \|\cdot\|)$ be a Banach space. Suppose that $\{x_n\}_{n\in\mathbb{N}}$ is a sequence in X that converges to $x \in X$ in the weak topology. . Suppose that $\{L_n\}_{n\in\mathbb{N}}$ is a sequence in X^* that converges weakly to $L \in X^*$. Is it necessarily the case that $\lim_{n\to\infty} L_n(x_n) = L(x)$? Prove this is true or give a counter-example. Is it necessarily the case that for some strictly increasing sequence $\{n_k\}_{k\in\mathbb{N}}$ in $\mathbb{N}$ that $\lim_{k\to\infty} L_{n_k}(x_{n_k}) = L(x)$? Prove this is true or give a counter-example.

14. Let $(X, \|\cdot\|)$ be a Banach space. The weak-$*$ topology on X^* is no stronger than the weak topology on X^*. When X is reflexive, the two topologoes are the same. Otherwise, the weak-$*$ topology on X^* is strictly weaker than the weak topology on X^*. Here is one way to see this: Show that a linear functional

ϕ on X^* is weak-$*$ continuous if and only if ϕ is in the image of the canonical embedding of X into X^{**}. Thus, when X is not refexive, there are continuous (and hence weakly continuous) linear functionals on X^* that are not weak-$*$ continuous.

15. Let $(X, \|\cdot\|)$ be a Banach space. A set $A \subset X^*$ is *total* in case the only vector $x \in X$ such that $L(x) = 0$ for all $L \in A$ is $x = 0$. Show that if X is separable, then there is a countable total set in X^*. Show that in this case, there is even a sequence $\{L_n\}_{n\in\mathbb{N}}$ of unit vectors in X^* such that for all $x \in X$, $\|x\| = \sup_{n\in\mathbb{N}}\{|L_n(x)|\}$. Show that ℓ^∞ is not separable, but that it does contain a countable total set.

16. Let $(X, \|\cdot\|)$ be a Banach space. Let $A \subset X$ be weakly compact. Show that A is bounded.

17. $(X, \|\cdot\|)$ be a Banach space whose dual X^* contains a countable total set, which, without loss of generlaity may be takesm to be a sequence $\{L_n\}_{n\in\mathbb{N}}$ od unit vectors. Let $A \subset X$ be weakly compact. Show that

$$\rho(x, y) = \sum_{n=1}^{\infty} 2^{-k}|L_n(y - x)|$$

defines a metric on A, and that the identity map on A is continuous from the relative weak topology on A to the metric topology on A. Then, using the fact that A is weakly compact, show that the identity map is also open, and hence that the relative weak topology and the metric topology coincide.

18. $(X, \|\cdot\|)$ be a Banach space, and let $A \subset X$ be weakly compact. Let $\{x_n\}_{n\in\mathbb{N}}$ be any sequence in A. Let Y be the norm-closed span of $\{x_n\}_{n\in\mathbb{N}}$, which is a separable subspace of X. Show that $A \cap Y$ is weakly compact, and then using the separability of Y, that the relative weak topology on $A \cap Y$ is metrizable, so that $A \cap Y$ is weakly sequentially compact. Conclude that in any Banach space $(X, \|\cdot\|)$, every weakly compact set A is weakly sequentially compact.

Chapter 3

Hilbert Space

3.1 Hilbert Space

3.1.1 Inner product spaces

The modern definition of a *Hilbert space* was given by John von Neumann in 1929 during the course of his work on the mathematical foundations of the then new quantum theory. He had gone to Götingen to work with Hilbert on problems in mathematical logic. When he arrived, he was drawn to a seminar on quantum theory, and embarked on a new direction.

3.1.2 Inner product spaces

3.1.1 DEFINITION (Sesqilinear form)**.** Let $\mathcal{H}$ be a complex vector space. A *sesqilinear form* on $\mathcal{H}$ is a function on $\mathcal{H} \times \mathcal{H}$ with values in $\mathbb{C}$ such that for fixed $f \in \mathcal{H}$, $g \mapsto \langle f, g\rangle$ is a linear functional on $\mathcal{H}$, and such that for all $f, g \in \mathcal{H}$, $\langle g, f\rangle = \overline{\langle f, g\rangle}$ The sesqilinear form is *positive definite* in case $\langle f, f\rangle > 0$ for all $f \neq 0$.

3.1.2 DEFINITION (Inner product space)**.** An *inner product space* $(\mathcal{H}, \langle\cdot,\cdot\rangle)$ is a complex vector space $\mathcal{H}$ equipped with a positive definite sesqilinear form $\langle\cdot,\cdot\rangle$ on $\mathcal{H} \times \mathcal{H}$. The *norm* $\|f\|$ of a vector $f \in \mathcal{H}$ is defined by

$$\|f\| = \sqrt{\langle f, f\rangle}\ . \tag{3.1.1}$$

The example behind these definitions is ℓ^2, the space of complex valued square-summable sequences where for two such sequences f, g,

$$\langle f, g\rangle = \sum_{n=1}^{\infty} \overline{f(n)} g(n)\ ,$$

which was basic to Hilbert's theory of "infinite matrices".

The fundamental theorem concerning inner product spaces is that the Cauchy-Schwarz inequality is satisfied:

3.1.3 THEOREM (Cauchy-Schwarz)**.** *Let $\mathcal{H}$ be a vector space, and let $\langle\cdot,\cdot\rangle$ be an inner product, possibly degenerate, on $\mathcal{H}$. Then for all $f,g \in \mathcal{H}$.*

$$|\langle f,g\rangle|^2 \le \langle f,f\rangle\langle g,g\rangle \tag{3.1.2}$$

and, when $\langle\cdot,\cdot\rangle$ is non-degenerate, so that $(\mathcal{H},\langle\cdot,\cdot\rangle)$ is an inner product space, there is equality in (3.1.2) if and only if $\{f,g\}$ is linearly dependent.

Proof. If either $f=0$ or $g=0$, equality holds in (3.1.2). Suppose that this is not the case. Then $\|f\| \neq 0$ and $\|g\| \neq 0$, and we may define $u = \|f\|^{-1}f$ and $v = e^{i\theta}\|g\|^{-1}g$ for $\theta \in [0,2\pi)$ to be chosen later. Then $\|u\| = \|v\| = 1$ and hence

$$\|u-v\|^2 = \langle u-v, u-v\rangle = \|u\|^2 + \|v\|^2 + 2\Re(\langle u,v\rangle) = 2(1+\Re(\langle u,v\rangle)) \ .$$

Now choose θ so that $\Re(\langle u,v\rangle) = |\langle u,v\rangle|$, and hence $|\langle u,v\rangle| = 1 - \dfrac{1}{2}\|u-v\|^2$. Multiplying through by $\|f\|\|g\|$, this becomes

$$|\langle f,g\rangle| \le \|f\|\|g\| - \frac{1}{2}\|\left(\|g\|f - e^{i\theta}\|f\|g\right)\|^2 \ . \tag{3.1.3}$$

This proves the inequality and shows, under the assumption that $\|h\| = 0$ only for $h=0$, that equality holds if and only if $\|g\|f = e^{i\theta}\|f\|g$, which is the cases if and only if $\{f,g\}$ is linearly dependent. □

A consequence of the Cauchy-Schwarz Inequality is that the function $f \mapsto \|f\|$ is sub-additive on $\mathcal{H}$. That is, for all $f,g \in \mathcal{H}$, $\|f+g\| \le \|f\| + \|g\|$.

3.1.4 DEFINITION (Unit vectors and orthogonality)**.** A vector u in an inner product space $\mathcal{H}$ is a *unit vector* in case $\|u\| = 1$. Two vectors $f,g \in \mathcal{H}$ are *orthogonal* in case $\langle f,g\rangle = 0$. A subset $\{u_j\}_{j\in\mathscr{J}}$ of $\mathcal{H}$ is *orthonormal* in case for all $j,k \in \mathscr{J}$, $\langle u_j,u_k\rangle = 0$ if $j \neq k$ while $\langle u_j,u_j\rangle = 1$.

By the Cauchy-Schwarz inequality, for any two unit vectors $u,v \in \mathcal{H}$, $\Re(\langle u,v\rangle) \in [-1,1]$, and hence it makes sense to define the angle between two unit vectors in $\mathcal{H}$ to be $\arccos(\Re(\langle u,v\rangle))$. The angle between two non-zero vectors f,g is defined to be the angle between their *normalizations* $\|f\|^{-1}f$ and $\|g\|^{-1}g$; which is consistent with Definition 3.1.4.

Using the the Cauchy-Schwarz Inequality, it is simple to prove that the function $f \mapsto \|f\|$ is sub-additive on $\mathcal{H}$. That is, for all $f,g \in \mathcal{H}$, $\|f+g\| \le \|f\| + \|g\|$, and $f \mapsto \|f\|$ is evidently homogeneous of degree one. Thus, our terminology is consistent, and an inner product space, $(\mathcal{H},\langle\cdot,\rangle)$, equipped with the norm (3.1.1) is a normed vector sapce:

3.1.5 THEOREM (Minkowski's inequality for inner product spaces)**.** *Let $(\mathcal{H},\langle\cdot,\cdot\rangle)$ be an inner product space. Then for all $f,g \in \mathcal{H}$,*

$$\|f+g\| \le \|f\| + \|g\| \tag{3.1.4}$$

and there is equality in (3.1.4) if and only if $\{f,g\}$ is linearly dependent.

Proof. for any $f,g \in \mathcal{H}$, by the Cauchy-Schwarz inequality,

$$\|f+g\|^2 = \langle f+g, f+g\rangle = \|f\|^2 + \|g\|^2 + 2\Re\langle f,g\rangle \le \|f\|^2 + \|g\|^2 + 2\|f\|\|g\| = (\|f\| + \|g\|)^2 \ .$$

The square root function is strictly monotone, and there is equality above if and only if there is equality in the Cauchy-Schwarz inequality. □

The metric associated to the inner product norm is often called the *inner-product metric*. We know for any two normed vectors spaces $(X, \|\cdot\|_X)$ and $(Y, \|\cdot\|_Y)$, a linear transformation T from X to Y is continuous if and only if it is bounded. Let $\mathcal{H}$ and $\mathcal{K}$ be two inner product spaces. Let $\mathscr{B}(\mathcal{H},\mathcal{K})$ denote the set of all bounded linear transformations from $\mathcal{H}$ to $\mathcal{K}$, and in the special case $\mathcal{K}=\mathcal{H}$, let $\mathscr{B}(\mathcal{H})$ denote $\mathscr{B}(\mathcal{H},\mathcal{H})$. As always, the function $T \mapsto \|T\|$ is evidently a norm on the vector space $\mathscr{B}(\mathcal{H},\mathcal{K})$ which is called the *operator norm*. Elements of $\mathscr{B}(\mathcal{H},\mathcal{K})$ will often be referred to as *operators*.

There are two important identities that hold in any complex inner product space $\mathcal{H}$, both of which are easily verified by direct computation: The *polarization identity* is

$$\begin{aligned}\langle f,g\rangle &= \frac{1}{4}[\langle f+g,f+g\rangle - \langle f-g,f-g\rangle - i\langle f+ig,f+ig\rangle + i\langle f-ig,f-ig\rangle]\\ &= \frac{1}{4}[\|f+g\|^2 - \|f-g\|^2 - i\|f+ig\|^2 + i\|f-ig\|^2]\ . \end{aligned} \tag{3.1.5}$$

The polarization identity shows that the *correspondence between inner products and norms is one-to-one*: Every inner product defines a norm, and the inner product may be recovered from the norm.

The *parallelogram identity* is

$$\left\|\frac{f+g}{2}\right\|^2 + \left\|\frac{f-g}{2}\right\|^2 = \frac{\|f\|^2+\|g\|^2}{2}\ . \tag{3.1.6}$$

This expresses a quantitative strict convexity property of the function $f \mapsto \|f\|^2$.

3.1.3 Hilbert spaces and the Projection Lemma

3.1.6 DEFINITION (Hilbert Space)**.** A *Hilbert space* is a complex vector space $\mathcal{H}$ equipped with a sesqilinear form $\langle\cdot,\cdot\rangle$ such that $\mathcal{H}$ is complete in its inner product metric. In particular, a Hilbert sapces is a Banach space.

By what has been explained earlier about general normed vector spaces, if $\mathcal{H}$ and $\mathcal{K}$ are both Hilbert spaces, then $\mathscr{B}(\mathcal{H},\mathcal{K})$ is complete in the operator norm, and hence is a Banach space.

The next theorem makes essential use of the completeness of Hilbert space.

3.1.7 THEOREM (Projection Lemma)**.** *Let K be a non-empty closed convex set in a Hilbert space $\mathcal{H}$. Then K contains a unique element of minimal norm. That is, there exists $f_0 \in K$ such that $\|f_0\| < \|f\|$ for all $f \in K$, $f \neq f_0$. Moreover, if $\{f_n\}_{n\in\mathbb{N}}$ is any sequence in K such that*

$$\lim_{n\to\infty} \|f_n\| = \inf\{\|f\| \ :\ f \in K\}\ ,$$

then $\lim_{n\to\infty} \|f_n - f_0\| = 0$.

Proof. Let $D := \inf\{\|f\| \ :\ f \in K\}$. If $D = 0$, then $0 \in K$ since K is closed, and this is the unique element of minimal norm. Hence we may suppose that $D > 0$. Let $\{f_n\}_{n\in\mathbb{N}}$ be a sequence in K such that $\lim_{n\to\infty}\|w_n\| = D$. By the parallelogram identity

$$\left\|\frac{f_m+f_n}{2}\right\|^2 + \left\|\frac{f_m-f_n}{2}\right\|^2 = \frac{\|f_m\|^2+\|f_n\|^2}{2}\ .$$

By the convexity of K, and the definition of D, $\left\|\frac{f_m + f_n}{2}\right\|^2 \geq D^2$ and so

$$\left\|\frac{f_m - f_n}{2}\right\|^2 = \frac{\left(\|f_m\|^2 - D^2\right) + \left(\|f_n\|^2 - D^2\right)}{2} .$$

By construction, the right side tends to zero, and so $\{f_n\}_{n\in\mathbb{N}}$ is a Cauchy sequence. By the completeness of $\mathcal{H}$, $\{f_n\}_{n\in\mathbb{N}}$ is a convergent sequence. Let f_0 denote the limit. By the continuity of the norm, $\|f_0\| = \lim_{n\to\infty}\|f_n\| = D$. Finally, if f_1 is any other vector in K with $\|f_1\| = D$, $(f_0 + f_1)/2 \in K$, so that $\|(f_0+f_1)/2\| \geq D$. Then by the parallelogram identity once more $\|(f_0 - f_1)/2\| = 0$, and so $f_0 = f_1$. This proves the uniqueness. □

3.1.4 Orthogonal complements

As a first application, we discuss *orthogonal complements.*

3.1.8 DEFINITION (Orthogonal complement)**.** Let $\mathcal{H}$ be a Hilbert space and $S \subset \mathcal{H}$. Then $S^\perp$, the *orthogonal complement of* S is the set

$$S^\perp = \bigcap_{g\in S}\{f \in \mathcal{H}\ ;\ \langle g, f\rangle = 0\} . \tag{3.1.7}$$

By the continuity of $f \mapsto \langle g, f\rangle$, for each g, $\{f \in \mathcal{H}\ ;\ \langle g, f\rangle = 0\}$ is closed, and hence $S^\perp$ is closed. Also it is evident that if $f_1, f_2 \in S^\perp$ and $\alpha_1, \alpha_2 \in \mathbb{C}$, for all $g \in S$,

$$\langle \alpha_1 f_1 + \alpha_2 f_2, g\rangle = \alpha_1\langle f_1, g\rangle + \alpha_2\langle f_2, g\rangle = 0 .$$

Hence $S^\perp$ is a subspace of $\mathcal{H}$ for all $S \subset \mathcal{H}$.

3.1.9 THEOREM. *Let $\mathcal{H}$ be a Hilbert space, and let $\mathcal{K}$ be a closed subspace of $\mathcal{H}$. Let $f \in H$. Then there exist unique vectors $f_0 \in \mathcal{K}$ and $f_1 \in \mathcal{K}^\perp$ such that $f = f_0 + f_1$. That is, $\mathcal{H} = \mathcal{K} \oplus \mathcal{K}^\perp$. Finally, define the distances $d(f, \mathcal{K}^\perp)$ and $d(f, \mathcal{K})$ from f to $\mathcal{K}^\perp$ and $\mathcal{K}$ respectively,*

$$d(f, \mathcal{K}^\perp) := \inf\{\|f - g\ :\ g \in \mathcal{K}^\perp\} \quad \text{and} \quad d(f, \mathcal{K}) := \inf\{\|f - g\|\ :\ g \in \mathcal{K}\} . \tag{3.1.8}$$

Then f_1 is the unique vector in $\mathcal{K}^\perp$ such that $\|f - f_1\| = d(f, \mathcal{K}^\perp)$, and f_0 is the unique vector in $\mathcal{K}$ such that $\|f - f_0\| = d(f, \mathcal{K})$.

Proof. We first show that if such a decomposition of f exists, then it is unique. Suppose that $f_0, g_0 \in \mathcal{K}$ and $f_1, g_1 \in \mathcal{K}^\perp$, and that $f = f_0 + f_1 = g_0 + g_1$. Then $f_0 - g_0 = g_1 - f_1$, and $f_0 - g_0 \in \mathcal{K}$ while $g_1 - f_1 \in \mathcal{K}^\perp$, so that $f_0 - g_0$ and $g_1 - f_1$ are orthogonal. Therefore

$$0 = \langle f_0 - g_0, g_1 - f_1\rangle = \langle f_0 - g_0, f_0 - g_0\rangle = \|f_0 - g_0\|^2 ,$$

and hence $f_0 = g_0$, from which it follows that $f_1 = g_1$.

To prove the existence of such a decomposition, Let $K = \{f - g\ :\ g \in \mathcal{K}\}$. Then K is a non-empty closed convex subset of $\mathcal{H}$, and hence it contains a unique element $f - g_0$ of minimal norm. By construction, for all $g \in \mathcal{K}$ and $t \in \mathbb{R}$, $f - (g_0 + tg) \in K$, and so the function $\varphi(t) = \|(f - g_0) - tg)\|^2$ has a maximum at $t = 0$. Differentiating,

$$0 = \varphi'(0) = 2\Re\langle f - g_0, g\rangle .$$

This shows that $f - g_0 \in \mathcal{K}^\perp$ and then $f = (f - g_0) + g_0$ where $f - g_0 \in \mathcal{K}^\perp$ and $g_0 \in \mathcal{K}$.

Finally, for any $g \in \mathcal{K}^\perp$, $\|f - g\|^2 = \|(f - f_1) + (f_1 - g)\|^2 = \|(f - f_1)\|^2 + \|(f_1 - g)\|^2$ since $f - f_1 = f_0 \in \mathcal{K}$ and $f_1 - g \in \mathcal{K}^\perp$. Therefore, $\|f - g\| \geq \|f - f_1\|$ with equality if and only if $f = f_1$. The same reasoning shows that for all $g \in \mathcal{K}$, $\|f - g\| \geq \|f - f_0\|$ with equality if and only if $g = f_0$. □

Let $\mathcal{K}$ be a closed non-zero, proper subspace of a Hilbert space $\mathcal{H}$. For any $f \in \mathcal{H}$, define Pf and $P^\perp f$ to be the unique elements of $\mathcal{K}$ and $\mathcal{K}^\perp$ respectively such that $f = Pf + P^\perp f$. By the uniqueness of Theorem 3.1.9, the transformations $f \mapsto Pf$ and $f \mapsto P^\perp f$ are both linear, and since $\|f\|^2 = \|Pf\|^2 + \|P^\perp f\|^2 \geq \max\{\|Pf\|^2, \|P^\perp f\|^2\}$, it is evident that $\|P\|, \|P^\perp\| \leq 1$. That is, $P, P^\perp \in \mathscr{B}(\mathcal{H})$. (Since $\mathcal{K}$ is a subspace of $\mathcal{H}$, we may regard $\mathscr{B}(\mathcal{H}, \mathcal{K})$ as a subspace of $\mathscr{B}(\mathcal{H})$, and likewise with $\mathscr{B}(\mathcal{H}, \mathcal{K}^\perp)$.) Moreover, since neither $\mathcal{K}$ nor $\mathcal{K}^\perp$ is the zero subspace, there are unit vectors u and v in $\mathcal{K}$ and $\mathcal{K}^\perp$ respectively such that $Pu = u$ and $P^\perp v = v$. Therefore, $\|P\| = \|P^\perp\| = 1$.

3.1.10 DEFINITION. Let $\mathscr{K}$ be a closed non-zero, proper subspace of a Hilbert space $\mathcal{H}$. The bounded linear transformations P and $P^\perp$ such that for all $f \in \mathcal{H}$, $Pf \in \mathcal{K}$, $P^\perp f \in \mathcal{K}^\perp$ and $f = Pf + P^\perp f$ are the *orthogonal projections of $\mathcal{H}$ onto $\mathcal{K}$ and $\mathcal{K}^\perp$ respectively.*

3.2 Duality in Hilbert space

3.2.1 The dual space of an inner product space

3.2.1 DEFINITION (The dual space of an inner product space)**.** Let $\mathcal{H}$ be an inner product space. The *dual space* $\mathcal{H}^*$ of $\mathcal{H}$ is the vector space space of all continuous linear functional on $\mathcal{H}$.

Let L be a continuous linear functional on an inner product space $\mathcal{H}$. As a linear transformation L from the normed space $\mathcal{H}$ to the normed space $\mathbb{C}$, L is continuous if and only if L is bounded, so that the continuity of L is equivalent to the condition that $\|L\|_* < \infty$ where $\|L\|_*$ defined by

$$\|L\|_* = \sup\{|L(u)| \ : \ \|u\| \leq 1\} \ . \tag{3.2.1}$$

By what we have said earlier about bounded linear transformations from one normed space to another, $\|\cdot\|_*$ is a norm on $\mathcal{H}^*$, and therefore $d_*(L, M) := \|L - M\|_*$ defined a metric on $\mathcal{H}^*$.

As an example of a bounded linear functional on an inner product space $\mathcal{H}$, consider any $g \in \mathcal{H}$, and define

$$L_g(f) = \langle g, f\rangle \ . \tag{3.2.2}$$

Then L_g is linear, and by the Cauchy-Schwarz inequality, $|L_g(f)| \leq \|g\|\|f\|$ with equality if $f = g$, showing that $\|L_g\| = \|g\|$. Thus the map $g \mapsto L_g$ is an isometry from $\mathcal{H}$ into $\mathcal{H}^*$. Note that it is not, however, linear, but *sequilinear*: For all $\alpha \in \mathbb{C}$, $L_{\alpha g} = \overline{\alpha} L_g$. We shall soon prove that *when $\mathcal{H}$ is a Hilbert space, every $L \in \mathcal{H}^*$ is of the form $L = L_g$ for some $g \in \mathcal{H}$.*

A number of facts about dial spaces that are, in the general setting of Banach spaces, conseqeunces of the Hahn-Banach Theorem can be proved directly in the Hilber space setting. For example, the fact that the dual of any Banach space separates points is a conseqeunce of the Hahn-Banach Theorem. In Hilbert space, matters are simpler: For all $f, g \in H$, $f \neq g$, $L_{f-g}(f - g) = \langle f - g, f - g\rangle = \|f - g\|^2 > 0$. Thus, $L_{f-g}f \neq L_{f-g}g$.

Likewise, we know that for any Banach space X, and all $x \in X$, there exists $L \in X^*$ such that $\|L\|_* = 1$ and $L(x) = \|x\|$. Again, in general, this is a consequence of the Hahn-Banach Theorem, but when X is a Hilbert space $\mathcal{H}$, we can be more explicit: If $f = 0$, we may take $L = L_g$ for any unit vector $g \in \mathcal{H}$. Otherwise, $f/\|f\|$ is a unit vector in $\mathcal{H}$, and hence $L_{f/\|f\|}$ is a unit vector in $\mathcal{H}^*$. Then

$$L_{f/\|f\|}(f) = \|f\|^{-1}\langle f, f\rangle = \|f\| \ .$$

3.2.2 The Riesz Representation Theorem and the dual space of a Hilbert space

3.2.2 THEOREM (Riesz Representation Theorem)**.** *Let $\mathcal{H}$ be a Hilbert space, and let $L \in \mathcal{H}^*$. There is a unique vector $g_L \in \mathcal{H}$ such that $L(f) = \langle g_L, f\rangle$ for all $f \in \mathcal{H}$, and $\|g_L\| = \|L\|_*$.*

Proof. If $L(f) = 0$ for all $f \in \mathcal{H}$, the assertion is trivial, so suppose that $\|L\|_* > 0$. Define K to be the set

$$K := \{f \in \mathcal{H} \ : \ \Re(L(f)) = \|L\|_* \ \} \ .$$

It is readily checked that this is a closed convex set in $\mathcal{H}$.

If $f \in K$, then $\|L\|_*\|f\| \geq |L(f)| \geq \Re(L(f)) = \|L\|_*$, and hence $\|f\| \geq 1$. On the other hand, by the definition of $\|L\|_*$, there is a sequence of unit vectors $\{u_n\}_{n\in\mathbb{N}}$ such that $|L(u_n)| \to \|L\|_*$. Then choosing $\theta_n \in [0, 2\pi)$ so that $e^{i\theta_n}L(u_n) = |L(u_n)|$, $v_n := e^{i\theta_n}\dfrac{\|L\|_*}{|L(u_n)|}u_n \in K$ and $\|v_n\| \to 1$. Thus, $\inf\{\|v\| \ : \ v \in K\} = 1$.

It now follows from the Projection Lemma that there is a (unique) unit vector $u_0 \in K$ with

$$\|L\|_* = \Re(L(u_0)) \ . \tag{3.2.3}$$

For all $f \in \mathcal{H}$, $\Re(L(f)) \leq |L(f)| \leq \|L\|\|f\|$, when f with $f \neq 0$,

$$\frac{\Re(L(f))}{\|f\|} \leq \|L\|_* = \frac{\Re(L(u_0))}{\|u_0\|} \ .$$

Hence, for any $g \in \mathcal{H}$, the function $\varphi : \left(-\dfrac{1}{2\|g\|}, \dfrac{1}{2\|g\|}\right) \to \mathbb{R}$ defined by $\varphi(t) := \dfrac{\Re(L(u_0 + tg))}{\|u_0 + tg\|}$ has a maximum at $t = 0$. One readily checks that φ is differentiable and computes

$$\varphi'(0) = \Re(L(g)) - \|L\|_*\Re(\langle u_0, g\rangle) \ .$$

Since the left hand side is zero for all g, $\Re(L(g)) = \|L\|_*\Re(\langle u_0, g\rangle)$ for all g. Replacing g by ig, the same is true of the imaginary parts, and so $L(g) = \langle \|L\|_*u_0, g\rangle$ for all g. Thus, $g_L = \|L\|_*u_0$ is such that $L(f) = \langle g_L, f\rangle$ for all $f \in \mathcal{H}$, and $\|g_L\| = \|L\|_*$.

If h_L were any other vector with $L(f) = \langle h_L, f\rangle$ for all $f \in \mathcal{H}$, we would have $\langle g_L - h_L, f\rangle = 0$ for all $f \in \mathcal{H}$, Taking $f = g_L - h_L$, we see that $\|g_L - h_L\|^2 = 0$, and so $h_L = g_L$, proving the uniqueness of g_L. □

The Riesz Representation Theorem allows us to identify a Hilbert space $\mathcal{H}$ with its dual space $\mathcal{H}^*$: That is, the sesquilinear mapping $L \to g_L$ is an isometry from $\mathcal{H}^*$ onto $\mathcal{H}$ whose inverse is the map $g \mapsto L_g$. A consequence of the Riesz Representation Theorem is that $\mathcal{H}$ is reflexive. This is more or less evident since the identification of $\mathcal{H}$ with $\mathcal{H}^*$ extends to identify $\mathcal{H}$ with $\mathcal{H}^{**}$. The fact that the identification of $\mathcal{H}$ with $\mathcal{H}^*$ is sesquilinear dooes nor matter since the composition of two sequilinear maps is linear. All the same, it is perhaps worth spelling out the short argument:

Let $\phi \in \mathcal{H}^{**}$. Then $g \mapsto \overline{\phi(L_g)}$ is a continuous linear funtional on $\mathcal{H}$. By Theorem 3.2.2, there is a unique $h \in \mathcal{H}$ such that for all $g \in \mathcal{H}$,

$$\overline{\phi(L_g)} = \langle h, g\rangle = \overline{\langle g, h\rangle} = \overline{L_g h} \ .$$

Since every $L \in \mathcal{H}^*$ is of the form $L = L_g$, by Theorem 3.2.2 a second time, $\phi(L) = L(h)$ for all $L \in \mathcal{H}^*$, and thus ϕ is an evaluation functional. That is, the canonical *linear* embedding of $\mathcal{H}$ into $\mathcal{H}^{**}$ is surjective.

An elementary but important application of the Riesz Representation Theorem concerns the *adjoint* A^* of an operator $A \in \mathscr{B}(\mathcal{H})$, $\mathcal{H}$ a Hilbert space. Let $A \in \mathscr{B}(\mathcal{H})$, and for $g \in \mathcal{H}$, let L_g be defined as in (3.2.2). Then $L_g \circ A$ is a continuous, and hence it is a bounded linear functional on $\mathcal{H}$.

By the Riesz Representation Theorem, since $L_g \circ A \in \mathcal{H}^*$, there exists a *unique* $h \in \mathcal{H}$ so that for all $f \in \mathcal{H}$,

$$\langle g, Af\rangle = L_g \circ A(f) = \langle h, f\rangle \ .$$

Therefore we may define a function $A^* : \mathcal{H} \to \mathcal{H}$ by defining A^*g to be the unique element of $\mathcal{H}$ such that

$$\langle g, Af\rangle = \langle A^*g, f\rangle \tag{3.2.4}$$

for all $f, g \in \mathcal{H}$.

Again because of the uniqueness, the linearity of A implies that A^* is a linear transformation on $\mathcal{H}$. Moreover, for all unit vectors $u, v \in \mathcal{H}$,

$$|\langle u, A^*v\rangle| = |\langle A^*v, u\rangle| = |\langle v, Au\rangle| \leq \|v\|\|Au\| \leq \|A\| \ .$$

Taking $u = \|A^*v\|^{-1}A^*v$, we obtain $\|A^*v\| \leq \|A\|$. Since v is an arbitrary unit vector in $\mathcal{H}$, this yields

$$\|A^*\| \leq \|A\| \ . \tag{3.2.5}$$

Therefore, $A^* \in \mathscr{B}(\mathcal{H})$.

Taking complex conjugates of both sides of (3.2.4), $\langle f, A^*g\rangle = \langle Af, g\rangle$, and then swapping the roles of f and g,

$$\langle g, A^*f\rangle = \langle Ag, f\rangle \tag{3.2.6}$$

for all $f, g \in \mathcal{H}$. Comparing (3.2.4) and (3.2.6), it is evident that $A^{**} := (A^*)^* = A$ for all bounded linear transformations A. Then by (3.2.5),

$$\|A\| = \|A^{**}\| \leq \|A^*\| \leq \|A\| \ .$$

This means that for all bounded linear transformations $\|A\|$,

$$\|A^*\| = \|A\| \ . \tag{3.2.7}$$

Summarizing, we have proved the following:

3.2.3 THEOREM. *Let $A \in \mathscr{B}(\mathcal{H})$, $\mathcal{H}$ a Hilbert space. Then there exists a unique bounded linear transformation A^* on $\mathcal{H}$ that that (3.2.4) is valid for all $f, g \in \mathcal{H}$. The map $A \mapsto A^*$ is conjugate linear, and satisfies $\|A^*\| = \|A\|$.*

Theorem 3.2.3 says that the map $A \mapsto A^*$ is a conjugate linear isometry on the Banach space $\mathscr{B}(\mathcal{H})$, and since $A^{**} = A$, it is an involution. Because of its canonical nature, it is often called *the involution* on $\mathscr{B}(\mathcal{H})$. The above considerations are readily extended to maps in $\mathscr{B}(\mathcal{H}, \mathcal{K})$ for two Hilbert spaces. This is left to the reader.

3.2.4 DEFINITION. Let $A \in \mathscr{B}(\mathcal{H})$, $\mathcal{H}$ a Hilbert space. Then the unique bounded linear transformation A^* on $\mathcal{H}$ that that (3.2.4) is valid for all $f, g \in \mathcal{H}$ is called the *adjoint* of A. $A \in \mathscr{B}(\mathcal{H})$ is called *self-adjoint* in case $A = A^*$, that is, in case for all $f, g \in \mathcal{H}$,

$$\langle g, Af \rangle = \langle Ag, f \rangle \ . \tag{3.2.8}$$

As an example, let $\mathcal{K}$ be a closed, proper, non-zero subspace of a Hilbert space $\mathcal{H}$. Let P be the orthogonal projection onto $\mathcal{K}$. Then for all $f, g \in \mathcal{H}$,

$$\langle f, Pg \rangle = \langle P^\perp f + Pf, Pg \rangle = \langle Pf, Pg \rangle = \langle Pf, Pg + P^\perp g \rangle = \langle Pf, g \rangle \ .$$

Thus P is self-adjoint, and for the same reason, so is $P^\perp$.

3.2.5 THEOREM (Hellinger-Toeplitz Theorem)**.** *Let A be a linear transformation from $\mathcal{H}$ to $\mathcal{H}$ defined everywhere on $\mathcal{H}$ and such that for all $f, g \in \mathcal{H}$, (3.2.8) is valid. Then $A \in \mathscr{B}(\mathcal{H})$.*

Proof. By the Closed Graph Theorem, it suffices to show that the graph of A is closed. Suppose that $\{f_n\}$ is a sequence in $\mathcal{H}$ such that $f = \lim_{n\to\infty} f_n$ and $g = \lim_{n\to\infty} Af_n$ both exist. Then for all $h \in \mathcal{H}$,

$$\langle h, g \rangle = \lim_{n\to\infty} \langle h, Af_n \rangle = \lim_{n\to\infty} \langle Ah, f_n \rangle = \langle Ah, f \rangle = \langle h, Af \rangle \ .$$

That is, for all $h \in \mathcal{H}$, $\langle h, g - Af \rangle = 0$, Taking $h = g - Af$, we conclude that $g = Af$, and the graph of A is closed. □

3.2.6 THEOREM (Gram-Schmidt)**.** *Let $\mathcal{H}$ be a Hilbert space, and let $\{f_j\}_{j\in\mathscr{J}}$ be a linearly independent subset of $\mathcal{H}$ where either $\mathscr{J} = \{1, \dots, N\}$ for some $N \in \mathbb{N}$, $N \leq \dim(\mathcal{H})$, or, in case $\mathcal{H}$ is infinite dimensional, $\mathscr{J} = \mathbb{N}$. Then there exists an orthonormal set $\{u_j\}_{j\in\mathscr{J}}$ such that for all $n \in \mathscr{J}$,*

$$\mathrm{span}(\{f_1, \dots, f_n\}) = \mathrm{span}(\{u_1, \dots, u_n\}) \ . \tag{3.2.9}$$

We may further require that $\langle f_n, u_n \rangle > 0$ for all n, and under this condition, $\{u_j\}_{j\in\mathscr{J}}$ is uniquely determined.

Proof. To have $\mathrm{span}(\{f_1\}) = \mathrm{span}(\{u_1\})$ and $\|u_1\| = 1$, we must choose $u_1 = \alpha_1 \|f_1\|^{-1} f_1$ for some $\alpha_1 \in \mathbb{C}$ with $|\alpha| = 1$. For each $n \in \mathscr{J}$, $n > 1$, let $\mathcal{K}_n = \mathrm{span}(\{f_1, \dots, f_n\})$. Since $\dim(\mathcal{K}_n) = n$, $\mathcal{K}_n$ is closed. Let P_n denote the orthogonal projection onto $\mathcal{K}_n$. Since $\{f_1, \dots, f_{n-1}\}$ spans $\mathcal{K}_{n-1}$, and since $\{f_j\}_{j\in\mathscr{J}}$ is linearly independent, $f_n \notin \mathcal{K}_{n-1}$, and hence $P_{n-1}^\perp f_n \neq 0$. Choose $\alpha_n \in \mathbb{C}$ with $|\alpha_n| = 1$, and define

$$u_n = \frac{\alpha_n}{\|P_{n-1}^\perp f_n\|} P_{n-1}^\perp f_n \ . \tag{3.2.10}$$

This formula is valid for all n including $n = 1$ if we define P_0 to be the identity operator.

Note that

$$f_n = P_{n-1}f_n + P_{n-1}^{\perp}f = P_{n-1}f_n + \frac{\|P_{n-1}^{\perp}f_n\|}{\alpha_n}u_n$$

and $P_{n-1}f_n \in \mathcal{K}_{n-1}$. Hence

$$\mathcal{K}_n = \mathrm{span}(\{f_1, \ldots, f_{n-1}, f_n\}) = \mathrm{span}(\mathcal{K}_{n-1} \cup \{f_n\})) = \mathrm{span}(\mathcal{K}_{n-1} \cup \{u_n\})) \ .$$

Therefore, if $\mathcal{K}_{n-1} = \mathrm{span}(\{u_1, \ldots, u_{n-1}\})$, then $\mathcal{K}_n = \mathrm{span}(\{u_1, \ldots, u_n\})$. Since evidently $\mathrm{span}\{f_1\}) = \mathrm{span}\{u_1\})$, induction shows that $\mathcal{K}_n = \mathrm{span}(\{u_1, \ldots, u_n\})$ for all n, and thus (3.2.9) is valid for all n, and for all n,

$$u_n \in \mathcal{K}_{n-1}^{\perp} = (\mathrm{span}(\{u_1, \ldots, u_n\}))^{\perp} \ .$$

That is, for all $j < n$ $\langle u_j, u_k\rangle = 0$ for all $n \in \mathscr{J}$. Since each u_j is a unit vector, $\{u_j\}_{j\in\mathscr{J}}$ is orthonormal.

For the uniqueness, note that u_n must be orthogonal to every vector in $\mathcal{K}_{n-1} = \mathrm{span}(\{u_1, \ldots, u_{n-1}\})$, and must belong to $\mathcal{K}_n = \mathrm{span}(\{f_1, \ldots, f_n\})$, there are $\beta_j \in \mathbb{C}$, $j = 1, \ldots, n$ such that $u_n = \sum_{j=1}^n \beta_j f_j$, and therefore

$$u_n = P_{n-1}^{\perp}u_n = P_{n-1}^{\perp}\left(\sum_{j=1}^n \beta_j f_j\right) = \beta_n P_{n-1}^{\perp}f_n \ .$$

and so u_n must have the form given in (3.2.10), and the only freedom in the choice of $\{u_j\}_{j\in\mathscr{J}}$ is the choice of the multiples α_j, but the further condition $\langle f_n, u_n\rangle > 0$ fixes $\alpha_j = 1$ for all j. □

In any infinite dimensional Hilbert space $\mathcal{H}$, there is an infinite linearly independent sequence $\{f_n\}_{n\in\mathbb{N}}$ in $\mathcal{H}$, and then by Theorem 3.2.6, there is an infinte orthonormal sequence $\{u_n\}_{n\in\mathbb{N}}$ in $\mathcal{H}$. Since for $m \neq n$, $\|u_n - u_m\| = \sqrt{2}$, no subsequence of this sequence is Cauchy, and hence this sequence has no convergent subsequence. This proves:

3.2.7 THEOREM. *In an infinite dimensional Hilbert space $\mathcal{H}$, $B := \{f \ : \ \|f\| \leq 1\}$, the closed unit ball, is not compact.*

3.3 The Hilbert space $L^2(X, \mathcal{M}, \mu)$

3.3.1 $L^2(X, \mathcal{M}, \mu)$ as an inner product space

So far, our only example of an infinite dimensional Hilbert space is ℓ^2. The Lebesgue theory of integration provided a vast new range of examples. If $(X, \mathcal{M}, \mu)$ is a measure space, the vector space of square integrable complex valued function f on X, identified under almost everywhere equivalence, has a natural inner product making it a Hilbert space. This is the content of the Riesz-Fisher Theorem. A particular case is that in which $X = \mathbb{N}$, $\mathcal{M} = 2^{\mathbb{N}}$, and μ is counting measure. In this case, $L^2(X, \mathcal{M}, \mu) = \ell^2$.

Let $(X, \mathcal{M}, \mu)$ be a measure space. As a set, $L^2(X, \mathcal{M}, \mu)$ consists of the equivalence classes, under equivalence almost everywhere with respect to μ, of functions on X that are $\mathcal{M}$-measurable and such that $\int_X |f|^2 \mathrm{d}\mu < \infty$. Clearly, if $z \in \mathbb{C}$ and $f \in L^2(X, \mathcal{M}, \mu)$, $|zf|^2 = |z|^2|f|^2$ is integrable, and if $f, g \in L^2(X, \mathcal{M}, \mu)$,

$$|f + g|^2 \leq (|f| + |g|)^2 \leq 2(|f|^2 + |g|^2) \ . \tag{3.3.1}$$

is integrable. Thus, $L^2(X, \mathcal{M}, \mu)$ is a vector space under the usual rules of addition and scalar multiplication for functions. Also, for $\alpha, \beta \in \mathbb{C}$, $2|\overline{\alpha}\beta| \leq |\alpha|^2 + |\beta|^2$ so that for $L^2(X, \mathcal{M}, \mu)$, $\overline{f}g \in L^1(X, \mathcal{M}, \mu)$.

3.3.1 DEFINITION (The L^2 inner product)**.** For $f, g \in L^2(X, \mathcal{M}, \mu)$, we define

$$\langle f, g \rangle = \int_X \overline{f} g \mathrm{d}\mu \ . \tag{3.3.2}$$

Note that $\langle \cdot, \cdot \rangle$ is a positive sesquilinear form on $L^2(X, \mathcal{M}, \mu)$, and it is non-degenralte since $\langle f, f \rangle = 0$ if and only if $\int_X |f|^2 \mathrm{d}\mu = 0$ if and only if $f = 0$ almost everywhere. Therefore, $L^2(X, \mathcal{M}, \mu)$ equipped with this inner product is an inner product sapce. We write d_2 to denote the metric corresponding to this inner product, and refer to is as the L^2 metric.

3.3.2 The Reisz-Fischer Theorem

3.3.2 THEOREM (Riesz-Fischer Theorem)**.** *$L^2(X, \mathcal{M}, \mu)$ equipped with the L^2 metric is complete, and hence a Hilbert space. Moreover, if $\{f_n\}_{n \in \mathbb{N}}$ is any Cauchy sequence in $L^2(X, \mathcal{M}, \mu)$, then there is a subsequence of $\{f_n\}_{n \in \mathbb{N}}$ that converges almost everywhere with respect to μ.*

Proof. Let $\{f_n\}_{n \in \mathbb{N}}$ be a Cauchy sequence in $L^2(X, \mathcal{M}, \mu)$. Recursively define an increasing sequence of natural numbers $\{n_k\}_{k \in \mathbb{N}}$ such that $\|f_n - f_{n_k}\|_2 \leq 2^{-k}$ for all $n \geq n_k$. Since $\{n_k\}_{k \in \mathbb{N}}$ is increasing, it follows that $\|f_{n_{k+1}} - f_{n_k}\|_2 \leq 2^{-k}$ for all k.

Now define $F_m = |f_{n_1}| + \displaystyle\sum_{k=1}^{m-1} |f_{n_k} - f_{n_{k-1}}|$. By Theorem 3.1.5, applied iteratively,

$$\|F_m\|_2 \leq \|f_{n_1}\|_2 + \sum_{k=1}^{m-1} \|f_{n_k} - f_{n_{k-1}}\|_2 \leq \|f_{n_1}\|_2 + 1 \ .$$

Thus, by the Lebesgue Monotone Convergence Theorem, $F := \lim\limits_{m \to \infty} F_m$ is square-integrable and

$$\int_X F^2 \mathrm{d}\mu \leq \|f_{n_1}\|_2 + 1 \ .$$

It follows that $F < \infty$ a.e. μ, and thus that $\displaystyle\sum_{k=1}^{\infty} (f_{n_k} - f_{n_{k-1}})$ is absolutely convergent a.e. μ. But since absolute convergence implies convergence, $\lim\limits_{m \to \infty} \left[f_{n_1} + \displaystyle\sum_{k=1}^{m-1} (f_{n_k} - f_{n_{k-1}}) \right] = \lim\limits_{m \to \infty} f_{n_m}$ exists almost everywhere. Call this limit f. As a point-wise limit of measurable functions, f is measurable. Also, $f \in L^2(X, \mathcal{M}, \mu)$ by Fatou's Lemma.

Next, $|f_{n_m} - f|^2 \leq 4F^2$, and since $4F^2$ is integrable, the Lebesgue Dominated Convergence Theorem implies that

$$\lim_{m \to \infty} \|f_{n_m} - f\|_2 = 0 \ .$$

Thus, a subsequence of the Cauchy sequence $\{f_n\}_{n \in \mathbb{N}}$ converges of f in the L^2 metric. But then the whole sequence converges to f. The subsequence $\{f_{n_m}\}_{m \in \mathbb{N}}$ converges to f a.e. μ. □

3.3.3 Landau's Theorem

The next result illustrates the way in which the various abstract results we have proved so far may be combined to prove a theorem that refers only to the Lebesgue theory of integration, but is not easy to prove using only the tools of that theory.

3.3.3 THEOREM (Landau's Theorem)**.** *Let $(X, \mathcal{M}, \mu)$ be a measure space such that for every set $B \in \mathcal{M}$ with $\mu(B) = \infty$, there is a set $A \in \mathcal{M}$, $A \subset B$, with $0 < \mu(A) < \infty$. Suppose f is a measurable function on X such that whenever g is a measurable function on M with $\int_X |g|^2 \mathrm{d}\mu < \infty$, $|fg|$ is integrable. Then $\int_X |f|^2 \mathrm{d}\mu < \infty$.*

Proof. The Riesz-Fischer Theorem identifies the set of measurable g such that $\int_X |g|^2 \mathrm{d}\mu < \infty$ with the elements of the Hilbert space $\mathcal{H} := L^2((X, \mathcal{M}, \mu)$, and then the hypotheses allow us to define a linear functional L on $\mathcal{H}$ the by $L(g) = \int_X \overline{f} g \mathrm{d}\mu$ for all $g \in \mathcal{H}$, For $n \in \mathbb{N}$, define $E_n \subset \mathcal{H}$ by

$$E_n := \left\{ g \ : \ \int_X \| fg|\mathrm{d}\mu \leq n \right\} .$$

Then E_n is closed. To see this, let $\{g_m\}_{m\in\mathbb{N}}$ be a sequence in E_n that converges to $g \in \mathcal{H}$. By the final part of Theorem 3.3.2, there is a subsequence $\{g_{m_k}\}_{k\in\mathbb{N}}$ that converges almost everywhere to g. By Fatou's Lemma,

$$\int |fg|\mathrm{d}\mu \leq \liminf_{k\to\infty} \int_X |fg_{m_k}|\mathrm{d}\mu \leq n .$$

Therefore, $g \in E_n$, and hence E_n is closed.

By hypothesis,$\mathcal{H} = \cup_{n=1}^{\infty} E_n$, and then by Baire's Theorem, there exists some n such that E_n has a non-empty interior. Hence for some $g_0 \in \mathcal{H}$ and some $r > 0$, $B(r, g_0) \subset E_n$; i.e., for all $g \in B(1, 0)$, and all $0 < s < r$, $g_0 + sg \in E_n$. Therefore

$$|L(g_0) + sL(g)| = \left| \int_X \overline{f}(g_0 + sg)\mathrm{d}\mu \right| \leq \int_X |f(g_0 + sg)|\mathrm{d}\mu \leq n .$$

It follows that $|L(g)| \leq (n + |L(g_0|)/r$ for all $g \in B(1, 0)$, and hence L is bounded. By the Reisz Representation Theorem, there exists a unique $f_0 \in \mathcal{H}$ such that $L(g) = \int_X \overline{f_0} g \mathrm{d}\mu$ for all $g \in \mathcal{H}$, and thus,

$$\int_X \overline{(f_0 - f)} g \mathrm{d}\mu = 0$$

for all $g \in \mathcal{H}$. This implies that $f = f_0$ almost everywhere. To see this, for each $n \in \mathcal{H}$ define the set $B_n = \{x \ : \ |f_0(x) - f(x)| \geq 1/n\}$. By hypothesis, even if $\mu(B_n) = \infty$, there exists a measurable set $A_n \subset B_n$ with $0 < \mu(A_n) < \infty$. Define g_n by

$$g_n(x) = \begin{cases} f_0(x) - f(x)/|f_0(x) - f(x)| & x \in A_n \\ 0 & x \notin A_n . \end{cases}$$

then $g_n \in \mathcal{H}$ and $\int_X \overline{f_0 - f} g_n \mathrm{d}\mu \geq \mu(A_n)/n$, which is a contradiction. Hence $\mu(B_n) = 0$ for all n. □

The condition on the measure space is much weaker than countable additivity, but the condition is necessary: If there exists a set $B \in \mathcal{M}$ with $\mu(B) = \infty$, and such that for all measurable $A \subset B$, either $\mu(A) = 0$ or $\mu(A) = \infty$, let f be the function 1_B, the indicator function of B. Since every $g \in L^2(X, \mathcal{M}, \mu)$ must equal zero almost everywhere on B,it follows that fg is zero almost everywhere for all $g \in L^2(X, \mathcal{M}, \mu)$, and therefore certainly fg is integrable. However, f is not square integrable.

3.4 Bessel's inequality and complete orthonormal sets

3.4.1 Best approximation in a Hilbert space and Bessel's inequality

Let $\mathcal{H}$ be a Hilbert space space, and let $\{u_1, \dots, u_n\}$ be an orthonormal set in $\mathcal{H}$. Let $\mathcal{K} = \text{span}(\{u_1, \dots, u_n\})$ which is finite dimensional and therefore closed. Let P denote the orthogonal projection onto $\mathcal{K}$. For any $f \in \mathcal{H}$, by Theorem 3.1.9, the best approximation to f by elements of $\mathcal{K}$ is given by Pf in the sense that $\|f - Pf\| < \|F - g\|$ for any $g \neq Pf$ in $\mathcal{K}$. This result will be more useful once we have a formula for Pf in terms of $\{u_1, \dots, u_n\}$. We now derive such a formula.

The general element of $\mathcal{K}$ has the form $\displaystyle\sum_{j=1}^n \alpha_j u_j$ for some complex numbers $\alpha_1, \dots, \alpha_n$. To determine the choice of these coefficients that gives the best approximation to f, we compute

$$\begin{aligned}
\left\| f - \sum_{j=1}^n \alpha_j u_j \right\|^2 &= \left\langle f - \sum_{j=1}^n \alpha_j u_j \ , \ f - \sum_{j=1}^n \alpha_j u_j \right\rangle \\
&= \|f\|^2 - \sum_{j=1}^n 2\Re(\overline{\alpha_j}\langle u_j, f\rangle) + \sum_{j=1} |\alpha_j|^2 \\
&= \|f\|^2 - \sum_{j=1}^n |\langle u_j, f\rangle|^2 + \sum_{j=1}^n |\alpha_j - \langle u_j, f\rangle|^2 \ . \qquad (3.4.1)
\end{aligned}$$

Evidently, the best choice is given by $\alpha_j = \langle u_j, f\rangle$ for each $= 1, \dots, n$, and therefore,

$$Pf = \sum_{j=1}^n \langle u_j, f\rangle u_j \ . \tag{3.4.2}$$

We summarize so far:

3.4.1 THEOREM. *Let $\{u_j\}_{j\in\mathbb{N}}$ be an onrthonormal sequence in an Hilbert sapce $\mathcal{H}$. Thren for all $n \in \mathbb{N}$,*

$$\left\| f - \sum_{j=1}^n \alpha_j u_j \right\| \leq \left\| f - \sum_{j=1}^n \langle u_j, f\rangle u_j \right\| \tag{3.4.3}$$

and there is equality if and only if $\alpha_j = \langle u_j, f\rangle$ for all j.

Making the choice $\alpha_j = \langle u_j, f\rangle$ for each $= 1, \dots, n$, we have

$$\left\| f - \sum_{j=1}^n \langle u_j, f\rangle u_j \right\|^2 = \|f\|^2 - \sum_{j=1}^n |\langle u_j, f\rangle|^2 \ . \tag{3.4.4}$$

Since the left hand side is non-negative, we have that for any finite orthonormal set $\{u_1, \dots, u_n\}$,

$$\sum_{j=1}^n |\langle u_j, f\rangle|^2 \leq \|f\|^2 \ . \tag{3.4.5}$$

Now suppose that $\mathcal{H}$ contains an uncountable orthonormal set $\{u_j\}_{j\in\mathscr{J}}$. (Hence $\mathscr{J}$ is some uncountable set.) For any $f \in \mathcal{H}$, if $|\langle f, u_j\rangle| > 0$ for uncountably many $j \in \mathscr{J}$, there is some $n \in \mathbb{N}$ such that $|\langle f, u_j\rangle| > 1/n$ for uncountably many $j \in \mathscr{J}$ since

$$\{j \ : \ |\langle f, u_j\rangle| > 0\} = \bigcup_{n=1}^\infty \{j \ : \ |\langle f, u_j\rangle| > 1/n\}$$

and a countable union of countable sets is countable. But then there is a sequence $\{u_{j.k}\}_{k\in\mathbb{N}}$ such that $\sum_{k=1}^{\infty}|\langle f, u_{j,k}\rangle|^2 = \infty$, and this contradicts (3.4.5). Therefore, even when $\mathcal{H}$ contains an uncountable orthonormal set $\{u_j\}_{j\in\mathscr{J}}$. $|\langle f, u_j\rangle| > 0$ only for countably many $j \in \mathscr{J}$, and then we have from (3.4.4) that

$$\sum_{j\in\mathscr{J}} |\langle u_j, f\rangle|^2 \le \|f\|^2 \ . \tag{3.4.6}$$

Summarizing, we have proved:

3.4.2 THEOREM (Bessel's Inequality)**.** *Let $\mathcal{H}$ be a Hilbert space, and let $\{u_j\}_{j\in\mathscr{J}}$ be an orthonormal set in $\mathcal{H}$. Then for all $f \in \mathcal{H}$, $|\langle f, u_j\rangle| > 0$ only for countably many $j \in \mathscr{J}$, and (3.4.6) is valid.*

3.4.2 Complete orthonormal sets

3.4.3 LEMMA. *Let $\{\alpha_j\}_{j\in\mathbb{N}}$ be any square-summable sequence of complex numbers; i.e, $\sum_{j=1}^{\infty}|\alpha_j|^2 < \infty$. Let $\{u_j\}_{j\in\mathbb{N}}$ be any orthonormal sequence in a Hilbert space $\mathcal{H}$. Then the sequence $\left\{\sum_{j=1}^{n}\alpha_j u_j\right\}_{n\in\mathbb{N}}$ is Cauchy and therefore this sequence has a unique limit $g \in \mathcal{H}$ so that*

$$\sum_{j=1}^{\infty}\alpha_j u_j = g \ . \tag{3.4.7}$$

Moreover, this sum converges to the same vector no matter how the terms are ordered.

Proof. For each $n \in \mathbb{N}$ define $f_n = \sum_{j=1}^{n}\alpha_j u_j$. Then for $n > m$,

$$\|f_n - f_m\|^2 = \left\|\sum_{j=m+1}^{n}\alpha_j u_j\right\|^2 = \sum_{j=m+1}^{n}|\alpha_j|^2 \le \sum_{j=m+1}^{\infty}|\alpha_j|^2 \ . \tag{3.4.8}$$

Since $\lim_{m\to\infty}\sum_{j=m+1}^{\infty}|\alpha_j|^2 = 0$, $\{f_n\}_{n\in\mathbb{N}}$ is a Cauchy sequence, and then since $\mathcal{H}$ is complete, there exists a unique $g \in \mathcal{H}$ such that $\lim_{n\to\infty}\|f_n - g\| = 0$. This proves (3.4.7). Since the condition $\sum_{j=1}^{\infty}|\alpha_j|^2 < \infty$ holds independent of how the terms in the sum are ordered, the convergence of the infinite sum in (3.4.7) holds independent of how the terms in the sum are ordered. For all $\epsilon > 0$, let J and K be finite subsets of $\mathbb{N}$ such that

$$\sum_{j\notin J}|\alpha_j|^2 < \epsilon^2/9 \quad \text{and} \quad \sum_{j\notin K}|\alpha_j|^2 < \epsilon^2/9 \ . \tag{3.4.9}$$

Then $\left\|\sum_{j\in J}\alpha_j u_j - \sum_{j\in J\cup K}\alpha_j u_j\right\| \le \frac{1}{3}\epsilon$, and $\left\|\sum_{j\in K}\alpha_j u_j - \sum_{j\in J\cup K}\alpha_j u_j\right\| \le \frac{1}{3}\epsilon$. Then by Minkowski's inequality, $\left\|\sum_{j\in J}\alpha_j u_j - \sum_{j\in K}\alpha_j u_j\right\| < \frac{2}{3}\epsilon$. Let $J = \{1,\dots,n\}$ for n sufficently large that (3.4.9) is valid. By (3.4.8), $\left\|g - \sum_{j\in J}\alpha_j u_j\right\| < \frac{1}{3}\epsilon$, and then $\left\|g - \sum_{j\in K}\alpha_j u_j\right\| < \epsilon$. This shows that the sequence of partial sums tends to g no matter how the terms are ordered. $\square$

Let $\{u_j\}_{j\in\mathscr{J}}$ be an orthonormal set in $\mathcal{H}$ and let $f \in \mathcal{H}$. By Theorem 3.4.2, $\displaystyle\sum_{j\in\mathscr{J}} |\langle u_j, f\rangle|^2 < \infty$, and then by Lemma 3.4.3,

$$g := \sum_{j\in\mathscr{J}} \langle u_j, f\rangle u_j \tag{3.4.10}$$

is a well-defined element of $\mathcal{H}$. Notice that for each $j \in \mathscr{J}$, $\langle u_j, g\rangle = \langle u_j, f\rangle$ and hence $\langle f - g, u_j\rangle = 0$ for all $j \in \mathscr{J}$. This brings us to the following definition:

3.4.4 DEFINITION (Complete orthonormal set)**.** Let $\mathcal{H}$ be a Hilbert space and let $\{u_j\}_{j\in\mathscr{J}}$ be an orthonormal set in $\mathcal{H}$. Then $\{u_j\}_{j\in\mathscr{J}}$ is a *complete orthonormal set in* $\mathcal{H}$ in case the only vector $f \in \mathcal{H}$ such that $\langle u_j, f\rangle = 0$ for all j is $f = 0$. A complete orthonormal set in $\mathcal{H}$ is also called an orthonormal basis for $\mathcal{H}$.

We have shown just above that if $\{u_j\}_{j\in\mathscr{J}}$ is any orthonormal set in $\mathcal{H}$ and f is any vector in $\mathcal{H}$, $g := \sum_{j\in\mathscr{J}} \langle u_j, f\rangle u_j$ is a well-defined vector in $\mathcal{H}$ such that $\langle f - g, u_j\rangle = 0$ for all $j \in \mathscr{J}$. If $\{u_j\}_{j\in\mathscr{J}}$ is complete, this means that $f - g = 0$, and hence for all $f \in \mathcal{H}$.

$$f = \sum_{j\in\mathscr{J}} \langle u_j, f\rangle u_j\ . \tag{3.4.11}$$

3.4.5 THEOREM (Parseval's Theorem)**.** *Let $\{u_j\}_{j\in\mathscr{J}}$ be a complete orthonormal set in a Hilbert space $\mathcal{H}$. Then for all $f \in H$, (3.4.11) is valid, and*

$$\|f\|^2 = \sum_{j\in\mathscr{J}} |\langle u_j, f\rangle|^2\ . \tag{3.4.12}$$

Proof. It remains only to prove (3.4.12). Note that

$$\left\|\sum_{j\in\mathscr{J}} \langle u_j, f\rangle u_j\right\|^2 = \sum_{j\in\mathscr{J}} |\langle u_j, f\rangle|^2\ ,$$

so if $\displaystyle\sum_{j\in\mathscr{J}} |\langle u_j, f\rangle|^2 \neq \|f\|^2$, then (3.4.11) cannot be valid, and this is a contradiction. □

3.4.3 Separability

A metric space is *separable* in case it contains a countable dense set. Hilbert spaces are, in particular, metric spaces and hence a separable Hilbert space $\mathcal{H}$ is one that contains a dense sequence $\{f_n\}_{n\in\mathbb{N}}$ of vectors. To avoid trivialities, suppose that $\mathcal{H}$ is infinite dimensional as well as separable. Let $\{f_n\}_{n\in\mathbb{N}}$ be any dense sequence in $\mathcal{H}$. Discard the vector f_n in case $f_n \in \mathrm{span}(\{f_1, \dots, f_{n-1}\})$. Let g_k denote the kth retained vector. Then $\{g_k\}_{k\in\mathbb{N}}$ is linearly independent and $\mathrm{span}\,(\{g_k\}_{k\in\mathbb{N}})$ is dense in $\mathcal{H}$.

Applying Theorem 3.2.6 to $\{g_k\}_{k\in\mathbb{N}}$ we obtain an orthonormal sequence $\{u_k\}_{k\in\mathbb{N}}$ such that $\mathrm{span}\,(\{u_k\}_{k\in\mathbb{N}}) = \mathrm{span}\,(\{g_k\}_{k\in\mathbb{N}})$, and hence such that $\mathrm{span}\,(\{u_k\}_{k\in\mathbb{N}})$ is dense. This shows that every separable Hilbert space $\mathcal{H}$ contains an orthonormal sequence $\{u_k\}_{k\in\mathbb{N}}$ whose span is dense in $\mathcal{H}$.

Such an orthonormal sequence is necessarily complete, as we show next:

3.4.6 LEMMA. *Let $\mathcal{H}$ be a Silbert sapce and let $\{u_n\}_{n\in\mathbb{N}}$ be an orthonormal sequence in $\mathcal{H}$ such that* $\mathrm{span}(\{u_n\}_{n\in\mathbb{N}})$ *is dense in $\mathcal{H}$. Then $\{u_n\}_{n\in\mathbb{N}}$ is complete.*

Proof. Suppose $f \in \mathcal{H}$ and $\langle u_k, f\rangle = 0$ for all $k \in \mathbb{N}$. Since $\mathrm{span}\,(\{u_k\}_{k\in\mathbb{N}})$ is dense in $\mathcal{H}$, for all $\epsilon > 0$, there exists for $n \in \mathbb{N}$ and coefficients $\alpha_1, \ldots, \alpha_n$ such that $\epsilon \geq \left\| f - \sum_{j=1}^{n} \alpha_j u_j \right\|$. Then using (**??**) and the orthogonality hypothesis,

$$\epsilon \geq \left\| f - \sum_{j=1}^{n} \alpha_j u_j \right\| \geq \left\| f - \sum_{j=1}^{n} \langle f, u_j\rangle u_j \right\| = \|f\| \ .$$

Since $\epsilon > 0$ is arbitrary, $\|f\| = 0$. □

3.4.7 THEOREM. *A Hilbert space $\mathcal{H}$ is separable if and only if exists there exists a sequence $\{u_k\}_{k\in\mathbb{N}}$ that is orthonormal and complete.*

Proof. We have already proved that if $\mathcal{H}$ is separable, then $\mathcal{H}$ contains a sequence $\{u_k\}_{k\in\mathbb{N}}$ that is orthonormal and complete. Therefore, suppose that $\mathcal{H}$ contains a sequence $\{u_k\}_{k\in\mathbb{N}}$ that is orthonormal and complete. Then for all $f \in \mathcal{H}$, $f = \sum_{k=1}^{\infty} \langle u_k, f\rangle u_k$ and

$$\left\| f - \sum_{k=1}^{\infty} \alpha_k u_k \right\|^2 = \sum_{k=1}^{\infty} |\langle u_k, f\rangle - \alpha_k|^2 \ .$$

For any $\epsilon > 0$, we may choose an $n \in \mathbb{N}$ and $\alpha_1, \ldots \alpha_n$, each of whose real and imaginary parts are rational, such that

$$\sum_{k=1}^{n} |\langle u_k, f\rangle - \alpha_k|^2 < \frac{1}{2}\epsilon \qquad \text{and} \qquad \sum_{k=n+1}^{\infty} |\langle u_k, f\rangle|^2 < \frac{1}{2}\epsilon \ .$$

It follows that with $g = \sum_{k=1}^{n} \alpha_k u_k$, $\|f - g\| < \epsilon$. Thus, the set of finite linear combinations of the vectors in $\{u_k\}_{k\in\mathbb{N}}$ with coefficients whose real and imaginary parts are rational is dense in $\mathcal{H}$. Evidently this set is countable, and hence $\mathcal{H}$ is separable. □

Notice that if $\mathcal{H}$ is any separable Hilbert space, and $\{u_j\}_{j\in\mathbb{N}}$ is any orthonormal basis in $\mathcal{H}$, then the transformation that sends $f \in \mathcal{H}$ to the sequence whose jth term is $\langle u_j, f\rangle$ is a linear isometry of $\mathcal{H}$ onto ℓ^2. That is, every separable Hilbert space can be mapped onto ℓ^2 by a linear isometry.

3.4.8 EXAMPLE (Fourier series)**.** *Let $\mathcal{H} = L^2(S^1, \mathcal{B}_{S^1}, m)$ be the Hilbert space of Borel functions $f(\theta)$ on the unit circle that are square integrable with respect to Lebesgue measure m on S^1 normalized so that $\mu(S^1) = 1$.*

It is then readily checked that with u_j defined by $u_n(\theta) = e^{in\theta}$, $\{u_n\}_{n\in\mathbb{Z}}$ is orthonormal. By the Stone-Wierstrass Theorem, every continuous function on S^1 can be approximated arbitrarily well in the uniform metric by a finite linear combination of the vectors in our orthonormal set. Since for continuous functions f, g on S^1,

$$\left(\int_{S^1} |f - g|^2 \mathrm{d}\mu \right)^{1/2} \leq \max_{\theta}\{|f(\theta) - g(\theta)|\} \ ,$$

the span of $\{u_n\}_{n\in\mathbb{Z}}$ is dense in the set continous function on S^1 equipped with the norm metric inherited from $\mathcal{H}$.

Since the continuous functions are dense in $\mathcal{H}$, and for any $f \in \mathcal{H}$ and any $\epsilon > 0$, we can find a continuous function g such that $\|f - g\| < \epsilon$, and then a function p in the span of $\{u_n\}_{n\in\mathbb{Z}}$ such that $\|g - p\| < \epsilon/2$. Then $\|f - p\| < \epsilon$, and hence the span of $\{u_n\}_{n\in\mathbb{Z}}$ is dense in $\mathcal{H}$. Then by Lemma 3.4.6, $\{u_n\}_{n\in\mathbb{Z}}$ is complete: Thus, $\{u_n\}_{n\in\mathbb{Z}}$ is orthonormal basis for $\mathcal{H}$, called the Fourier basis. *It follows that for each $f \in \mathcal{H}$,*

$$f = \lim_{n\to\infty} \sum_{j=-n}^{n} \langle u_j, f\rangle u_j \ .$$

The sequence $\{\langle u_j, f\rangle\}_{j\in\mathbb{Z}}$ is called the sequence of Fourier coefficients *of f. By Parseval's identiy,*

$$\|f\|^2 = \sum_{n\in\mathbb{Z}} |\langle u_n, f\rangle|^2 \ .$$

The map sending f into the (doubly) infnite sequence $\{\langle u_n, f\rangle\}_{n\in\mathbb{Z}}$ is then a a linear isometry from $\mathcal{H}$ into the Hilbert space of square summable sequences indexed by $\mathbb{Z}$, which we can identify with ℓ^2 using any bijection between $\mathbb{N}$ and $\mathbb{Z}$. In fact this isometry is a bijection: As we have seen, if $\{\alpha_n\}_{n\in\mathbb{Z}}$ is any sqaure summable sequence, then $g = \sum_{n\in\mathbb{Z}} \alpha_n u_u$ is a well defined element of $\mathcal{H}$, and for each $n \in \mathbb{Z}$, $\langle u_n, g\rangle = \alpha_n$. This isometric linear bijection between $\mathcal{H}$ onto ℓ^2 is the Fourier transform.

3.4.9 EXAMPLE (Orthogonal polynomials)**.** *For $a, b \in \mathbb{R}$, $a < b$, let $\mathcal{B}$ the Borel σ-algebra on $[a, b]$, and let μ be any finite Borel measure on $[a, b]$. Let $\mathcal{H} = L^1([a, b], \mathcal{B}, \mu)$. Continuous functions are dense in $\mathcal{H}$, and by the Stone-Wierstrass Theorem, for every continuous function f on $[a, b]$ and every $\epsilon > 0$, there is a polynomial $p(x)$ such that $\max\{|f(x) = p(x)| \ : \ x \in [a, b]\} < \epsilon$, and then by the Cauchy-Schwarz inequality, $\|f - p\| \leq \sqrt{\epsilon\mu([a, b])}$. Without loss of generality, we may take the coefficients of p to have rational real and imaginary parts and the set of such polynomials is countable. Therefore $\mathcal{H}$ is separable.*

Moreover, this proves that the sequence of monomials $\{x^{n-1}\}_{n\in\mathbb{N}}$ has a dense span in $\mathcal{H}$, and therefore, the orthonormal sequence $\{u_n\}_{n\in\mathbb{N}}$ that is obtained from $\{x^{n-1}\}_{n\in\mathbb{N}}$ via the Gram-Schmidt Theorem is an orthonormal basis for $\mathcal{H}$ such that each u_n is a polynomial of degree $n - 1$. The elements of this basis are uniquely determined up to a multiple by a unit complex number, and conventionally one chooses the multiple so that the leading coefficient; i.e., the multiple of x^{n-1}, is positive. This is the canonical orthonormal polynomial basis *for $\mathcal{H}$.*

3.4.4 Partial isometries and unitary operators

3.4.10 DEFINITION. Let $\mathcal{H}$ and $\mathcal{K}$ be s Hilbert spaces. An *isometry* from $\mathcal{H}$ into $\mathcal{K}$ is an operator $U \in \mathscr{B}(\mathcal{H}.\mathcal{K})$ such that $\|Uf\|_{\mathcal{K}} = \|f\|_{\mathcal{H}}$ for all $f \in \mathcal{H}$. An isometry U from $\mathcal{H}$ to $\mathcal{K}$ is *unitary* in case U is surjective. Two Hilbert spaces $\mathcal{H}$ and $\mathcal{K}$ are *unitarily equivalent* in case there is a unitary transformation from $\mathcal{H}$ onto $\mathcal{K}$. A partial isometry in $\mathscr{B}(\mathcal{H}, \mathcal{K})$ is an operator U such that the restriction of U to $(\ker(U))^\perp$ is an isometry from $(\ker(U))^\perp$ into $\mathcal{K}$.

Note that the composition of unitary operators is unitary since the composition of invertible transformations is invertible, and the composition of isometries is an isometry. It follows easily that the inverse of a unitary transformation is also unitary, so that unitary equivalence is, indeed, an equivalence relation on the set of Hilbert spaces. As the next example shows, all separable Hilbert spaces belong to the same equivalence class, which is represented by ℓ^2.

3.4.11 EXAMPLE. *Let $\mathcal{H}$ be a separable Hilbert space, and let $\{u_n\}_{n\in\mathbb{N}}$ be an orthonormal sequence in $\mathcal{H}$. Define a map U from $\mathcal{H}$ to ℓ^2 by $U(\{\alpha_n\}_{n\in\mathbb{N}}) = \sum_{n=1}^{\infty} \alpha_n u_n$, which is well-defined by Lemma 3.4.3. Since $\|U(\{\alpha_n\}_{n\in\mathbb{N}})\|^2 = \sum_{n=1}^{\infty} |\alpha_n|^2$, U is an isometry.*

Suppose moreover that $\{u_n\}_{n\in\mathbb{N}}$ is not only an orthonormal sequence, but that it is an orthonormal basis for $\mathcal{H}$. Then by Theorem 3.4.5, $f = \sum_{n=1}^{\infty} \langle u_n, f\rangle u_n = U(\{\langle u_n, f\rangle\}_{n\in\mathbb{N}})$ so that U is surjective, and hence unitary. Hence $\mathcal{H}$ is unitarily equivalent to ℓ^2. Since $\mathcal{H}$ is an arbitrary separable Hilbert space, by the remarks preceding the example, all separable Hilbert spaces are unitarily equivalent.

3.4.12 LEMMA. *Let $\mathcal{H}$ and $\mathcal{K}$ be Hilbert spaces, and let $U \in \mathscr{B}(\mathcal{H}, \mathcal{K})$. Then U is an isometry if and only if for all $f, g \in \mathcal{H}$, $\langle Uf, Ug\rangle_{\mathcal{K}} = \langle f, g\rangle_{\mathcal{H}}$.*

Proof. Suppose that for all $f, g \in \mathcal{H}$, $\langle Uf, Ug\rangle_{\mathcal{K}} = \langle f, g\rangle_{\mathcal{H}}$. Then taking $g = f$, $\|Uf\|_{\mathcal{K}} = \|f\|_{\mathcal{H}}$ so that U is an isometry. Conversely, if $\|Uh\|_{\mathcal{K}} = \|h\|_{\mathcal{H}}$ for all $h \in \mathcal{H}$,

$$\Re\left(\langle f, g\rangle_{\mathcal{H}}\right) = \frac{1}{4}(\|f+g\|_{\mathcal{H}} - \|f-g\|_{\mathcal{H}}) = \frac{1}{4}(\|Uf+Ug\|_{\mathcal{K}} - \|Uf-Ug\|_{\mathcal{K}}) = \Re\left(\langle Uf, Ug\rangle_{\mathcal{K}}\right) \ .$$

Replacing g by $e^{i\theta}g$, we conclude that for all $f, g \in \mathcal{H}$, $\langle Uf, Ug\rangle_{\mathcal{K}} = \langle f, g\rangle_{\mathcal{H}}$. $\square$

3.4.13 THEOREM. *Let $\mathcal{H}$ and $\mathcal{K}$ be Hilbert spaces, and let $U \in \mathscr{B}(\mathcal{H}, \mathcal{K})$. Then U is unitary if and only if $U^*U = I_{\mathcal{H}}$, and $UU^* = I_{\mathcal{K}}$.*

Proof. If U is unitary, by Lemma 3.4.15, for all $f, g \in \mathcal{H}$,

$$\langle f, g\rangle_{\mathcal{H}} = \langle Uf, Ug\rangle_{\mathcal{K}} = \langle f, U^*Ug\rangle_{\mathcal{H}} \ .$$

This means that $g - U^*Ug$ is orthogonal to all $f \in \mathcal{H}$, and so $g - U^*Ug = 0$. Since $g \in \mathcal{H}$ is arbitrary, this means that $U^*U = I_{\mathcal{H}}$. Hence U^* is the inverse of U, and so $UU^* = I_{\mathcal{K}}$.

Conversely, suppose that $U^*U = I_{\mathcal{H}}$ and $UU^* = I_{\mathcal{K}}$. Then for all $f, g \in \mathcal{H}$,

$$\langle f, g\rangle_{\mathcal{H}} = \langle f, U^*Ug\rangle_{\mathcal{H}} = \langle Uf, Ug\rangle_{\mathcal{K}} \ ,$$

so that U is an isometry into $\mathcal{K}$, Then since for any $f \in \mathcal{K}$, $f = UU^*f = U(U^*f)$, $\mathrm{ran}(U) = \mathcal{K}$, so that U is surjective and hence unitary. $\square$

3.4.14 DEFINITION. Let $\mathcal{H}$ and $\mathcal{K}$ be two Hilbert spaces, and let $U \in \mathscr{B}(\mathcal{H}, \mathcal{K})$ be a partial isometry. Let $\mathcal{H}_1 := (\ker(U))^\perp$ and let $\mathcal{K}_1 := \mathrm{ran}(U)$, which is closed. Then $\mathcal{K}_1$ is called the *initial space of* U, and $\mathcal{H}_1$ is called the *final space of* U.

Let $U \in \mathscr{B}(\mathcal{H}, \mathcal{K})$ be a partial isometry with initial space $\mathcal{H}_1$ and final space $\mathcal{K}_1$. Then the restriction of U to $\mathcal{H}_1$ is a unitary from $\mathcal{H}_1$ onto $\mathcal{K}_1$. Let P be the orthogonal projection in $\mathcal{H}$ onto $\mathcal{H}_1$, and let Q be the orthogonal projection in $\mathcal{K}$ onto $\mathcal{K}_1$. By Theorem 3.4.13, the restriction of U^*U is the identity on $\mathcal{H}_1$, and hence $U^*UP = P$. Since the range of $P^\perp$ is $\mathcal{H}_1^\perp = \ker(U)$, $U^*UP^\perp = 0$. Hence $U^*U = (U^*U(P + P^\perp) = P$. Likewise, $UU^* = Q$.

Now consider any $U \in \mathscr{B}(\mathcal{H}, \mathcal{K})$ such that $U^*U = P$, where P is the orthogonal projection onto some closed subspace $\mathcal{H}_1$ of $\mathcal{H}$. Then for all $f \in \mathcal{H}_1$, $\|Uf\|^2_{\mathcal{K}} = \langle f, U^*Uf\rangle_{\mathcal{H}} = \langle f, Pf\rangle_{\mathcal{H}} = \|f\|^2_{\mathcal{H}}$. Hence U is a partial isometry. This proves:

3.4.15 THEOREM. *Let $\mathcal{H}$ and $\mathcal{K}$ be Hilbert spaces. An operator $U \in \mathscr{B}(\mathcal{H}, \mathcal{K})$ is a partial isometry if and only if U^*U is an orthogonal projection in $\mathcal{H}$, in which case UU^* is an orthogonal projection in $\mathcal{K}$.*

Let $\mathcal{H}$ and $\mathcal{K}$ be two Hilbert spaces. If $\mathcal{H}$ and $\mathcal{K}$ are not unitarily equivalent, then there are no unitary operators in $\mathscr{B}(\mathcal{H}, \mathcal{K})$. If $\mathcal{H}$ and $\mathcal{K}$ are unitarily equivalent, let V be any particular unitary operator in $\mathscr{B}(\mathcal{H}, \mathcal{K})$. Then for all $T \in \mathscr{B}(\mathcal{H}, \mathcal{K})$, $V^*T \in \mathscr{B}(\mathcal{H})$, and V^*T is unitary if and only if T is unitary. Thus, the study of unitary operators on $\mathscr{B}(\mathcal{H}, \mathcal{K})$ readily reduces to the study of unitary operators on $\mathscr{B}(\mathcal{H})$. We henceforth focus on this case.

Suppose that $U \in \mathscr{B}(\mathcal{H})$ is an isometry, but that U is not unitary, so that $UU^* \neq I$. It turns out that UU^* is the orthogonal projection onto the range of U, as we now exlain: Let $V := \mathrm{ran}(U)$, which is closed since U is an isometry, and $V^\perp = \ker(U^*)$. Let P be the orthogonal projection onto the range of U, and $f, g \in \mathcal{H}$. By definition, there exists $h \in \mathcal{H}$ such that $Pg = Uh$. Then

$$\langle f, Pg \rangle = \langle f, Uh \rangle = \langle U^* f, h \rangle = \langle UU^* f, Uh \rangle = \langle UU^* f, g \rangle \ .$$

since $f, g \in \mathcal{H}$ are arbitrary, it follows that $UU^* = P$. We have proved:

3.4.16 THEOREM. *Let $\mathcal{H}$ be a Hilbert space and let $U \in \mathscr{B}(\mathcal{H})$ be an isometry, Then UU^* is the orthogonal prjection on the range of U.*

3.4.17 DEFINITION (Partial isometry)**.** Let $\mathcal{H}$ be a Hilbert space. An operator $U \in \mathscr{B}(\mathcal{H})$ is a *partial isometry* in case the restriction of U to $(\ker(U))^\perp$ is an isometry into $\mathcal{H}$.

3.4.18 DEFINITION (Ortogonal projection)**.** Let $\mathcal{H}$ be a Hilbert space. An operator $P \in \mathscr{B}(\mathcal{H})$ is an *orthogonal proejction* in case $P = P^*$ and $P^2 = P$.

Let P be an orthogonal projection in $\mathscr{B}(\mathcal{H})$, and let $V = \mathrm{ran}(P)$. Since $(\ker(P))^\perp = \mathrm{ran}(P^*) = \mathrm{ran}(P)$, the restriction of P to $(\ker(P))^\perp$ is an isometry, and hence P is a partial isometry.

3.4.19 THEOREM. *Let $\mathcal{H}$ be a Hilbert space. $U \in \mathscr{B}(\mathcal{H})$ is a partial isometry if and only if UU^* is an orthogonal projection.*

Proof. Suppose that U is a partial isometry. Then UU^* is self adjoint, and

$$(UU^*)^2 = U(U^*U)U^* = UU^* \ ,$$

so that UU^* is an orthogonal projection. Conversely, suppose that UU^* is an orthogonal projection.

□

3.5 The weak topology on a Hilbert space

3.5.1 Weak topology and weak convergence

As a Banach space, a Hilbert space $\mathcal{H}$ may be equipped with its weak topology, and since every Hilbert space $\mathcal{H}$ is a reflexive Banach space, as a consequence of the Riesz Representation Theorem, this weak topolgoy conicides with the weak-$*$ topology on $\mathcal{H}$ considered as the dual of $\mathcal{H}$ through the canonical sesquilinear isometry of $\mathcal{H}$ onto $\mathcal{H}$. By the Alaoglu's Theorem, it then follows that the norm-closed unit ball B in $\mathcal{H}$ is weakly compact, while we have seen that in an infinite dimensional Hilbert space B is never compact in the norm topology.

Specializing the Banach space construction to the Hilbert space setting, the weak topology on a Hilbert space $\mathcal{H}$ is the weakest topology for which all of the functions $f \mapsto \langle g, f\rangle$, $g \in \mathcal{H}$, are continuous, and a neighborhood base of 0 is given as follows:

Let $\mathcal{F} = \{f_1, \dots, f_n\}$ be a finite subset of $\mathcal{H}$, and let $\epsilon > 0$. Define the set $V_{\mathcal{F},\epsilon}$ by

$$V_{\mathcal{F},\epsilon} = \{g \in \mathcal{H} \ : \ |\langle f, g\rangle| < \epsilon \text{ for all } f \in \mathcal{F}\} \tag{3.5.1}$$

Note that each $V_{\mathcal{F},\epsilon}$ is convex, balanced, and absorbing. Let $\mathscr{V}$ denote the collection of all such set $V_{\mathcal{F},\epsilon}$, and $\mathcal{F}$ ranges over the finite subsets of $\mathcal{H}$ and ϵ ranges over $(0, \infty)$. Since every $L \in \mathcal{H}^*$ is of the form L_f for some $f \in \mathcal{H}$, by Theorem 3.2.2 once more, this is the weak topology for $\mathcal{H}$ considered as a Banach space. Since $\mathcal{H}^*$ separates points, the weak topology is Hausforff. Also, as in any Banach space, the norm-closed unit ball B is weakly closed.

As a direct consequence of our general Banach space results, we have:

3.5.1 THEOREM. *Let $\mathcal{H}$ be a Hilbert space. A sequence $\{f_n\}_{n\in\mathbb{N}}$ in $\mathcal{H}$ has the limit $f \in \mathcal{H}$ under the weak topology if and only if for all $g \in \mathcal{H}$,*

$$\langle g, f\rangle = \lim_{n\to\infty} \langle g, f_n\rangle. \tag{3.5.2}$$

Every weakly convergence sequence $\{f_n\}_{n\in\mathbb{N}}$ is bounded: $\sup_{n\in\mathbb{N}}\{\|f_n\|\} < \infty$.

3.5.2 Alaoglu's Theorem for Hilbert Space

3.5.2 THEOREM (Alaoglu's Theorem for Hilbert Space)**.** *Let $\mathcal{H}$ be a Hilbert space and let $B = \{f \in \mathcal{H} \ : \ \|f\| \leq 1\}$. Then B is compact in the weak topology on $\mathcal{H}$.*

Proof. Since $\mathcal{H}$ is reflexive, the weak-$*$ topology on $\mathcal{H}^*$ coincides with the weak topology on $\mathcal{H}$ under the idetification of $\mathcal{H}$ and $\mathcal{H}^*$ through the canonical sesquilinear embedding. Alaoglu's Theorem in the general Banach space setting then yields the result. □

Sequential compactness is the same as compactness for metric topologies, but not in general. In an infinite dimensional Banah space X, the weak topology is never metrizable, but when the dual X^* is separable, then the relative weak topology is metrizable on bounded subsets of X. This is not the case when $\mathcal{H}$ is not separable, but nonethless, bounded weakly closed subsets of of a Hilbert sapce are always weakly sequentially compact.

3.5.3 THEOREM. *Let $\mathcal{H}$ be a Hilbert space, and let B be the norm-closed unit ball. For every $r \in (0, 1)$, rB is weakly sequentially compact.*

Proof. Consider any sequence $\{f_n\}_{n\in N}$ in a Hilbert space $\mathcal{H}$ such that for some $f_n \in rB$ for all $n \in \mathbb{N}$. Let $\mathcal{K}$ be the norm closure of the span on $\{f_n\}_{n\in N}$. The set of all finite linear combintations of vectors in $\{f_n\}_{n\in N}$ with coefficients that have rational real and imaginary parts is countable and dense in $\mathcal{K}$. Hence $\mathcal{K}$, as a subspace of $\mathcal{H}$, is a Hilbert space in its own right, and is a separable Hilbert space. Since the relaitve weak topology on $rB \cap \mathcal{K}$ is metrizable, there is a subsequence $\{f_{n)_k}\}_{k\in N}$ that converges weakly to some $f \in rB \cap \mathcal{K}$. Then for all $g \in \mathcal{K}$, $\lim_{n\to\infty}\langle g, f_{n_k}\rangle = \langle g, f\rangle$ while for all $g \in \mathcal{K}^\perp$, $\langle g, f_{n_k}\rangle = \langle g, f\rangle = 0$ for all n. It follows that for all $g \in \mathcal{H}$, $\lim_{n\to\infty}\langle g, f_{n_k}\rangle = \langle g, f\rangle$, and thus $\{f_{n)_k}\}_{k\in N}$ converges weakly to f in $\mathcal{H}$. □

Corollary 3.5.3 will be useful one we have useful methods for identifying weakly closed subsets of $\mathcal{H}$. We already know, from Alaoglu's Theorem, that norm closed unit ball B is weakly compact, and therefore weakly closed. In the next section we shall prove a significant generalization of this: Every norm closed convex set is weakly closed.

3.5.3 Separation theorems

3.5.4 THEOREM. *Let K be a non-empty, norm-closed convex set in a Hilbert space $\mathcal{H}$, and let $f \in \mathcal{H}$ with $f \notin K$. Then there exists $\epsilon > 0$ and $g_0 \in \mathcal{H}$ such that*

$$\Re(\langle g_0, h\rangle) \geq \epsilon + \Re(\langle g_0, f\rangle) \qquad \text{for all } h \in K \ . \tag{3.5.3}$$

3.5.5 Remark. This result is an immediate consequence of the Hahn-Banach Separation Theorem. However, like everything else pertaining to the Hahn-Banach Theorem, in Hilbert space there is a simpler method of proof that we now give.

Proof of Thereom 3.5.4. Let $K_f = K - f = \{h - f \ : \ h \in K\}$. Since K_f is a non-empty closed convex set, the Projection Lemma provides the existence of a unique element $g_0 \in K_f$ of minimal norm, and since $f \notin K$, $g_0 \neq 0$. Hence for all $h \in K$ and $t > 0$,

$$\|t(h-f) + (1-t)g_0\|^2 \geq \|g_0\|^2$$

The left hand side equals $\|g_0 + t(h-f-g_0)\|^2 = \|g_0\|^2 + 2t\Re(\langle g_0, h-f-g_0\rangle) + t^2\|h-f-g_0\|^2$. Therefore,

$$\Re(\langle g_0, h-f-g_0\rangle) + \frac{t}{2}\|h-f-g_0\|^2 \geq 0$$

for all $t \in (0,1)$, and hence $\Re(\langle g_0, h-f-g_0\rangle) \geq 0$. This is the same as $\Re(\langle g_0, h-f\rangle) \geq \|g_0\|^2$, which, for $\epsilon = \|g_0\|^2 > 0$, yields (3.5.3). □

For $g \in \mathcal{H}$ and $\lambda \in \mathbb{R}$, define the *half-space*

$$H_{g,\lambda} = \{h \in \mathcal{H} \ : \ \Re(\langle g, h\rangle) \geq \lambda\} \ . \tag{3.5.4}$$

since the function $h \mapsto \Re(\langle g, h\rangle)$ is weakly continuous, $H_{g,\lambda}$ is weakly closed. Theorem 3.5.4 says that if K be a non-empty, closed convex set in a Hilbert space $\mathcal{H}$, and $f \in \mathcal{H}$ but $f \notin K$, there is a closed half space $H_{g,\lambda}$ such that $K \subset H_{g,\lambda}$ but $f \notin H_{g,\lambda}$. Therefore,

$$K = \bigcap\{H_{g,\lambda} \ : \ K \subset H_{g,\lambda}\} \ .$$

This displays K as the intersection of weakly closed sets, and we have proved:

3.5.6 THEOREM. *Let $\mathcal{H}$ be a Hilbert space. Every norm closed convex set $K \subset \mathcal{H}$ is weakly closed.*

3.5.4 From weak convergence to norm convergence

Combining corollary 3.5.3 and Theorem 3.5.6, we conclude that in a closed, bounded convex subset K of a separable Hilbert space $\mathcal{H}$, one can extract a weakly convergent subsequence $\{f_{n_k}\}_{k\in\mathbb{N}}$ from any infinite sequence $\{g_n\}_{n\in\mathbb{N}}$ in K.

It is useful to know when such a sequence also converges in the norm topology. The next theorem, provides a useful criterion.

3.5.7 THEOREM. *Let $\mathcal{H}$ be a Hilbert space, and let $\{f_n\}_{n\in N}$ be a weakly convergent sequence in $\mathcal{H}$ with limit f. Then*

$$\|f\| \leq \liminf_{n\to\infty} \|f_n\| \ , \tag{3.5.5}$$

and if $\|f\| = \lim_{n\to\infty} \|f_n\|$, then $\lim_{n\to\infty} \|f - f_n\| = 0$. That is, weak convergence, together with convergence of the norms, implies norm convergence.

Proof. Suppose on the contrary that $\|f\| > \liminf_{n\to\infty} \|f_n\|$. Then for some $\epsilon > 0$, there is a subsequence $\{f_{n_k}\}_{k\in\mathbb{N}}$ such that $\|f_{n_k}\| \leq \|f\| - \epsilon$ for all $k \in \mathbb{N}$. But since the subsequence also converges weakly to f,

$$\|f\|^2 = \Re(\langle f, f\rangle) = \lim_{k\to\infty} \Re(\langle f, f_{n_k}\rangle) \leq \|f\|\|f_{n_k}\| \leq \|f\|(\|f\| - \epsilon) \ ,$$

which is impossible. This proves (3.5.5).

Now suppose that $\|f\| = \lim_{n\to\infty} \|f_n\|$. Since

$$\|f - f_n\|^2 = \|f\|^2 + \|f_n\|^2 - 2\Re(\langle f, f_n\rangle) \ ,$$

$$\lim_{n\to\infty} \|f - f_n\|^2 = \|f\|^2 + \lim_{n\to\infty} \|f_n\|^2 - 2\lim_{n\to\infty} 2\Re(\langle f, f_n\rangle) = 0 \ .$$

□

3.6 Excercises

1. Let $\mathcal{H}$ be a separable, infinite dimensional Hilbert space, and let $\{u_n\}_{n\in\mathbb{N}}$ be an orthonormal basis for $\mathcal{H}$. Let $\{c_j\}_{\mathrm{J}\in\mathbb{N}}$ be a given sequence of non-negative numbers, and define

Let $C \subset \mathcal{H}$ be defined by

$$C = \{f \in \mathcal{H} \ : \ \|f\| \leq 1 \quad \text{and} \quad |\langle u_j, f\rangle| \leq c_j \quad \text{for all } j\} \ .$$

Show that C is always closed and bounded, but is compact if and only if $\sum_{j=1}^{\infty} c_j^2 < \infty$. Taking each $c_j = 1$, C becomes the unit ball in $\mathcal{H}$, and thus the unit ball is not compact.

2. For real valued square integrable functions f on $[-1,1]$, compute

$$\max\{\int_{[-1,1]} x^3 f(x)\mathrm{d}m \ : \ \int_{[-1,1]} x^j f(x)\mathrm{d}m = 0 \quad \text{for} \quad j = 0,1,2 \quad \text{and} \quad \int_{[-1,1]} f^2(x)\mathrm{d}m = 1\}$$

3. Show that if E is any Borel set in $(0, 2\pi]$ then

$$\lim_{j\to\infty} \int_E \cos(jx)\mathrm{d}m = \lim_{j\to\infty} \int_E \sin(jx)\mathrm{d}m = 0 \ .$$

Next, consider any increasing sequence $\{n_k\}$ of the natural numbers. Define E to be the set of all x for which

$$\lim_{k\to\infty} \sin(n_k x) \quad \text{exists} .$$

Show that $m(E) = 0$. (The identity $2\sin^2 x = 1 - \cos(2x)$ and the first part may prove useful.)

4. Show that in any Hilbert sapce $\mathcal{H}$, there is a continuous curve $t \mapsto f(t) \in \mathcal{H}$ defined for $t \in 0, \infty)$ such that for all $r < s < t$, $\langle f(t) - f(s), f(s) - f(r)\rangle = 0$. That is, the curve is constantly making rangle angle turns. Do this first wehn $\mathcal{H} = L^2([0,\infty), \mathscr{B}, \mu)$ where μ is Lebesgue measure, and then deduce the general case as a consequence.

5. Let $\mathcal{H}$ be a seperable Hilbert sapce and let $\{u_n\}_{n\in\mathbb{N}}$ be an orthonormal basis for $\mathcal{H}$. Let $\{v_n\}_{n\in\mathbb{N}}$ be another orthonormal sequence in $\mathcal{H}$.

(a) Suppose that $\sum_{n=1}^\infty \|u_n - v_n\|^2 < 1$. Use Bessel's inequality and Parseval's identity to show that $\{v_n\}_{n\in\mathbb{N}}$ is complete; i.e., that $\{v_n\}_{n\in\mathbb{N}}$ is also an orthonormal basis for $\mathcal{H}$.

(b) Suppose that $\sum_{n=1}^\infty \|u_n - v_n\|^2 < \infty$. Pick N large enough that $\sum_{n=N+1}^\infty \|u_n - v_n\|^2 < 1$. Show that if $f \in \mathcal{H}$ is orthognal to u_n for $1 \le n \le N$, and is orthogonal to v_n for $n \ge N+1$, then $g = 0$. Then let V denote the closed span of $\{v_{N+1}, v_{N+2}, \dots\}$, and show that $V^\perp$ is a subsapce of dimension N.

(c) Show that when $\sum_{n=1}^\infty \|u_n - v_n\|^2 < \infty$, $\{v_n\}_{n\in\mathbb{N}}$ is an orthonormal basis.

Commentary: Exercise 5 is based on a result of Birkhoff and Rota who apply it to show completeness of the orthonormal sequences of Sturm-Liouville eigenfunctions. It is known, going back to Liouville himself, that if $\{v_n\}_{n\in\mathbb{N}}$ is the normalized sequence of eigenfunctions of a Sturm-Liouville operator with some boundary considtions, and $\{u_n\}_{n\in\mathbb{N}}$ is the Fourier basis for the same boundaty conditions, then for some constant $C < \infty$, $\|u_n - v_n\|^2 \le Cn^{-2}$ for all n. Hence, the condition in part **(c)** is applicable, and the completeness of $\{v_n\}_{n\in\mathbb{N}}$ follows from the known completeness of $\{u_n\}_{n\in\mathbb{N}}$.

6. Let $\mathcal{H}$ be a Hilbert space, and let $\{u_n\}_{n\in\mathbb{N}}$ be orthonormal in $\mathcal{H}$. Let $\{v_n\}_{n\in\mathbb{N}}$ be any sequence of vectors in $\mathcal{H}$.

(a) Prove that the following are equivalent:

(1) There exists some $C \in (0,\infty)$, whenever $N \in \mathbb{N}$ and $\{\alpha_1, \dots, \alpha_N\} \in \mathbb{C}^N$,

$$\left\|\sum_{n=1}^N \alpha_n v_n\right\|^2 \le C \sum_{n=1}^N |\alpha_n|^2 . \tag{$*$}$$

(2) For all $f \in \mathcal{H}$, $Af := \sum_{n=1}^\infty \langle f, u_n\rangle v_n$ is norm-convergent, and $f \mapsto Af$ is a bounded linear operator on $\mathcal{H}$.

Show moreover that in this case $\|A\|$ is the least value of C for which $(*)$ is valid.

(b) Prove that $(*)$ is satisfed in case $\{v_n\}_{n\in\mathbb{N}}$ is norm-bounded, and $\sum_{m\neq n} |\langle v_m, v_n\rangle|^2 < \infty$.

7. Let $\mathcal{H}$ be a Hilbert space, and let V and W be closed subspaces of $\mathcal{H}$. Let P_V be the orthogonal projection onto V, and let P_W be the orthogonal projection onto W. Let P denote the ortohogonal projection onto $V \cap W$. Define $T = P_V P_W$.

(a) Show that if Q is the orthogonal projection onto any closed subspace of $\mathcal{H}$, then for all $f \in \mathcal{H}$,

$$\|Qf - f\|^2 = \|f\|^2 - \|Qf\|^2 .$$

(b) Use the telscoping sum $Tf = (P_V P_W f - P_W f) + (P_W f - f)$ and part **(a)** to show that for all $f \in H$,

$$\|Tf - f\|^2 \leq 2(\|f\| - \|Tf\|^2) \ .$$

(c) Show that for all $f \in \mathcal{H}$, $\lim_{n\to\infty} \|T^n f\|$ exists, and then that $\lim_{n\to\infty} \|T^n(T-I)f\| = 0$.

(d) Show that $\ker(T^* - I) = V \cap W$, and that for $g \in \text{ran}(T - I)$, $\lim_{n\to\infty} T^n g = 0$. Finally, prove that for all $f \in \mathcal{H}$, $\lim_{n\to\infty} T^n f = Pf$. (This result is due to von Neumann; the proof suggested by the components of the problem is due to Kakutani. This proof may be readily genralized to products of arbitrarily many projections.)

8. Let $\mathcal{H}$ be a Hilbert space. The *closed convex hull* of a $A \subset \mathcal{H}$, denoted $\overline{\text{co}}(A)$, is the intersection over all of the closed convex sets containing A.

(a) For $A \subset \mathcal{H}$, $g \in \mathcal{H}$ is said to be a *finite convex combination of elements of* A in case for some finite sets $\{f_1, \dots, f_n\} \subset A$ and $\{t_1, \dots, t_n\} \subset [0,1]$, $\sum_{k=1}^n t_k = 1$ and $\sum_{k=1}^n t_k f_k = g$. Show that for all $A \subset \mathcal{H}$, $\overline{\text{co}}(A)$ is the norm closure of the set of finite convex combinations of elements of A.

(b) Let $\{f_n\}_{n\in\mathbb{N}}$ be a weakly convergent sequence in $\mathcal{H}$ with weak limit f. For each $N \in$
N, let

$$K_N := \overline{\text{co}}\left(\{f_n\}_{n\geq N}\right) \ .$$

Show that each K_N is bounded and weakly compact, and that $f \in K_N$ for all $N \in N$. Then show that there exists a sequence $\{g_n\}_{n\in\mathbb{N}}$ such that for each $n \in \mathbb{N}$, g_n is a finite convex combination vectors in $\{f_n\}_{n\geq N}$ and such that $\lim_{n\to\infty} \|g_n - f\| = 0$.

9. (a) Let T be an operator on a Hilbert space $\mathcal{H}$ such that $\langle f, Tf\rangle \in \mathbb{R}$ for all $f \in \mathcal{H}$. Let $S = T - T^*$, Show that $\langle f, Sf\rangle = 0$ for all $f \in \mathcal{H}$.

(b) Let S be an operator on a complex Hilbert space $\mathcal{H}$ such that $\langle f, Sf\rangle = 0$ for all $f \in \mathcal{H}$ considering $f = g + h$ and $f = g + ih$, show that $\langle g, Sh\rangle = 0$, and then that $S = 0$.

(c) A bounded operator T Hilbert space $\mathcal{H}$ is said to be *positive* in case $\langle f, Af\rangle \geq 0$ for all $f \in \mathcal{H}$. Show that every positive bounded operator T on a complex Hilbert space is slef-adjoint; i.e., $T = T^*$.

10. Let T be a positive bounded operator on a complex Hilbert space $\mathcal{H}$. (See Exercise 9.)

(a) For all $f, g \in \mathcal{H}$, define a sesquilinear form $\langle \cdot, \cdot\rangle_T$ on $\mathcal{H}$ by $\langle f, Tg\rangle_T = \langle f, g\rangle$. Show that $\langle \cdot, \cdot\rangle_T$ is an inner product on $\mathcal{H}$.

(b) Define a function F on $\mathcal{H}$ by $F(f) = \frac{1}{2}\langle f, Tf\rangle = \langle f, Tg\rangle_T$. Show that F is convex on $\mathcal{H}$.

(c) Show that the function F is sequentially weakly lower semicontinuous on $\mathcal{H}$. That is, if $\{f_n\}_{n\in\mathbb{N}}$ converges weakly to f, then $F(f) \leq \liminf_{n\to\infty} F(f_n)$. To do this, use part **(b)** and Exercise 8.

11. An operator T on a Hilbert space H is *uniformly positive* or *coercive* in case for some $\epsilon > 0$, $\langle f, Tf\rangle \geq \epsilon\|f\|^2$ for all $f \in \mathcal{H}$. Let T be coercive on $\mathcal{H}$.

(a) Show that for all $f \in \mathcal{H}$ there is a unique $g_f \in \mathcal{H}$ such that for all $g \neq g_f$,

$$\Re\langle f, g\rangle - F(g) < \Re\langle f, g_f\rangle - F(g_f) \ .$$

and that $Tg_f = f$ and $\|g_f\| \leq \epsilon^{-1} f$.

(b) Show that T has a bounded inverse and that

$$\frac{1}{2}\langle g, T^{-1}g\rangle = \sup_{f\in\mathcal{H}}\{\Re\langle f, g\rangle - \langle f, Tf\rangle\} \ .$$

12. Let $T \in \mathscr{B}(\mathcal{H})$ be self-adjoint. Show that $T + iI$ is invertible. Then define an operator U_T by

$$U_T := (T - iI)(T + iI)^{-1} \ .$$

Show that U_T is unitary, and that $U_T - I$ is invertible. Finally, show that $T = i(I + U_T)(I - U_T)^{-1}$, and that if U is unitary and $U - I$ is invertible, then $i(I + U)(I - U)^{-1}$ is self adjoint. Thus the map $T \mapsto U_T$ is a one-to-one map from the set of self adjoint memebers of $\mathscr{B}(\mathcal{H})$ onto the set of unitary operators $U \in \mathscr{B}(\mathcal{H})$ such that $U - I$ is invertible.

Chapter 4

Compact operators on Hilbert Space

4.0.1 Norms of self-adjoint operators on Hilbert space

Let $T \in \mathscr{B}(\mathcal{H})$, $\mathcal{H}$ a Hilbert space. Since for all $f \in \mathcal{H}$, there is a unit vector v such that $\langle v, f\rangle = \|f\|$, and since $|\langle v, f\rangle| \leq \|f\|$ for all unit vectors v, $\|Tu\| = \sup\{|\langle v, Tu\rangle| \ : \|v\| = 1\}$. Therefore,

$$\|T\| = \sup\{|\langle v, Tu\rangle| \ : \|u\| = 1 \ , \|v\| = 1\} \ . \tag{4.0.1}$$

When T is self-adjoint; i.e., when $T = T^*$, there is a simpler formula:

4.0.1 LEMMA. *Let $T \in \mathscr{B}(\mathcal{H})$, $T = T^*$. Then*

$$\|T\| = \sup\{|\langle u, Tu\rangle| \ : \|u\| = 1 \ \} \ . \tag{4.0.2}$$

Proof. Temporarily define $C_T := \sup\{|\langle u, Tu\rangle| \ : \|u\| = 1 \ \}$. Let u, v be unit vectors in $\mathcal{H}$ such that $v \neq \pm u$. Since T is self-adjoint,

$$\begin{aligned} \langle u+v, T(u+v)\rangle &= \langle u, Tu\rangle + \langle v, Tv\rangle + 2\Re(\langle v, Tu\rangle) \\ \langle u-v, T(u-v)\rangle &= \langle u, Tu\rangle + \langle v, Tv\rangle - 2\Re(\langle v, Tu\rangle) \ . \end{aligned}$$

Therefore,

$$\Re(\langle v, Tu\rangle) = \frac{1}{4}\left(\langle u+v, T(u+v)\rangle - \langle u-v, T(u-v)\rangle\right) \ .$$

Defining $f = u+v$ and $g = u-v$, neither of which is zero, we obtain

$$\begin{aligned} \Re(\langle v, Tu\rangle) &= \frac{1}{4}\left(\|f\|^2\frac{\langle f, Tf\rangle}{\|f\|^2} - \|g\|^2\frac{\langle g, Tg\rangle}{\|g\|^2}\right) \\ &\leq \frac{1}{4}(\|f\|^2 + \|g\|^2)C_T = C_T \ . \end{aligned}$$

Replacing v by $e^{i\theta}v$ and varying θ, we obtain that $|\langle v, Tu\rangle| \leq C_T$ for all unit vectors u and v, the excluded case $v = \pm u$ being trivial. Now (4.0.2) follows from (4.0.1). □

If $T \in \mathscr{B}(\mathcal{H})$, then $\|T^*\| = \|T\|$, which follows from (4.0.1) since $|\langle v, T^*u\rangle| = |\langle u, Tv\rangle|$. Also, for all unit vectors u and all $S, T \in \mathscr{B}(\mathcal{H})$, $\|STu\| \leq \|S\|\|Tu\| \leq \|S\|\|T\|$ so that $\|ST\| \leq \|S\|\|T\|$. In particular, $\|T^2\| \leq \|T\|^2$, and then by a simple induction, $\|T^n\| \leq \|T\|^n$ for all $n \in \mathbb{N}$. The inequality can be strict: For example, if T is nilpotent of order n, so that $T^n = 0$, but $T \neq 0$, then $0 = \|T^n\| < \|T\|^n$.

4.0.2 THEOREM. *For all $T \in \mathscr{B}(\mathcal{H})$, $\mathcal{H}$ a Hilbert space, $\|T^*T\| = \|T\|^2$.*

Proof. Since T^*T is self-adjoint, $\|T^*T\| = \sup\{|\langle u, T^*Tu\rangle| \ : \|u\| = 1\ \}$. However, $\langle u, T^*Tu\rangle = \langle Tu, Tu\rangle = \|Tu\|^2$, so that

$$\|T^*T\| = \sup\{\|Tu\|^2 \ : \|u\| = 1\ \} = \|T\|^2 \ .$$

□

In particular, if T is self-adjoint, then $\|T^2\| = \|T\|^2$, and then it follows that $\|T^n\| = \|T\|^n$ for all $n \in \mathbb{N}$. The identity $\|T^*T\| = \|T\|^2$ is known as the C^* *algebra identity* because of its crucial role in the Gelfand-Naimark theory of C^* algebras.

4.0.2 Compact operators

4.0.3 DEFINITION (Compact operator)**.** An operator $T \in \mathscr{B}(\mathcal{H})$, $\mathcal{H}$ a Hilbert space, is *compact* in case whenever $\{f_n\}_{n\in\mathbb{N}}$ is a weakly convergent sequence in $\mathcal{H}$, $\{Tf_n\}_{n\in\mathbb{N}}$ is a strongly convergent sequence in $\mathcal{H}$.

4.0.4 EXAMPLE. *Let $(\Omega, \mathscr{M}, \mu)$ be a measure space, and let $\mathcal{H} = L^2(\Omega, \mathscr{M}, \mu)$. Let K be a square integrable function on the product space $(\Omega \times \Omega, \mathscr{M} \otimes \mathscr{M}, \mu \otimes \mu)$ and define*

$$\|K\|^2 = \int_{\Omega\times\Omega} |K|^2 \mathrm{d}\mu \otimes \mu \ .$$

Then for each $h \in \mathcal{H}$, and each $x \in \Omega$,

$$\left| \int_\Omega K(x,y)h(y)\mathrm{d}\mu(y) \right| \le \left(\int_\Omega |K(x,y)|^2 \mathrm{d}\mu(y) \right)^{1/2} \|h\| \ . \tag{4.0.3}$$

By Fubini's Theorem, the right hand side is a square integrable function of x, and hence for all $f \in \mathcal{H}$, the function Kf defined by

$$Kf(x) := \int_\Omega K(x,y)f(y)\mathrm{d}\mu(y) \tag{4.0.4}$$

belongs to $\mathcal{H}$, and moreover, $\|Kf\| \le \|K\|\|f\|$. Therefore, the map $f \mapsto Kf$ is a bounded linear transformation on $\mathcal{H}$.

In fact, K is compact. To see this, let $\{f_n\}$ be a sequence that converges weakly to $f \in \mathcal{H}$. By Fubini's Theorem, for almost every $x \in X$, $y \mapsto K(x,y)$ is square integrable, and thus we may write $Kf = \langle K(x,\cdot), f\rangle$, and then by the weak convergence, $Kf(x) = \lim_{n\to\infty} Kf_n(x)$ for almost every x. Moreover, since weakly convergence sequences are uniformly bounded, there exists $C < \infty$ such that $\|f_n\| \le C$ for all n, and then $\|f\| \le C$ as well since $\|f\| \le \liminf_{n\to\infty} \|f_n\|$. Therefore, by (4.0.3)

$$|K(f - f_n)(x)|^2 \le 4C^2 \left(\int_\Omega |K(x,y)|^2 \mathrm{d}\mu(y) \right) \ .$$

Then by the Lebesgue Dominated Convergence Theorem, $\lim_{n\to\infty} \|Kf - Kf_n\|^2 = 0$. Thus, $\{Kf_n\}_{n\in\mathbb{N}}$ converges strongly to f. This proves that K is a compact operator.

The class of compact operators considered in this example is called the class of Hilbert-Schmidt integral operators.

An even simpler example is given by the class of *finite rank operators* on a Hilbert space $\mathcal{H}$. First, we fix a useful notational convention. Given two vectors $f, g \in \mathcal{H}$, let $|g\rangle\langle f|$ denote the operator on $\mathcal{H}$ given by

$$|g\rangle\langle f|h = \langle f, h\rangle g \quad \text{for all } h \in \mathcal{H}\ . \tag{4.0.5}$$

An operator $T \in \mathscr{B}(\mathcal{H})$ has finite rank if its range, $\mathrm{ran}(T)$, is a finite dimensional subspace of $\mathcal{H}$, or equivalently, if the orthogonal complement of its null-space, $\ker(T)^\perp$, is finite dimensional. If T is finite rank and $\{f_1, \dots, f_m\}$ is an orthonormal basis for $\ker(T)^\perp$, then we may write T in the form

$$T = \sum_{j=1}^m |g_j\rangle\langle f_j|\ , \tag{4.0.6}$$

where for each j, $g_j = Tf_j$. Conversely, every operator of the form (4.0.6), even without the assumption that $\{f_1, \dots, f_m\}$ is orthonormal, is finite rank. It is very simple to show that every finite rank operators is compact; this is left to the reader.

4.0.5 THEOREM. *Let $T \in \mathscr{C}(\mathcal{H})$. Then $T^* \in \mathscr{C}(\mathcal{H})$, and for all $S \in \mathscr{B}(H)$, ST and TS belong to $\mathscr{C}(H)$. Finally, $\mathscr{C}(\mathcal{H})$ is an operator norm closed subspace of $\mathscr{B}(\mathcal{H})$.*

Proof. To show that $T^* \in \mathscr{C}(\mathcal{H})$ whenever $T \in \mathscr{C}(\mathcal{H})$, let $T \in \mathscr{C}(\mathcal{H})$ and let $\{f_n\}_{n\in\mathbb{N}}$ be a sequence in $\mathcal{H}$ that converges weakly to $f \in \mathcal{H}$. We must show that $\lim_{n\to\infty} \|T^*(f_n - f)\| = 0$. If this is not the case, then for some $\epsilon > 0$, there is a subsequence $\{f_{n_k}\}_{k\in\mathbb{N}}$ such that $\|T^*(f_{n_k} - f)\| \geq \epsilon$ for all k. Passing to this subsequence, $\|T^*(f_k - f)\| \geq \epsilon$ for all k. Then there exists a sequence $\{u_k\}_{k\in\mathbb{N}}$ of unit vectors such that

$$|\langle Tu_k, f_k - f\rangle| = |\langle u_k, T^*(f_k - f)\rangle| = \|T^*(f_k - f)\| \geq \epsilon$$

for all k. The norm closed unit ball B in $\mathcal{H}$ is weakly sequentially compact, and hence there exists a (further) subsequence $\{u_{k_\ell}\}_{\ell\in\mathbb{N}}$ that converges weakly to some $u \in B$. Since T is compact, $\lim_{\ell\to\infty} \|T(u_{k_\ell} - u)\| = 0$ and $\|f_{k_\ell} - f\|$ is bounded uniformly in ℓ. Therefore, for all sufficiently large ℓ, $\|T(u_{k_\ell} - u)\|\|f_{k_\ell} - f\| < \epsilon/2$, and then for all such ℓ,

$$\begin{aligned} |\langle Tu, f_{k_\ell} - f\rangle| &\geq |\langle Tu_{k_\ell}, f_{k_\ell} - f\rangle| - |\langle T(u_{k_\ell} - u), f_{k_\ell} - f\rangle| \\ &\geq |\langle Tu_{k_\ell}, f_{k_\ell} - f\rangle| - \|T(u_{k_\ell} - u)\|\|f_{k_\ell} - f\| \geq \frac{1}{2}\epsilon\ . \end{aligned}$$

However this is impossible since $\{f_n\}_{n\in N}$ converges weakly to f. This contradiction shows that T^* is compact.

It is evident that since S takes norm convergent sequences to norm convergent sequences, then ST is compact for all $S \in \mathscr{B}(\mathcal{H})$ and $T \in \mathscr{C}(\mathcal{H})$. Since $TS = (S^*T^*)^*$, the first part of the proof shows that TS is compact for all $S \in \mathscr{B}(\mathcal{H})$ and $T \in \mathscr{C}(\mathcal{H})$.

Now let $\{T_n\}_{n\in\mathbb{N}}$ be a norm convergent sequence in $\mathscr{C}(H)$ and let T be the limit in $\mathscr{B}(\mathcal{H})$. We must show that $T \in \mathscr{C}(\mathcal{H})$. Let $\{f_k\}_{k\in\mathbb{N}}$ be a weakly convergent sequence with limit f. Then there is a finite constant C such that $\|f_k\| \leq C$ for all k, and also $\|f\| \leq C$.

Pick $\epsilon > 0$, and then pick N so that $\|T_n - T\| < \epsilon$ whenever $n \geq N$. Then for all k, ℓ,

$$\|Tf_k - Tf_\ell\| \leq \|(T - T_N)f_k\| + \|T_N(f_k - f_\ell)\| + \|(T - T_N)f_\ell\| \leq C\epsilon + \|T_N(f_k - f_\ell)\| + C\epsilon\ .$$

Since T_N is compact, there is a finite M so that whenever $k, \ell \geq M$, $\|T_N(f_k - f_\ell)\| < \epsilon$. Altogether,

$$k, \ell \geq M \quad \Rightarrow \quad \|Tf_k - Tf_\ell\| \leq (2C+1)\epsilon \ .$$

Since $\epsilon > 0$ is arbitrary, this shows that $\{Tf_k\}$ is a Cauchy sequence. Let g denote the limit. Then for all $h \in \mathcal{H}$,

$$\langle h, g\rangle = \lim_{n\to\infty} \langle h, Tf_n\rangle = \lim_{n\to\infty} \langle T^*h, f_n\rangle = \langle T^*h, f\rangle = \langle h, Tf\rangle \ ,$$

and therefore $Tf = g = \lim_{n\to\infty} Tf_n$. □

A number $\lambda \in \mathbb{C}$ is called an *eigenvalue* of $T \in \mathscr{B}(\mathcal{H})$ in case there is a non-zero vector $f \in \mathcal{H}$ such that $Tf = \lambda f$, and in this case, f is called an *eigenvector* of T with the eigenvalue λ.

If λ is an eigenvalue of T, the corresponding *eigenspace* $\mathcal{H}_\lambda$ is the subspace of $\mathcal{H}$ spanned by all of the eigenvectors of T with eigenvalue λ. That is,

$$\mathcal{H}_\lambda = \ker(\lambda I - T) \ ,$$

which shows that $\mathcal{H}_\lambda$ is always closed.

4.0.6 THEOREM. *Let T be a compact operator on a Hilbert space $\mathcal{H}$. If if λ is an non-zero eigenvalue of T, then* $\dim(\ker(\lambda I - T)) < \infty$. *Moreover, for each $r > 0$, there are at most finitely many $\lambda \in \mathbb{C}$ such that λ is an eigenvalue of T and $|\lambda| \geq r$.*

Proof. If for any non-zero λ, $\dim(\ker(\lambda I - T)) = \infty$, then there exists an orthonormal sequence $\{u_n\}_{n\in\mathbb{N}}$ of eigenvectors of T, $Tu_n = \lambda_n u_n$, such that $\inf_{n\in\mathbb{N}}\{|\lambda_n|\} > 0$. Then $\{u_n\}_{n\in\mathbb{N}}$ converges weakly to zero, but $\{Tu_n\}_{n\in\mathbb{N}}$ does not converge strongly. If T is self-adjoint, so that eigenvectors with distinct eigenvalues are necessarily orthogonal, essentially the same argument can be used for the second part.

To prove the second part in general, suppose that there are infinitely many eigenvalues $\{\lambda_n\}_{n\in\mathbb{N}}$ with $|\lambda_n| \geq r$ for all n. For each n, let u_n be a unit vector with $Tu_n = \lambda_n u_n$. Passing to a subsequence, we may suppose that u_n converges weakly to $u \in B$ as $n \to \infty$. Define $\mathcal{K}_n = \mathrm{span}(\{u_1, \dots, u_n\})$. Since any set of eigenvectors with distinct eigenvalues is linearly independent, $\mathcal{K}_{n+1} \cap \mathcal{K}_n^\perp \neq \{0\}$ for any n. Define $v_1 = u_1$, and for all $n \in \mathbb{N}$, choose any unit vector $v_{n+1} \in \mathcal{K}_{n+1} \cap \mathcal{K}_n^\perp$. Since v_n converges weakly to 0 as $n \to \infty$, Then $\lim_{n\to\infty} \|Tv_n\| = 0$. Now note that $(\lambda_n I - T)v_n \in \mathcal{K}_{n-1}$ for each $n \geq 2$. Hence v_n is orthogonal to $(\lambda_n I - T)v_n$, and hence

$$\|Tv_n\|^2 = \|(T - \lambda_n)v_n + \lambda_n v_n\|^2 = \|(T - \lambda_n)v_n\|^2 + \|\lambda_n v_n\|^2 \geq r \ .$$

This contradiction shows that there do not exist infinitely many eigenvalues λ with $|\lambda| \geq r$. □

4.0.3 The Hilbert-Schmidt Spectral Theorem

In an infinite dimensional Hilbert space, bounded operators, even bounded self-adjoint operators, need not have any eigenvalues at all. For example, let $\mathcal{H} := L^2([0,1], \mathscr{B}, \mu)$ where μ is Lebesgue measure. Define $Tf(t) = tf(t)$. Then it is easily checked that $\|T\| = 1$ and $T = T^*$. However, if $Tf = \lambda f$, then $(t - \lambda)f(t) = 0$ for almost every t, and this is impossible unless $f = 0$. However, for compact self-adjoint operators, the situation is different.

4.0.7 THEOREM. *Let T be a self-adjoint compact operator on a Hilbert space $\mathcal{H}$. Then either $\|T\|$ or $-\|T\|$ is an eigenvalue of T (or both are).*

Proof. By Lemma 4.0.1, $\|T\| = \sup\{|\langle u, Tu\rangle| \ : \|u\| = 1\ \}$. Therefore, either

$$\|T\| = \sup\{\langle u, Tu\rangle \ : \|u\| = 1\ \} \quad \text{or} \quad \|T\| = \sup\{-\langle u, Tu\rangle \ : \|u\| = 1\ \}\ , \tag{4.0.7}$$

or both. Suppose first that $\|T\| = \sup\{\langle u, Tu\rangle \ : \|u\| = 1\ \}$. We may assume that $T \neq 0$ to avoid trivialities.

It follows that there exists a sequence of unit vectors $\{u_n\}_{n\in\mathbb{N}}$ such that $\lim_{n\to\infty}\langle u_n, Tu_n\rangle = \|T\|$. Since every bounded sequence contains a weakly convergent subsequence, we may select a subsequence $\{u_{n_k}\}_{k\in\mathbb{N}}$ that converges weakly to some $u \in \mathcal{H}$ with $\|u\| \leq 1$.

We now show that in fact $\|u\| = 1$ and $\langle u, Tu\rangle = \|T\|$. Note that

$$\langle u, Tu\rangle - \langle u_{n_k}, Tu_{n_k}\rangle = \langle u - u_{n_k}, Tu\rangle + \langle u_{n_k}, T(u - u_{n_k})\rangle\ .$$

Since $\{u - u_{n_k}\}_{k\in\mathbb{N}}$ converges weakly to 0, $\lim_{k\to\infty}\langle u - u_{n_k}, Tu\rangle = 0$. Moreover, since T is compact, $\{T(u - u_{n_k})\}_{k\in\mathbb{N}}$ converges strongly to zero and hence

$$\limsup_{k\to\infty} |\langle u_{n_k}, T(u - u_{n_k})\rangle| \leq \limsup_{k\to\infty} \|T(u - u_{n_k})\| = 0\ .$$

Therefore

$$\langle u, Tu\rangle - \|T\| = \lim_{k\to\infty}\left(\langle u, Tu\rangle - \langle u_{n_k}, Tu_{n_k}\rangle\right) = 0\ .$$

Hence $\|T\| = \langle u, Tu\rangle$. By the Cauchy-Schwarz inequality, $\|T\| = |\langle u, Tu\rangle| \leq \|T\|\|u\|^2$. Since $\|u\| \leq 1$, we must have $\|u\| = 1$.

Now that we have found a unit vector u such that $\langle u, Tu\rangle = \|T\| \geq \langle v, Tv\rangle$ for all unit vectors $v \in \mathcal{H}$, let h be any unit vector, and define the function φ on $(-1/2, 1/2)$ by

$$\varphi(t) = \frac{\langle u + th, T(u + th)\rangle}{\|u + th\|^2}\ , \tag{4.0.8}$$

and note that $\varphi(0) \geq \varphi(t)$ for all $t \in (-1/2, 1/2)$. Since T is self adjoint, and since $\langle u, Tu\rangle = \|T\|$,

$$\begin{aligned}
\langle u + th, T(u + th)\rangle &= \langle u, Tu\rangle + t\langle h, Tu\rangle + t\langle u, Th\rangle + t^2\langle h, Th\rangle \\
&= \|T\| + t\langle h, Tu\rangle + t\langle Tu, h\rangle + t^2\langle h, Th\rangle \\
&= \|T\| + 2t\Re(\langle Tu, h\rangle) + t^2\langle h, Th\rangle\ .
\end{aligned}$$

Therefore,

$$\varphi(t) = \frac{\|T\| + 2t\Re(\langle Tu, h\rangle + t^2\langle h, Th\rangle}{1 + t2\Re(\langle u, h\rangle) + t^2}\ ,$$

Computing the derivative, we find $0 = \Re(\langle Tu, h\rangle - \|T\|\Re(\langle u, h\rangle)$. Replacing h by ih, we see that also $\Im(\langle Tu, h\rangle = \|T\|\Im(\langle u, h\rangle$, and altogether that $\langle Tu - \|T\|u, h\rangle = 0$ for all unit vectors h, and hence for all vectors h. Taking $h = Tu - \|T\|u$, we conclude that $Tu = \|T\|u$. Thus, u is an eigenvector of T with eigenvalue $\|T\|$.

Now suppose that the second alternative in (4.0.7) is valid. This is the same as $\|T\| = -\inf\{\langle u, Tu\rangle \ : \|u\| = 1\}$. The same reasoning proves the existence of a unit vector u such that $-\|T\| = \langle u, Tu\rangle$. Defining $\varphi(t)$ exactly as in (4.0.8), we have this time that $\varphi(0) \leq \varphi(t)$ for all $t \in (-1/2, 1/2)$. Again, this means $\varphi'(0) = 0$, and computing the derivative as above we find that $Tu = -\|T\|u$. $\square$

The following simple lemmas will be frequently useful.

4.0.8 LEMMA. *Let T be a self-adjoint operator on a Hilbert space $\mathcal{H}$. Suppose that $\mathcal{K}$ is a subspace of $\mathcal{H}$ such that $\mathscr{K}$ is invariant under T, meaning that $Tf \in \mathcal{K}$ for all $f \in \mathcal{K}$. Then $\mathcal{K}^\perp$ is also invariant under T.*

Proof. Let $f \in V$, and $g \in V^\perp$. Then $0 = \langle Tf, g\rangle = \langle f, Tg\rangle$ so that $Tg \in V^\perp$. □

In particular, if T is a self-adjoint operator on $\mathcal{H}$, and $\mathcal{K} = \ker(T)$, then $\mathcal{K}^\perp$ is invariant under T, and being a closed subspace of $\mathcal{H}$, $K^\perp$ is a Hilbert space in its own right. If T is a compact self-adjoint operator, the restriction of T to $\mathcal{K}^\perp$, $T|_{\mathcal{K}^\perp}$, is an injective compact self-adjoint operator on $\mathcal{K}^\perp$.

Next, any eigenvalues of a self-adjoint operator on a Hilbert space $\mathcal{H}$ are necessarily real:

4.0.9 LEMMA. *Let T be a self-adjoint operator on a Hilbert space $\mathcal{H}$. Suppose that $f \neq 0$ and that $Tf = \lambda f$ for some $\lambda \in \mathbb{C}$. Then $\lambda \in \mathbb{R}$.*

Proof. Let $u = \|f\|^{-1}f$. Then $\lambda = \langle u, Tu\rangle = \langle Tu, u\rangle = \overline{\langle u, Tu\rangle} = \overline{\lambda}$. □

4.0.10 THEOREM (Hilbert-Schmidt Spectral Theorem)**.** *Let T be a self-adjoint compact operator on a Hilbert space $\mathcal{H}$. Let $\mathcal{K} = \ker(T)$. If $\dim(\mathcal{K}^\perp) =: m < \infty$, define the index set $\mathscr{J}$ to be $\{1, \dots, m\}$. Otherwise, define $\mathscr{J} := \mathbb{N}$. Then there exists an orthonormal basis $\{u_n\}_{n\in\mathscr{J}}$ for $\mathcal{K}^\perp$ consisting of eigenvectors of T with $Tu_j = \lambda_j u_j$ for all $j \in \mathbb{N}$ such that λ_j is real, $|\lambda_k| \le |\lambda_j|$ for all $j < k \in \mathscr{J}$ and $|\lambda_1| = \|T\|$. Moreover,*

$$T = \sum_{j\in\mathscr{J}} \lambda_j |u_j\rangle\langle u_j| \ , \tag{4.0.9}$$

where, in case $\mathscr{J} = \mathbb{N}$, the sum converges in the operator norm and $\lim_{n\to\infty} \lambda_n = 0$.

Proof. By restricting T to $(\ker(T))^\perp$, we may assume without loss of generality that $\ker(T) = \{0\}$, which we do in order to simplify the notation.

By Theorem 4.0.7, either $\|T\|$ or $-\|T\|$ is an eigenvalue of T, or both are. Let u_1 be an eigenvector of T with eigenvalue λ_1, where either $\lambda_1 = \|T\|$ or $\lambda_1 = -\|T\|$. If $\dim(\mathcal{H}) = 1$, we are done.

Otherwise, define $\mathcal{K}_1 = \mathrm{span}(\{u_1\})$. Then $\mathcal{K}_1$ is invariant under T, and then by Lemma 4.0.8, $\mathcal{K}_1^\perp$ is a invariant under T. Since the orthogonal complement of any set is closed, $\mathcal{K}_1^\perp$ is closed, and hence is a Hilbert space. Since $\mathcal{K}_1^\perp$ is invariant under T, the restriction of T to $\mathcal{K}_1$, $T\big|_{\mathcal{K}_1^\perp}$, is a self-adjoint compact operator on $\mathcal{K}_1^\perp$.

Therefore, we may apply Theorem 4.0.7 to $T\big|_{\mathcal{K}_1^\perp}$: Either $\left\|T\big|_{\mathcal{K}_1^\perp}\right\|$ or $-\left\|T\big|_{\mathcal{K}_1^\perp}\right\|$ is an eigenvalue of $T\big|_{\mathcal{K}_1^\perp}$. Choose u_2 to be an eigenvector of $T\big|_{\mathcal{K}_1^\perp}$ with eigenvalue $\lambda_2 = \pm\left\|T\big|_{\mathcal{K}_1^\perp}\right\|$. Evidently, $\left\|T\big|_{\mathcal{K}_1^\perp}\right\| \le \|T\|$ and hence $|\lambda_2| \le |\lambda_1|$. If $\dim(\mathcal{H}) = 2$, we are done. Otherwise we iterate.

Suppose we have found an orthonormal set $\{u_1, \dots, u_m\}$ consisting of eigenvectors of T with $Tu_j = \lambda u_j$ and $|\lambda_j| \le |\lambda_k|$ for all $1 \le j < k \le m$. Suppose that $\dim(\mathcal{H}) > m$. Define $\mathcal{K}_m := \mathrm{span}(\{u_1, \dots, u_m\})$, which, being spanned by eigenvectors, is evidently invariant under T. By Lemma 4.0.8, $\mathcal{K}_m^\perp$ is a closed subspace of $\mathcal{H}$ that is invariant under T.

Therefore, we may apply Theorem 4.0.7 to $T\big|_{\mathcal{K}_m^\perp}$: Either $\left\|T\big|_{\mathcal{K}_m^\perp}\right\|$ or $-\left\|T\big|_{\mathcal{K}_m^\perp}\right\|$ is an eigenvalue of $T\big|_{\mathcal{K}_m^\perp}$. Choose u_{m+1} to be an eigenvector of $T\big|_{\mathcal{K}_m^\perp}$ with eigenvalue $\lambda_{m+1} = \pm\left\|T\big|_{\mathcal{K}_m^\perp}\right\|$. Then $Tu_{m+1} =$

$\lambda_{m+1}u_{m+1}$ so that u_{m+1} is also an eigenvector of T. Evidently,

$$\left\|T|_{\mathcal{K}_m^\perp}\right\| \leq \left\|T|_{\mathcal{K}_{m-1}^\perp}\right\|$$

and hence $|\lambda_{m+1}| \leq |\lambda_m|$. Thus, $\{u_1, \dots, u_{m+1}\}$ is an orthonormal set consisting of eigenvectors of T with $Tu_j = \lambda u_j$ and $|\lambda_j| \leq |\lambda_k|$ for all $1 \leq j < k \leq m+1$.

If $\dim(\mathcal{H}) := N < \infty$, the inductive construction terminates in N steps producing an orthonormal basis for $\mathcal{H}$. Otherwise it continues indefinitely producing an infinite orthonormal sequence $\{u_n\}_{n\in\mathbb{N}}$ with $Tu_n = \lambda_n u_n$ and $n \mapsto |\lambda_n|$ non-increasing on $\mathbb{N}$. By Theorem 4.0.6, for each $r > 0$, there can be only finitely many n such that $|\lambda_n| \geq r$. It follows that $\lim_{n\to\infty} \lambda_n = 0$.

We now claim that $\{u_n\}_{n\in\mathbb{N}}$ is complete. To see this, let $\mathcal{K}$ denote the orthogonal complement of the span of $\{u_n\}_{n\in\mathbb{N}}$, and note that since the span of $\{u_n\}_{n\in\mathbb{N}}$ is invariant under T, so is $\mathcal{K}$. Since $\ker(T) = \{0\}$, if $\mathcal{K} \neq \{0\}$, $\|T|_{\mathcal{K}}\| > 0$. We could then apply Theorem 4.0.7 to produce an unit vector u that is an eigenvector of T with $Tu = \pm\|T|_{\mathcal{K}}\|u$, and with $\langle u, u_n\rangle = 0$ for all n. But this is impossible since by the construction of $\{u_n\}_{n\in\mathbb{N}}$, for all unit vectors $u \in \mathcal{H}$ with $\langle u, u_j\rangle = 0$ for $j = 1, \dots, m$, $|\langle u, Tu\rangle| \leq |\lambda_m|$ and we have seen that $\lim_{n\to\infty} |\lambda_n| = 0$. Hence $\mathcal{K} = \{0\}$, which means that $\{u_n\}_{n\in\mathbb{N}}$ is complete.

Since $\{u_j\}_{j\in\mathscr{J}}$ is an orthonormal basis for $\mathcal{H}$, we have that for all $f \in \mathcal{H}$, $f = \sum_{j\in\mathscr{J}} \langle u_j, f\rangle u_j$, and hence

$$Tf = T\left(\sum_{j\in\mathscr{J}} \langle u_j, f\rangle u_j\right) = \sum_{j\in\mathscr{J}} \langle u_j, f\rangle Tu_j = \sum_{j\in\mathscr{J}} \lambda_j |u_j\rangle\langle u_j| f \ .$$

If $\mathscr{J}$ is finite, this proves (4.0.9).

For the case $\mathscr{J} = \mathbb{N}$, we now show that the sum in (4.0.9) converges in the operator norm. To do this, for each $n \in \mathbb{N}$, define $T_n = \displaystyle\sum_{j=1}^{n} \lambda_j |u_j\rangle\langle u_j|$. Then all $n > m$ and unit vectors u,

$$|\langle u, (T_n - T_m)u\rangle| = \left|\sum_{j=m+1}^{n} \lambda_j |\langle u_j, u\rangle|^2\right| \leq |\lambda_{m+1}| \sum_{j=1}^{\infty} |\langle u_j, u\rangle|^2 = |\lambda_{m+1}| \ .$$

By Lemma 4.0.1, $\|T_n - T_m\| \leq |\lambda_{m+1}|$ and then since $\lim_{n\to\infty} \lambda_n = 0$, $\{T_n\}_{n\in\mathbb{N}}$ is a Cauchy sequence in $\mathscr{B}(\mathcal{H})$. This sequence therefore converges in operator norm to a limit that must agree with T on the dense span of $\{u_j\}_{j\in N}$, and must therefore be T. This proves (4.0.9) in the infinite dimensional case. □

There are several corollaries:

4.0.11 COROLLARY. *For all self-adjoint $T \in \mathscr{C}(\mathcal{H})$, there exists an orthonormal basis for $\mathcal{H}$ consisting of eigenvectors of $\mathcal{H}$.*

Proof. Let $\mathcal{K} := \ker(T)$, which is a closed subspace of $\mathcal{H}$, and therefore a Hilbert space in its own right. Combine any orthonormal basis for this space with the orthonormal basis of $\mathcal{K}^\perp$ that is provided by Theorem 4.0.10. □

4.0.12 COROLLARY. *$\mathscr{C}(\mathcal{H})$ is the operator norm closure of the set $\mathscr{F}(\mathcal{H})$ of finite rank operators on $\mathcal{H}$.*

Proof. We have seen that $\mathscr{F}(\mathcal{H}) \subset \mathscr{C}(\mathcal{H})$, and that $\mathscr{C}(\mathcal{H})$ is closed so that $\overline{\mathscr{F}(\mathcal{H})} \subset \mathscr{C}(\mathcal{H})$. The formula (4.0.9) displays every self adjoint compact operator T as the operator norm limit of finite rank operators, and if T is compact, so are $R := T + T^*$ and $S := i(T^* - T)$. As self-adjoint compact operators, R and S can be approximated in operator norm by finite rank operators, but then since $T = R + iS$, so can T. □

If $T \in \mathscr{C}(H)$, and $T = T^*$, and φ is a continuous function on defined on $[-\|T\|, \|T\|]$, then we define

$$\varphi(T) = \sum_{j\in\mathbb{N}} \varphi(\lambda_j)|u_j\rangle\langle u_j| \ .$$

4.0.4 The Fredholm Alternative

The Fundamental Theorem of Linear Algebra says that a linear transformation T between finite dimensional vector spaces V and W is invertible if and only if V and W have the same dimension *and* T is either injective or surjective. In other words, for linear maps between vector spaces of the same finite dimension, infectivity implies subjectivity and *vice-versa*.

In infinite dimensions, this is not true in general, but there is an important case in which is is true.

4.0.13 THEOREM. *Let T be a compact operator on a Hilbert space $\mathcal{H}$. Then $(I - T)$ is invertible if and only if $(I - T)$ is injective, and $(I - T)$ is invertible if and only if $(I - T)$ is surjective. Moreover*

$$\dim(\ker(I - T)) = \dim((\operatorname{ran}(I - T))^{\perp}) \ . \tag{4.0.10}$$

4.0.14 Remark. The first part of the theorem can be expressed as saying that either $(I - T)$ is invertible, or else 1 is an eigenvalue of T. This is the *Fredholm alternative.* For $\lambda \neq 0$, $(\lambda I - T) = \lambda(I - \lambda^{-1}T)$, and hence $(\lambda I - T)$ is invertible if and only if $(I - \lambda^{-1}T)$ is invertible, and $\lambda^{-1}T$ is compact if and only if T is compact. Hence if T is compact, and $\lambda \neq 0$, either λ is an eigenvalue of T or else $\lambda I - T$ is invertible.

4.0.15 LEMMA. *Let T be a compact operator on a Hilbert space $\mathcal{H}$. Then either there exists $C > 0$ such that for all $f \in \mathcal{H}$,*

$$\|(I - T)f\| \geq C\|f\| \tag{4.0.11}$$

or else $\ker(I - T) \neq \{0\}$.

Proof. Suppose that there is no $C > 0$ such that (4.0.11) is valid for all $f \in \mathcal{H}$. The there exists a sequence of unit vectors $\{u_n\}_{n\in\mathbb{N}}$ such that $\lim_{n\to\infty} \|(I - T)u_n\| = 0$. Since the closed unit ball B in $\mathcal{H}$ is weakly sequentially compact, by passing to a subsequence, we may assume that $\{u_n\}_{n\in\mathbb{N}}$ converges weakly to some $u \in B$, and then, since T is compact, that $\lim_{n\to\infty} \|Tu_n - Tu\| = 0$. Then since

$$|1 - \|Tu_n\|| = |\|u_n\| - \|Tu_n\|| \leq \|u_n - Tu_n\| = \|(I - T)u_n\| \ ,$$

$\lim_{n\to\infty} \|Tu_n\| = 1$, and hence $\|Tu\| = 1$.

Since T commutes with $(I - T)$, and is bounded, the hypothesis that $\lim_{n\to\infty} \|(I - T)u_n\| = 0$ implies that

$$0 = \lim_{n\to\infty} \|T(I - T)u_n\| = \lim_{n\to\infty} \|(I - T)Tu_n\| = \|(I - T)Tu\| \ .$$

Since $\|Tu\| = 1$, this means that Tu is a non-zero vector in $\ker(I - T)$. □

4.0.16 LEMMA. *Let T be a compact operator on a Hilbert space $\mathcal{H}$. If* $\ker(I-T)=\{0\}$*, then* $\mathrm{ran}(I-T)$ *is closed.*

Proof. Let $\{g_n\}_{n\in\mathbb{N}}$ be a sequence in $\mathrm{ran}(I-T)$ that converges in norm to some $g\in\mathcal{H}$. We must show that $g\in\mathrm{ran}(I-T)$. Since $(I-T)$ is injective, for each $n\in\mathbb{N}$, there is a unique $f_n\in\mathcal{H}$ such that $(I-T)f_n=g_n$. Then by (4.0.11), $\{f_n\}_{n\in\mathbb{N}}$ is a Cauchy sequence in $\mathcal{H}$, and hence there exists $f\in\mathcal{H}$ such that $\lim_{n\to\infty}\|f_n-f\|=0$. Then

$$g=\lim_{n\to\infty}(I-T)f_n=(I-T)f\ ,$$

showing that $g\in\mathrm{ran}(I-T)$. $\square$

4.0.17 LEMMA. *Let $T\in\mathscr{B}(\mathcal{H})$ Then*

$$(\mathrm{ran}(T))^\perp=\ker(T^*)\ . \tag{4.0.12}$$

Proof. Let $f\in\ker(T^*)$. For all $g\in\mathcal{H}$, $0=\langle T^*f,g\rangle=\langle f,Tg\rangle$, and hence $f\perp Tg$. Thus $\ker(T^*)\subset(\mathrm{ran}(T))^\perp$. Let $f\in(\mathrm{ran}(T))^\perp$. For all $g\in\mathcal{H}$, $0=\langle f,Tg\rangle=\langle T^*f,g\rangle$, and hence $T^*f=0$. Thus, $(\mathrm{ran}(T))^\perp\subset\ker(T^*)$. $\square$

Proof of Theorem 4.0.13. Suppose that $I-T$ is injective. By Lemma 4.0.16, $V_1:=\mathrm{ran}(I-T)$ is a closed subspace of $\mathcal{H}$, and hence a Hilbert space. Since T is a continuous vector space isomorphism of H onto V_1, it has a bounded inverse: By Lemma 4.0.15, the unique f such the $(I-T)f=g$; i.e., $(I-T)^{-1}g$ satisfies $\|(I-T)^{-1}g\|\le C^{-1}\|g\|$, showing that $(I-T)^{-1}\in\mathscr{B}(\mathcal{H})$. (One could also invoke the Open Mapping Theorem.)

Suppose that V_1 is a proper subspace of $\mathcal{H}$. Then since $I-T$ is injective, $V_2:=(I-T)V_1$ is a proper subspace of $V_1=(I-T)\mathcal{H}$, and it is closed since $(I-T)$ is a topological homeomorphism. We inductively define $V_{n+1}:=(I-T)V_n$ for each $n\in\mathbb{N}$, and then, as above, we have that each V_{n+1} is a proper, closed subspace of V_n.

Then $V_n\cap V_{n+1}^\perp$ is a non-zero subspace for all $n\in\mathbb{N}$. Choose any unit vector $u_n\in V_n\cap V_{n+1}^\perp$. Since $V_n\subset V_m$ for all $n\ge m$, $\{u_n\}_{n\in\mathbb{N}}$ is an orthonormal sequence. Since $u_n\in V_n$ and $Tu_n-u_n\in V_{n+1}$, u_n and Tu_n-u_n are orthogonal, and hence

$$\|Tu_n\|^2=\|(Tu_n-u_n)+u_n\|^2=\|Tu_n-u_n\|^2+\|u_n\|^2\ge 1\ .$$

However, since $\{u_n\}_{n\in\mathbb{N}}$ is orthonormal, it converges weakly to 0, and then since T is compact $\lim_{n\to\infty}\|Tu_n\|=0$. This contradiction shows that $\mathrm{ran}(I-T)=V_1=\mathcal{H}$, and hence that $I-T$ is surjective onto $\mathcal{H}$ as well as injective, and we have already seen that $(I-T)^{-1}\in\mathscr{B}(\mathcal{H})$. This proves that $I-T$ is invertible if and only if it is injective.

Next, suppose that $I-T$ is surjective. Then by Lemma 4.0.17, $I-T^*=(I-T)^*$ is injective, and since T^* is also compact, what we have proved above shows that $I-T^*$ is invertible in $\mathscr{B}(\mathcal{H})$. But then $I-T$ is invertible in $\mathscr{B}(\mathcal{H})$, with inverse $((I-T^*)^{-1})^*$.

Finally, suppose that $I-T$ is not invertible. Then $\ker(I-T)$ and $(\mathrm{ran}(I-T))^\perp=\ker(I-T^*)$ are eigenspace of the compact operators T and T^*, and hence are finite dimensional. Let $\{u_1,\dots,u_m\}$ be any orthonormal basis of $\ker(I-T)$, and let $\{v_1,\dots,v_n\}$ be any orthonormal basis of $(\mathrm{ran}(I-T))^\perp$. Let

$p := \min\{m, n\}$, and define $F = \sum_{j=1}^{p} |v_j\rangle\langle u_j|$. Note that $(I - T + F)$ is injective if and only if if $p = m$, and $(I - T + F)$ is surjective if and only if $p = n$. Then since $T - F$ is compact, by what we have proved above, $p = m = n$, and this proves (4.0.10). □

4.0.18 Remark. For $S \in \mathscr{B}(H)$, the *nullity of* S is defined by $\text{nullity}(S) = \dim(\ker(S))$, and the *rank of* S is defined by $\text{rank}(S) = \dim(\text{ran}(S))$. When $\mathcal{H} = \mathbb{C}^n$, so that we may identity $\mathscr{B}(\mathcal{H})$ with the $n \times n$ matrices, we have the simple identity that for any subspace $\mathcal{K}$ of $\mathcal{H}$, $\dim(\mathcal{K}) + \dim(\mathcal{K}^\perp) = n$, and hence $\text{rank}(S) = n - \dim((\text{ran}(S))^\perp)$. Defining $T = I - S$, so that $S = I - T$, and noting that in finite dimensions every linear operator is compact, the identity (4.0.10) of Theorem 4.0.13 says that $\text{nullity}(S) + \text{rank}(S) = n$.

By Lemma 4.0.17, and what we have said above, and equivalent formulation is that

$$\text{nullity}(S) = \text{nullity}(S^*) \ . \tag{4.0.13}$$

This formulation has the advantage of not referring explicitly to the dimension, and as Theorem 4.0.13 shows, it remains true in infinite dimensions when $S = I - T$ with T compact. For $\lambda \neq 0$, write $S = \lambda^{-1}(\lambda I - T)$. Then by (4.0.13), $\text{nullity}(\lambda I - T) = \text{nullity}(\lambda^* I - T^*)$, and hence if T is compact operator, then λ is a eigenvalue of T if and only if λ^* is an eigenvalues of T^* and in that case, these eigenvalues have the same (finite) geometric multiplicity, just as in finite dimensions.

Chapter 5

Convexity

5.1 Convex functions on $\mathbb{R}$

5.1.1 Continuity and lower-semicontinuity of convex functions

5.1.1 DEFINITION (Convex Function). A function ϕ on defined on a real vector space X with values in $(-\infty, \infty]$ is *convex* in case for all $x, y \in C$ and all $\lambda \in (0,1)$,

$$\phi(\lambda x + (1-\lambda)y) \leq \lambda\phi(x) + (1-\lambda)\phi(y) \ , \tag{5.1.1}$$

and ϕ is *strictly convex* in case this inequality is strict for all $x \neq y$ and all $\lambda \in (0,1)$. A convex function ϕ on X is *proper* in case $\phi(x) < \infty$ for at least one $x \in X$.

5.1.2 DEFINITION (Epigraph). Let ϕ be a function from some set X to $(-\infty, \infty]$. The *epigraph of* ϕ, $\mathrm{Epi}(\phi)$, is the subset of $X \times \mathbb{R}$ consisting of points (x,t) such that $\phi(x) \geq t$. Note that $\mathrm{Epi}(\phi) = \emptyset$ if and only if $\phi(x) = \infty$ for all x, and that ϕ is convex if and only if $\mathrm{Epi}(\phi)$ is a convex subset of $X \times \mathbb{R}$.

5.1.3 LEMMA. *Let $(X, \mathscr{O})$ be a topological vector space. Then a function ϕ on X with values in $(-\infty, \infty]$ is convex if and only if* $\mathrm{Epi}(\phi)$ *is convex, and is lower-semicontinuous if and only if* $\mathrm{Epi}(\phi)$ *is closed in the product topology.*

Proof. This is elementary and is left to the reader. □

Let $(X, \mathscr{O})$ be a topological vector space, and let ϕ be a convex function on X. Then $\mathrm{Epi}(\phi)$ is convex, and since the closure of a convex set is convex, $\overline{\mathrm{Epi}(\phi)}$ is closed and convex. Define a convex function $\overline{\phi}$ by

$$\overline{\phi}(x) = \begin{cases} \infty & \{x\} \times \mathbb{R} \cap \overline{\mathrm{Epi}(\phi)} = \emptyset \\ \inf\{t \ : \ (x,t) \in x \in \overline{\mathrm{Epi}(\phi)}\} & \text{otherwise} \ . \end{cases}$$

Then $\overline{\phi}$ is a lower-semicontinuous convex function such that $\overline{\phi}(x) \leq \phi(x)$ for all x. By construction and Lemma 5.1.3, $\overline{\phi}$ is the largest (in the usual ordering) lower-semicontinuous convex function that is dominated by ϕ pointwise. The function $\overline{\phi}$ is *lower-semicontinuous regularization of* ϕ.

5.1.4 EXAMPLE. *Define* $\phi : \mathbb{R} \to (-\infty, \infty]$ *by*

$$\phi(x) = \begin{cases} 0 & x \in (-1,1) \\ 1 & x \in \{-1,1\} \\ \infty & x \notin [-1,1] \ . \end{cases} \qquad \textit{Then} \qquad \overline{\phi}(x) = \begin{cases} 0 & x \in [-1,1] \\ \infty & x \notin [-1,1] \ . \end{cases}$$

5.1.2 Convex functions on $\mathbb{R}$

The special case in which the vector space X is simply $\mathbb{R}$ is especially important. In this case the set C on which ϕ is finite is an interval.

5.1.5 THEOREM. *Let ϕ be a real valued convex function on an interval $C \subset \mathbb{R}$. Then ϕ is continuous on the interior of C. Moreover, for $a, b, c, d \in C$*

$$b - a = d - c \quad \text{and} \quad a < c \quad \Rightarrow \quad \phi(b) - \phi(a) \le \phi(d) - \phi(c) \ . \tag{5.1.2}$$

That is, the increment of ϕ over an interval increases as the interval is translated to the right. If ϕ is strictly convex, the inequality in (5.1.2) is strict. Conversely, let ϕ be any function that is continuous and finite on an open interval C, and is such that (5.1.2) is valid for all $a, b, c, d \in C$. Then ϕ is convex.

Proof. We first prove (5.1.2) Note that $b, c \in [a, d]$, and hence for some $\lambda, \beta \in (0,1)$, $b = \lambda a + (1-\lambda)d$ and $c = \beta a + (1-\beta)d$. Solving for λ and β, we find

$$\lambda = 1 - \beta = \frac{d-b}{d-a} \ .$$

Since $\lambda = 1-\beta$, $b = \lambda a + (1-\lambda)d$ and $c = \lambda d + (1-\lambda)a$. It now follows from the definition of convexity that

$$\phi(b) \le \lambda\phi(a) + (1-\lambda)\phi(d) \qquad \text{and} \qquad \phi(c) = \lambda\phi(d) + (1-\lambda)\phi(a) \ .$$

Adding these two inequalities yields $\phi(b) + \phi(c) \le \phi(a) + \phi(d)$, with strict inequality if ϕ is strictly convex, and this proves the first assertion, and evidently if ϕ is strictly convex, all of the inequalities are strict.

Fix any a in the interior of C; we shall show that ϕ is continuous at a. For any $b \in C$, $b \neq a$, use a telescoping sum expansion to write

$$\phi(b) - \phi(a) = \sum_{j=1}^{n} \left[\phi\left(a + (b-a)\frac{j}{n} \right) - \phi\left(a + (b-a)\frac{j-1}{n} \right) \right] \ . \tag{5.1.3}$$

Choose $\delta > 0$ so that $a \pm \delta \in C$. By (5.1.2), for $b = a + \delta$, the first term in the sum is the least, and hence $\phi(a + \delta/n) - \phi(a) \le \dfrac{\phi(a+\delta) - \phi(a)}{n}$ Likewise, for $b = a - \delta$, the same reasoning yields $\dfrac{\phi(a) - \phi(a-\delta)}{n} \le \phi(a) - \phi(a - \delta/n)$. Again by (5.1.2), $\phi(a) - \phi(a-\delta) \le \phi(a+\delta/n) - \phi(a)$. Altogether,

$$\frac{\phi(a) - \phi(a-\delta)}{n} \le \phi(a) - \phi(a - \delta/n) \le \phi(a + \delta/n) - \phi(a) \le \frac{\phi(a+\delta) - \phi(a)}{n} \ . \tag{5.1.4}$$

Taking $n \to \infty$, we conclude $\lim_{n\to\infty} \phi(a - \delta/n) = \phi(a) = \lim_{n\to\infty} \phi(a + \delta/n)$.

We now claim that $\lim_{x \downarrow a} \phi(x) = \phi(a)$. Suppose that $\limsup_{x\downarrow a} \phi(x) > \phi(a)$. Then for some $\epsilon > 0$, there is an infinite sequence $\{t_m\}_{m\in\mathbb{N}}$ contained in $(a, a+\delta)$ such that $\lim_{m\to\infty} t_m = a$ and $\phi(t_m) \ge \phi(a) + \epsilon$

for all $m \in \mathbb{N}$. Choose n so that $\phi(a+\delta/n) \leq \phi(a)+\epsilon/2$. Then there exists $m \in \mathbb{N}$ such that $t_m \in (a, a+\delta/n)$ and thus $\lambda \in (0,1)$ such that $t_m = \lambda a + (1-\lambda)(a+\delta/n)$. Then $\phi(t_m) \leq \lambda\phi(a) + (1-\lambda)(\phi(a)+\epsilon/2) < \phi(a)+\epsilon/2$. This contradiction shows that $\limsup_{x\downarrow a}\phi(x) \leq \phi(a)$.

Next, suppose that $\liminf_{x\downarrow a}\phi(x) < \phi(a)$. Then for some $\epsilon > 0$, there is an infinite sequence $\{t_m\}_{m\in\mathbb{N}}$ contained in $(a, a+\delta)$ such that $\lim_{m\to\infty} t_m = a$ and $\phi(t_m) \leq \phi(a) - \epsilon$ for all $m \in \mathbb{N}$. Choose n so that $\phi(a+\delta/n) \geq \phi(a)-\epsilon/2$. Choose $n \in \mathbb{N}$ so that $\phi(a+\delta/n) \geq \phi(a)-\epsilon/2$, and $a+\delta/n < t_1$. Choose m so that $t_m < a+\delta/n$. Then $a+\delta/n \in (t_m, t_1)$ and there exists $\lambda \in (0,1)$ such that $a+\delta/n = \lambda t_m + (1-\lambda)t_1$. Then $\phi(a+\delta/n) \leq \lambda\phi(t_m) + (1-\lambda)\phi(t_m) \ < \phi(a) - \epsilon/2$. This contradiction shows that $\liminf_{x\downarrow a}\phi(x) \geq \phi(a)$.

Altogether, we have shown that ϕ is right continuous at a. We could repeat the same analysis to show that ϕ is also left continuous at a, but observe that the function $\psi(x) := \phi(2a-x)$ is convex and finite on an open interval about a. (Epi(ψ) is just the reflection of Epi(ϕ) about the vertical line $x = a$). By what we have just proved, ψ is right continuous at a. But then since ϕ is the reflection of ϕ about $x = a$, ϕ is left continuous at $x = a$. Altogether, the continuity of ϕ is proved.

For the converse, let $x < y \in C$ and $\lambda \in (0,1)$. We must show that when (5.1.2) is valid for all $a, b, c, d \in C$, then $\phi(\lambda x + (1-\lambda)y) \leq \lambda\phi(x) + (1-\lambda)\phi(y)$. By the continuity of ϕ, it suffices to do this when λ is a dyadic rational; i.e., $\lambda = k/2^n$ for some $k, n \in \mathbb{N}$ with $k < 2^n$. For $n = 1$, define $z = (x+y)/2$ and note that

$$\frac{1}{2}\phi(x) + \frac{1}{2}\phi(y) - \phi\left(\frac{x+y}{2}\right) = \frac{1}{2}\left(\phi(y) - \phi(z)\right) - \frac{1}{2}\left(\phi(z) - \phi(x)\right) \geq 0$$

because $y - z = z - x$. That is, when (5.1.2) is valid for all $a, b, c, d \in C$, then

$$\phi(\lambda x + (1-\lambda)y) \leq \lambda\phi(x) + (1-\lambda)\phi(y) \tag{5.1.5}$$

for all $x, y \in C$ and $\lambda = 1/2$.

This is the first step of an inductive proof that the same is true whenever $\lambda = j2^{-m}$, $j, m \in \mathbb{N}$ and $j < 2^m$. We suppose that this has been shown whenever $m < n$.

Now fix $\lambda = j/2^{-n} \in (0,1)$. Let $k, \ell \in \mathbb{N}$ such that $k + \ell = j$ and $k, \ell \leq 2^{n-1}$. (If j is even take $k = \ell = j/2$, and if j is odd, take k to be the integer part of $j/2$.) Then for all x, y,

$$\lambda x + (1-\lambda)y = \frac{jx + (2^n - j)y}{2^n} = \frac{1}{2}\left(\frac{kx + (2^{n-1}-k)y}{2^{n-1}}\right) + \frac{1}{2}\left(\frac{\ell x + (2^{n-1}-\ell)y}{2^{n-1}}\right) \tag{5.1.6}$$

Since (5.1.5) is true for $\lambda = 1/2$,

$$\phi(\lambda x + (1-\lambda)y) \leq \frac{1}{2}\phi\left(\frac{kx + (2^{n-1}-k)y}{2^{n-1}}\right) + \frac{1}{2}\phi\left(\frac{\ell x + (2^{n-1}-\ell)y}{2^{n-1}}\right) .$$

By the inductive hypothesis,

$$\phi\left(\frac{kx + (2^{n-1}-k)y}{2^{n-1}}\right) \leq \frac{k}{2^{n-1}}\phi(x) + \frac{2^{n-1}-k}{2^{n-1}}\phi(y)$$

and likewise with ℓ in place of k. Using these inequalities in (5.1.6) shows that (5.1.5) is valid for $\lambda = j2^{-n}$, completing the inductive proof. $\square$

5.1.3 The subgradient of a convex function

5.1.6 THEOREM. *Let ϕ be a real valued convex function on an interval $C \subset \mathbb{R}$. For all $a, b, c \in C$, $a < b < c$,*

$$\frac{\phi(c) - \phi(a)}{c - a} \geq \frac{\phi(b) - \phi(a)}{b - a} \ , \tag{5.1.7}$$

If ϕ is strictly convex, the inequality in (5.1.7) is strict.

Proof. Because ϕ is continuous, it suffices to consider rational values of $b - a$ and $c - a$. Choosing a common denominator n, we can write $b - a = \dfrac{k}{n}$ and $c - a = \dfrac{m}{n}$ with $m > k$. Define a sequence $\{a_j\}$ by

$$a_j = \phi\left(a + \frac{j}{n}\right) - \phi\left(a + \frac{j-1}{n}\right) \ .$$

By (5.1.2), this is an increasing sequence, and hence,

$$\frac{\phi(b) - \phi(a)}{b - a} = n\frac{\phi(a + k/n) - \phi(a)}{k} = n\left(\frac{1}{k}\sum_{j=1}^{k} a_j\right) \ .$$

Likewise, $\dfrac{\phi(c) - \phi(a)}{c - a} = n\left(\dfrac{1}{m}\displaystyle\sum_{j=1}^{m} a_j\right)$. Since a_j increases with j, $\dfrac{1}{k}\displaystyle\sum_{j=1}^{k} a_j \leq \dfrac{1}{m}\displaystyle\sum_{j=1}^{m} a_j$ for $m \geq k$. This proves (5.1.7). By theorem 5.1.5, if ϕ is strictly convex, then $a_{j+1} > a_j$ for all j, and then there is strict inequality in (5.1.7). $\square$

Let ϕ be convex on $\mathbb{R}$ and finite on C, For each $s \in C^\circ$, define

$$\sigma_+^\phi(s) = \lim_{h \to 0^+} \frac{\phi(s + h) - \phi(s)}{h} \quad \text{and} \quad \sigma_-^\phi(s) = \lim_{h \to 0^+} \frac{\phi(s) - \phi(s - h)}{h} \ . \tag{5.1.8}$$

The limit defining σ_+^ϕ exists by (5.1.7), and then limit defining σ_-^ϕ also exists by (5.1.7), but applied to the convex functions $\phi(-s)$. We refer to $\sigma_+^\phi(s)$ as the *right derivative of ϕ at s*, and to $\sigma_-^\phi(s)$ as the *left derivative of ϕ at s*. If is clear from Theorem 5.1.5 that in general, $\sigma_+^\phi(s) \geq \sigma_-^\phi(s)$. Evidently ϕ is differentiable at s if and only if $\sigma_+^\phi(s) = \sigma_-^\phi(s)$, so that in this case, $\phi'(s) = \sigma_+^\phi(s) = \sigma_-^\phi(s)$ represents the slope of ϕ at s. Since the left derivative of ϕ at s is minus the right derivative of $t \mapsto \phi(-t)$ at $t = -s$, we may economize in the formulation of the following theorem by referring only to right derivatives.

5.1.7 THEOREM (One-sided Derivatives)**.** *Let ϕ be a convex function on $\mathbb{R}$ that is finite on an interval C. for all $a, b \in C^\circ$, $a < b$.*

$$\sigma_+^\phi(a) \leq \sigma_-^\phi(b) \tag{5.1.9}$$

and for all $s \in [\sigma_-^\phi(a), \sigma_+^\phi(a)]$,

$$\phi(b) \geq \phi(a) + s(b - a) \ . \tag{5.1.10}$$

Moreover, if ϕ is strictly convex, then both of these inequalities are strict. Finally, for all $a \in C^\circ$,

$$\sigma_+^\phi(a) = \inf_{b > a} \sigma_+^\phi(b) \quad \text{and} \quad \sigma_-^\phi(a) = \sup_{b < a} \sigma_-^\phi(b) \ . \tag{5.1.11}$$

In other words, σ_+^ϕ is a right-continuous *non-decreasing function, and σ_-^ϕ is a* left-continuous *non-decreasing function.*

Proof. By Theorem 5.1.5, for all $0 < h < b-a$, $\phi(a+h) - \phi(a) \leq \phi(b) - \phi(b-h)$. Dividing by h and taking the limit $h \downarrow 0$ yields (5.1.9).

Next, by the telescoping sum identity (5.1.3) and Theorem 5.1.5 to obtain, for $b > a$,

$$\phi(b) - \phi(a) \geq n(\phi(a + (b-a)/n) - \phi(a)) \ . \tag{5.1.12}$$

Multiplying by $1 = (b-a)/(b-a)$, yields

$$\phi(b) \geq \phi(a) + (b-a)\left[\frac{\phi(a + (b-a)/n) - \phi(a)}{(b-a)/n}\right]$$

Taking the limit $n \to \infty$ yields $\phi(b) \geq \phi(a) + \sigma_+^\phi(a)(b-a)$.

For $b < a$, (5.1.3) and Theorem 5.1.5 yield

$$\phi(a) - \phi(b) \leq n(\phi(a) - \phi(a - (a-b)/n)) \ . \tag{5.1.13}$$

Multiplying and dividing by $-1 = (b-a)/(a-b)$, yileds

$$\phi(b) \geq \phi(a) + (b-a)\left[\frac{\phi(a) - \phi(a - (a-b)/n)}{(a-b)/n}\right]$$

Taking the limit $n \to \infty$ yields $\phi(b) \geq \phi(a) + \sigma_-^\phi(a)(b-a)$. Therefore, for any $s \in [\sigma_-^\phi(a), \sigma_+^\phi(a)]$, (5.1.10) is valid.

Next, for $a \in C^\circ$, choose $\epsilon > 0$, and then $h > 0$ so that $\sigma_+^\phi(a) + \epsilon \geq (\phi(a+h) - \phi(a))/h$. By the continuity of ϕ, there is a $\delta > 0$ so that $(\phi(a+h) - \phi(a))/h + \epsilon \geq (\phi(a+\delta+h) - \phi(a+\delta))/h$. Therefore,

$$\sigma_+^\phi(a) + 2\epsilon \geq (\phi(a+\delta+h) - \phi(a+\delta))/h \geq \sigma_+^\phi(a+\delta)$$

The fact that for all $\delta > 0$, $\sigma_+^\phi(a) \leq \sigma_+^\phi(a_\delta)$ is an immediate consequence of (5.1.9). This proves the first identity in (5.1.11). The second is proved in the same manner. □

Theorem 5.1.7 has the following interpretation: Let ϕ be convex and finite on an interval $C \subset \mathbb{R}$ with non-empty interior C^o. Then for all $x_0 \in C^o$, and all $s \in [\sigma_-^\phi(a), \sigma_+^\phi(a)]$, the affine function $h(x) = s(x - x_0) + \phi(x_0)$ satisfies $h(x) \leq \phi(x)$ for all x, with $h(x_0) = \phi(x_0)$: The graph of h, a line, lies below the graph of ϕ but touches it at the point $(x_0, \phi(x_0))$. Such a line is called a *supporting line* for the graph of ϕ.

The inequality (5.1.10) is called the "above the tangent line inequality" because it expresses the fact that the graph of ϕ lies everywere above its tangent line at each pontx_0 where ϕ is differentiaible – and more generally lies above each of its supporing lines.

5.1.8 DEFINITION (Subgradient)**.** Let ϕ be a convex function on $\mathbb{R}$ that is finite on an interval C. For $x \in C$, the *subgradient of* ϕ *at* x, $\partial\phi(x)$, is the set of numbers s such that

$$\phi(y) \geq \phi(x) + s(y-x)$$

for all $y \in \mathbb{R}$. In other words, $s \in \partial\phi(x)$ if and only if the line with slope s that passes through $(x, \phi(x))$ is a supporting line for ϕ. If $\phi(x) = \infty$, we define $\partial\phi(x) = \emptyset$. For $A \subset \mathbb{R}$, define

$$\partial\phi(A) = \bigcup_{x \in A} \partial\phi(x) \ . \tag{5.1.14}$$

5.1.9 LEMMA. *Let ϕ be a convex function on $\mathbb{R}$ that is finite on in a interval C. Then for all $x \in C^o$, $\partial\phi(x) = [\sigma_-^\phi(x), \sigma_+^\phi(x)]$. In particular, for $x \in C^o$, $\partial\phi(x) \neq \emptyset$.*

Proof. This follows immediately from (5.1.10). □

5.1.10 EXAMPLE. *Let $\phi(x) = \begin{cases} a & x = x_0 \\ \infty & x \neq x_0 \end{cases}$. Then ϕ is lower-semicontinuous, and for all $s, x \in \mathbb{R}$, $\phi(x) \leq \phi(x_0) + s(x - x_0)$ so that $\partial\phi(x_0) = \mathbb{R}$ and for $x \neq x_0$, $\partial\phi(x) = \emptyset$. In this case, we have $\partial\phi(\mathbb{R}) = \mathbb{R}$, but whole contribution to the union comes from the single point x_0.*

For a complementary example, consider $\psi(x) = a(x - x_0)$ for given $a, x_0 \in \mathbb{R}$. This is a continuous convex function. Since ψ is differentialbe at each point x with $\psi'(x) = a$, $\partial\psi(x) = \{a\}$ for all $x \in \mathbb{R}$, and hence $\partial\psi(\mathbb{R}) = \{a\}$.

We close this secton with a simple but imprtant application of the “above the tangent line inequality” (5.1.10):

5.1.11 THEOREM (Jensen's Inequality)**.** *Let $(\Omega, \mathcal{M}, \mu)$ be a measure space with $\mu(\Omega) = 1$. Then for all real valued convex functions ϕ on $\mathbb{R}$, and all measurable function f, the negative part of $\phi(|f|)$ is integrable, so that $\int_\Omega \phi(|f|)\mathrm{d}\mu$ is well defined, though it may be infinite. Moreover,*

$$\phi\left(\int_\Omega |f|\mathrm{d}\mu\right) \leq \int_\Omega \phi(|f|)\mathrm{d}\mu \ . \tag{5.1.15}$$

and if ϕ is strictly convex there is equality if and only if

$$|f(x)| = \int_\Omega |f|\mathrm{d}\mu \tag{5.1.16}$$

for almost every x.

Proof. Let $a := \int_\Omega |f|\mathrm{d}\mu$. By (5.1.10), $\phi(|f(x)|) \geq \phi(a) + \sigma_+^\phi(a)(|f(x)| - a)$. Integrating this pointwise inequality yields (5.1.15). If ϕ is strictly convex, this pointwise inequality is strict wherever $|f(x)| \neq a$, and hence the inequality (5.1.15) is strict unless $|f(x)| = a$ almost everywhere. □

5.1.4 The Legendre Transform

Let ϕ be a proper convex function. Let C be the set on which ϕ is finite, which is not empty since ϕ is proper. Since ϕ is convex, C is an interval. If C has empty interior, then $C = [x_0, x_0]$ for some x_0, and then ϕ is of the form given in Example 5.1.10; $\phi(x_0) = a$ for some $a, x_0 \in R$, and $\phi(x) = \infty$ otherwise. In this case we have seen that $\partial\phi(x_0) = \mathbb{R}$. Otherwise, the set C has a non-empty interior C^o, and then by Lemma 5.1.9, for each $x \in C^o$, $\partial\phi(x) \neq \emptyset$.

5.1.12 DEFINITION (Legendre Transform)**.** Let ϕ be a proper convex function on $\mathbb{R}$. The *Legendre transform* ϕ^* of ϕ is the function defined by

$$\phi^*(y) = \sup_{x \in \mathbb{R}} \{yx - \phi(x)\} \ . \tag{5.1.17}$$

which takes values in $(-\infty, \infty]$.

Notice that ϕ^* is lower-semicontinuous since it is the pointwise supremem of a set of lower-semicontinuous functions (They are actually continuous, and even affine, in this case). Also, since each affine function is trivially convex, and since the pointwise supremum of any set of convex functions is convex, ϕ^* is convex. We now show that ϕ^* is proper.

By the remarks made at the beginning of this section, $\partial\phi(\mathbb{R}) \neq \emptyset$. Pick some point x_0 such that $\partial\phi(x_0) \neq \emptyset$, and then pick $y \in \partial\phi(x_0)$. Then by (5.1.10), for all $x \in \mathbb{R}$, $\phi(x) \geq \phi(x_0) + y(x - x_0)$. Therefore,

$$yx - \phi(x) \leq yx - (\phi(x_0) - y(x - x_0)) = yx_0 - \phi(x_0) \ .$$

Therefore, taking the supremum over all $x \in \mathbb{R}$, $\phi^*(y) \leq yx_0 - \phi(x_0) < \infty$. Hence, ϕ^* is proper. We have proved:

5.1.13 THEOREM. *Let ϕ be a proper convex function on $\mathbb{R}$. The its Legendre transform ϕ^* is a proper, lower-semicontinuous convex function on $\mathbb{R}$, and ϕ^* is finite at every point in $\partial\phi(\mathbb{R})$.*

5.1.14 EXAMPLE. *For $1 < p < \infty$, define $\phi_p(x) := p^{-1}|x|^p$. Note that ϕ_p is continuously differentiable with $\phi_p'(x) = |x|^{p-1}\mathrm{sgn}(x)$ which is strictly monotone increasing. Hence $\lim_{x\uparrow\infty} \sigma_+^\phi(x) = \lim_{x\uparrow\infty} \phi_p'(x) = \infty$, and $\lim_{x\downarrow-\infty} \sigma_-^\phi(x) = \lim_{x\downarrow-\infty} \phi_p'(x) = -\infty$. Thus, $\partial\phi_p(\mathbb{R}) = \mathbb{R}$ and ϕ_p^* will be defined on all of $\mathbb{R}$.*

To compute it, note that for all $y \in \mathbb{R}$, the function $x \mapsto xy - \phi(x)$ is continuously differentiable, and its derivative is $y - |x|^{p-1}\mathrm{sgn}(x)$, and hence the unique maximum occurs where x has the same sign as y, and $|x| = |y|^{1/(p-1)}$. Evaluating $xy - \phi_p(x)$ as this x, we find

$$\phi_p^*(y) = |y|^{1+1/(p-1)} - \frac{1}{p}|y|^{p/(p-1)} = \frac{p-1}{p}|y|^{p/(p-1)} \ .$$

Hence if we define $q = p/(p-1)$, we have that

$$\frac{1}{p} + \frac{1}{q} = 1 \qquad \text{and} \qquad \phi_p^* = \phi_q \ . \tag{5.1.18}$$

Since the relation (5.1.18) is symmetric in p and q, $(\phi_p^*)^* = \phi_p$. We shall soon see that this is no coincidence.

5.1.15 EXAMPLE. *Now consider the two limiting functions $\phi_1(x) = \lim_{p\downarrow1} \phi_p(x) = |x|$, and $\phi_\infty(x) = \lim_{p\uparrow1} \phi_p(x)$, so that*

$$\phi_\infty(x) := \begin{cases} 0 & x \in [-1,1] \\ \infty & x \notin [-1,1] \ . \end{cases} \tag{5.1.19}$$

Both ϕ_1 and ϕ_∞ are lower-semicontinuous convex functions. (In fact, ϕ_1 is even continuous.) Fix $y \in \mathbb{R}$, and note that $yx - \phi(x) = (y\mathrm{sgn}(x) - 1)|x| \leq (|y| - 1)|x|$, with equality when $\mathrm{sgn}(x) = \mathrm{sgn}(y)$. It follows that $\sup_{x\in\mathbb{R}}\{yx - \phi_1(x)\} = \phi_\infty(y)$. That is, $\phi_1^* = \phi_\infty$.

It is also true that $\phi_\infty^* = \phi_1$. To see this, fix $x \in \mathbb{R}$, and note that $xy - \phi_\infty(y) = xy$ for $|y| \leq 1$ and $xy - \phi_\infty(y) = -\infty$ for $|y| > 1$. Hence

$$\phi_\infty^*(x) = \sup\{xy \ : \ |y| \leq 1\} = |x| = \phi_1(x) \ .$$

We shall see below that the fact that $\phi_1^{**} = \phi_1$ is no coincidence.

5.1.16 THEOREM (Young's Inequality)**.** *Let ϕ be a proper convex function, and let ϕ^* be the Legendre transform of ϕ. Then for all $x, y \in \mathbb{R}$,*

$$xy \le \phi(x) + \phi^*(y) \ . \tag{5.1.20}$$

Moreover, there is equality in (5.1.20) if and only if $y \in \partial\phi(x)$, and in this case, $x \in \partial\phi^*(y)$.

Proof. The inequality (5.1.20) is an immediate consequence of the definition (5.1.17). Suppose for some $x_0, y_0 \in \mathbb{R}$, $x_0 y_0 = \phi(x_0) + \phi^*(y_0)$, so that necessarily $\phi(x_0)$ and $\phi^*(y_0)$ are both finite. By (5.1.20), for all $x, y \in \mathbb{R}$,

$$\phi(x) + \phi^*(y) - xy \ge \phi(x_0) + \phi^*(y_0) - x_0 y_0 \ .$$

Setting $y = y_0$, and cancelling $\phi^*(y_0)$ from both sides yields $\phi(x) \ge \phi(x_0) + y_0(x - x_0)$, which shows that $y_0 \in \partial\phi(x_0)$. Setting $x = x_0$ and cancelling $\phi(x_0)$ from both sides yields $\phi^*(y) \ge \phi^*(y_0) + x_0(y - y_0)$ which shows that $x_0 \in \partial\phi^*(y_0)$

Conversely, if $y_0 \in \partial\phi(x_0)$, then by definition, $\phi(x) \ge \phi(x_0) + y_0(x - x_0)$ for all x, and hence

$$x_0 y_0 \ge \phi(x_0) + y_0 x - \phi(x) \ .$$

Taking the supremum over x, we find $x_0 y_0 \ge \phi(x_0) + \phi^*(y_0)$. Together with (5.1.20), this proves that $x_0 y_0 = \phi(x_0) + \phi^*(y_0)$. □

Suppose that ϕ and ϕ^* are both continuously differentiable. For instance, this is the case when $\phi(x) = p^{-1}|x|^p$, $p \in (1, \infty)$, so that $\phi^*(y) = q^{-1}|y|^q$, $q = p/(p-1)$. Then for all x, y, $\partial\phi(x) = \{\phi'(x)\}$ and $\partial\phi^*(y) = \{(\phi^*)'(y)\}$. Then the statement about cases of equality in Young's inequality says that

$$[\phi^*]'(\phi'(x)) = x \qquad \text{and} \qquad \phi'[(\phi^*)'(y)] = y \ .$$

In other words, the functions ϕ' and $(\phi^*)'$ are inverse to one another. For the dual pair in Example 5.1.14, this can be checked by simple computations.

5.1.17 THEOREM (Fenchel-Moreau Theorem)**.** *Let ϕ be a proper, lower-semicontinuous function on $\mathbb{R}$ and let ϕ^* be its Legendre transform. Let ϕ^{**} be the Legendre transform of ϕ^*. Then*

$$\phi^{**} = \phi \ . \tag{5.1.21}$$

Proof. Example 5.1.10 takes care of the cases in which ϕ is finite only at one single point. Therefore, suppose ϕ is finite on an interval C with $C^o = (a, b)$. Let $x \in (a, b)$ and $y \in \partial\phi(x)$. Then $\phi(x) + \phi^*(y) = xy$, and $x \in \partial\phi^*(y)$. By Young's inequality applied to the pair ϕ^* and ϕ^{**}, $xy = \phi^*(y) + \phi^{**}(x)$, and

$$\phi(x) + \phi^*(y) = xy = \phi^*(y) + \phi^{**}(x) \ .$$

By Theorem 5.1.13, $\phi^*(y) < \infty$, so that it may be cancelled from both sides. This proves that at all points of (a, b), $\phi(x) = \phi^{**}(x)$. Since both ϕ and ϕ^{**} are lower-semicontinuous. $\phi = \phi^{**}$ on $[a, b]$. □

5.2 L^p norms and their close relatives

5.2.1 The L^p norm and Hölder's Inequality

5.2.1 DEFINITION ($L^p(\Omega,\mathcal{M},\mu)$)**.** Let $(\Omega,\mathcal{M},\mu)$ be a measure space. For $1 \le p < \infty$, $L^p(\Omega,\mathcal{M},\mu)$ consists of the equivalence classes, identified under equivalence almost everywhere, of measurable functions f such that $|f|^p$ is integrable. The L^p *norm* is the function $f \mapsto \|f\|_p$ where

$$\|f\|_p := \left(\int_\Omega |f|^p \mathrm{d}\mu\right)^{1/p} . \tag{5.2.1}$$

As we have already observed, the function $t \mapsto t^p$ is convex on $[0,\infty)$, and hence for $f,g \in L^p(\Omega,\mathcal{M},\mu)$, $x \in \Omega$,

$$\left|\frac{|f(x)|+|g(x)|}{2}\right|^p \le \frac{1}{2}|f(x)|^p + \frac{1}{2}|g(x)|^p .$$

This shows that $|f+g|^p$ is integrable whenever $|f|^p$ and $|g|^p$ are, and thus $L^p(\Omega,\mathcal{M},\mu)$ is closed under vector addition. It is also evidently closed under scalar multiplication, and thus is a vector space over $\mathbb{C}$. The next inequality will allow us to show that the L^p norm is, in fact, a norm on this vector space.

5.2.2 THEOREM (Hölder's Inequality)**.** *Let $(\Omega,\mathcal{M},\mu)$ be a measure space. Let $1 < p,q < \infty$, with $\frac{1}{p}+\frac{1}{q} = 1$. Let f and g be functions on $(\Omega,\mathcal{M},\mu)$ such that $|f|^q$ and $|g|^p$ are integrable. Then fg is integrable, and*

$$\int_\Omega |fg|\mathrm{d}\mu \le \|f\|_q\|g\|_p . \tag{5.2.2}$$

There is equality in (5.2.2) if and only if for almost every x,

$$\|g\|_p^p|f(x)|^q = \|f\|_q^q|g(x)|^p . \tag{5.2.3}$$

Proof. If either $\int_\Omega |f|^q\mathrm{d}\mu = 0$ or $\int_\Omega |g|^p\mathrm{d}\mu = 0$, then (5.2.2) is true for trivial reasons. Therefore, suppose that both integrals are strictly positive.

Apply Young's inequality with the dual pair ϕ_p and $\phi_q = \phi_p^*$ from Example 5.1.14. By (5.1.18), for any $a > 0$, and all $x \in \Omega$,

$$|f(x)||g(x)| = (a|f(x)|)\left(\frac{1}{a}|g(x)|\right) \le a^q\frac{1}{q}|f(x)|^q + a^{-p}\frac{1}{p}|g(x)|^p . \tag{5.2.4}$$

By Theorem 5.1.16, there is equality if and only if $a|f(x)| \in \partial\phi_p(a^{-1}|g(x)|)$, and hence in case

$$a|f(x)| = \phi_p'(a^{-1}|g(x)|) = a^{1-p}|g(x)|^{p-1} . \tag{5.2.5}$$

Integrating both sides of (5.2.4),

$$\int_\Omega |fg|\mathrm{d}\mu \le a^q\left(\frac{1}{q}\int_\Omega |f|^q\mathrm{d}\mu\right) + a^{-p}\left(\frac{1}{p}\int_\Omega |g|^p\mathrm{d}\mu\right) = a^q\frac{1}{q}\|f\|_q^q + a^{-p}\frac{1}{p}\|g\|_p^p . \tag{5.2.6}$$

We now choose the value of a so as to make the right hand side as small as possible. A simple calculus exercise shows that the best choice is $a = \|f\|_q^{-1/p}\|g\|_p^{1/q}$. With this choice of a, (5.2.6) becomes (5.2.2), and (5.2.5) becomes $\|g\|_p^{p/q}|f(x)| = \|f\|_q|g(x)|^{p-1}$, and since $q = p/(p-1)$, raising both sides to the qth power yields (5.2.3). □

Consider any non-zero $f \in L^p(\Omega, \mathcal{M}, \mu)$, $1 < p < \infty$, and let $q = p/(p-1)$. By Hölder's inequality, for all $h \in L^q(\Omega, \mathcal{M}, \mu)$ with $\|h\|_q = 1$,

$$\Re\left(\int_\Omega \overline{h} f \mathrm{d}\mu\right) \leq \int_\Omega |h||f|\mathrm{d}\mu \leq \|f\|_p \ ,$$

and there is equality in the second inequality if and only if $\|f\|_p^p |h(x)|^q = \|h\|_q^q |f(x)|^p = |f(x)|^p$ for almost every x. In this case, $|h(x)| = \|f\|_p^{1-p}|f(x)|^{p-1}$. There will be equality in the first inequality if and only if $\Re(\overline{h(x)}f(x)) = |h(x)||f(x)|$ almost everywhere. This forces that

$$g(x) := \|f\|_p^{1-p}|f(x)|^{p-1}\mathrm{sgn}(f(x)) \ ,$$

almost everywhere. (The *signum function*, $z \mapsto \mathrm{sgn}(z)$ on $\mathbb{C}$ is defined by $z \in \mathbb{C}$, define $\mathrm{sgn}(z) = z/|z|$ for $z \neq 0$, and $\mathrm{sgn}(0) = 0$.)

5.2.3 DEFINITION. For $1 < p < \infty$, $q = p/(p-1)$ define the function D_p mapping $L^p(\Omega, \mathcal{M}, \mu)\backslash\{0\}$ to the unit sphere in $L^q(\Omega, \mathcal{M}, \mu)$ by

$$D_p(f) = \|f\|_p^{1-p}|f|^{p-1}\mathrm{sgn}(f) \ . \tag{5.2.7}$$

$f \mapsto D_p(f)$ is called the *gradient map* for reasons that will become clear.

Altogether, we have proved:

5.2.4 THEOREM. *Let $1 < p < \infty$, $q = p/(p-1)$. For all $f \in L^p(\Omega, \mathcal{M}, \mu)$,*

$$\|f\|_p = \sup\left\{ \Re\left(\int_\Omega \overline{h} f \mathrm{d}\mu\right) \ : \ h \in L^p(\Omega, \mathcal{M}, \mu), \|h\|_q = 1 \right\} \tag{5.2.8}$$

Moreover, the supremum in (5.2.8) is a maximum, and when $f \neq 0$, the unique maximizer is $h = D_p(f)$.

5.2.5 THEOREM (Minkowski's Inequality)**.** *Let $1 < p < \infty$, $q = p/(p-1)$. For all $f, g \in L^p(\Omega, \mathcal{M}, \mu)$,*

$$\|f+g\|_p \leq \|f\|_p + \|g\|_p \tag{5.2.9}$$

and there is equality if and only if either $f = 0$, or else g is a non-negative multiple of f.

Proof. We may suppose that neither $f = 0$ nor $g = 0$. By Theorem 5.2.4,

$$\|f+g\|_p = \int_\Omega \overline{D_p(f+g)}(f+g)\mathrm{d}\mu = \int_\Omega \overline{D_p(f+g)} f \mathrm{d}\mu + \int_\Omega \overline{D_p(f+g)} g \mathrm{d}\mu \leq \|f\|_p + \|g\|_p \ .$$

There is equality if and only if $D_p(f+g) = D_p(f) = D_p(g)$. The second equality forces $\mathrm{sgn}(f) = \mathrm{sgn}(g)$ and $|f|^{p-1} = a|g|^{p-1}$ for some $a > 0$, and hence $|f| = a^{1/(p-1)}|g|$. $\square$

It is easy to prove that for $1 < p < \infty$, $L^p(\Omega, \mathcal{M}, \mu)$ is a complete metric space, and hence a Banach space. In fact the proof is very much like the one we have already given for completeness of $L^2(\Omega, \mathcal{M}, \mu)$, and it extends to a much wider class of norms that we now introduce.

5.2.2 Orlicz spaces

Throughout this section, let $(\Omega, \mathcal{M}, \mu)$ be a given measure space, and define $L^0(\Omega, \mathcal{M}, \mu)$ to be the vector space of measurable complex valued functions f on Ω, identified under almost-everywhere equivalence.

5.2.6 DEFINITION. An *Orlicz function* ϕ is a convex lower-semicontinuous function on $\mathbb{R}$ such that is symmetric ($\phi(-x) = \phi(x)$ for all x), with $\phi(0) = 0$, $\phi(1) < \infty$ with $1 \in \partial\phi(1)$, and $\lim_{x\uparrow\infty} \phi(x) = \infty$. For any Orlicz function ϕ, define

$$B_\phi = \left\{ f \in L^0(\Omega, \mathcal{M}, \mu) : \int_\Omega \phi(|f|)\mathrm{d}\mu \le \phi(1) \right\} .$$

Define L_ϕ to be the subspace of $L^0(\Omega, \mathcal{M}, \mu)$ spanned by B_ϕ.

Every Orlicz function ϕ is non-negative: Since ϕ is convex and even, $\phi(0) \le \frac{1}{2}(\phi(x) + \phi(-x)) = \phi(x)$, Hence 0 is a minimizer for any symmetric convex function, and since ϕ is a Orlicz function, $\phi(0) = 0$. Evidently the x-axis is a supporting line for the graph of ϕ, and so $0 \in \partial\phi(0)$.

We claim that ϕ^* is also an Orlicz function. First, ϕ^* is evidently convex proper, symmetric and lower semicontinuous. By the remark above, $\phi^*(0) \le \phi^*(y)$ for all y. Since $0 \in \partial\phi(0)$, the conditions for equality in Young's inequality give us $0 = \phi(0) + \phi^*(0)$, so that $\phi^*(0) = 0$. Also, since $1 \in \partial\phi(1)$, the conditions for equality in Young's inequality give $1 = \phi(1) + \phi^*(1)$, so that $\phi^*(1)$ and $\phi(1)$ are not only finite, but both lie in $[0, 1]$, and $1 \in \partial\phi^*(1)$. Finally, since $1 \in \partial\phi^*(1)$, $\phi^*(y) \ge \phi^*(1) + (y - 1)$, and hence $\lim_{y\uparrow\infty} \phi^*(y) = \infty$. Summarizing, we have:

5.2.7 LEMMA. *Let ϕ be an Orlicz function, and let ϕ^* be its Legendre transform. Then ϕ^* is also an Orlicz function and*

$$\phi(1) + \phi^*(1) = 1 . \tag{5.2.10}$$

5.2.8 EXAMPLE. *Let $1 \le p < \infty$, and let $\phi_p(x) = p^{-1}|x|^p$, which is easily seen to be an Orlicz function since it is continuously differentiable at $x = 1$, and $\phi_p'(1) = 1$, showing that $\partial\phi_p(1) = \{1\}$. A measurable function f belongs to B_{ϕ_p} if and only if*

$$\frac{1}{p}\int_\Omega |f|^p\mathrm{d}\mu \le \frac{1}{p} ,$$

and hence $f \in B_{\phi_p}$ is and only if $\|f\|_p \le 1$. Thus, B_{ϕ_p} is precisely the closed unit ball in $L^p(\Omega, \mathcal{M}, \mu)$. Therefore, $L_{\phi_p} = L^p(\Omega, \mathcal{M}, \mu)$. As we have seen in Example 5.1.15, $\phi_1^* = \phi_\infty$ where ϕ_∞ is defined in (5.1.19). Then $\phi_\infty(1) = 0$, and hence

$$\int_\Omega \phi_\infty(|f|)\mathrm{d}\mu \le \phi_\infty(1)$$

if and only if $\phi_\infty(|f|) = 0$ almost everywhere, and this is the case if an only if the essential supremum $\|f\|_\infty$ of $|f|$ belongs to $[0, 1]$. The subspace of $L^0(\Omega, \mathcal{M}, \mu)$ of functions with $\|f\|_\infty < \infty$ defines the space $L^\infty(\Omega, \mathcal{M}, \mu)$, and hence $L^\infty(\Omega, \mathcal{M}, \mu) = L_{\phi_\infty}$. In particular, each L^p space, $1 \le p \le \infty$ is an Orlicz space. We now introduce norms which will turn out to be the L^p norms on the L^p spaces.

5.2.9 LEMMA. *Let ϕ be any Orlicz function. Then B_ϕ is a balanced convex subset of $L^0(\Omega, \mathcal{M}, \mu)$, and $\cap_{r>0} rB_\phi = \{0\}$.*

Since B_ϕ is absorbing in L_ϕ by definition, the function $f \mapsto \|f\|_\phi$ in L_ϕ given by

$$\|f\|_\phi = \inf\{r > 0 \ : f \in rB_\phi\ \} \tag{5.2.11}$$

is a norm on L_ϕ called the *Luxembourg norm on* L_ϕ. Notice that

$$f \in rB_\phi \quad \Longleftrightarrow \quad \frac{1}{r} f \in B_\phi \quad \Longleftrightarrow \quad \int_\Omega \phi\left(\frac{|f|}{r}\right) \mathrm{d}\mu \leq \phi(1) \ .$$

Therefore, an equivalent form of (5.2.11) is

$$\|f\|_\phi = \inf\left\{ r > 0 \ : \ \int_\Omega \phi\left(\frac{|f|}{r}\right) \mathrm{d}\mu \leq \phi(1) \ \right\} \ . \tag{5.2.12}$$

5.2.10 EXAMPLE. *Continuing with Example 5.2.8, it is clear that* $\|\cdot\|_{\phi_p} = \|\cdot\|_p$.

5.2.11 LEMMA. *Let ϕ be an Orlicz function. For all $f \in L_\phi$,*

$$\int_\Omega \phi\left(\frac{f}{\|f\|_\phi}\right) \mathrm{d}\mu \leq \phi(1) \ . \tag{5.2.13}$$

and for all non-negative measurable functions f and g,

$$f \leq g \quad \Rightarrow \quad \|f\|_\phi \leq \|g\|_\phi \ , \tag{5.2.14}$$

Proof. By (5.2.12), for all $n \in \mathbb{N}$, $\displaystyle\int_\Omega \phi\left(\frac{|f|}{\|f\|_\phi + 1/n}\right) \mathrm{d}\mu \leq \phi(1)$, and then (5.2.13) follows from Fatou's Lemma and the lower-semicontinuity of ϕ.

Next, since ϕ is monotone on $[0,\infty)$, for all $t > 0$, $\displaystyle\int_\Omega \phi\left(\frac{|f|}{t}\right) \mathrm{d}\mu \leq \int_\Omega \phi\left(\frac{|g|}{t}\right) \mathrm{d}\mu$. From this and (5.2.12), we obtain (5.2.14). □

The next theorem shows that when ϕ is an Orlicz function, $(L_\phi, \|\cdot\|_\phi)$ is complete, and hence is a Banach space. By what has been shown in the examples above, this extends the Riesz-Fischer Theorem from L^2 to L^p for all $1 \leq p \leq \infty$.

5.2.12 THEOREM (Completeness of Orlicz spaces)**.** *Let ϕ be an Orlicz function, and let $(\Omega, \mathcal{M}, \mu)$ be a measure space, Let $(L_\phi, \|\cdot\|_\phi)$ be the Orlicz space associated to ϕ and $(\Omega, \mathcal{M}, \mu)$. Then $(L_\phi, \|\cdot\|_\phi)$ is a Banach space, and from every Cauchy sequence $\{f_n\}_{n\in\mathbb{N}}$ in L_ϕ, one may extract a subsequence that converges almost everywhere.*

Proof. For each k, pick n_k so that $j, \ell \geq n_k \Rightarrow \|f_j - f_\ell\|_\phi \leq 2^{-k}$. Without loss of generality, we may suppose that $n_{k+1} > n_k$ for each k. For $N \in \mathbb{N}$, define the function $F_N(x)$ by

$$F_N(x) = |f_{n_1}(x)| + \sum_{k=1}^N |f_{n_{k+1}}(x) - f_{n_k}(x)| \ .$$

By Minkowski's inequality,

$$\|F_N\|_\phi \leq \|f_{n_1}\|_\phi + \sum_{k=1}^N \|f_{n_{k+1}} - f_{n_k}\|_\phi \leq \|f_{n_1}\|_\phi + \sum_{k=1}^N 2^{-k} \leq \|f_{n_1}\|_\phi + 1 \ .$$

Let $c = \|f_{n_1}\|_\phi + 1$. By (5.2.12), $\int_\Omega \phi\left(\frac{F_N(x)}{c}\right) \mathrm{d}x \leq \phi(1)$. Define $F(x) = \lim_{N\to\infty} F_N(x)$. Since ϕ is lower-semicontinuous, $\phi\left(\frac{F(x)}{c}\right) \leq \liminf_{N\to\infty} \phi\left(\frac{F_N(x)}{c}\right)$, and then by Fatou's Lemma,

$$\int_\Omega \phi\left(\frac{F(x)}{c}\right) \mathrm{d}x \leq \liminf_{N\to\infty} \int_\Omega \phi\left(\frac{F_N(x)}{c}\right) \mathrm{d}x \leq \phi(1) \ .$$

Hence $\|F\|_\phi \leq c$, and since $\lim_{t\uparrow\infty} \phi(t) = \infty$, F is finite almost everywhere.

Next define f by

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

where the sum on the right converges absolutely wherever F is finite; that is, almost everywhere. Since $|f| \leq F$, $f \in L_\phi$ by Lemma 5.2.11.

By the definition of the subsequence $\{f_{n_j}\}_{j\in\mathbb{N}}$, for all $k \in \mathbb{N}$, $\int_\Omega \phi\left(\frac{|f_{n_\ell} - f_{n_k}|}{2^{-k}}\right) \mathrm{d}\mu \leq \phi(1)$. Again since ϕ is lower-semicontinuous, $\phi\left(\frac{|f - f_{n_k}|}{2^{-k}}\right) \leq \liminf_{\ell\to\infty} \phi\left(\frac{|f_{n_x} - f_{n_k}|}{2^{-k}}\right)$, and by Fatou's Lemma once more, $\|f - f_{n_k}\|_\phi \leq 2^{-k}$.

$$\lim_{k\to\infty} \|f_{n_k} - f\|_\phi = 0 \ .$$

Since the sequence is Cauchy, we have this convergence along the whole sequence as well. □

5.2.3 Duality in Orlicz spaces

Let ϕ be an Orlicz function, and let ϕ^* be its Legendre transform. Any such pair of Orlicz functions is called a *dual pair of Orlicz functions*. Associated to such a pair of Orlicz functions is the pair of normed spaces $(L_\phi, \|\cdot\|_\phi)$ and $(L_{\phi^*}, \|\cdot\|_{\phi^*})$.

What is the relation between L_ϕ^* and L_{ϕ^*}? The first step towards answering this question provided by the *generalized Hölder inequality*:

5.2.13 THEOREM (Generalized Hölder's Inequality)**.** *Let ϕ, ϕ^* be any dual pair of Orlicz functions. Then for all functions $f \in L_\phi$ and $g \in L_{\phi^*}$, fg is measurable and*

$$\int_\Omega |fg| \mathrm{d}\mu \leq \|f\|_\phi \|g\|_{\phi^*} \ . \tag{5.2.15}$$

There is equality in (5.2.15) if and only if for almost every x, $\frac{|g(x)|}{\|g\|_{\phi^*}} \in \partial\phi\left(\frac{|f(x)|}{\|f\|_\phi}\right)$,

$$\int_\Omega \phi\left(\frac{|f|}{\|f\|_\phi}\right) \mathrm{d}\mu = \phi(1) \qquad \text{and} \qquad \int_\Omega \phi^*\left(\frac{|g|}{\|g\|_{\phi^*}}\right) \mathrm{d}\mu = \phi^*(1) \tag{5.2.16}$$

The following lemma will be useful here and elsewhere:

5.2.14 LEMMA. *Let ϕ be an Orlicz function. Lor all $a, b > 0$, define the function $\phi_{a,b}$ by $\phi_{a,b}(s) = b\phi(as)$. Then*

$$\phi_{a,b}^*(t) = b\phi^*\left(\frac{t}{ab}\right) \ .$$

Proof.

$$\phi^*_{a,b}(t) = \sup_{s\geq 0}\{ts - b\phi(as)\} = \sup_{s\geq 0}\left\{b\frac{t}{ab}as - b\phi(as)\right\} = b\phi^*\left(\frac{t}{ab}\right) \ .$$

□

Proof of Theorem 5.2.13. By Young's inequality applied to $\phi_{a,b}$ with $a, b > 0$, we have from Lemma 5.2.14 that

$$st \leq b\phi(as) + b\phi^*\left(\frac{t}{ab}\right)$$

for all $s, t \geq 0$, and there is equality only if $t \in \partial\phi_{a,b}(s)$ Then for any $f \in L_\phi$ and $g \in L_{\phi^*}$,

$$\int_\Omega |fg|\mathrm{d}\mu \leq b\int_\Omega \phi(|af|)\mathrm{d}\mu + b\int_\Omega \phi^*\left(\frac{|g|}{ab}\right)\mathrm{d}\mu \ .$$

By Lemma 5.2.11, $\int_\Omega \phi\left(\frac{|f|}{\|f\|_\phi}\right)\mathrm{d}\mu \leq \phi(1)$ and $\int_\Omega \phi^*\left(\frac{|g|}{\|g\|_{\phi^*}}\right)\mathrm{d}\mu \leq \phi^*(1)$. Choosing $a = \frac{1}{\|f\|_\phi}$ and $b = \|f\|_\phi\|g\|_{\phi^*}$, we obtain

$$\int_\Omega |fg|\mathrm{d}\mu \leq \|f\|_\phi\|g\|_{\phi^*}\left(\phi(1) + \phi^*(1)\right) = \|f\|_\phi\|g\|_{\phi^*} \ ,$$

where we have used the fact that $1 = \phi(1) + \phi^*(1)$ since $1 \in \partial\phi(1)$. There is equality if and only if (5.2.16) is valid and equality holds in the application of Young's inequality at almost every x, which means that $|g(x)| \in \partial\phi_{a,b}(|f(x)|)$. Then since $s \in \partial\phi_{a,b}(t)$ if and only if $\frac{s}{ab} \in \partial\phi(at)$, there is equality in (5.2.15) if and only if (**??**) is valid. □

Suppose that ϕ and ϕ^* are a dual pair of Orlicz functions. For all $g \in L_{\phi^*}$, define the linear functional L_g on L_ϕ by

$$L_g(f) = \int_\Omega \overline{g}f\mathrm{d}\mu \ , \tag{5.2.17}$$

which is well-defined by Theorem 5.2.13, and in fact, by Theorem 5.2.13,

$$|L_g(f)| \leq \|f\|_\phi\|g\|_{\phi^*}.$$

Therefore, $L_g \in L^*_\phi$, and

$$\|L_g\|_{L^*_\phi} \leq \|g\|_{\phi^*} \ .$$

Thus, the mapping $g \mapsto L_g$ is a linear contraction from L_{ϕ^*} into L^*_ϕ. It is not hard to show that it is injective, at least when μ has the property that every measurable set with positive measure contains a measurable set with finite positive measure. One might hope that this map would also be surjective under these same mild conditions. In that case, by the Open Mapping Theorem, it would be a Banach space isomorphism, and thus we would identify L^*_ϕ with L_{ϕ^*}. But then the same argument would identify $L^*_{\phi^*}$ with $L_{\phi^{**}} = L_\phi$, and we would have that the natural injection of L_ϕ into L^{**}_ϕ would be a Banach space isomorphism, i.e., that L_ϕ would be *reflexive*. This is not true in general: It fails for ϕ_1 and ϕ_∞, but it is the case for ϕ_p, $1 < p < \infty$. A powerful key to this and other issues lies in the notion of *uniform convexity*, to which we now turn.

5.3 Uniform convexity and uniform smoothness

5.3.1 Uniform convexity

The unit ball of any normed vector space V is convex, though it need not be strictly convex, which would mean that for all unit vectors u and v in V,

$$\|u - v\| > 0 \Rightarrow \|(u+v)/2\| < 1 \ . \tag{5.3.1}$$

Indeed, strict convexity fails for $L^1(\Omega, \mathcal{M}, \mu)$ and $L^\infty(\Omega, \mathcal{M}, \mu)$, even for a two-point measure space.

To see this in $L^1(\Omega, \mathcal{M}, \mu)$, take any two *non-negative* unit vectors $u(x)$ and $v(x)$. Then of course

$$\left\|\frac{u+v}{2}\right\| = \frac{1}{2}\int_\Omega (u(x)+v(x))\mathrm{d}\mu = \frac{1}{2}\int_\Omega u(x)\mathrm{d}\mu + \frac{1}{2}\int_\Omega v(x)\mathrm{d}\mu = 1 \ .$$

To see this in $L^\infty(\Omega, \mathcal{M}, \mu)$, take any two $u(x)$ and $v(x)$ to be the indicator functions of two measurable sets E and F respectively such that $\mu(E \cap F) > 0$. Then $u(x)$ and $v(x)$ are both unit vectors in $L^\infty(\Omega, \mathcal{M}, \mu)$, as is their average, $(u+v)/2$.

In some normed spaces however, a *uniform* version of strict convexity holds, and this has significant consequences.

5.3.1 DEFINITION (Uniform convexity)**.** Let $(V, \|\cdot\|)$ be a normed vector space. The *modulus of convexity* of $(V, \|\cdot\|)$ is the function δ_V defined by

$$\delta_V(\epsilon) = \inf\left\{1 - \left\|\frac{v+w}{2}\right\| \ : \ \|v-w\| \geq 2\epsilon \right\} \tag{5.3.2}$$

for $0 < \epsilon \leq 1$. We say that V is *uniformly convex* in case $\delta_V(\epsilon) > 0$ for all $0 < \epsilon < 1$.

Since $v = \dfrac{v+w}{2} + \dfrac{v-w}{2}$, $1 - \left\|\dfrac{v+w}{2}\right\| \leq \left\|\dfrac{v-w}{2}\right\|$

$$\delta_V(\epsilon) \leq \epsilon \tag{5.3.3}$$

for all $\epsilon \in (0,1]$. In fact, we shall soon see that $\lim_{\epsilon\to 0}\epsilon^{-1}\delta_V(\epsilon) = 0$. By definition, the function $\delta_V : [0,1] \to [0,1]$ is monotone non-decreasing.

5.3.2 LEMMA. *Let $(V, \|\cdot\|)$ be a uniformly conves normed vector space. Let $\{u_n\}_{n\in\mathbb{N}}$ and $\{v_n\}_{n\in\mathbb{N}}$ be teo sequences of unit vectors. Then*

$$\lim_{n\to\infty}\left\|\frac{u_n+v_n}{2}\right\| = 1 \quad \Rightarrow \quad \lim_{n\to\infty}\|u_n - v_n\| = 0 \ . \tag{5.3.4}$$

Proof. An immediate consequence of the definition is that for any two unit vectors u and v,

$$1 - \left\|\frac{u+v}{2}\right\| \geq \delta_V\left(\frac{\|u-v\|}{2}\right) \ . \tag{5.3.5}$$

Therefore, if $\displaystyle\lim_{n\to\infty}\left\|\frac{v+w}{2}\right\| = 1$, then by (5.3.5) $\displaystyle\lim_{n\to\infty}\delta_V\left(\frac{\|u_n-v_n\|}{2}\right) = 0$. If V is unifromly convex, then since $\epsilon \mapsto \delta_V(\epsilon)$ is strictly positve for $\epsilon > 0$ and monotone increasing, $\lim_{n\to\infty}\|u_n - v_n\| = 0$. □

Lemma 5.3.2 is the basis of many aplications of uniform convexity. Before we begin to apply uniform convexity, we should first explain the uniformly convex normed spaces exist. By what we have seen above, neither $L^1(M, \mathcal{M}, \mu)$ nor $L^\infty(M, \mathcal{M}, \mu)$ is uniformly convex. It turns out, however, that for $1 < p < \infty$, $L^p(M, \mathcal{M}, \mu)$ *is* uniformly convex. This is easiest to show for $L^2(M, \mathcal{M}, \mu)$, and we begin with that:

For any f and g in we have the *parallelogram identity* $\|f - g\|_2^2 + \|f + g\|_2^2 = 2\|f\|_2^2 + 2\|g\|_2^2$. Take f and g to be unit vectors. Divide through by 4 to obtain

$$\left\|\frac{f+g}{2}\right\|_2^2 + \left\|\frac{f-g}{2}\right\|_2^2 = \frac{\|f\|_2^2 + \|g\|_2^2}{2} = 1 \ .$$

Therefore, $\left\|\frac{f+g}{2}\right\|_2 = \sqrt{1 - \left\|\frac{f-g}{2}\right\|_2^2}$. For any number a with $0 < a < 1$, $\sqrt{1-a} < 1 - a/2$, and hence,

$$\left\|\frac{f+g}{2}\right\|_2 \leq 1 - \frac{1}{2}\left\|\frac{f-g}{2}\right\|_2^2 \ .$$

Since f and g are arbitrary unit vecotors, this gives us the exact modulus of convexity for $L^2(M, \mathcal{M}, \mu)$, namely

$$\delta_{L^2}(\epsilon) = 1 - \sqrt{1 - \epsilon^2} \geq \frac{1}{2}\epsilon^2 \ . \tag{5.3.6}$$

5.3.2 First applications of uniform convexity

5.3.3 THEOREM (Convergence of norms plus weak convergence yields strong convergence). *Let V be a uniformly convex normed space. Let $\{f_n\}_{n\in\mathbb{N}}$ converge weakly to f in V. Then $\{f_n\}_{n\in\mathbb{N}}$ is strongly convergent if and only if* $\|f\| = \lim_{n\to\infty} \|f_n\|$.

Proof. If $\{f_n\}_{n\in\mathbb{N}}$ is strongly convergent, it must converge strongly to f, and then $\|f\| = \lim_{n\to\infty} \|f_n\|$.

The converse is more subtle, and it is here that uniform convexity comes in. If $f = 0$, the strong convergence is obvious. This case aside, suppose that $\|f\| > 0$, and then, dividing through by $\|f\|$, that $\|f\| = 1$. Since $\lim_{n\to\infty} \|f_n\| = \|f\| = 1$, we may delete a finite number of terms from the sequence to arrange that $\|f_n\| \neq 0$ so any n.

Consider the sequence $\{g_n\}_{n\in\mathbb{N}}$ where

$$g_n = \frac{f_n/\|f_n\| + f}{2} \ .$$

Since $\lim_{n\to\infty} \|f_n\| = \|f\| = 1$, $\{g_n\}_{n\in\mathbb{N}}$ also converges weakly to f. By the weak lower semicontinuity of the norms, $\liminf_{n\to\infty} \|g_n\| \geq \|f\| = 1$. By Minkowski's inequality, $1 = \frac{\|f_n\|/\|f_n\| + \|f\|}{2} \geq \|g_n\|$. Altogether, $\lim_{n\to\infty} \|g_n\| = 1$. Then by Lemma 5.3.2, $\lim_{n\to\infty} \|f_n/\|f_n\| - f\| = 0$. But $\|f_n - f\| \leq \frac{|\|f_n\| - 1|}{\|f_n\|} + \|f_n/\|f_n\| - f\|$. Hence it follows that $\lim_{n\to\infty} \|f_n - f\| = 0$. □

Our next application is very important: It is the generalization of the Projection Lemma to general uniformly convex spaces.

5.3.4 THEOREM (Projection Lemma for uniformly convex spaces). *Let V be a uniformly convex Banach space, and let K be a non-empty, closed convex set in V. Then there exists a unique element of minimal norm in K. That is, there exists an element $v \in K$ with $\|v\| < \|w\|$ for all $w \in K$ with $w \neq v$. Moreover, if $\{v_n\}_{n\in\mathbb{N}}$ is any sequence in K such that* $\lim_{n\to\infty} \|v_n\| = \|v\|$, $\lim_{n\to\infty} \|v_n - v\| = 0$.

Proof. Let $D = \inf\{\|w\| \mid w \in K\}$. Let $\{v_n\}_{n\in\mathbb{N}}$ be any sequence in K with $\lim_{n\to\infty} \|v_n\| = D$. If $D = 0$, $\lim_{n\to\infty} v_n = 0$. Since K is closed, this means $0 \in K$, and this is our unique element of minimal norm. Therefore, assume that $D > 0$.

Normalize the v_n to obtain unit vectors, as needed for the application of uniform convexity. Let $u_n = v_n/\|v_n\|$. For large n, we have that $u_n \approx v_n/D$ since $\lim_{k\to\infty} \|v_k\| = D$. Indeed, adding and subtracting,

$$u_n = \frac{1}{D}v_n + \frac{D - \|v_n\|}{D\|v_n\|}v_n \ .$$

Therefore, for any m and n,

$$\begin{aligned} \frac{1}{D}\left\|\frac{v_n + v_m}{2}\right\| &= \left\|\frac{u_n + u_m}{2} - \frac{D - \|v_n\|}{2D\|v_n\|}v_n - \frac{D - \|v_m\|}{2D\|v_m\|}v_m\right\| \\ &\le \left\|\frac{u_n + u_m}{2}\right\| + \frac{(\|v_n\| - D) + (\|v_m\| - D)}{2D} \ . \end{aligned}$$

Since K is convex, $(v_n + v_m)/2 \in K$ and hence $\|(v_n + v_m)/2\| \ge D$. Therefore,

$$\frac{(\|v_n\| - D) + (\|v_m\| - D)}{2D} \ge 1 - \left\|\frac{u_n + u_m}{2}\right\| \ge \delta_V\left(\frac{u_n - u_m}{2}\right) \ .$$

Since $\displaystyle\lim_{m,n\to\infty} \frac{(\|v_n\| - D) + (\|v_m\| - D)}{2D} = 0$, $\lim_{m,n\to\infty} \|u_n - u_m\| = 0$, as in the proof of Lemma 5.3.2. Hence $\{u_n\}_{n\in\mathbb{N}}$ is a Cauchy sequence. Since V is complete, $\{u_n\}_{n\in\mathbb{N}}$ converges in norm to $u \in V$, and this implies that $\{v_n\}_{n\in\mathbb{N}}$ converges in norm to $v := Du$. Since K is closed, $Du \in K$, and since $\|u\| = 1$, $\|v\| = \|Du\| = D$. This proves the existence of an element v of K with minimal norm, and that $\lim_{n\to\infty} v_n = v$.

To prove the uniqueness, let $\tilde{v}$ also be in K with $\|\tilde{v}\| = D$. Define $v_n = n$ for v even and $v_n = \tilde{v}$ for n odd. By what we proved above $\{v_n\}_{n\in\mathbb{N}}$ converges, and hence $\tilde{v} = v$. $\square$

Recall that for any normed space V, and any $v \in V$, there exists an $f \in V^*$ with $\|f\|_* = 1$ and $f(v) = \|v\|$. This is a consequence of the Hahn-Banach Theorem. However, given $f \in V^*$, there may or may not be any unit vector u in V such that $f(u) = \|f\|_*$, as we have seen in the case of $V = \mathcal{C}([0,1])$ with the uniform norm. If V is uniformly convex, things are much better.

5.3.5 THEOREM (Uniform Convexity and Unit Normal Vectors)**.** *Let V be a uniformly convex Banach space, and let L be any non–zero linear functional in V^*. Then there is a unique unit vector $v_L \in V$ so that*

$$L(v_L) = \|L\|_* \ .$$

Moreover, the function $L \mapsto v_L$ from V^* to V is continuous at $L \neq 0$, and in fact, for all non zero $L, M \in V^*$,

$$\|L - M\|_* \le \|Lf\|_*\delta_V(\epsilon) \ \Rightarrow \ \|v_L - v_M\| \le 2\epsilon \ . \tag{5.3.7}$$

The vector whose existence is asserted by the theorem is called the *unit normal vector at L* for reasons that will soon be explained.

Proof. Let K be given by $K = \{\ v \in V \ :\ L(v) = \|L\|_*\ \}$. K is closed, convex and non–empty. By the projection lemma, K contains a unique element v of minimal norm.

Note that $\|L\|_* = L(v) \le \|L\|_*\|v\|$, so $\|v\| \ge 1$. By the definition of $\|L\|_*$, and homogeneity of L, for all $\epsilon > 0$, there is a unit vector u with $L(u) > \|L\|_* - \epsilon$. Then taling $\epsilon < \|L\|_*$ and defining $w := \|L\|_*(L(u))^1 u$, $L(w) = \|L\|_*$ and $\|w\| \le \|L\|_*(\|L\|_* - \epsilon)^1$. Hence $\inf_{v\in K}\{\|v\|\} \le 1$, Then $\|v\| = 1$, and $L(v) = \|L\|$.

This proves existence and unqueness of v_L. We now show that $L \mapsto v_L$ is continuous. Let $L, M \in V^*$ be given, $L, M \ne 0$, and let v_L and v_M be the corresponding unit vectors in V. Then

$$\begin{aligned}
\|L+M\|_*\|v_L+v_M\| &\ge \Re\left((L+M)(v_L+v_M)\right) \\
&= \Re\left(L(v_L)+L(v_M)+M(v_L)+M(v_M)\right) \\
&= 2(\|L\|_*+\|M\|_*)+\Re\left(L(v_M)+M(v_L)-L(v_L)-M(v_M)\right) \\
&= 2(\|L\|_*+\|M\|_*)-\Re\left((L-M)(v_L-v_M)\right) \\
&\ge 2\|L+M\|_*-\|L-M\|_*\|v_L-v_M\| \ . \qquad (5.3.8)
\end{aligned}$$

Dividing through by $2\|L+M\|_*$ snd rearranging terms,

$$\left(\frac{\|L-M\|_*}{\|L+M\|_*}\right)\left\|\frac{v_L-v_M}{2}\right\| \ge 1 - \left\|\frac{v_L+v_M}{2}\right\| \ge \delta_V\left(\left\|\frac{v_L-v_M}{2}\right\|\right) \ .$$

Therefore,

$$\frac{\|L-M\|_*}{2\|L\|_*-\|L-M\|_*} \ge \left\|\frac{v_L-v_M}{2}\right\|^{-1}\delta_V\left(\left\|\frac{v_L-v_M}{2}\right\|\right) \ . \qquad (5.3.9)$$

for all $M \in V^*$ with $\|M-L\|_* < \|L\|_*$. Since in any case $\|v_L - v_M\| \le 2$, we also have the cruder but simpler inequality

$$\frac{\|L-M\|_*}{2\|L\|_*-\|L-M\|_*} \ge \delta_V\left(\left\|\frac{v_L-v_M}{2}\right\|\right) \ . \qquad (5.3.10)$$

If $\epsilon \in (0,1)$, and $\|L-M\|_* < \delta_V(\epsilon)\|L\|_*$, then the left side of (5.3.10) is at most $\delta_V(\epsilon)$, and hence $\|v_f - v_g\| \le 2\epsilon$, which proves (5.3.7). □

5.3.6 Remark. Since for any two unit vectors $v, w \in V$, there exist unit vectors $L, M \in V^*$ such that $\|L\| = \|M\| = 1$, $L(v) = 1$ and $M(w) = 1$, so that $v = v_L$ and $w = v_M$, follows from (5.3.9) that $\lim_{\epsilon\downarrow 0}\epsilon^{-1}\delta_V(\epsilon) = 0$, as mentioned earlier.

5.3.3 Uniform smoothness

Let $(V, \|\cdot\|)$ be a normed vector space. A functional F on V is *Frechét differentiable* at $u \in V$ in case there is a linear functional $L_{F,u} \in V^*$ so that

$$F(u+v) - F(u) = L_{F,u}(v) + o(\|v\|)$$

or, in other words, if

$$\lim_{v\to 0}\frac{|F(u+v)-F(u)-L_{F,u}(v)|}{\|v\|} = 0 \ , \qquad (5.3.11)$$

where the limit is taken in the norm sense.

There is another notion of differentiability, corresponding to the usual directional derivative. A functional F is said to be *Gateaux differentiable* at $u \in V$ in case for there is a linear functional $L_{F,u} \in V^*$ so that for each $v \in V$,

$$F(u+tv) - F(u) = tL_{F,u}(v) + o(t)$$

or, in other words, if

$$\lim_{v\to 0}\frac{|F(u+tv)-F(u)-tL_{F,u}(v)|}{t}=0\ . \tag{5.3.12}$$

If a functional F is Frechét differentiable, then it is Gateaux differentiable, and the two derivatives coincide. However, there are functionals that are Gateaux differentiable, but not Frechét differentiable.

To check differentiability from the definition, one to know the derivative $L_{F,u}$, and this is impracticle in many situations. There is, however, a necessary condition for differentiability that can be stated solely in terms of F itself. If F is Frechét differentiable at u, then for any w,

$$F(u+w)-F(u)=L_{F,u}(w)+o(\|w\|)$$

and

$$F(u-w)-F(u)=-L_{F,u}(w)+o(\|w\|)\ .$$

Summing, the terms involving $L_{F,u}$ cancel, and we have

$$F(u+w)+F(u-w)-2F(u)=o(\|w\|)\ .$$

In particular, a necessary condition for Frechét differentiability the norm functional on a Banach space is that

$$\left\|\frac{u+w}{2}\right\|+\left\|\frac{u-w}{2}\right\|-\|u\|=o(\|w\|)\ .$$

Taking u to be a unit vector, and writing $w=tv$, v a unit vector, bring us to:

5.3.7 DEFINITION (Uniform Smoothness)**.** Let V be a Banach space with norm $\|\cdot\|$. The *modulus of smoothness* of V is the function $\rho_V(\tau)$ defined by

$$\rho_V(\tau)=\sup\left\{\left\|\frac{u+\tau v}{2}\right\|+\left\|\frac{u-\tau v}{2}\right\|-1\ :\ \|u\|=\|v\|=1\right\} \tag{5.3.13}$$

for each $\tau\geq 0$. Then V is said to be *uniformly smooth* in case $\rho_V(\tau)=o(\tau)$, i.e., if

$$\lim_{\tau\to 0}\frac{\rho_V(\tau)}{\tau}=0\ . \tag{5.3.14}$$

It is easy to see that uniform smoothness fails for $L^1(\Omega,\mathcal{M},\mu)$ and $L^\infty(\Omega,\mathcal{M}\mu)$, even for a two-point measure space, while $L^2(\Omega,\mathcal{M},\mu)$ is uniformly smooth. This is left as an exercise. In fact, it is a good exercise to compute the moduli of smoothness for these spaces. The results are:

(1) When $V=L^1(\Omega,\mathcal{M},\mu)$, $\delta_V(\epsilon)=0$ *and* $\rho_V(\tau)=\tau$.

(2) When $V=L^2(\Omega,\mathcal{M},\mu)$, $\delta_V(\epsilon)=1-\sqrt{1-\epsilon^2}$ *and* $\rho_V(\tau)=\sqrt{1+\tau^2}-1$.

(3) When $V=L^\infty(\Omega,\mathcal{M},\mu)$, $\delta_V(\epsilon)=0$ *and* $\rho_V(\tau)=\tau$.

There is a close relation between uniform convexity and uniform smoothness. In order to specifying it, we introduce the notion of a *dual pair* of Banach spaces.

5.3.8 DEFINITION (Dual Pairs). A *dual pair* of Banach spaces is a pair of Banach spaces V and W with norms $\|\cdot\|_V$ and $\|\cdot\|_W$ respectively, and a bilinear form $\langle\cdot,\cdot\rangle$ on $V\times W$ so that for all $v\in V$,

$$\|v\|_V = \sup\{\ |\langle v,w\rangle|\ :\ w\in W\ ,\ \|w\|_W\le 1\ \} \tag{5.3.15}$$

and for all $w\in W$

$$\|w\|_W = \sup\{\ |\langle v,w\rangle|\ :\ v\in V\ ,\ \|v\|_V\le 1\ \}\ . \tag{5.3.16}$$

The primary example is that in which $W=V^*$, and

$$\langle v,w\rangle = w(v)\ .$$

Then (5.3.16) holds by the definition of the norm on V^*, while (5.3.15) holds by the Hahn-Banach Theorem, which asserts the existence of a $w\in V^*$ with $\|w\|=1$ and $w(v)=\|v\|$.

When V and W are a dual pair, there is a map from V into W^* which assigns to v the linear functional $f_v(\cdot)=\langle v,\cdot\rangle$. By (5.3.16), and the definition of the dual norm $\|\cdot\|_*$, $\|f_w\|_*=\|w\|_W$. Hence the map $w\mapsto\langle\cdot,w\rangle$, which is clearly linear, is also an isometry.

However, it need not be the case that its image is all of V^*. In summary:

5.3.9 LEMMA. *When V and W are a dual pair, W may be identified with a subspace (which may be proper) of V^* through the isometric linear transformation*

$$w\mapsto\langle\cdot,w\rangle\ .$$

We now prove that when V and W are a dual pair, the moduli of smoothness and convexity of the one space can be determined from those of the other.

5.3.10 THEOREM (Lindenstrauss–Day Theorem). *Let V and W be a dual pair of Banach spaces. Then*

$$\rho_W(\tau) = \sup_{0\le\epsilon\le 1}\{\ \epsilon\tau-\delta_V(\epsilon)\ \}\ . \tag{5.3.17}$$

Consequently, W is uniformly smooth if and only if V is uniformly convex.

5.3.11 Remark. It is easy to deduce the final statement in Theorem 5.3.10 from the formula (5.3.17): By (5.3.17), for all $\epsilon,\tau\in(0,1)$, $\rho_W(\tau)+\delta_V(\epsilon)\ge\tau\epsilon$, so that

$$\delta_V(\epsilon)\ge\tau\left(\epsilon-\frac{\rho_W(\tau)}{\tau}\right)\ .$$

If $\rho_W(\tau)=o(\tau)$ there is some $\tau>0$ such that $\rho(\tau)/\tau<\epsilon/2$, and then $\delta(\epsilon)>\tau\epsilon/2>0$. Thus, uniform smootness of W implies uniform convexity of V.

Now suppose that W is not uniformly smooth. By (5.3.17), ρ_W is convex and finite on $[0,1]$. It follows that $\lim_{\tau\to 0}\rho(\tau)/\tau =: a$ exists and belongs to $\partial\rho_W(0)$. Hence $\rho_W(\tau)\ge a\tau$ for all $\tau\in[0,1]$. Then for all $r,\tau\in(0,1)$, there exists $\epsilon_{r,\tau}\in[0,1]$ nearly achieving the supremum defining $\rho_W(\tau)$, such that $ra\tau<\epsilon_{r,\tau}\tau-\delta_V(\epsilon_{r,\tau})$. Thus, $\delta_V(\epsilon_{r,\tau})<(\epsilon_{r,\tau}-ra)\tau$. The right side must be positive, so that $\epsilon_{r,\tau}>rA$. Since δ_V is montone and $\epsilon_{r\tau}\le 1$, then $\delta_v(ra)\le\tau$. Thus, $\delta_V(\epsilon)=0$ for all $\epsilon<a$.

Before proving (5.3.17), we give three simple but important applications of Theorem 5.3.10.

5.3.12 THEOREM (Uniqueness and continuity of unit tangent functionals)**.** *If V is a uniformly smooth Banach space, then for each non-zero $v \in V$, there exists a unique unit vector $L_v \in V^*$ such that $L_v(v) = \|v\|$. Moreover, the map $v \mapsto L_v$ is continuous in the norm topologies.*

Proof. The Hahn-Banach Theorem tell us that the linear functional L_v exists; the points to be shown are the uniqueness and the continuity. By the Lindenstrauss-Day Theorem, V^* is uniformly convex. If L and M are two unit vectors in V^* such that $L(v) = M(v) = \|v\|$, then

$$\|v\| = \frac{1}{2}(L+M)(v) \leq \frac{1}{2}\|L+M\|_*\|v\| \leq \|v\| - \|v\|\delta_{V^*}\left(\frac{1}{2}\|L-M\|_*\right) \ .$$

Hence $L = M$. This proves uniqueness.

Now that the map $v \mapsto L_v$ is well-defined, we proceed as in (5.3.8) to obtain

$$\|L_v + L_w\|_*\|v+w\| \geq 2\|w+w\| - \|L_v - L_w\|_*\|v-w\| \ .$$

Dividing thtough by $2\|v+w\|$, rearranging terms, and using $\|v+w\| \geq 2\|v\| - \|v-w\|$, all very much as in the proof of Theorem!5.3.5, we obtain

$$\frac{\|v-w\|}{2\|v\| - \|v-w\|} \geq \left(\left\|\frac{L_v - L_w}{2}\right\|\right)^{-1} \delta_{V^*}\left(\left\|\frac{L_v - L_w}{2}\right\|\right) \geq \delta_{V^*}\left(\left\|\frac{L_v - L_w}{2}\right\|\right) \ ,$$

and this proves the continuity.

□

5.3.13 THEOREM (Differentiability of the Norm)**.** *Let V be a uniformly smooth Banach space. Then the norm on V is continuously Frechét differentiable at all $v \neq 0$ in V, and the derivative is given by $\Re \circ L_v$, where L_v is the unique unit vector in V^* with $L_v(v) = \|v\|$*

Proof. Since V^* is uniformly convex, for each $u \in V$, there exists a unique unit vector $L_u \in V^*$ so that $L_u(u) = \|u\|$. Hence,

$$\|v+w\| = f_{v+w}(v+w) = \Re\left(f_{v+w}(v+w)\right) = \Re\left(L_{v+w}(v)\right) + \Re\left(L_{v+w}(w)\right) \leq \|v\| + \Re\left(L_{v+w}(w)\right) \ .$$

On the other hand,

$$\|v+w\| \geq \Re\left(L_v(v+w)\right) = \Re\left(L_v(v)\right) + \Re\left(L_v(w)\right) = \|v\| + \Re\left(L_v(w)\right) \ .$$

Altogether,

$$0 \leq \|v+w\| - \|v\| - \Re\left(L_v(w)\right) \leq \Re\left(L_{v+w}(w)\right) - \Re\left(L_v(w)\right) \leq \|L_{v+w} - L_v\|_*\|w\| \ .$$

Hence $|\|v+w\| - \|v\| - \Re\left(L_v(w)\right)| \leq \|L_{v+w} - L_v\|_*\|w\| = o(\|w\|)$ by Theorem 5.3.12. □

5.3.14 EXAMPLE. *Let $(\Omega, \mathcal{M}, \mu)$ be a measure space, and for $1 < p < \infty$, let $L^p = L^p(\Omega, \mathcal{M}, \mu)$. We have seen that for non-zero $f \in L^p$, the unique unit vector u in $L^{p/(p-1)}$ such that $\int_\Omega uf\mathrm{d}\mu = \|f\|_p$ is given by*

$$u = \|f\|_p^{1-p}|f|^{p-1}\overline{\mathrm{sgn}(f)} \ .$$

By Theorem 5.3.13, the map $f \mapsto \|f\|_p^{1-p}|f|^{p-1}\overline{\mathrm{sgn}(f)}$ is continuous from $L^p\backslash\{0\}$ to $L^{p/(p-1)}$.

5.3.15 THEOREM (Millman). *A uniformly convex Banach space is reflexive.*

Proof. By the Lindenstrauss–Day Theorem, V^{**} is uniformly convex. Fix any unit vector $\phi \in V^{**}$, For each $n \in \mathbb{N}$, pick a unit vector $L_n \in V^*$ so that $\phi(L_n) > 1 - 1/n$. Since V is uniformly convex, there is a unit vector $v_n \in V$ with $L_n(v_n) = 1$. Let ϕ_{v_n} be the image of v_n in V^{**} under the cannonical embedding of V into V^{**}. Since $\phi_{v_n}(L_n) = L_n(v_n) = 1$ for all n,

$$\left\|\frac{\phi + \phi_{v_n}}{2}\right\|_{**} \geq \left(\frac{\phi + \phi_{v_n}}{2}\right)(L_n) \geq 1 - \frac{1}{2n} \ .$$

By Lemma 5.3.2, $\lim_{n\to\infty} \|\phi - \phi_{v_n}\| = 0$. Then image of V under its canonical embedding into V^{**} is claosed, and hence ϕ belongs to this image. Since ϕ is an arbitrary unit vector in V^{**}, the canonical embedding is surjective. □

Proof of Theorem 5.3.10. We will use f and g to denote elements of W, and u and v to denote elements of V. We will leave subscripts off the norms as this convention makes it clear which norm is intended.

On account of the remark following Theorem 5.3.17, it remains to prove (5.3.17). We first show that $\rho_W(\tau) + \delta_V(\epsilon) \geq \tau\epsilon$ for all $\tau, \epsilon \in (0, 1)$. Fix any $\tau, \epsilon \in (0, 1)$. Take any u and v in V with $\|u\| = \|v\| = 1$ and $\|u - v\| \geq 2\epsilon$.

Since V and W are a dual pair, for any $\eta > 0$, there are unit vectors f and g in W with

$$\langle f, (u + v)/2\rangle \geq \left\|\frac{u+v}{2}\right\| - \eta \qquad \text{and} \qquad \langle f, (u - v)/2\rangle \geq \left\|\frac{u-v}{2}\right\| - \eta \ .$$

Then

$$\begin{aligned}
\rho_W(\tau) &\geq \left\|\frac{f+\tau g}{2}\right\| + \left\|\frac{f-\tau g}{2}\right\| - 1 \\
&\geq \langle (f+\tau g)/2, u\rangle + \langle (f-\tau g)/2, v\rangle - 1 \\
&= \langle f, (u+v)/2\rangle + \tau\langle g, (u-v)/2\rangle - 1 \\
&\geq \left\|\frac{u+v}{2}\right\| + \tau\left\|\frac{u-v}{2}\right\| - 1 - 2\eta \geq \left\|\frac{u+v}{2}\right\| + \tau\epsilon - 1 - 2\eta
\end{aligned} \tag{5.3.18}$$

Hence $\rho_W(\tau) + \left(1 - \left\|\frac{u+v}{2}\right\|\right) \geq \tau\epsilon - 2\eta$. By the definition of δ_V, and the fact that $\eta > 0$ is arbitrary, this proves that $\rho_W(\tau) + \delta_V(\epsilon) \geq \tau\epsilon$ for all $\tau, \epsilon \in (0, 1)$.

The second step is to prove an upper bound on ρ_W. To do this, fix any $\tau > 0$ and any unit vectors f and g in W. Fix any $\eta > 0$, and choose unit vectors u_τ and v_τ in V with

$$\langle (f+\tau g), u_\tau\rangle \geq \|f + \tau g\| - \eta \qquad \text{and} \qquad \langle (f-\tau g), v_\tau\rangle \geq \|f - \tau g\| - \eta \ . \tag{5.3.19}$$

Then

$$\begin{aligned}
\left\|\frac{f+\tau g}{2}\right\| + \left\|\frac{f-\tau g}{2}\right\| &\leq \frac{\langle (f+\tau g), u_\tau\rangle}{2} + \frac{\langle (f-\tau g), v_\tau\rangle}{2} + \eta \\
&= \frac{\langle f, u_\tau + v_\tau\rangle}{2} + \tau\frac{\langle g, u_\tau - v_\tau\rangle}{2} + \eta \leq \left\|\frac{u_\tau + v_\tau}{2}\right\| + \tau\left\|\frac{u_\tau - v_\tau}{2}\right\| + \eta
\end{aligned}$$

Now define $\epsilon_\tau := \left\|\frac{u_\tau - v_\tau}{2}\right\|$ so that $0 \leq \epsilon_\tau \leq 1$ and $\left\|\frac{u_\tau + v_\tau}{2}\right\| \leq 1 - \delta_V(\epsilon_\tau)$. Therefore,

$$\left(\left\|\frac{f+\tau g}{2}\right\| + \left\|\frac{f-\tau g}{2}\right\| - 1\right) \leq \tau\epsilon_\tau - \delta_V(\epsilon_\tau) + \eta \ . \tag{5.3.20}$$

By the definition of ρ_W, $\rho_W(\tau) \le \tau\epsilon_\tau - \delta_V(\epsilon_\tau) + \eta \le \sup_{0\le\epsilon\le1} \{\ \epsilon\tau - \delta_V(\epsilon)\ \} + \eta$. Since $\eta > 0$ is arbitrary, this proves that $\rho_W(\tau) \le \sup_{0\le\epsilon\le1}\{\ \epsilon\tau - \delta_V(\epsilon)\ \}$, and together with the lower bound (5.3.20), this proves (5.3.17). □

5.4 Uniform convexity and smoothness in L^p spaces

5.4.1 Uniform convexity in L^p, $1 < p < \infty$

In any Hilbert space, in particular $L^2(\Omega, \mathcal{M}, \mu)$, we have the parallelogram identity:

$$\left\|\frac{f+g}{2}\right\|_2^2 + \left\|\frac{f-g}{2}\right\|_2^2 = \frac{\|f\|_2^2 + \|g\|_2^2}{2}\ .$$

If f and g are unit vectors, this yields

$$\left\|\frac{f+g}{2}\right\|_2 \le \left(1 - \left\|\frac{f-g}{2}\right\|_2^2\right)^{1/2} \le 1 - \frac{1}{2}\left\|\frac{f-g}{2}\right\|_2^2\ ,$$

and hence $\delta_{L^2}(\epsilon) \ge \dfrac{1}{2}\epsilon^2$.

There is a close analog of the paralleleogram law in $L^p(\Omega, \mathcal{M}, \mu)$, $p > 2$: Recall that for counting measure, $\|f\|_p \ge \|f\|_q$ for $p < q$, while for any probability measure, $\|f\|_p \le \|f\|_q$ for $p < q$.

Therefore, for all $a, b > 0$, when $p > 2$,

$$\left(\left|\frac{a+b}{2}\right|^p + \left|\frac{a-b}{2}\right|^p\right)^{1/p} \le \left(\left|\frac{a+b}{2}\right|^2 + \left|\frac{a-b}{2}\right|^2\right)^{1/2} = \left(\frac{a^2+b^2}{2}\right)^{1/2} \le \left(\frac{a^p+b^p}{2}\right)^{1/p}\ .$$

Next, for all $z, w \in \mathbb{C}$, and all $p \ge 2$,

$$|z+w|^p + |z-w|^p \le ||z| + |w||^p + ||z| - |w||^p\ .$$

Combining, we obtain, $\left|\dfrac{f(x)+g(x)}{2}\right|^p + \left|\dfrac{f(x)-g(x)}{2}\right|^p \le \dfrac{|f(x)|^p + |g(x)|^p}{2}$. Integrating in x yields *Clarkson's inequality*:

$$\left\|\frac{f+g}{2}\right\|_p^p + \left\|\frac{f-g}{2}\right\|_p^p \le \frac{\|f\|_p^p + \|g\|_p^p}{2}\ .$$

If f and g are unit vectors, this becomes

$$\left\|\frac{f+g}{2}\right\|_p \le \left(1 - \left\|\frac{f-g}{2}\right\|_p^p\right)^{1/p} \le 1 - \frac{1}{p}\left\|\frac{f-g}{2}\right\|_p^p\ .$$

Thus we see that

$$\delta_{L^p}(\epsilon) \ge \frac{1}{p}\epsilon^p\ .$$

Uniform convexity for $1 < p < 2$ is more subtle, and the result is somewhat surprising: It turns out that for $1 < p \le 2$,

$$\delta_{L^p}(\epsilon) \ge \frac{p-1}{2}\epsilon^2\ .$$

Notice that the exponent is 2, as in the Hilbert space case. However, as p decreases towards 1, the constant $(p-1)/2$ decreases to zero. Both the exponent 2, and the constant $(p-1)/2$ are best possible, and both have significant implications. The result is can be obtained from a result of Hanner, who *exactly computed* $\delta_{L^p}(\epsilon)$ for all $1 < p < \infty$. The final remark in his 1955 paper is (in slightly different notation) that

$$\delta_{L^p}(\epsilon) = \frac{p-1}{2}\epsilon^2 + \mathcal{O}(\epsilon^3) \ , \tag{5.4.1}$$

which certinaly shows that $\delta_{L^p}(\epsilon)$ is bounded below by some multiple of ϵ^2. The fact that the remainder term is positive and may be droped, yielding the asserted lower bound, may be folklore, but appears in work by Ball and Pisier in the 1990's. They used Hanner's exact computationof δ_{L^p} and controlled the sign of the remainder term in (5.4.1). However, the sharp bound may be proved directly, as we now explain.

Let f and g be simple functions of the form

$$f(x) = \sum_{j=1}^n z_j 1_{A_j}(x) \qquad \text{and} \qquad g(x) = \sum_{j=1}^n w_j 1_{A_j}(x) \ ,$$

where for each j, $z_j w_j^*$ is not real. This guarantees that $z_j + tw_j \neq 0$ for any real t, and thus for all $x \in \cup_{j=1}^n A_j$, and all $t \in \mathbb{R}$, $f(x) + tg(x) \neq 0$. Define

$$Y(t) = \|f + tg\|_p^p \qquad \text{and} \qquad q = \frac{p}{2} \ ,$$

so that $\|f + tg\|_p^2 = Y^{1/q}(t)$. Differentiating twice,

$$\begin{aligned} \frac{\mathrm{d}^2}{\mathrm{d}t^2}\|f+tg\|_p^2 &= \frac{1}{q}\left(\frac{1}{q} - 1\right) Y^{1/q-2}(Y')^2 + \frac{1}{q}Y^{1/q-1}Y'' \\ &\geq \frac{1}{q}Y^{1/q-1}Y'' \end{aligned}$$

A simple calculation yields $Y''(t) \geq p(p-1)\int |f + tg|^{2q-2}|g|^2\mathrm{d}\mu$. To this we apply the *reverse Hölder inequality*, which says that for $0 < r < 1$ and $s = r/(r-1)$, whenever $a_j \geq 0$ for $j = 1, \ldots, n$, and $b_j > 0$ for $j = 1, \ldots, n$,

$$\sum_{j=1}^n a_j b_j \geq \left(\sum_{j=1}^n a_j^r\right)^{1/r} \left(\sum_{j=1}^n b_j^s\right)^{1/s} .$$

The result is that, for all t, $\dfrac{\mathrm{d}^2}{\mathrm{d}t^2}\|f + tg\|_p^2 \geq 2(p-1)\|g\|_p^2$. Let $\psi''(t) \geq 2c$ for all t, and define

$$\varphi(t) := \psi(t) + ct(1-t) \ .$$

Then φ is convex, and thus

$$\varphi(1/2) \leq \frac{\varphi(0) + \varphi(1)}{2} \ , \qquad \text{that is,} \qquad \psi(1/2) + \frac{c}{4} \leq \frac{\psi(0) + \psi(1)}{2} \ .$$

We conclude that with f and g as above,

$$\left\|f + \frac{1}{2}g\right\|_p^2 + \frac{p-1}{4}\|g\|_p^2 \leq \frac{\|f\|_p^2 + \|f+g\|_p^2}{2} \ .$$

The simple function approximation is now easily removed.

Now let u and v be vectors in L^p space, $1 < p \le 2$, and let $f = u$ and $g = v - u$. Then

$$\left\|\frac{u+v}{2}\right\|_p^2 + (p-1)\left\|\frac{u-v}{2}\right\|_p^2 \le \frac{\|u\|_p^2 + \|v\|_p^2}{2} \ .$$

If u and v are unit vectors, the right hand side is 1, and this implies

$$\left\|\frac{u+v}{2}\right\|_p \le 1 - \frac{p-1}{2}\left\|\frac{u-v}{2}\right\|_p^2 \ ,$$

which proves that

$$\delta_{L^p}(\epsilon) \ge \frac{p-1}{2}\epsilon^2 \ .$$

We now have the following result:

5.4.1 THEOREM (Uniform convexity of L^p, $1 < p < \infty$)**.** *For any measure space $(M, \mathcal{M}, \mu)$ and any $L^p(M, \mathcal{M}, \mu)$ is uniformly convex. For $1 < p < 2$, one has the bound*

$$\delta_{L^p}(\epsilon) \ \ge \frac{p-1}{2}\epsilon^2 \tag{5.4.2}$$

while for $2 < p < \infty$, one has the bound

$$\delta_{L^p}(\epsilon) \ \ge \frac{1}{p}\epsilon^p \tag{5.4.3}$$

Proof. In the discussion just above, we have proved the bounds on uniform convexity, and the left halves of (5.4.2) and (5.4.3). □

5.4.2 COROLLARY. *For $1 < p < \infty$, L^p is reflexive.*

Proof. This is an immediate consequence of the uniform convexity of L^p, $1 < p < \infty$ and Millman's Theorem. □

The main result of the next section given an even stronger result: It identifies the dual of L^p, $1 < p < \infty$ with L^q where $q = p/(p-1)$.

5.4.2 The Riesz Representation Theorem in L^p, $1 < p < \infty$

.

We are now give two proofs of the Riesz Representation Theorem for L^p, $1 < p < \infty$.

5.4.3 THEOREM (Riesz Representation Theorem for L^p, $1 < p < \infty$)**.** *Let $(\Omega, \mathcal{M}, \mu)$ be any measure space. Let $1 \le p \le \infty$, and let $q = p/(1-p)$. Then the map from L^q into $(L^p)^*$ given by $g \mapsto \varphi_g$ where*

$$\varphi_g(f) = \int_\Omega fg\mathrm{d}\mu \ , \qquad f \in L^p \ ,$$

is an isometry from L^q into $(L^p)^*$.

We give two proofs. As a preface to both of them, note that by Theroem 5.2.4, or Hölder's inequality with the cases of equality, that for every $g \in L^q$, $\|\varphi_g\|_{(L^p)^*} = \|g\|_q$, and hence $g \mapsto \varphi_g$ is an isometric map *into* $(L^p)^*$. It only remains to be shown that this map is *onto* $(L^p)^*$.

First Proof of Theorem 5.4.3. Let $1 < p < \infty$. Let V be the range of the mapping $g \mapsto \varphi_g$ in $(L^p)^*$. Since the map is an isometry, and since L^q is complete, V is a closed subspace of $(L^p)^*$. If V is a proper subspace of $(L^p)^*$, there exists a non-zero $\varphi \in (L^p)^* \backslash V$ and then by the Hahn-Banach Theorem, there is an $L \in (L^p)^{**}$ such that

$$L(\varphi) = \|\varphi\|_{(L^p)^*} \neq 0 \ , \tag{5.4.4}$$

and $L(\varphi_g) = 0$ for all $g \in L^q$.

However, by Millman's Theorem, since L^p is uniformly convex, it is reflexive, and so there exists an $f \in L^p$ so that

$$L(\psi) = \psi(f) \qquad \text{for all} \quad \psi \in (L^p)^* \ .$$

Therefore, for all $g \in L^q$, since $\varphi_g \in V$, $0 = L(\varphi_g) = \varphi_g(f) = \int_\Omega gf\mathrm{d}\mu$. But since

$$\|f\|_p = \sup\left\{ \int_\Omega gf\mathrm{d}\mu \ : \ \|g\|_q = 1 \right\} \ ,$$

it would follow that $\|f\|_p = 0$, and hence $L = 0$. This is contradicts (5.4.4), and hence V is not a proper subspace of $(L^p)^*$. □

Second Proof of Theorem 5.4.3. Let $1 < p < \infty$. Since L^p is uniformly convex, for each $\varphi \in (L^p)^*$, there exists a unique $f_\varphi \in L^p$ with $f_\varphi \in L^p$ and $\varphi(f_\varphi) = \|\varphi\|_{(L^p)^*}$. Then, for any $g \in L^p$, the function $t \mapsto \varphi\left(\dfrac{f_\varphi + tg}{\|f_\varphi + tg\|_p}\right)$ has a maximum at $t = 0$.

If we assume for the moment that $t \mapsto \|f_\varphi + tg\|_p$ is differentiable at $t = 0$, then

$$0 = \frac{\mathrm{d}}{\mathrm{d}t}\varphi\left(\frac{f_\varphi + tg}{\|f_\varphi + tg\|_p}\right)\Bigg|_{t=0} = \Re\varphi(g) - \|\varphi\|_{(L^p)^*}\frac{\mathrm{d}}{\mathrm{d}t}\|f_\varphi + tg\|_p\Bigg|_{t=0} \ .$$

By the convexity of $x \mapsto |x|^p$, for all $0 < t < 1$ and all $x \in M$,

$$|f_\varphi(x)|^p - |f_\varphi(x) - g(x)|^p \leq \frac{|f_\varphi(x) + tg(x)|^p - |f_\varphi(x)|^p}{t} \leq |f_\varphi(x) + g(x)|^p - |f_\varphi(x)|^p \ .$$

Then, since $|f_\varphi - g|^p$, $|f_\varphi - g|^p$ and $|f_\varphi|^p$ are all integrable, The Dominated Convergence Theorem yields us

$$\lim_{t\to 0}\int_\Omega \frac{|f_\varphi(x) + tg(x)|^p - |f_\varphi(x)|^p}{t}\mathrm{d}\mu = \int_\Omega \lim_{t\to 0}\frac{|f_\varphi(x) + tg(x)|^p - |f_\varphi(x)|^p}{t}\mathrm{d}\mu \ .$$

Now one easily computes that for all x,

$$\lim_{t\to 0}\frac{|f_\varphi(x) + tg(x)|^p - |f_\varphi(x)|^p}{t} = \Re|f_\varphi|^{p-2}\overline{f_\varphi}(x)g(x) \ .$$

Thus, for all $g \in L^q$, we have

$$\Re\varphi(g) = \|\varphi\|_{(L^p)^*}\int_\Omega \Re|f_\varphi|^{p-2}\overline{f_\varphi}(x)g(x)\mathrm{d}\mu \ .$$

Substituting g by ig, we obtain the same result for the imaginary part, and hence

$$\varphi(g) = \|\varphi\|_{(L^p)^*}\int_\Omega |f_\varphi|^{p-2}\overline{f_\varphi}(x)g(x)\mathrm{d}\mu \ .$$

It is now easily checked that $|f_\varphi|^{p-2}\overline{f_\varphi}$ is a unit vector in L^q, and hence φ is in the range of our isometry into $(L^p)^*$. But since φ is an arbitrary element of $(L^p)^*$, we see that our isometry is onto $(L^p)^*$. □

5.4.4 COROLLARY (Riesz Representation Theorem for L^1). *Let $(\Omega, \mathcal{M}, \mu)$ be a σ-finite measure space. The map $g \mapsto \varphi_g \in (L^1)^*$ from L^∞ to $(L^1)^*$ given by $\phi_g(f) = \int_\Omega gf\mathrm{d}\mu$ is an isometry onto from L^∞ $(L^1)^*$.*

Proof. Suppose first that $\mu(\Omega) = 1$. By Jensen's inequality, for all measurable f and all $1 \le p < q$,

$$\left(\int_\Omega |f|^p \mathrm{d}\mu\right)^{q/p} \le \int_\Omega |f|^q \mathrm{d}\mu \ .$$

Therefore, $L^q \subset L^p \subset L^1$, and for any $f \in L^q$, $\|f\|_1 \le \|f\|_p \le \|f\|_q$.

Let L be any bounded linear functional on L^1. Let $f \in L^q$. By the inclusion proved above, $|L(f)| \le \|L\|_*\|f\|_1 \le \|L\|_*\|f\|_q$. By Theorem 5.4.3, there is a unique $g_q \in L^{q/(q-1)}$ such that $\|g_q\|_{q/(q-1)} \le \|L\|_*$, and such that $L(f) = \int_\Omega fg_q\mathrm{d}\mu$ for all $f \in L^q$. By the same reasoning, for all $1 < p < q$, there is a unique $g_p \in L^{p/(p-1)}$ such that $\|g_p\|_{p/(p-1)} \le \|L\|_*$, and such that $L(f) = \int_\Omega fg_p\mathrm{d}\mu$ for all $f \in L^p$. By the inclusion of L^q in L^p and the uniqueness, $g_q = g_p$, and hence g_p is independent of $p > 1$. Therefore, we drop the subscript and denote this function by g. We have that $\|g\|_{p/(p-1)} \le \|L\|_*$ for all p, and since $\lim_{p\downarrow 1} \|g\|_{p/(p-1)} = \|g\|_\infty$, $\|g\|_\infty \le \|L\|_\infty$.

For all $f \in L^1$ and $n \in \mathbb{N}$, define $f_n(x) = f(x)$ if $|f(x)| < n$ and $f(x) = 0$ otherwise. Then by the Lebesgue Dominated Convergence Theorem, $\lim_{n\to\infty} \|f_n - f\|_1 = 0$, and for each n, $fn \in L^p$ for all $1 \le p \le \infty$. Hence

$$L(f) = \lim_{n\to\infty} L(f_n) = \lim_{n\to\infty} \int_\Omega f_n g\mathrm{d}\mu = \int_\Omega fg\mathrm{d}\mu$$

by the Lebesgue Dominated Convergence Theorem once more. Thus, the map $g \mapsto \varphi_g$, which we know to be an isometry from L^∞ to $(L^1)^*$ is surjective.

It is a simple matter to extend this proof to the case in which μ is finite, and then to the case in which μ is *sigma*-finite noting that if $E \subset \Omega$ is any subset of Ω, the restriction of L to to the subspace of L^1 consisting of functions that vanish outside E has a norm that is not greater than $\|L\|_*$ independent of the choice of E. The details are left to the reader. □

5.4.3 Uniform smoothness in L^p, $1 < p < \infty$.

As a consequence of Theorem 5.4.1 the Lindenstrauss-Day Theorem, we obtain the following result:

5.4.5 THEOREM (Uniform smoothness of L^p, $1 < p < \infty$). *For any measure space $(M, \mathcal{M}, \mu)$ and any $L^p(M, \mathcal{M}, \mu)$ is uniformly smooth. For $1 < p < 2$,one has the bound*

$$\rho_{L^p}(\tau) \le \frac{1}{2(p-1)}\tau^2 \ , \tag{5.4.5}$$

while for $2 < p < \infty$ and $q = p/(p-1)$, one has the bound

$$\rho_{L^p}(\tau) \le \frac{1}{q}\tau^q \ . \tag{5.4.6}$$

Proof. The uniform smoothness follows directly from Theorems 5.3.10 and 5.4.1. To obtain (5.4.5) use (5.3.17) to deduce

$$\rho_{L^p}(\tau) \le \sup_{0\le\epsilon\le 1} \{\epsilon\tau - \delta_{L^q}(\epsilon)\ \} \ .$$

Then by (5.4.2) and a simple calculation, one obtains and (5.4.5). The proof of (5.4.6) is similar. □

This result has useful implications the theory of partial differential equations. Very often the solution of a partial differential equation give rise to a curve $f(t, cdot)$ where for each $t > 0$, $f(t, cdot) \in L^p$ for some L^p space – ona all of $\mathbb{R}$, in some bounded set Ω, or soemthing else. Let is simple write $f(t)$ to denote the t-dependent element of L^p. That is $t \mapsto f(t)$ is a *curve* in L^p.

5.4.6 DEFINITION. Let $(\Omega, \mathcal{M}, \mu)$ be anty meaasure space. Let $(a,b) \subset \mathbb{R}$, $1 \leq p \leq \infty$, and let $t \mapsto f(t)$ be a function from (a,b) to $L^p := L^p(\Omega, \mathcal{M}, \mu)$. The map $t \mapsto f(t)$ is *strongly differentiable* at $t_0 \in (a,b)$ in case for some $v(t_0) \in L^p$,

$$\lim_{h\to 0} \frac{1}{|h|}\|f(t_0+h) - f(t_0) - v(t_0)h\|_p = 0\ .$$

In this case we define $v(t_0)$ to be the *derivative* of $t \mapsto f(t)$ at $t = t_0$, ad we write $\dfrac{\mathrm{d}}{\mathrm{d}t}f(t_0) = v(t_0)$. The map $t \mapsto f(t)$ is *strongly continously differentiable on* (a,b) in case it is differentiable at each $t \in (a,b)$, and $t \mapsto v(t)$ is norm-continuous in L^p.

5.4.7 THEOREM (The chain-rule in L^p)**.** *Let $(\Omega, \mathcal{M}, \mu)$ be anty meaasure space. Let $(a,b) \subset \mathbb{R}$, $1 < p < \infty$, and let $t \mapsto f(t)$ be a function from (a,b) to $L^p := L^p(\Omega, \mathcal{M}, \mu)$. If $t \mapsto f(t)$ is strongly differentialble at each $t \in (a,b)$, then $t \mapsto \|f(t)\|_p$ is differetniable at each $t \in (a,b)$. If $t \mapsto f(t)$ is strongly continuously differentialble on (a,b), then $t \mapsto \|f(t)\|_p$ is continuosly differetniable on (a,b).*

Proof. For any $t_0 \in (a,b)$ and $h \neq 0$ such that $t_0 + h \in (a,b)$, the Frechét differentiability of the L^p norm that follows from the uniform smoothness of L^p means that for all $\epsilon > 0$, there is a $\delta_\epsilon > 0$ such that

$$\begin{aligned}&\|f(t_0+h) - f(t_0)\|_p < \delta_\epsilon \quad \Rightarrow \\ &\qquad\qquad \Big|\|f(t_0+h)\|_p - \|f(t_0)\|_p - \langle D_p(f(t_0)), f(t_0+h) - f(t_0)\rangle\Big| < \epsilon\|f(t_0+h) - f(t_0)\|_p\ .\end{aligned}$$

Then since $t \mapsto f(t)$ is strongly differentiable, for all $\epsilon > 0$, there is an $\eta_\epsilon > 0$ such that

$$|h| < \eta_\epsilon \quad \Rightarrow \qquad \frac{1}{|h|}\|f(t_0+h) - f(t_0) - hv(t_0)\| < \epsilon$$

Then for all h with $|h| < \eta_\epsilon$, $\|f(t_0+h) - f(t_0)\| \leq |h|(\|v(t_0)\| + \epsilon)$. Decreasing η_ϵ as necessary, we can ensure that the right hand side is less than δ_ϵ. Then for all h with $|h| < \eta_\epsilon$,

$$\Big|\|f(t_0+h)\|_p - \|f(t_0)\|_p - h\langle D_p(f(t_0)), v(t_0)\rangle\Big| \leq \epsilon|h|(\|v(t_0)\| + \epsilon) + |h|\epsilon\ .$$

Thus,

$$\limsup_{|h|\to 0} \frac{1}{|h|}\Big|\|f(t_0+h)\|_p - \|f(t_0)\|_p - h\langle D_p(f(t_0)), v(t_0)\rangle\Big| \leq \epsilon|(\|v(t_0)\| + \epsilon) + \epsilon\ .$$

Since $\epsilon > 0$ is arbitrary,

$$\lim_{h\to 0} \frac{\|f(t_0+h)\|_p - \|f(t_0)\|_p}{h} = \langle D_p(f(t_0)), v(t_0)\rangle\ .$$

Therefore, if $t \mapsto f(t)$ is strongly differentiable at t_0, then $t \mapsto \|f(t)\|_p$ is differentiable at $t = t_0$.

By Theorem 5.3.13, it follows that since $t \mapsto f(t)$ is continuous into L^p, $t \mapsto D_p(f(t))$ is continuous into $L^{p/(p-1)}$. Then if $t \mapsto v(t)$ is continuous into L^p, $t \mapsto \langle D_p(f(t)), v(t)\rangle$ is continuous into $\mathbb{R}$. $\square$

Chapter 6

Topics in Classical Analysis

6.1 Convolution and related operations

6.1.1 Markov kernels

6.1.1 DEFINITION. Let $(\Omega, \mathcal{M}, \mu)$ be a measure space. A *Markov kernel* on $(\Omega, \mathcal{M}, \mu)$ is a non-negative function K on $(\Omega \times \Omega, \mathcal{M} \otimes \mathcal{M})$ such that for each $x \in \Omega$,

$$\int_\Omega K(x,y)\mathrm{d}\mu(y) = 1 \ . \tag{6.1.1}$$

That is, for each x, $K(x,y)\mathrm{d}\mu(y)$ is a probability measure on $(\Omega, \mathcal{M})$. A Markov kernel K is *doubly stochastic* in case both

$$\int_\Omega K(x,y)\mathrm{d}\mu(y) = 1 \quad \text{and} \quad \int_\Omega K(x,y)\mathrm{d}\mu(x) = 1 \tag{6.1.2}$$

for all x, y. A Markov kernel K is *symmetric* in case $K(x,y) = K(y,x)$ for all $x, y \in \Omega$. Every symmetric Markov kernel is doubly stochastic.

In probability theory, Markov kernels arise in the description of Markov jump process with a *state space* Ω. When a jump from state x occurs, the probability of jumping from x into a measurable set $E \subset \Omega$ is given by $\int_E K(x,y)\mathrm{d}\mu(y)$. Markov kernels play a prominent role in analysis as well; the results in the rest of this section give a first indication of why this is the case.

6.1.2 THEOREM. *Let K be a doubly stochastic Markov kernel on $(\Omega, \mathcal{M}, \mu)$. For each $p \in [1,\infty]$, define a linear operator P_K on $L^p(\Omega, \mathcal{M}, \mu) \cap L^\infty(\Omega, \mathcal{M}, \mu)$ by*

$$P_K f(x) = \int_\Omega K(x,y) f(y)\mathrm{d}\mu(y) \ . \tag{6.1.3}$$

Then $\|P_K f\|_p \leq \|f\|_p$, so that P_K extends by continuity to a contraction on $L^p(\Omega, \mathcal{M}, \mu)$.

Proof. Suppose f is bounded and measurable. Then for each x, f is integrable with respect to $K(x,y)\mathrm{d}\mu(y)$, and

$$|P_K f(x))| = \left|\int_\Omega K(x,y) f(y)\mathrm{d}\mu(y)\right| \leq \int_\Omega K(x,y)\|f\|_\infty \mathrm{d}\mu(y) \leq \|f\|_\infty \ .$$

Now suppose that for $p \in [1,\infty)$, $f \in L^p(\Omega, \mathcal{M}, \mu)$. Since $t \mapsto |t|^p$ is convex on $\mathbb{R}$, Jensen's inequality implies

$$\left|\int_\Omega K(x,y)f(y)\mathrm{d}\mu(y)\right|^p \le \left(\int_\Omega K(x,y)|f(y)|\mathrm{d}\mu(y)\right)^p \le \int_\Omega K(x,y)|f(y)|^p\mathrm{d}y\ .$$

Then by the Fubini-Toninelli Theorem,

$$\|P_K f\|_p^p = \int_\Omega \left|\int_\Omega K(x,y)f(y)\mathrm{d}\mu(y)\right|^p \mathrm{d}\mu(x) \le \int_\Omega \left(\int_\Omega K(x,y)\mathrm{d}\mu(x)\right)|f(y)|^p\mathrm{d}\mu(y) = \|f\|_p^p\ .$$

□

Let K be a symmetric Markov kernel on $(\Omega, \mathcal{M}, \mu)$, and let P_K also denote the corresponding contraction $\mathcal{H} := L^2(\Omega, \mathcal{M}, \mu)$. It is easy to see that P_K is self-adjoint.

Convolution operators provide an important class of examples. Let ρ be a non-negative Borel function on $\mathbb{R}$ such that $\int_\mathbb{R} \rho\mathrm{d}x = 1$. In this case, we say that ρ is a probability density on $(\mathbb{R}, \mathcal{B}, \mathrm{d}x)$. Given such a function ρ, define $K(x,y) = \rho(x-y)$. Then it is evident that $K(x,y)$ is a doubly stochastic Markov kernel. Therefore, for each $p \in [1,\infty]$ we may define an operator T_ρ acting on $L^p(\mathbb{R}, \mathcal{B}, \mathrm{d}x)$ by

$$T_\rho f(x) = \int_\mathbb{R} \rho(x-y)f(y)\mathrm{d}y =: \rho * f(x)\ . \tag{6.1.4}$$

Then $\rho * f$ is called the convolution of ρ and f, and T_ρ is the operator of convolution by ρ. As an immediate consequence of Theorem 6.1.2 we have that for any probability density ρ, and $p \in [1,\infty]$ and and $f \in L^p(\mathbb{R}, \mathcal{B}, \mathrm{d}x)$,

$$\|\rho * f\|_p \le \|f\|_p\ . \tag{6.1.5}$$

If ρ is any probability density, then for all $\lambda > 0$,

$$\rho_\lambda(x) = \lambda^{-1}\rho(x/\lambda) \tag{6.1.6}$$

is also a probability density. If ρ has support in a compact interval $[-L, L]$, ρ_λ has support in a the interval $[-\lambda L, \lambda L]$. Let f be a continuous compactly supported function on $\mathbb{R}$. Then for all $\epsilon > 0$, there is a $\delta > 0$ so that $|x-y| < \delta \Rightarrow |f(x) - f(y)| < \epsilon$. Then for λ such that $\lambda L < \delta$,

$$|\rho_\lambda * f(x) - f(x)| = \left|\int_\mathbb{R} \rho_\lambda(x-y)(f(y) - f(x))\mathrm{d}y\right| \le \int_\mathbb{R} \rho_\lambda(x-y)|f(y) - f(x)|\mathrm{d}y \le \epsilon$$

since $\rho_\lambda(x-y) = 0$ if $|y-x| \ge \delta$. Also note that if f is supported in the interval $[-R, R]$, then $\rho_\lambda * f$ is supported in the interval $[-\lambda L - R, \lambda L + R]$ since $\rho_\lambda(y-x)f(y) = 0$ if $|x| > R + \lambda L$. It follows that $|\rho_\lambda * f - f|$ is bounded by ϵ and is supported in $[-R-1, R+1]$ for all λ such that $\lambda L < \min\{\delta, 1\}$. For such λ, $\|\rho_\lambda f - f\|_p \le \epsilon(2R+2)^{1/p}$. We have proved an important special case of the following theorem:

6.1.3 THEOREM. *For any probability density ρ on $\mathbb{R}$, let ρ_λ be defined for $\lambda > 0$ by (6.1.6). Then for all $p \in [1,\infty)$, and all $f \in L^p(\mathbb{R}, \mathcal{M}, \mathrm{d}x)$,*

$$\lim_{\lambda to 0} \|\rho_\lambda * f - f\|_p = 0\ .$$

Proof. Define $\eta_L = \int_{[-L,L]^c} \rho(x)\mathrm{d}x$, and note that $\lim_{L\to\infty} \eta_L = 0$. For all L large enough that $\eta_L < 1$, Define

$$\rho^{(L)}(x) := (1-\eta_L)^{-1}1_{[-L,L]}(x)\rho(x) \quad \text{and} \quad r^{(L)}(x) := \eta_L^{-1}1_{[-L,L]^c}(x)\rho(x)\ .$$

Then evidently both $\rho^{(L)}$ and $r^{(L)}$ are probability densities $\rho = (1-\eta_L)\rho^{(L)} + \eta_L r^{(L)}$ is a convex combination of them. Hence for all $f \in L^p(\mathbb{R}, \mathcal{M}, \mathrm{d}x)$, Likewise, for all $\lambda > 0$, $\rho_\lambda = (1-\eta_L)\rho_\lambda^{(L)} + \eta_L r_\lambda^{(L)}$, and then for all $p \in [1,\infty)$, and all $f \in L^p(\mathbb{R}, \mathcal{M}, \mathrm{d}x)$,

$$\begin{aligned} \|\rho_\lambda * f - f\|_p &= \|(1-\eta_L)(\rho_\lambda^{(L)} * f - f) + \eta_L(r_\lambda^{(L)} * f - f)\|_p \\ &\leq (1-\eta_L)\|\rho_\lambda^{(L)} * f - f\|_p + \eta_L\|r_\lambda^{(L)} * f - f\|_p \\ &\leq (1-\eta_L)\|\rho_\lambda^{(L)} * f - f\|_p + \eta_L 2\|f\|_p \ , \end{aligned}$$

where we used (6.1.5) in the last step,

Now fix $\epsilon > 0$, and choose L sufficiently large that $\eta_L 2\|f\|_p < \epsilon/3$. Since $C_c(\mathbb{R})$ is dense in $L^p(\mathbb{R}, \mathcal{B}, \mathrm{d}x)$, we may choose $g \in C_c(\mathbb{R})$ such that $\|f - g\|_p < \epsilon/3$. Then using Minkowski's inequality and (6.1.5),

$$\|\rho_\lambda^{(L)} * f - f\|_p \leq \|\rho_\lambda^{(L)} * (f-g)\|_p + \|\rho_\lambda^{(L)} * g - g\|_p + \|f-g\|_p \leq \|\rho_\lambda^{(L)} * g - g\|_p + 2\epsilon/3 \ .$$

Altogether, we have that

$$\|\rho_\lambda * f - f\|_p \leq \|\rho_\lambda^{(L)} * g - g\|_p + \epsilon \ ,$$

and since ρ_L has support in $[-L, L]$ and since $g \in C_c(\mathbb{R})$, by what we have explained just before the statement of the theorem, $\lim_{\lambda\to 0} \|\rho_\lambda^{(L)} * g - g\|_p = 0$. Since $\epsilon > 0$ is arbitrary, the proof is complete. □

6.1.2 The basic facts about convolution

Throughout this section, L^p denotes $L^p(\mathbb{R}, \mathcal{B}, \mathrm{d}x)$. Let $f, g \in L^1 \cap L^\infty$. Then the integral

$$g * f(x) := \int_{\mathbb{R}} g(x-y) f(y) \mathrm{d}y \tag{6.1.7}$$

converges for all x, and we have the point-wise inequality

$$|g * f(x)| \leq |g| * |f|(x) \tag{6.1.8}$$

If $g \neq 0$, so that $\|g\|_1 \neq 0$, $\rho := \|g\|_1^{-1}|g|$ is a probability density. Therefore, by (6.1.5), for all $p \in [1,\infty]$,

$$\|g * f\|_p \leq \||g| * |f|\|_p = \|g\|_1 \|\rho * |f|\|_p \leq \|g\|_1 \|f\|_p \ . \tag{6.1.9}$$

In particular, $g * f \in L^1 \cap L^\infty$. Thus, convolution is a product on $L^1 \cap L^\infty$, making it an algebra. By the Fubini-Toninelli Theorem, and the change of variables $w = y - x$,

$$\int_{\mathbb{R}} f * g \mathrm{d}x = \int_{\mathbb{R}} \left(\int_{\mathbb{R}} g(x-y) f(y) \mathrm{d}x \right) \mathrm{d}y = \left(\int_{\mathbb{R}} g(w) \mathrm{d}w \right) \left(\int_{\mathbb{R}} f(y) \mathrm{d}y \right) \ , \tag{6.1.10}$$

so that when f and g are non-negative, $\|f * g\|_1 = \|f\|_1 \|g\|_1$.

Making the change of variables $y = x - w$,

$$g * f(x) = \int_{\mathbb{R}} g(x-y) f(y) \mathrm{d}y = \int_{\mathbb{R}} f(x-w) g(w) \mathrm{d}w = f * g(x) \tag{6.1.11}$$

Thus the convolution product on $L^1 \cap L^\infty$ is commutative. It is also associative: Let $f, g, h \in L^1 \cap L^\infty$. We have already seen that $f * g \in L^1 \cap L^\infty$. Then making the change of variables $y = u - w$, and applying the Fubini-Toninelli Theorem,

$$\begin{aligned}(f * g) * h(x) &= \int_{\mathbb{R}} \left(\int_{\mathbb{R}} f(x-w-y)g(y)\mathrm{d}y \right) h(w)\mathrm{d}w \\ &= \int_{\mathbb{R}} f(x-u) \left(\int_{\mathbb{R}} g(u-w)h(w)\mathrm{d}w \right) \mathrm{d}u = f * (g * h)(x)\ . \end{aligned} \tag{6.1.12}$$

There is an important complement to (6.1.9): Since $L^1 \cap L^\infty \subset L^p$ for all $p \in [1,\infty]$, for $p \in (1,\infty)$, $g \in L^{p/(p-1)}$ and $f \in L^p$, and then by Hölder's inequality, for all x,

$$|g * f(x)| \leq \int_{\mathbb{R}} |g(x-y)||f(y)|\mathrm{d}y \leq \|g\|_{p/(p-1)}\|f\|_p\ .$$

Therefore,

$$\|g * f\|_\infty \leq \|g\|_{p/(p-1)}\|f\|_p\ . \tag{6.1.13}$$

For all $\lambda > 0$ and all $f \in L^1$, define

$$f_\lambda(x) = \lambda^{-1} f(x/\lambda) \tag{6.1.14}$$

which reduces to (6.1.6) when f is a probability density. A simple change of variables shows that $\|f_\lambda\|_1 = \|f\|_1$ for all $\lambda > 0$. A simple computation shows that for all $f, g \in L^1$,

$$g_\lambda * f_\lambda(x) = \lambda^{-2}\int_{\mathbb{R}} g(x/\lambda - y/\lambda)f(y/\lambda)\mathrm{d}y = \lambda^{-1} f * g(x/\lambda) = (g*f)_\lambda(x)\ . \tag{6.1.15}$$

Another simple computation shows that for $f \in L^1 \cap L^\infty$, $p \in [1,\infty)$ and $\lambda > 0$, with f_λ defined in (6.1.14),

$$\|f_\lambda\|_p = \lambda^{1/p-1}\|f\|_p\ . \tag{6.1.16}$$

Now suppose that for some $p, q, r \in [1,\infty)$, there is a finite constant C such that for all $f, g \in L^1 \cap L^\infty$,

$$\|g * f\|_r \leq C\|g\|_q\|f\|_p\ . \tag{6.1.17}$$

Then replacing f and g by f_λ and g_λ, and using (6.1.15),

$$\|g*f\|_r = \lambda^{1-1/r}\|(g*f)_\lambda\|_r = \lambda^{1-1/r}\|g_\lambda * f_\lambda\|_r \leq C\|g_\lambda\|_q\|f_\lambda\|_p = \lambda^{1/q+1/p-1/r-1}C\|g\|_q\|f\|_p\ .$$

Since the left side is independent of λ, the right had side must be independent of λ also, since if $1/q+1/p-1/r-1 > 0$, taking $\lambda \to 0$, we would conclude that $\|g*f\|_r = 0$, and thus that $g*f = 0$ almost everywhere. If $1/q + 1/p - 1/r - 1 < 0$, taking $\lambda \to \infty$ brings us to the same conclusion. But for non-negative f, g, $\|g * f\|_1 = \|g\|_1\|f\|_1$, and so $g * f = 0$ almost everywhere implies that $f = g = 0$.

Therefore, there is no finite constant C for which the inequality (6.1.17) can hold in general unless

$$\frac{1}{p} + \frac{1}{q} = \frac{1}{r} + 1\ . \tag{6.1.18}$$

It turns out that when (6.1.18) is satisfied, then (6.1.17) is also valid in general, with a constant C no greater than 1. This is Young's inequality for convolution. We have already seen several important cases, namely the case in which $r = p, q = 1$, which is (6.1.9) and the case $r = \infty$,$q = p/(p-1)$, which is (6.1.13).

There are several ways to prove the remaining cases of Young's inequality with the constant $C = 1$. The inequality actually holds in these cases with an optimal constant C that is strictly less than 1, We shall prove this sharp form of Young's inequality, which is due to Beckner and Brascamo and Lieb later in the Chapter. The main tool used in the proof that we give is the heat semigroup, our next topic. For now, we close this section with the following lemma that summarizes our main conclusions.

6.1.4 LEMMA. *The convolution product on $L^1 \cap L^\infty$ is commutative and associative, and the inequalities (6.1.9) and (6.1.13) for all $f, g \in L^1 \cap L^\infty$. Moreover, since $L^1 \cap L^\infty$ is dense in L^r for all $1 < r < \infty$, the map $(g, f) \mapsto g * f$ extends by continuity to a map from $L^1 \times L^p$ to L^p for which (6.1.9) is valid, and to a map from $L^{p/(p-1)} \times L^p$ to L^∞ for which (6.1.13) is valid.*

6.1.3 The Heat semigroup

The Gaussian probability density $\gamma(x)$ associated to the *normal distribution* of probability theory is given by

$$\gamma(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \ .$$

For $t > 0$, define $\gamma_t(x) = (2\pi t)^{-1/2} e^{-x^2/2t}$. This is the Gaussian probability density with mean 0 and variance t. That is,

$$\int_{\mathbb{R}} \gamma_t(x) \mathrm{d}x = 1 \ , \qquad \int_{\mathbb{R}} x\gamma_t(x) \mathrm{d}x = 0 \quad \text{and} \quad \int_{\mathbb{R}} x^2 \gamma_t(x) \mathrm{d}x = t \ . \tag{6.1.19}$$

Note that the definition $\gamma_t(x) = t^{-1/2}\gamma(x/\sqrt{t})$ duffers slightly from (6.1.6) due to the square roots on the right. In the present context, this scaling will turn out to be more natural. The following lemma explains why.

6.1.5 LEMMA. *For all $s, t > 0$,*

$$\gamma_t * \gamma_s = \gamma_{t+s} \ . \tag{6.1.20}$$

Proof. Note that $\gamma_t(x-y)\gamma_s(y) = \dfrac{1}{2\pi\sqrt{st}} e^{-[(x-y)^2/t + y^2/s]/2}$, and

$$\frac{(x-y)^2}{t} + \frac{y^2}{s} = \frac{t+s}{st}\left(y - \frac{s}{t+s}x\right)^2 + \frac{1}{t+s}x^2 \ .$$

$$\int_{\mathbb{R}} \gamma_t(x-y)\gamma_s(y)\mathrm{d}y = \gamma_{s+t}(x)\sqrt{\frac{s+t}{2\pi st}} \int_{\mathbb{R}} e^{-(t+s)(y-sx/(t+s))^2/(2st)}\mathrm{d}y = \gamma_{s+t}(x) \ .$$

□

6.1.6 DEFINITION (Heat semigroup)**.** For each $t > 0$, define an operator P_t on $L^1 \cap L^\infty$ by

$$P_t f = \gamma_t * f \ . \tag{6.1.21}$$

For $t > 0$, define $P_0 = I$. Then the family of operators $\{P_t\}_{t \geq 0}$ constitute the *heat semigroup*. The name will be justified shortly.

6.1.7 THEOREM. *Let $\{P_t\}_{t\geq 0}$ be the heat semigroup on $L^1 \cap L^\infty$. Then for all $s, t \geq 0$,*

$$P_s P_t = P_{s+t} \ . \tag{6.1.22}$$

For all $p \in [1, \infty]$, P_t extends by continuity to an element of $\mathscr{B}(L^p)$, and, for all $f \in L^p$, $\|P_t f\|_p \leq \|f\|_p$. Moreover, for all $p \in [1, \infty)$ and all $f \in L^p$, $\lim_{t\to 0} \|P_t f - f\|_p = 0$. Finally, for all $r > p \in [1, \infty]$, and all $t > 0$, P_t extends by continuity to an element of $\mathscr{B}(L^p, L^r)$ with operator norm

$$\|P_t\|_{p\to r} \leq \|\gamma_t\|_{p/(p-1)}^{1-p/r} \ . \tag{6.1.23}$$

6.1.8 Remark. Since evidently for all $t > 0$, $\gamma_t \in L^1 \cap L^\infty$, $\gamma_t \in L^p$ for all $p \in [1, \infty]$. The precise value of $\|\gamma_t\|_p$ is readily computed for all $p \in [1, \infty]$; see the exercises. The main point that is relevant in the next sections is the for all $t > 0$, the range of P_t acting on L^p lies in $L^p \cap L^\infty$.

Proof. By Lemma 6.1.4 and the definition of the heat semigroup, and then Lemma 6.1.5, for all $f \in L^1 \cap L^\infty$,

$$P_s P_t f = \gamma_s * (\gamma_t * f) = (\gamma_s * \gamma_t) * f = \gamma_{s+t} * f = P_{s+t} f \ .$$

This proves (6.1.22). The inequality $\|P_t f\|_p \leq \|f\|_p$ is a direct consequence of (6.1.5) for $f \in L^1 \cap L^\infty$, and then evidently P_t extends by continuity to an element of $\mathscr{B}(L^p)$. The statement concerning $\lim_{t\to 0} \|P_t f - f\|_p$ is a direct consequence of Theorem 6.1.3. finally, by (6.1.13),

$$\|P_t f\|_\infty \leq \|\gamma_t\|_{p/(p-1)} \|f\|_p \ ,$$

which proves (6.1.23) for $r = \infty$. For $r \in (p, \infty)$ $|P_t f(x)|^r \leq |P_t f(x)|^p \|P_t f\|_\infty^{r-p}$, and hence

$$\|P_t f\|_r^r \leq \|P_t f\|_p^p \|P_t f\|_\infty^{r-p} \leq \|f\|_p^p (\|\gamma_t\|_{p/(p-1)} \|f\|_p)^{r-p} \ ,$$

which proves (6.1.23) in general. □

The case $p = 2$ will be particularly important in what follows. Since for all $f, g \in L^2$, $t > 0$,

$$\langle f, P_t g\rangle_{L^2} = \int_{\mathbb{R}} \overline{f} P_t g(x) \mathrm{d}x = \int_{\mathbb{R}} \int_{\mathbb{R}} \overline{f(x)} \gamma_t(x - y) g(y) \mathrm{d}x \mathrm{d}y = \langle P_t f, g\rangle_{L^2} \ ,$$

it follows that P_t is self-adjoint on L^2.

Theorem 6.1.7 says that for each $p \in [1, \infty]$ and each $f \in L^p$, $P_t f$ converges to f in the L^p norm as $t \to 0$. as $t \to 0$. Later in this chapter we will have the means to easily generate examples showing that for all $t, \epsilon > 0$ and all $1 \leq p < \infty$, there exists $h \in L^p$, $\|h\|_p = 1$ and $\|P_t h - h\|_p > 1 - \epsilon$. This brings us to the notion of the *strong operator topology*

6.1.9 DEFINITION (The strong operator topology)**.** Let $(X, \|\cdot\|_X)$ and $(Y, \|\cdot\|_Y)$ be Banach spaces. For each $x \in X$, define the map ψ_x on $\mathscr{B}(X, Y)$ by $\psi_x(T) = \|Tx\|_Y$

The *strong operator topology* on $\mathscr{B}(X, Y)$ is the weakest topology on $\mathscr{B}(X, Y)$ making each of the map ψ_x continuous.

By what we have seen in our study of topological vector spaces, this topology makes $\mathscr{B}(X, Y)$ a topological vector space, an a neighborhood base at the origin is given by the sets

$$V_{x_1, \dots, x_n, \epsilon} = \bigcap_{j=1}^{n} \{T \in \mathscr{B}(X, Y) \ : \ \|Tx_j\| < \epsilon\}$$

for $\epsilon > 0$ and finite sets $\{x_1, \dots, x_n\} \subset X$.

6.1.10 DEFINITION (Strongly continuous semigroup)**.** Let $(X, \|\cdot\|)$ be a Banach space. A *strongly continuous semigroup* on X is a set $\{P_f \ : \ t \geq 0\} \subset \mathscr{B}(X)$ such that $P_0 = I$, $t \mapsto P_t$ is continuous with respect to the strong operator topology on $\mathscr{B}(X)$, and such that for all $s, t \geq 0$, $P_t P_s = P_{s+t}$.

6.1.11 EXAMPLE (The heat semigroup on L^p)**.** *For each $1 \leq p \leq \infty$, $t > 0$, let $P_t \in \mathscr{B}(L^p)$ be defined as in (6.1.23) and extended by continuity. Then by Theorem 6.1.7, $t \mapsto P_t$ is a strongly continuous semigroup on L^p. It is called the* heat semigroup on *L^p.*

The heat semigroup gets its name from the fact that, considered as a function of $t > 0$ and $x \in \mathbb{R}$, $\gamma_t(x)$ satisfies the *heat equation*:

$$\frac{\partial}{\partial t}\gamma_t(x) = \frac{1}{2}\frac{\partial^2}{\partial x^2}\gamma_t(x) \ . \tag{6.1.24}$$

Indeed, a calculation gives $\frac{\partial}{\partial t}\gamma_t(x) = \frac{x^2}{2t^2}\gamma_t(x)$, which, as a function of x, belongs to L^q for all $1 \le q \le 1$, Moreover, for all $0 < s, t$, $s \neq t$,

$$\gamma_t(x) - \gamma_s(x) = \int_s^t \frac{x^2}{2r^2}\gamma_r(x)\mathrm{d}r = (t-s)\frac{x^2}{2t^2}\gamma_t(x) + \int_s^t \left[\frac{x^2}{2r^2}\gamma_r(x) - \frac{x^2}{2t^2}\gamma_t(x)\right]\mathrm{d}r$$

Simple estimates now yield, for all $1 \le q \le \infty$,

$$\left\|\gamma_t - \gamma_s - (t-s)\tfrac{\partial}{\partial t}\gamma_t\right\|_q = o(|t-s|) \ . \tag{6.1.25}$$

Using (6.1.25) for $q = 1$, we have that for all $f \in L^p$ if we define $f(t,x) = P_t f(x)$,

$$\lim_{s\to t}\frac{f(s,x) - f(t,x)}{s-t} = \frac{\partial}{\partial t}\gamma_t * f(x)$$

in the L^p norm, making use once more of (6.1.5). We may also use (6.1.25) with $q = p/(p-1)$ to prove point-wise convergence of this limit. This shows that $t \mapsto f(t,x)$ is differentiable for $t > 0$, and $\frac{\partial}{\partial t}f(t,x) = \frac{\partial}{\partial t}\gamma_t * f(x)$. The argument can be repeated to show that for all $m \in \mathbb{N}$,

$$\frac{\partial^m}{\partial t^m}f(t,x) = \frac{\partial^m}{\partial t^m}\gamma_t * f \ .$$

A similar argument shows that $x \mapsto f(t,x)$ is infinitely differentiable everywhere on $\mathbb{R}$, and in fact

$$\frac{\partial^m}{\partial x^m}f(t,x) = \frac{\partial^m}{\partial x^m}\gamma_t * f \ .$$

Together with (6.1.24) we have proved:

6.1.12 THEOREM. *Let $1 \le p < \infty$, $f \in L^p$, and define $f(t,x) = P_t f(x)$. Then $f(\cdot,\cdot)$ is C^∞ on $(0,\infty)\times\mathbb{R}$, satisfies the heat equation*

$$\frac{\partial}{\partial t}f(t,x) = \frac{1}{2}\frac{\partial^2}{\partial x^2}f(t,x) \tag{6.1.26}$$

on $(0,\infty)\times\mathbb{R}$, and $\lim_{t\to 0}\|f(t,\cdot) - f\|_p = 0$.

6.1.13 LEMMA. *For all $t > 0$, and $1 \le p \le 2$, P_t is injective on L^p.*

Proof. Suppose that $f \in L^2$ and $P_t f = 0$. Then $P_{t/2}f \in L^2$, and by by Theorem 6.1.7, and the fact that each $P_{t/2}$ is self-adjoint,

$$0 = \langle f, P_t f\rangle = \langle f, P_{t/2}P_{t/2}f\rangle = \langle P_{t/2}f, P_{t/2}f\rangle = \|P_{t/2}f\|_2^2 \ .$$

Hence $P_{t/2}f = 0$ as well. A simple induction shows that $P_{t/2^n}f = 0$ for all n. Then by Theorem 6.1.7,

$$0 = \lim_{n\to\infty}\|P_{t/2^n}f - f\| = \|f\| \ ,$$

so that $f = 0$.

Now suppose that $f \in L^p$, $1 \le p < 2$ and $P_t f = 0$. By the semigroup property, for all $0 < s < t$ $P_t f = P_{t-s}(P_s f)$, and by Then by Theorem 6.1.7, $P_s f \in L^2$. Therefore, by what was proved just above, $P_s f = 0$ for all $0 < s < t$, and so $f = \lim_{s\to 0} P_s f$. □

6.2 Hermite polynomials and the Mehler semigroup

6.2.1 Hermite polynomials

Let ν be the normalized Gaussian probability measure that has the density $\gamma(x) := (2\pi)^{-1/2}e^{-|x|^2/2}$ with respect to Lebesgue measure $\mathrm{d}x$. Let $\mathcal{H} = L^2(\mathbb{R}, \mathcal{B}, \nu)$. In this section, $\|\cdot\|$ denotes the norm on $\mathcal{H}$, and $\langle\cdot,\cdot\rangle$ denotes the inner product on $\mathcal{H}$. Also, in this section, L^p denotes $L^p(\mathbb{R}, \mathcal{B}, \nu)$.

6.2.1 DEFINITION (Hermite polynomials)**.** For each $n \in \mathbb{N}$, let $V_n = \mathrm{span}(\{1, x, \dots, x^n\})$. Let P_n be the orthogonal projection in $\mathcal{H}$ onto V_n. For $n \in \mathbb{N}$, define the polynomial $p_n(x) = x^n$, and define $p_o(x) = 1$. Note that $\|H_n\| \neq 0$. For all $n \geq 0$, the nth *Hermite polynomial* H_n is given by

$$H_n := p_n - P_{n-1}p_n \tag{6.2.1}$$

and the nth *normalized Hermite polynomial* h_n is defined by

$$h_n := \|H_n\|^{-1}H_n \ . \tag{6.2.2}$$

By construction H_n is orthogonal to H_k for all $k \leq n$, and hence $\langle H_n, H_m\rangle = 0$ for all $n \neq m$. It follows that $\{h_n\}_{n\geq 0}$ is an orthonormal sequence in $\mathcal{H}$.

6.2.2 THEOREM. *For $1 \leq p < \infty$, the set $\mathscr{P}$ of polynomial functions is dense in $L^p(\mathbb{R}, \mathcal{B}, \nu)$.*

Theorem 6.2.2 has the following immediate corollary:

6.2.3 COROLLARY. *The normalized hermite polynomials $\{h_n\}_{n\geq 0}$ are an orthonormal basis for $L^2(\mathbb{R}, \mathcal{B}, \mathrm{d}\nu)$.*

Before proving Theorem 6.2.2, we prove two simple lemmas

6.2.4 LEMMA. *For $y \in \mathbb{R}$, define the function $g_y = e^{yx}$. Then $g_y \in L^p$ for all $p \in [1, \infty)$, and defining $p_n(x) = x^n$, the point-wise series expansion $g_y(x) = \displaystyle\sum_{n=0}^{\infty} \frac{1}{n!} y^n p_n(x)$ converges absolutely in $L^p(\mathbb{R}, \mathcal{B}, \nu)$. That is, for all $y \in \mathbb{R}$, and all $p \in [1, \infty)$,*

$$\sum_{n=0}^{\infty} \frac{1}{n!}|y|^n \|p_n\|_p < \infty \ . \tag{6.2.3}$$

Proof.

$$\int_{\mathbb{R}} |g_y|^p \gamma(x)\mathrm{d}x = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{xyp - x^2/2} = \frac{e^{y^2p^2/2}}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{-(x-yp)/2} = e^{y^2p^2/2} \ .$$

Thus $g_y \in L^p(\mathbb{R}, \mathcal{B}, \nu)$. Since ν is a probability measure, if $q > p$, then $\|p_n\|_q \geq \|p_n\|_p$ for all n. Therefore, it suffices to prove (6.2.3) when p is an even integer. Let us consider the case $p = 2k$, $k \in \mathbb{N}$. Then

$$\|p_n\|_{2k}^{2k} = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} x^{2kn} e^{-|x|^2/2}\mathrm{d}x = (2nk-1)!! = (2nk-1)(2mk-3)\cdots 1 \ .$$

(See the exercises for the evaluations ation of this integral.) The crude estimate $\|p_n\|_{2k}^{2k} \leq (2nk)^{nk}$. which is true since $\|p_n\|_{2k}^{2k}$ is the product of nk terms each of which is less than $2nk$, suffices for our purposes, and gives us $\|p_n\|_{2k} \leq (2nk)^{n/2} = (2k)^{n/2}n^{n/2}$. Therefore,

$$\sum_{n=0}^{\infty} \frac{1}{n!}|y|^n \|p_n\|_{2k} \leq \sum_{n=0}^{\infty} \frac{n^{n/2}}{n!} \left(\sqrt{2k}|y|\right)^n \ .$$

Since the power series $\displaystyle\sum_{n=0}^{\infty} \frac{n^{n/2}}{n!} z^n$ has an infnite radius of convergence, (6.2.3) is true for all y. $\square$

The next lemma relates integration against g_y with respect to $\mathrm{d}\nu$ to the action of the heat semigroup on $L^p(\mathbb{R}, \mathcal{B}, \mathrm{d}x)$.

6.2.5 LEMMA. *For $q \in [1, \infty)$ and $h \in L^q$ and define the function $\widetilde{h}(x) = \gamma^{1/q}(x)h(x)$ so that*

$$\int_{\mathbb{R}} |\widetilde{h}(x)|^q dx = \int_{\mathbb{R}} |h(x)|^q d\nu \ . \tag{6.2.4}$$

$$(2\pi)^{(1-q)/2q} e^{-qy^2/2} \int_{\mathbb{R}} h(x) g_y(x) \mathrm{d}\nu = P_q \widetilde{h}(qy) \ , \tag{6.2.5}$$

where $\{P_t\}_{t \geq 0}$ is the heat semigroup.

Proof. Computing,

$$\begin{aligned} \int_{\mathbb{R}} h(x) g_y(x) \mathrm{d}\nu &= \int_{\mathbb{R}} \widetilde{h}(x) \gamma^{1/q}(x) g_y(x) \mathrm{d}x \\ &= (2\pi)^{-1/2q} \int_{\mathbb{R}} \widetilde{h}(x) e^{-x^2/2q + xy} \mathrm{d}x = (2\pi)^{-1/2q} e^{qy^2/2} \int_{\mathbb{R}} \widetilde{h}(x) e^{-(x-qy))^2/2q} \mathrm{d}x \ . \end{aligned}$$

Rearranging terms yields (6.2.5). $\square$

Proof of Theorem 6.2.2. Suppose that $\mathscr{P}$ is not dense in $L^p(\mathbb{R}, \mathcal{B}, \nu)$. Then by the Hahn-Banach Theorem and the Riesz Representation Theorem, there would exist a unit vector $h \in L^q$, $q = p/(p-1)$, such that $\int_{\mathbb{R}} hf \mathrm{d}\nu = 0$ for all $f \in \mathscr{P}$. By Lemma 6.2.4, g_y is the L^p norm limit of a sequence of polynomials, and therefore $\displaystyle\int_{\mathbb{R}} hg_y \mathrm{d}\nu = 0$ for all $y \in \mathbb{R}$.

Then with $\widetilde{h}(x) = \gamma^{1/q}(x)h(x)$, Lemma 6.2.4 says that $P_q\widetilde{h} = 0$. For $q \in [1, 2]$, P_q is injective on L^q by Lemma 6.1.13. Hence $\widetilde{h} = 0$, and then by (6.2.4), $h = 0$, which is a contradiction. Hence for $2 \in [1, 2]$ there is no unit vector $h \in L^q$ such that $\int_{\mathbb{R}} hf \mathrm{d}\nu = 0$ for all $f \in \mathscr{P}$. Since $q = p/(p-1)$, $q \in [1, 2]$ exactly when $p \in [2, \infty]$, and therefore $\mathscr{P}$ is dense in L^p for $p \in [2, \infty]$.

For $p \in [1, 2)$, $f \in L^p$, and $\epsilon > 0$, pick $g \in L^2$ such that $\|g - f\|_p < \epsilon/2$. Then pick $h \in \mathscr{P}$ such that $\|g - h\|_2 < \epsilon/2$. Then since ν is a probability measure,

$$\|f - h\|_p \leq \|f - g\|_p + \|g - h\|_p \leq \|f - g\|_p + \|g - h\|_2 \leq \epsilon$$

This proves the density of $\mathscr{P}$ in L^p for $p \in [1, 2)$. $\square$

6.2.2 The Mehler semigroup

The Mehler semigroup was introduced by Gustav Mehler in 1866, shortly after Charles Hermite had introduce the Hermite polynomials in 1864. It arises very naturally in the study of the normal distribution $\mathrm{d}\nu$, as we now explain.

For any $(x, y) \in \mathbb{R}^2$ and any $\theta \in [0, 2\pi)$ introduce new coordinates (u, v) on $\mathbb{R}^2$ by

$$u = \cos\theta x + \sin\theta y \qquad \text{and} \qquad v = -\sin\theta x + \cos\theta y$$

so that

$$x = \cos\theta u - \sin\theta v \qquad \text{and} \qquad y = \sin\theta u + \cos\theta v \ .$$

Since (u, v) is obtained by rotating (x, y) (clockwise) through the angle θ, $u^2 + v^2 = x^2 + y^2$ hence

$$\gamma(x)\gamma(y) = \frac{1}{2\pi} e^{-(x^2+y^2)/2} = \frac{1}{2\pi} e^{-(u^2+v^2)/2} = \gamma(u)\gamma(v) \ . \tag{6.2.6}$$

Now let g and f be any polynomials. Then for any $\theta \in [0, 2\pi)$, $g(x)$ and $f(\cos\theta x + \sin\theta y)$ are polynomials in x and y (in the first case with a trvial dependence on y). Therefore we may integrate their product over $\mathbb{R}^2$ with respect to $\gamma(x)\gamma(y)\mathrm{d}x\mathrm{d}y$. We obtain:

$$\int_{\mathbb{R}^2} g(x) f(\cos\theta x + \sin\theta y)\gamma(x)\gamma(y)\mathrm{d}x\mathrm{d}y = \int_{\mathbb{R}} \left(g(x) \int_{\mathbb{R}} f(\cos\theta x + \sin\theta y)\mathrm{d}\nu(y) \right) \mathrm{d}\nu(x) \ . \tag{6.2.7}$$

Now let us write the left side using the rotated cooridnates u, v. Since a rotation has unit Jacobian, (6.2.6) implies

$$\begin{aligned} \int_{\mathbb{R}^2} g(x) f(\cos\theta x + \sin\theta y)\gamma(x)\gamma(y)\mathrm{d}x\mathrm{d}y &= \int_{\mathbb{R}^2} g(\cos\theta u - \sin\theta v) f(u)\gamma(u)\gamma(v)\mathrm{d}u\mathrm{d}v \\ &= \int_{\mathbb{R}} \left(\int_{\mathbb{R}} g(\cos\theta u - \sin\theta v)\mathrm{d}\nu(v) \right) f(u)\mathrm{d}\nu(u) \\ &= \int_{\mathbb{R}} \left(\int_{\mathbb{R}} g(\cos\theta u + \sin\theta v)\mathrm{d}\nu(v) \right) f(u)\mathrm{d}\nu(u) \end{aligned}$$

where in the last line we have used the fact that $\gamma(-v) = \gamma(v)$. Combining this with (6.2.7) and replacing g by its complex conjugate $\overline{g}$, we have the identitiy

$$\int_{\mathbb{R}} \left(\overline{g}(x) \int_{\mathbb{R}} f(\cos\theta x + \sin\theta y)\mathrm{d}\nu(y) \right) \mathrm{d}\nu(x) = \int_{\mathbb{R}} \left(\int_{\mathbb{R}} \overline{g}(\cos\theta u + \sin\theta v)\mathrm{d}\nu(v) \right) f(u)\mathrm{d}\nu(u) \ . \tag{6.2.8}$$

Before porceeding further, let us simplify our notation. Consider angles $\theta \in [0, \pi/2)$ and define $\lambda = \cos\theta$ so that $\sin\theta = \sqrt{1-\lambda^2}$. Then

$$\int_{\mathbb{R}} f(\cos\theta x + \sin\theta y)\mathrm{d}\nu(y) = \int_{\mathbb{R}} f(\lambda x + \sqrt{1-\lambda^2} y)\mathrm{d}\nu(y)$$

This brings us to the following definition:

6.2.6 DEFINITION (Mehler operator)**.** For $\lambda \in [0, 1)$ S_λ is the transformation defined on the polynomials $\mathscr{P}$ by

$$S_\lambda f(x) = \int_{\mathbb{R}} f(\lambda x + \sqrt{1-\lambda^2} y)\gamma(y)\mathrm{d}y \ . \tag{6.2.9}$$

6.2.7 LEMMA. *For all $\lambda \in [0, 1)$, if f is a polynomial of degree less than or equal to $n \in \mathbb{N}$, then so is $S_\lambda f$.*

Proof. Let $p_n(x) := x^n$. By the definition (6.2.9),

$$\begin{aligned} S_\lambda p_n(x) &= \sum_{m=0}^{n} x^{n-m}(1-\lambda^2)^{m/2}\lambda^{n-m}\binom{n}{m}\left(\int_{\mathbb{R}} y^m \gamma(y)\mathrm{d}y \right) \\ &= \sum_{m=0\ ,\ m \text{ even}}^{n} \left[(1-\lambda^2)^{m/2}\lambda^{n-m}\binom{n}{m}(m-1)!! \right] p_{n-m}(x) \ , \end{aligned} \tag{6.2.10}$$

where we have used the fact that $\int_{\mathbb{R}} y^m \gamma(y)\mathrm{d}y = (m-1)!! = (m-1)(m-3)\cdots 1$ for m even and is zero for m odd. This displays $S_\lambda p_n$ as a polynomial of degree n (unless $\lambda = 0$ in which case $S_\lambda p_n$ is constant), and proves the lemma. $\square$

In particular, for all $f \in \mathscr{P}$, $S_\lambda f \in \mathscr{P} \subset \mathcal{H} := L^2(\mathbb{R}, \mathcal{B}, \nu)$, and we may now rewrite (6.2.8) as

$$\langle g, S_\lambda f \rangle_{\mathcal{H}} = \langle S_\lambda g, f \rangle_{\mathcal{H}} \ . \tag{6.2.11}$$

By Lemma 6.2.7, for each $n \in \mathbb{N}$, $V_n := \mathrm{span}(\{1, x, \dots, x^n\})$ is in the domain of S_λ for each $\lambda \in [0,1)$, and is invariant under S_λ. Thus, S_λ define a linear transformation on the finite dimensional vector space V_n, and then (6.2.11) says that this linear transformation is self-adjoint with respect to the inner product that V_n inherits from $\mathcal{H}$ as a subspace of $\mathcal{H}$.

It follows that there exists an orthonormal basis of V_n consisting of eigenvectors of S_λ. Since for all $n \in \mathbb{N}$, $V_{n-1} \subset V_n$, both V_{n-1} and its orthogonal complelement in V_n, are invariant under S_λ because S_λ is self-adjoint. However, by definition, the orthogonal complement of V_{n-1} in V_n is the one dimensional space spanned by h_n, the normalized nth Hermite polynomial. Hence h_n is an eigenvector of S_λ for each $n \in \mathbb{N}$. It is obvious from the definition that $S_\lambda h_0 = h_0$. We have proved the first part of the following theorem:

6.2.8 THEOREM. *The Hermite basis $\{h_n\}_{n\geq 0}$ of $\mathcal{H}$ is a complete orthonormal set consisting eigenvectors of the Mehler operators S_λ: For each $\lambda \in (0,1)$ and each $n > 0$*

$$S_\lambda h_n = \lambda^n h_n \ . \tag{6.2.12}$$

Moreover, for all $f \in \mathscr{P}$, $\|S_\lambda f\|_{\mathcal{H}} \leq \|f\|$ so that S_λ extends by continuity from its its definition on $\mathscr{P}$ to a contraction in $\mathscr{B}(\mathcal{H})$.

Proof. Single out the $m = 0$ term in (6.2.10) to obtain

$$S_\lambda p_n(x) = \lambda^n x^n + \text{lower order} \ . \tag{6.2.13}$$

Then since $H_n(x) = p_n(x) + \text{lower order}$, and since $S_\lambda V_{n-1} \subset V_{n-1}$,

$$S_\lambda H_n = S_\lambda p_n + \text{lower order} = \lambda^n p_n + \text{lower order} = \lambda^n H_n + \text{lower order} \ . \tag{6.2.14}$$

Since $\{h_0, \dots, h_n\}$ spans V_n, $S_\lambda H_n = \displaystyle\sum_{m=0}^{n} \langle h_m, S_\lambda H_n \rangle h_m = \sum_{m=0}^{n} \langle S_\lambda h_m, H_n \rangle$, using the fact that S_λ is self-adjoint. Then since $S_\lambda h_m \in V_m$, for $m < n$, $\langle S_\lambda h_m, H_n \rangle = 0$. Therefore, $S_\lambda H_n$ is a multiple of H_n, and the lower order terms on the right hand side of (6.2.14) must all be zero.

For the final part, let $f = \sum_{n=0}^{N} \alpha_n h_n$ be the expansion of $f \in \mathscr{P}$ in terms of the Hermite basis. Then $S_\lambda f = \sum_{n=0}^{N} \alpha_n \lambda^n h_n$, and hence $\|S_\lambda f\|_{\mathcal{H}}^2 = \displaystyle\sum_{n=0}^{N} \lambda^{2n} |\alpha_n|^2 \leq \sum_{n=0}^{N} |\alpha_n|^2 = \|f\|_{\mathcal{H}}^2$. □

6.2.9 LEMMA. *For all $\lambda, \mu \in (0,1)$,*

$$S_\mu S_\lambda = S_{\lambda\mu} \ , \tag{6.2.15}$$

holds as an identity in $\mathscr{B}(\mathcal{H})$, and for all $f \in \mathcal{H}$,

$$\lim_{\lambda \to 1} \|S_\lambda f - f\|_H = 0 \ . \tag{6.2.16}$$

Proof. For $f \in \mathcal{H}$, consider the expansion $f = \sum_{n=1}^{\infty} \langle h_n, f\rangle h_n$. Then

$$S_\lambda f = \sum_{n=0}^{\infty} \langle h_n, f\rangle S_\lambda h_n = \sum_{n=0}^{\infty} \langle h_n, f\rangle \lambda^n h_n \ .$$

Therefore $S_\mu S_\lambda f = \displaystyle\sum_{n=0}^{\infty} \langle h_n, f\rangle (\mu\lambda)^n h_n = S_{\mu\lambda} f$ Since $L^2 \cap L^p$ is dense in L^p for all $p \in [1,\infty)$, this proves (6.2.15).

To prove (6.2.16), we first consider the case in which f is a polynomial. Again let $p_n(x) := x^n$, and combine (6.2.10) with Minkowski's inequality to conclude

$$\|S_\lambda p_n - p_n\|_{\mathcal{H}} \le (1-\lambda^n)\|p_n\|_{\mathcal{H}} + \sum_{m=0, m \text{ even}}^{n} (1-\lambda^2)^{m/2} \binom{n}{m} (m-1)!! \|p_{n-m}\|_H$$

This shows that $\lim_{\lambda\to 1} \|S_\lambda p_n - p_n\|_{\mathcal{H}} = 0$. It follows that for all $g \in \mathscr{P}$, $\lim_{\lambda\to 1} \|S_\lambda g - g\|_H = 0$.

Next fix $\epsilon > 0$ and $f \in \mathcal{H}$. Let $g \in \mathscr{P}$ be such that $\|g - f\|_H < \epsilon$. Then since $\|S_\lambda\| = 1$,

$$\|S_\lambda f - f\|_{\mathcal{H}} \le \|S_\lambda(f-g)\|_{\mathcal{H}} + \|S_\lambda g - g\|_{\mathcal{H}} + \|t - g\| \le 2\epsilon + \|S_\lambda g - g\|_{\mathcal{H}} \ .$$

For λ sufficiently close to 1, $\|S_\lambda g - g\|_{\mathcal{H}} < \epsilon$, and then $\|S_\lambda f - f\|_{\mathcal{H}} < 3\epsilon$. This proves (6.2.16). □

6.2.10 Remark. One can reparameterize the family of operators $\{S_\lambda \ : \ \lambda \in (0,1)\}$ by defining $M_t = S_{e^{-t}}$ and $M_0 = I$, the identitiy. Then the results we have proved so far show that $\{M_t\}_{t\ge 0}$ is a strongly continuous semigroup on $\mathcal{H}$. This is further developed in the exercises, and justifies the title of this secion, but for our purpose, the present parameterization is more suitable.

6.2.11 LEMMA. *For all $f \in \mathscr{P}$, define*

$$\mathscr{N} f(x) = -\frac{\partial^2}{\partial x^2} f(x) + x\frac{\partial}{\partial x} f(x) \ . \tag{6.2.17}$$

Then for all $\lambda \mapsto S_\lambda f$ is left differentiable in $\mathcal{H}$ at $\lambda = 1$, and

$$\lim_{\lambda\uparrow 1} \frac{S_\lambda f - f}{\lambda - 1} = \mathscr{N} f \ , \tag{6.2.18}$$

where the limit is taken in $\mathcal{H}$, and the same formula is valid point-wise.

Proof. Consider (6.2.10), and explicitly evaluate the terms for $m = 0$ and $m = 2$ in the sum on the right to obtain

$$\begin{aligned} S_\lambda p_n - p_n \quad &= \quad (\lambda^n - 1)p_n + (1-\lambda^2)\lambda^{n-2}\frac{n(n-1)}{2} p_{n-2} \\ &+ \quad \sum_{m=4\ ,\ m \text{ even}}^{n} \left[(1-\lambda^2)^{m/2}\lambda^{n-m}\binom{n}{m}(m-1)!!\right] p_{n-m} \ . \end{aligned} \tag{6.2.19}$$

All of the terms in the second line of (6.2.19) contain at least two factors of $(1-\lambda)$, and hence

$$\lim_{\lambda\uparrow 1} \frac{S_\lambda p_n - p_n}{\lambda - 1} = \lim_{\lambda\uparrow 1}\left[\frac{\lambda^n - 1}{\lambda - 1} p_n - (1+\lambda)\lambda^{n-2}\frac{n(n-1)}{2} p_{n-2}\right] = \mathscr{N} p_n$$

since

$$np_n(x) - n(n-1)p_{n-2}(x) = -\frac{\partial^2}{\partial x^2} p_n(x) + x\frac{\partial}{\partial x} p_n(x) \ .$$

Since the sum has only finitely many terms, the limit may be taken in $\mathcal{H}$ or pointwise. By linearity, these formulae and conclusions extend to general $f \in \mathscr{P}$. □

Applying (6.2.18) with $f = h_n$ yields an important identity:

$$\mathscr{N} h_n = \lim_{\lambda \uparrow 1} \frac{S_\lambda h_n - h_n}{\lambda - 1} = \lim_{\lambda \uparrow 1} \frac{\lambda^n - 1}{\lambda - 1} h_n = n h_n \ .$$

That is, the Hermite basis $\{h_n\}_{n\geq 0}$ is an orthonormal basis if $\mathcal{H}$ consisting of eigenfunctions of $\mathscr{N}$. This is the key to the proof of a number of basic identities concerning the Hemite polynomials.

6.2.12 LEMMA. *For all $f, g \in \mathscr{P}$,*

$$\langle f, \mathscr{N} g\rangle_{\mathcal{H}} = \left\langle \tfrac{\mathrm{d}}{\mathrm{d}x} f, \tfrac{\mathrm{d}}{\mathrm{d}x} g \right\rangle_{\mathcal{H}} \tag{6.2.20}$$

and

$$\langle f, \tfrac{\mathrm{d}}{\mathrm{d}x} g\rangle_{\mathcal{H}} = \left\langle \left(x - \tfrac{\mathrm{d}}{\mathrm{d}x}\right) f, g \right\rangle_{\mathcal{H}} \tag{6.2.21}$$

Proof. This is simply an integration by parts calculation. □

6.2.13 THEOREM. *Define $h_{-1} = 0$. Then for all $n \geq 0$, and all $x \in \mathbb{R}$,*

$$\tfrac{\mathrm{d}}{\mathrm{d}x} h_n(x) = \sqrt{n} h_{n-1}(x) \qquad \text{and} \qquad \left(x - \tfrac{\mathrm{d}}{\mathrm{d}x}\right) h_n(x) = \sqrt{n+1} h_{n+1}(x) \ . \tag{6.2.22}$$

Moroever,

$$x h_n(x) = \sqrt{n} h_{n-1}(x) + \sqrt{n+1} h_{n+1}(x) \ . \tag{6.2.23}$$

Proof. Since $\frac{\mathrm{d}}{\mathrm{d}x} h_n$ is a polynomial of degree $n-1$, it may be expanded in the Hermite basis using only $\{h_0, \dots, h_{m-1}\}$:

$$\tfrac{\mathrm{d}}{\mathrm{d}x} h_n = \sum_{m=1}^{n-1} \langle h_m, \tfrac{\mathrm{d}}{\mathrm{d}x} h_n\rangle_{\mathcal{H}} h_m \ . \tag{6.2.24}$$

By (6.2.21), for $m < n-1$,

$$\langle h_m, \tfrac{\mathrm{d}}{\mathrm{d}x} h_n\rangle_{\mathcal{H}} = \left\langle \left(x - \tfrac{\mathrm{d}}{\mathrm{d}x}\right) h_m, h_n \right\rangle_{\mathcal{H}} = 0$$

since when $m < n-1$, $\left(x - \frac{\mathrm{d}}{\mathrm{d}x}\right) h_m$ is a polynomial of degree at most $n-1$, and is therefore orthogonal to h_n. Hence (6.2.25) simplifies to

$$\tfrac{\mathrm{d}}{\mathrm{d}x} h_n = \langle h_{n-1}, \tfrac{\mathrm{d}}{\mathrm{d}x} h_n\rangle_{\mathcal{H}} h_{n-1} \ . \tag{6.2.25}$$

Combining this with (6.2.20),

$$n = \langle h_n, \mathscr{N} h_n\rangle_{\mathcal{H}} = \left\langle \tfrac{\mathrm{d}}{\mathrm{d}x} h_n, \tfrac{\mathrm{d}}{\mathrm{d}x} h_n \right\rangle_{\mathcal{H}} = \left|\langle h_{n-1}, \tfrac{\mathrm{d}}{\mathrm{d}x} h_n\rangle_{\mathcal{H}}\right|^2 \ . \tag{6.2.26}$$

Since the leading coefficient of h_n is positive for all n, combining (6.2.25) and (6.2.26) yields the first identity in (6.2.22).

It now follows from (6.2.21) that for all $m, n \geq 0$,

$$\sqrt{m} \delta_{n,m-1} = \langle h_n, \tfrac{\mathrm{d}}{\mathrm{d}x} h_m\rangle_{\mathcal{H}} = \left\langle \left(x - \tfrac{\mathrm{d}}{\mathrm{d}x}\right) h_n, h_m \right\rangle_{\mathcal{H}} \tag{6.2.27}$$

and then by the completeness of the Hermite polynomials, this means that for all $n \geq 0$,

$$\left(x - \tfrac{\mathrm{d}}{\mathrm{d}x}\right) h_n = \sqrt{n+1} h_{n+1} \ .$$

Combining the identities in (6.2.22), we obtain (6.2.23). □

Theorem 6.2.13 can be used to prove a number of useful identities concerning the Hermit polynomials, including explict expressions for the coeeficinents and, more important for us here, useful pointwise upper bounds. The rest of this section is not needed in what follows, but illuminates it, and is a beautiful example of classical analysis.

6.2.14 LEMMA. *For all $n \geq 0$, the normalized Hermite polynomial h_n is realted to its un-normalized parent H_n through*

$$H_n = \sqrt{n!} h_n \ , \tag{6.2.28}$$

so that $\|H_n\|_2^2 = n!$. Moreover,

$$H_n(x) = \left(x - \frac{\mathrm{d}}{\mathrm{d}x}\right)^n 1 \tag{6.2.29}$$

and

$$\frac{\mathrm{d}}{\mathrm{d}x} H_n(x) = n H_{n-1}(x) \tag{6.2.30}$$

Proof. since $h_0 = 1$, a simple inductive argument using the inequality on the right in (6.2.22) shows that

$$\left(x - \frac{\mathrm{d}}{\mathrm{d}x}\right)^n 1 = \sqrt{n!} h_n \ .$$

The left hand side is a polynomial for which the coefficient of x^n is 1. Hence the left hand side is the multiple of h_n in which the coeffcient of x^n is 1, which is H_n. This proves (6.2.28) and (6.2.29). Finally, (6.2.30) follows immediately from (6.2.28) and the inequality on the left in (6.2.22). □

6.2.15 LEMMA (Addition formula)**.** *For all $x, y \in \mathbb{R}$, $n \geq 0$.*

$$H_n(x + y) = \sum_{m=0}^{n} \binom{n}{m} H_{n-m}(x) y^m \ . \tag{6.2.31}$$

Proof. By Taylor's Theorem, $H_n(x+y) = \displaystyle\sum_{m=0}^{n} \frac{1}{m!} H_n^{(m)}(x) y^m$ where $H_n^{(m)}$ denotes the mth derivative of H_n. By (6.2.30), $H_n^{(m)} = \dfrac{n!}{(n-m)!} H_{n-m}$. □

6.2.16 LEMMA (Kapetyn's formula)**.** *For all $n \geq 0$,*

$$\gamma * H_n(x) = x^n \ . \tag{6.2.32}$$

Proof. We compute, using the addition formula,

$$\gamma * H_n(x) = \int_{\mathbb{R}} \gamma(y) H_n(x-y) \mathrm{d}y = \sum_{m=0}^{n} \binom{n}{m} x^m \int_{\mathbb{R}} H_{n-m}(y)\gamma(y)\mathrm{d}y = \sum_{m=0}^{n} \binom{n}{m} x^m \delta_{m,n} = x^n \ ,$$

where we have used the orthogonality of the Hermite polynomials and the fact that $H_0 = 1$. □

6.2.17 LEMMA (Rodrigue's formula)**.** *For all $n \geq 0$,*

$$H_n(x) = (-1)^n e^{x^2/2} \left(\frac{\mathrm{d}}{\mathrm{d}x}\right)^n e^{-x^2/2} \ . \tag{6.2.33}$$

Proof. We first prove a useful auxilary result. For each $y \in \mathbb{R}$, the function $f_y(x) = e^{xy-y^2/2}$ belongs to $L^2(\mathbb{R}, \mathcal{B}, \nu)$. Therefore it has a convergent expansion in terms of the Hermite polynomials:

$$e^{xy-y^2/2} = \sum_{n=0}^{\infty} \beta_n(y) H_n(x) \ . \tag{6.2.34}$$

for $\lambda \in (0,1)$, we compute

$$S_\lambda f_y(x) = e^{-y^2/2} \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{\lambda xy + \sqrt{1-\lambda^2}uy} e^{-u^2/2} \mathrm{d}u = f_{\lambda y}(x) \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{-(u-\sqrt{1-\lambda^2}y)^2/2} \mathrm{d}u = f_{\lambda y}(x) \ .$$

From (6.2.34), and $S_\lambda H_n = \lambda^n H)_n$, we also have

$$S_\lambda f_y(x) = \sum_{n=0}^{\infty} \beta_n(y) \lambda^n H_n(x) = \sum_{n=0}^{\infty} \beta_n(\lambda y) H_n(x) \ .$$

Defining $\alpha_n := \beta_n(1)$, we then have that for all $\lambda \in (0,1)$,

$$e^{x\lambda - \lambda^2/2} = \sum_{n=0}^{\infty} \alpha_n \lambda^n H_n(x) \ . \tag{6.2.35}$$

Multiply both sides by $H_m(x)$ and integrate against $\gamma(x)$ to obtain, using Lemma 6.2.14,

$$\gamma * H_m(\lambda) = \alpha_m \lambda^m m! \ .$$

The using Kapetyn's formula, we conclude $\alpha_m = 1/m!$, and hence

$$e^{x\lambda - \lambda^2/2} = \sum_{n=0}^{\infty} \frac{\lambda^n}{n!} H_n(x) \ . \tag{6.2.36}$$

This is the *generating function* for the Hermite polynomials. Multiplying through by $e^{-x^2/2}$, we obtain

$$e^{-(x-\lambda)^2/2} = \sum_{n=0}^{\infty} \frac{\lambda^n}{n!} H_n(x) e^{-x^2/2} \ .$$

Therefore,

$$\left(\frac{\mathrm{d}}{\mathrm{d}x}\right)^n e^{-x^2/2} = (-1)^n \left(\frac{\mathrm{d}}{\mathrm{d}\lambda}\right)^n e^{-(x-\lambda)^2/2}\bigg|_{\lambda=0} = H_n(x) e^{-x^2/2} \ .$$

□

6.2.18 LEMMA (Mehler's formula)**.** *For all* $n \geq 0$*,*

$$\frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{-(ix-y)^2/2} (iy)^n \mathrm{d}y = H_n(x) \ . \tag{6.2.37}$$

Proof.

$$\begin{aligned}
\frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{-(ix-y)^2/2} (iy)^n \mathrm{d}y &= e^{x^2/2} \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} (iy)^n e^{ixy} e^{-y^2/2} \mathrm{d}y \\
&= e^{x^2/2} \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} (iy)^n e^{ixy} e^{-y^2/2} \mathrm{d}y \\
&= e^{x^2/2} \left(\frac{\mathrm{d}}{\mathrm{d}x}\right)^n \left(\frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{ixy} e^{-y^2/2} \mathrm{d}y\right) \\
&= e^{x^2/2} \left(\frac{\mathrm{d}}{\mathrm{d}x}\right)^n e^{-x^2/2} = (-1)^n H_n(x) \ ,
\end{aligned}$$

where we have used Rodrigue's formula in the last line. □

6.2.19 LEMMA (Galbrun's inequality). *For all $n \geq 1$, and all $x \in \mathbb{R}$,*

$$|h_n(x)| \leq \sqrt{\frac{2}{\pi}} e^{x^2/2} \tag{6.2.38}$$

uniformly in n.

Proof. By Mehler's formula (6.2.18),

$$|H_n(x)| \leq \frac{e^{x^2/2}}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{-y^2/2} |y|^n \mathrm{d}y = \frac{e^{x^2/2}}{\sqrt{\pi}} 2^{n/2} \int_0^\infty u^{(n-1)/2} e^{-u} \mathrm{d}u = \frac{e^{x^2/2}}{\sqrt{\pi}} 2^{n/2} \Gamma\left(\frac{n+1}{2}\right) .$$

Define $\alpha_n := \dfrac{2^{n/2}}{\sqrt{n!}} \Gamma\left(\dfrac{n+1}{2}\right)$. Then a simple calcluation shows that $\alpha_n = \sqrt{(n-1)/n}\alpha_{n-2}$. Since $\alpha_1 = \sqrt{2}$ and $\alpha_2 = \sqrt{\pi/2} < \alpha_1$, it follws that $\alpha_n \leq \sqrt{2}$ for all $n \in \mathbb{N}$, Recalling that $H_n = \sqrt{n!} h_n$, the inequality is proved. □

6.2.20 Remark. A deeper inequality due to Hille and Cramér indepndently says that there is a constant C (explicitly given) such that $|h_n(x)| \leq C e^{x^2/4}$ uniformly in n and x. This follws from an analysis of the Sturm-Liouville equation satsified by $h_n(x) e^{-x^2/4}$, and Hille actually proves the stronger result $|h_n(x)| \leq C e^{x^2/4} n^{-1/12}$. However, Galbrun's simpler inequality already tells us something of interest here. The eigenfunction expansion of the operator S_λ gives $S_\lambda = \sum_{k=1}^\infty |h_n\rangle\langle h_n|$ and Galbrun's inequality sats that the series

$$\sum_{n=1}^\infty \lambda^n h_n(x) h_n(y)$$

converges uniformly on compact subsets of $\mathbb{R}^2$, as well as in $L^2(\mathbb{R}^2, \mathcal{B}, \nu \otimes \nu)$.

6.2.3 The Mehler kernel

It turns out that the Mehler operator S_λ that was defined through (6.2.9) can be written in terms of a Markov kernel:

6.2.21 THEOREM. *For all $\lambda \in [0, 1)$, and all $f \in \mathscr{P}$,*

$$S_\lambda f(x) = \int_{\mathbb{R}} K_\lambda(x, z) f(z) \gamma(z) \mathrm{d}z \ . \tag{6.2.39}$$

where

$$K_\lambda(x, z) = \frac{1}{\sqrt{1-\lambda^2}} \exp\left[-\frac{\lambda^2}{2(1-\lambda^2)}\left(z^2 + x^2 - \frac{2}{\lambda} zx\right)\right] . \tag{6.2.40}$$

Proof. We make a change of varialbles taking z to be the argunment of f in (6.2.9). Define $z = \lambda x + \sqrt{1-\lambda^2} y$ so that $\begin{pmatrix} x \\ z \end{pmatrix} = \begin{bmatrix} 1 & 0 \\ \lambda & \sqrt{1-\lambda^2} \end{bmatrix} \begin{pmatrix} x \\ y \end{pmatrix}$. Computing the Jacobian of the transformation, we find $\mathrm{d}x\mathrm{d}y = (1-\lambda^2)^{-1/2} \mathrm{d}x\mathrm{d}z$. Also,

$$x^2 + y^2 = \frac{\lambda^2}{1-\lambda^2}\left(z^2 + x^2 - \frac{2}{\lambda} zx\right) + x^2 + z^2 \ ,$$

Hence

$$\gamma(x)\gamma(y) = \exp\left[-\frac{\lambda^2}{2(1-\lambda^2)}\left(z^2 + x^2 - \frac{2}{\lambda} zx\right)\right] \gamma(x)\gamma(z) \ .$$

and for $g, f \in \mathscr{P}$, $\int_{\mathbb{R}} g(x)S_\lambda f(x)\gamma(x)\mathrm{d}x = \int_{\mathbb{R}}\int_{\mathbb{R}} g(x)K_\lambda(x,z)f(z)\gamma(x)\gamma(z)\mathrm{d}z\mathrm{d}x$ where $K_\lambda(x,y)$ is given by (6.2.40). Since the $\mathscr{P}$ is dense in $\mathcal{H}$, this hold for all $g \in \mathcal{H}$ and this proves (6.2.39) □

Note that $K_\lambda(x,z)\gamma(z) = \frac{1}{\sqrt{2\pi}}\frac{1}{\sqrt{1-\lambda^2}}\exp\left[-\frac{1}{2(1-\lambda^2)}|z-\lambda x|^2\right] = \gamma_{1-\lambda^2}(z-\lambda x)$. By (6.1.19) and symmetry,

$$\int_{\mathbb{R}} K_\lambda(x,z)\gamma(z)\mathrm{d}z = \int_{\mathbb{R}} \gamma(x)K_\lambda(x,z)\mathrm{d}x = 1 \ . \tag{6.2.41}$$

This shows that $K_\lambda(x,y)$ is a symmetric (and therefore doubly stochastic) Markov kernel on $(\mathbb{R},\mathcal{B},\mathrm{d}\nu)$. Since S_λ is the operators associated to K_λ, Theorem 6.1.2 immediately yields:

6.2.22 LEMMA. *For all $p \in [1,\infty]$, all continuous polynomaily bounded functions f, and all λ in$(0,1)$,*

$$\|S_\lambda f\|_p \le \|f\|_p \ . \tag{6.2.42}$$

Combining this with the density of $\mathscr{P}$ in L^p for $p \in [1,\infty)$, we have that for each $\lambda \in (0,1)$, S_λ extends by continuity to a bounded linear operator of norm one on L^p. We denote the extension also by S_λ. The L^p theory is further develooped in the exercises.

6.3 The Fourier Transform

6.3.1 The Hermite functions

The map $U : f \mapsto \sqrt{\gamma}f$ is evidently a unitary transformation from $L^2(\mathbb{R},\mathcal{B},\nu)$ to $L^2(\mathbb{R},\mathcal{B},\mathrm{d}x)$. The image under U of the Hermite polynomial basis $\{h_n\}_{n\ge 0}$ for $L^2(\mathbb{R},\mathcal{B},\nu)$ is evidently an orthonormal basis for $L^2(\mathbb{R},\mathcal{B},\mathrm{d}x)$

6.3.1 DEFINITION (Hermite functions)**.** For integers $n \ge 0$, define $g_n = \sqrt{\gamma}h_n$. Then g_n is the nth *Hermite function* and $\{g_n\}_{n\ge 0}$ is the *Hermite function basis* of $L^2(\mathbb{R},\mathcal{B},\mathrm{d}x)$.

6.3.2 LEMMA. *For all $n \ge 0$,*

$$\frac{\mathrm{d}}{\mathrm{d}x}g_n(x) = \frac{1}{2}\left(\sqrt{n}g_{n-1} - \sqrt{n+1}g_{n+1}\right) \tag{6.3.1}$$

and

$$\frac{x}{2}g_n(x) = \frac{1}{2}\left(\sqrt{n}g_{n-1} + \sqrt{n+1}g_{n+1}\right) \tag{6.3.2}$$

Proof. We compute using (6.2.22) and (6.2.23):

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}x}g_n(x) &= \sqrt{\gamma(x)}\frac{\mathrm{d}}{\mathrm{d}x}h_n(x) + \left(\frac{\mathrm{d}}{\mathrm{d}x}\sqrt{\gamma(x)}\right)h_n(x) \\
&= \sqrt{\gamma(x)}\sqrt{n}h_{n-1}(x) - \frac{x}{2}\sqrt{\gamma(x)}h_n(x) \\
&= \sqrt{\gamma(x)}\sqrt{n}h_{n-1}(x) - \frac{1}{2}\sqrt{\gamma(x)}\left(\sqrt{n}h_{n-1}(x) + \sqrt{n+1}h_{n+1}(x)\right) \ ,
\end{aligned}$$

and this proves (6.3.1). Even more simply, (6.3.2) follows from (6.2.23). □

6.3.3 DEFINITION (Fourier Transform on $L^2(\mathbb{R}, \mathcal{B}, \mathrm{d}x)$)**.** Define a unitary transformation $\mathcal{F}$ on $L^2(\mathbb{R}, \mathcal{B}, \mathrm{d}x)$ by and

$$\mathcal{F} g_n = (-i)^n g_n \tag{6.3.3}$$

for all $n \geq 0$. This is the *Fourier transform on* $L^2(\mathbb{R}, \mathcal{B}, \mathrm{d}x)$

In what follows, we write $\dfrac{x}{2}$ to denote the operator of multiplication by $x/2$. Then by the definition of $\mathcal{F}$ and Lemma 6.3.2, for all $n \geq 0$

$$\frac{x}{2}\mathcal{F} g_n = (-i)^n \frac{1}{2}\left(\sqrt{n} g_{n-1} + \sqrt{n+1} g_{n+1}\right) = \frac{1}{2}\left(\sqrt{n}(-i)i\mathcal{F} g_{n-1} - \sqrt{n+1}(-i)i\mathcal{F} g_{n+1}\right) = -i\mathcal{F}\frac{\mathrm{d}}{\mathrm{d}x} g_n \ .$$

In other words,

$$\mathcal{F}^* \circ \frac{x}{2} \circ \mathcal{F} g_n = \frac{1}{i}\frac{\mathrm{d}}{\mathrm{d}x} g_n \qquad \text{and} \qquad \mathcal{F} \circ \frac{1}{i}\frac{\mathrm{d}}{\mathrm{d}x} \circ \mathcal{F}^* g_n = \frac{x}{2} g_n \ . \tag{6.3.4}$$

This identity is the primary source of the utility of the Fourier transform; it *diagonalizes differentiation* in the sense that it unitarily identifies differentiation with a multiplication operator. The formulae (6.3.4) would be more useful as a pair of identities between operators. To reformulate them as such, we have to take into account that neither multiplication by $x/2$ nor differentiation are defined on all of $L^2(\mathbb{R}, \mathcal{B}, \mathrm{d}x)$. Therefore, we introduce a Hilbert space on which they are defined.

6.3.4 DEFINITION. The Hilbert space $\mathcal{H}_1$ consists of the functions $f \in \mathrm{L}^2(\mathbb{R}, \mathcal{B}, \mathrm{d}x)$ having the Hermite function expansion

$$f = \sum_{n=0}^{\infty} \alpha_n g_n$$

such that $\sum_{n=0}^{\infty}(n+1)|\alpha_n|^2 < \infty$. For f, g in $\mathcal{H}_1$, the $\mathcal{H}_1$ inner product is defined by

$$\langle f, g\rangle_{\mathcal{H}_1} = \sum_{n=0}^{\infty}(n+1)\overline{\alpha_n}\beta_n \ ,$$

where $\alpha_n = \langle g_n, f\rangle_{L^2}$ and $\beta_n = \langle g_n, g\rangle_{L^2}$. We denote the norm on $\mathcal{H}_1$ by $\|\cdot\|_{\mathcal{H}_1}$. Observe that $\mathcal{H}_1$, considered as a subspace of L^2, is dense in L^2.

Consider any finite linear combination $f = \sum_{n=0}^{m} \alpha_n g_n$ for Hermite functions. Then

$$\frac{\mathrm{d}}{\mathrm{d}x} f = \frac{1}{2}\sum_{n=0}^{m} \alpha_n \left(\sqrt{n} g_{n-1} - \sqrt{n+1} g_{n+1}\right)$$

Now a simple computation shows that $\left\|\dfrac{\mathrm{d}}{\mathrm{d}x} f\right\|_{L^2} \leq \|f\|_{\mathcal{H}_1}$. Likewise, $\left\|\dfrac{x}{2} f\right\|_{L^2} \leq \|f\|_{\mathcal{H}_1}$. Therefore, both $\dfrac{\mathrm{d}}{\mathrm{d}x}$ and $\dfrac{x}{2}$ belong to $\mathscr{B}(\mathcal{H}_1, L^2)$. Since $\{(n+1)^{-1/2} g_n\}_{n\geq 0}$ is an orthonormal basis for $\mathcal{H}_1$, $\mathcal{F}$ is unitary on $\mathcal{H}_1$ as well as on L^2. Therefore we have:

6.3.5 THEOREM. *The operators* $\dfrac{\mathrm{d}}{\mathrm{d}x}$ *and* $\dfrac{x}{2}$ *in* $\mathscr{B}(\mathcal{H}_1, L^2)$ *are related through the Fourier transform* $\mathcal{F}$ *by*

$$\mathcal{F}^* \circ \frac{x}{2} \circ \mathcal{F} = \frac{1}{i}\frac{\mathrm{d}}{\mathrm{d}x} \qquad \text{and} \qquad \mathcal{F} \circ \frac{1}{i}\frac{\mathrm{d}}{\mathrm{d}x} \circ \mathcal{F}^* = \frac{x}{2} \ . \tag{6.3.5}$$

The Fourier transform was originally introduced in a rather different way. We now explain the connection with the original definition. Let D be the closed unit disk in $\mathbb{C}$: $D := \{z \in \mathbb{C} \ : \ |z| \le 1\}$. For $z \in D$ we define

$$S_z = \sum_{n=0}^{\infty} z^n |h_n\rangle\langle h_n| \ . \tag{6.3.6}$$

For $z \in D^o$, the series in (6.3.6) converges in the operator norm, while for $z = e^{i\theta} \in \partial D$, the series converges in the strong operator topology and $S_{e^{i\theta}}$ is unitary.

This family of operators can be transplanted from $L^2(\mathbb{R}, \mathcal{B}, \nu)$ to $L^2(\mathbb{R}, \mathcal{B}, \mathrm{d}x)$ using the unitary transformation $U : f \mapsto \sqrt{\gamma} f$. For $z \in D$ we define $T_z = U M_z U^*$, which is the same as

$$T_z = \sum_{n=1}^{\infty} z^n |g_n\rangle\langle g_n| \ . \tag{6.3.7}$$

As before, the series in (6.3.7) converges in the operator norm, while for $z = e^{i\theta} \in \partial D$, the series converges in the strong operator topology and $T_{e^{i\theta}}$ is unitary. Now note that $\mathcal{F} = T_i$.

6.3.6 LEMMA. *The map $z \mapsto T_z$ is continuous from D^o into $\mathscr{B}(L^2)$ with the norm topology on the range, and it is continuous from D into $\mathscr{B}(L^2)$ with the strong operator topology on the range.*

Proof. Consider $z, w \in D^o$, Then for some $r \in (0,1)$, $|z|, |w| \le r$. Since $\||g_b\rangle\langle g_n|\| = 1$, Minkowski's inequality and the identity

$$z^n - w^n = (z - w) \sum_{m=0}^{n-1} z^{n-1-m} w^m \tag{6.3.8}$$

gives us

$$\|T_z - T_w\| \le \sum_{n=0}^{\infty} |z^n - w^n| \le |z - w| \left(\sum_{n=0}^{\infty} n r^n \right) = |z - w| \frac{r}{(1-r)^2} \ .$$

This proves the norm continuity into $\mathscr{B}(L^2)$.

Next, fix $f \in L^2$, and for $n \ge 0$, define $\alpha_n = \langle g_n, f\rangle_{L^2}$ so that $\|f\|_2^2 = \sum_{n=0}^{\infty} |\alpha_n|^2$. Fix $z \in D$, which we may as well assume to belong to ∂D. We must show that for all $\epsilon > 0$, there exists $\delta > 0$ such that for all $w \in D$ with $|w - z| < \delta$, $\|T_w f - T_z f\|_2 < \epsilon$. Pick $\epsilon > 0$, and then N such that $\sum_{n=N+1}^{\infty} |\alpha_n|^2 < \epsilon/4$. Then, using (6.3.8) once more in the last line

$$\begin{aligned} \|T_w f - T_z f\|_2^2 &\le \sum_{n=1}^{N} |z^n - w^n||\alpha_n|^2 + \sum_{n=N+1}^{\infty} |z^n - w^n||\alpha_n|^2 \\ &\le \|f\|_2^2 \sum_{n=1}^{N} |z^n - w^n| + 2 \sum_{n=N+1}^{\infty} |\alpha_n|^2 \\ &\le |z - w| \|f\|_2^2 \sum_{n=1}^{N} (n-1) + \frac{\epsilon}{2} \ . \end{aligned}$$

Thus we may tale $\delta = \|f\|_2^2 N(N-1)\epsilon/4$.

□

For $z = \lambda \in (0,1)$, the operator S_z has the representation (6.2.40), in terms of the kernel $K_\lambda(x,y)$ defined there, and consequently, for any $f, g \in L^2(\mathbb{R}, \mathcal{B}, \mathrm{d}x)$ and $\lambda \in (0,1)$,

$$\langle f, T_\lambda g\rangle_{L^2(\mathrm{d}x)} = \int_{\mathbb{R}} \int_{\mathbb{R}} \overline{f(x)} \sqrt{\gamma}(x) K_\lambda(x,y) \sqrt{\gamma}(y) g(y) \mathrm{d}x\mathrm{d}y = \int_{\mathbb{R}} \int_{\mathbb{R}} \overline{f(x)} M_\lambda(x,y) g(y) \mathrm{d}x\mathrm{d}y$$

where
$$M_\lambda(x,y) := \frac{1}{\sqrt{2\pi}}\frac{1}{\sqrt{1-\lambda^2}}\exp\left[-\frac{1+\lambda^2}{4(1-\lambda^2)}\left(y^2+x^2\right)+\frac{\lambda}{1-\lambda^2}yx\right] . \tag{6.3.9}$$
The from (6.3.7) we obtain the identity
$$\frac{1}{\sqrt{2\pi}}\frac{1}{\sqrt{1-\lambda^2}}\exp\left[-\frac{1+\lambda^2}{4(1-\lambda^2)}\left(y^2+x^2\right)+\frac{\lambda}{1-\lambda^2}yx\right] = \sum_{n=0}^\infty \lambda^n g_n(x)g_n(y) . \tag{6.3.10}$$
By Galbrun's Inequality (6.2.19), $|g_n(x)g_n(y)|$ is bounded uniformly in $n\in\mathbb{N}$ and x,y in any compact set of the plane. (Bt the deeper Cramér-Hille inequality, which we have not proved but do not need, it is in fact boundend informly in $n\in\mathbb{N}$ and $x,y\in\mathbb{R}^2$.) Therefore, $z\mapsto\sum_{n=0}^\infty z^n g_n(x)g_n(y)$ converges uniformly on the set
$$\{(z,x,y)\ :|z|<r\ ,\ \sqrt{x^2+y^2}<R\ \}$$
for all $r\in(0,1)$ and all $R\in(0,\infty)$. In particular, both sides of (6.3.10) are equal and anlytic in λ for $\lambda\in(0,1)$. Note that $\Re(1-z^2)>0$ for $z\in D^0$, and hence $\sqrt{1-z^2}$ is analytic on D^o.) Therefore, we may replace $\lambda\in(0,1)$ by $z\in D^o$ in (6.3.10) to define $M_z(x,y)$, and then with this definition, for all $f\in L^2$, $z\in D$,
$$T_zf(x) = \int_{\mathbb{R}} M_z(x,y)f(y)\mathrm{d}x\mathrm{d}y . \tag{6.3.11}$$

By the previous lemma, $T_{-i}f=\lim_{z\to -i}T_zf$ in the norm topology on L^2, Hence if we let $\{z_n\}_{n\in\mathbb{N}}$ be any sequence in D^o with $\lim_{n\to\infty}z_n=-i$, we have that
$$T_{-i}f(x) = \lim_{n\to\infty}\int_{\mathbb{R}} M_{z_n}(x,y)f(y)\mathrm{d}x\mathrm{d}y$$
Since $M_{-i}(x,y)=\dfrac{1}{2\sqrt{\pi}}e^{-ixy/2}$, if we formally take the limit under the integral sign, we obtain
$$\mathcal{F}f(x)=T_tf(x)=\frac{1}{2\sqrt{\pi}}\int_{\mathbb{R}}e^{-ixy/2}f(y)\mathrm{d}y\ ,$$
If $f\in L^1\cap L^2$, there is nothing formal: We may take the limit $n\to\infty$ two ways. Using the fact that $f\in L^2$, we have the L^2 norm convergence of $T_{z_n}f$ to T_if, and hence we have convergence almost everywhere along a subsequence. Next, since $f\in L^1$, and $|M_z(x,y)|$ is uniformly bounded by a constant for z in a neighborhood of i, we may apply the Lebesgue Dominated Convergence Theorem to conclude the point-wise convergence
$$\lim_{z\to -i}T_zf(x)=\lim_{z\to -i}\int_{\mathbb{R}}M_{z_n}(x,y)f(y)\mathrm{d}x\mathrm{d}y=\frac{1}{2\sqrt{\pi}}\int_{\mathbb{R}}e^{-ixy/2}f(y)\mathrm{d}y\ .$$
We have proved:

6.3.7 THEOREM. *For all $f\in L^1(\mathbb{R},\mathcal{B},\mathrm{d}x)\cap L^2(\mathbb{R},\mathcal{B},\mathrm{d}x)$, the Fourier transform $\mathcal{F}f$ of f is given by*
$$\mathcal{F}f(x)=\frac{1}{2\sqrt{\pi}}\int_{\mathbb{R}}e^{-ixy/2}f(y)\mathrm{d}y \tag{6.3.12}$$

There is a one-parameter family of unitary transformations on $L^2(\mathbb{R},\mathcal{B},\mathrm{d}x)$ by composing $\mathcal{F}$ with a unitary scale transformation: For λ^0, and $f\in L^2(\mathbb{R},\mathcal{B},\mathrm{d}x)$, define
$$\delta_\lambda f(x)=\sqrt{\lambda}f(\lambda x)\ . \tag{6.3.13}$$

Then δ_λ is unitary, and therefore so it $\delta_\lambda\mathcal{F}$, and we have

$$\delta_\lambda\mathcal{F}f(x) = \frac{\sqrt{\lambda}}{2\sqrt{\pi}}\int_{\mathbb{R}} e^{-i\lambda xy/2}f(y)\mathrm{d}y \tag{6.3.14}$$

The choices $\lambda = 2$ and $\lambda = 4\pi$ are traditional in various fields, and the resulting operators are also known as the Fourier transform. The latter has many advantages; For $f \in L^2(\mathbb{R}, \mathcal{B}, \mathrm{d}x)$, define $\widehat{f}$ by $\widehat{f} := \delta_{4\pi}\mathcal{F}f$. Then for $f \in L^1(\mathbb{R}, \mathcal{B}, \mathrm{d}x) \cap L^2(\mathbb{R}, \mathcal{B}, \mathrm{d}x)$

$$\widehat{f}(x) = \int_{\mathbb{R}} e^{-i2\pi xy}f(y)\mathrm{d}y \ . \tag{6.3.15}$$

6.3.8 Remark. We could have arrived directly at (6.3.15) had we used a different Gaussian density γ_t to define our reference measure ν, and thus the Hermite polynomials. The choice $t = 1$ for the density of ν makes the variance of ν equal to one, so that ν is what probabilists call the "normal distribution" on $\mathbb{R}$. This "normal" choice gives the "normal" formulas for the "normal" Hermite polynomials, though of course other Hermite polynomials can be defined as the orthonormal sequence in $L^2(\mathbb{R}, \mathcal{B}, \gamma_t(x)\mathrm{d}x)$ that one obtains by applying the Gram-Schmidt algorithm to the sequence of non-negative powers of x. Then one obtains identities analogous to those found in Theorem 6.2.13, which are the basis of the key properties of the Fourier transform given in Theorem 6.3.5. Had we use $\gamma_t e^{-x^2/2t}\mathrm{d}x$ for $t = \frac{1}{2\pi}$ as our reference measure; that is, $e^{-\pi|x^2}\mathrm{d}x$, we would have arrived directly at (6.3.15) as the definition of the Fourier transform.

Partly for the reasons explained above, it is useful to record the simple interaction of $\mathcal{F}$ with several other one parameter families of unitary operations on $L^2(\mathbb{R}, \mathcal{B}, \mathrm{d}x)$. Another is the unitary group of translation τ_y, $y \in \mathbb{R}$ where $\tau_y f(x) = f(x-y)$. Another is the group of *phase transformations* ϕ_y given

$$\phi_y f(x) = e^{-ixy}f(x) \ . \tag{6.3.16}$$

In the formulation of the following lemma, we will use the physics convention of writing k for the variable that is the argument of $\mathcal{F}f$. *In this context, k ranges over $\mathbb{R}$ and not $\mathbb{N}$ or $\mathbb{Z}$.*

6.3.9 LEMMA. *For $\lambda > 0$, let δ_λ be given by (6.3.13). Then*

$$\delta_\lambda \circ \mathcal{F} = \mathcal{F} \circ \delta_{\lambda^{-1}} \ . \tag{6.3.17}$$

Moreover for all $y \in \mathbb{R}$, let τ_y be the translation operator on $L^2(\mathbb{R}, \mathcal{B}, \mathrm{d}x)$, $\tau_y f(x) = f(x-y)$, and let ϕ_y be the phase transformation defined in (6.3.16). Then

$$\mathcal{F} \circ \tau_y = \phi_{y/2} \circ \mathcal{F} \ . \tag{6.3.18}$$

Proof. For $f \in L^1(\mathbb{R}, \mathcal{B}, \mathrm{d}x) \cap L^2(\mathbb{R}, \mathcal{B}, \mathrm{d}x)$, making the change of variables $x = u/\lambda$,

$$\delta_\lambda\mathcal{F}f(x) = \frac{\sqrt{\lambda}}{2\sqrt{\pi}}\int_{\mathbb{R}} e^{-i\lambda kx/2}f(y)\mathrm{d}x = \delta_\lambda\mathcal{F}f(k) = \frac{\sqrt{\lambda}}{2\sqrt{\pi}}\int_{\mathbb{R}} e^{-i\lambda k(u/\lambda)/2}f(u/\lambda)\frac{1}{\lambda}\mathrm{d}u = \mathcal{F}\delta_{\lambda^{-1}}f(k) \ .$$

Likewise, making the change of variable $u = x - y$,

$$\mathcal{F}\tau_y f(k) = \frac{1}{2\sqrt{\pi}}\int_{\mathbb{R}} e^{-ikx/2}f(x-y)\mathrm{d}x = \frac{1}{2\sqrt{\pi}}\int_{\mathbb{R}} e^{-ik(u+y)/2}f(u)\mathrm{d}u = e^{-iky/2}\mathcal{F}f(k) = \phi_{y/2}\mathcal{F}f(k) \ .$$

$\square$

Another way to rephrase (6.3.18) is

$$\mathcal{F} \circ \tau_y \circ F^* = \phi_{y/2} \ , \tag{6.3.19}$$

so that $\mathcal{F}$ "diagonalizes" translation, in the sense that it relates translation to a unitary "multiplication operator" (multiplication by $e^{-iky/2}$) as specified in (6.3.19).

6.3.10 THEOREM. *For any function $h \in L^1 \cap L^\infty$ define*

$$\widehat{f}(k) = \int_{\mathbb{R}} e^{-i2\pi kx} f(x)\mathrm{d}x \ . \tag{6.3.20}$$

Then for $f, g \in L^1 \cap L^\infty$,

$$\widehat{f * g}(k) = \widehat{f}(k)\widehat{g}(k) \ . \tag{6.3.21}$$

Proof. We compute

$$\begin{aligned}\widehat{f * g}(k) &= \int_{\mathbb{R}} e^{-i2\pi kx} \left(\int_{\mathbb{R}} f(x-y)g(y)\mathrm{d}y \right) \mathrm{d}x \\ &= \int_{\mathbb{R}} \int_{\mathbb{R}} e^{-i2\pi k(x-y)} f(x-y) e^{-i2\pi ky} g(y)\mathrm{d}x\mathrm{d}y = \widehat{f}(k)\widehat{g}(k) \ .\end{aligned}$$

□

Since $\gamma_t(x)e^{i2\pi kx} = \frac{1}{\sqrt{2\pi t}} \exp(-\frac{1}{2t}(x - 2\pi ikt)^2) \exp(2\pi^2k^2t)$

$$\widehat{\gamma_t}(k) = \exp(2\pi^2k^2t) \ . \tag{6.3.22}$$

In particular, taking $t = (2\pi)^{-1}$ so that $\gamma_{1/2pi}(x) = e^{-\pi x^2}$,

$$\widehat{\gamma_{1/2\pi}}(k) = \gamma_{1/2\pi}(k) \ . \tag{6.3.23}$$

That is, $\gamma_{1/2\pi}$ is a fixed point of the map $f \mapsto \widehat{f}$, as we have seen earlier.

For $f \in L^1 \cap L^2$, recall that $P_tf(x) = \gamma_t * f * x)$, where $\{P_t\}_{t\geq 0}$ is the heat semigroup. Then by Theorem 6.3.10,

$$\widehat{P_tf}(k) = \widehat{\gamma_t}(k)\widehat{f}(k) = \exp(2\pi^2k^2t)\widehat{f}(k) \ .$$

Since $L^1 \cap L^2$ is dense in L^2, the formula $\widehat{P_tf}(k) = \exp(2\pi^2k^2t)\widehat{f}(k)$ extends by continuity to all $f \in L^2$. Since $f \mapsto \widehat{f}$ is unitary and P_t is contractive on L^2, $P_tf = 0$ implies that $\widehat{P_tf}(k) = 0$ for almost every k, and then the identity above shows that $\widehat{f}(k) = 0$ for almost every k. By the unitarity of $f \mapsto \widehat{f}$, this means that $f = 0$. This gives another proof of the infectivity of the heat semigroup operators on L^2, however, it relies on the unitarity of $f \mapsto \widehat{f}$, and in the approach to this taken her, we have relied on this infectivity to prove the unitarity of $f \mapsto \widehat{f}$, but there are other approaches in which the order can be inverted.

6.3.2 Higher dimensions

So far, we have discussed the Fourier transform and convolution for functions on $\mathbb{R}$. It is easy to extend our results to functions on $\mathbb{R}^n$. In this section, $L^p(\mathbb{R}^n)$ denotes $L^p(\mathbb{R}^n, \mathcal{B}, \mathrm{d}x)$.

We begin with the Fourier transform and $n = 2$. For integers $n, m \geq 0$ define

$$g_m \otimes g_m(x_1, x_2) = g_m(x_1)g_n(x_2) \ . \tag{6.3.24}$$

It is then evident that $\{g_m \otimes g_m\}_{m,n\geq 0}$ is an orthonormal basis of $L^2(\mathbb{R}^2)$.

For a function $f \in L^1(\mathbb{R}^2) \cap L^2(\mathbb{R}^2)$ define

$$\widehat{f}(k) = \int_{\mathbb{R}^2} e^{-i2\pi k\cdot x} f(x)\mathrm{d}x \ , \tag{6.3.25}$$

where $k \cdot x$ denotes the standard inner product in $\mathbb{R}^2$, so that $k \cdot x = k_1x_1 + k_2x_2$ and

$$e^{-i2\pi k\cdot x} = e^{-i2\pi k_1x_1}e^{-i2\pi k_2x_2} \ .$$

It follows that

$$\begin{aligned} \widehat{g_m \otimes g_n}(k_1.k_2) &= \left(\int_{\mathbb{R}} e^{-i2\pi k_1x_1} g_m(x_1)\mathrm{d}x_1\right)\left(\int_{\mathbb{R}} e^{-i2\pi k_2x_2} g_n(x_2)\mathrm{d}x_2\right) \\ &= \widehat{g}_m(k_1)\widehat{g}_n(k_2) = (-i)^{m+n} g_m \otimes g_n(k_1.k_2) \ . \end{aligned}$$

Since $\{(-i)^{m+n}g_m \otimes g_m\}_{m,n\geq 0}$ is also an orthonormal basis of $L^2(\mathbb{R}^2)$, the map $f \mapsto \widehat{f}$ defined in (6.3.25) is unitary on $L^2(\mathbb{R}^n)$.

The same considerations extend in the obvious way to functions on $\mathbb{R}^n$ for arbitrary $n \in N$. For $f \in L^1(\mathbb{R}^n) \cap L^2(\mathbb{R}^n)$, we define

$$\widehat{f}(k) = \int_{\mathbb{R}^n} e^{-i2\pi k\cdot x} f(x)\mathrm{d}x \ , \tag{6.3.26}$$

and then $f \mapsto \widehat{f}$ extends by continuity to a unitary transformation on $L^2(\mathbb{R}^n)$. Even more simply,

$$|\widehat{f}(k)| \leq \int_{\mathbb{R}^n} |f(x)|\mathrm{d}x = \|f\|_1$$

so that $f \mapsto \widehat{f}$ extends by continuity to a contraction from $L^1(\mathbb{R}^n)$ to $L^\infty(\mathbb{R}^n)$. In summary, we have two operator norm bounds for the map $f \mapsto \widehat{f}$: It extends by continuity to an element of $\mathscr{B}(L^2(\mathbb{R}^n), L^2(\mathbb{R}^n))$ with operator norm 1, and it extends by continuity to an element of $\mathscr{B}(L^1(\mathbb{R}^n), L^\infty(\mathbb{R}^n))$ with operator norm 1

Likewise, if $f \in L^1(\mathbb{R}^n)$, $f \neq 0$, define $\rho(x) = \|f\|_1^{-1}|f(x)|$ so that ρ is a probability density on $\mathbb{R}^n$ and $K(x,y) := \rho(x-y)$ is a doubly stochastic Markov Kernel. It then follows directly from Theorem 6.1.2 that for all $g \in L^1(\mathbb{R}^n) \cap L^\infty(\mathbb{R}^n)$, that if we define

$$f * g(x) := \int_{\mathbb{R}^n} f(x-y)g(y)\mathrm{d}y = \|f\|_1 \int_{\mathbb{R}^n} K(x,y)g(y)\mathrm{d}y \ ,$$

then for all $p \in [1,\infty]$, then $\|f^*g\|_p \leq \|f\|_1\|g\|_p$. Even more simply if we assume that both f and g belong to $L^1(\mathbb{R}^n) \cap L^\infty(\mathbb{R}^n)$, and therefore to $L^p(\mathbb{R}^n)$ for all $p \in [1,\infty]$, H "older's inequality gives us, for each x, and each p,

$$|f * g(x)| \leq \|f\|_{p/(p-1)}\|g\|_p \ .$$

We may rephrase the last two inequalities as a pair of operator bounds. For $g \in L^p(\mathbb{R}^n)$ define an operator T_g on $g \in L^1(\mathbb{R}^n) \cap L^\infty(\mathbb{R}^n)$ $T_gf = f * g$, Then we have $\|T_gf\|_p \leq \|g\|_p\|f\|_1$ and $\|T_gf\|_\infty \leq \|g\|_p\|f\|_{p/(p-1)}$. Therefore, T_g extends by continuity to an element of $\mathscr{B}(L^1(\mathbb{R}^n), L^p(\mathbb{R}^n))$ an its operator norm in this space is no more that $\|g\|_p$. Likewise, T_g extends by continuity to an element of $\mathscr{B}(L^{p/(p-1)}(\mathbb{R}^n), L^\infty(\mathbb{R}^n))$ an its operator norm in this space is no more that $\|g\|_p$.

Thus for both the Fourier transform and for convolution we have operator norm bounds in $\mathscr{B}(L^p(\mathbb{R}^n), L^q(\mathbb{R}^n))$ for two specific pair of values of p and q. In the next section we shall see how to "interpolate" between such pairs of inequalities producing a one parameter family of inequalities.

6.3.3 The Riesz-Thorin Interpolation Theorem

6.3.11 THEOREM. *Let $(\Omega, \mathcal{M}, \mu)$ be an arbitrary measure space, and for $p \in [1,\infty]$, let L^p denote $L^p(\Omega, \mathcal{M}, \mu)$. For $p_1, p_2 \in [1,\infty]$, The T be a linear operator defined on $L^{p_1} \cap L^{p_2}$. Suppose that for $q_1, q_2 \in [1,\infty]$ there exists finite constants C_1, C_2 such that for all $f \in L^{p_1} \cap L^{p_2}$,*

$$\|Tf\|_{q_1} \le C_1 \|f\|_{p_1} \quad \text{and} \quad \|Tf\|_{q_2} \le C_1 \|f\|_{p_2} \ . \tag{6.3.27}$$

Then for all $\lambda \in [0,1]$, T extends to a bounded linear transformation from $L^{p(\lambda)}$ to $L^{q(\lambda)}$ where

$$\frac{1}{p(\lambda)} = (1-\lambda)\frac{1}{p_1} + \lambda \frac{1}{p_2} \quad \text{and} \quad \frac{1}{q(\lambda)} = (1-\lambda)\frac{1}{q_1} + \lambda\frac{1}{q_2} \ , \tag{6.3.28}$$

and for all $f \in L^{p(\lambda)}$,

$$\|Tf\|_{q(\lambda)} \le C_1^{1-\lambda} C_2^{\lambda} \|f\|_{p(\lambda)} \ . \tag{6.3.29}$$

This has a number of consequences; we give two of these in which the measure space is $(\mathbb{R}^n, \mathcal{B}, \mathrm{d}x)$ before giving the proof.

6.3.12 THEOREM (Huassdoerff-Young-Titschmarsh). *Define the Fourier transform $f \mapsto \widehat{f}$ on $L^1(\mathbb{R}^n, \mathcal{B}, \mathrm{d}x) \cap L^2(\mathbb{R}^n, \mathcal{B}, \mathrm{d}x)$ by*

$$\widehat{f}(k) = \int_{\mathbb{R}^n} e^{-i2\pi k\cdot x} f(x)\mathrm{d}x \ . \tag{6.3.30}$$

Then for all $p \in [1,2]$, $f \mapsto \widehat{f}$ extends to a contraction from $L^p(\mathbb{R}, \mathcal{B}, \mathrm{d}x)$ to $L^q(\mathbb{R}, \mathcal{B}, \mathrm{d}x)$ where $q = p/(p-1)$.

Proof. It is evident that $|\widehat{f}(k)| \le \int_R |f(x)|\mathrm{d}x = \|f\|_1$ so that $\||\widehat{f}\|_\infty \le \|f\|_1$. By the unitarity of the Fourier transform on L^2, $\|\widehat{f}\|_2 = \|f\|_2$. Therefore (6.3.27) is valid with $p_1 = 1$, $q_1 = \infty$, $C_1 = 1$, and $p_2 = q_2 = 1$, $C_2 = 1$.

For all $p \in (1,2)$ write

$$\frac{1}{p} = \frac{1}{p(\lambda)} = (1-\lambda)1 + \lambda\frac{1}{2} = 1 + \frac{\lambda}{2} \quad \text{and hence} \quad \lambda = 2\left(1 - \frac{1}{p}\right) = \frac{2}{q} \ .$$

Then, with this value of λ, $1/q(\lambda) = (1-\lambda)(1/\infty) + \lambda(1/2) = 1/q$. Since $C_1 = C_2 = 1$, (**??**) becomes $\|\widehat{f}\|_q \le \|f\|_p$. □

6.3.13 THEOREM (Young's Inequality for Convolution). *For all $f, g \in L^1(\mathbb{R}, \mathcal{B}, \mathrm{d}x) \cap L^\infty(\mathbb{R}, \mathcal{B}, \mathrm{d}x)$ and all $r, s, t \in [1,\infty]$ such that*

$$\frac{1}{s} + \frac{1}{t} = 1 + \frac{1}{r} \ , \tag{6.3.31}$$

$$\|f * g\|_r \le \|f\|_s \|g\|_t \ . \tag{6.3.32}$$

Proof. Fix $g \in L^1 \cap L^\infty$ and let T denote the operator on $L^1 \cap L^\infty$ defined by

$$Tf = g * f \ .$$

Then for $t \in (1,\infty)$, the inequality $\|g * f\|_t \le \|f\|_1 \|g\|_t$ and be written as $\|Tf\|_t \le \|g\|_p \|f\|_1$, while the inequality $\|g * f\|_\infty \le \|f\|_{t/(t-1)} \|g\|_t$ can be written as $\|Tf\|_\infty \le \|g\|_t \|f\|_{t/(t-1}$. Therefore (6.3.27) is valid with

$$p_1 = 1 \ , \ q_1 = t \ , \ C_1 = \|g\|_p \quad \text{and} \quad p_2 = t/(t-1) \ , \ q_2 = \infty \ , \ C_2 = \|g\|_t \ .$$

Note from (6.3.31) that $s \le t/(t-1)$ with equality if and only if $r = \infty$. Therefore, $s \in [p_1, p_2]$. Define $\lambda \in [0,1]$ so that

$$\frac{1}{s} = \frac{1}{p(\lambda)} = (1-\lambda) + \lambda\frac{t-1}{t} \quad \text{and hence} \quad 1-\lambda = \frac{1}{t}\left(\frac{1}{s} + \frac{1}{t} - 1\right) .$$

Then for this λ,

$$\frac{1}{q(\lambda)} = (1-\lambda)\frac{1}{t} + \lambda\frac{1}{\infty} = \frac{1}{s} + \frac{1}{t} - 1 = \frac{1}{r} .$$

It follows from the Riesz-Thorin Theorem that T extends to map from L^s to L^r with norm $\|g\|_t$, and this yields (6.3.32). □

We now turn to the basic lemma upon which the Riesz-Thorin Theorem rests. For $x \in \mathbb{C}^n$, $p \in [1,\infty]$, let $\|x\|_p$ denote the ℓ^p norm on $\mathbb{C}^n$. Consider the bilinear form $A(x,y)$ on C^n given by

$$A(x,y) = \sum_{j,k=1}^{n} x_j a_{j,k} y_k .$$

For $1 \le p, q \le 1$, define

$$C(1/p, 1/q) = \sup\left\{|A(x,y)| \ : \ \|x\|_p = \|y\|_q = 1\right\} .$$

It will be convenient to write $\alpha = \dfrac{1}{p}$ and $\beta = \dfrac{1}{q}$. Notice then that $0 \le \alpha, \beta \le 1$.

6.3.14 LEMMA. *[Riesz–Thorin Interpolation Lemma] The function* $(\alpha,\beta) \mapsto \ln\left(C(\alpha,\beta)\right)$ *is convex on* $[0,1] \times [0,1]$.

6.3.15 Remark. Marcel Riesz proved this in 1927 for $\alpha + \beta \ge 1$; i.e., in the upper right half of the unit square. His method was a direct assault on the maximization problem using Lagrange multipliers. This approach worked only in the restricted domain $\alpha + \beta \ge 1$, and he even conjectured that this wasn't die to his method, but that the result wasn't true in general except in this case. About a decade later, his student Thorin proved the full result by a completely different method. It is the lower–left half of the square that is of most interest to us.

Proof. Fix any two pairs of numbers (α_1, β_1) and (α_2, β_2) in $[0,1] \times [0,1]$. Let $p_1 = 1/\alpha_1$, $q_1 = 1/\beta_1$, $p_2 = 1/\alpha_2$ and $q_2 = 1/\beta_2$ with the obvious meaning if any of the denominators vanish. Next, any x with $\|x\|_{p_1} = 1$ can be written as

$$(x_1, x_2, \ldots, x_n) = (e^{i\phi_1} b_1^{\alpha_1}, e^{i\phi_2} b_2^{\alpha_1}, \ldots, e^{i\phi_1} b_n^{\alpha_1}) ,$$

where each $b_j \ge 0$ and $\sum_{j=1}^{n} b_j = 1$. Let $\mathcal{P}_n$ denote the set of all n–tuples $(d_1, d_2, \ldots, d_n)$ of non–negative numbers such that $\sum_{j=1}^{n} d_j = 1$, so that we may express the condition on the finite sequence $\{b_j\}$ as $\{b_j\} \in \mathcal{P}_n$. Similarly y with $\|y\|_{q_1} = 1$ can be written as

$$(y_1, y_2, \ldots, y_n) = (e^{i\psi_1} c_1^{\beta_1}, e^{i\psi_2} c_2^{\beta_1}, \ldots, e^{i\psi_1} c_n^{\beta_1}) ,$$

where $\{c_j\} \in \mathcal{P}_n$.

Then we can rewrite the definition of $C(\alpha_1, \beta_1)$ as

$$C(\alpha_1,\beta_1) = \sup\left\{ \sum_{j,k=1}^{n} a_{j,k} e^{i(\psi_k - \phi_j)} b_j^{\alpha_1} c_k^{\beta_1} \ : \ \{b_j\},\{c_k\} \in \mathcal{P}_n, \{\phi_j\}, \{\psi_k\} \right\} .$$

It will be convenient to take the supremum in two parts, as follows:

$$C(\alpha_1,\beta_1) = \sup\left\{ C(\alpha_1,\beta_1,\{\phi_j\},\{\psi_k\}) \ : \ \{\phi_j\},\{\psi_k\} \right\}$$

where

$$C(\alpha_1,\beta_1,\{\phi_j\},\{\psi_k\}) = \sup\left\{ \sum_{j,k=1}^{n} \tilde{a}_{j,k} b_j^{\alpha_1} c_k^{\beta_1} \mid \{b_j\},\{c_k\} \in \mathcal{P}_n \right\} \quad \text{and} \quad \tilde{a}_{j,k} = a_{j,k} e^{i(\psi_k - \phi_j)} .$$

We can do the same for (α_2) and (β_2), and all we are changing is the exponents in the sum $\sum_{j,k=1}^{n} \tilde{a}_{j,k} b_j^{\alpha} c_k^{\beta}$. In particular, we optimize over the *same* sets of sequences $\{b_j\}$, $\{c_k\}$, $\{\phi_j\}$ and $\{\psi_k\}$ in each case. Thorin's idea is now to interpolate between (α_1, β_1) and (α_2, β_2) with a *complex variable* z as follows: For all z belonging to the infinite strip given by

$$0 \le \mathcal{R}(z) \le 1$$

define

$$f(z) = \frac{\sum_{j,k=1}^{n} \tilde{a}_{j,k} b_j^{(1-z)\alpha_1 + z\alpha_2} c_k^{(1-z)\beta_1 + z\beta_2}}{C(\alpha_1,\beta_1)^{1-z} C(\alpha_2,\beta_2)^{z}} .$$

Now for any positive number a, $a^z = e^{(\ln a)z}$ is an analytic function on the entire complex plane, and has no zeros. For this reason, $f(z)$ is an analytic function everywhere on its domain $0 \le \mathcal{R}(z) \le 1$. Also, on the left boundary of this domain; i.e., where $\mathcal{R}(z) = 0$,

$$|f(z)| = \frac{\left| \sum_{j,k=1}^{n} a_{j,k} e^{i(\tilde{\psi}_k - \tilde{\phi}_j)} b_j^{\alpha_1} c_k^{\beta_1} \right|}{C(\alpha_1,\beta_1)}$$

where the $\tilde{\phi}_j$ and $\tilde{\psi}_k$ include additional phases involving the imaginary part of z. By definition, the numerator is no more than $C(\alpha_1, \beta_1)$, and hence we arrive at the conclusion that $|f(z)| \le 1$ for all z with $\mathcal{R}(z) = 0$. That is,

$$\sup_{y \in R} |f(iy)| \le 1 .$$

In the exact same way, we see that $|f(z)| \le 1$ for all z with $\mathcal{R}(z) = 1$. That is,

$$\sup_{y \in R} |f(1+iy)| \le 1 .$$

Now we are in a position to apply the *maximum modulus principle* of complex analysis to conclude that $|f(z)|$ is largest on the boundary of the strip $0 \le \mathcal{R}(z) \le 1$, and hence that $|f(z)| \le 1$ for all such z.

We will come back to the maximum modulus principle argument shortly to make the proof self contained, but let us suppose for the moment that it has been established that $|f(z)| \le 1$ whenever

$0 \le \mathcal{R}(z) \le 1$. We are now done with the complex interpolation; take $z = \lambda$ with $0 < \lambda < 1$. Then $|f(\lambda)| \le 1$ implies that

$$\left|\sum_{j,k=1}^{n} \tilde{a}_{j,k} b_j^{(1-\lambda)\alpha_1+\lambda\alpha_2} c_k^{(1-\lambda)\beta_1+\lambda\beta_2}\right| \le C(\alpha_1,\beta_1)^{1-\lambda} C(\alpha_2,\beta_2)^{\lambda} \ .$$

Since this is true for each $\{b_j\}, \{c_k\} \in \mathcal{P}_n$, and each $\{\phi_j\}$ and $\{\psi_k\}$, we can take the supremum to arrive at

$$C((1-\lambda)\alpha_1 + \lambda\alpha_2, (1-t)\beta_1 + t\beta_2) \le C(\alpha_1,\beta_1)^{1-\lambda} C(\alpha_2,\beta_2)^{\lambda}$$

and taking the natural logarithm, we obtain the stated convexity result.

We now return to the part of the argument involving the maximum modulus principle. First modify $f(z)$ by multiplying by $e^{z^2/n-1/n}$. For $z = x + iy$ with $0 \le x \le 1$, the real part of z^2 is $x^2 - y^2 \le 1$, and so for all z with $0 \le \mathcal{R}(z) \le 1$, $|e^{z^2/n-1/n}| \le 1$. In fact, the real part of z^2 becomes strongly negative for y large, and so $g(z) = f(z)e^{z^2/n-1/n}$ vanishes as z tends to infinity in $0 \le \mathcal{R}(z) \le 1$, and it is also clearly analytic there. In particular, it is continuous, and so there is a point z_0 in this domain so that

$$|g(z_0)| \ge |g(z)| \tag{6.3.33}$$

for all z with $0 \le \mathcal{R}(z) \le 1$. We claim that a point z_0 satisfying (6.3.33) lies on the boundary of the strip. To see this, consider *any* point z_0 satisfying 6.3.33). If it is not already on the boundary, appeal to the Cauchy integral formula

$$g(z_0) = \frac{1}{2\pi i}\int_0^{2\pi} g(z_0 + re^{i\phi})\mathrm{d}\phi$$

where r is the distance from z_0 to the boundary; i.e., $r = \min\{\mathcal{R}(z), 1 - \mathcal{R}(z)\}$. Then

$$|g(z_0)| \le \frac{1}{2\pi}\int_0^{2\pi} |g(z_0 + re^{i\phi})|\mathrm{d}\phi \le \frac{1}{2\pi}\int_0^{2\pi} |g(z_0)|\mathrm{d}\phi = |g(z_0)|$$

where in the first inequality we simply to absolute values, and in the second one we used (6.3.33) with $z = z_0 + re^{i\phi}$. The conclusion is that $|g(z_0 + re^{i\phi})| = |g(z_0)|$ for almost every ϕ, and then by continuity, for all ϕ. For either $\phi = 0$ or $\phi = \pi$, $z_0 + re^{i\phi}$ lies on the boundary, and the claim is proved. Hence $|g(z)|$ is maximized on the boundary, and so

$$|g(z)| = |f(z)e^{z^2/n-1/n}| \le 1$$

for all $0 \le \mathcal{R}(z) \le 1$. Taking the limit in which n tends to infinity, we obtain the desired result. □

Proof of Theorem 6.3.11. Given any t with $0 < \lambda < 1$, and any simple function f, we we first observe that $f \in L^{p_1} \cap L^{p_2}$, and hence Tf is defined. By the variational characterization of the L^p norms,

$$\|T_f\|_{q(t)} = \sup\left\{\left|\int_{R^d} gTf\mathrm{d}\mu\right| \ : \ \|g\|_{q(t)'} = 1\right\}$$

where $1/q(t)' + 1/q(t) = 1$.

By the density of simple functions, for any $\epsilon > 0$, there is a simple function g with $\|g\|_{q(t)'} = 1$ so that

$$\|T_f\|_{q(t)} \le \left|\int_{R^d} gTf\mathrm{d}\mu\right| + \epsilon \ .$$

Now by further partitioning the subsets on which the simple functions f and g take on their various values, we may write them in the form

$$f(x) = \sum_{j=1}^{n} \frac{f_j}{\mu(A_j)^{1/p(t)}} 1_{A_j}(x) \qquad \text{and} \qquad g(x) = \sum_{k=1}^{n} \frac{g_k}{\mu(A_k)^{1/q'(t)}} 1_{A_k}(x)$$

with the same n and the same family of *disjoint* measurable sets A_j. Notice that with the way the coefficients are defined,

$$\|f\|_{p(t)} = \left(\sum_{j=1}^{n} |f_j|^{p(t)} \right)^{1/p(t)} \qquad \text{and} \qquad \|g\|_{q'(t)} = \left(\sum_{k=1}^{n} |g_k|^{q'(t)} \right)^{1/q'(t)}$$

Then $\displaystyle\int_{R^d} gTf\mathrm{d}\mu = \sum_{j,k=1}^{n} a_{j,k} f_j g_k$ where $a_{j,k} = \displaystyle\int_{R^d} 1_{A_k} T 1_{A_j} \mathrm{d}\mu$. This reduces the Corollary to the Reisz–Thorin Lemma and, since $\epsilon > 0$ is arbitrary, the theorem is proved. $\square$

Chapter 7

The Reisz-Markoff Theorem and Related Topics

7.1 Locally compact Hausdorff spaces

7.1.1 DEFINITION. A topological space $(X,\mathcal{U})$ is *locally compact* in case every point of X has a neighborhood with compact closure.

When $(X,\mathcal{U})$ is locally compact, every neighborhood U of x contain another neighborhood of x that has compact closure: Let V be any neighborhood of x that has compact closure, and then $V \cap U \subset U$ and $\overline{V \cap U} \subset \overline{V}$ which is compact.

For example, $\mathbb{R}^n$ is compact since for each x, $\overline{B_1(x)}$ is compact. Also, it is evident that any compact space is locally compact. However, an infinite dimensional Hilbert space with its norm topology is not locally compact: As we have seen, the closed unit ball – and thus any closed ball – in such a space fails to be compact, and every closed neighborhood must contain a closed ball.

In a locally compact Hausdorff space, one can separate compact sets K and points $y \in K^c$ as follows: For each $x \in K$, let V_x be a neighborhood of x with compact closure, and let U_x be a neighborhood of y such that $V_x \cap U_x = \emptyset$, which is possible since X is Hausdorff. Then $\{V_x \ : \ x \in K\}$ is an open cover of K so that there exist $\{x_1, \dots, x_n\} \subset K$ such that $K \subset \cup_{j=1}^n V_{x_j} =: V$. let $U = \cup_{j=1}^n U_{x_j}$. Since $\overline{V} \subset \cup_{j=1}^n \overline{V_{x_j}}$ which is compact, V has compact closure, and since $y \notin \overline{V_{x_j}}$ for any j, $y \notin \overline{V}$. In summary:

7.1.2 LEMMA. *Let $(X,\mathcal{U})$ be a locally compact Hausdorff space. Suppose $K \subset X$ is compact and $y \notin K$. Then there exists disjoint open sets V and U such that $K \subset V$, $y \in U$, and $\overline{V}$ is compact.*

The slightly more elaborate result in the next lemma is fundamental.

7.1.3 LEMMA. *Let $(X,\mathcal{U})$ be a locally compact Hausdorff space. Suppose $K \subset U \subset X$ with K compact and U open. Then there exists an open set V with compact closure such that*

$$K \subset V \subset \overline{V} \subset U \ . \tag{7.1.1}$$

Proof of Lemma 7.1.3. If $y \in U^c$, then $y \in K^c$, and using Lemma 7.1.2 we may choose for each $y \in U^c$ disjoint open sets V_y and W_y such that $K \subset V_y$, $y \in W_y$ and $\overline{V_y}$ is compact. Since $y \notin \overline{V_y}$,

$$\bigcap_{y \in U^c} U^c \cap \overline{V_y} = \emptyset \ .$$

and each set in the intersection is compact. Therefore there exist $\{y_1, \ldots, y_n\} \subset U^c$ such that

$$\bigcap_{j=1,\ldots,n} U^c \cap \overline{V_{v_j}} = \emptyset \ .$$

Define $V = \cap_{j=1}^n V_{y_j}$ which is open and contains K. Since $\overline{V} \subset \cap_{j=1}^n \overline{V_{y_j}}$, which is compact and disjoint from U^c, $\overline{V}$ is compact and $\overline{V} \cap U^c = \emptyset$, which is the same as $\overline{V} \subset U$. □

The main feature of locally compact Hausdorff spaces that makes it possible to develop a rich theory linking topology and integration for them is that locally compact Hausdorff spaces are rich in continuous functions in the sense that we now explain. We shall write

$$K \prec f$$

to mean that K is a compact subset of X, and that f is a continuous and compactly supported function on X with values in $[0,1]$ and with $f(x) = 1$ for all $x \in K$.

We shall write

$$f \prec U$$

to mean that U is a compact subset of X, and that f is a continuous and compactly supported function on X with values in $[0,1]$ and that the support of f is contained in U.

7.1.4 THEOREM (Urysohn's Lemma)**.** *Let $(X, \mathcal{U})$ be a locally compact Hausdorff space. Suppose $K \subset U \subset X$ with K compact and U open. Then there exists a continuous function f such that*

$$K \prec f \prec U \ . \tag{7.1.2}$$

Proof. First, pick an open G with compact closure such that $K \subset G \subset \overline{G} \subset U$, which we may do by Lemma 7.1.3. In the next step we construct a sequence of open sets $\{V_s\}$ indexed by the dyadic rational numbers s in $(0,1)$ such that for each $t > s$,

$$K \subset V_t \subset \overline{V_t} \subset V_s \subset G \ .$$

We proceed inductively. By what we have shown in the first step, there exists an open set $V_{1/2}$ such that

$$K \subset V_{1/2} \subset \overline{V_{1/2}} \subset G \ .$$

For the same reason, there exist open sets $V_{1/4}$ and $V_{3/4}$ such that

$$K \subset V_{3/4} \subset \overline{V_{3/4}} \subset V_{1/2} \qquad \text{and} \qquad \overline{V_{1/2}} \subset V_{1/4} \subset \overline{V_{1/4}} \subset G \ .$$

Now an obvious induction argument provides the construction.

Having constructed our sequence $\{V_s\}$, s ranging over the dyadic rationals in $(0,1)$, we are ready to define f: First, set $V_0 := X$. Next, for $x \in X$, define

$$f(x) = \sup\{s \ : \ x \in V_s\} \ .$$

If $x \in K$, then $x \in V_s$ for all s, and hence $f(x) = 1$, and if $x \in G^c$, then $f(x) = 0$. Hence the support of f is contained in $\overline{G}$, which is compact. Finally, we claim that f is continuous. It suffices to show that for all λ, $f^{-1}((\lambda,\infty))$ is open and that and $f^{-1}([\lambda,\infty))$ is closed.

First, we claim that

$$f^{-1}((\lambda,\infty)) = \bigcup_{s>\lambda} V_s \ ,$$

which is open. To see this, note that if $f(x) > \lambda$ then for some $s > \lambda$, $x \in V_s$, and so x belongs to the union on the right. On the other hand, if x belongs to the union on the right, then $x \in V_s$ for some $s > \lambda$, and then $f(x) > \lambda$

Second, we claim that

$$f^{-1}([\lambda,\infty)) = \bigcap_{s<\lambda} \overline{V_s} \ ,$$

To see this, note that if $f(x) \geq \lambda$, then for all $s < \lambda$, $x \in V_s$, and hence $x \in \overline{V_s}$, and so x belongs to the intersection on the right. On the other hand, if x belongs to the intersection on the right, then for all $r < \lambda$, there is an $s > r$ so that $x \in \overline{V_s}$. Since $\overline{V_s} \subset V_r$, $x \in V_r$ and $f(x) \geq r$. Since $r < \lambda$ is arbitrary, $f(x) \geq \lambda$. □

7.1.5 THEOREM (Continuous partitions of unity)**.** *Let $(X,\mathcal{U})$ be a locally compact Hausdorff space. Let K be a compact subset of X, and let $\{U_1,\dots,U_n\}$ be a finite open cover of K. Then there exist functions g_j, $j = 1,\dots,n$ such that $g_j \prec U_j$ for each j and such that $K \prec \sum_{j=1}^n g_j$.*

Proof. Each $x \in K$ belongs to some U_j, and applying Lemma 7.1.3 to $\{x\} \subset U_j$, we can choose an open set W_x with compact closure such that $x \in W_x \subset \overline{W_x} \subset U_j$. Since K is compact, we may cover K by finitely many such sets $\{W_{x_1},\dots,W_{x_m}\}$. Let C_ℓ be the union of those $\overline{W_{x_k}}$ that were constructed choosing $W_x \subset U_\ell$. Then C_ℓ is compact, and $C_\ell \subset U_\ell$, and $K \subset \cup_{\ell=1}^n C_\ell$.

By Urysohn's Lemma, there exists f_ℓ such that $C_\ell \prec f_\ell \prec U_\ell$. How consider the function

$$h := 1 - \prod_{j=1}^n (1 - f_j) \ .$$

Clearly, h is continuous and $0 \leq h \leq 1$. If $x \in K$, $x \in C_j$ for some j, and then $f_j(x) = 1$ so that $h(x) = 1$.

We next claim that

$$h = f_1 + \sum_{j=1}^{n-1} f_j \prod_{k=1}^{j-1} (1 - f_k) \ .$$

To see this note that

$$1 - \prod_{j=1}^n (1 - f_j) = f_1 + (1 - f_1) - \prod_{j=1}^n (1 - f_j) = f_1 + (1 - f_1)\left[1 - \prod_{j=2}^n (1 - f_j)\right] \ ,$$

and then apply the obvious induction argument.

Finally, we define

$$g_1 = f_1 \qquad \text{and} \qquad g_j = f_j \prod_{k=1}^{j-1}(1 - f_k) \quad \text{for} \quad j = 1, \dots, n-1 \ .$$

Thus, $h = \sum_{j=1}^n g_j$ and $g_j \prec U_j$ for each j. □

7.1.1 Some spaces of continuous functions

Given a topological space $(X, \mathcal{U})$, there are three natural normed vector spaces of continuous functions:

7.1.6 DEFINITION (Spaces of continuous functions)**.** Let $(X, \mathcal{U})$ be a topological space. We define:

(i) $\mathcal{C}_c(X)$ is the normed vector space of continuous, compactly supported functions f on X with values in $\mathbb{C}$ on which the norm, $\|\cdot\|_\infty$, is given by $\|f\|_\infty = \sup\{\ |f(x)|\ :\ x \in X\ \}$.

(ii) $\mathcal{C}_0(X)$ is the normed vector space of continuous functions f on X with values in $\mathbb{C}$ such that for each $\epsilon > 0$, there exists a compact set $K_{\epsilon,f}$ such that $|f(x)| < \epsilon$ for all x outside of $K_{\epsilon,f}$. The norm is once again $\|\cdot\|_\infty$, given by $\|f\|_\infty = \sup\{\ |f(x)|\ :\ x \in X\ \}$.

(iii) $\mathcal{C}_b(X)$ is the normed vector space of continuous functions f on X with values in $\mathbb{C}$ such that $\|f\|_\infty = \sup\{\ |f(x)|\ :\ x \in X\ \} < \infty$. The norm is once again $\|\cdot\|_\infty$.

When X is not compact, $\mathcal{C}_c(X)$ is not complete, but if $(X, \mathcal{U})$ is a locally compact Hausdorff space, then it is dense in $\mathcal{C}_0(X)$, and both $\mathcal{C}_0(X)$ and $\mathcal{C}_b(X)$ are Banach spaces.

7.1.7 THEOREM. *$\mathcal{C}_0(X)$ equipped with the sup norm is a Banach space. If X is a locally compact Hausdorff space, then the subspace $\mathcal{C}_c(X)$ is dense.*

Proof. If $\{f_n\}$ is a Cauchy sequence in $\mathcal{C}_0(X)$, then $\{f_n(x)\}$ is a Cauchy sequence in $\mathbb{C}$. Hence the limit $\lim_{n\to\infty} f_n(x)$ exists for each x, and we define a function f by

$$f(x) = \lim_{n\to\infty} f_n(x) \ .$$

Given the uniform convergence, it is easy to check, using an $\epsilon/3$ argument, that $f \in \mathcal{C}_0(X)$ so that $\mathcal{C}_0(X)$ is complete.

To see that $\mathcal{C}_c(X)$ is dense, pick $f \in \mathcal{C}_0(X)$ and $\epsilon > 0$. Let K_ϵ be a compact set such that $|f(x)| \leq \epsilon$ for all $x \notin K_\epsilon$. Because X is locally compact, it is possible to find an open set U containing K_ϵ such that U has compact closure. Then by Urysohn's Lemma, there exists a continuous function g with $K \prec g \prec U$. Then $fg = f$ on K and and $|fg - f| = |f||g-1| \leq |f|$ everywhere, so that $|fg - f| \leq \epsilon$ on K_ϵ^c. In particular, $\|fg - f\|_\infty \leq \epsilon$, and $fg \in \mathcal{C}_c(X)$. □

7.2 The Riesz-Markoff Theorem

7.2.1 Radon measures

Radon measures, are, roughly speaking, the class of Borel measures on a locally compact Hausdorff space for which the measure theory and the topology are “nicely compatible”.

7.2.1 DEFINITION (Inner and outer regularity)**.** Let $(X,\mathcal{U})$ be a topological space. A Borel measure μ on X is *outer regular* in case for each Borel set E

$$\mu(E) = \inf\{\ \mu(U)\ :\ E \subset U\ ,\ U \text{ open }\}\ . \tag{7.2.1}$$

A Borel measure μ is *inner regular* in case for each Borel set E

$$\mu(E) = \sup\{\ \mu(K)\ :\ K \subset E\ ,\ K \text{ compact }\}\ . \tag{7.2.2}$$

A Borel measure μ is *inner regular for open sets* in case (7.2.2) holds for all *open* E. A Borel measure is *regular* if it is both inner and outer regular.

7.2.2 DEFINITION (Radon measure)**.** A *Radon measure* on a topological space $(X,\mathcal{U})$ is a positive Borel measure μ on X such that $\mu(K) < \infty$ for all compact sets $K \subset X$ that is outer regular and inner regular for open sets.

The next results demonstrate the compatibility of the topology and the measure theory for Radon measures.

7.2.3 THEOREM. *Let $(X,\mathcal{U})$ be a locally compact Hausdorff space, and let μ be a Radon measure on X. Then $\mathcal{C}_c(X)$ is dense in $L^p(X,\mathcal{B},\mu)$ for all $p \in [1,\infty)$.*

Proof. Since $L^1 \cap L^\infty$ is dense in Lp for all $p \in [1,\infty)$, it suffices to deal with $p = 1$. We know that integrable simple functions are dense in $L^1(\mu)$. Therefore, it suffices to show that whenever E is a Borel set with $\mu(E) < \infty$, for all $\epsilon > 0$, there exists $f \in \mathcal{C}_c(X)$ such that

$$\int_X |f - 1_E|\mathrm{d}\mu \le \epsilon\ .$$

Since μ is outer regular, there exists U open with $E \subset U$ such that $\mu(U) \le \mu(E) + \epsilon$. Since μ is inner regular for open sets, there is a compact set $K \subset U$ such that $\mu(K) \ge \mu(U) - \epsilon/$.

By Urysohn's Lemma, there exists $f \in \mathcal{C}_c(X)$ such that $K \prec f \prec U$. But then $|f - 1_U| \le 1_{U\cap K^c}$, and so

$$\|f - 1_E\|_1 \le \|f - 1_U\|_1 + \|1_U - 1_E\|_1 \le \mu(U\backslash E) + \mu(U \cap K^c) \le 2\epsilon\ .$$

□

7.2.4 THEOREM. *Let $(X,\mathcal{U})$ be a locally compact Hausdorff space. Let μ and ν be a Radon measures on X such that*

$$\int_X f\mathrm{d}\mu = \int_X f\mathrm{d}\nu$$

for all $f \in C_c(X)$. Then $\mu = \nu$.

Proof. Let U be open, and let $f \prec U$. For all compact $K \subset U$, Urysohn's lemma provides g with $K \prec g \prec U$, and hence

$$\mu(K) \le \int_X g\mathrm{d}\mu = \int_X g\mathrm{d}\nu = \nu(U)\ .$$

By the inner regularity of μ on open sets, $\mu(U) = \sup\{\mu(K)\ :\ K \subset U, K \text{ compact}\}$. Hence $\mu(U) \le \nu(U)$ By symmetry, $\nu(U) \le \mu(U)$ so that $\mu(U) = \nu(U)$ for all open U, and then by the outer regularity of μ and ν, $\mu(E) = \nu(E)$ for all Borel sets E. □

7.2.5 THEOREM. *Let $(X,\mathcal{U})$ be a locally compact Hausdorff space. Every σ-finite Radon measure μ on X is regular.*

Proof. Let E be a Borel set with $\mu(E) < \infty$. Pick $\epsilon > 0$. By the outer regularity of μ, there is an open set U such that $E \subset U$ and $\mu(U) \le \mu(E) + \epsilon$. For the same reason, there is an open set V such that $U\backslash E \subset V$ and $\mu(V) \le \mu(U\backslash E) + \epsilon \le 2\epsilon$.

By the inner regularity of μ for open sets, there is a compact set $C \subset U$ such that $\mu(U) \le \mu(C) + \epsilon$.

Define $K := C \cap V^c$ which is compact. By construction, every point $x \in U$ that is not in E lies in V, so that $C \cap V^c \subset E$. Therefore,

$$\mu(K) \ge \mu(C) - \mu(V) \ge \mu(U) - 2\epsilon \ .$$

Hence $\mu(K) \le \mu(E) - 2\epsilon$, and since $\epsilon > 0$ is arbitrary, $\mu(E) = \sup\{\ \mu(K)\ :\ K \subset E\ ,\ K \text{ compact }\}$.

Finally, if $\mu(E) = \infty$, there is an increasing sequence of sets E_n with $\mu(E_n) < \infty$ whose union is E and such that $\lim_{n\to\infty} \mu(E_n) = \infty$. By what was proved above, within each E_n there exists a compact K_n with $\mu(K_n) \ge \mu(E_n) - 1$. Each K_n is contained in E and $\lim_{n\to\infty} \mu(K_n) = \infty$ □

7.2.2 The Riesz-Markov Theorem for locally compact Hausdorff spaces

Throughout this section, let $(X,\mathcal{U})$ be a locally compact Hausdorff space. The spaces $\mathcal{C}_c(X)$, $\mathcal{C}_0(X)$ and $\mathcal{C}_b(X)$ are more than topological spaces: They contain a distinguished *cone* of non-negative elements: We say that *a function f on X is non-negative* in case for each x, $f(x) \in \mathbb{R}$, and $f(x) \ge 0$. In this case we write $f \ge 0$.

Let L be a linear functional on $\mathcal{C}_c(X)$. We say that L *is a positive linear functional* on $\mathcal{C}_c(X)$ in case

$$f \ge 0 \quad \Rightarrow \quad L(f) \ge 0 \ .$$

Evidently. if f is real, $L(f)$ is real, and $|L(f)| \le L(|f|)$. If $f = g + ih$ where g and h are real, $|L(f)| = |L(g) + iL(h)| \le L(|f|)$.

There is a close connection between the topology on $\mathcal{C}_c(X)$ and the partial order structure on $\mathcal{C}_c(X)$ induced by its cone of positive elements.

7.2.6 THEOREM. *Let L be a positive linear functional on $\mathcal{C}_c(X)$. Then for each compact $K \subset X$, there exists a finite constant C_K such that*

$$|f| \prec K \quad \Rightarrow \quad |L(f)| \le C_K \|f\|_\infty \ .$$

Proof. The uniqueness is immediate from Theorem 7.2.4, and we turn to existence. By Lemma 7.1.3, there exists an open set U with compact closure $\overline{U}$ such that $K \subset U \subset \overline{U}$, and then by Urysohn's Lemma, there exists a continuous function φ on X such that $K \prec \varphi \prec U$.

Then evidently $\|f\|_\infty \varphi - |f| \ge 0$, and hence $L(\|f\|_\infty \varphi - |f|) \ge 0$. Thus,

$$|L(f)| \le L(|f|) \le L(\|f\|_\infty \varphi) = L(\varphi)\|f\|_\infty \ .$$

Thus,

$$\sup\{|L(f)|\ :\ |f| \prec K\ , \|f\|_\infty \le 1\ \} \le L(\varphi)\|f\|_\infty \ .$$

We may take $C_K = L(f)$, or, better yet, $C_K = \inf\{L(f)\ :\ K \prec f\ \}$. □

To construct an example of a positive linear functional on $\mathcal{C}_c(X)$, let μ be a Borel measure on X that is finite on every compact set $K \subset X$. If $f \in \mathcal{C}_c(X)$, let K denote the compact support of f. Then $0 \leq |f| \leq \|f\|_\infty 1_K$, and since $\mu(K) < \infty$, f is integrable. Thus, we may define

$$L_\mu(f) = \int_X f \mathrm{d}\mu \ ,$$

which is therefore a linear functional on $\mathcal{C}_c(X)$, and evidently it is positive.

The Reisz-Markov Theorem asserts that *every* example is of this type. Moreover, we shall show that one consequence of the Reisz-Markov Theorem is that every Borel measure on X such that $\mu(K) < \infty$ for all compact K automatically has certain regularity properties:

7.2.7 THEOREM (Riesz-Markov Theorem). *Let L be any positive linear functional on $\mathcal{C}_c(X)$. Then there exists a unique Radon measure μ such that*

$$L(f) = \int_X f \mathrm{d}\mu \quad \text{for all} \quad f \in \mathcal{C}_c(X) \ . \tag{7.2.3}$$

Moreover,

$$\mu(U) = \sup\{ \ L(f) \ : \ f \prec U \ , f \in \mathcal{C}_c(X) \ \} \tag{7.2.4}$$

for all open sets U, and

$$\mu(K) = \inf\{ \ L(f) \ : \ K \prec f \ , f \in \mathcal{C}_c(X) \ \} \tag{7.2.5}$$

for all compact sets K.

Proof. Step 1: Use L to construct an outer measure μ^*. We define a set function μ^* on open subsets of X by

$$\mu^*(U) = \sup\{ \ L(f) \ : \ f \prec U \ , f \in \mathcal{C}_c(X) \ \}$$

for open sets U, and then on arbitrary subsets E of X by

$$\mu^*(E) = \inf\{ \ \mu^*(U) \ : \ E \subset U \ , \ U \text{ open} \ \} \ .$$

It is clear that $\mu^*(\emptyset) = 0$, and that if $A \subset B$, then $\mu^*(A) \leq \mu^*(B)$. Therefore, to show that μ^* is an outer measure, we must show that for any sequence $\{E_n\}_{n\in\mathbb{N}}$ of subsets of X,

$$\mu^* \left(\bigcup_{n=1}^\infty E_n \right) \leq \sum_{n=1}^\infty \mu^*(E_n) \ .$$

Let E denote $\cup_{n=1}^\infty E_n$. It suffices to consider the case in which $\mu^*(E_n) < \infty$ for all n.

Pick any $\epsilon > 0$. Then by construction, there exists an open set U_n with $E_n \subset U_n$, and $\mu^*(U_n) \leq \mu^*(E_n) + 2^{-n}\epsilon$. But then

$$E \subset U := \bigcup_{n=1}^\infty U_n \qquad \text{and} \qquad \sum_{n=1}^\infty \mu^*(U_n) \leq \sum_{n=1}^\infty \mu^*(E_n) + \epsilon \ .$$

It therefore suffices to prove that

$$\mu^*\left(U\right) \leq \sum_{n=1}^\infty \mu^*(U_n) \ . \tag{7.2.6}$$

To do this, consider any $f \in \mathcal{C}_c$ such that $f \prec U$. Let K denote the support of f. Then $\{U_n\}_{n\in\mathbb{N}}$ is an open cover of K, and so there exists a finite sub cover, which we may take to be $\{U_1, \dots, U_N\}$.

Let $\{h_1, \dots, h_N\}$ be a partition of unity on K, subordinate to the open cover $\{U_1, \dots, U_N\}$. Then $f = \sum_{n=1}^{N} f h_n$ and $f h_n \prec U_n \quad n = 1, \dots, N$. It follows that

$$L(f) = \sum_{n=1}^{N} L(fh_n) \le \sum_{n=1}^{N} \mu^*(U_n) \le \sum_{n=1}^{\infty} \mu^*(U_n) \ .$$

Since, $f \in \mathcal{C}_c$ such that $f \prec U$ is arbitrary, we obtain (7.2.6).

Step 2: Caratheodory σ-algebra contains all open sets, and hence all Borel sets.

Let U be open, and let $E \subset X$ be arbitrary. We must show that

$$\mu^*(E) \ge \mu^*(E \cap U) + \mu^*(E \cap U^c) \ . \tag{7.2.7}$$

Since $\mu^*(E)$ is the infimum of $\mu^*(V)$, V open with $E \subset V$, it suffices to prove (7.2.7) when $E = V$, where V is open. We may also suppose that $\mu^*(V) < \infty$. Note that $U \cap V$ is open. Hence for any $\epsilon > 0$, there is an $f \prec V \cap U$ so that $L(f) \ge \mu^*(V \cap U) - \epsilon$. Let K denote the support of f. Since $K \subset U$, $U^c \subset K^c$, and so $V \cap U^c \subset V \cap K^c$, which is open. Choose $g \prec V \cap K^c$ so that $L(g) \ge \mu^*(V \cap K^c) - \epsilon$. Then $f + g$ has compact support contained in V and since the supports of f and g are disjoint, $f + g \prec V$. Therefore,

$$\mu^*(V) \ge L(f+g) = L(f) + L(g) \ge \mu^*(V \cap U) + \mu^*(V \cap U^c) - 2\epsilon \ .$$

At this point, we know that the Caratheodory σ-algebra contains the Borel σ-algebra $\mathcal{B}(X)$, and that the restriction of μ^* to $\mathcal{B}(X)$ is countably additive. We define μ to be this restriction.

Step 3: Compact sets have finite measure

Let U be any open set containing K such that U has compact closure. Such sets exist by Lemma 7.1.3. By Urysohn's Lemma, there exists an $f \in \mathcal{C}_c(X)$ such that $\overline{U} \prec f$. Thus, if $g \prec U$, $g \le f$, and so $L(g) \le L(f)$, and hence $\mu(U) \le L(f)$ since $g \prec U$ is arbitrary. Therefore, $\mu(K) \le \mu(U) < \infty$.

Step 4: μ is inner regular for open sets

Let K be compact. We claim that for K compact,

$$\mu(K) = \inf\{L(f) \ : \ K \prec f \ \} \tag{7.2.8}$$

Fix $\epsilon > 0$. Then there exists an open set U so that $K \subset U$ and $\mu(K) \ge \mu(U) - \epsilon$. By Urysohn's Lemma, there exists a function f with $K \prec f \prec U$. Then $L(f) \le \mu(U) \le \mu(K) + \epsilon$. Since $\epsilon > 0$ is arbitrary,

$$\mu(K) \ge \inf\{L(f) \ : \ K \prec f \ \} \ .$$

Next, suppose that $K \prec f$. Then for any $0 < \epsilon < 1$, Let $U_\epsilon = \{x \ : \ f(x) > 1 = \epsilon\}$. Then U_ϵ is open and has compact support, and $K \subset U_\epsilon$.

There exists a function g such that $g \prec U_\epsilon$ and $L(g) \ge \mu(U_\epsilon) - \epsilon$. Note that $f \ge fg \ge (1-\epsilon)g$. Therefore

$$L(f) \ge (1-\epsilon)L(g) \ge (1-\epsilon)[\mu(U_\epsilon) - \epsilon] \ge (1-\epsilon)[\mu(K) - \epsilon] \ .$$

Since $0 < \epsilon < 1$ is arbitrary, $\mu(K) \leq L(f)$ whenever $K \prec f$. This completes the proof of (7.2.8).

Now let V be open. Suppose $\mu(V) < \infty$. Then for every $\epsilon > 0$, there exists a function $f \prec V$ such that $\mu(V) - \epsilon \leq L(f)$. Let K be the support of f. By (7.2.8), there is a function g with $K \prec g$ such that $L(g) \leq \mu(K) + \epsilon$. But since $f \leq 1_K \leq g$,

$$\mu(V) - \epsilon \leq L(f) \leq L(g) \leq \mu(K) + \epsilon \ .$$

The same sort of reasoning shows that if $\mu(V)$ is infinite, we can find compact sets in V of arbitrarily large measure.

Step 5: For $f \in \mathcal{C}_c(X)$, $\int_X f \mathrm{d}\mu = L(F)$.

It suffices to treat the case in which f takes values in $[0,1]$. For each $t \in (0,1]$, define $K_t = \{x \ : f(x) \geq t\}$. Since $f \in \mathcal{C}_c(X)$, each K_t is compact Let K_0 denote the support of f.

For any $0 \leq a < b \leq 1$, define $f_{[a,b]} = (f-a)_+ \wedge b$. Then for all $0 < t < a$,

$$K_b \prec \frac{1}{b-a} f_{[a,b]} \leq 1_{K_a} \ .$$

Therefore, for all open U with $K_a \subset U$, $\mu(K_b) \leq \dfrac{1}{b-a} L(f_{[a,b]})\mu(U)$. By outer regularity,

$$\mu(K_b) \leq \frac{1}{b-a} L(f_{[a,b]}) \leq \mu(K_a) \ .$$

Then for any $n \in \mathbb{N}$, $f = \displaystyle\sum_{j=1}^n f_{[(j-1)/n, j/n]}$, and so $L(f) = \displaystyle\sum_{j=1}^n L\left(f_{[(j-1)/n, j/n]}\right)$. Therefore,

$$\sum_{j=1}^n \frac{1}{n}\mu(K_{j/n}) \leq L(f) \leq \sum_{j=1}^n \frac{1}{n}\mu(K_{(j-1)/n}) \ .$$

Since $\lim_{n\to\infty}\left(\sum_{j=1}^n \frac{1}{n}\mu(K_{j/n})\right) = \lim_{n\to\infty}\left(\sum_{j=1}^n \frac{1}{n}\mu(K_{j/n})\right) = \int_X f \mathrm{d}\mu$, the proof is complete. □

7.2.8 DEFINITION (Radon space)**.** A topological space $(X, \mathcal{U})$ is a *Radon space* in case every Borel measure on X that is finite on all compact sets is regular, and hence a Radon measure.

7.2.9 DEFINITION (Polish space)**.** A topological space $(X, \mathcal{U})$ is a *Polish space* in case it is separable and homeomorphic to a metric space.

The term "Polish space" recognizes the work of a group of Polish mathematicians including Kuratowski, Sierpinski and Tarski, who proved a number results pertaining to the concept. Note that a Polish space is necessarily Hausdorff. Therefore, a locally compact Polish space is, in particular, a locally compact Hausdorff space. For example, $\mathbb{R}^n$ is a locally compact Polish space, as is any Riemannain manifold.

7.2.10 LEMMA. *Let $(X, \mathcal{U})$ be a locally compact Polish space. Then every open set U in X is the countable union of compact sets in X. In particular, X is the countable union of compact sets.*

Proof. Let ρ be a metric that induces the topology on $(X, \mathcal{U})$, and let $B(r,x)$ denote the open ball of radius $r > 0$ about $x \in X$. Since $(X, \mathcal{U})$ is locally compact, for each $x \in X$, there exists some $r > 0$ $\overline{B(r,x)}$ is compact . (It is not excluded that $\overline{B(r,x)}$ is compact for *all* $r > 0$; this is the case in $\mathbb{R}^n$, for example.) Define $r_x = \sup\{r > 0 \ : \ \overline{B(r,x)} \text{ is compact}\}$. If $0 < \delta < r_x$ and $\rho(y,x) < \delta$, then for all r with $\delta < r < r_x$, $B(r-\delta, y) \subset B(r,x)$ and $\overline{B(r,x)}$ is compact. Hence $r_y \geq r_x - \delta$.

Likewise, given the open set U and $x \in U$ define $d_x = \sup\{s > 0 \ : \ B(s,x) \subset U\}$. That is, d_x is the distance from x to U^c. Just as above, one shows that if $\rho(y,x) < \delta < d_x$, then $d_y \geq d_x - \delta$.

Let $\{x_n\}_{n\in\mathbb{N}}$ be a dense sequence in the open set U. For $n \in \mathbb{N}$, define

$$K_n := \overline{B(\min\{d_{x_n} + r_{x_n}\}/2, x_n)} \ .$$

Then K_n is compact and contained in U. For any $x \in U$, there is some n such that $\rho(x_n, x) < \frac{1}{4}\min\{d_x, r_x\}$. Then $r_{x_n} \geq \frac{3}{4}r_x$ and $d_{x_n} \geq \frac{3}{4}d_x$, so that $\min\{d_{x_n} + r_{x_n}\}/2 > \rho(x_n, x)$. It follows that $x \in K_n$, Since $x \in U$ is arbitrary, $U = \cup_{n=1}^\infty K_n$. □

7.2.11 THEOREM. *Let $(X,\mathcal{U})$ be a locally compact Polish space. Then $(X,\mathcal{U})$ is a Radon space.*

Proof. Let μ be any Borel measure on X that is finite on every compact set. Then each $f \in C_c(X)$ is integrable with respect to μ, and hence $f \mapsto \int_X f \mathrm{d}\mu =: L(f)$ is a well-defined positive linear functional on $C_c(X)$. By the Riesz-Markov Theorem, there exists a Radon measure ν such that for all $f \in C_c(X)$,

$$\int_X f\mathrm{d}\mu = \int_X f\mathrm{d}\nu \ . \tag{7.2.9}$$

Since by Lemma 7.2.10, $(X, \mathcal{B}, \nu)$ is σ-finite, Theorem 7.2.5 then says that ν is regular. We next show that μ and ν agree on all open sets U.

Let U be open and write $U = \cup_{n=1}^\infty K_n$ where K_n is compact, which is possible by Lemma 7.2.10. By Uryson's lemma, for each $n \in \mathbb{N}$ there exists $f_n \in C_c(X)$ such that $\cup_{j=1}^n K_n \prec f_n \prec U$, Then define $g_n := \max\{f_1, \dots, f_n\}$. Then $g_n \uparrow 1_U$.

We have seen that there is a monotone increasing sequence $\{g_n\}_{n\in\mathbb{N}}$ with $g_n \uparrow 1_U$. By the Lebesgue Monotone Convergence Theorem and (7.2.9),

$$\mu(U) = \lim_{n\to\infty} \int_X g_n \mathrm{d}\mu = \lim_{n\to\infty} \int_X g_n \mathrm{d}\nu = \nu(U) \ .$$

Hence μ and ν agree on all open sets.

Now let K be compact. By Lemma 7.1.3, there is an open set V such that $K \subset V$ and $\overline{V}$ is compact. It follows that $\nu(V) = \mu(V) < \infty$. Then since $K = V\backslash(V\backslash K)$, and since V and $V\backslash K$ are both open

$$\mu(K) = \mu(V) - \mu(V\backslash K) = \nu(V) - \nu(V\backslash K) = \nu(K) \ .$$

Hence μ and ν agree on all compact sets. Let E be any Borel set. Since ν is regular,

$$\nu(E) = \sup\{\nu(K) \ : \ K \subset E \ , K \text{ compact } \} = \sup\{\mu(K) \ : \ K \subset E \ , K \text{ compact } \} \leq \mu(E) \ .$$

Hence if $\nu(E)$ is infinite, so is $\mu(E)$. Suppose that $\nu(E) < \infty$, and pick $\epsilon > 0$. By the regularity of ν, there exist K compact and U open such that $K \subset E \subset U$ and $\nu(U) - \nu(K) < \epsilon$. Then because $\mu(K) = \nu(K)$ and $\mu(U) = \nu(U)$, both $\mu(E)$ and $\nu(E)$ lie in the interval $[\nu(K), \nu(K) + \epsilon]$, and hence $|\mu(E) - \nu(E)| \leq \epsilon$. Since $\epsilon > 0$, $\mu(E) = \nu(E)$. Hence μ and ν agree on all Borel sets. □

7.2.3 The Hahn-Saks Theorem

Throughout this section, $(X,\mathcal{U})$ is a locally compact Hausdorff space, $\mathcal{C}_c(X)$ denotes the *real* normed space of continuous compactly supported *real valued* functions on X, and $\mathcal{C}_0(X)$ denotes the *real* Banach space consisting of all uniform limits of functions in $\mathcal{C}_c(X)$.

The real vector spaces $\mathcal{C}_c(X)$ and $\mathcal{C}_0(X)$ are *ordered* vector spaces: We say $f \geq g$ in case $f(x)-g(x) \geq 0$ for all x. Let $\mathcal{C}_c^+(X)$ and $\mathcal{C}_0^+(X)$ denote the sets of point-wise non-negative functions in $\mathcal{C}_c(X)$ and $\mathcal{C}_0(X)$ respectively. Then $f \geq g$ in $\mathcal{C}_c(X)$ if and only if $f - g \in \mathcal{C}_c^+(X)$, and likewise for the order in $\mathcal{C}_0(X)$.

Our goal is to concretely identify the elements of the dual space $(\mathcal{C}_0(X))^*$ in terms of measures on X. An element of $(\mathcal{C}_0(X))^*$ is *positive* in case $L(f) \geq 0$ wherever f is a non-negative function in $\mathbb{C}_0(X)$. Restricting such a functional L to $C_c(X)$, yields a positive linear functional on $C_c(X)$, and then by the Riesz-Markoff Theorem, there is a Radon measure μ_L on X and such that

$$L(f) = \int_X f \mathrm{d}\mu_L \tag{7.2.10}$$

for all $f \in \mathcal{C}_c(X)$.

7.2.12 LEMMA. *Let L be a positive linear functional on $\mathbb{C}_0(X)$, and let μ_L be the Radon measure on X such that (7.2.10) is valid for all $f \in \mathbb{C}_c(X)$. Then $\mu_L(X) \leq \|L\|$, and hence (7.2.10) extends by continuity to all of $C_0(X)$.*

Proof. Since μ_L is inner regular on open sets, there exists an increasing sequence $\{K_n\}_{n\in\mathbb{N}}$ such that $\mu_L(K_n) \uparrow \mu_L(X)$. By Urysohn's Lemma, for each n there exists $f_n \in V_c(X)$ such that $K_n \prec f_n$. Then $\mu_L(K_n) \leq \displaystyle\int_X f_n \mathrm{d}\mu_L = L(f_n) \leq \|L\|$. □

Suppose that $L \in (\mathcal{C}_0(X))^*$ can be written as the difference of two *positive* linear functionals,

$$L = L_1 - L_2 \ . \tag{7.2.11}$$

Then by Lemma 7.2.12, there are finite Radon measures μ_1 and μ_2 such that for all $f \in C_0(X)$,

$$L(f) = \int_X f \mathrm{d}\mu_1 - \int_X f \mathrm{d}\mu_2 \ . \tag{7.2.12}$$

Define $\nu = \mu_1 + \mu_2$, which is also a finite Radon measure. Evidently μ_1 and μ_2 are absolutely continuous with respect to ν, and hence there exist non-negative functions $h_1, h_2 \in L^1(X,\mathcal{B},\nu)$ such that $\mathrm{d}\mu_1 = h_1\mathrm{d}\nu$ and $\mathrm{d}\mu_2 = h_2\mathrm{d}\nu$. Define $\widetilde{h}_1 := h_1 - h_1 \wedge h_2$ and $\widetilde{h}_2 := h_2 - h_1 \wedge h_2$. Let $A = \{x \ : \ h_1 \wedge h_2(x) = h_1(x)\}$ and observe that $\widetilde{h}_1(x) = 0$ for all $x \in A$, while $\widetilde{h}_2(x) = 0$ for all $x \in A^c$. Then since $h_1 - h_2 = \widetilde{h}_1 - \widetilde{h}_2$, (7.2.12) becomes

$$L(f) = \int_X f h_1 \mathrm{d}\nu - \int_X f h_2 \mathrm{d}\nu = \int_X f(\widetilde{h}_1 - \widetilde{h}_2)\mathrm{d}\nu \ . \tag{7.2.13}$$

Define measures $\mathrm{d}\widetilde{\mu}_1 := \widetilde{h}_1\mathrm{d}\nu$ and $\widetilde{\mu}_2 := \widetilde{h}_2\mathrm{d}\nu$. Then $\widetilde{\mu}_1(A) = 0$ and $\widetilde{\mu}_2(A^c) = 0$ so that $\widetilde{\mu}_1$ and $\widetilde{\mu}_2$ are mutually singular. Since any Borel measure that is absolutely continuous with respect to a Radon measure is itself as Radon measure, $\widetilde{\mu}_1$ and $\widetilde{\mu}_2$ are also Radon measures.

In summary, if there exists a decomposition of $L \in (\mathcal{C}_0(X))^*$ as the difference of two positive elements of $(\mathcal{C}_0(X))^*$, then there exists a pair of *mutually singular* finite Radon measures μ_+ and μ_- such that for all $f \in C_0(X)$,

$$L(f) = \int_X f \mathrm{d}\mu_+ - \int_X f \mathrm{d}\mu_- \ . \tag{7.2.14}$$

Such a decomposition is necessarily unique: Let A be a Borel set such that $\mu_+(A) = 0$ and $\mu_-(A^c) = 0$. Pick $\epsilon > 0$, and then $h \in C_c(X)$ such that such that $\|h - 1_A\|_{L^1(X,\mathcal{B},\mu_+ + \mu_-)} < \epsilon$. Without loss of generality, we may assume that $0 \le h \le 1$. Then for any $0 \le g \le f$ in $C_0(X)$, by the choice of h,

$$\begin{aligned} L(g) &= L((1-h)g) + L(hg) \\ &= \int_X (1-h)g\mathrm{d}\mu_+ - \int_X (1-h)g\mathrm{d}\mu_- + \int_X hg\mathrm{d}\mu_+ - \int_X hg\mathrm{d}\mu_- \\ &\le \int_X (1-h)g\mathrm{d}\mu_+ + \|f\|_\infty \epsilon \\ &\le \int_X (1-h)f\mathrm{d}\mu_+ + \|f\|_\infty \epsilon \le \int_X f\mathrm{d}\mu_+ + 2\|f\|_\infty \epsilon \end{aligned}$$

This shows that

$$\sup\{\ L(g)\ :\ 0 \le g \le f\ \} \le \int_X f\mathrm{d}\mu_+\ . \tag{7.2.15}$$

On the other hand,

$$L((1-h)f) = \int_X (1-h)f\mathrm{d}\mu_+ - \int_X (1-h)f\mathrm{d}\mu_- \ge \int_X f\mathrm{d}\mu_+ - 2\|f\|_\infty \epsilon\ .$$

Since $0 \le (1-h)f \le f$, equality holds in (7.2.15) and therefore, the linear functional $f \mapsto \displaystyle\int_X f\mathrm{d}\mu_+$ is uniquely determined by L under the assumption that (7.2.14) with μ_+ and μ_- mutually singular. Then by Theorem 7.2.4, μ_+ is then uniquely determined by L. We have proved:

7.2.13 LEMMA. *Let $L \in (\mathcal{C}_0(X))^*$ have a decomposition as the difference of two* positive *linear functionals as in (7.2.12). Then there is a unique pair of mutually singular Radon measures μ_+, μ_- such that (7.2.14) is valid for all $f \in C_0(X)$, and moreover, μ_+ is determined through L by (7.2.15).*

It is now a simple matter to show that, in fact, *every* $L \in (\mathcal{C}_0(X))^*$ can be written as the difference of two *positive* linear functionals as in (7.2.12). The identity (7.2.15) gives us a candidate for the components of the decomposition:

7.2.14 LEMMA. *Let $L \in (\mathcal{C}_0(X))^*$. For $f \in \mathcal{C}_0^+(X)$, define*

$$L_+(f) = \sup\{\ L(g)\ :\ 0 \le g \le f\ \}\ . \tag{7.2.16}$$

For general $f \in \mathcal{C}_0(X)$, define

$$L_+(f) = L_+(f_+) - L_+(f_-)\ . \tag{7.2.17}$$

where f_+ and f_- are, respectively, the positive and negative parts of f. Then $L_+(f) \in (\mathcal{C}_0(X))^*$, and both L_+ and $L_- := L - L_+$ are positive, so that $L = L_+ - L_-$ is a decomposition of L into the difference between two positive linear functionals.

Proof. We first show that for $f_1, f_2 \in \mathcal{C}_0^+(X)$, $L_+(f_1 + f_2) = L_+(f_1) + L_+(f_2)$. First, let $0 \le g_1 \le f_1$ and $0 \le g_2 \le f_2$. Then $0 \le g_1 + g_2 \le f_1 + f_2$ so that $L_+(f_1 + f_2) \ge L(g_1 + g_2) = L(g_1) + L(g_2)$ Taking the supremum over $g_1 \le f_1$ and $g_2 \le f_2$ yields $L_+(f_1 + f_2) \ge L_+(f_1) + L_+(f_2)$.

Fix $\epsilon > 0$, and choose $g \in \mathcal{C}_0^+(X)$ with $0 \le g \le f_1 + f_2$ so that $L_+(f_1 + f_2) \le L(g) + \epsilon$. Define $g_1 = f_1 \wedge g$ and $g_2 = g - g_1$. Then $0 \le g_1 \le f_1$, $0 \le g_2 \le f_2$ and $g_1 + g_2 = g$. Therefore

$$L_+(f_1 + f_2) \le L(g) + \epsilon \le L(g_1) + L(g_2) + \epsilon \le L_+(f_1) + L_+(f_2) + \epsilon \ .$$

Since $\epsilon > 0$ is arbitrary, $L_+(f_1 + f_2) \le L_+(f_1) + L_+(f_2)$. Together with what we proved above, this shows $L_+(f_1 + f_2) = L_+(f_1) + L_+(f_2)$.

To see that L_+ as extended to all of $C_0(X)$ by (7.2.17), is additive, we observe that if $f = g_1 - g_2$ and $f = h_1 - h_2$ are two ways of writing $f \in C_0(X)$ as a difference of two elements in $C_0^+(X)$, then $g_1 + h_2 = h_1 + g_2$. By what was proved above,

$$L_+(g_1) + L_+(h_2) = L_+(g_1 + h_2) = L_+(h_1 + g_2) = L_+(h_1) + L_+(g_2) \ ,$$

and hence

$$L_+(g_1) - L_+(g_2) = L_+(h_1) - L_+(h_2) \ . \tag{7.2.18}$$

Therefore, one can replace the specific decomposition $f = f_+ - f_-$ in (7.2.17) by any other decomposition of f into the difference between elements of $C_0^+(X)$, and the result is the same. Then for $f, g \in \mathcal{C}_0(X)$, $f + g = (f_+ + g_+) - (f_- + g_-)$ is one way of $f + g$ as the difference of elements of $\mathcal{C}_0^+(X)$,

$$L_+(f+g) = L_+(f_+ + g_+) - L_+(f_- + g_-) = (L_+(f_+) - L_+(f_-)) - (L_+(g_+) - L_+(g_-)) = L_+(f) + L_+(g) \ .$$

This together with the evident fact that $L_+(\alpha f) = \alpha L_+(f)$ for all $\alpha \in \mathbb{R}$ and $f \in \mathcal{C}_0(X)$ shows that L_+ is a linear functional.

By (7.2.17) and then (7.2.16) that

$$\|L_+\| = \sup\{L_+(f) \ : \ 0 \le f \le 1\} \le \sup\{L(g) \ : \ 0 \le g \le 1\} \le \|L\| \ .$$

Thus, L_+ is a bounded linear functional and It is evident from (7.2.16) that $L_+(f) \ge 0$ for all $f \in \mathcal{C}_0(X)$, so that it is also positive.

Finally, defining $L_- = L_+ - L$, we have that for all $f \in \mathcal{C}_0^+(X)$, $L_-(f) = L_+(f) - L(f) \ge 0$ since $L_+(f) \ge L(f)$. Hence L_- is a positive linear functional. Since $L = L_+ - L_-$, this shows that every element L of $(\mathcal{C}_0(X))^*$ can be written as the difference of two positive linear functionals on $\mathcal{C}_0(X)$. □

7.2.15 LEMMA. *Let $L \in (C_0(X))^*$, and let μ_+, μ_- be the unique pair of mutually singular Radon measures on X such that (7.2.14) is valid for all $f \in C_0(X)$. Then*

$$\|L\| := \mu_+(X) + \mu_-(X) \ . \tag{7.2.19}$$

Proof. For all $f \in C_0(X)$ with $\|f\|_\infty \le 1$, $|L(f)| \le \int_X |f| \mathrm{d}\mu_+ + \int_X |f| \mathrm{d}\mu_- \le \mu_+(X) + \mu_-(X)$. On the other hand, let A be a Borel set such that $\mu_+(A) = 0$ and $\mu_-(A^c) = 0$. Pick $\epsilon > 0$, and then $h \in C_c(X)$ such that such that $\|h - (1_{A^c} - 1_A)\|_{L^1(X,\mathcal{B},\mu_+ + \mu_-)} < \epsilon$. Without loss of generality, we may assume that $-1 \le h \le 1$. Then

$$\begin{aligned} \|L\| \ge L(h) &= \int_X h \mathrm{d}\mu_+ - \int_X h \mathrm{d}\mu_- \\ &\ge \int_X (1_{A^c} - 1_A) \mathrm{d}\mu_+ - \int_X (1_{A^c} - 1_A) \mathrm{d}\mu_- - 2\epsilon = \mu_+(X) + \mu_-(X) - 2\epsilon \ . \end{aligned}$$

□

Collecting results from the lemmas, we have proved:

7.2.16 THEOREM (Riesz Representation Theorem for $((C_0(X))^*)$**.** *Let* $(X,\mathcal{U})$ *be a locally compact Hausdorff space. Then for each* $L \in ((C_0(X))^*$ *there exists a unique pair of mutually singular finite Radon measures* μ_+ *and* μ_- *such that (7.2.14) is valid for all* $f \in C_0(X)$ *and such that the norm* $\|L\|$ *of* L *is given by (7.2.19).*

A number of consequences of this theorem deserve further discussion. We begin with a definition:

7.2.17 DEFINITION. A *signed measure* on X is a real valued function μ on the Borel σ–algebra of X such that there exist two positive finite Borel measures μ_1 and μ_2 such that for all Borel sets E, $\mu(E) = \mu_1(E) - \mu_2(E)$.

The set of signed measures is evidently a real vector space. (Complex measures are defined in the analogous way, and would constitute a complex measure space.) We denote the real vector space of signed measures on X by $\mathcal{M}(X)$.

For a bounded Borel function f, define the integral $\int_X f\mathrm{d}\mu$ by $\displaystyle\int_X f\mathrm{d}\mu = \int_X f\mathrm{d}\mu_1 - \int_X f\mathrm{d}\mu_2$. This gives us a continuous linear functional L on $\mathcal{C}_0(X)$ where $L(f) = \displaystyle\int_X f\mathrm{d}\mu$. Theorem 7.2.16 then gives us the existence of uniquely determined positive Borel measures μ_+ and μ_- that are mutually singular – i.e., supported on disjoint sets – and such that

$$\mu(E) = \mu_+(E) - \mu_-(E)$$

for all Borel sets E in X.

7.2.18 DEFINITION. For any signed measure μ, the positive measure $|\mu|$ given by

$$|\mu| = \mu_+ + \mu_-$$

is called the *total variation measure* of μ, and and the function $\mu \mapsto \|\mu\|_{\mathrm{TV}}$ where

$$\|\mu\|_{\mathrm{TV}} = \mu_+(X) + \mu_-(X)$$

is called the *total variation norm* of μ.

It is easy to see from our analysis above that

$$\|\mu\|_{\mathrm{TV}} = \sup\left\{ \int_X f\mathrm{d}\mu \;\middle|\; f \in C_c(X)\ ,\ -1 \le f \le 1 \right\}\ ,$$

and from this the Minkowski inequality is easily seen to hold, so that $\|\cdot\|_{\mathrm{TV}}$ is actually a norm, as the name indicates.

We know that the dual of a Banach space is complete in the dual norm, and so $\mathcal{M}(X)$ is complete in the total variation norm. Moreover, the map $L \mapsto \mu_+ - \mu_-$ where μ_+ and μ_- are related to L as in Theorem 7.2.16 is evidently an isometric isomorphism of $(\mathcal{C}_0(X))^*$ onto $\mathcal{M}(X)$.

The Banach space $C_0(X)$ is not reflexive expect when X is very simple. As long as there exists a single Borel set E that is not both open and closed, we may define a linear functional on $\mathcal{M}$ by

$$\Lambda(\mu) = \mu(E) = \int_X 1_E \mathrm{d}\mu\ .$$

Clearly Λ is linear and

$$|\Lambda(E)| \leq |\mu|(E) \leq |\mu|(X) = \|\mu\|_{\mathrm{TV}} \ ,$$

so Λ is indeed bounded, and so is an element of $(\mathcal{M}(X))^*$. But if there were a function $g \in \mathcal{C}_0(X)$ for which $\Lambda(\mu) = \int_X g\mathrm{d}\mu$ for all $\mu \in \mathcal{M}$, we would have $g(x) = 1_E(x)$ for all x since the point mass δ_x that concentrates unit mass at x belongs to $\mathcal{M}$ for each $x \in X$. But since E is not both open and closed, 1_E is not continuous.

7.2.4 Wiener measure

In this section, Ω denotes the set of continuous functions $\omega : [0,\infty) \to \mathbb{R}$ such that $\omega(0) = 0$. We equip Ω with the topology of uniform convergence on compact subsets of $[0,\infty)$. This makes it a Frechét space, and in particular, a Hausdorff space. However, Ω is not locally compact. For each $t \in [0,\infty)$, let X_t denote the *evaluation functional* $X_t(\omega) = \omega(t)$. Let $\mathcal{F}$ be the σ-algebra on Ω generated by the X_t, $t \geq 0$. *Wiener measure* is a probability measure ν_W on $(\Omega, \mathcal{F})$ such that if $E_1, \dots, E_n$ are Borel sets in $\mathbb{R}$, and $0 \leq t_1 < \cdots < t_n$, a particle performing "Browninan motion" starting from x at time $t = 0$ is in E_j at time t_j for each $j = 1, \dots, n$ with probability

$$\nu_W(\{\omega \in \Omega \ : \ \omega(t_j) \in E_j \ , \quad j = 1, \dots, n \ \}) = \int_{E_1 \times \cdot \times E_n} \prod_{j=1}^n \gamma_{t_j - t_j - 1}(x_{j-1} - x_j)\mathrm{d}x_1 \cdots \mathrm{d}x_n \qquad (7.2.20)$$

where $\gamma_t(x)$ is the Gaussian probability density used to define the heat semigroup, $t_0 := 0$ and $x_0 := 0$. (We restrict ourselves to one dimension only to keep the notation simple. Everything we say in this section about "Brownian motion" in $\mathbb{R}$ extends readily to Brownian motion in $\mathbb{R}^n$ for any $n \in \mathbb{N}$.)

The formula on the right side of (7.2.20) for the probabilities of such events follows from Einstein's work on Browninan motion in 1905, in which he related Brownian motion to diffusion and the heat equation. Einstein's precise explanation for Brownian motion in terms of molecular collisions − at a time when the very existence of atoms and molecules was still a matter of dispute − made it possible to determine Avogadro's number, the number of atoms of hydrogen in one gram of hydrogen, by making observations through a microscope of pollen-sized particles undergoing Brownian motion. This was actually done in 1908 by Jean Baptiste Perrin, and he was awarded the Nobel prize in 1926 for his experimental work. Einstein received the prize in 1921 for his theoretical work. A simplified version of Einstein's formulae, leaving out the constants that make it possible to determine Avogadro's number by looking though a microscope, says that the probability that a Brownian particle starting at 0 at time $t = 0$ is in E_j at time t_j is given in terms of the heat kernel $\gamma_t(x - y)$ by

$$\int_{E_1 \times \cdot \times E_n} \prod_{j=1}^n \gamma_{t_j - t_j - 1}(x_{j-1} - x_j)\mathrm{d}x_1 \cdots \mathrm{d}x_n \qquad (7.2.21)$$

where $t_0 := 0$ and $x_0 := 0$, which is the formula on the right side of (7.2.20). For example if $n = 2$, this reduces to, using the Fubini-Tonelli Theorem,

$$\int_{E_1} \gamma_{t_1}(x) \left(\int_{E_2} \gamma_{t_2 - t_1}(x - y)\mathrm{d}y \right) \mathrm{d}x \ .$$

The inner integral gives the probability of the Brownian particle making a transition from x to E_2 in time $t_2 - t_1$, and $\gamma_{t_1}(x)$ is the probability density for the Brownian particle making a transition from 0 to x in time t_1. The composite formula is justified on account of the increments of the motion being "statistically independent".

We shall not go further into the physical origins of the formula on the right side of (7.2.20), but turn to the mathematical question of whether or not there exists a probability measure ν_W on Ω, the infinite dimensional "path space" for the Brownian particle, such that (7.2.20) is valid.

The sets of the form $\{\omega \in \Omega \ : \ \omega(t_j) \in E_j \ , \quad j = 1, \dots, n \ \}$ are very special in $\mathcal{F}$, and the existence of a countably additive probability measure $(\Omega, \mathcal{F})$ that assigns the specified probabilities to these sets is far from trivial. It is therefore somewhat amazing that Norbert Wiener constructed the measure ν_W on $(\Omega, \mathcal{F})$ in 1923, well before Kolmogorov had even given his measure-theoretic formulation of probability theory.

In this section we give a proof of Wiener's Theorem due to Edward Nelson that makes use of the Riesz-Markov Theorem, and the regularity of the measures that it provides. There are two main parts to the proof: In the first part, we embed Ω into a *compact* Hausdorff space, and use the formula on the right hand side of (7.2.20) to define a positive linear functional on $\mathcal{C}(X) = \mathcal{C}_c(X)$. The Riesz-Markoff Theorem then provides a regular Borel probability measure μ_X on X. In the second second part, we show that Ω is a Borel set, and that $\mu_W(\Omega) = 1$. We then obtain ν_W by restricting μ_W to Ω. It is in the course of the proof that $\mu_W(\Omega) = 1$ that we make essential use of the inner regularity of μ_W to deal with the fact that continuity of a function ω from $[0, \infty)$ into $\mathbb{R}$ depends on the behavior of ω at uncountably many points.

This is not the only way to construct Wiener measure, and indeed Wiener's paper predates the work of Markoff on which this approach depends. However, it is flexible and powerful, and can be used to construct many other measures on "path spaces". It will be clear that very little specific information about the heat kernel is used in the proof.

Our first task, which is essentially notational, is to recast the formula (7.2.21) in terms of a probability measure on $\mathbb{R}^n$. Let S be an arbitrary finite subset of distinct elements t_j of $(0, \infty)$ arrange in increasing order: $S = \{t_1, \dots, t_n\}$ with $t_j \le t_{j+1}$ for $j = 1, \dots, n-1$.

Given such a set S, define a measure $\mu_{x,S}$ on $\mathbb{R}^n$ by

$$\mathrm{d}\mu_{x,S} = \prod_{j=1}^{n} \gamma_{t_j - t_j - 1}(x_{j-1} - x_j)\mathrm{d}x_1 \cdots \mathrm{d}x_n \ . \tag{7.2.22}$$

where $t_0 := 0$ and $x_0 := x$. (We need the more general formula in which x_0 is arbitrary and not set equal to 0 for reasons that are explained below.) For example, if $S = \{t_1, t_2\}$, then

$$\mathrm{d}\mu_{x,\{t_1,t_2\}} = \gamma_{t_1}(x - x_1))\gamma(x_1 - x_2)\mathrm{d}x_1\mathrm{d}x_2 \ . \tag{7.2.23}$$

Integrating first in x_2, and then in x_1 and using the fact that $\int_{\mathbb{R}} \gamma_t(x)\mathrm{d}x = 1$ for all t, one readily sees that $\mu_{x,\{t_1,t_2\}}$ is a probability measure. The same reasoning shows that for all S, $\mu_{x,S}$ is a probability measure. For Borel set E_1 and E_2, $\mu_{x,\{t_1,t_2\}}(E_1 \times E_2)$ is the probability that the particle, initially at x, is in E_1 at time t_1, and then in E_2 at time t_2. Likewise,

$$\mu_{x,\{t_1,\dots,t_n\}}(E_1 \times \cdots \times E_n)$$

is to be thought of as the probability the particle, initially at x, is in E_j at time t_j for $j = 1, \ldots, n$. Notice that if some $E_j = \mathbb{R}$, the condition that the particle is in E_j is vacuous given that the particle must be *somewhere* in $\mathbb{R}$. This corresponds to an important consistence condition of the family of measures $\{\mu_{x,S}\}$ indexed by the finite ordered sets S Let $S = \{t_1, \ldots, t_n\}$ and for some $j = 1, \ldots, n$, let $S' := S\backslash\{t_j\}$ ordered as above.

Let $E_1, \ldots, E_n$ be n Borel sets in $\mathbb{R}$. Let $E = E_1 \times \cdot \times E_n$ and let E' be the Cartesian product of $\{E_1, \ldots, E_n\}\backslash\{E_j\}$ taken in the order induced by the subscripts. Then since

$$\int_{\mathbb{R}} \gamma_{s-t_{j-1}}(x_{j-1} - y)\gamma_{t_j - s}(y - x_j)\mathrm{d}y = \gamma_{t_j - t_{j-1}}(x_{j-1} - x_j) \ ,$$

doing the integration over the x_j first on the left we find that in case $E_j = \mathbb{R}$,

$$\mu_{x,S}\left(E\right) = \mu_{x,S'}\left(E'\right) \ . \tag{7.2.24}$$

There is another important relation satisfied by these measures, this times with x and S both being variable. Let $S = \{t_1, \ldots, t_n\}$ be given, $n \geq 2$. Let $1 \leq k \leq n-1$ be given and let f be a bounded Borel function on $\mathbb{R}^k$ and let g be a bounded Borel function on $\mathbb{R}^{n-k}$. Let $S' = \{t_1, \ldots, t_k\}$ and let $S'' := \{t_{k+1}, \ldots, t_n\}$. Then doing the integral over $x_{k+1}, \ldots, x_n$ first, we find that

$$\int_{\mathbb{R}^n} f(x_1, \ldots, x_k) g(x_{k+1}, \ldots, x_n)\mathrm{d}\mu_{x,S} = \int_{\mathbb{R}^k} f(x_1, \ldots, x_k)\left(\int_{R^{n-k}} g(x_{k+1}, \ldots, x_n)\mathrm{d}\mu_{x_k,S''}\right)\mathrm{d}\mu_{x,S'} \ . \tag{7.2.25}$$

In probabilistic terms, one may regard the functions $X_j : (x_1, \ldots, x_n) \mapsto x_j$, $j = 1, \ldots, n$ as "random variables" on the probability space $(\mathbb{R}^n, \mathcal{B}, \mu_{x,S})$. The product structure in (7.2.25) can then be interpreted as saying that "given X_k, the future variables $X_{k+1}, \ldots, X_n$ are statistically independent of the past variables'$X_1, \ldots, X_{k-1}$." This is the *Markov property.* In what follows we do not need to know precisely what "given X_k" means; it involves the notion of conditional probability. We shall only make direct analytic use of the factorization identity (7.2.25). However, it would be a grave injustice not to at least mention mention in passing the Markov property at this point. For our purposes, the Markov property of the measure $\mu_{x,S}$ is precisely the factorization formula (7.2.25) relating it to the measures $\mu_{x,S'}$ and $\mu_{x_k,S''}$.

7.2.19 THEOREM. *There exists a unique probability measure ν_W on $(\Omega, \mathcal{F})$ such that if φ is any function on $\mathbb{R}^n$ and $S = \{t_1, \ldots, t_n\}$, $0 < t_0 < \cdots < t_n$, then*

$$\int_{\Omega} \varphi(\omega(t_1), \ldots, \omega(t_n))\mathrm{d}\nu_W(\omega) = \int_{\mathbb{R}^n} \varphi(x_1, \ldots, x_n)\mathrm{d}\mu_{0,\{t_1,\ldots,t_n\}} \ .$$

Proof of Theorem 7.2.19. Let $\dot{\mathbb{R}}$ denote the one-point compactification of $\mathbb{R}$, and let X denote the Cartesian product $(\dot{\mathbb{R}})^{[0,\infty)}$ with the product topology. Then by Tychonov's Theorem, X is a compact Hausdorff space. The general element ω of X is an arbitrary function from $[0, \infty)$ into $\dot{\mathbb{R}}$.

Let $\mathcal{A}$ denote the set of all functions f on X of the form

$$f(\omega) = \phi(\omega(t_1), \ldots, \omega(t_n)) \tag{7.2.26}$$

for some $n \in \mathbb{N}$, some $S = \{t_1, \ldots, t_n\}$, and some bounded continuous function $\varphi : \mathbb{R}^n \to \mathbb{R}$. Then $\mathcal{A}$ is an algebra consisting of continuous functions on X which contains the constant function 1 and separates points. By the Stone Wierstrass Theorem, it is dense in $\mathcal{C}(X)$.

We now define a linear functional L on $\mathcal{A}$ by

$$L(f) := \int_{\mathbb{R}}^{n} \phi(x_1, \ldots, x_n) \mathrm{d}\mu_{0,\{t_1,\ldots,t_n\}} \tag{7.2.27}$$

where f and ϕ are related by (7.2.26). Any given function in $\mathcal{A}$ has many different representations of the form (7.2.26) since one can always enlarge the set $\{t_1, \ldots, t_n\}$ but then have ϕ depends trivially on the inserted coordinates. Because of the consistency relation (7.2.24), the right hand side of (7.2.27) is independent of the choice of the representative. Hence $f \mapsto L(f)$ is a well-defined linear functional on $\mathcal{A}$, and evidently when $f \in \mathcal{A}$ is given by (7.2.26), $f(\omega) \geq 0$ for all ω if and only if $\varphi(x_1, \ldots, x_n) \geq 0$ for all $x_1, \ldots, x_n$. Therefore, since each $\mu_{x,S}$ is a probability measure, L is positive on $\mathcal{A}$, and for all $f \in \mathcal{A}$, $|L(f)| \leq \|f\|_\infty$. Thus L has a unique extension by continuity from the dense sub algebra $\mathcal{A}$ to all of $\mathcal{C}(X)$, and evidently this extension, still denotes by L, is positive with $\|L\| = 1$.

We may now invoke the Riesz-Markov Theorem to assert the existence of a unique regular Borel measure μ_W on X such that $L(f) = \displaystyle\int_X f(\omega)\mathrm{d}\mu_W(\omega)$ for all $f \in \mathcal{C}(X)$.

The proof will be completed by showing that Ω is a Borel set in X, and that $\mu_W(\Omega) = 1$. Whether $\omega \in X$ is continuous or not depends on the behavior of ω at uncountably many values of t. The fact that μ_W is regular, specifically inner regular, will be crucial for estimating the probabilities of sets depending on the behavior of ω at all of the uncountably many points in an interval $[a, b]$ in terms of sets depending on the behavior of ω at only finitely many points in $[a, b]$.

For $\epsilon, \delta > 0$, define

$$\rho(\epsilon, \delta) := \sup_{0<t<\delta} \int_{|x|>\epsilon} \gamma_t(y)dy = \int_{|x|>\epsilon} \gamma_\delta(y)dy \tag{7.2.28}$$

It is easy to see that as $\delta \downarrow 0$, $\rho(\epsilon, \delta) = o(\delta)$. In fact, $\rho(\epsilon, \delta) = o(\delta^n)$ for any $n \in \mathbb{N}$, and even that is not all. However, all we shall use is that $\rho(\epsilon, \delta) = o(\delta)$. The rest of the proof is broken into several steps.

Step 1. Let $\epsilon, \delta > 0$. Let $0 < t_1 < \cdots < t_n$ with $t_n - t_1 < \delta$. Define

$$A = \{\omega \ : \ |\omega(t_1) - \omega(t_j)| > \epsilon \quad \text{for some } j = 2, \ldots, n \ \}$$

We show in this step that $\mu_W(A) \leq 2\rho(\epsilon, \delta/2)$, independent of n. This is based on the Markov property. Define

$$B := \{\omega \ : \ |\omega(t_1) - \omega(t_n)| > \epsilon/2\}$$

and then for $j = 2, \ldots n$, define

$$C_j := \{\omega \ : \ |\omega(t_1) - \omega(t_j)| > \epsilon \quad \text{and} \quad |\omega(t_1) - \omega(t_i)| \leq \epsilon \quad \text{for} \quad i \leq j-1\}$$

and

$$D_j := \{\omega \ : \ |\omega(t_j) - \omega(t_n)| > \epsilon/2\} \ .$$

Note that if $\omega \in A$, there is some least value of j such that $|\omega(t_1) - \omega(t_j)| > \epsilon$, and then, if $\omega \notin B$, $j \leq n-1$, and $|\omega(t_j) - \omega(t_n)| > \epsilon/2$. That is,

$$A = B \bigcup_{j=2}^{n-1} (C_j \cap D_j) \ .$$

Note that 1_{C_j} can be written in the form $1_{C_j}(\omega) = \varphi(\omega(t_1), \ldots \omega(t_{j-1}))$ where φ is the characteristic function of an open set on $\mathbb{R}^{j-1}$, and that 1_{C_j} can be written in the form $1_{C_j}(\omega) = \psi(\omega(t_j), \omega(t_n))$ where

φ is the characteristic function of an open set on $\mathbb{R}^2$. Then by the Markov property (7.2.25) and the definition of μ_W,

$$\mu_W(C_j \cap D_j) = \int_{\mathbb{R}^{j-1}} \varphi(x_1, \cdots, x_{j-1}) \left(\int_{\mathbb{R}^2} \psi(x_j, x_n) \mathrm{d}\mu_{x_j, \{t_n - t_j\}} \right) \mathrm{d}\mu_{0,\{t_1,\ldots,t_j\}} \tag{7.2.29}$$

Since

$$\int_{\mathbb{R}^2} \psi(x_j, x_n) \mathrm{d}\mu_{x_j,\{t_n-t_j\}} = \int_{|x|>\epsilon/2} \gamma_{t_n-t_j}(x) \mathrm{d}x \leq \rho(\epsilon/2, \delta) \ ,$$

(7.2.29) yields $\mu_W(C_j \cap D_j) \leq \rho(\epsilon/2, \delta)\mu_W(C_j)$. Then since the sets C_j are mutually disjoint, $\mu_W(\cup_{j=2}^n C_j) \leq 1$. Thus, $\mu_W(A) \leq \mu_W(B) + \rho(\epsilon/2, \delta) \leq 2\rho(\epsilon/2, \delta)$.

Step 2. Fix $\epsilon, \delta > 0$. Fix $a < b \in [0, \infty)$ with $b - a < \delta$. Define the set

$$E(a, b, \epsilon) := \{\omega \ : |\omega(s) - \omega(t)| \geq 2\epsilon \quad \text{for some} \quad s, t \in [a, b] \ \} \ .$$

In this step we show that $\mu_W(E(a, b, \epsilon)) \leq 2\rho(\epsilon/2, \delta)$.

To do this, let S be a finite subset of $[a, b]$, and define

$$E(a, b, \epsilon, S) := \{\omega \ : |\omega(s) - \omega(t)| \geq 2\epsilon \quad \text{for some} \quad s, t \in S \ \} \ .$$

Note that each $E(a, b, \epsilon, S)$ is open, and $E(a, b, \epsilon)$ is the union of all of the $E(a, b, \epsilon, S)$ as S ranges over all finite subsets S of $[a, b]$. Since μ_W is regular, for all $\eta > 0$, there is a compact set $K \subset E(a, b, \epsilon)$ such that $\mu_W(K) \geq\subset E(a, b, \epsilon) - \eta$. Since the sets of the form $E(a, b, \epsilon, S)$, $S \subset [a, b]$ finite, are an open cover of K, there exists a finite sub-cover. But any finite union of sets of the form $E(a, b, \epsilon, S)$ is again of this form. Hence there exists a finite $S_\eta \subset [a, b]$ such that $\mu_W(E(a, b, \epsilon, S_\eta) \geq \mu_W(E(a, b, \epsilon)$. It follows that

$$\mu_W(E(a, b, \epsilon) = \sup\{\mu_W(E(a, b, \epsilon, S) \ : \ S \subset [a, b] \text{ finite } \} \ . \tag{7.2.30}$$

Now for any finite set $S = \{t_1, \ldots, t_n\}$ if for some $1 \leq i < j \leq n$, $|\omega(t_i) - \omega(t_j)| > 2\epsilon$, then either $|\omega(t_1) - \omega(t_i)| > \epsilon$ or $|\omega(t_1) - \omega(t_j)| > \epsilon$. Hence the bound proved in Step 1 implies that $\mu_W(E(a, b, \epsilon, S) \leq 2\rho(\epsilon/2, \delta)$.

Step 3. Fix $\epsilon, \delta > 0$ with $1/\delta \in \mathbb{N}$. Let $k \in \mathbb{N}$. Define

$$F(k, \epsilon, \delta) := \{\omega \ : \ |\omega(t) - \omega(s)| > 4\epsilon \quad \text{for some} \quad t, s \in [0, k] \ \} \ .$$

In this step we show that $\mu_W(F(k, \epsilon, \delta)) \leq 2k\rho(\epsilon/2\delta)/\delta$.

To do this, write $[0, k]$ as the union of k/δ subintervals of the form $[(j-1)\delta, j\delta]$, $j = 1, \ldots, k/\delta$. If $\omega \in F(k, \epsilon, \delta)$, $|\omega(t) - \omega(s)| > 4\epsilon$ for some s, t in the same or adjacent intervals. But that means that $|\omega(u) - \omega(v)| > 2\epsilon$ for some u, v belonging to one of these intervals. Thus, ω belongs to $E((j-1)\delta, j\delta, \epsilon)$, as defined in Step 2, for some $j = 1, \ldots, k/\delta$. This proves the desired bound.

Step 4. We complete the proof. A function $\omega : [0, \infty] \to \dot{\mathbb{R}}$ is continuous with $\omega(0) \in \mathbb{R}$ if and only if its restriction to each $[0, k]$, $k \in \mathbb{N}$ is uniformly continuous, meaning that for for all $\epsilon > 0$, there is a $\delta > 0$ so that $|\omega(s) - \omega(t)| < 4\epsilon$ whenever $|s - t| \leq \delta$. That is, ω is continuous if and only if for each $k \in \mathbb{N}$, ω belongs to

$$\bigcap_{\epsilon>0} \bigcup_{\delta>0} F^c(k, \epsilon, \delta) \ ,$$

and $F(k,\epsilon,\delta)$ is defined as in the previous step. Moreover, we can restrict ϵ and δ to be the reciprocals on positive integers so that the intersection and union are countable. Since each $F(k,\epsilon,\delta)$ is open, in X,

$$\Omega = \bigcap_{k\in\mathbb{N}} \bigcap_{\epsilon>0} \bigcup_{\delta>0} F^c(k,\epsilon,\delta) \ ,$$

is a Borel set. It follows that Ω^c is a countable union of sets of the form $\bigcap\limits_{\delta>0} F(k,\epsilon,\delta)$ and

$$\mu_W\left(\bigcap_{\delta>0} F(k,\epsilon,\delta)\right) = \lim_{\delta\downarrow 0} \mu_W(F(k,\epsilon,\delta)) = 0$$

by the estimate of Step 3 since $\rho(\epsilon,\delta) = o(\delta)$. Thus Ω is a Borel subset of X, and $\mu_W(\Omega) = 1$. We define ν_W to be the restriction of μ_W to Ω. □

7.3 Functions of Bounded Variation

7.3.1 The classical definition

The class of *functions of bounded variation* on an interval $I \subset \mathbb{R}$ was introduced by Camile Jordan in 1881 in an investigation of the point-wise convergence of Fourier Series. The notion turns out to be quite useful in many contexts, and we shall give a modern development of the theory that readily admits generalization to functions on open sets in $\mathbb{R}^n$, which is crucial for many modern applications. However, first we recall the classical definition:

7.3.1 DEFINITION (Bounded variation)**.** Let I be an interval in $\mathbb{R}$. Let $\mathcal{P}_I$ denote the set of all ordered sets $\{x_0, x_1, \dots, x_n\} \subset I$, $n \in \mathbb{N}$, with $x_{j-1} < x_j$ for all $j = 1, \dots, n$. For any real valued function h defined on I, define

$$TV(h;I) = \sup\left\{ \sum_{j=1}^{n} |h(x_j) - h(x_{j-1})| \ : \ \{x_0, x_1, \dots, x_n\} \in \mathcal{P}_I \right\} . \tag{7.3.1}$$

A function $h : I \to \mathbb{R}$ is of *bounded variation on* I in case $TV(h;I) < \infty$.

7.3.2 EXAMPLE. *Let h be continuously differentiable and such that the derivative h' is integrable on an interval (a,b). Then for any $\{x_0, x_1, \dots, x_n\} \in \mathcal{P}_{(a,b)}$,*

$$\sum_{j=1}^{n} |h(x_j) - h(x_{j-1})| = \sum_{j=1}^{n} \left| \int_{x_{j-1}}^{x_j} h'(y)\mathrm{d}y \right| \le \int_a^b |h'(x)|\mathrm{d}x \ ,$$

and hence $TV(h;(a,b)) \le \int_a^b |h'(x)|\mathrm{d}x$. In fact, it is not hard to show that actually equality holds in this inequality.

However, functions of bounded variation need not be continuous. Consider for example the function f on $\mathbb{R}$ defined by

$$h(x) = \begin{cases} 0 & x < 0 \\ 1 & x \ge 0 \ . \end{cases}$$

Evidently $TV(h,\mathbb{R}) = 1$. More generally, any monotone non-decreasing function has bounded variation on any interval on which it is bounded.

We shall soon we that every bounded variation function on any interval $[a, b]$ is the difference of two monotone non-decreasing functions; this is the *Jordan decomposition.* The Jordan decomposition is closely related to the Hahn-Saks decomposition of a signed Borel measure into its positive and negative parts. This means that any function of bounded variation is measurable, and in fact, has left and right limits (which need not be equal) at every point. In particular, functions of bounded variation are measurable. In the modern development presented here, we shall deduce the Jordan Decomposition from the Hahn-Saks decomposition, but we need to know from the outset that functions of bounded variation are at least measurable, which we show after proving a simple lemma that will be useful again.

7.3.3 LEMMA. *Let $\epsilon > 0$. Let h be a real valued function defined on the interval $[c, d]$ such that $TV(h, [c, d]) < \epsilon$. Let $\widetilde{h}$ be the linear interpolation of h between c and d, That is, for any $\lambda \in [0, 1]$ define $x_\lambda = (1-\lambda)c + \lambda d$, and $\widetilde{h}(x_\lambda) = (1-\lambda)h(c) + \lambda h(d)$. Then $|\widetilde{h}(x) - h(x)| < \epsilon$ for all $x \in [c, d]$.*

Proof. By hypothesis, for any $x \in (c, d)$ $|h(x) - h(c)| + |h(x) - h(d)| \le TV(h, [c, d]) < \epsilon$. By the convexity of $t \mapsto |t|$,

$$|h(x_\lambda) - (1-\lambda)h(c) - \lambda h(d)| \le (1-\lambda)|h(x_\lambda) - h(c)| + \lambda|h(x_\lambda) - h(d)| \le \epsilon \ .$$

□

Then next simple observation to make is that if $b < c < d$, $TV(h; [b, d]) = TV(h; [b, c]) + TV(h; [c, d])$. This may be iterated in the obvious way.

7.3.4 LEMMA. *Let m denote Lebesgue measure on $\mathbb{R}$. Let h be a real valued function defined on the interval $[c, d]$ such that $TV(h, [c, d]) < \infty$. For each $\epsilon > 0$ there is a piecewise linear function h_ϵ and a Borel set $E \subset [c, d]$ with $m(E) < \epsilon$ such that for all $x \in [c, d] \cap E^c$, $|h(x) - h_\epsilon(x)| < \epsilon$ and $|h(x) - h_\epsilon(x)| \le TV(h, [c, d])$ for* all *$x \in [a, d]$. Moreover, $TV(h_\epsilon, [c, d]) \le TV(h, [c, d])$*

Proof. Choose $k \in \mathbb{N}$ such that $\frac{1}{k}TV(h; [c, d]) < \epsilon$. Then choose n so large that $k(d-c)/n < \epsilon$. Divide $[c, d]$ into n closed intervals $\{I_j \ : \ j = 1, \dots, n\}$ of equal length, with consecutive intervals overlapping at endpoints only. Since $\sum_{j=1}^{n} TV(h; I_j) = TV(h; [c, d])$, $TV(h; I_j) \ge \frac{1}{k}TV(h; I_j)$ for at most k values of j. Let E be the union of any such intervals, and then by the choice of n, $m(E) < \epsilon$. On each I_j not included in E, h is uniformly within ϵ of its linear interpolation between the endpoints $\{x_0, \dots, x_n\}$ of the intervals of I_j according to Lemma 7.3.3. Thus if we define h_ϵ be the linear interpolation of h at the endpoints of the intervals I_j, $j = 1, \dots, n$, $|h(x) - h_\epsilon(x)| < \epsilon$ for $x \notin E$. Finally, it is clear that since h_ϵ is linear between x_{j-1} and x_j for each $j = 1, \dots, n$,

$$TV(h_\epsilon) = \sum_{j=1}^{n} |h(x_j - h(x_{j-1})| \le TV(h, [c, d]) \ .$$

□

7.3.5 Remark. Given any real valued function h on $[c, d]$ such that $TV(h, [c, d]) < \infty$ Lemma 7.3.4 provides a sequence of continuous functions converging to h in measure on $[c, d]$, and hence h is measurable on $[c, d]$.

In what follows, (a,b) denotes any open interval in $\mathbb{R}$. In particular, it is not excluded that either $a=-\infty$, $b=\infty$, or both. Let $\mathcal{C}_c^1((a,b))$ denote the set of continuously differentiable real-valued functions with compact support in (a,b). Then $\mathcal{C}_c^1((a,b))$ is dense in $\mathcal{C}_c((a,b))$. This is easily shown by convolving $f \in \mathcal{C}_c((a,b))$ by a smooth probability density supported in some sufficiently small interval $[-\delta,\delta]$. In fact, the same argument shows that $\mathcal{C}_c^\infty((a,b))$, the set of infinitely differentiable real-valued functions with compact support in (a,b), is dense in $\mathcal{C}_c((a,b))$. For $f \in \mathcal{C}_c^1((a,b))$, let f' denote the derivative of f. Throughout the rest of this section, $\mathcal{B}$ denotes the Borel σ-algebra of (a,b), and m denote Lebesgue measure on (a,b), though we shall often write integrals with respect to Lebesgue measure using $\mathrm{d}x$ in place of $\mathrm{d}m$.

Now let h be a real-valued function on (a,b) with $TV(h;(a,b)) < \infty$. In particular, h is uniformly bounded on (a,b), and so if both a and b are finite, h is integrable with respect to Lebesgue measure m since it is also measurable by the remark following Lemma 7.3.4. If either $a=-\infty$ or $b=\infty$ this need not be the case, and then we further suppose that h is integrable. The next lemma leads to the modern characterization of functions of bounded variation.

7.3.6 LEMMA. *Let h be an integrable real-valued function on (a,b) with $TV(h;(a,b)) < \infty$. Let μ denote the finite Borel measure on (a,b) that is absolutely continuous with respect to Lebesgue measure on (a,b) with Radon-Nikodym derivative h. That is, $d\mu = h\mathrm{d}x$. Define $L \in (\mathcal{C}_0((a,b)))^*$ by*

$$L(f) = \int_a^b f\mathrm{d}\mu \ . \tag{7.3.2}$$

Then for all $f \in \mathcal{C}_c^1((a,b))$,

$$|L(f')| \le TV(h;[a,b])\|f\|_\infty \ . \tag{7.3.3}$$

Proof. Let $f \in \mathcal{C}_c^1((a,b))$, and let $[c,d]$ contain the support of f. Let $\epsilon > 0$. By Lemma 7.3.4, there exists a piecewise linear function h_ϵ on $[c,d]$ such that $|h_\epsilon(x) - h(x)| \le TV(h,[c,d])$ for all $x \in [c,d]$ and such that $|h_\epsilon(x) - h(x)| \le \epsilon$ outside a set E with $m(E) < \epsilon$. Therefore,

$$L(f') = \int_a^b f'(x)h_\epsilon(x)\mathrm{d}x + \int_a^b f'(x)(h - h_\epsilon((x)\mathrm{d}x \ ,$$

and hence

$$\begin{aligned}
\left| L(f') - \int_a^b f'(x)h_\epsilon(x)\mathrm{d}x \right| &\le \int_c^d |f'(x)||h - h_\epsilon((x)|\mathrm{d}x \\
&\le \|f'\|_\infty \int_{[c,d]\setminus E} |h - h_\epsilon((x)|\mathrm{d}x + \|f'\|_\infty \int_E |h - h_\epsilon((x)|\mathrm{d}x \\
&\le \epsilon\|f'\|_\infty(d - c + TV(h,[c,d])) \ .
\end{aligned}$$

Then if $\{x_0,\dots,x_n\}$ denotes set in $\mathcal{P}_{[c,d]}$ such that h_ϵ is the linear interpolation of h through $\{x_0,\dots,x_n\}$, integration by parts yields

$$\left| \int_a^b f'(x)h_\epsilon(x)\mathrm{d}x \right| = \left| \sum_{j=1}^n \frac{h(x_j) - h(x_{j-1})}{x_j - x_{j-1}} \int_{x_{j-1}}^{x_j} f(y)\mathrm{d}y \right| \le \|f\|_\infty \sum_{j=1}^n |h(x_j) - h(x_{j-1})| \ .$$

Combining the last two estimates, $|L(f')| \le TV(h;[a,b])\|f\|_\infty(1 + \epsilon(d-c) + TV(h,[c,d]))$. Since $\epsilon > 0$ is arbitrary, (7.3.3) is proved. □

7.3.2 The modern characterization

The construction in the previous lemma show how integrable functions of bounded variation give rise to a special class of linear functionals L on $\mathcal{C}_0((a,b))$: Those for which there exists $C < \infty$ such that for all $f \in \mathcal{C}_c^1((a,b))$

$$|L(f')| \le C\|f\|_\infty \ . \tag{7.3.4}$$

We show in this section that every such functional L $\mathcal{C}_0((a,b))$ arises in this way: There is a unique integrable function h of bounded variation on (a,b) such that $L(f) = \int_{(a,b)} fh\mathrm{d}x$ for all $f \in \mathcal{C}_0((a,b))$. In the course of proving this, we shall obtain further information about the class of functions of bounded variation. We begin with a lemma that provides a crucial "generalized integration by parts" formula.

7.3.7 LEMMA. *$L \in (\mathcal{C}_0((a,b)))^*$ be such that for some $C < \infty$ and all all $f \in \mathcal{C}_c^1((a,b))$, (7.3.4) is valid. Then there exists a unique pair of signed Borel measures μ and ν on (a,b) such that for all $f \in \mathcal{C}_0((a,b))$*

$$L(f) = \int_{(a,b)} f\mathrm{d}\mu \tag{7.3.5}$$

and for all $f \in \mathcal{C}_c^1((a,b))$

$$L(f') = -\int_{(a,b)} f\mathrm{d}\nu \ . \tag{7.3.6}$$

Moreover, $\|\nu\|_{\mathrm{TV}} \le C$, and for all $f \in \mathcal{C}_c^1((a,b))$,

$$\int_{(a,b)} f'\mathrm{d}\mu = -\int_{(a,b)} f\mathrm{d}\nu \ . \tag{7.3.7}$$

Proof. Direct application of the Riesz Representation Theorem for $(\mathcal{C}_0((a,b)))^*$, provides the signed measure μ such that (7.3.5) is valid. Next, note that $\mathcal{C}_c^1((a,b))$ is dense in $\mathcal{C}_0((a,b))$ in the uniform norm, and the functional $f \mapsto L(f')$ is linear since differentiation and L are both linear. It is bounded on the dense subspace $\mathcal{C}_c^1((a,b))$ by (7.3.4), and hence it extends by continuity to a linear functional M on all of $\mathcal{C}_0((a,b))$ and $\|M\|_* \le C$. Now a second application of the Riesz Representation Theorem for $(\mathcal{C}_0((a,b)))^*$ provides the signed measure ν such that (7.3.6) is valid. Finally, (7.3.7) follows directly from (7.3.5) and (7.3.6). □

7.3.8 EXAMPLE. *Let h be a continuously differentiable function on (a,b) such that $\int_a^b |h'(x)|\mathrm{d}x = C < \infty$. Define*

$$L(f) = \int_a^b h(x)f(x)\mathrm{d}x \tag{7.3.8}$$

for all $f \in \mathcal{C}_c((a,b))$. Then, integrating by parts, for $f \in \mathcal{C}_c^1((a,b))$,

$$L(f') = \int_a^b h(x)f'(x)\mathrm{d}x = -\int_a^b f(x)h'(x)\mathrm{d}x \tag{7.3.9}$$

and hence $|L(f')| \le C\|f\|_\infty$ so that (7.3.4) is satisfied with $C = \displaystyle\int_a^b |h'(x)|\mathrm{d}x$. Moreover, in this case the measure ν is evidently given by

$$\mathrm{d}\nu = -h'(x)\mathrm{d}x \ .$$

We may regard $-\nu$ as the "generalized derivative" of the bounded variation function h even when h is not differentiable.

7.3.9 LEMMA. *Let L, C, μ and ν be as in Lemma 7.3.7. If $b < \infty$, then the limits*

$$\lim_{c\uparrow b} \frac{1}{b-c}\mu(c,b) =: h(b) \quad \text{and} \quad \lim_{c\downarrow a} \frac{1}{c-a}\mu(c,b) =: h(a) \tag{7.3.10}$$

exit and

$$|h(a)|, |h(b)| \le \frac{1}{b-a}\|\mu\|_{\mathrm{TV}} + C \ . \tag{7.3.11}$$

Proof. Let f be a C^1 function that monotonically increases from 0 to 1 such that f' has support contained in $[a', b'] \subset (a,b)$. For each $c \in (b', b)$ define

$$f_c(x) = \begin{cases} f(x) & x \le c \\ 1 - x/(b-c) & c < x < b \ . \end{cases}$$

While f_c does not belong to $C^1_c((a,b))$, it is easy to approximate f_c by a sequence $\{g_{c,k}\}_{k\in\mathbb{N}}$ of such functions that converge uniformly to f_c and whose derivatives converge at each x to

$$f'(x) - \frac{1}{b-c}1_{(c,b)}(x)$$

and are uniformly bounded in absolute value by $\max\{\|f'\|_\infty, (b-c)^{-1}\}$. Applying (7.3.7) to $g_{c,k}$ and taking the limit $k \to \infty$, we obtain

$$\int_{(a,b)} f'\mathrm{d}\mu - \frac{\mu((c,b))}{b-c} = -\int_{(a,b)} f_c\mathrm{d}\nu \ . \tag{7.3.12}$$

By the Lebesgue Dominated Convergence Theorem,

$$\lim_{c\uparrow b}\int_{(a,b)} f_c\mathrm{d}\nu = \int_{(a,b)} f\mathrm{d}\nu \ , \tag{7.3.13}$$

and since the integral on the left hand side of (7.3.12) is independent of c, the first limit in (7.3.10) exists. Moreover, for each c and k, $|L(g'_{n,k})| \le C\|g_{n,k}\|_\infty = C$, and hence for all c,

$$\left|\int_{(a,b)} f'\mathrm{d}\mu - \frac{\mu((c,b))}{b-c}\right| \le C \ . \tag{7.3.14}$$

Moreover, f can be chosen so that $\|f'\|_\infty$ is arbitrarily close to $(b'-a)^{-1}$, and we may take b' arbitrarily close to c. In particular, if $a = -\infty$, we can choose f with $\|f'\|_\infty$ is arbitrarily close to 0. Then (7.3.14) yields

$$\frac{|\mu((c,b))|}{b-c} \le (c-a)^{-1}\|\mu\|_{\mathrm{TV}} + C \ ,$$

and this proves (7.3.11) for $h(b)$. The statements involving $h(a)$ are valid by symmetry. □

7.3.10 LEMMA. *Let L, C, μ and ν be as in Lemma 7.3.7. If $b < \infty$ let $h(b)$ be defined as in (7.3.10), and if $b = \infty$, define $h(b) = 0$. Let f be any $C^1((a,b))$ function with $f'(x) \ge 0$ for all x and $\lim_{x\downarrow a} f(x) = 0$ and $\|f\|_\infty = \lim_{x\uparrow b} f(x) < \infty$. Then*

$$\int_{(a,b)} f'\mathrm{d}\mu = -\int_{(a,b)} f\mathrm{d}\nu + h(b)\|f\|_\infty \ . \tag{7.3.15}$$

Proof. Note that $\|f'\|_{L^1(m)} = \|f\|_{L^\infty(m)}$. Since $f(x) = \int_a^x f'(y)\mathrm{d}y$, making a change in f' that has a small $L^1(m)$ norm results in a change in f that is correspondingly small in the uniform norm. We may therefore suppose without loss of generality that f' has compact support in $[a', b'] \subset (a, b)$. By homogeneity, we may assume that $\|f\|_\infty = 1$. In case $b < \infty$, we may then proceed as in the last lemma, and then (7.3.10), (7.3.12) and (7.3.13) yield (7.3.15).

In case $b = \infty$, pick $\epsilon > 0$. Since ν is a finite Radon measure, by increasing b' as necessary, we may suppose that $\nu((b', b)) < \epsilon$. Let $g \in C_c^1((a, b))$ be such that $g'(x) = f'(x)$ for $x \leq b'$, and such that $|g'(x)| < \epsilon$ for $x > b'$, which is easy to do since $(b', b) = (b', \infty)$. Define $g(x) = \int_a^x g'(y)\mathrm{d}y$, and note that for $x \leq b'$, $g(x) = f(x)$, and we may arrange that $\|g - f\|_\infty = \|f\|_\infty$. Then by (7.3.7),

$$\begin{aligned}
\int_{(a,b)} f'\mathrm{d}\mu &= \int_{(a,b)} g'\mathrm{d}\mu - \int_{(b',b)} g'\mathrm{d}\mu \\
&= -\int_{(a,b)} g\mathrm{d}\nu - \int_{(b',b)} g'\mathrm{d}\mu \\
&= -\int_{(a,b)} f\mathrm{d}\nu + \int_{(b',b)} (f-g)\mathrm{d}\nu - \int_{(b',b)} g'\mathrm{d}\mu
\end{aligned}$$

Now note that $\left|\int_{(b',b)} (f-g)\mathrm{d}\nu\right| \leq \|f\|_\infty \nu(b', b)$ and $\left|\int_{(b',b)} g'\mathrm{d}\nu\right| \leq \epsilon\|\nu\|_{\mathrm{TV}}$. Thus, recalling that $\|\nu\|_{\mathrm{TV}} \leq C$,

$$\left|\int_{(a,b)} f'\mathrm{d}\mu + \int_{(a,b)} f\mathrm{d}\nu\right| \leq \epsilon(\|f\|_\infty + C) \ .$$

Since ϵ is arbitrary, the identity (7.3.15) is proved. ☐

7.3.11 THEOREM. *$L \in (\mathcal{C}_0((a, b)))^*$ be such that for some $C < \infty$ and all all $g \in \mathcal{C}_c^1((a, b))$, (7.3.4) is valid. Let μ and ν be the unique signed Borel measures on (a, b) such that (7.3.5) and (7.3.6) are valid for all $f \in \mathcal{C}_0((a, b)$ and all $f \in \mathcal{C}_c^1((a, b)$ respectively.*

Then μ is absolutely continuous with respect to Lebesgue measure. Let h denote the Radon-Nikodymn derivative of μ with respect to Lebesgue measure so that by the Lebesgue Differentiation Theorem, at almost every x,

$$h(x) = \lim_{y \downarrow x} \frac{\mu((x, y)}{y - x} \ . \tag{7.3.16}$$

Then there is a preferred representative of the a.e. equivalence class of h such that (7.3.16) is valid for every *$x \in (a, b)$, and with this version of h,*

$$TV(h; (a, b)) \leq C \ , \tag{7.3.17}$$

and for all $a \leq x < y < b$,

$$h(x) - h(y) = \nu((x, y]) \ . \tag{7.3.18}$$

In particular, h is right continuous and has a left limit at each point $x \in (a, b)$.

Before giving the proof, we recall a basic measure theoretic result that we shall use. Let $\mathcal{A}$ denote the half-open interval algebra which consists of all finite disjoint unions of sets of the form $(c, d]$ with $a \leq c < d < b$ or of the form (c, b) where $a < c < b$. By a standard application of the Monotone Class Theorem, for every positive Radon measure λ on $\mathbb{R}$ and every Borel set E, for each $\epsilon > 0$, there is a set $A \in \mathcal{A}$ such that $\lambda(E \Delta A) < \epsilon$.

Proof. Let $\mu = \mu_+ - \mu_-$ be the unique decomposition of μ into a different of mutually singular finite (positive) Borel measures, and let $|\mu| = \mu_+ + \mu_-$, and likewise define ν_+ and ν_- in terms of ν. Let m denote Lebesgue measure and note that $|\mu| + m$ is a Radon measure.

Let E be a Borel set such that $\mu_-(E) = 0$ and $\mu_+(E^c) = 0$. Then for any $\epsilon > 0$, there exists $A_\epsilon \in \mathcal{A}$ such that

$$(|\mu| + m)(E \Delta A_\epsilon) < \epsilon \ .$$

Now let F be any Borel subset of E and suppose that $\mu_+(F) > 0$. For $\epsilon > 0$, let $B \in \mathcal{A}$ be such that $\mu(F \Delta B) < \epsilon$. Then since $F \subset E$, $1_B 1_{A_\epsilon} = (1_B - 1_F)1_{A_\epsilon} + 1_F(1_{A_\epsilon} - 1_E) + 1_F$. It follows that

$$(|\mu| + m)(B \cap A_\epsilon - F) \leq 2\epsilon \ . \tag{7.3.19}$$

Since $B \cap A_\epsilon \in \mathcal{A}$, it can be written in as $B \cap A_\epsilon = \displaystyle\sum_{j-1}^{n} (c_j, d_j]$, where $c_1 < d_1 < c_2 < \cdots < d_n$. At the cost of another ϵ, we may assume that $c_1 > -\infty$ and $d_n < \infty$, which is automatic in case a and b are finite.

Now for $j = 1, \ldots, n$ choose disjoint open intervals U_j such that $[c_j, d_j] \subset U_j$ and such that

$$\sum_{j=1}^{n} (|\mu| + m)(U_j) \leq (|\mu| + m)(B \cap A_\epsilon) + \epsilon \ , \tag{7.3.20}$$

which is possible by the outer regularity of $|\mu| + m$. For each j, choose $[c_j, d_j] \prec g_j \prec U_j$, and define $g := \sum_{j=1}^{n} g_j$. Define $f(x) = \displaystyle\int_a^x \sum_{j=1}^{n} g_j(y)\mathrm{d}y$.

Then

$$\|f\|_\infty \leq \sum_{j=1}^{n} m(U_j) \leq \epsilon + m(B \cap A_\epsilon) \leq 3\epsilon + m(F) \ . \tag{7.3.21}$$

Also, with $U := \cap_{j=1}^{n} U_j$,

$$\int_{(a,b)} f' \mathrm{d}\mu \geq \mu_+(B \cap A_\epsilon) - \mu_-(U) \ .$$

By (7.3.20), $\mu_-(U) \leq \mu_-(A_\epsilon) + \epsilon \leq 2\epsilon$. By (7.3.19), $\mu_+(B \cap A_\epsilon) \geq \mu_+(F) - \epsilon$. Altogether, using Lemma 7.3.10 together with the positivity of f,

$$\begin{aligned} \mu_+(F) &\leq 3\epsilon + \int_{(a,b)} f' \mathrm{d}\mu \\ &\leq 3\epsilon + \int_{(a,b)} f \mathrm{d}\nu_- + h(b)\|f\|_\infty \\ &\leq 3\epsilon + (C + h(b))\|f\|_\infty \ . \end{aligned}$$

Combining this with (7.3.11) and (7.3.21), we have

$$\mu_+(F) \leq 3\epsilon + (2C + (b-a)^{-1})(3\epsilon + m(F)) \ .$$

Since $\epsilon > 0$ is arbitrary, $\mu_+(F) \leq (2C + (b-a)^{-1}\|\mu\|_{\mathrm{TV}})m(F)$. and then since F is arbitrary, this proves that μ_+ is absolutely continuous with respect to Lebesgue measure, and then using the Lebesgue Differentiation Theorem that the Radon-Nikodymn derivative h satisfies $\|h\|_\infty \leq 2C + (b-a)^{-1}\|\mu\|_{\mathrm{TV}}$.

For the final part, given any $a \leq x < y < b$, for n sufficiently large that and $y - 1/n > x$, define the "ramp function" f_n approximation of $1_{(x,y]}$ by

$$f_n(t) = \begin{cases} n(t-x) & x < t \leq x + 1/n \\ 1 & x + 1/n \leq t \leq y \\ 1 - n(t-y) & y < t \leq y + 1/n \end{cases}$$

and $f_n(t) = 0$ for all other t. Notice that $\lim_{n\to\infty} f_n(t) = 1_{(x,y]}(t)$ for all t. Though f is not in $C^1_c((a,b))$, a simple approximation argument such as we made in the proof of Lemma 7.3.9 shows that (7.3.7) is nonetheless valid and therefore

$$n\mu((x - 1/n, x)) - n\mu((y - 1/n, y)) = -\int f_n \mathrm{d}\nu \ .$$

Taking the limit $n \to \infty$, we obtain $h(y) - h(x) = \nu((x, y])$. It now follows that when $a < x_0 < x_1 \cdots x_n < b$, $\sum_{j=1}^{n} |h(x_j) - h(x_{j-1})| = \sum_{j=1}^{n} |\nu((x_{j-1}, x_j]) \leq \|\nu\|_{\mathrm{TV}}$. □

7.3.12 COROLLARY. *Let h be integrable on (a,b) and suppose that $TV(h;(a,b)) < \infty$. Then in the a.e. equivalence class of h there is a preferred representative, also denoted by h, such that h is right continuous and has a left limit at each point $x \in (a,b)$. Moreover, h is the difference of two monotone non-decreasing right continuous functions $h = h_+ - h_-$.*

Proof. The first part is a direct consequence of Lemma 7.3.6 and Theorem 7.3.11. Let ν be the signed measure associated to h as in Theorem 7.3.11. Let $\nu = nu_+ - \nu_-$ be the Hahn-Saks decomposition of ν into its positive and negative parts. Then since for all $a < x < b$, $h(x) - h(a) = \nu_-((a,x]) - \nu_+((a,x])$. Define $h_+(x) = h(a) + \nu_-((a,x])$ and define $h_-(x) = \nu_+((a,x])$. □

7.3.3 The Banach space $BV((a,b))$

7.3.13 DEFINITION (BV((a,b)))**.** The real normed space $BV((a,b))$, called the space of BV functions on (a,b) is the vector space of real integrable functions h on (a,b) such that $TV(h;(a,b)) < \infty$. For $h \in BV((a,b))$, let $\mu = h\mathrm{d}x$, so that by Lemma 7.3.6 and Lemma **??**, there is a signed Borel measure ν with $\|\nu\|_{\mathrm{TV}} = TV(h;(a,b)) < \infty$, and such that (7.3.7) is valid for all $f \in \mathcal{C}^1_c((a,b))$. Then we define the BV norm of h as

$$\|h\|_{\mathrm{BV}} = \|h\|_{L^1(m)} + \|\nu\|_{\mathrm{TV}} = \|\mu\|_{\mathrm{TV}} + \|\nu\|_{\mathrm{TV}} \tag{7.3.22}$$

By the uniqueness in Lemma 7.3.7 and the linearity of (7.3.7), the map $h \mapsto \nu$ is linear, and so it is evident that $\|\cdot\|_{\mathrm{BV}}$ is a norm on $BV((a,b))$. To see that $BV((a,b))$ is complete in this norm, Let $\{h_n\}_{n\in\mathbb{N}}$ be a Cauchy sequence in $BV((a,b))$. Then $\{h_n\}_{n\in\mathbb{N}}$ is also a Cauchy sequence in $L^1((a,b),\mathcal{B},m)$ and so there exist $h \in L^1((a,b),\mathcal{B},m)$ such that $\lim_{n\to\infty} \|h_n - h\|_{L^1(m)} = 0$. For each $n \in \mathbb{N}$, let μ_n be the signed Borel measures such that

$$\int_{(a,b)} f' h_n \mathrm{d}x = -\int_{(a,b)} f \mathrm{d}\nu_n$$

for all $f \in \mathcal{C}_0((a,b))$.

Since the space of signed Borel measures on $\mathcal{C}_0((a,b))$ is, like every dual space, complete, and since by the definition of the BV norm $\{\nu_n\}_{n\in\mathbb{N}}$ is a Cauchy sequence in the total variation norm, there exists a signed measure ν such that $\lim_{n\to\infty} \|\nu_n - \nu\|_{\mathrm{TV}} = 0$. It follows that

$$\int_{(a,b)} f' h \mathrm{d}x = -\int_{(a,b)} f \mathrm{d}\nu$$

for all $f \in \mathcal{C}_0((a,b))$, and hence $h \in BV((a,b))$ and $\lim_{n\to\infty} \|h_n - h\|_{\mathrm{BV}} = 0$.

There is in fact a better way to see that $BV((a,b))$ is a Banach space: It turns out the $BV((a,b))$ is the dual of another Banach space. Let Y be the Banach space of of all pairs (f_1, f_2) with $f_1, f_2 \in \mathcal{C}_0((a,b))$ and with the norm

$$\|(f_1, f_2)\|_Y = \|f_1\|_\infty + \|f_2\|_\infty \ .$$

The dual space Y^* is the space of all pairs (μ_1, μ_2) of signed Borel measures on (a,b) with the norm

$$\|(\mu_1, \mu_2)\|_Y = \|\mu_1\|_{\mathrm{TV}} + \|\mu_2\|_{\mathrm{TV}} \ .$$

Now let Z be closure of the subspace of Y consisting on (f_1, f_2) with $f_1, f_2 \in \mathcal{C}_c^1((a,b))$ and $f_1 = -f_2'$. The annihilator of Z is the subspace of Y^* consisting of pairs (μ, ν) such that

$$\int_{(a,b)} f' \mathrm{d}\mu = \int_{(a,b)} f' \mathrm{d}\nu \tag{7.3.23}$$

for all $f \in \mathcal{C}_c^1((a,b))$. By Theorem **??**, (μ, ν) belongs to the annihilator of Z precisely when $\mu = h\mathrm{d}x$ where $h \in BV((a,b))$, and then μ is the unique signed measure associated to h such (7.3.23) is valid for all $f \in \mathcal{C}_c^1((a,b))$, and in this case

$$\|h\|_{\mathrm{BV}} = \|\mu\|_{\mathrm{TV}} + \|\nu\|_{\mathrm{TV}} = \|(\mu, \nu)\|_{Y^*} \ .$$

Since Z is a closed subspace of a Banach space Y, the annihilator of Z in Y^* is the dual of Y/Z by a general result in the theory of Banach spaces. Therefore, $BV((a,b))$ is the dual of Y/Z. By the Banach-Aloglu Theorem, the unit closed ball in $BV((a,b))$ is compact in the weak-$*$ topology induced on $BV((a,b))$ by Y/Z.

The space $BV((a,b))$ is not separable: For each $y \in (a,b)$, let h_y be the step function with $h_y(x) = 1$ for $y \geq x$ and $h_y(x) = 0$ otherwise. Then for $y \neq z$, $\|h_y - h_z\|_{\mathrm{BV}} \geq 2$.