MAA6617 COURSE NOTES
SPRING 2017

Contents

20. Normed vector spaces

Let $\mathcal{X}$ be a vector space over a field $\mathbb{K}$ (in this course we always have either $\mathbb{K} = \mathbb{R}$ or $\mathbb{K} = \mathbb{C}$).

Definition 20.1. A *norm* on $\mathcal{X}$ is a function $\|\cdot\| : \mathcal{X} \to \mathbb{R}$ satisfying:

(i) (positivity) $\|x\| \geq 0$ for all $x \in \mathcal{X}$, and $\|x\| = 0$ if and only if $x = 0$;
(ii) (homogeneity) $\|kx\| = |k|\|x\|$ for all $x \in \mathcal{X}$ and $k \in \mathbb{K}$, and
(iii) (triangle inequality) $\|x + y\| \leq \|x\| + \|y\|$ for all $x, y \in \mathcal{X}$.

$\triangleleft$

Using these three properties it is straightforward to check that the quantity

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

defines a metric on $\mathcal{X}$. The resulting topology is the *norm topology.* The next proposition is simple but fundamental; it says that the norm and the vector space operations are continuous in the norm topology.

Proposition 20.2 (Continuity of vector space operations)**.** *Let $\mathcal{X}$ be a normed vector space over $\mathbb{K}$.*

a) *If (x_n) converges to x in $\mathcal{X}$, then $(\|x_n\|)$ converges to $\|x\|$ in $\mathbb{R}$.*
b) *If (k_n) converges to k in $\mathbb{K}$ and (x_n) converges to x in $\mathcal{X}$, then $(k_n x_n)$ converges to kx in $\mathcal{X}$.*
c) *If (x_n) converges to x and (y_n) converges to y in $\mathcal{X}$, then $(x_n + y_n)$ converges to $x+y$ in $\mathcal{X}$.*

Proof. The proofs follow readily from the properties of the norm, and are left as exercises. □

Two norms $\|\cdot\|_1, \|\cdot\|_2$ on $\mathcal{X}$ are *equivalent* if there exist absolute constants $C, c > 0$ such that

$$c\|x\|_1 \leq \|x\|_2 \leq C\|x\|_1 \quad \text{for all } x \in \mathcal{X}.$$

Equivalent norms determine the same topology on $\mathcal{X}$ and the same Cauchy sequences (Problem 20.2). A normed space is a *Banach space* if it is complete in the norm topology. It follows that if $\mathcal{X}$ is equipped with two equivalent norms $\|\cdot\|_1, \|\cdot\|_2$ then it is complete (a Banach space) in one norm if and only if it is complete in the other.

The following proposition gives a convenient criterion for a normed vector space to be complete. A series $\sum_{n=1}^\infty x_n$ in $\mathcal{X}$ is *absolutely convergent* if $\sum_{n=1}^\infty \|x_n\| < \infty$. The series *converges* in $\mathcal{X}$ if the limit $\lim_{N\to\infty} \sum_{n=1}^N x_n$ exists in $\mathcal{X}$ (in the norm topology). (Quite explicitly, the series $\sum_{n=1}^\infty x_n$ converges to $x \in \mathcal{X}$ if $\lim_{N\to\infty} \left\| x - \sum_{n=1}^N x_n \right\| = 0$.)

Proposition 20.3. *A normed space $(\mathcal{X}, \|\cdot\|)$ is complete if and only if every absolutely convergent series in $\mathcal{X}$ is convergent.*

Proof. First suppose $\mathcal{X}$ is complete and $\sum_{n=1}^\infty x_n$ is absolutely convergent. Write $s_N = \sum_{n=1}^N x_n$ for the N^{th} partial sum and let $\epsilon > 0$ be given. Since $\sum_{n=1}^\infty \|x_n\|$ is convergent, there exists an L such that $\sum_{n=L}^\infty \|x_n\| < \epsilon$. If $N > M \geq L$, then

$$\|s_N - s_M\| = \left\| \sum_{n=M+1}^N x_n \right\| \leq \sum_{n=M+1}^N \|x_n\| < \epsilon.$$

Thus the sequence (s_N) is Cauchy in $\mathcal{X}$, hence convergent by hypothesis.

Conversely, suppose every absolutely convergent series in $\mathcal{X}$ is convergent. Given a Cauchy sequence (x_n) from X, choose a super-Cauchy subsequence (y_k); i.e., $(y_k = x_{n_k})_k$

and

$$\sum_{k=1}^{\infty} \|y_{k+1} - y_k\| < \infty.$$

(To do this, first choose N_1 such that $\|x_n - x_m\| < 2^{-1}$ for all $n, m \geq N_1$. Next choose $N_2 > N_1$ such that $\|x_n - x_m\| < 2^{-2}$ for all $n, m \geq N_2$. Continuing in this way recursively defines an increasing sequence of integers $(N_k)_{k=1}^{\infty}$ such that $\|x_n - x_m\| < 2^{-k}$ for all $n, m \geq N_k$. Set $y_k = x_{N_k}$.) The series $\sum_{k=1}^{\infty}(y_{k+1} - y_k)$ is absolutely convergent and hence, by hypothesis, convergent in $\mathcal{X}$. In other words, the sequence $y_k - y_1$ of partial sums converges in $\mathcal{X}$ which means that (x_n) has a convergent subsequence. The proof is finished by invoking a standard fact about convergence in metric spaces: if (x_n) is a Cauchy sequence which has a convergent subsequence, then the full sequence converges. □

20.1. **Examples.**

(a) Of course, $\mathbb{K}^n$ with the usual Euclidean norm $\|(x_1, \dots x_n)\| = \left(\sum_{k=1}^{n} |x_k|^2\right)^{1/2}$ is a Banach space. The vector space $\mathbb{K}^n$ can also be equipped with the ℓ^p*-norms*

$$\|(x_1, \dots x_n)\|_p := \left(\sum_{k=1}^{n} |x_k|^p\right)^{1/p}$$

for $1 \leq p < \infty$, and the ℓ^∞*-norm*

$$\|(x_1, \dots x_n)\|_\infty := \max(|x_1|, \dots |x_n|).$$

It is not too hard to show that all of the ℓ^p norms $(1 \leq p \leq \infty)$ are equivalent on $\mathbb{K}^n$ (though the constants c, C depend on the dimension n). It turns out that *any* two norms on a finite-dimensional vector space are equivalent. As a corollary, every finite-dimensional normed space is a Banach space. See Problem 20.3.

(b) (Sequence spaces) Define

$$c_0 := \{f : \mathbb{N} \to \mathbb{K} | \lim_{m\to\infty} |f(m)| = 0\}$$
$$\ell^\infty := \{f : \mathbb{N} \to \mathbb{K} | \sup_{m\in\mathbb{N}} |f(m)| < \infty\}$$
$$\ell^1 := \{f : \mathbb{N} \to \mathbb{K} | \sum_{m=0}^{\infty} |f(m)| < \infty\}.$$

It is a simple exercise to check that each of these is a vector space (a subspace of the vector space of all functions $f : \mathbb{N} \to \mathbb{K}$). Define, for functions $f : \mathbb{N} \to \mathbb{K}$,

$$\|f\|_\infty := \sup_m |f(m)|$$
$$\|f\|_1 := \sum_{m=1}^{\infty} |f(m)|.$$

Then $\|f\|_\infty$ defines a norm on both c_0 and ℓ^∞, and $\|f\|_1$ is a norm on ℓ^1. Equipped with these respective norms, each is a Banach space. We sketch the proof for c_0. Verification of the other two assertions is left as exercises (Problem 20.4).

The key observation is that (f_n) converges to f in the $\|\cdot\|_\infty$ norm if and only if (f_n) converges to f uniformly as functions on $\mathbb{N}$. Suppose (f_n) is a Cauchy sequence in c_0. Then the sequence of functions f_n is uniformly Cauchy on $\mathbb{N}$, and in particular converges pointwise to a function f. To check completeness it will suffice to show that also $f \in c_0$. To this end, let $\epsilon > 0$ be given. There is an N so that $\|f_n - f_N\| < \epsilon$ for $n \geq N$. Thus, for all m and all $n \geq N$, $|f_n(m) - f_N(m)| < \epsilon$ and thus, $|f(m) - f_N(m)| \leq \epsilon$ for all m. There is an M so that $|f_N(m)| < \epsilon$ for $m \geq M$. Hence, for such m,

$$|f(m)| \leq |f_N(m)| + \epsilon < 2\epsilon,$$

and the conclusion follows.

Along with these spaces it is also helpful to consider the vector space

$$c_{00} := \{f : \mathbb{N} \to \mathbb{K} | f(n) = 0 \text{ for all but finitely many } n\}$$

Notice that c_{00} is a vector subspace of each of c_0, ℓ^1 and ℓ^∞. Thus it can be equipped with either the $\|\cdot\|_\infty$ or $\|\cdot\|_1$ norms. It is not complete in either of these norms, however. What is true is that c_{00} is *dense* in c_0 and ℓ^1 (but not in ℓ^∞). (See Problem 20.9).

(c) (L^1 spaces) Let $(X, \mathscr{M}, m)$ be a measure space. The quantity

$$\|f\|_1 := \int_X |f|\, dm$$

defines a norm on $L^1(m)$, provided we agree to identify f and g when $f = g$ a.e. (Indeed the chief motivation for making this identification is that it makes $\|\cdot\|_1$ into a norm. Note that ℓ^1 from the last example is a special case of this (what is the measure space?))

Proposition 20.4. *$L^1(m)$ is a Banach space.*

Proof. It suffices to verify the hypotheses of Proposition 20.3. If $\sum_{n=1}^\infty f_n$ is absolutely convergent, then the function $g := \sum_{n=1}^\infty |f_n|$ belongs to L^1 and is thus finite m-a.e. In particular the sequence of partial sums $s_N = \sum_{n=1}^N f_n(x)$ is a sequence of measurable functions with $|s_N| \leq g$ which converges to f. Hence by the DCT and its corollary, $f \in L^1$ and the partial sums $(s_N)_N$ converges to f in L^1. □

(d) (L^p spaces) Again let $(X, \mathscr{M}, m)$ be a measure space. For $1 \leq p < \infty$ let $L^p(m)$ denote the set of measurable functions f for which

$$\|f\|_p := \left(\int_X |f|^p\, dm\right)^{1/p} < \infty$$

(again we identify f and g when $f = g$ a.e.). It turns out that this quantity is a norm on $L^p(m)$, and $L^p(m)$ is complete, though we will not prove this yet (it is not immediately obvious that the triangle inequality holds when $p > 1$). The sequence space ℓ^p is defined analogously: it is the set of $f : \mathbb{N} \to \mathbb{K}$ for which

$$\|f\|_p := \left(\sum_{n=1}^{\infty} |f(n)|^p \right)^{1/p} < \infty$$

and this quantity is a norm making ℓ^p into a Banach space.

When $p = \infty$, we define $L^\infty(m)$ to be the set of all functions $f : X \to \mathbb{K}$ with the following property: there exists $M > 0$ such that

$$|f(x)| \leq M \quad \text{for } m-\text{a.e. } x \in X; \tag{1}$$

as for the other L^p spaces we identify f and g when there are equal a.e. When $f \in L^\infty$, let $\|f\|_\infty$ be the smallest M for which (1) holds. Then $\|\cdot\|_\infty$ is a norm making $L^\infty(m)$ into a Banach space.

(e) ($C(X)$ spaces) Let X be a compact metric space and let $C(X)$ denote the set of continuous functions $f : X \to \mathbb{K}$. It is a standard fact from advanced calculus that the quantity $\|f\|_\infty := \sup_{x\in X} |f(x)|$ is a norm on $C(X)$. A sequence is Cauchy in this norm if and only if it is uniformly Cauchy. It is thus also a standard fact that $C(X)$ is complete in this norm—completeness just means that a uniformly Cauchy sequence of continuous functions on X converges uniformly to a continuous function.

This example can be generalized somewhat: let X be a locally compact metric space. Say a function $f : X \to \mathbb{K}$ *vanishes at infinity* if for every $\epsilon > 0$, there exists a compact set $K \subset X$ such that $\sup_{x\notin K} |f(x)| < \epsilon$. Let $C_0(X)$ denote the set of continuous functions $f : X \to \mathbb{K}$ that vanish at infinity. Then $C_0(X)$ is a vector space, the quantity $\|f\|_\infty := \sup_{x\in X} |f(x)|$ is a norm on $C_0(X)$, and $C_0(X)$ is complete in this norm. (Note that c_0 from above is a special case.)

(f) (Subspaces and direct sums) If $(\mathcal{X}, \|\cdot\|)$ is a normed vector space and $\mathcal{Y} \subset \mathcal{X}$ is a vector subspace, then the restriction of $\|\cdot\|$ to $\mathcal{Y}$ is clearly a norm on $\mathcal{Y}$. If $\mathcal{X}$ is a Banach space, then $(\mathcal{Y}, \|\cdot\|)$ is a Banach space if and only if $\mathcal{Y}$ is *closed* in the norm topology of $\mathcal{X}$. (This is just a standard fact about metric spaces—a subspace of a complete metric space is complete in the restricted metric if and only if it is closed.)

If $\mathcal{X}, \mathcal{Y}$ are vector spaces then the *algebraic direct sum* is the vector space of ordered pairs

$$\mathcal{X} \oplus \mathcal{Y} := \{(x, y) : x \in \mathcal{X}, y \in \mathcal{Y}\}$$

with entrywise operations. If $\mathcal{X}$, $\mathcal{Y}$ are equipped with norms $\|\cdot\|_{\mathcal{X}}, \|\cdot\|_{\mathcal{Y}}$, then each of the quantities

$$\|(x, y)\|_\infty := \max(\|x\|_{\mathcal{X}}, \|y\|_{\mathcal{Y}}),$$
$$\|(x, y)\|_1 := \|x\|_{\mathcal{X}} + \|y\|_{\mathcal{Y}}$$

is a norm on $\mathcal{X} \oplus \mathcal{Y}$. These two norms are equivalent; indeed it follows from the definitions that

$$\|(x, y)\|_\infty \leq \|(x, y)\|_1 \leq 2\|(x, y)\|_\infty.$$

If $\mathcal{X}$ and $\mathcal{Y}$ are both complete, then $\mathcal{X} \oplus \mathcal{Y}$ is complete in both of these norms. The resulting Banach spaces are denoted $\mathcal{X} \oplus_\infty \mathcal{Y}$, $\mathcal{X} \oplus_1 \mathcal{Y}$ respectively.

(g) (Quotient spaces) If $\mathcal{X}$ is a normed vector space and $\mathcal{M}$ is a proper subspace, then one can form the *algebraic quotient* $\mathcal{X}/\mathcal{M}$, defined as the collection of distinct cosets $\{x + \mathcal{M} : x \in \mathcal{X}\}$. From linear algebra, $\mathcal{X}/\mathcal{M}$ is a vector space under the standard operations. If $\mathcal{M}$ is a *closed* subspace of $\mathcal{X}$, then the quantity

$$\|x + \mathcal{M}\| := \inf_{y \in \mathcal{M}} \|x - y\|$$

is a norm on $\mathcal{X}/\mathcal{M}$, called the *quotient norm*. (Geometrically, $\|x + \mathcal{M}\|$ is the distance in $\mathcal{X}$ from x to the closed set $\mathcal{M}$.) It turns out that if $\mathcal{X}$ is complete, so is $\mathcal{X}/\mathcal{M}$. See Problem 20.20.

More examples are given in the exercises and further examples will appear after the development of some theory.

20.2. Linear transformations between normed spaces.

Definition 20.5. Let $\mathcal{X}, \mathcal{Y}$ be normed vector spaces. A linear transformation $T : \mathcal{X} \to \mathcal{Y}$ is *bounded* if there exists a constant $C > 0$ such that $\|Tx\|_{\mathcal{Y}} \leq C\|x\|_{\mathcal{X}}$ for all $x \in \mathcal{X}$. $\triangleleft$

Remark 20.6. Note that in this definition it would suffice to require that $\|Tx\|_{\mathcal{Y}} \leq C\|x\|_{\mathcal{X}}$ just for all $x \neq 0$, or for all x with $\|x\|_{\mathcal{X}} = 1$ (why?) $\diamond$

The importance of boundedness and the following simple proposition is hard to overstate. Recall, a mapping $f : X \to Y$ between metric spaces is *Lipschitz continuous* if there is a constant $C > 0$ such that $d(f(x), f(y)) \leq C d(x, y)$ for all $x, y \in X$. A simple exercise shows Lipschitz continuity implies continuity.

Proposition 20.7. *If $T : \mathcal{X} \to \mathcal{Y}$ is a linear transformation between normed spaces, then the following are equivalent:*

(i) T is bounded.
(ii) T is Lipschitz continuous.
(iii) T is uniformly continuous.
(iv) T is continuous.
(v) T is continuous at 0.

Moreover, in this case,

$$\begin{aligned} \|T\| &:= \sup\{\|Tx\| : \|x\| = 1\} \\ &= \sup\{\frac{\|Tx\|}{\|x\|} : x \neq 0\} \\ &= \inf\{C : \|Tx\| \leq C\|x\| \text{ for all } x \in \mathcal{X}\} \end{aligned}$$

and $\|T\|$ is the smallest number (the infimum is attained) such that

$$\|Tx\| \leq \|T\| \, \|x\|, \tag{2}$$

for all $x \in \mathcal{X}$.

Proof. Suppose T is bounded so that there exists a $C > 0$ such that $\|Tx\| \le C\|x\|$ for all $x \in \mathcal{X}$. Thus, if $x, y \in \mathcal{X}$, then, $\|Tx - Ty\| = \|T(x-y)\| \le C\|x-y\|$ by linearity of T. Hence (i) implies (ii). The implications (ii) implies (iii) implies (iv) implies (v) are evident. The proof of (v) implies (i) exploits the homogeneity of the norm and the linearity of T. By hypothesis, with $\epsilon = 1$ there exists $\delta > 0$ such that if $\|x\| < \delta$, then $\|Tx\| < 1$. Fix a nonzero vector $x \in \mathcal{X}$ and a real number $0 < \lambda < \delta$. The vector $\lambda x/\|x\|$ has norm less than δ, so

$$\left\| T\left(\frac{\lambda x}{\|x\|}\right)\right\| = \lambda \frac{\|Tx\|}{\|x\|} < 1.$$

Rearranging this we find $\|Tx\| \le (1/\lambda)\|x\|$ for all $x \neq 0$, which shows T is bounded; in fact we can take $C = \frac{1}{\delta}$.

The rest of the proof is left as an exercise. □

The set of all bounded linear operators from $\mathcal{X}$ to $\mathcal{Y}$ is denoted $B(\mathcal{X}, \mathcal{Y})$. It is a vector space under the operations of pointwise addition and scalar multiplication. The quantity $\|T\|$ is called the *operator norm* of T.

Proposition 20.8. *For normed vector spaces $\mathcal{X}$ and $\mathcal{Y}$, the operator norm makes $B(\mathcal{X}, \mathcal{Y})$ into a normed vector space, which is complete if $\mathcal{Y}$ is complete.*

Proof. That $B(\mathcal{X}, \mathcal{Y})$ is a normed vector space follows readily from the definitions and is left as an exercise. Suppose now $\mathcal{Y}$ is complete, and let T_n be a Cauchy sequence in $B(\mathcal{X}, \mathcal{Y})$. For each $x \in \mathcal{X}$, we have

$$\|T_n x - T_m x\| = \|(T_n - T_m)x\| \le \|T_n - T_m\|\|x\|$$

which shows that $(T_n x)$ is a Cauchy sequence in $\mathcal{Y}$. By hypothesis, $T_n x$ converges in $\mathcal{Y}$. Define $T : \mathcal{X} \to \mathcal{Y}$ by setting $Tx := y$. It is straightforward to check that T is linear.

Let B denote the closed unit ball in $\mathcal{X}$. The sequence $(T_n|_B)$ is uniformly Cauchy and converges pointwise to $T|_B$. Hence $T|_B$ is continuous and $(T_n|_B)$ converges uniformly to $T|_B$. It follows that T is continuous at 0 and therefore T is continuous. Since $\|T_n - T\| = \sup\{\|(T_n - T)x\| : x \in B\}$ it follows that (T_n) converges to T in $B(\mathcal{X}, \mathcal{Y})$. □

If $T \in B(\mathcal{X}, \mathcal{Y})$ and $S \in B(\mathcal{Y}, \mathcal{Z})$ then two applications of the the inequality (2) gives, for $x \in \mathcal{X}$,

$$\|STx\| \le \|S\|\|Tx\| \le \|S\|\|T\|\|x\|$$

and it follows that $ST \in B(\mathcal{X}, \mathcal{Z})$ and $\|ST\| \le \|S\|\|T\|$. In the special case that $\mathcal{Y} = \mathcal{X}$ is complete, $B(\mathcal{X}) := B(\mathcal{X}, \mathcal{X})$ is an example of a *Banach algebra*.

The following proposition is very useful in constructing bounded operators—at least when the codomain is complete. Namely, it suffices to define the operator (and show that it is bounded) on a dense subspace.

Proposition 20.9 (Extending bounded operators)**.** *Let $\mathcal{X}$, $\mathcal{Y}$ be normed vector spaces with $\mathcal{Y}$ complete, and $\mathcal{E} \subset \mathcal{X}$ a dense linear subspace. If $T : \mathcal{E} \to \mathcal{Y}$ is a bounded linear* *operator, then there exists a unique bounded linear operator* $\widetilde{T} : \mathcal{X} \to \mathcal{Y}$ *extending* T *(so* $\widetilde{T}|_{\mathcal{E}} = T$*), and* $\|\widetilde{T}\| = \|T\|$.

Proof. Recall, if X, Y are metric spaces, Y is complete, $D \subset X$ is dense and $f : D \to Y$ is uniformly continuous, then f has a unique continuous extension $\tilde{f} : X \to Y$. Moreover, this extension can be defined as follows. Given $x \in X$, choose a sequence (x_n) from D converging to x and let $\tilde{f}(x) = \lim f(x_n)$ (that the sequence $f(x_n)$ is Cauchy follows from uniform continuity; that it converges from the assumption that $\mathcal{Y}$ is complete and finally it is an exercise to show $\tilde{f}(x)$ is well defined independent of the choice of (x_n)). Thus, it only remains to verify that the extension $\tilde{T}$ of T is in fact linear and $\|T\| = \|\tilde{T}\|$, both of which are routine exercises. □

Remark: The completeness of $\mathcal{Y}$ is essential in the above proposition; Problem 20.11 suggests a counterexample.

A bounded linear transformation $T \in B(\mathcal{X}, \mathcal{Y})$ is said to be *invertible* if it is bijective (automatically T^{-1} exists and is a linear transformation) and T^{-1} is bounded from $\mathcal{Y}$ to $\mathcal{X}$. Two normed spaces $\mathcal{X}, \mathcal{Y}$ are said to be *(boundedly) isomorphic* if there exists an invertible linear transformation $T : \mathcal{X} \to \mathcal{Y}$.

An operator $T : \mathcal{X} \to \mathcal{Y}$ such that $\|Tx\| = \|x\|$ for all $x \in \mathcal{X}$ is an *isometry.* Note that an isometry is automatically injective and if it is also surjective then it is automatically invertible and T^{-1} is also an isometry. An isometry need not be surjective, however. The normed vector spaces are *isometrically isomorphic* if there is an invertible isometry $T : \mathcal{X} \to \mathcal{Y}$.

20.3. **Examples.**

(a) If $\mathcal{X}$ is a finite-dimensional normed space and $\mathcal{Y}$ is any normed space, then every linear transformation $T : \mathcal{X} \to \mathcal{Y}$ is bounded.

(b) Let $\mathcal{X}$ be c_{00} equipped with the $\|\cdot\|_1$ norm, and $\mathcal{Y}$ be c_{00} equipped with the $\|\cdot\|_\infty$ norm. Then the identity map $\mathrm{id} : c_{00} \to c_{00}$ is bounded as an operator from $\mathcal{X}$ to $\mathcal{Y}$ (in fact its norm is equal to 1), but is unbounded as an operator from $\mathcal{Y}$ to $\mathcal{X}$.

(c) Consider c_{00} with the $\|\cdot\|_\infty$ norm. Let $a : \mathbb{N} \to \mathbb{K}$ be any function and define a linear transformation $T_a : c_{00} \to c_{00}$ by
$$T_a f(n) = a(n) f(n). \tag{3}$$
The mapping T_a is bounded if and only if $M = \sup_{n\in\mathbb{N}} |a(n)| < \infty$, in which case $\|T_a\| = M$. In this case, T_a extends uniquely to a bounded operator from c_0 to c_0, and one may check that the formula (3) defines the extension. All of these claims remain true if we use the $\|\cdot\|_1$ norm instead of the $\|\cdot\|_\infty$ norm. In this case, we get a bounded operator from ℓ^1 to itself.

(d) Define $S : \ell^1 \to \ell^1$ by $Sf(1) = 0$ and $Sf(n) = f(n-1)$ when $n > 1$ for $f \in \ell^1$. (Viewing f as a sequence, S shifts the sequence one place to the right and fills in a

0 in the first position). This S is an isometry, but is not surjective. In contrast, if $\mathcal{X}$ is finite-dimensional, then the rank-nullity theorem from linear algebra guarantees that every injective linear map $T : \mathcal{X} \to \mathcal{X}$ is also surjective.

(e) Let $C^\infty([0,1])$ denote the space of functions on $[0,1]$ with continuous derivatives of all orders. The differentiation map $f \to \frac{df}{dx}$ is a linear transformation from $C^\infty([0,1])$ to itself. Since $\frac{d}{dx}e^{tx} = te^{tx}$ for all real t, we find that there is *no* norm on $C^\infty([0,1])$ for which $\frac{d}{dx}$ is bounded.

20.4. **Problems.**

Problem 20.1. Prove Proposition 20.2.

Problem 20.2. Prove equivalent norms define the same topology and the same Cauchy sequences.

Problem 20.3. (a) Prove all norms on a finite dimensional vector space $\mathcal{X}$ are equivalent. (Hint: fix a basis $e_1, \dots e_n$ for $\mathcal{X}$ and define $\|\sum a_k e_k\|_1 := \sum |a_k|$. Compare any given norm $\|\cdot\|$ to this one. Begin by proving that the "unit sphere" $S = \{x \in \mathcal{X} : \|x\|_1 = 1\}$ is compact in the $\|\cdot\|_1$ topology.)

(b) Combine the result of part (a) with the result of Problem 20.2 to conclude that every finite-dimensional normed vector space is complete.

(c) Let $\mathcal{X}$ be a normed vector space and $\mathcal{M} \subset \mathcal{X}$ a finite-dimensional subspace. Prove $\mathcal{M}$ is closed in $\mathcal{X}$.

Problem 20.4. Finish the proofs from Example 20.1(b).

Problem 20.5. A function $f : [0,1] \to \mathbb{K}$ is called *Lipschitz continuous* if there exists a constant C such that

$$|f(x) - f(y)| \leq C|x - y|$$

for all $x, y \in [0,1]$. Define $\|f\|_{Lip}$ to be the best possible constant in this inequality. That is,

$$\|f\|_{Lip} := \sup_{x \neq y} \frac{|f(x) - f(y)|}{|x - y|}$$

Let Lip$[0,1]$ denote the set of all Lipschitz continuous functions on $[0,1]$. Prove $\|f\| := |f(0)| + \|f\|_{Lip}$ is a norm on Lip$[0,1]$, and that Lip$[0,1]$ is complete in this norm.

Problem 20.6. Let $C^1([0,1])$ denote the space of all functions $f : [0,1] \to \mathbb{R}$ such that f is differentiable in $(0,1)$ and f' extends continuously to $[0,1]$. Prove

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

is a norm on $C^1([0,1])$ and that C^1 is complete in this norm. Do the same for the norm $\|f\| := |f(0)| + \|f'\|_\infty$. (Is $\|f'\|_\infty$ a norm on C^1?)

Problem 20.7. Let $(X, \mathscr{M})$ be a measurable space. Let $M(X)$ denote the (real) vector space of all signed measures on $(X, \mathscr{M})$. Prove the *total variation norm* $\|\mu\| := |\mu|(X)$ is a norm on $M(X)$, and $M(X)$ is complete in this norm.

Problem 20.8. Prove, if $\mathcal{X}, \mathcal{Y}$ are normed spaces, then the operator norm is a norm on $B(\mathcal{X}, \mathcal{Y})$.

Problem 20.9. Prove c_{00} is dense in c_0 and ℓ^1. (That is, given $f \in c_0$ there is a sequence f_n in c_{00} such that $\|f_n - f\|_\infty \to 0$, and the analogous statement for ℓ^1.) Using these facts, or otherwise, prove that c_{00} is *not* dense in ℓ^∞. (In fact there exists $f \in \ell^\infty$ with $\|f\|_\infty = 1$ such that $\|f - g\|_\infty \geq 1$ for all $g \in c_{00}$.)

Problem 20.10. Prove c_{00} is not complete in the $\|\cdot\|_1$ or $\|\cdot\|_\infty$ norms. (After we have studied the Baire Category theorem, you will be asked to prove that there is *no* norm on c_{00} making it complete.)

Problem 20.11. Consider c_0 and c_{00} equipped with the $\|\cdot\|_\infty$ norm. Prove there is no bounded operator $T : c_0 \to c_{00}$ such that $T|_{c_{00}}$ is the identity map. (Thus the conclusion of Proposition 20.9 can fail if $\mathcal{Y}$ is not complete.)

Problem 20.12. Prove the $\|\cdot\|_1$ and $\|\cdot\|_\infty$ norms on c_{00} are not equivalent. Conclude from your proof that the identity map on c_{00} is bounded from the $\|\cdot\|_1$ norm to the $\|\cdot\|_\infty$ norm, but not the other way around.

Problem 20.13. a) Prove $f \in C_0(\mathbb{R}^n)$ if and only if f is continuous and $\lim_{|x|\to\infty} |f(x)| = 0$. b) Let $C_c(\mathbb{R}^n)$ denote the set of continuous, compactly supported functions on $\mathbb{R}^n$. Prove $C_c(\mathbb{R}^n)$ is dense in $C_0(\mathbb{R}^n)$ (where $C_0(\mathbb{R}^n)$ is equipped with sup norm).

Problem 20.14. Prove, if $\mathcal{X}, \mathcal{Y}$ are normed spaces and $\mathcal{X}$ is finite dimensional, then every linear transformation $T : \mathcal{X} \to \mathcal{Y}$ is bounded.

Problem 20.15. Prove the claims in Example 20.3(c).

Problem 20.16. Let $g : \mathbb{R} \to \mathbb{K}$ be a (Lebesgue) measurable function. The map $Mg : f \to gf$ is a linear transformation on the space of measurable functions. Prove, if $g \notin L^\infty(\mathbb{R})$, then there is an $f \in L^1(\mathbb{R})$ such that $gf \notin L^1(\mathbb{R})$. Conversely, show if $g \in L^\infty(\mathbb{R})$, then M_g is bounded from $L^1(\mathbb{R})$ to itself and $\|M_g\| = \|g\|_\infty$.

Problem 20.17. Prove the claims about direct sums in Example 20.1(f).

Problem 20.18. Let $\mathcal{X}$ be a normed vector space and $\mathcal{M}$ a proper *closed* subspace. Prove for every $\epsilon > 0$, there exists $x \in \mathcal{X}$ such that $\|x\| = 1$ and $\inf_{y\in\mathcal{M}} \|x - y\| > 1 - \epsilon$. (Hint: take any $u \in \mathcal{X} \setminus \mathcal{M}$ and let $a = \inf_{y\in\mathcal{M}} \|u - y\|$. Choose $\delta > 0$ small enough so that $\frac{a}{a+\delta} > 1 - \epsilon$, and then choose $v \in \mathcal{M}$ so that $\|u - v\| < a + \delta$. Finally let $x = \frac{u-v}{\|u-v\|}$.)

Note that the distance to a (closed) subspace need not be attained. Here is an example. Consider the Banach space $C([0, 1])$ (with the sup norm of course and either real or complex valued functions) and the closed subspace

$$T = \{f \in C([0, 1]) : f(0) = 0 = \int_0^1 f \, dt\}.$$

Using machinery in the next section it will be evident that T is a closed subspace of $C([0, 1])$. For now, it can be easily verified directly. Let g denote the function $g(t) = t$.

Verify that, for $f \in T$, that

$$\frac{1}{2} = \int g\,dt = \int (g - f)\,dt \le \|g - f\|_\infty.$$

In particular, the distance from g to T is at least $\frac{1}{2}$.

Note that the function $h = x - \frac{1}{2}$, while not in T, satisfies $\|g - h\|_\infty = \frac{1}{2}$.

On the other hand, for any $\epsilon > 0$ there is an $f \in T$ so that $\|g - f\|_\infty \le \frac{1}{2} + \epsilon$ (simply modify h appropriately). Thus, the distance from g to T is $\frac{1}{2}$. Now verify, using the inequality above, that h is the only element of $C([0,1])$ such that $\int h\,dt = 0$ and $\|g - h\|_\infty = \frac{1}{2}$.

Problem 20.19. Prove, if $\mathcal{X}$ is an infinite-dimensional normed space, then the unit ball $ball(\mathcal{X}) := \{x \in \mathcal{X} : \|x\| \le 1\}$ is not compact in the norm topology. (Hint: use the result of Problem 20.18 to construct inductively a sequence of vectors $x_n \in \mathcal{X}$ such that $\|x_n\| = 1$ for all n and $\|x_n - x_m\| \ge \frac{1}{2}$ for all $m < n$.)

Problem 20.20. (The quotient norm) Let $\mathcal{X}$ be a normed space and $\mathcal{M}$ a proper closed subspace.

a) Prove the quotient norm is a norm (see Example 20.1(g)).
b) Show that the quotient map $x \to x + \mathcal{M}$ has norm 1. (Use Problem 20.18.)
c) Prove, if $\mathcal{X}$ is complete, so is $\mathcal{X}/\mathcal{M}$.

Problem 20.21. A normed vector space $\mathcal{X}$ is called *separable* if it is separable as a metric space (that is, there is a countable subset of $\mathcal{X}$ which is dense in the norm topology). Prove c_0 and ℓ^1 are separable, but ℓ^∞ is not. (Hint: for ℓ^∞, show that there is an uncountable collection of elements $\{f_\alpha\}$ such that $\|f_\alpha - f_\beta\| = 1$ for $\alpha \ne \beta$.)

21. Linear functionals and the Hahn-Banach theorem

If there is a “fundamental theorem of functional analysis,” it is the Hahn-Banach theorem. The theorem is somewhat abstract-looking at first, but its importance will be clear after studying some of its corollaries.

Let $\mathcal{X}$ be a normed vector space over the field $\mathbb{K}$. A *linear functional* on $\mathcal{X}$ is a linear map $L : \mathcal{X} \to \mathbb{K}$. As one might expect, we are especially interested in bounded linear functionals. Since $\mathbb{K} = \mathbb{R}$ or $\mathbb{C}$ is complete, the vector space of bounded linear functionals $B(\mathcal{X}, \mathbb{K})$ is itself a Banach space (complete normed vector space). This space is called the *dual space* of $\mathcal{X}$ and is denoted $\mathcal{X}^*$. It is not yet obvious that $\mathcal{X}^*$ need be non-trivial (that is, that there are any bounded linear functionals on $\mathcal{X}$ besides 0). One corollary of the Hahn-Banach theorem will be that there always exist many linear functionals on any normed space $\mathcal{X}$ (in particular, enough to separate points of $\mathcal{X}$).

21.1. **Examples.** For each of the sequence spaces c_0, ℓ^1, ℓ^∞, for each n the map $f \to f(n)$ is a bounded linear functional. If we fix $g \in \ell^1$, then the functional $L_g : c_0 \to \mathbb{K}$

defined by

$$L_g(f) := \sum_{n=0}^{\infty} f(n)g(n)$$

is bounded, since

$$|L_g(f)| \leq \sum_{n=0}^{\infty} |f(n)g(n)| \leq \|f\|_\infty \sum_{n=0}^{\infty} |g(n)| = \|g\|_1 \|f\|_\infty.$$

This inequality shows that $\|L_g\| \leq \|g\|_1$. In fact, equality holds, and every bounded linear functional on c_0 is of this form:

Proposition 21.1. *The map $\Phi : \ell^1 \to c_0^*$ defined by $\Phi(g) = L_g$ is an isometric isomorphism from ℓ^1 onto the dual space c_0^*.*

Proof. We have already seen that each $g \in \ell^1$ gives rise to a bounded linear functional $L_g \in c_0^*$ via

$$L_g(f) := \sum_{n=0}^{\infty} g(n)f(n)$$

and that $\|L_g\| \leq \|g\|_1$. We will prove simultaneously that this map is onto and that $\|L_g\| \geq \|g\|_1$.

Let $L \in c_0^*$, we will first show that there is unique $g \in \ell^1$ so that $L = L_g$. Let $e_n \in c_0$ be the indicator function of n, that is

$$e_n(m) = \delta_{nm}.$$

Define a function $g : \mathbb{N} \to \mathbb{K}$ by

$$g(n) = L(e_n).$$

We claim that $g \in \ell^1$ and $L = L_g$. To see this, fix an integer N and let $h \in c_{00}$ be the function

$$h(n) = \begin{cases} \overline{g(n)}/|g(n)| & \text{if } n \leq N \text{ and } g(n) \neq 0 \\ 0 & \text{otherwise} \end{cases}.$$

By definition $h \in c_0$ and $\|h\|_\infty \leq 1$. Note that $h = \sum_{n=0}^{N} h(n)e_n$. Now

$$\sum_{n=0}^{N} |g(n)| = \sum_{n=0}^{N} h(n)g(n) = L(h) = |L(h)| \leq \|L\|\|h\| \leq \|L\|.$$

It follows that $g \in \ell^1$ and $\|g\|_1 \leq \|L\|$. Moreover, the same calculation shows that $L = L_g$ when restricted to c_{00}, so by the uniqueness of extensions of bounded operators, $L = L_g$. Thus the map $g \to L_g$ is onto.

Finally, if $h \in \ell^1$, then $L = L_h \in c_0^*$ and hence there is a $g \in \ell^1$ such that $L = L_g$ and $\|L\| \geq \|g\|_1$. Evidently $g = h$ and hence $\|L_h\| \geq \|h\|_1$. Therefore $\|h\|_1 = \|L(h)\|$ for $h \in \ell^1$; i.e., Φ is an isometry (and is in particular one-one). $\square$

Proposition 21.2. *$(\ell^1)^*$ is isometrically isomorphic to ℓ^∞.*

Proof. The proof follows the same lines as the proof of the previous proposition; the details are left as an exercise. ☐

The same mapping $g \to L_g$ also shows that every $g \in \ell^1$ gives a bounded linear functional on ℓ^∞, but it turns out these do not exhaust $(\ell^\infty)^*$ (see Problem 21.11).

If $f \in L^1(m)$ and g is a bounded measurable function with $\sup_{x\in X} |g(x)| = M$, then the map

$$L_g(f) := \int_X fg\, dm$$

is a bounded linear functional of norm at most M. We will prove in Section 24 that the norm is in fact equal to M, and every bounded linear functional on $L^1(m)$ is of this type (at least when m is σ-finite).

If X is a compact metric space and μ is a finite, signed Borel measure on X, then

$$L_\mu(f) := \int_X f\, d\mu$$

is a bounded linear functional on $C_{\mathbb{R}}(X)$ with norm $\|\mu\| = |\mu|(X)$ (see Problem 21.8). A version of the Riesz Markov Theorem says the converse is true too.

Theorem 21.3 (Riesz-Markov)**.** *Suppose X is a compact Hausdorff space. If $\lambda \in C(X)^*$, then there exists a unique regular Borel measure σ such that, for $f \in C(X)$,*

$$\lambda(f) = \int_X f\, d\sigma.$$

The result is true with both real and complex scalars. Focusing on the case of real scalars, the strategy is to write λ as the difference of positive linear functionals on $C(X)$ and apply the Riesz-Markov Theorem (twice). (For complex scalars, write λ in terms of its real and imaginary parts and apply the result in the real case (twice)).

A function $f \in C(X)$ is *positive* (really nonnegative) if $f(x) \geq 0$ for all $x \in X$, written $f \geq 0$. Let $C(X)^+$ denote the positive elements of $C(X)$. Given linear functionals $\lambda, \rho \in C(X)^*$, the inequality $\lambda \leq \rho$ means that $\lambda(f) \leq \rho(f)$ for all $f \in C(X)^+$.

Lemma 21.4. *For $\lambda \in C(X)^*$ there exists a positive Borel measure μ on X such that $L_\mu \geq \lambda$.*

From the lemma, both L_μ and $L_\mu - \lambda$ are positive linear functionals. Hence there exists a positive regular Borel measure ν such that $L_\mu - \lambda = L_\nu$ and consequently

$$\lambda = L_\mu - L_\nu = L_{\mu-\nu}.$$

Proof of the lemma. For $f \in C(X)^+$, define

$$\Lambda(f) = \sup\{\lambda(u) : 0 \leq u \leq f\}.$$

The sup is finite and at most $\|\lambda\|\,\|f\|_\infty$ since, $|\lambda(u)| \leq \|\lambda\|\,\|u\|_\infty \leq \|\lambda\|\,\|f\|_\infty$. In particular, $|\Lambda(1)\| \leq \|\lambda\|$.

Note that $\Lambda(cf) = c\Lambda(f)$ for $c \geq 0$ (and of course $f \in C(X)^+$). Now suppose $f, g \in C(X)^+$. It is evident that

$$\Lambda(f+g) \geq \Lambda(f) + \Lambda(g).$$

The proof of the reverse inequality takes advantage of the *lattice structure* of $C(X)$. Namely, if $u, v \in C(X)$, then so is $\min\{u, v\}$ (defined pointwise). Suppose $0 \leq u \leq f+g$. Let $h = \min\{u, f\}$ and note that $u \leq h+g$ and therefore, $0 \leq u-h \leq g$. It follows that

$$\lambda(u) - \lambda(h) = \lambda(u-h) \leq \Lambda(g).$$

Likewise, $0 \leq h \leq f$ and thus,

$$\lambda(h) \leq \Lambda(f).$$

Adding the last two inequalities gives,

$$\lambda(u) = \lambda(h) + \lambda(u-h) \leq \Lambda(f) + \Lambda(g)$$

and the inequality $\Lambda(f+g) \leq \Lambda(f) + \Lambda(g)$ follows.

Now suppose $f_\pm, g_\pm \in C(X)^+$ and $f_+ - f_- = g_+ - g_-$. In this case $f_+ + g_- = g_+ + f_-$ and thus

$$\Lambda(f_+) + \Lambda(g_-) = \Lambda(f_+ + g_-) = \Lambda(g_+ + f_-) = \Lambda(g_+) + \Lambda(f_-).$$

Rearranging shows that Λ extends (in a well defined fashion) to $C(X)$ by writing $f \in C(X)$ as the difference (in any way) of two positive functions.

It is now an exercise to verify that Λ (now extended to $C(X)$) is in fact linear. Thus Λ a finite positive linear functional on $C(X)$ and therefore has a unique representation as a regular positive Borel measure on X. □

21.2. **The Hahn-Banach Extension Theorem.** To state and prove the Hahn-Banach Extension Theorem, we first work in the setting $\mathbb{K} = \mathbb{R}$, then extend our results to the complex case.

Definition 21.5. Let $\mathcal{X}$ be a real vector space. A *Minkowski functional* is a function $p : \mathcal{X} \to \mathbb{R}$ such that $p(x+y) \leq p(x) + p(y)$ and $p(\lambda x) = \lambda p(x)$ for all $x, y \in \mathcal{X}$ and nonnegative $\lambda \in \mathbb{R}$. ◁

For example, if $L : \mathcal{X} \to \mathbb{R}$ is any linear functional, then the function $p(x) := |L(x)|$ is a Minkowski functional. Also $p(x) = \|x\|$ is a Minkowski functional. More generally, $p(x) = \|x\|$ is a Minkowski functional whenever $\|x\|$ is a *seminorm* on $\mathcal{X}$. (A seminorm is a function obeying all the requirements of a norm, except we allow $\|x\| = 0$ for nonzero x. Note that both of the given examples of Minkowski functionals come from seminorms.)

Theorem 21.6 (The Hahn-Banach Theorem, real version)**.** *Let $\mathcal{X}$ be a vector space over $\mathbb{R}$, p a Minkowski functional on $\mathcal{X}$, and $\mathcal{M}$ a subspace of $\mathcal{X}$. If L a linear functional on $\mathcal{M}$ such that $L(x) \leq p(x)$ for all $x \in \mathcal{M}$, then there exists a linear functional L' on $\mathcal{X}$ such that*

(i) $L'|_{\mathcal{M}} = L$ (L' extends L)
(ii) $L'(x) \leq p(x)$ for all $x \in \mathcal{X}$ (L' is dominated by p).

Proof. The idea is to show that the extension can be done one dimension at a time and then infer the existence of an extension to the whole space by appeal to Zorn's lemma. We may of course assume $\mathcal{M} \neq \mathcal{X}$. So, fix a vector $x \in \mathcal{X}\setminus\mathcal{M}$ and consider the subspace $\mathcal{M} + \mathbb{R}x \subset \mathcal{X}$. For any $m_1, m_2 \in \mathcal{M}$, by hypothesis,

$$L(m_1) + L(m_2) = L(m_1 + m_2) \leq p(m_1 + m_2) \leq p(m_1 - x) + p(m_2 + x).$$

Rearranging gives, for $m_1, m_2 \in \mathcal{M}$,

$$L(m_1) - p(m_1 - x) \leq p(m_2 + x) - L(m_2)$$

and thus

$$\sup_{m\in\mathcal{M}} \{L(m) - p(m - x)\} \leq \inf_{m\in\mathcal{M}} \{p(m + x) - L(m)\}.$$

Now choose any real number λ satisfying

$$\sup_{m\in\mathcal{M}} \{L(m) - p(m - x)\} \leq \lambda \leq \inf_{m\in\mathcal{M}} \{p(m + x) - L(m)\}.$$

In particular, for $m \in \mathcal{M}$,

$$\begin{aligned} L(m) - \lambda &\leq p(m - x) \\ L(m) + \lambda &\leq p(m + x). \end{aligned} \tag{4}$$

Let $\mathcal{N} = \mathcal{M} + \mathbb{R}x$ and define $L' : \mathcal{N} \to \mathbb{R}$ by $L'(m + tx) = L(m) + t\lambda$ for $m \in \mathcal{M}$ and $t \in \mathbb{R}$. Thus L' is linear and agrees with L on $\mathcal{M}$ by definition. We now check that $L'(y) \leq p(y)$ for all $y \in \mathcal{M} + \mathbb{R}x$. Accordingly, suppose $m \in \mathcal{M}$, $t \in \mathbb{R}$ and let $y = m + tx$. If $t = 0$ there is nothing to prove. If $t > 0$, then, in view of equation (4),

$$L'(y) = L'(m + tx) = t\left(L(\frac{m}{t}) + \lambda\right) \leq t\,p(\frac{m}{t} + x) = p(m + tx) = p(y)$$

and a similar estimate shows that $L'(m + tx) \leq p(m + tx)$ for $t < 0$.

We have thus successfully extended L to $\mathcal{M} + \mathbb{R}x$. To finish, let $\mathcal{L}$ denote the set of pairs $(L', \mathcal{N})$ where $\mathcal{N}$ is a subspace of $\mathcal{X}$ containing $\mathcal{M}$, and L' is an extension of L to $\mathcal{N}$ obeying $L'(y) \leq p(y)$ on $\mathcal{N}$. Declare $(L'_1, \mathcal{N}_1) \preceq (L'_2, \mathcal{N}_2)$ if $\mathcal{N}_1 \subset \mathcal{N}_2$ and $L'_2|_{\mathcal{N}_1} = L'_1$. This relation $\preceq$ is a partial order on $\mathcal{L}$. An exercise shows, given any increasing chain $(L'_\alpha, \mathcal{N}_\alpha)$ in $\mathcal{L}$, it has as an upper bound $(L', \mathcal{N})$ in $\mathcal{L}$, where $\mathcal{N} := \bigcup_\alpha \mathcal{N}_\alpha$ and $L(n_\alpha) := L'_\alpha(n_\alpha)$ for $n_\alpha \in \mathcal{N}_\alpha$. By Zorn's lemma the collection $\mathcal{L}$ has a maximal element $(L', \mathcal{N})$ with respect to the order $\preceq$. Since it always possible to extend to a strictly larger subspace, the maximal element must have $\mathcal{N} = \mathcal{X}$, and the proof is finished. □

The proof is a typical application of Zorn's lemma - one knows how to carry out a construction one step a time, but there is no clear way to do it all at once.

In the special case that p is a seminorm, since $L(-x) = -L(x)$ and $p(-x) = p(x)$ the inequality $L \leq p$ is equivalent to $|L| \leq p$.

Corollary 21.7. *Suppose $\mathcal{X}$ is a normed vector space over $\mathbb{R}$, $\mathcal{M}$ is a subspace, and L is a bounded linear functional on $\mathcal{M}$. If $C \geq 0$ and $|L(x)| \leq C\|x\|$ for all $x \in \mathcal{M}$, then there exists a bounded linear functional L' on $\mathcal{X}$ extending L such that $\|L'\| \leq C$.*

Proof. Apply the Hahn-Banach theorem with the Minkowski functional $p(x) = C\|x\|$. □

Before obtaining further corollaries, we extend these results to the complex case. First, if $\mathcal{X}$ is a vector space over $\mathbb{C}$ then trivially it is also a vector space over $\mathbb{R}$, and there is a simple relationship between the $\mathbb{R}$- and $\mathbb{C}$-linear functionals.

Proposition 21.8. *Let $\mathcal{X}$ be a vector space over $\mathbb{C}$. If $L : \mathcal{X} \to \mathbb{C}$ is a $\mathbb{C}$-linear functional, then $u(x) = ReL(x)$ defines an $\mathbb{R}$-linear functional on $\mathcal{X}$ and $L(x) = u(x) - iu(ix)$. Conversely, if $u : \mathcal{X} \to \mathbb{R}$ is $\mathbb{R}$-linear then $L(x) := u(x) - iu(ix)$ is $\mathbb{C}$-linear. If in addition $p : \mathcal{X} \to \mathbb{R}$ is a seminorm, then $|u(x)| \le p(x)$ for all $x \in \mathcal{X}$ if and only if $|L(x)| \le p(x)$ for all $x \in \mathcal{X}$.*

Proof. Problem 21.5. □

Theorem 21.9 (The Hahn-Banach Theorem, complex version)**.** *Let $\mathcal{X}$ be a vector space over $\mathbb{C}$, p a seminorm on $\mathcal{X}$, and $\mathcal{M}$ a subspace of $\mathcal{X}$. If $L : \mathcal{M} \to \mathbb{C}$ is a $\mathbb{C}$-linear functional satisfying $|L(x)| \le p(x)$ for all $x \in \mathcal{M}$, then there exists a $\mathbb{C}$-linear functional $L' : \mathcal{X} \to \mathbb{C}$ such that*

(i) $L'|_{\mathcal{M}} = L$ *and*
(ii) $|L'(x)| \le p(x)$ *for all* $x \in \mathcal{X}$.

Proof. The proof consists of applying the real Hahn-Banach theorem to extend the $\mathbb{R}$-linear functional $u = \mathrm{Re}L$ to a functional $u' : \mathcal{X} \to \mathbb{R}$ and then defining L' from u' as in Proposition 21.8. The details are left as an exercise. □

The following corollaries are quite important, and when the Hahn-Banach theorem is applied it is usually in one of the following forms:

Corollary 21.10. *Let $\mathcal{X}$ be a normed vector space.*

(i) *If $\mathcal{M} \subset \mathcal{X}$ is a subspace and $L : \mathscr{M} \to \mathbb{K}$ is a bounded linear functional, then there exists a bounded linear functional $L' : \mathcal{X} \to \mathbb{K}$ such that $L'|_{\mathscr{M}} = L$ and $\|L'\| = \|L\|$.*
(ii) (Linear functionals detect norms) *If $x \in \mathcal{X}$ is nonzero, there exists $L \in \mathcal{X}^*$ with $\|L\| = 1$ such that $L(x) = \|x\|$.*
(iii) (Linear functionals separate points) *If $x \neq y$ in $\mathcal{X}$, there exists $L \in \mathcal{X}^*$ such that $L(x) \neq L(y)$.*
(iv) (Linear functionals detect distance to subspaces) *If $\mathcal{M} \subset \mathcal{X}$ is a closed subspace and $x \in \mathcal{X} \setminus \mathcal{M}$, there exists $L \in \mathcal{X}^*$ such that*
  (a) $L|_{\mathcal{M}} = 0$;
  (b) $\|L\| = 1$; *and*
  (c) $L(x) = \mathit{dist}(x, \mathcal{M}) = \inf_{y \in \mathcal{M}} \|x - y\| > 0$.

Proof. (i): Consider the (semi)norm $p(x) = \|L\|\,\|x\|$. By construction, $|L(x)| \le p(x)$ for $x \in \mathcal{M}$. Hence, there is a linear functional L' on $\mathcal{X}$ such that $L'|_{\mathcal{M}} = L$ and $|L'(x)| \le p(x)$ for all $x \in \mathcal{X}$. In particular, $\|L'\| \le \|L\|$. On the other hand, $\|L'\| \ge \|L\|$ since L' agrees with L on $\mathcal{M}$.

(ii): Let $\mathcal{M}$ be the one-dimensional subspace of $\mathcal{X}$ spanned by x. Define a functional $L : \mathcal{M} \to \mathbb{K}$ by $L(t\frac{x}{\|x\|}) = t$. In particular, $|L(y)| = \|y\|$ for $y \in \mathcal{M}$ and thus $\|L\| = 1$. By (i), the functional L extends to a functional (still denoted L) on $\mathcal{X}$ such that $\|L\| = 1$.

(iii): Apply (ii) to the vector $x - y$.

(iv): Let $\delta = \operatorname{dist}(x, \mathcal{M})$. Since $\mathcal{M}$ is closed, $\delta > 0$. Define a functional $L : \mathcal{M} + \mathbb{K}x \to \mathbb{K}$ by $L(y + tx) = t\delta$. Since for $t \neq 0$

$$\|y + tx\| = |t|\|t^{-1}y + x\| \geq |t|\delta = |L(y + tx)|,$$

by Hahn-Banach we can extend L to a functional $L \in \mathcal{X}^*$ with $\|L\| \leq 1$. □

Needless to say, the proof of the Hahn-Banach theorem is thoroughly non-constructive, and in general it is an important (and often difficult) problem, given a normed space $\mathcal{X}$, to find some concrete description of the dual space $\mathcal{X}^*$. Usually this means finding a Banach space $\mathcal{Y}$ and a bounded (or, better, isometric) isomorphism $T : \mathcal{Y} \to \mathcal{X}^*$.

Note that since $\mathcal{X}^*$ is a normed space, we can form its dual, denoted $\mathcal{X}^{**}$, and called the *bidual* of $\mathcal{X}$. There is a canonical relationship between $\mathcal{X}$ and $\mathcal{X}^{**}$. Each fixed $x \in \mathcal{X}$ gives rise to a linear functional $\hat{x} : \mathcal{X}^* \to \mathbb{K}$ via evaluation,

$$\hat{x}(L) := L(x).$$

Since $|\hat{x}(L)| = |L(x)| \leq \|L\| \, \|x\|$, the linear functional $\hat{x} \in \mathcal{X}^{**}$ and $\|\hat{x}\| \leq \|x\|$.

Corollary 21.11. *(Embedding in the bidual) The map $x \to \hat{x}$ is an isometric linear map from $\mathcal{X}$ into $\mathcal{X}^{**}$.*

Proof. First, from the definition we see that

$$|\hat{x}(L)| = |L(x)| \leq \|L\|\|x\|$$

so $\hat{x} \in \mathcal{X}^{**}$ and $\|\hat{x}\| \leq \|x\|$. It is straightforward to check (recalling that the L's are linear) that the map $x \to \hat{x}$ is linear. Finally, to show that $\|\hat{x}\| = \|x\|$, fix a nonzero $x \in \mathcal{X}$. From Corollary 21.10(i) there exists $L \in \mathcal{X}^*$ with $\|L\| = 1$ and $L(x) = \|x\|$. But then for this x and L, we have $|\hat{x}(L)| = |L(x)| = \|x\|$ so $\|\hat{x}\| \geq \|x\|$, and the proof is complete. □

Definition 21.12. A Banach space $\mathcal{X}$ is called *reflexive* if the map $\hat{\ } : \mathcal{X} \to \mathcal{X}^{**}$ is surjective. ◁

In other words, $\mathcal{X}$ is reflexive if the map $\hat{\ }$ is an (isometric) isomorphism of $\mathcal{X}$ with $\mathcal{X}^{**}$. For example, every finite dimensional Banach space is reflexive (Problem 21.6). Reflexive spaces often have nice properties. For instance, the distance from a point to a (closed) subspace is attained. On the other hand, by Propositions 21.1 and 21.2, c_0^{**} is isometrically isomorphic to ℓ^∞. In Problem 21.7 you will show that c_0 is not isometrically isomorphic to ℓ^∞ and so c_0 is not reflexive. After we have studied the L^p and ℓ^p spaces in more detail, we will see that L^p is reflexive for $1 < p < \infty$.

The embedding into the bidual has many applications; one of the most basic is the following.

Proposition 21.13 (Completion of normed spaces)**.** *If $\mathcal{X}$ is a normed vector space, then there is a Banach space $\overline{\mathcal{X}}$ and in isometric map $\iota : \mathcal{X} \to \overline{\mathcal{X}}$ such that the image $\iota(\mathcal{X})$ is dense in $\overline{\mathcal{X}}$.*

Proof. Embed $\mathcal{X}$ into $\mathcal{X}^{**}$ via the map $x \to \hat{x}$ and let $\overline{\mathcal{X}}$ be the closure of the image of $\mathcal{X}$ in $\mathcal{X}^{**}$. Since $\overline{\mathcal{X}}$ is a closed subspace of a complete space, it is complete. □

The space $\overline{\mathcal{X}}$ is called the *completion* of $\mathcal{X}$. It is unique in the sense that if $\mathcal{Y}$ is another Banach space and $j : \mathcal{X} \to \mathcal{Y}$ embeds $\mathcal{X}$ isometrically as a dense subspace of $\mathcal{Y}$, then $\mathcal{Y}$ is isometrically isomorphic to $\overline{\mathcal{X}}$. The proof of this fact is left as an exercise.

21.3. **Dual spaces and adjoint operators.** Let $\mathcal{X}, \mathcal{Y}$ be normed spaces with duals $\mathcal{X}^*, \mathcal{Y}^*$. If $T : \mathcal{X} \to \mathcal{Y}$ is a linear transformation and $f : \mathcal{Y} \to \mathbb{K}$ is a linear functional, then $T^*f : \mathcal{X} \to \mathbb{K}$ defined by

$$(T^*f)(x) = f(Tx) \tag{5}$$

is a linear functional on $\mathcal{X}$. If T and f are both continuous (that is, bounded) then the composition T^*f is bounded, and more is true:

Theorem 21.14. *Let $T : \mathcal{X} \to \mathcal{Y}$ be a bounded linear transformation. For $f \in \mathcal{Y}^*$, define T^*f by the formula (5). Then:*

i) *T^*f belongs to $\mathcal{X}^*$, and T^* is a linear map from $\mathcal{Y}^*$ into $\mathcal{X}^*$.*
ii) *$T^* : \mathcal{Y}^* \to \mathcal{X}^*$ is bounded and $\|T^*\| = \|T\|$.*

Proof. Since T is assumed bounded, for a fixed $f \in \mathcal{Y}^*$ and all $x \in \mathcal{X}$

$$|T^*f(x)| = |f(Tx)| \le \|f\|\|Tx\| \le \|f\|\|T\|\|x\|.$$

It follows that T^*f is bounded on $\mathcal{X}$ (thus, belongs to $\mathcal{X}^*$) and $\|T^*f\| \le \|f\|\|T\|$. Thus T^* maps $\mathcal{Y}^*$ into $\mathcal{X}^*$ and it is straightforward to verify that T^* is linear, which proves (i). Moreover, this inequality also shows that T^* is bounded and $\|T^*\| \le \|T\|$.

It remains to show $\|T^*\| \ge \|T\|$. Toward this end, let $0 < \epsilon < 1$ be given and choose $x \in \mathcal{X}$ with $\|x\| = 1$ and $\|Tx\| > (1-\epsilon)\|T\|$. Now consider Tx. By the Hahn-Banach theorem, there exists $f \in \mathcal{Y}^*$ such that $\|f\| = 1$ and $f(Tx) = \|Tx\|$. For this f,

$$\|T^*\| \ge \|T^*f\| \ge |T^*f(x)| = |f(Tx)| = \|Tx\| > (1-\epsilon)\|T\|.$$

Hence, $\|T^*\| \ge (1-\epsilon)\|T\|$. Since ϵ was arbitrary, $\|T^*\| \ge \|T\|$. □

21.4. **Duality for Sub and Quotient Spaces.** The Hahn-Banach Theorem allows for the identification of the duals of subspaces and quotients of Banach spaces. Informally, the dual of a subspace is a quotient and the dual of a quotient is a subspace. The precise results are stated below for complex scalars, but they hold also for real scalars.

Given a (closed) subspace $\mathcal{M}$ of the Banach space $\mathcal{X}$, let π denote the map from $\mathcal{X}$ to the quotient $\mathcal{X}/\mathcal{M}$. Recall (see Problem 20.20), the quotient is a Banach space with the norm,

$$\|z\| = \inf\{\|y\| : \pi(y) = z\}.$$

In particular, if $x \in \mathcal{X}$, then

$$\|\pi(x)\| = \inf\{\|x - m\| : m \in \mathcal{M}\}.$$

It is evident from the construction that π is continuous and $\|\pi\| \le 1$. Further, by Problem 20.18 (or see Proposition 21.15 below) if $\mathcal{M}$ is a proper (closed) subspace, then $\|\pi\| = 1$. In particular, $\pi^* : (\mathcal{X}/\mathcal{M})^* \to \mathcal{X}^*$ defined by $\pi^*\lambda = \lambda \circ \pi$ is also continuous. Moreover, if $x \in \mathcal{M}$, then

$$\pi^*\lambda(x) = 0.$$

Let

$$\mathcal{M}^\perp = \{f \in \mathcal{X}^* : f(x) = 0 \text{ for all } x \in \mathcal{M}\}.$$

($\mathcal{M}^\perp$ is called the *annihilator* of $\mathcal{M}$ in $\mathcal{X}^*$.) Recall, given $x \in \mathcal{X}$, the element $\hat{x} \in \mathcal{X}^{**}$ is defined by $\hat{x}(\tau) = \tau(x)$. In particular,

$$\mathcal{M}^\perp = \cap_{x \in \mathcal{M}} \ker(\hat{x})$$

and thus $\mathcal{M}^\perp$ is a closed subspace of $\mathcal{X}^*$. Further, if $\lambda \in (\mathcal{X}/\mathcal{M})^*$, then $\pi^*\lambda \in \mathcal{M}^\perp$.

Proposition 21.15 (The dual of a quotient)**.** *The mapping $\psi : (\mathcal{X}/\mathcal{M})^* \to \mathcal{M}^\perp$ defined by*

$$\psi(\lambda) = \pi^*\lambda$$

is an isometric isomorphism; i.e., the mapping $\pi^* : (\mathcal{X}/\mathcal{M})^* \to \mathcal{X}^*$ is an isometric isomorphism onto $\mathcal{M}^\perp$.

Informally, the proposition is expressed as $(\mathcal{X}/\mathcal{M})^* = \mathcal{M}^\perp$.

Proof. The linearity of ψ follows from Theorem 21.14 as does $\|\psi\| = \|\pi\| \le 1$. To prove that ψ is isometric, let $\lambda \in (\mathcal{X}/\mathcal{M})^*$ be given. Automatically, $\|\psi(\lambda)\| \le \|\lambda\|$. To prove the reverse inequality, fix $r > 1$. Let $q \in \mathcal{X}/\mathcal{M}$ with $\|q\| = 1$ be given. There exists an $x \in X$ such that $\|x\| < r$ and $\pi(x) = q$. Hence,

$$|\lambda(q)| = |\lambda(\pi(x))\| = \|\psi(\lambda)(x)\| \le \|\psi(\lambda)\|\,\|x\| < r\|\psi(\lambda)\|.$$

Taking the supremum over such q shows $\|\lambda\| \le r\|\psi(\lambda)\|$. Finally, since $1 < r$ is arbitrary, $\|\lambda\| \le \|\psi(\lambda)\|$.

To prove that ψ is onto, and complete the proof, let $\tau \in \mathcal{M}^\perp$ be given. Fix $q \in \mathcal{X}/\mathcal{M}$. If $x, y \in \mathcal{X}$ and $\pi(x) = q = \pi(y)$, then $\tau(x) = \tau(y)$. Hence, the mapping $\lambda : \mathcal{X}/\mathcal{M} \to \mathbb{C}$ defined by $\lambda(q) = \tau(x)$ is well defined. That λ is linear is left as an exercise. To see that λ is continuous, observe that

$$|\lambda(q)| = |\tau(x)| \le \|\tau\|\,\|x\|,$$

for each $x \in \mathcal{X}$ such that $\pi(x) = q$. Taking the infimum over such x gives shows

$$|\lambda(q)| \le \|\tau\|\,\|q\|.$$

□

Since $\mathcal{M}^\perp$ is closed in $\mathcal{X}^*$, the quotient space $\mathcal{X}^*/\mathcal{M}^\perp$ is a Banach space. Let $\rho : \mathcal{X}^* \to \mathcal{X}^*/\mathcal{M}^\perp$ denote the quotient mapping. Suppose $\lambda \in \mathcal{M}^*$. If f and g are two extensions of λ to bounded linear functionals on $\mathcal{X}^*$, then $f(x) - g(x) = 0$ for $x \in \mathcal{M}$. Hence $f - g \in \mathcal{M}^\perp$ or equivalently, $\rho(f) = \rho(g)$. Consequently, the mapping $\varphi : \mathcal{M}^* \to \mathcal{X}^*/\mathcal{M}^\perp$ defined by $\varphi(\lambda) = \rho(f)$ (where f is any bounded extension of λ to $\mathcal{X}$) is well defined. It is easily verified that φ is linear. Moreover, $\varphi(f|_\mathcal{M}) = \rho(f)$ for $f \in \mathcal{X}^*$ and therefore φ is onto.

Proposition 21.16 (The dual of a subspace)**.** *The mapping $\varphi : \mathcal{M}^* \to \mathcal{X}^*/\mathcal{M}^\perp$ is an isometric isomorphism.*

Proof. It remains to show that φ is an isometry, a fact that is an easy consequence of the Hahn-Banach Theorem. Fix $\lambda \in \mathcal{M}^*$ and let $q = \varphi(\lambda)$. If f is any bounded extension of λ to $\mathcal{X}^*$, then $\|f\| \geq \|\lambda\|$. Hence,

$$\begin{aligned}\|\varphi(\lambda)\| =& \|q\| \\ =& \inf\{\|f\| : f \in \mathcal{X}^*, \quad \rho(f) = q\} \\ =& \inf\{\|f\| : f \in \mathcal{X}^*, \quad f|_\mathcal{M} = \lambda\} \\ \geq& \|\lambda\|.\end{aligned}$$

On the other hand, by the Hahn-Banach Theorem there is a bounded extension g with $\|g\| = \|\lambda\|$. Thus $\|\lambda\| \leq \|q\|$. □

A special case of the following useful fact was used in the proofs above. If $\mathcal{X}, \mathcal{Y}$ are vector spaces and $T : \mathcal{X} \to \mathcal{Y}$ is linear and $\mathcal{M}$ is a subspace of the kernel of T, then T induces a linear map $\tilde{T} : \mathcal{X}/\mathcal{M} \to \mathcal{Y}$. A canonical choice is $\mathcal{M} = \ker(T)$ in which case $\tilde{T}$ is one-one. If $\mathcal{X}$ is a Banach space, $\mathcal{Y}$ is a normed vector space and $\mathcal{M}$ is closed, then $\mathcal{X}/\mathcal{M}$ is a Banach space.

Lemma 21.17. *If $\mathcal{X}$ is a Banach space, $\mathcal{M}$ is a (closed) subspace, $\mathcal{Y}$ is a normed vector space and $T : \mathcal{X} \to \mathcal{Y}$ is continuous, then the mapping $\tilde{T}$ is bounded and $\|\tilde{T}\| = \|T\|$.*

Proof. Let $\pi : \mathcal{X} \to \mathcal{X}/\mathcal{M}$ denote the quotient map and observe that $\tilde{T}\pi = T$. Fix $q \in \mathcal{X}/\mathcal{M}$. For any $x \in \mathcal{X}$ such that $\pi(x) = q$,

$$\|\tilde{T}q\| = \|\tilde{T}(\pi(x))\| = \|Tx\| \leq \|T\|\,\|x\|.$$

Taking the infimum on x such that $\pi(x) = q$ gives,

$$\|\tilde{T}q\| \leq \|T\|\,\|q\|.$$

□

Remark 21.18. ◇

21.5. **Problems.**

Problem 21.1. Prove, if $\mathcal{X}$ is any normed vector space, $\{x_1, \dots x_n\}$ is a linearly independent set in $\mathcal{X}$, and $\alpha_1, \dots \alpha_n$ are scalars, then there exists a bounded linear functional f on $\mathcal{X}$ such that $f(x_j) = \alpha_j$ for $j = 1, \dots n$. (Recall linear maps from a finite dimensional normed vector space to a normed vector space are bounded.)

Problem 21.2. Let $\mathcal{X}, \mathcal{Y}$ be normed spaces and $T : \mathcal{X} \to \mathcal{Y}$ a linear transformation. Prove T is bounded if and only if there exists a constant C such that for all $x \in \mathcal{X}$ and $f \in \mathcal{Y}^*$,

$$|f(Tx)| \le C\|f\|\|x\|; \tag{6}$$

in which case $\|T\|$ is equal to the best possible C in (6).

Problem 21.3. Let $\mathcal{X}$ be a normed vector space. Show that if $\mathcal{M}$ is a closed subspace of $\mathcal{X}$ and $x \notin \mathcal{M}$, then $\mathcal{M} + \mathbb{K}x$ is closed. Use this result to give another proof that every finite-dimensional subspace of $\mathcal{X}$ is closed.

Problem 21.4. Prove, if $\mathcal{M}$ is a *finite-dimensional* subspace of a Banach space $\mathcal{X}$, then there exists a closed subspace $\mathcal{N} \subset \mathcal{X}$ such that $\mathcal{M} \cap \mathcal{N} = \{0\}$ and $\mathcal{M} + \mathcal{N} = \mathcal{X}$. (In other words, every $x \in \mathcal{X}$ can be written uniquely as $x = y + z$ with $y \in \mathcal{M}$, $z \in \mathcal{N}$.) *Hint:* Choose a basis $x_1, \dots x_n$ for $\mathcal{M}$ and construct bounded linear functionals $f_1, \dots f_n$ on $\mathcal{X}$ such that $f_i(x_j) = \delta_{ij}$. Now let $\mathcal{N} = \cap_{i=1}^n \ker f_i$. (Warning: this conclusion can fail badly if $\mathcal{M}$ is not assumed finite dimensional, even if $\mathcal{M}$ is still assumed closed. Perhaps the first known example is that c_0 is not *complemented* in ℓ^∞, though it is nontrivial to prove.)

Problem 21.5. Prove Proposition 21.8.

Problem 21.6. Prove every finite-dimensional Banach space is reflexive.

Problem 21.7. Let B denote the subset of ℓ^∞ consisting of sequences which take values in $\{-1, 1\}$. Show that any two (distinct) points of B are a distance 2 apart. Show, if C is a countable subset of ℓ^∞, then there exists a $b \in B$ such that $\|b - c\| \ge 1$ for all $c \in C$. Conclude ℓ^∞ is not separable. Prove there is no isometric isomorphism $\Lambda : c_0 \to \ell^\infty$. As a corollary, conclude that c_0 is not reflexive. (Of course, saying $c_0 \ne \ell^\infty$ via the canonical embedding is much weaker than saying there is no isometric isomorphism between c_0 and ℓ^∞.)

Problem 21.8. Prove, if μ is a finite regular (signed) Borel measure on a compact Hausdorff space, then the linear function $L_\mu : C(X) \to \mathbb{R}$ defined by

$$L_\mu(f) = \int_X f \, d\mu$$

is bounded (continuous) and $\|L_\mu\| = \|\mu\| := |\mu|(X)$. (See the Riesz-Markov Theorem for positive linear functionals.)

Problem 21.9. Let $\mathcal{X}$ and $\mathcal{Y}$ be normed vector spaces and $T \in L(\mathcal{X}, \mathcal{Y})$.

a) Consider $T^{**} : \mathcal{X}^{**} \to \mathcal{Y}^{**}$. Identifying $\mathcal{X}, \mathcal{Y}$ with their images in $\mathcal{X}^{**}$ and $\mathcal{Y}^{**}$, show that $T^{**}|_{\mathcal{X}} = T$.

b) Prove T^* is injective if and only if the range of T is dense in $\mathcal{Y}$.
c) Prove that if the range of T^* is dense in $\mathcal{X}^*$, then T is injective; if $\mathcal{X}$ is reflexive then the converse is true.

Problem 21.10. Prove, if $\mathcal{X}$ is a Banach space and $\mathcal{X}^*$ is separable, then $\mathcal{X}$ is separable. [*Hint:* let $\{f_n\}$ be a countable dense subset of $\mathcal{X}^*$. For each n choose x_n such that $\|x_n\| = 1$ and $|f_n(x_n)| \geq \frac{1}{2}\|f_n\|$. Show that the set of $\mathbb{Q}$-linear combinations of $\{x_n\}$ is dense in $\mathcal{X}$.]

Problem 21.11. a) Prove there exists a bounded linear functional $L \in (\ell^\infty)^*$ with the following property: whenever $f \in \ell^\infty$ and $\lim_{n\to\infty} f(n)$ exists, then $L(f)$ is equal to this limit. (Hint: first show that the set of such f forms a subspace $\mathcal{M} \subset \ell^\infty$).
b) Show that such a functional L is not equal to L_g for any $g \in \ell^1$; thus the map $T : \ell^1 \to (\ell^\infty)^*$ given by $T(g) = L_g$ is not surjective.
c) Give another proof that T is not surjective, using Problem 21.10.

22. The Baire Category Theorem and applications

Recall, a set D in a metric space X is *dense* if $\overline{D} = X$. Thus D is dense if and only if D^c does not contain a nonempty open set if and only if it has nontrivial intersection with every nonempty open set. A topological space X is called a *Baire space* if it has the following property: if $(U_n)_{n=1}^\infty$ is a countable sequence of open dense subsets of X, then the intersection $\cap_{n=1}^\infty U_n$ is dense in X.

Theorem 22.1 (The Baire Category Theorem)**.** *Every complete metric space X is a Baire space. In other words, if X is a complete metric space and if $(U_n)_{n=1}^\infty$ is a sequence of open dense subsets of X, then $\cap_{n=1}^\infty U_n$ is dense in X.*

Theorem 22.1 is true if X is a locally compact Hausdorff space and there are connections between the Baire Category Theorem and the axiom of choice.

A subset $E \subset X$ is *nowhere dense* if its closure has empty interior. Equivalently, $\overline{E}^c$ is open and dense. A set F in a metric space X is *first category* (or *meager*) if it can be expressed as the countable union of nowhere dense sets. In particular, a countable union of first category sets is first category. A set G is *second category* if it is not first category. For applications, the following corollary often suffices.

Corollary 22.2. *If X is a complete metric space, then X is not a countable union of nowhere dense sets; i.e., X is of second category in itself.*

Proof. Take complements and apply Theorem 22.1. ☐

Thus, the Baire property is used as a kind of pigeonhole principle: the "thick" Baire space X cannot be expressed as a countable union of the "thin" nowhere dense sets E_n. Equivalently, if X is Baire and $X = \bigcup_n E_n$, then at least one of the E_n is somewhere dense.

The following lemma should be familiar from advanced calculus.

Lemma 22.3. *Let X be a complete metric space and suppose (C_n) is a sequence of subsets of X. If*

(i) each C_n is nonempty;
(ii) (C_n) is nested decreasing;
(iii) each C_n is closed; and
(iv) $(diam(C_n))$ converges to 0,

then then there is an $x \in X$ such that

$$\{x\} = \cap C_n.$$

Moreover, if $x_n \in C_n$, then (x_n) converges to some x.

Proof. Let $(U_n)_{n=1}^{\infty}$ be a sequence of open dense sets in X and let $I = \cap U_n$. To prove I is dense, it suffices to show that I has nontrivial intersection with every nonempty open set W. Fix such a W. Since U_1 is dense, there is a point $x_1 \in W \cap U_1$. Since U_1 and W are open, there is a radius $0 < r_1 < 1$ such that the $\overline{B(x_1, r_1)}$ is contained in $W \cap U_1$. Similarly, since U_2 is dense and open there is a point $x_2 \in B(x_1, r_1) \cap U_2$ and a radius $0 < r_2 < \frac{1}{2}$ such that

$$\overline{B(x_2, r_2)} \subset B(x_1, r_1) \cap U_2 \subset W \cap U_1 \cap U_2.$$

Continuing inductively, since each U_n is dense and open there is a sequence of points $(x_n)_{n=1}^{\infty}$ and radii $0 < r_n < \frac{1}{n}$ such that

$$\overline{B(x_n, r_n)} \subset B(x_{n-1}, r_{n-1}) \cap U_n \subset W \cap (\cap_{j=1}^{n} U_n).$$

The sequence of sets $(\overline{B(x_n, r_n)})$ satisfies the hypothesis of Lemma 22.3 and X is compact. Hence there is an $x \in X$ such that

$$x \in \cap_n \overline{B(x_n, r_n)} \subset W \cap I.$$

□

We now give three important applications of the Baire category theorem in functional analysis. These are the Principle of Uniform boundedness (also known as the Banach-Steinhaus theorem), the Open Mapping Theorem, and the Closed Graph Theorem. (In learning these theorems, keep careful track of what completeness hypotheses are needed.)

Theorem 22.4 (The Principle of Uniform Boundedness (PUB))**.** *Suppose $\mathcal{X}, \mathcal{Y}$ be normed spaces and $\{T_\alpha : \alpha \in A\} \subset B(\mathcal{X}, \mathcal{Y})$ is a collection of bounded linear transformations from $\mathcal{X}$ to $\mathcal{Y}$. Let B denote the set*

$$B := \{x \in X : M(x) := \sup_{\alpha} \|T_\alpha x\| < \infty\}. \tag{7}$$

If B is of the second category (not a countable union of nowhere dense sets) in X, then

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

In particular, if X is complete and if the collection $\{T_\alpha : \alpha \in A\}$ is pointwise bounded, then it is uniformly bounded.

Proof. For each integer $n \geq 1$ consider the set

$$V_n := \{x \in \mathcal{X} : M(x) > n\}.$$

Since each T_α is bounded, the sets V_n are open. (Indeed, for each α the map $x \to \|T_\alpha x\|$ is continuous from $\mathcal{X}$ to $\mathbb{R}$, so if $\|T_\alpha x\| > n$ for some α then also $\|T_\alpha y\| > n$ for all y sufficiently close to x.) Let E_n denote the complement of V_n and observe that $B = \cup_{n=1}^\infty E_n$. Since B is assumed to be of the second category, there is an N such that $\overline{E_N}^\circ$ is not empty. Since E_N is closed, it follows that E_N has nonempty interior; i.e., there is an $x_0 \in E_N$ and $r > 0$ so that $x_0 - x \in E_N$ for all $\|x\| < r$. Thus, for every α and every $\|x\| < r$,

$$\|T_\alpha x\| \leq \|T_\alpha(x - x_0)\| + \|T_\alpha x_0\| \leq N + N.$$

That is, if $\|x\| < r$, then $M(x) \leq 2N$. By rescaling we conclude that if $\|x\| < 1$, then $\|T_\alpha x\| \leq 2N/r$ for all α and thus $\sup_\alpha \|T_\alpha\| \leq 2N/r < \infty$. □

Given a subset B of a vector space $\mathcal{X}$ and a scalar $s \in \mathbb{K}$, let $sB = \{sb : b \in B\}$. Similarly, for $x \in \mathcal{X}$, let $B - x = \{b - x : b \in B\}$. Let $\mathcal{X}, \mathcal{Y}$ be normed vector spaces and suppose $T : \mathcal{X} \to \mathcal{Y}$ is linear. If $B \subset \mathcal{X}$ and $s \in \mathbb{K}$ is nonzero, then $T(sB) = sT(B)$ and further, an easy argument shows $\overline{T(sB)} = s\,\overline{T(B)}$. It is also immediate that if B is open, then so is $B - x$.

Recall that if X, Y are topological spaces, a mapping $f : X \to Y$ is called *open* if $f(U)$ is open in Y whenever U is open in X. In particular, if f is a bijection, then f is open if and only if f^{-1} is continuous. In the case of normed linear spaces the condition that a linear map be open can be refined somewhat.

Lemma 22.5 (Translation and Dilation lemma)**.** *Let $\mathcal{X}, \mathcal{Y}$ be normed vector spaces, let B denote the open unit ball of $\mathcal{X}$, and let $T : \mathcal{X} \to \mathcal{Y}$ be a linear map. The following are equivalent.*

(i) The map T is open;
(ii) $T(B)$ contains an open ball centered to 0;
(iii) there is an $s > 0$ such that $T(sB)$ contains an open ball centered to 0; and
(iv) $T(sB)$ contains an open ball centered to 0 for each $s > 0$.

Proof. This result is more or less immediate from the fact, for fixed z_0 and $r \in \mathbb{K}$, that the translation map $z \to z + z_0$ and the dilation map $z \to rz$ are continuous in a normed vector space. The implication (i) implies (ii) is immediate. The fact that $T(sB) = sT(B)$ for $s > 0$ readily shows (ii), (iii) and (iv) are equivalent.

To finish the proof it suffice to show (iv) implies (i). Accordingly, suppose (iv) holds and let $U \subset X$ be a given open set. To prove that $T(U)$ is open, let $y \in T(U)$ be given. There is an $x \in U$ such that $T(x) = y$. There is an $s > 0$ such that the ball $B(x, s)$ lies in U; i.e., $B(x,s) \subset U$. The ball $sB = B(0,s) = B(x,r) - x$ is an open ball centered to 0. By hypothesis there is an $r > 0$ such that $B(0,r) \subset T(B(0,s))$. By linearity of T,

$$\begin{aligned} B(y,r) =& B(0,r) + y \subset T(B(0,s)) + y \\ =& T(B(0,s)) + T(x) = T(B(0,s)+x) = T(B(x,s)) \subset T(U). \end{aligned}$$

Thus $T(U)$ is open. □

Theorem 22.6 (Open Mapping)**.** *Suppose that $\mathcal{X}$ is a Banach space, $\mathcal{Y}$ is a normed vector space and $T : \mathcal{X} \to \mathcal{Y}$ is bounded. If the range of T is of second category, then*

(i) $T(\mathcal{X}) = \mathcal{Y}$;
(ii) $\mathcal{Y}$ is complete (so a Banach space); and
(iii) T is open.

In particular, if $\mathcal{X}, \mathcal{Y}$ are Banach spaces, and $T : \mathcal{X} \to \mathcal{Y}$ is bounded and onto, then T is an open map.

Proof. Let $B(x,r)$ denote the open ball of radius r centered at x in $\mathcal{X}$. Trivially $\mathcal{X} = \bigcup_{n=1}^\infty B(0,n)$ and thus $T(X) = \bigcup_{n=1}^\infty T(B(0,n))$. Since the range of T is assumed second category, there is an N such that $T(B(0,N))$ is second category and hence not nowhere dense. In other words, $\overline{T(B(0,N))}$ has nonempty interior. By scaling (see Lemma 22.5), $\overline{T(B(0,1))}$ has nonempty interior. Hence, there exists $p \in \mathcal{Y}$ and $r > 0$ such that $\overline{T(B(0,1))}$ contains the open ball $B^{\mathcal{Y}}(p,r)$. (Here the superscript $\mathcal{Y}$ is used to emphasize this ball is in $\mathcal{Y}$.) It follows that for all $\|y\| < r$,

$$y = -p + (y+p) \in \overline{T(B(0,2))}.$$

In other words,

$$B^{\mathcal{Y}}(0,r) \subset \overline{T(B(0,2))}.$$

By scaling, it follows that, for $n \in \mathbb{N}$,

$$B^{\mathcal{Y}}(0, \frac{r}{2^{n+1}}) \subset \overline{T(B(0, \frac{1}{2^n}))}.$$

We will use the hypothesis that $\mathcal{X}$ iscomplete to prove $B^{\mathcal{Y}}(0, \frac{r}{4}) \subset T(B(0,1))$. Accordingly let y such that $\|y\| < \frac{r}{4}$ be given. Since y is in the closure of $T(B(0,\frac{1}{2}))$, there is a $y_1 \in T(B(0,\frac{1}{2}))$ such that $\|y - y_1\| < \frac{r}{8}$. Since $y - y_1 \in B^{\mathcal{Y}}(0,\frac{r}{8})$ it is is in the closure of $T(B(0,\frac{1}{4}))$. Thus there is a $y_2 \in T(B(0,\frac{1}{4}))$ such that $\|(y-y_1) - y_2\| < \frac{r}{16}$. Continuing in this fashion produces a sequence $(y_j)_{j=1}^\infty$ from $\mathcal{Y}$ such that,

(a) $\|y - \sum_{j=1}^n y_j\| \le \frac{r}{2^{n+2}}$; and
(b) $y_n \in T(B(0, \frac{1}{2^n}$

for all n. It follows that $\sum_{j=1}^\infty y_j$ converges to y. Further, for each j there is an $x_j \in B(0, \frac{1}{2^j})$ such that $y_j = Tx_j$. Since

$$\|x\| \le \sum_{j=1}^\infty \|x_j\| < \sum_{k=1}^\infty 2^{-k} = 1,$$

the series $\sum_{j=1}^{\infty} x_j$ converges to some $x \in B(0,1)$. It follows that $y = Tx$ by continuity of T. Consequently $y \in T(B(0,1))$ and the proof of (i) is complete.

To prove (ii), let $\mathcal{M}$ denote the kernel of T and $\tilde{T}$ the mapping $\tilde{T} : \mathcal{X}/\mathcal{M} \to \mathcal{Y}$ determined by $\tilde{T}\pi = T$. By Lemma 21.17, $\tilde{T}$ is continuous and one-one. Further its range is the same as the range of T, namely $\mathcal{Y}$, and is thus second category. Hence, by what has already been proved, $\tilde{T}$ is an open map. and consequently $\tilde{T}^{-1}$ is continuous. Hence $\mathcal{X}/\mathcal{M}$ and $\mathcal{Y}$ are isomorphic (though of course not necessarily isometric) as normed vector spaces. Therefore, since $\mathcal{X}/\mathcal{M}$ is complete, so is $\mathcal{Y}$. □

Note that the proof of item (ii) in the Open Mapping Theorem shows, in the case that in the case that T is one-one and its range is of second category, that T is onto and its inverse is continuous. In particular, if $T : \mathcal{X} \to \mathcal{Y}$ is a continuous bijection and $\mathcal{Y}$ is a Banach space (so the range of T is second category), then T^{-1} is continuous.

Corollary 22.7 (The Banach Isomorphism Theorem)**.** *If $\mathcal{X}, \mathcal{Y}$ are Banach spaces and $T : \mathcal{X} \to \mathcal{Y}$ is a bounded bijection, then T^{-1} is also bounded (hence, T is an isomorphism).*

To state the final result of this section, we need a few more definitions. Let $\mathcal{X}, \mathcal{Y}$ be normed spaces. The Cartesian product $\mathcal{X} \times \mathcal{Y}$ is a topological space in the *product topology*. A set is open in the product topology if and only if it can be written as a union of products of open sets. Alternately, a set O is open if and only if for each $z = (x,y) \in O$ there exists open sets $U \subset \mathcal{X}$ and $V \subset \mathcal{Y}$ such that $z \in U \times V \subset O$. It is not too hard to show that $\mathcal{X} \times \mathcal{Y}$ is metrizable (in fact the product topology can be realized by norming $\mathcal{X} \times \mathcal{Y}$, e.g. with the norm $\|(x,y)\| := \max(\|x\|, \|y\|)$). It is easy to see that a sequence $z_n = (x_n, y_n)$ converges if and only if both (x_n) and (y_n) converge. Further, if $\mathcal{X}, \mathcal{Y}$ are both Banach spaces (complete), then $\mathcal{X} \times \mathcal{Y}$ is also complete and hence a Banach space. The space $\mathcal{X} \times \mathcal{Y}$ is equipped with the coordinate projections $\pi_{\mathcal{X}}(x,y) = x, \pi_{\mathcal{Y}}(x,y) = y$. It is clear from the definition of the product topology that these maps are continuous (in fact the product topology is the coarsest topology for which the coordinate projections are continuous).

Given a linear map $T : \mathcal{X} \to \mathcal{Y}$, its *graph* is the set

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

Observe that since T is a linear map, $G(T)$ is a linear subspace of $\mathcal{X} \times \mathcal{Y}$. The transformation T is called *closed* if $G(T)$ is a closed subset of $\mathcal{X} \times \mathcal{Y}$. It is an easy exercise to show that $G(T)$ is closed if and only if whenever (x_n, Tx_n) converges to (x,y), we have $y = Tx$. Problem 22.2 gives an example where $G(T)$ is closed, but T is not continuous. On the other hand, if $\mathcal{X}, \mathcal{Y}$ are complete (Banach spaces), then $G(T)$ is closed if and only if T is continuous.

Theorem 22.8 (The Closed Graph Theorem)**.** *If $\mathcal{X}, \mathcal{Y}$ are Banach spaces and $T : \mathcal{X} \to \mathcal{Y}$ is closed, then T is bounded.*

Proof. Let π_1, π_2 be the coordinate projections $\pi_{\mathcal{X}}, \pi_{\mathcal{Y}}$ restricted to $G(T)$; explicitly $\pi_1(x, Tx) = x$ and $\pi_2(x, Tx) = Tx$. Note that π_1 is a bijection between $G(T)$ and $\mathcal{X}$ and in particular $\pi_1^{-1}(x) = (x, Tx)$. By hypothesis $G(T)$ is a closed subset of a Banach space and hence a Banach space. Thus π_1 is a bounded linear bijection between Banach spaces and therefore, by Corollary 22.7, $\pi_1^{-1} : X \to G(T)$ is bounded. Since π_2 is bounded, $\pi_2 \circ \pi_1^{-1} : X \to Y$ is continuous and $\pi_2 \circ \pi_1^{-1}(x) = \pi_2(x, Tx) = Tx$.

□

22.1. **Problems.**

Problem 22.1. Show that there exists a sequence of open, dense subsets $U_n \subset \mathbb{R}$ such that $m(\bigcap_{n=1}^\infty U_n) = 0$.

Problem 22.2. Consider the linear subspace $\mathcal{D} \subset c_0$ defined by

$$\mathcal{D} = \{f \in c_0 : \lim_{n\to\infty} |nf(n)| = 0\}$$

and the linear transformation $T : \mathcal{D} \to c_0$ defined by $(Tf)(n) = nf(n)$.

a) Prove T is closed, but not bounded. b) Prove T is bijective and $T^{-1} : c_0 \to \mathcal{D}$ is bounded (and surjective), but not open. c) What can be said of $\mathcal{D}$ as a subset of c_0?

Problem 22.3. Suppose $\mathcal{X}$ is a vector space equipped with two norms $\|\cdot\|_1, \|\cdot\|_2$ such that $\|\cdot\|_1 \le \|\cdot\|_2$. Prove that if $\mathcal{X}$ is complete in both norms, then the two norms are equivalent.

Problem 22.4. Let $\mathcal{X}, \mathcal{Y}$ be Banach spaces. Provisionally, say that a linear transformation $T : \mathcal{X} \to \mathcal{Y}$ is *weakly bounded* if $f \circ T \in \mathcal{X}^*$ whenever $f \in \mathcal{Y}^*$. Prove, if T is weakly bounded, then T is bounded.

Problem 22.5. Let $\mathcal{X}, \mathcal{Y}$ be Banach spaces. Suppose (T_n) is a sequence in $B(\mathcal{X}, \mathcal{Y})$ and $\lim_n T_n x$ exists for every $x \in \mathcal{X}$. Prove, if T is defined by $Tx = \lim_n T_n x$, then T is bounded.

Problem 22.6. Suppose that $\mathcal{X}$ is a vector space with a countably infinite basis. (That is, there is a linearly independent set $\{x_n\} \subset \mathcal{X}$ such that every vector $x \in \mathcal{X}$ is expressed uniquely as a *finite* linear combination of the x_n's.) Prove there is no norm on $\mathcal{X}$ under which it is complete. (Hint: consider the finite-dimensional subspaces $\mathcal{X}_n := \operatorname{span}\{x_1, \dots x_n\}$.)

Problem 22.7. The Baire Category Theorem can be used to prove the existence of (very many!) continuous, nowhere differentiable functions on $[0,1]$. To see this, let E_n denote the set of all functions $f \in C[0,1]$ for which there exists $x_0 \in [0,1]$ (which may depend on f) such that $|f(x) - f(x_0)| \le n|x - x_0|$ for all $x \in [0,1]$. Prove the sets E_n are nowhere dense in $C[0,1]$; the Baire Category Theorem then shows that the set of nowhere differentiable functions is second category. (To see that E_n is nowhere dense, approximate an arbitrary continuous function f uniformly by piecewise linear functions g, whose pieces have slopes greater than $2n$ in absolute value. Any function sufficiently close to such a g will not lie in E_n.)

Problem 22.8. Let $L^2([0,1])$ denote the Lebesgue measurable functions $f : [0,1] \to \mathbb{C}$ such that $|f|^2$ is in $L^1([0,1])$. It turns out, as we will see later, that $L^2([0,1])$ is a linear manifold (subspace of the vector space $L^1([0,1])$), though this fact is not needed for this problem.

Let $g_n : [0,1] \to \mathbb{R}$ denote the function which takes the value n on $[0, \frac{1}{n^3}]$ and 0 elsewhere. Show,

(i) if $f \in L^2([0,1])$, then $\lim_{n\to\infty} \int g_n f \, dm = 0$;
(ii) $L_n : L^1([0,1]) \to \mathbb{C}$ defined by $L_n(f) = \int g_n f \, dm$ is bounded, and $\|L_g\| = n$;
(iii) conclude $L^2([0,1])$ is of the first category in $L^1([0,1])$.

Problem 22.9. A *Banach space of functions* on a set X is a vector subspace B of the space of complex-valued functions on X with a norm $\|\cdot\|$ making B a Banach space such that, for each $x \in X$, the mapping $E_x : B \to \mathbb{C}$ defined by $E_x(f) = f(x)$ is continuous (bounded) and if $f(x) = 0$ for all $x \in X$, then $f = 0$.

Suppose $g : X \to \mathbb{C}$. Show, if $gf \in B$ for each $f \in B$, then the linear map $M_g : B \to B$ defined by $M_g f = gf$ is bounded.

Problem 22.10. Suppose $\mathcal{X}$ is a Banach space and $\mathcal{M}$ and $\mathcal{N}$ are closed subspaces. Show, if for each $x \in \mathcal{X}$ there exist unique $m \in \mathcal{M}$ and $n \in \mathcal{N}$ such that

$$x = m + n,$$

then the mappings $P : \mathcal{X} \to \mathcal{M}$ defined by $Px = m$ is bounded.

23. Hilbert spaces

23.1. Inner products.

Let $\mathbb{K}$ denote either $\mathbb{C}$ or $\mathbb{R}$.

Definition 23.1. Let X be a vector space over $\mathbb{K}$. An *inner product* on X is a function $u : X \times X \to \mathbb{K}$ such that, for all $x, y, z \in X$ and all $\alpha, \beta \in \mathbb{K}$,

1) $u(x,x) \geq 0$ and $u(x,x) = 0$ if and only if $x = 0$.
2) $u(x,y) = \overline{u(y,x)}$
3) $u(\alpha x + \beta y, z) = \alpha u(x,z) + \beta u(y,z)$.

Notice that 2) and 3) together imply

4) $u(x, \alpha y + \beta z) = \overline{\alpha} u(x,y) + \overline{\beta} u(x,z)$.

◁

Remark 23.2. A function u satisfying only 3) and 4) is called a *bilinear form* (when $\mathbb{K} = \mathbb{R}$) or a *sesquilinear form* (when $\mathbb{K} = \mathbb{C}$). In this case, if 2) is also satisfied then u is called *symmetric* ($\mathbb{R}$) or *Hermitian* ($\mathbb{C}$). A Hermitian or symmetric form satisfying $u(x,x) \geq 0$ for all x is called *positive semidefinite* or a *pre-inner product.*

Typically, u is written $\langle \cdot, \cdot \rangle$ so that $u(x,y) = \langle x, y \rangle$.

Finally, observe if u is a bilinear (resp. sesquilinear) form, then each $y \in X$ induces a linear functional on X defined by $x \mapsto u(x,y)$. ◇

Theorem 23.3 (The Cauchy-Schwarz inequality)**.** *Suppose $\langle \cdot, \cdot \rangle$ is a pre-inner product on the vector space X.*

(i)
$$|\langle x,y\rangle|^2 \le \langle x,x\rangle\,\langle y,y\rangle. \tag{8}$$
(ii) *If $x \in M$ and $y \in X$, then $\langle x,y\rangle = 0$.*
(iii) *The set*
$$M = \{x \in X : \langle x,x\rangle = 0\}$$
is a subspace of X.
(iv) *Let $Y = X/M$ and let $\pi : X \to Y$ denote the quotient map. The form $[p,q] = \langle x,y\rangle$ where $x,y \in X$ are any choices of vectors such that $\pi(x) = p$ and $\pi(y) = q$ is well defined and an inner product on Y.*

Proof. Fix $x,y \in X$. For $t \in \mathbb{R}$, let $\lambda = t\langle x,y\rangle$ and compute, using the nonnegativity assumption
$$\begin{aligned} 0 \le& \langle x-\lambda y, x-\lambda y\rangle \\ =& \langle x,x\rangle - 2t|\langle x,y\rangle|^2 + t^2|\langle x,y\rangle|^2\,\langle y,y\rangle \\ =:& P(t). \end{aligned}$$
Since $P(t)$ is a nonnegative quadratic, its discriminant is nonpositive; i.e.,
$$|\langle x,y\rangle|^4 \le \langle x,x\rangle\,\langle y,y\rangle\,|\langle x,y\rangle|^2$$
and the Cauchy-Schwarz inequality (item (i)) follows.

If $x \in M$ and $y \in X$, then (i) immediately implies $\langle x,y\rangle = 0$, which proves (iii). In particular, if both $x,y \in M$ and $c \in \mathbb{C}$, then
$$\langle x+cy, x+cy\rangle = \langle x,x\rangle + c\langle y,x\rangle + \overline{c}\langle x,y\rangle + |c|^2\langle y,y\rangle = 0$$
and so M is a subpace.

Item (iv) is an exercise in definition chasing. □

23.2. **Examples.**

$\mathbb{K}^n$: It is easy to check that the standard scalar product on $\mathbb{R}^n$ is an inner product; it is defined as usual by
$$\langle x,y\rangle = \sum_{j=1}^n x_j y_j \tag{9}$$
where we have written $x = (x_1, \dots x_n)$; $y = (y_1, \dots y_n)$. Similarly, the standard inner product of vectors $z = (z_1, \dots z_n)$, $w = (w_1, \dots w_n)$ in $\mathbb{C}^n$ is given by
$$\langle z,w\rangle = \sum_{j=1}^n z_j \overline{w_j}. \tag{10}$$
(Note that it is necessary to take complex conjugates of the w's to obtain positive definiteness.)

$\ell^2(\mathbb{N})$: Let

$$\ell^2(\mathbb{N}) = \{(a_1, a_2, \dots a_n, \dots) \mid a_n \in \mathbb{K},\ \sum_{j=1}^{\infty} |a_n|^2 < \infty\}.$$

We may view ℓ^2 as a subset of the vector space $\mathscr{S}$ of all sequence (with domain $\mathbb{N}$) with entrywise addition and scalar multiplication. Define, for sequences $a = (a_1, a_2, \dots)$ and $b = (b_1, b_2, \dots)$ in ℓ^2,

$$\sum_{n=1}^{\infty} a_n \overline{b_n} \tag{11}$$

is seen to converge absolutely using the comparison test and the inequality $2|a_n b_n| \le |a_n|^2 + |b_n|^2$. From here it is not hard to prove that ℓ^2 is closed under the vector space operations of $\mathscr{S}$ and is hence a vector space; and further that (11) defines an inner product, $\langle a, b\rangle = \sum a_n \overline{b_n}$, called the *standard inner product* on ℓ^2.

$L^2(\mu)$: Generalizing the previous example, let $(X, \mathscr{M}, \mu)$ be a measure space. Consider the set of all measurable functions $f : X \to \mathbb{K}$ such that

$$\int_X |f|^2 \, d\mu < \infty$$

The space $L^2(\mu)$ is defined to be this set, modulo the equivalence relation which declares f equivalent to g if $f = g$ almost everywhere. From the inequality $2|f\overline{g}| \le |f|^2 + |g|^2$ it follows that $L^2(\mu)$ is a vector space and that we can define the inner product on $L^2(\mu)$ by

$$\langle f, g\rangle = \int_X f\overline{g} \, d\mu. \tag{12}$$

23.3. **Norms.** Given a vector space X over $\mathbb{K}$ and a semi-inner product $\langle \cdot, \cdot \rangle$, define for each $x \in X$

$$\|x\| := \sqrt{\langle x, x\rangle}. \tag{13}$$

This quantity should act something like a "length" of the vector x. Clearly $\|x\| \ge 0$ for all x, and moreover we have:

Theorem 23.4. *Let X be a semi-inner product space over $\mathbb{K}$, with $\|\cdot\|$ defined by equation (13). Then for all $x, y \in X$ and $\alpha \in \mathbb{K}$,*

(a) $\|x + y\| \le \|x\| + \|y\|$
(b) $\|\alpha x\| = |\alpha|\|x\|$

If $\langle \cdot, \cdot \rangle$ is an inner product, then also

(c) $\|x\| = 0$ *if and only if* $x = 0$ *and thus* $\|\cdot\|$ *is a norm on* X.

Proof. For all $x, y \in X$ we have

$$\begin{aligned} \|x+y\|^2 &= \langle x+y, x+y\rangle \\ &= \|x\|^2 + 2\mathrm{Re}\,\langle x,y\rangle + \|y\|^2 \\ &\le \|x\|^2 + 2|\langle x,y\rangle| + \|y\|^2 \\ &\le \|x\|^2 + 2\|x\|\|y\| + \|y\|^2 \qquad (14) \\ &= (\|x\| + \|y\|)^2 \end{aligned}$$

where we have used the Cauchy-Schwarz inequality in (14). Taking square roots finishes the proof of (b). Items (a) and (c) are left as exercises. □

When X is an inner product space, the quantity $\|x\|$ will be called the *norm* of x. Property (1) will be referred to as the *triangle inequality.* On $\mathbb{R}^n$,

$$\|x\| = (x_1^2 + \cdots + x_n^2)^{1/2},$$

the usual Euclidean norm.

Lemma 23.5. *Let H be an inner product space equipped with the norm topology. If (x_n) converges to x and (y_n) converges to y in H, then $(\langle x_n, y_n\rangle)$ converges to $\langle x, y\rangle$.*

Proof. By Cauchy-Schwarz,

$$|\langle x_n, y_n\rangle - \langle x,y\rangle| \le |\langle x_n, y_n - y\rangle| + |\langle x_n - x, y\rangle| \le \|x_n\|\|y_n - y\| + \|x_n - x\|\|y\| \to 0,$$

since $\|x_n - x\|, \|y_n - y\| \to 0$ and the sequence $\|x_n\|$ is bounded. □

23.4. **Orthogonality.** In this section we show that many of the basic features of the Euclidean geometry of $\mathbb{K}^n$ extend naturally to the setting of an inner product space.

Definition 23.6. Let H be an inner product space.

(i) Two vectors $x, y \in H$ are *orthogonal* if $\langle x,y\rangle = 0$, written $x \perp y$.
(ii) Two subsets A, B of H are *orthogonal* if $x \perp y$ for all $x \in A$ and $y \in B$, written $A \perp B$.
(iii) A subset A of H is *orthogonal* if $x \perp y$ for each $x, y \in A$ with $x \ne y$ and is *orthonormal* if also $\langle x, x\rangle = 1$ for all $x \in A$.
(iv) The *orthogonal complement* of a subset E of H is

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

◁

The proof of the following lemma is an easy exercise.

Lemma 23.7. *If E is a subset of an inner product space H, then*

(i) $E^\perp$ is a closed subspace of H;
(ii) $E \cap E^\perp = (0)$; and
(iii) $E \subset (E^\perp)^\perp = E^{\perp\perp}$.

Theorem 23.8 (The Pythagorean Theorem)**.** *If H is an inner product space and $f_1, \dots f_n$ are mutually orthogonal vectors in H, then*

$$\|f_1 + \cdots + f_n\|^2 = \|f_1\|^2 + \cdots + \|f_n\|^2.$$

Proof. When $n = 2$, we have

$$\begin{aligned}\|f_1 + f_2\|^2 &= \|f_1\|^2 + \langle f_1, f_2\rangle + \langle f_2, f_1\rangle + \|f_2\|^2 \\ &= \|f_1\|^2 + \|f_2\|^2.\end{aligned}$$

The general case follows by induction. □

Theorem 23.9 (The Parallelogram Law)**.** *If H is an inner product space and $f, g \in H$, then*

$$\|f + g\|^2 + \|f - g\|^2 = 2(\|f\|^2 + \|g\|^2). \tag{15}$$

Proof. As usual, from the definition of the norm we have

$$\|f \pm g\|^2 = \|f\|^2 + \|g\|^2 \pm 2\text{Re}\, \langle f, g\rangle.$$

Now add. □

Subtracting, instead of adding, in the proof of the Parallelogram Law gives the polarization identity

$$\|f + g\|^2 - \|f - g\|^2 = 4\text{Re}\, \langle f, g\rangle$$

in the case $\mathbb{K} = \mathbb{R}$.

Theorem 23.10 (The Polarization identity)**.** *If H is an inner product space over $\mathbb{R}$, then*

$$\langle f, g\rangle = \frac{1}{4}\left(\|f + g\|^2 - \|f - g\|^2\right). \tag{16}$$

If H is a complex Hilbert space, then

$$\langle f, g\rangle = \frac{1}{4}[\|f + g\|^2 - \|f - g\|^2 + i\|f + ig\|^2 - i\|f - ig\|^2] \tag{17}$$

Remark: An elementary (but tricky) theorem of von Neumann says, in the real case, that if H is any vector space equipped with a norm $\|\cdot\|$ such that the parallelogram law (15) holds for all $f, g \in H$, then H is an inner product space; the inner product is then given by the formula (16). (The proof is simply to *define* the inner product by equation (16), and check that it is indeed an inner product.) There is of course a complex version as well.

23.5. **Completeness.**

Definition 23.11. A *Hilbert space* over $\mathbb{K}$ is an inner product space X over $\mathbb{K}$ that is complete in the metric $d(x, y) = \|x - y\|$. (Here, as usual, $\mathbb{K}$ is either $\mathbb{C}$ or $\mathbb{R}$.) ◁

That the inner product space $\mathbb{K}^n$ are complete (and hence Hilbert spaces) is known from elementary analysis. (Note that the complex case follows from the real case, since the Euclidean norms are equal under the natural isomorphism $\mathbb{C}^n \cong \mathbb{R}^{2n}$.)

Theorem 23.12. *$L^2(\mu)$ is complete.*

Proof. We use Proposition 20.3. Suppose f_k is a sequence in $L^2(\mu)$ and $\sum_{k=1}^\infty \|f_k\| = B < \infty$. Define

$$G_n = \sum_{k=1}^n |f_k| \quad \text{and} \quad G = \sum_{k=1}^\infty |f_k|.$$

By the triangle inequality, $\|G_n\| \le \sum_{k=1}^n \|f_k\| \le B$ for all n. Thus, by the Monotone Convergence Theorem and the Cauchy-Schwarz inequality,

$$\int_X G^2\, d\mu = \lim_{n\to\infty} \int_X G_n^2\, d\mu \le B^2. \tag{18}$$

Thus G belongs to $L^2(\mu)$ and in particular $G(x) < \infty$ for almost every x. Hence, by the definition of G, the sum

$$\sum_{k=1}^\infty f_k(x)$$

converges absolutely for almost every x. Hence there is a measurable function f such that this sum converges a.e. to f. By construction, $|f| \le G$ and thus $f \in L^2(\mu)$. Moreover, for all n we have

$$\left| f - \sum_{k=1}^n f_k \right|^2 \le (2G)^2.$$

Equation (18) says that G^2 is integrable, so we can apply the Dominated Convergence Theorem to obtain

$$\lim_{n\to\infty} \left\| f - \sum_{k=1}^n f_k \right\|^2 = \lim_{n\to\infty} \int_X \left| f - \sum_{k=1}^n f_k \right|^2 d\mu = 0.$$

□

23.6. **Best approximation.** The results of the preceding subsection used only the inner product, but to go further we will need to invoke completeness. From now on, then, we work only with Hilbert spaces. We begin with a fundamental approximation theorem. Recall that if X is a vector space over $\mathbb{K}$, a subset $K \subseteq X$ is called *convex* if whenever $a, b \in K$ and $0 \le t \le 1$, we have $(1-t)a + tb \in K$ as well. (Geometrically, this means that when a, b lie in K, so does the line segment joining them.)

Theorem 23.13. *Suppose H is a Hilbert space. If $K \subseteq H$ is a closed, convex, nonempty set, and $h \in H$, then there exists a unique vector $k_0 \in K$ such that*

$$\|h - k_0\| = \operatorname{dist}(h, K) := \inf\{\|h - k\| : k \in K\}.$$

Proof. Let $d = \operatorname{dist}(h, K) = \inf_{k\in K} \|h - k\|$. First observe, if $x, y \in K$, then so is $v - \frac{x+y}{2}$ and in particular, $\|h - v\|^2 \geq d$. Hence, by the parallelogram law,

$$\begin{aligned} \left\|\frac{x-y}{2}\right\|^2 =& \frac{1}{2}\left(\|x-h\|^2 + \|y-h\|^2\right) - \left\|\frac{x+y}{2} - h\right\|^2 \\ \leq& \frac{1}{2}\left(\|x-h\|^2 + \|y-h\|^2\right) - d. \end{aligned} \tag{19}$$

By assumption, there exists a sequence (k_n) in K so that $(\|k_n - h\|)$ converges to d. Given $\epsilon > 0$ choose N such that for all $n \geq N$, $\|k_n - h\|^2 < d^2 + \frac{1}{4}\epsilon^2$. By (19), if $m, n \geq N$ then

$$\left\|\frac{k_m - k_n}{2}\right\|^2 < \frac{1}{2}(2d^2 + \frac{1}{2}\epsilon^2) - d^2 = \frac{1}{4}\epsilon^2.$$

Consequently $\|k_m - k_n\| < \epsilon$ for $m, n \geq N$ and (k_n) is a Cauchy sequence. Since H is complete, (k_n) converges to a limit k_0, and since K is closed, $k_0 \in K$. Since $(k_n - h)$ converges to $(k_0 - h)$ and $\|k_n - h\|$ converges to d it follows, by continuity of the norm, that $\|k_0 - h\| = d$.

It remains to show that k_0 is the unique element of K with this property. If $k' \in K$ and $\|k' - h\| = d$, then another application of $v = (k_0 + k')/2$ belongs to K, and $\|v - h\| \geq d$. By the parallelogram law again, equation (19) gives

$$0 \leq \left\|\frac{k_0 - k'}{2}\right\|^2 \leq \frac{1}{2}\left(\|k_0 - h\|^2 + \|k' - h\|^2\right) - d = \frac{1}{2}(d^2 + d^2) - d^2 = 0.$$

Hence $k_0 = k'$. □

The most important application of the preceding approximation theorem is in the case when $K = M$ is a closed subspace of the Hilbert space H. (Note that a subspace is always convex). What is significant is that in the case of a subspace, the minimizer k_0 has an elegant geometric description, namely, it is obtained by "dropping a perpendicular" from h to M. This is the content of the next theorem.

Since we will use the notation often, let us write $M \leq H$ to mean that M is a closed subspace of H.

Theorem 23.14. *Suppose H is a Hilbert space, $M \leq H$, and $h \in H$. If f_0 is the unique element of M such that $\|h - f_0\| = \operatorname{dist}(h, M)$, then $(h - f_0) \perp M$. Conversely, if $f_0 \in M$ and $(h - f_0) \perp M$, then $\|h - f_0\| = \operatorname{dist}(h, M)$.*

Proof. Let $f_0 \in M$ with $\|h - f_0\| = \operatorname{dist}(h, M)$ be given. Given $f \in M$, for $t \in \mathbb{R}$, let $\lambda = t\langle h - f_0, f\rangle$. Since $f_0 + \lambda f \in M$,

$$\begin{aligned} 0 \leq& \|h - (f_0 + \lambda f)\|^2 - \|h - f_0\|^2 \\ =& \|(h - f_0) + \lambda f\|^2 - \|h - f_0\|^2 \\ =& 2\Re\lambda \langle h - f_0, f\rangle + |\lambda|^2\|f\|^2 \\ =& [2t + t^2]\, \|f\|^2\, |\langle h - f_0, f\rangle|^2 \end{aligned}$$

for all t. Thus $\langle h - f_0, f\rangle = 0$.

Conversely, suppose $f_0 \in M$ and $(h - f_0) \perp M$. In particular, we have $(h - f_0) \perp (f_0 - f)$ for all $f \in M$. Therefore, for all $f \in M$

$$\begin{aligned} \|h - f\|^2 &= \|(h - f_0) + (f_0 - f)\|^2 && (20) \\ &= \|h - f_0\|^2 + \|f_0 - f\|^2 \quad \text{(why?)} && (21) \\ &\geq \|h - f_0\|^2. && (22) \end{aligned}$$

Thus $\|h - f_0\| = \operatorname{dist}(h, M)$. □

Corollary 23.15. *If $M \leq H$, then $(M^\perp)^\perp = M$.*

Proof. By Lemma 23.7, $M \subset (M^\perp)^\perp$. Now suppose that $x \in (M^\perp)^\perp$. By Theorem 23.14 applied to x and M, there exists $m \in M$ such that $x - m \in M^\perp$. On the other hand, both x and m are in $(M^\perp)^\perp$ and thus by Lemma 23.7, $x - m \in (M^\perp)^\perp$. Thus $x - m = 0$ by Lemma 23.7, and $x \in M$. □

Corollary 23.16. *If E is a subset of H, then $(E^\perp)^\perp$ is equal to the smallest closed subspace of H containing E.*

Proof. Evidently $E \subset (E^\perp)^\perp$. Further $(E^\perp)^\perp$ is a closed subspace. If M is a closed subspace containing E, then $E^\perp \supset M^\perp$ and thus $E \subset (E^\perp)^\perp \subset (M^\perp)^\perp = M$. The result follows. □

Corollary 23.17. *A vector subspace E of a Hilbert space H is dense in H if and only if $E^\perp = (0)$.*

Lemma 23.18. *Suppose $M, N \leq H$. If M and N are orthogonal, then $M + N$ is closed.*

Given subspaces $M, N \leq H$ of a Hilbert space H, then notation $M \oplus N$ is used for $M + N$ in the case M and N are closed subspaces and $M \perp N$. Hence, $M \oplus N$ indicates that M, N are orthogonal closed subspaces of H.

Proof. It suffices to prove that $M + N$ is complete. Accordingly suppose $(m_k + n_k)$ is a Cauchy sequence from $M + N$. From orthogonality, for $k, \ell \in \mathbb{N}$,

$$\|m_k - m_\ell\|^2 + \|n_k - n_\ell\|^2 = \|(m_k + n_k) - (m_\ell + n_\ell)\|^2$$

and hence (m_k) and (n_k) are both Cauchy. Since H is complete and M, N are closed, M and N are each complete. Thus (m_k) converges to some $m \in M$ and (n_k) converges to some $n \in N$ and thus $(m_k + n_k)$ converges to $m + n \in M + N$. □

Corollary 23.19. *If $M \leq H$, then $H = M \oplus M^\perp$.*

Proof. Given $x \in H$, there exists $m \in M$ such that $x - m \in M^\perp$ by Theorem 23.14. Hence $x = m + (x - m) \in M \oplus M^\perp$. □

23.7. **The Riesz Representation Theorem.** In this section we investigate the dual H^* of a Hilbert space H. One way to construct bounded linear functionals on Hilbert space is as follows. Given a vector $g \in H$ define,

$$L_g(h) = \langle h, g \rangle.$$

Indeed, linearity of L is just the linearity of the inner product in the first entry, and the boundedness of L follows from the Cauchy-Schwarz inequality,

$$|L_g(h)| = |\langle h, g \rangle| \leq \|g\| \|h\|.$$

So $\|L_g\| \leq \|g\|$, but in fact it is easy to see that $\|L_g\| = \|g\|$; just apply L_g to the unit vector $g/\|g\|$. Hence, $L : H \to H^*$ defined by $g \mapsto L_g$ is a *conjugate linear* isometry (thus linear in the case of real scalars).

In fact, it is clear from linear algebra that every linear functional on $\mathbb{K}^n$ takes the form L_g. More generally, every *bounded* linear functional on a Hilbert space has the form just described.

Theorem 23.20 (The Riesz RepresentationTheorem)**.** *If H is a Hilbert space and $\lambda : H \to \mathbb{K}$ is a bounded linear functional, then there exists a unique vector $g \in H$ such that $\lambda = L_g$. Thus the conjugate linear mapping L is isometric and onto.*

Proof. It has already been established that L is isometric and in particular one-one. Thus it only remains to show L is onto. Accordingly, let $\lambda \in H^*$ be given. If $\lambda = 0$, then $\lambda = L_0$. So, assume $\lambda \neq 0$. Since λ is continuous by Proposition 20.7, $\ker \lambda = \lambda^{-1}(\{0\})$ is a *proper, closed* subspace of H. Thus, by Theorem 23.14 (or Corollary 23.19) there exists a nonzero vector $f \in (\ker \lambda)^\perp$ and by rescaling we may assume $\lambda(f) = 1$.

Given $h \in H$, observe

$$\lambda(h - \lambda(h)f) = \lambda(h) - \lambda(h)\lambda(f) = 0.$$

Thus $h - \lambda(h)f \in \ker \lambda$ and consequently,

$$\begin{aligned} 0 &= \langle h - \lambda(h)f, f \rangle \\ &= \langle h, g \rangle - \lambda(h)\langle f, f \rangle. \end{aligned}$$

Thus $\lambda = L_g$, where $g = \frac{f}{\|f\|^2}$ and the proof is complete. □

23.8. **Duality for Hilbert space.** In the case $\mathbb{K} = \mathbb{R}$ the Riesz representation theorem identifies H^* with H. In the case $\mathbb{K} = \mathbb{C}$, the mapping sending $\lambda \in H^*$ to the vector h_0 is *conjugate* linear and thus H^* is not exactly H (under this map). However, it is customary when working in complex Hilbert space not to make this distinction. In particular, it is a simple matter to identify the adjoint of a bounded operator $T : H \to H$ as an operator $T^* : H \to H$. (See Theorem 21.14.) Namely, for $h \in H$, define T^*h as follows. Observe that the mapping $\lambda : H \to \mathbb{C}$ defined by $\lambda(f) = \langle Tf, h \rangle$ is (linear and) continuous. Hence, there is a vector T^*h such that

$$\langle Tf, h \rangle = \lambda(f) = \langle f, T^*h \rangle.$$

Conversely, if $S : H \to H$ is linear and

$$\langle Tf, h\rangle = \langle f, Sh\rangle$$

for all $f, h \in H$, then $S = T^*$. In particular T^* is completely determined by $\langle Tf, h\rangle = \langle f, T^*h\rangle$ for all $f, h \in H$. Further properties of the adjoints on Hilbert space appear in Problem 23.2.

Returning to Theorem 23.14, if $M \leq H$ and $h \in H$, there exists a unique $f_0 \in M$ such that $(h - f_0) \perp M$. We thus obtain a well-defined function $P : H \to H$ (or, we could write $P : H \to M$) defined by

$$Ph = f_0. \tag{23}$$

That is, Ph is characterized by $Ph \in M$ and $(h - Ph) \perp M$. If the space M needs to be emphasized we will write P_M for P.

A bounded operator Q on a Hilbert space H (meaning $Q : H \to H$ is linear and bounded) is a *projection* if $Q^* = Q$ and $Q^2 = Q$.

Theorem 23.21. *Suppose $M \leq H$. The mapping $P = P_M$ is a projection with range M. Moreover, if Q is a projection with range N,then*

(i) if $h \in N$, then $Qh = h$;
(ii) $\|Qh\| \leq \|h\|$ for all $h \in H$;
(iii) N is closed;
(iv) $N^\perp$ is the kernel of Q; and
(v) $I - Q$ is a projection with range $N^\perp$.

The mapping P is called the *orthogonal projection of H on M* and, for $h \in H$, the vector Ph is the *orthogonal projection* of h onto M.

Proof. In view of Corollary 23.19, $M + M^\perp = H$ and $M \cap M^\perp = (0)$ from which it follows readily that P is a linear map.

Evidently P maps into M and if $f \in M$ then $Pf = f$ and hence P maps onto M and $PPf = Pf$ (and so $P^2 = P$). It follows that $P(I - P) = 0$ and therefore the ranges of P and $(I - P)$ are orthogonal.

If $h \in H$, then $h = (h - Ph) + Ph$. But $(h - Ph) \perp M$ and $Ph \in M$, and thus, by the Pythagorean Theorem

$$\|h\|^2 = \|h - Ph\|^2 + \|Ph\|^2.$$

Hence $\|Ph\| \leq \|h\|$. In particular, P is a bounded operator on H. (See also Problem 22.10.)

Given $f \in H$,

$$\begin{aligned}\mathbb{R} \ni \langle Pf, Pf\rangle =& \langle Pf, (I - P)f + Pf\rangle \\ =& \langle Pf, f\rangle \\ =& \langle f, P^*f\rangle \\ =& \langle P^*f, f\rangle,\end{aligned}$$

where $\langle Pf, Pf\rangle$ is real is used in the last equality. Now, if T is any operator on H such that $\langle Tf, f\rangle = 0$ for all $f \in H$, then equation (17) applied to Tf and g (a polarization argument) gives $\langle Tf, g\rangle = 0$ for all f and g. Hence $T = 0$. Applying this fact to $P - P^*$ shows $P - P^* = 0$. Thus all the claims about P have now been proved.

Turning to Q, suppose Q is a projection and let N denote the range of Q. Since $Q^2 = Q$ it follows that $Qh = h$ for $h \in N$ (the range of Q).

An easy computation shows that $Q(I - Q) = 0$. Thus if $h, f \in H$, then

$$\langle Qh, (I - Q)f\rangle = \langle h, Q(I - Q)f\rangle = 0.$$

Choosing $f = h$, it follows that $h = Qh + (I - Q)h$ is an orthogonal decomposition and hence $\|Qh\| \le \|h\|$.

If (h_n) is a sequence from the range of Q which converges to $h \in H$, then, by continuity of Q, the sequence $(h_n = Qh_n)$ converges to Qh and thus $h = Qh$ so that the range of Q is closed.

Next, $f \in N^\perp$ if and only if

$$0 = \langle Qh, f\rangle = \langle h, Qf\rangle$$

for every $h \in H$; if and only if $Qf = 0$. Thus $N^\perp = \ker(Q)$.

Finally, an easy argument shows $I - Q$ is a projection. In particular, f is in the range of $I - Q$ if an donly if $(I - Q)f = f$. On the other hand $(I - Q)f = f$ if and only if $Qf = 0$. Thus the range of $I - Q$ is the kernel of Q.

□

23.9. **Orthonormal Sets and Bases.**

Definition 23.22. Let H be a Hilbert space. A set $E \subset H$ is called *orthonormal* if

a) $\|e\| = 1$ for all $e \in E$, and
b) if $e, f \in E$ and $e \ne f$, then $e \perp f$.

An orthonormal set E is called *maximal* if it is not contained in any larger orthonormal set. A maximal orthonormal set is called an *(orthonormal) basis* for H. ◁

Proposition 23.23. *An orthonormal set E is maximal if and only if the only vector orthogonal to E is the zero vector. Equivalently, an orthonormal set E is maximal if and only if the span of E is dense in H.*

To prove the proposition use $H = \overline{E} \oplus E^\perp$.

Remark 23.24. It must be stressed that a basis in the above sense need *not* be a basis in the sense of linear algebra; i.e., a basis for H as a vector space. In particular, it is always true that an orthonormal set is linearly independent (Exercise: prove this statement), but in general an orthonormal basis need not span H. In fact, if E is an infinite orthonormal subset of H, then E does not span H. See Problem 22.6. ⋄

Example 23.25. Here are some common examples of orthonormal bases.

(a) Of course the standard basis $\{e_1,\dots,e_n\}$ is an orthonormal basis of $\mathbb{K}^n$.

(b) In much the same way we get a orthonormal basis of $\ell^2(\mathbb{N})$; for each n define

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

It is straightforward to check that the set $E=\{e_n\}_{n=1}^\infty$ is orthonormal. In fact, it is a basis. To see this, notice that if $h:\mathbb{N}\to\mathbb{K}$ belongs to $\ell^2(\mathbb{N})$, then $\langle h,e_n\rangle = h(n)$, and hence if $h\perp E$, we have $h(n)=0$ for all n, so $h=0$.

(c) Let $H=L^2[0,1]$. Consider for $n\in\mathbb{Z}$ the functions $e_n(x)=e^{2i\pi nx}$. An easy exercise shows this set is orthonormal. Though not obvious, it is in fact a basis. (See Problem 23.7.)

△

23.10. **Basis expansions.** Our ultimate goal in this section is to show that vectors in Hilbert space admit expansions as (possibly infinite) linear combinations of basis vectors.

Theorem 23.26. *Let $\{e_1,\dots e_n\}$ be an orthonormal set in H, and let $M=\operatorname{span}\{e_1,\dots e_n\}$. The orthogonal projection $P=P_M$ onto M is given by, for $h\in H$,*

$$Ph=\sum_{j=1}^n \langle h,e_j\rangle e_j. \tag{24}$$

Proof. Given $h\in H$, let $g=\sum_{j=1}^n\langle h,e_j\rangle e_j$. Since $g\in M$, it suffices to show $(h-g)\perp M$. For $1\le m\le n$,

$$\begin{aligned}\langle h-g,e_m\rangle &=\langle h,e_m\rangle - \left\langle \sum_{j=1}^n \langle h,e_j\rangle e_j, e_m\right\rangle \\ &= \langle h,e_m\rangle - \sum_{j=1}^n \langle h,e_j\rangle\langle e_j,e_m\rangle \\ &= \langle h,e_m\rangle - \langle h,e_m\rangle = 0.\end{aligned}$$

It follows that f is orthogonal to $\{e_1,\dots,e_n\}$ and hence to M. □

Theorem 23.27 (Gram-Schmidt process)**.** *If $(f_n)_{n=1}^\infty$ is a linearly independent sequence in H, then there exists an orthonormal sequence $(e_n)_{n=1}^\infty$ such that for each n,* $\operatorname{span}\{f_1,\dots f_n\}=\operatorname{span}\{e_1,\dots e_n\}$.

Proof. Put $e_1=f_1/\|f_1\|$. Assuming $e_1,\dots e_n$ have been constructed satisfying the conditions of the theorem, let $g_{n+1}=\sum_{j=1}^n\langle f_{n+1},e_j\rangle e_j$, the orthogonal projection of f_{n+1} onto $M_n=\operatorname{span}\{e_1,\dots,e_n\}$. Thus g_{n+1} is orthogonal to M_n and not 0. Let $e_{n+1}=\frac{g_n}{\|g_{n+1}\|}$. □

It follows from the Gram-Schmidt process that H is finite dimensional as a Hilbert space if and only if H is finite dimensional as a vector space (and these dimensions agree).

Theorem 23.28 (Bessel's inequality)**.** *If $\{e_1, e_2, \dots\}$ is an orthonormal sequence in H, then for all $h \in H$*

$$\sum_{n=1}^{\infty} |\langle h, e_n\rangle|^2 \leq \|h\|^2. \tag{25}$$

Proof. For $N \in \mathbb{N}^+$, let P_N denote the projection onto $M_N = \operatorname{span}(\{e_1, \dots, e_N\})$. Given $h \in H$, Theorem 23.26 and orthogonality gives,

$$\begin{aligned}\|h\|^2 &= \|P_N h + (I - P_N)h\|^2 \\ &= \|P_n h\|^2 + \|(I - P_N)h\|^2 \\ &\geq \|P_N h\|^2 \\ &= \sum_{j=1}^{N} |\langle h, e_j\rangle|^2.\end{aligned}$$

Since the inequality holds for all N, the proof is complete. □

Corollary 23.29. *If $E \subset H$ is an orthonormal set and $h \in H$, then $\langle h, e\rangle$ is nonzero for at most countably many $e \in E$.*

Proof. Fix $h \in H$ and a positive integer N, and define

$$E_N = \{e \in E : |\langle h, e\rangle| \geq \frac{1}{N}\}.$$

We claim that E_N is finite. If not, then it contains a countably infinite subset $\{e_1, e_2, \dots\}$. Applying Bessel's inequality to h and this subset, we get

$$\|h\|^2 \geq \sum_{n=1}^{\infty} |\langle h, e_n\rangle|^2 \geq \sum_{n=1}^{\infty} \frac{1}{N} = +\infty,$$

a contradiction. Hence,

$$\{e \in E : \langle h, e\rangle \neq 0\} = \bigcup_{N=1}^{\infty} E_N$$

is a countable union of finite sets, and therefore countable. □

Theorem 23.30. *Suppose $E \subset H$ is an orthonormal set and let $\mathcal{F}$ denote the collection of finite subsets of E. If $h \in H$, then*

$$\sup\{\sum_{e \in F} |\langle h, e\rangle|^2 : F \in \mathcal{F}\} \leq \|h\|^2. \tag{26}$$

At this point we pause to discuss convergence of infinite series in Hilbert space. We have already encountered ordinary convergence and absolute convergence in our discussion of completeness: recall that the series $\sum_{n=1}^{\infty} h_n$ *converges* if $\lim_{N\to\infty} \sum_{n=1}^{N} h_n$ exists; its limit h is called the sum of the series. The series *converges absolutely* if $\sum_{n=1}^{\infty} \|h_n\| < \infty$ and absolute convergence implies convergence.

The series $\sum_{n=1}^{\infty} h_n$ is *unconditionally convergent* if there exists an $h \in H$ such that for each bijection $\varphi : \mathbb{N} \to \mathbb{N}$ the series $\sum_{n=1}^{\infty} h_{\varphi(n)}$ converges to h. (In other words, every reordering of the series converges, and to the same sum.) Of course absolute convergence implies unconditional convergence. For ordinary scalar series, or in a finite dimensional Hilbert space such as $\mathbb{K}^n$, unconditional convergence implies absolute convergence; however in infinite dimensional Hilbert space unconditional convergence need not imply absolute convergence as the example following Theorem 23.31 shows.

Theorem 23.31. *Suppose $E = \{e_1, e_2, \dots\} \subset H$ is a countable orthonormal set and (a_n) is a sequence of complex numbers. The following are equivalent.*

(i) the series $\sum_{j=1}^{\infty} a_j e_j$ converges;
(ii) $\sum_{j=1}^{\infty} |a_j|^2$ converges; and
(iii) the series $\sum_{j=1}^{\infty} a_j e_j$ converges unconditionally.

Further, if $h \in H$, then the series

$$\sum_{j=1}^{\infty} \langle h, e_j \rangle e_j \tag{27}$$

is unconditionally convergent and, letting g denote the (unconditional) sum,

$$\langle g, e_j \rangle = \langle h, e_j \rangle.$$

Suppose $\{e_1, e_2, \dots\}$ is a countable orthonormal set in a Hilbert space H. The series

$$\sum_{j=1}^{\infty} \frac{1}{j} e_j$$

is Cauchy (verify this as an exercise) and hence converges to some $h \in H$. Moreover, $\langle h, e_j \rangle = \frac{1}{j}$. Hence the series above converges unconditionally to h by Theorem 23.31. On the other hand, this series does not converge absolutely and hence unconditional convergence does not imply absolute convergence.

Proof. Let s_n denote the partial sums of the series $\sum_{j=1}^{\infty} a_j \, e_j$,

$$s_n = \sum_{j=1}^{n} a_j \, e_j.$$

Since H is complete, the series $\sum_{j=1}^{\infty} a_j e_j$ converges if and only if for each $\epsilon > 0$ there is an N so that for all $m \geq n \geq N$,

$$\|s_m - s_n\|^2 = \sum_{j=n+1}^{m} |a_j|^2 < \epsilon \tag{28}$$

if and only if the series $\sum_{j=N+1}^{m} |a_j|^2 < \epsilon$ converges. Hence items (i) and it:squares are equivalent.

Now suppose $\sum_{j=1}^{\infty} |a_j|^2$ converges and let $\varphi : \mathbb{N} \to \mathbb{N}$. For numerical series, absolute convergence implies conditional convergence. Hence $\sum_{j=1}^{\infty} \|a_{\varphi(j)}|^2$ converges and therefore, using the equivalence between item (ii) implies (i) it follows that the series

$$\sum_{k=1}^{\infty} a_{\varphi(k)}\, e_{\varphi(k)}$$

converges to some g', the limit of the partial sums

$$s'_n = \sum_{j=1}^{n} a_{\varphi(j)}\, e_{\varphi(j)}.$$

It remains to show $g' = g$.

Given $\epsilon > 0$, choose N so that (28) holds. Now choose $M \geq N$ so that

$$\{1, 2, \dots N\} \subseteq \{\varphi(1), \varphi(2), \dots \varphi(M)\}.$$

For $n \geq M$, let G_n be the symmetric difference of the sets $\{1, \dots n\}$ and $\{\varphi(1), \dots \varphi(n)\}$ (that is, their union minus their intersection). Since $n \geq M$, the set $G_n \subset \{N+1, N+2, \dots\}$. It follows that

$$\|s_n - s'_n\|^2 = \|\sum_{k \in G_n} \pm a_k e_k\|^2 \tag{29}$$

$$= \sum_{k \in G_n} |a_k|^2 \tag{30}$$

$$\leq \sum_{N+1}^{\infty} |a_j|^2 \tag{31}$$

$$< \epsilon. \tag{32}$$

Thus $g' = g$. Hence item (ii) implies item (iii) and the proof of the first part of the theorem is complete.

For $h \in H$ Bessel's inequality implies the convergence of $\sum |\langle h, e_j\rangle|^2$ and thus, by what has already been proved, the series $\sum \langle h, e_j\rangle\, e_j$ converges unconditionally to some $g \in H$. To complete the proof, given $\epsilon > 0$, choose N so that if $n \geq N$, then

$$\|g - \sum_{j=1}^{n} \langle h, e_j\rangle e_j\| < \epsilon.$$

It follows that, by the Cauchy-Schwarz inequality, for $m \leq n$,

$$|\langle g - \sum_{j=1}^{n} \langle h, e_j\rangle e_j, e_m\rangle|^2 \leq \|g - \sum_{j=1}^{n} \langle h, e_j\rangle e_j\|\, \|e_m\| < \epsilon.$$

On the other hand,

$$\langle g - \sum_{j=1}^{n} \langle h, e_j \rangle e_j, e_m \rangle = \langle g, e_m \rangle - \langle h, e_m \rangle$$

and the desired conclusion follows. □

There is another notion of convergence in Hilbert space. Let I be an index set and let $\mathcal{F}$ denote the collection of finite subsets of I. Given $\{h_i : i \in I\}$, a collection of elements of H, the series

$$\sum_{i \in I} h_i$$

converges as a net if there exists $h \in H$ such that for every $\epsilon > 0$ there exists an $F \in \mathcal{F}$ such that for every $F \subset G \in \mathcal{F}$,

$$\| \sum_{i \in G} h_i - h \| < \epsilon.$$

Proposition 23.32. *If E is an orthonormal subset of a Hilbert space H and $h \in H$, then*

$$\sum_{e \in E} \langle h, e \rangle e$$

converges (as a net). Moreover, if g is the limit (as a net), then

$$\langle g, e \rangle = \langle h, e \rangle.$$

If (h_j) is a sequence from H and

$$\sum_{j \in \mathbb{N}} h_j$$

converges (as a net) to some $h \in H$, then

$$\sum_{j=1}^{\infty} h_j$$

converges unconditionally to h.

Proof. Let $E_0 = \{e \in E : \langle h, e \rangle = 0\}$. From Bessel's inequality, E_0 is at most countable. Suppose E_0 is countable and choose an enumeration, $E_0 = \{e_1, e_2, \dots\}$. The series

$$\sum_{j=1}^{\infty} \langle h, e_j \rangle e_j$$

is unconditionally convergent to some $g \in H$ and moreover $\langle g, e_j \rangle = \langle h, e_j \rangle$ for all j. Given $\epsilon > 0$, there is an N so that if $n \geq N$, then

$$|g - \sum_{j=1}^{n} \langle h, e_j \rangle e_j| < \epsilon$$

and, from Bessel's inequality,

$$\sum_{j=N}^{\infty} |\langle g, e_j\rangle|^2 < \epsilon^2.$$

Let $F = \{e_1, \dots, e_N\}$. If $G \subset E$ is finite and $F \subset G$, then

$$\begin{aligned}\|g - \sum_{e\in G} \langle g, e\rangle\| \le& \|g - \sum_{j=1}^{N}\langle g, e_j\rangle\| + \|\sum_{e\in G\setminus F}\langle g, e\rangle e\| \\ \le& \epsilon + [\sum_{j=N}^{M} |\langle g, e_j\rangle|^2]^{\frac{1}{2}} \\ \le& 2\epsilon,\end{aligned}$$

where $M \in \mathbb{N}$ is chosen so that if $m > M$, then $e_m \notin G$. Hence $\sum_{e\in E}\langle h, e\rangle$ converges as a net to g. Further, by construction, $\langle g, e\rangle = \langle h, e\rangle$ for $e \in E_0$. On the other hand, if $e \notin E_0$, then, for each n,

$$\langle \sum_{j=1}^{n} \langle h, e_j\rangle e_j, e\rangle = 0,$$

and thus $\langle g, e\rangle = 0 = \langle h, e\rangle$. □

Theorem 23.33. *If $E \subset H$ is an orthonormal set, then the following are equivalent:*

(a) E is a (orthonormal) basis for H;

(b) $h = \sum_{e\in E}\langle h, e\rangle e$ for each $h \in H$;

(c) $\langle g, h\rangle = \sum_{e\in E}\langle g, e\rangle\langle e, h\rangle$ for each $g, h \in H$; and

(d) $\|h\|^2 = \sum_{e\in E} |\langle h, e\rangle|^2$ for each $h \in H$.

Proof. Suppose E is an orthonormal set in H and $h \in H$. In this case,

$$\sum_{e\in E}\langle h, e\rangle e$$

converges (as a net) to some $g \in H$ and moreover $\langle g, e\rangle = \langle h, e\rangle$ for all $e \in E$. Suppose $g \neq h$ and let $f = \frac{g-h}{\|g-h\|}$ so that f is a unit vector. If $e \in E$, then

$$\langle f, e\rangle = \frac{1}{\|g - h\|}\langle g - h, e\rangle = 0,$$

and thus E is not maximal. Hence (a) implies (b).

Now suppose (b) holds and let $h, g \in H$ be given. Given ϵ, choose a finite subset F of E such that if $F \subset G \subset E$, then

$$\|h - \sum_{e\in G}\langle h, e\rangle e\|,\ \|g - \sum_{e\in G}\langle g, e\rangle e\| < \sqrt{\epsilon}$$

and observe, using the Cauchy-Schwarz inequality,

$$\begin{aligned}\epsilon >& \big|\langle h - \sum_{e\in G}\langle h,e\rangle e, g - \sum_{f\in G}\langle g,e\rangle e\big| \\ =& \big|\langle h,g\rangle - \sum_{e\in G}\langle h,e\rangle\,\langle e,g\rangle\big|.\end{aligned}$$

Hence (b) implies (c).

Item (d) follows from (c) by choosing $g = h$. Finally, suppose that (a) does not hold. In that case there exists a unit vector $h \in H$ such that h is orthogonal to E. In particular,

$$\sum_{e\in E} |\langle h,e\rangle| = 0$$

and (d) does not hold. □

Theorem 23.34. *Every Hilbert space $H \neq (0)$ has an orthonormal basis.*

Proof. The proof is essentially the same as the Zorn's lemma proof that every vector space has a basis. Let H be a Hilbert space and $\mathcal{E}$ the collection of orthonormal subsets of H, partially ordered by inclusion. Since $H \neq (0)$, the collection $\mathcal{E}$ is not empty. If (E_α) is an ascending chain in $\mathcal{E}$, then it is straightforward to verify that $\cup_\alpha E_\alpha$ is an orthonormal set, and is an upper bound for (E_α). Thus by Zorn's lemma, $\mathcal{E}$ has a maximal element. □

Remark 23.35. If H has a finite orthonormal basis $E = \{e_1, \dots, e_n\}$, then by Theorem 23.33(b), E spans (in the sense of linear algebra) and is therefore a vector space (Hamel) basis for H. Hence H has dimension n as a vector space and further every orthonormal basis of H has exactly n elements.

On the other hand, if H has an infinite orthonormal basis E, then it contains an infinite linearly independent set (the basis E) and so has infinite dimension as a vector space. ◇

Theorem 23.36. *Any two bases of a Hilbert space H have the same cardinality.*

Proof. Let E, F be two orthonormal bases. If E is finite, then H is finite dimensional as a vector space and if E is infinite, then H is infinite dimensional as a vector space. Since all basis of a vector space have the same cardinality, it follows that either both E and F are finite or they are both infinite. If they are both finite, then they are both vector space bases and thus have the same number of elements. So, assume both E and F are infinite. We will show, if F is orthonormal and E is an infinite set and a basis, then the cardinality of F is at most that of E.

Fix $e \in E$ and consider the set

$$F_e = \{f \in F \mid \langle f,e\rangle \neq 0\}.$$

Since F is a basis, each F_e is at most countable, and since E is a basis, each $f \in F$ belongs to at least one F_e. Thus $\bigcup_{e\in E} F_e = F$, and

$$|F| = \left|\bigcup_{e\in E} F_e\right| \leq |E| \cdot \aleph_0 = |E|$$

where the last equality holds since E is infinite.

If both E and F are infinite bases, then by symmetry, $|F| \leq |E|$ and the proof is complete. □

In light of this theorem, we make the following definition.

Definition 23.37. The *(orthogonal) dimension* of a Hilbert space H is the cardinality of any orthonormal basis, and is denoted $\dim H$. If $\dim H$ is finite or countable, H is *separable* or *separable Hilbert space.* ◁

23.11. **Weak convergence.** In addition to the norm topology, Hilbert spaces carry another topology called the *weak topology.* In these notes we will stick to the seperable case and just study weakly convergent sequences.

Definition 23.38. Let H be a seperable Hilbert space. A sequence (h_n) in H *converges weakly to* $h \in H$ if for all $g \in H$,

$$\langle h_n, g\rangle \to \langle h, g\rangle.$$

◁

The Cauchy-Schwarz inequality implies if (h_n) converges to h in norm, then (h_n) converges weakly to h. However, when H is infinite-dimensional, the converse can fail. For instance, let $\{e_n\}_{n=1}^{\infty}$ be an orthonormal basis for H. Then (e_n) converges to 0 weakly. (The proof is an exercise, see Problem 23.9). On the other hand, the sequence (e_n) is not norm convergent, since it is not Cauchy. In this section weak convergence is characterized as "bounded coordinate-wise convergence" and it is shown that the unit ball of a separable Hilbert space is weakly sequentially compact.

Proposition 23.39. *Let H be a Hilbert space with orthonormal basis $\{e_j\}_{j=1}^{\infty}$. A sequence (h_n) in H is weakly convergent if and only if*

i) $\sup_n \|h_n\| < \infty$, *and*
ii) $\lim_n \langle h_n, e_j\rangle$ *exists for each* j.

Proof. Suppose (h_n) converges to h weakly. For each n

$$L_n(g) = \langle g, h_n\rangle$$

is a bounded linear functional on H. Since, for fixed g, the sequence $|L_n(g)|$ converges, it is bounded. Thus, the family of linear functionals (L_n) is pointwise bounded and hence, by the Principle of Uniform boundedness, $\sup \|h_n\| = \sup \|L_n\| < \infty$, showing (i) holds. Item (ii) is immediate from the definition of weak convergence.

Conversely, suppose (i) and (ii) hold, let $M = \sup \|h_n\|$. Define

$$h_j = \lim\langle h_n, e_j\rangle.$$

We will show that $\sum_j |h_j|^2 \le M$ (so that the series $\sum h_j e_j$ is norm convergent in H); we then define h to be the sum of this series and show that $h_n \to h$ weakly.

For positive integers J and all n,

$$\sum_{j=1}^{N} |\langle h_n, e_j\rangle|^2 \le \|h_n\|^2 \le M^2$$

by Bessel's inequality. Thus,

$$\sum_{j=1}^{J} |h_j|^2 = \sum_{j=1}^{J} \lim_n |\langle h_n, e_j\rangle|^2 = \lim_n \sum_{j=1}^{J} |\langle h_n, e_j\rangle|^2 \le M^2.$$

Thus $\sum_j |h_j|^2 \le M$ and therefore the series $\sum_j h_j e_j$ is norm convergent to some $h \in H$ such that $\langle h, e_j\rangle = h_j$ by Theorem 23.31. By Theorem 23.33, $\|h\| \le M$.

Now we prove that (h_n) converges to h weakly. Fix $g \in H$ and let $\epsilon > 0$ be given. Since $g = \sum_j \langle g, e_j\rangle e_j$ (where the series is norm convergent) there exists an positive integer J large enough so that

$$\left\| g - \sum_{j=1}^{J} \langle g, e_j\rangle e_j \right\| = \left\| \sum_{j=J+1}^{\infty} \langle g, e_j\rangle e_j \right\| < \epsilon.$$

Let $g_0 = \sum_{j=1}^{J} \langle g, e_j\rangle e_j$, write $g = g_0 + g_1$, observe $\|g_1\| < \epsilon$ and estimate,

$$|\langle h_n - h, g\rangle| \le |\langle h_n - h, g_0\rangle| + |\langle h_n - h, g_1\rangle|.$$

By (ii), the first term on the right hand side goes to 0 with n, since g_0 is a finite sum of e_j's. By Cauchy-Schwarz, the second term is bounded by $2M\epsilon$. As ϵ was arbitrary, we see that the left-hand side goes to 0 with n. □

It turns out, if (h_n) converges to h weakly, then $\|h\| \le \liminf \|h_n\|$ and further, still assuming (h_n) converges weakly to h, $\|h\| = \lim \|h_n\|$ if and only if (h_n) converges to h in norm. See Problem 23.9.

Theorem 23.40 (Weak compactness of the unit ball in Hilbert space)**.** *If (h_n) is a bounded sequence in a separable Hilbert space H, then (h_n) has a weakly convergent subsequence.*

Theorem 23.40 holds without the separability hypothesis, but the proof is much simpler with the hypothesis.

Proof. Using the previous proposition, it suffices to fix an orthonromal basis (e_j) and produce a subsequence $(h_{n_k})_k$ such that $\langle h_{n_k}, e_j\rangle$ converges for each j. This is a standard "diagonalization" argument, and the details are left as an exercise (Problem 23.11) □

23.12. **Problems.**

Problem 23.1. Prove the complex form of the polarization identity: if H is a Hilbert space over $\mathbb{C}$, then for all $g, h \in H$

$$\langle g, h\rangle = \frac{1}{4}\left(\|g+h\|^2 - \|g-h\|^2 + i\|g+ih\|^2 - i\|g-ih\|^2\right)$$

Problem 23.2. (Adjoint operators) Let H be a Hilbert space and $T : H \to H$ a bounded linear operator.

a) Prove there is a unique bounded operator $T^* : H \to H$ satisfying $\langle Tg, h\rangle = \langle g, T^*h\rangle$ for all $g, h \in H$, and $\|T^*\| = \|T\|$.
b) Prove, if $S, T \in B(H)$, then $(aS + T)^* = \overline{a}S^* + T^*$ for all $a \in \mathbb{K}$, and that $T^{**} = T$.
c) Prove $\|T^*T\| = \|T\|^2$.
d) Prove $\ker T$ is a closed subspace of H, $\overline{(\operatorname{ran} T)} = (\ker T^*)^\perp$ and $\ker T^* = (\operatorname{ran} T)^\perp$.

Problem 23.3. Let H, K be Hilbert spaces. A linear transformation $T : H \to K$ is called *isometric* if $\|Th\| = \|h\|$ for all $h \in H$, and *unitary* if it is a surjective isometry. Prove the following:

a) T is an isometry if and only if $\langle Tg, Th\rangle = \langle g, h\rangle$ for all $g, h \in H$, if and only if $T^*T = I$ (here I denotes the identity operator on H).
b) T is unitary if and only if T is invertible and $T^{-1} = T^*$, if and only if $T^*T = TT^* = I$.
c) Prove, if $E \subset H$ is an orthonormal set and T is an isometry, then $T(E)$ is an orthonormal set in K.
d) Prove, if H is finite-dimensional, then every isometry $T : H \to H$ is unitary.
e) Consider the *shift operator* $S \in B(\ell^2(\mathbb{N}))$ defined by
$$S(a_0, a_1, a_2, \dots) = (0, a_0, a_1, \dots) \tag{33}$$
Prove S is an isometry, but not unitary. Compute S^* and SS^*.

Problem 23.4. For any set J, let $\ell^2(J)$ denote the set of all functions $f : J \to \mathbb{K}$ such that $\sum_{j\in J} |f(j)|^2 < \infty$. Then $\ell^2(J)$ is a Hilbert space.

a) Prove $\ell^2(I)$ is isometrically isomorphic to $\ell^2(J)$ if and only if I and J have the same cardinality. (Hint: use Problem 23.3(c).)
b) Prove, if H is any Hilbert space, then H is isometrically isomorphic to $\ell^2(J)$ for some set J.

Problem 23.5. Let $(X, \mathscr{M}, \mu)$ be a σ-finite measure space. Prove the simple functions that belong to $L^2(\mu)$ are dense in $L^2(\mu)$.

Problem 23.6. (The Fourier basis) Prove the set $E = \{e_n(t) := e^{2\pi i n t} | n \in \mathbb{Z}\}$ is an orthonormal basis for $L^2[0, 1]$. (Hint: use the Stone-Weierstrass theorem to prove that the set of trigonometric polynomials $P = \{\sum_{n=-M}^{N} c_n e^{2\pi i n t}\}$ is uniformly dense in the space of continuous functions f on $[0, 1]$ that satisfy $f(0) = f(1)$. Then show that this space of continuous functions is dense in $L^2[0,1]$. Finally show that if f_n is a sequence in $L^2[0,1]$ and $f_n \to f$ uniformly, then also $f_n \to f$ in the L^2 norm.)

Problem 23.7. Let $(g_n)_{n\in\mathbb{N}}$ be an orthonormal basis for $L^2[0,1]$, and extend each function to $\mathbb{R}$ by declaring it to be 0 off of $[0,1]$. Prove the functions $h_{mn}(x) := \mathbf{1}_{[m,m+1]}(x)g_n(x-m)$, $n \in \mathbb{N}, m \in \mathbb{Z}$ form an orthonormal basis for $L^2(\mathbb{R})$. (Thus $L^2(\mathbb{R})$ is separable.)

Problem 23.8. Let $(X, \mathscr{M}, \mu), (Y, \mathscr{N}, \nu)$ are σ-finite measure spaces, and let $\mu \times \nu$ denote the product measure. Prove, if (f_m) and (g_n) are orthonormal bases for $L^2(\mu), L^2(\nu)$ respectively, then the collection of functions $\{h_{mn}(x,y) = f_m(x)g_n(y)\}$ is an orthonromal basis for $L^2(\mu\times\nu)$. (Problem 23.5 may be helpful.) Use this result to construct an orthonormal basis for $L^2(\mathbb{R}^n)$, and conclude that $L^2(\mathbb{R}^n)$ is separable.

Problem 23.9. (Weak Convergence)

a) Prove, if (h_n) converges to h in norm, then also (h_n) converges to h weakly. (Hint: Cauchy-Schwarz.)
b) Prove, if H is infinite-dimensional, and (e_n) is an orthonormal sequence in H, then $e_n \to 0$ weakly, but $e_n \not\to 0$ in norm. (Thus weak convergence does not imply norm convergence.)
c) Prove (h_n) converges to h in norm if and only if (h_n) converges to h weakly and $\|h_n\| \to \|h\|$.
d) Prove if (h_n) converges to h weakly, then $\|h\| \le \liminf \|h_n\|$.

Problem 23.10. Suppose H is countably infinite-dimensional (separable Hilbert space). Prove, if $h \in H$ and $\|h\| \le 1$, then there is a sequence h_n in H with $\|h_n\| = 1$ for all n, and (h_n) converges to h weakly, but h_n does not converge to h strongly.

Problem 23.11. Prove Theorem 23.40.

Problem 23.12. Prove, if (a_n) is a sequence of complex numbers, then the following are equivalent.

(1) $\sum_{n\in\mathbb{N}} a_n$ converges as a net;
(2) $\sum_{n=1}^\infty a_n$ converges unconditionally;
(3) $\sum_{n=1}^\infty a_n$ converges absolutely.

Problem 23.13. Suppose (h_n) is a sequence from a Hilbert space H. Show, if $\sum_{n=1}^\infty h_n$ converges absolutely, then $\sum_{n=1}^\infty h_n$ converges unconditionally and as a net.

24. L^p spaces

Definition 24.1. Let $(X, \mathscr{M}, \mu)$ be a measure space. For $0 < p < \infty$, let $L^p(\mu)$ denote the space of measurable functions $f : X \to \mathbb{C}$ which satisfy

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

where f and g are identified when $f = g$ a.e. ◁

The inequality

$$|f+g|^p \le (2\max(|f|,|g|))^p \le 2^p(|f|^p + |g|^p)$$

and monotonicity of the integral imply that L^p is a vector space for all $0 < p < \infty$.

Proposition 24.2. *If $1 \le p < \infty$, then $\|f\|_p$ is a norm on L^p.*

Proof. Trivially $\|f\|_p \ge 0$, and if $\|f\|_p = 0$ then $f = 0$ a.e., which means $f = 0$ since we have identified functions that agree almost everywhere. The homogoeneity $\|cf\|_p = |c|\|f\|_p$ is evident from the definition. To prove the triangle inequality, let $f, g \in L^p$ and make the following reductions. First, since $|f+g|^p \le (|f|+|g|)^p$, we can assume $f, g \ge 0$. Using homogeneity, scale f and g by the same constant factor so that $\|f\|_p + \|g\|_p = 1$. Choose $t = \|f\|_p$. Thus $f = tF, g = (1-t)G$ with $F, G \in L^p$ and $\|F\|_p = \|G\|_p = 1$. Proving the triangle inequality has now been reduced to proving

$$\int_X |tF + (1-t)G|^p \, d\mu \le 1. \tag{34}$$

The function $x \to |x|^p$ is convex on $[0, +\infty)$, that is,

$$|tx_1 + (1-t)x_2|^p \le t|x_1|^p + (1-t)|x_2|^p \tag{35}$$

for all $x_1, x_2 \ge 0$. Applying the inequality of Equation (35) to Equation (34) gives

$$\int_X |tF + (1-t)G|^p \, d\mu \le \int_X t|F|^p + (1-t)|G|^p \, d\mu = 1$$

as desired. □

Remark 24.3. Note that the proof of the triangle inequality breaks down when $p < 1$ because the function $x \to |x|^p$ is not convex for these p. In fact, one can use the non-convexity to show that the triangle inequality fails in this range. Indeed since $x \to |x|^p$ is concave, $a^p + b^p > (a+b)^p$ for all $a, b > 0$. Thus, for disjoint sets E, F of finite positive measure,

$$\|\mathbf{1}_E + \mathbf{1}_F\|_p = (\mu(E) + \mu(F))^{1/p} > \mu(E)^{1/p} + \mu(F)^{1/p} = \|\mathbf{1}_E\|_p + \|\mathbf{1}_F\|_p.$$

◇

When $p = \infty$, recall that $L^\infty(\mu)$ is the space of all essentially bounded measurable functions, again identifying functions that agree almost everywhere. Recall too, a function is *essentially bounded* if there exists a number $M < \infty$ such that $|f(x)| \le M$ for a.e. $x \in X$ (equivalently, $\mu(\{|f| > M\}) = 0$) and $\|f\|_\infty$ is the smallest M with this property. In particular we have

$$\|f\|_\infty = \inf\{M : |f(x)| \le M \text{ a.e.}\} = \sup\{M : \mu(\{x : |f(x)| > M\}) > 0\}.$$

It is straightforward to check that $\|f\|_\infty$ is a norm on $L^\infty(\mu)$, and that $f_n \to f$ in L^∞ if and only if f_n converges to f essentially uniformly; and it follows from this that L^∞ is complete. (Problem 24.5).

Example 24.4. If $f = A\mathbf{1}_E$ with $A > 0$ and $0 < \mu(E) < \infty$, then $f \in L^p$ for every $0 < p \leq \infty$. In fact $\|f\|_p = A\mu(E)^{1/p}$ for finite p, and $\|f\|_\infty = A$. Thus $\|f\|_\infty = \lim_{p\to\infty} \|f\|_p$. This result is true more generally (see Problem 24.4). △

The proof that L^p is complete for $1 \leq p < \infty$ is essentially the same as the ones already given in the case of L^1 and L^2 and is left as an exercise.

Theorem 24.5. *For $1 \leq p \leq \infty$, L^p is a Banach space.*

For $p < \infty$, a function f is in L^p if and only if $|f|^p \in L^1$. Thus, an application of Markov's inequality to $|f|^p$ gives the following result.

Proposition 24.6 (Chebyshev's inequality)**.** *Suppose $f \in L^p(\mu)$ and $t > 0$. Then*

$$\mu(\{x : |f| > t\}) \leq \frac{1}{t^p} \int_X |f|^p \, d\mu = \left(\frac{\|f\|_p}{t}\right)^p. \tag{36}$$

Proposition 24.7. *Simple functions are dense in L^p for all $1 \leq p \leq \infty$.*

Proof. We prove the $1 \leq p < \infty$ case and leave $p = \infty$ as an exercise. Let $f \in L^p$. It is straightforward to check that $\mathrm{Re}f, \mathrm{Im}f$ are in L^p, as are the positive and negative parts when f is real valued. Thus, because L^p is a normed space, it may be assumed that f is unsigned. Next, by dominated convergence we see that f can be approximated in L^p by bounded functions (take $f_n = \min(f, n)$ and note that $0 \leq (f - f_n)^p \leq 2^p f^p$ and converges pointwise to 0), so we can assume f is bounded; and again by dominated convergence we can approximate f in L^p by functions supported on sets of finite measure (take f_n to be $\mathbf{1}_{E_n} f$, where $E_n = \{|f| > \frac{1}{n}\}$). Thus we may assume f is nonnegative, bounded, and supported on a set of finite measure. But in this case f can be approximated essentially uniformly by simple functions, and it is easy to verify that if $f_n \to f$ essentially uniformly on a finite measure space, then (f_n) converges to f in L^p also (the proof is the same as in the L^1 case). □

When we write $L^p(\mathbb{R}^n)$ we always refer to Lebesgue measure. We then have the following density result for $L^p(\mathbb{R}^n)$; the proof is essentially the same as the L^1 case.

Proposition 24.8. *For $1 \leq p < \infty$ the space of continuous functions of compact support $C_c(\mathbb{R}^n)$ is dense in $L^p(\mathbb{R}^n)$.*

It should be clear that $C_c(\mathbb{R}^n)$ is *not* dense in $L^\infty(\mathbb{R}^n)$ (why?)

Proof Sketch. Consider the case $n = 1$ and suppose $f \in L^p(\mathbb{R})$. Given $\epsilon > 0$, there is an L^1 simple function ψ such $\|\psi - f\|_1 < \epsilon$ by Proposition 24.7. It suffices (by linearity) to assume $\psi = \mathbf{1}_E$ for a set E with $m(E) < \infty$. By Littlewood's first principle, we can find a set A, a finite union of disjoint open intervals $A = \bigcup_{j=1}^n (a_j, b_j)$, such that $m(A\Delta E) < \epsilon^p$. It follows that $\|\mathbf{1}_A - \mathbf{1}_E\|_p = \|\mathbf{1}_{A\Delta E}\|_p < \epsilon$. Let g_j denote the a continuous, piecewise linear function such that $|g_j(x)| \leq 1$ for all x and is equal to 1 for $x \in (a_j, b_j)$ and 0 for $x < a_j - \frac{1}{2}\frac{\epsilon^p}{2^{jp}}$ and $x > b_j + \frac{1}{2}\frac{\epsilon^p}{2^{jp}}$. (Draw a picture.) Then for each interval $I_j = (a_j, b_j)$, we have $\|g_j - \mathbf{1}_{I_j}\|_p < \epsilon 2^{-j}$. Let $g = \sum_{j=1}^n g_j$. Then g is continuous and belongs to L^1, and from the triangle inequality we have $\|g - \mathbf{1}_A\|_p \le \sum_{j=1}^n \|g_j - \mathbf{1}_{I_j}\|_p < \epsilon$. It follows that $\|g - \mathbf{1}_E\|_p < 2\epsilon$ and hence

$$\|f - g\|_p \le \|f - \mathbf{1}_E\|_p + \|\mathbf{1}_E - \mathbf{1}_A\|_p + \|\mathbf{1}_A - g\|_p < 3\epsilon$$

and the proof is complete in the case $n = 1$.

In higher dimensions, the same approximation scheme works; it suffices (using linearity, Littlewood's first principle, and the $\epsilon/2^n$ trick as before) to approximate the indicator function of a box $B = I_1 \times \cdots \times I_n$; again a piecewise linear function which is 1 on the box and 0 outside a suitably small neighborhood of the box suffices. The details are left as an exercise. □

24.1. **Duality in L^p spaces.** In this section it is shown that, for $1 \le p < \infty$, the dual of the Banach space L^p is L^q, where $\frac{1}{p} + \frac{1}{q} = 1$ (interpret $1/\infty = 0$). The starting point is Hölder's inequality.

Theorem 24.9 (Hölder's inequality)**.** *Suppose* $1 \le p, q \le \infty$ *and* $\frac{1}{p} + \frac{1}{q} = 1$ *(interpret* $1/\infty = 0$*). If* $f \in L^p$ *and* $g \in L^q$*, then* $fg \in L^1$*, and*

$$\|fg\|_1 \le \|f\|_p \|g\|_q.$$

The proof of Hölder's inequality uses Young's inequality.

Lemma 24.10 (Young's inequality)**.** *If* a, b *are nonnegative numbers and* $1 < p, q < \infty$ *satisfy* $\frac{1}{p} + \frac{1}{q} = 1$*, then*

$$ab \le \frac{a^p}{p} + \frac{b^q}{q}.$$

Proof. If a or b is 0 the result is evident. So suppose $a, b > 0$. Begin by verifying, for $1 \le p < \infty$ and $t \in \mathbb{R}$ such that $1 + pt \ge 0$, that $1 + pt \le (1+t)^p$. Let $u = a^p$ and $v = b^q$ and choose t so that $1 + pt = \frac{u}{v}$. Now,

$$\begin{aligned} a\, b =& v(1+pt)^{\frac{1}{p}} \\ \le& v(1+t) \\ =& v\left[\frac{\frac{u}{v} - 1}{p} + 1\right] \\ =& \frac{u}{p} + v(1 - \frac{1}{p}) \\ =& \frac{a^p}{p} + \frac{b^q}{q}. \end{aligned}$$

□

Proof of Theorem 24.9. The proof is very simple in the case $p = \infty$ or $q = \infty$ (in which case the other exponent is 1), so suppose $1 < p, q < \infty$ and f, g are both nonzero. By homogeneity we may normalize so that $\|f\|_p = \|g\|_q = 1$. We must now show that

$$\int_X |fg|\, d\mu \leq 1. \tag{37}$$

Applying Lemma 24.10 gives

$$|f(x)g(x)| \leq \frac{1}{p}|f(x)|^p + \frac{1}{q}|g(x)|^q. \tag{38}$$

Integrating (38) with respect to μ and applying the normalizations on p, q, f, g gives (37). □

Example 24.11. One can get a more intuitive feel for what Hölder's inequality says by examining it in the case of step functions. Let E, F be sets of finite, positive measure and put $f = \mathbf{1}_E, g = \mathbf{1}_F$. Then $\|fg\|_1 = \mu(E \cap F)$ and

$$\|f\|_p \|q\|_q = \mu(E)^{1/p}\mu(F)^{1/q},$$

so Hölder's inequality can be proved easily in this case using the relation $\frac{1}{p} + \frac{1}{q} = 1$ and the fact that $\mu(E \cap F) \leq \min(\mu(E), \mu(F))$. △

Remark 24.12. A more general version of Hölder's inequality says the following: let $1 \leq p, q \leq \infty$ and $\frac{1}{r} = \frac{1}{p} + \frac{1}{q}$. If $f \in L^p$ and $g \in L^q$, then $fg \in L^r$, and

$$\|fg\|_r \leq \|f\|_p \|g\|_q \tag{39}$$

See Problem 24.1. ◇

Numbers $1 \leq p, q \leq \infty$ such that $\frac{1}{p} + \frac{1}{q} = 1$ are *conjugate exponents*. Given p, its conjugate index q is of course uniquely determined, $q = \frac{p}{p-1}$. There are numerous relations between p and q. For instance $p(q-1) = q$ and $(q-1)(p-1) = 1$. Hölder's inequality implies thate each $g \in L^q$ determines a bounded linear functional $L_g : L^p(\mu) \to \mathbb{C}$ by

$$\lambda_g(f) = \int_X f\overline{g}\, d\mu$$

and moreover $\|\lambda_g\| \leq \|g\|_q$. One of the most important facts about L^p spaces is that the converse is true for μ σ-finite and $1 \leq p < \infty$.

Theorem 24.13. *Suppose $1 \leq p < \infty$, $\frac{1}{p} + \frac{1}{q} = 1$ and μ is a σ-finite measure. If $\lambda : L^p(\mu) \to \mathbb{C}$ is a bounded linear functional, then there exists a unique $g \in L^q(\mu)$ such that $\lambda = \lambda_g$. Moreover $\|\lambda\| = \|g\|_q$.*

The following two lemmas will be used to pass from the case of a finite measure to that of a σ-finite measure in the proof of Theorem 24.13.

Lemma 24.14. *If μ is a σ-finite measure on a measurable space $(X, \mathscr{M})$, then there exists a measurable function $w \in L^1(\mu)$ such that $0 < w(x) < 1$ for all x.*

Proof. Write $X = \cup_{n=1}^\infty X_n$, a countable union of disjoint measurable sets of finite measure. Let $w_n = \frac{1}{2^n}\frac{1}{1+\mu(X_n)}\mathbf{1}_{X_n}$ and $w = \sum_{n=1}^\infty w_n$. □

Lemma 24.15. *Suppose μ is a σ-finite measure, $w \in L^1(\mu)$ and $0 < w(x) < 1$ for all x. Let τ denote the measure $w\, d\mu$.*

For $1 < p < \infty$, a measurable function f is in $L^p(\mu)$ if and only if $g = w^{\frac{1}{p}} f \in L^p(\tau)$ and in this case $\|f\|_p = \|g\|_p$; i.e., the mapping $\Phi_p : L^p(\mu) \to L^p(\tau)$ defined by

$$\Phi_p(f) = w^{\frac{1}{p}} f$$

is a (linear) isometric isomorphism.

Proof. It is easy to check that Φ_p is isometric with inverse Ψ defined by $\Psi f = w^{-\frac{1}{p}} f$. $\square$

Proof of Theorem 24.13. Uniqueness of g is clear, since if h is also in $L^q(\mu)$ and $\lambda_g = \lambda_h$, then, for any measureable set E of finite measure, $\lambda_{g-h}(\mathbf{1}_E) = 0$ and it follows, using the σ-finiteness assumption, that $g = h$ μ-a.e.

The idea of the proof is to use the linear functional λ to define a set function $\mathscr{M} \ni E \to \lambda(\mathbf{1}_E)$, prove this set function is a measure absolutely continuous to μ and choose g via the Radon-Nikodym theorem. Finally it must be shown that this g belongs to L^q. The details follow.

Suppose for now that μ is finite; i.e., $(X, \mathscr{M}, \mu)$ is a finite measure space. In this case, if $E \in \mathscr{M}$, then $\mathbf{1}_E \in L^p(\mu)$. Let $u = \mathrm{Re}(\lambda)$. Since λ is bounded, u is also bounded, hence continuous. Define $\nu : \mathscr{M} \to \mathbb{R}$ by

$$\nu(E) := u(\mathbf{1}_E).$$

To prove that ν is a countably additive, suppose $(E_n)_{n=1}^\infty$ is a sequence of disjoint measurable sets and let $E = \cup_{j=1}^\infty E_j$. Let $s_n = \sum_{j=1}^n \mathbf{1}_{E_j}$. In particular, (s_n) increases pointwise with limit $s = \mathbf{1}_E$. Further, $0 \le (s - s_n)^p \le 2^p s$ and thus, by dominated convergence, (s_n) converges to s in $L^p(\mu)$. Using continuity and (real) linearity of u,

$$\nu(E) = u(s) = \lim_{n\to\infty} u(s_n) = \lim_{n\to\infty} \sum_{j=1}^n \nu(E_j) = \sum_{j=1}^\infty \nu(E_j).$$

Evidently $\nu(\emptyset) = 0$ and hence ν defines a finite signed measure on $(X, \mathscr{M})$.

If $\mu(E) = 0$, then $\mathbf{1}_E = 0$ in L^p so $\nu(E) = 0$. Thus, the measure ν is absolutely continuous with respect to μ. Let g_1 denote the Radon-Nikodym derivative $d\nu/d\mu$. Since ν is finite, the Radon-Nikodym theorem gives $g_1 \in L^1(\mu)$. Applying the same arguments to $\mathrm{Im}(\lambda)$, we define g_2 similarly, and put $g = g_1 + i g_2$. In particular, for $E \in \mathscr{M}$,

$$\lambda(\mathbf{1}_E) = \int_E g\, d\mu. \tag{40}$$

Temporarily, view λ_g as defined (and continuous) on $L^\infty(\mu)$.

Now suppose f is a bounded unsigned measurable function. Since μ is a finite measure, f is in $L^p(\mu)$ as well as $L^\infty(\mu)$. Hence, both $\lambda_g(f)$ and $\lambda(f)$ are defined. If s

is a measurable simple function, then by equation (40),

$$\lambda(s) = \int_X sg\, d\mu.$$

There exists a sequence (s_n) of measurable simple functions $0 \le s_n \le f$ such that s_n converges to f uniformly and therefore in both $L^p(\mu)$ and $L^\infty(\mu)$. It follows

$$\lambda(f) = \lim \lambda(s_n) = \lambda_g(s_n) = \lambda_g(f).$$

It now follows that if f is bounded and measurable, then $\lambda(f) = \lambda_g(f)$.

To prove $g \in L^q$, first assume $p > 1$. For positive integers N, let $E_N = \{|g| \le N\}$ and let $g_N = g\mathbf{1}_{E_n}$. Thus g_N is bounded and so is $f_N = \overline{g_N}^{\frac{q}{2}} g_N^{\frac{q}{2}-1}$. By $(L^p(\mu))$ continuity of λ,

$$\|f_N\|_p \,\|\lambda\| \ge |\lambda(f_N)| = |\lambda_g(f_N)| = \int_X f_N g\, d\mu = \int_X |g_N|^q\, d\mu = \|g_N\|_q^q. \tag{41}$$

On the other hand, $\|f_N\|_p = \|g_N\|_q^{q-1}$, and combining this equality with (41) we see that $\|g_N\|_q \le \|\lambda\|$. By monotone congvergence, $\|g\|_q \le \|\lambda\| < \infty$. Now that we know $g \in L^q$, it follows that λ_g is continuous. It also agrees with λ on unsigned bounded functions and therefore on simple functions. Since simple functions are dense in $L^q(\mu)$, the conclusion $\lambda = \lambda_g$ follows. Further, $\|\lambda_g\| = \|\lambda\| \ge \|g\|_q$. and since Hölder gives the reverse inequality, $\|\lambda_g\| = \|g\|_q$.

In the $p = 1$ case, put $E_t = \{|g| > t\}$. If $\mu(E_t) > 0$ for all $t > 0$, let $f_t = \frac{\overline{g}}{g\mu(E_t)}\mathbf{1}_{E_t}$. Then $\|f_t\|_1 = 1$ for all t, while

$$\|\lambda_g\| = \|\lambda_g\|\|f_t\| \ge \int_X f_t g\, d\mu = \frac{1}{\mu(E_t)} \int_{E_t} g\, d\mu \ge t \tag{42}$$

for all t, which is a contradiction once $t > \|\lambda_g\|$. It follows that $g \in L^\infty$ and $\mu(E_t) = 0$ for all $t > \|\lambda_g\|$, so $\|g\|_\infty \le \|\lambda_g\|$.

For the σ-finite case and still with $p > 1$, suppose $X = \cup_{n=1}^\infty X_n$ where the X_n are measureable and of finite measure. For $1 < p < \infty$, let w be as in Lemma 24.14. Likewise, let $\tau = w\, d\mu$. By Lemma 24.15 the mapping $\Phi_p : L^p(\tau) \to L^p(\mu)$ defined by $\Phi_p h = w^{\frac{1}{p}} h$ is a linear isometric isomorphism. Thus, $\psi = \lambda(\Phi_p(f))$ is a bounded linear functional on $L^p(\tau)$. Since τ is a finite measure, by what is already proved, there is a $G \in L^q(\tau)$ such that $\psi = \lambda_G$. Let $g = \Phi_q G = w^{\frac{1}{q}} G$. Thus $g \in L^q(\mu)$ and $\|g\|_q = \|G\|_q$ by Lemma 24.15. Moreover, if $f \in L^p(\mu)$, then $\Phi_p^{-1} f = h = w^{-\frac{1}{p}} f \in L^p(\tau)$ and

$$\begin{aligned}\lambda(f) =& \psi(h) = \lambda_G(h)\\ =& \int h\, G\, d\tau = \int h\, G\, w\, d\mu\\ =& \int w^{\frac{1}{p}} h\, (w^{\frac{1}{q}} G)\, d\mu = \int f\, g\, d\mu = \lambda_g(f).\end{aligned}$$

Extending from finite to σ-finite in the case $p = 1$ is left to the gentle reader. □

Remark 24.16. In the case $(X, \mathscr{M}, \mu)$ is σ-finite and $1 < p < \infty$, Theorem 24.13 can be used to test whether a given measurable function $f : X \to \mathbb{C}$ is in L^p. Namely, letting q denote the conjugate exponent to p, if for each $g \in L^q(\mu)$ the function fg is integrable and the linear functional $\lambda : L^q(\mu) \to \mathbb{C}$ defined by $\lambda(g) = \int fg\, d\mu$ is bounded, then there is an $h \in L^p(\mu)$ such that $\lambda = \lambda_h$. It follows that $\int_E f\, d\mu = \int_E h\, d\mu$ for all measurable sets E of finite measure. Therefore $f = h$ a.e. μ and hence $f \in L^p(\mu)$. $\diamond$

Recall, a Banach space X is reflexive if the canonical isometric inclusion of X into X^{**} is onto. Recall this inclusion is given by mapping $x \in X$ to the linear functional $\hat{x}$ defined by $\hat{x}(\lambda) = \lambda(x)$ for $\lambda \in X^*$.

Corollary 24.17. *For $1 < p < \infty$, L^p is reflexive.*

Proof. Let $\Phi : L^p \to (L^p)^{**}$ denote the canonical inclusion. Thus, for $g \in L^p$, the functional $\Phi(g)$ is given by $\Phi(g)(\lambda) = \lambda(g)$ for $\lambda \in (L^p)^*$. It suffices to show Φ is onto. Accordingly, let $\sigma \in X^{**}$ be given. Let $\Lambda : L^q \to (L^p)^*$ denote the isometric isomorphism $\Lambda(g) = \lambda_g$ given by Theorem 24.13. It follows that $\sigma \circ \Lambda : L^q \to \mathbb{C}$ is a continuous linear functional and hence there is a $g \in L^p(\mu)$ such that $\lambda_g = \sigma \circ \Lambda$. Now let $\lambda \in (L^p)^*$ be given. There is an $h \in L^q(\mu)$ such that $\lambda = \Lambda(h)$. Thus, for $\psi \in L^p$,

$$\lambda(\psi) = \int \psi h\, d\mu$$

It follows that

$$\begin{aligned}\sigma(\lambda) &= \sigma(\Lambda(h)) = \lambda_g(h) \\ &= \int gh\, d\mu = \lambda(g) = \Phi(g)(\lambda).\end{aligned}$$

$\square$

Theorem 24.13 above fails in general when $p = 1$. The canonical inclusion of L^1 into $(L^\infty)^*$ is defined by sending $g \in L^1$ to the bounded linear functional on L^∞ defined by the formula $\lambda_g(f) = \int_X fg\, d\mu$. Unless μ is a finite sum of atoms, there exist bounded linear functionals on L^∞ that are not of this form. An abstract way to see this is that in when μ is σ-finite, $L^1(\mu)$ is separable, but if $L^\infty(\mu)$ is separable, then $(X, \mathscr{M}, \mu)$ is finitely atomic; i.e., $X = \cup_{j=1}^N A_j$ where $\mu(A_j) > 0$ and if $B \subset A_j$ and $\mu(B) < \mu(A_j)$, then $\mu(B) = 0$. Thus if it were the case that $(L^\infty)^* \cong L^1$, then L^∞ would be separable by the result in Problem 21.10, a contradiction. Problem 24.10 gives a somewhat more explicit argument in the case of $L^\infty(\mathbb{R})$.

24.2. **Distribution functions and weak L^p.** Let $(X, \mathscr{M}, \mu)$ be a measure space and $f : X \to \mathbb{C}$ a measurable function. The *distribution function* of f is the function $\lambda_f : (0, +\infty) \to [0, +\infty]$ defined by

$$\lambda_f(t) = \mu(\{x : |f(x)| > t\}).$$

To begin building an intuition about λ_f, we have the following lemma.

Lemma 24.18. *Let $f = \sum_{j=1}^{n} c_j \mathbf{1}_{E_j}$ be a simple function with the E_j disjoint measurable sets, and the c_j ordered as $0 < c_1 < c_2 < \cdots < c_n < +\infty$. Then*

$$\lambda_f(t) = \begin{cases} \mu(E_n) + \cdots + \mu(E_2) + \mu(E_1), & 0 \le t < c_1, \\ \mu(E_n) + \cdots + \mu(E_2), & c_1 \le t < c_2, \\ \cdots & , \cdots \\ \mu(E_n), & c_{n-1} \le t < c_n, \\ 0 & t \ge c_n \end{cases}$$

Proof. Problem 24.11. □

The basic properties of λ_f are collected in the following proposition:

Proposition 24.19. a) *λ_f is decreasing and right continuous.*
b) *[Monotonicity] If $|f| \le |g|$ a.e., then $\lambda_f \le \lambda_g$ everywhere.*
c) *[Monotone convergence] If $|f_n|$ increases to $|f|$ pointwise a.e., then λ_{f_n} increases to λ_f.*
d) *[Subadditivity] If $f = g + h$, then $\lambda_f(t) \le \lambda_g(t/2) + \lambda_h(t/2)$.*

Proof. Let $E(t, f) = \{x : |f(x)| > t\}$. Thus $\lambda_f(t) = \mu(E(t, f))$. In particular, λ_f is decreasing since $E(t, f) \subset E(s, f)$ when $s < t$. Since $E(t, f)$ is the increasing union of $E(t + \frac{1}{n}, f)$, monotone convergence for sets gives

$$\lambda_f(E(t, f)) = \mu(\cup_{n=1}^{\infty} E(t + \frac{1}{n}, f)) = \lim_{n\to\infty} E(t + \frac{1}{n}, f) = \lim_{n\to\infty} \lambda_f(t + \frac{1}{n}).$$

Now, if (t_m) is any sequence converging to t for which $t \le t_m$ for all m, then, given $\epsilon > 0$ choose N so that $\lambda_f(t) - \lambda_f(t_N) < \epsilon$. Choose M such that $t + \frac{1}{M} > t_m \ge t$ for $m \ge M$ and note that, since λ_f is decreasing,

$$0 \le \lambda_f(t) - \lambda_f(t_m) \le \lambda_f(t) - \lambda_f(t_N) < \epsilon$$

to complete the proof of item (a).

Item (b) is immediate from the definition. For (c) we have that for each fixed $t > 0$, $E(t, f)$ is the increasing union of the $E(t, f_n)$. Now argue as in the proof of part (a). Finally (d) is a pigeonhole argument: if $|f(x)| > t$, then either $|g(x)| > t/2$ or $|h(x)| > t/2$. Hence,

$$E(t, f) \subset E(\frac{t}{2}, g) + E(\frac{t}{2}, h).$$

□

The main use of the distribution function is to convert integrals of functions of f into integrals against the measure induced by λ_f. Indeed, the function $\lambda_f(t)$ is decreasing and right continuous on $[0, +\infty]$ and hence defines a (negative) Borel measure ν on $[0, +\infty]$ by

$$\nu((a, b]) := \lambda_f(a) - \lambda_f(b)$$

and passing to the Caratheodory extension. Thus if $\varphi : [0, +\infty] \to \mathbb{C}$ is a Borel function we can consider integrals $\int \varphi \, d\nu = \int \varphi \, d\lambda_f$. The following formula relates these integrals to f and μ.

Proposition 24.20. *Suppose $\lambda_f(t) < \infty$ for all $t > 0$ and $\varphi \geq 0$ is an unsigned Borel function. Then*

$$\int_X \varphi(|f|) \, d\mu = -\int_0^\infty \varphi(t) \, d\lambda_f(t). \tag{43}$$

Proof. Let $0 \leq a < b$ and $\varphi = \mathbf{1}_{(a,b]}$. Then $\varphi(|f|) = \mathbf{1}_{a<|f|\leq b}$, so

$$\int_X \varphi(|f|) \, d\mu = \mu(a < |f| \leq b) = \lambda_f(b) - \lambda_f(a);$$

on the other hand

$$-\int_0^\infty \varphi(t) \, d\lambda_f(t) = -\nu((a,b]) = -(\lambda_f(a) - \lambda_f(b))$$

so (43) holds when $\varphi = \mathbf{1}_{(a,b]}$. Since both sides are linear in φ, it also holds for simple functions, and then for all unsigned Borel functions by monotone convergence. □

The most important case of the above is $\varphi(t) = t^p$, since it will allow us to obtain very useful expressions for the L^p integrals $\int_X |f|^p \, d\mu$. In fact what is most useful is not (43) itself but its "integrated-by-parts" form:

Proposition 24.21. *If $0 < p < \infty$ then*

$$\int_X |f|^p \, d\mu = p \int_0^\infty t^{p-1} \lambda_f(t) \, dt.$$

Proof. This can be proved using the previous proposition and integration by parts for Lebesgue-Stieltjes measures, or directly as follows. If $\lambda_f(t) = +\infty$ for some t then both integrals are infinite. Otherwise, first let f be a simple function; then the identity can be verified directly using Lemma 24.18. For general f, take a sequence of simple functions $0 \leq f_n$ increasing to $|f|$; then the formula holds by Lemma 24.18, Proposition 24.19(c), and monotone convergence. □

The distribution function is used to define the so-called "weak L^p" spaces, as follows: first observe that if $f \in L^p$, then from Chebyshev's inequality we have

$$\mu(\{|f| > t\}) \leq \frac{1}{t^p} \int_X |f|^p \, d\mu$$

or rearranging

$$t^p \lambda_f(t) \leq \|f\|_p^p \quad \text{for all } t > 0.$$

For general f, say that f belongs to *weak L^p* if

$$(\sup_{t>0} t^p \lambda_f(t))^{1/p} := [f]_p < \infty.$$

From what was just said, if $f \in L^p$, then f belongs to weak L^p, but the converse does not hold. The standard example is $f(x) = x^{-1/p}$ on $(0, \infty)$. On the other hand, weak L^p functions are "almost" in L^p, in the sense that if we use Proposition 24.21 we find

$$\int_X |f|^p \, d\mu = p \int_0^\infty t^{p-1} \lambda_f(t) \, dt \leq [f]_p \int_0^\infty t^{-1} \, dt$$

and the integral is just barely divergent.

24.3. **The Hardy-Littlewood maximal function redux.** As an illustration of the usefulness of the distribution function (and the associated idea of splitting L^p functions into their "small" and "large" parts, we reconsider the Hardy-Littlewood maximal function. (This and the next subsection follow sections I.1 and I.4 of *Singular Integrals and Differentiability Properties of Functions* by Eli Stein.) Let f be a locally integrable function on $\mathbb{R}^n$, then

$$(Mf)(x) := \sup_{r>0} \frac{1}{m(B(r,x))} \int_{B(r,x)} |f(y)| \, dy$$

In the language of distribution functions, the Hardy-Littlewood maximal theorem (Theorem 16.4) says that if $f \in L^1(\mathbb{R}^n)$, then Mf belongs to weak L^1, with $[Mf]_1 \leq 3^n \|f\|_1$. We now investigate what happens for $f \in L^p$, $1 < p \leq \infty$, and find the situation is rather better. First, for $p = \infty$ it is trivial that $\|Mf\|_\infty \leq \|f\|_\infty$. For finite p we have:

Theorem 24.22. *If $f \in L^p(\mathbb{R}^n)$, $1 < p < \infty$, then $Mf \in L^p$ and*

$$\|Mf\|_p \leq 2 \left(\frac{3^n p}{p-1} \right)^{1/p} \|f\|_p$$

Proof. Let $f \in L^p(\mathbb{R}^n)$. We fix a parameter $t > 0$ and use it to cut off f: let

$$f_1(x) := \begin{cases} f(x) & \text{if } f(x) \geq t/2, \\ 0 & \text{otherwise} \end{cases}.$$

Note that by vertical truncation, $f_1 \in L^1$, and we have $|f(x)| \leq |f_1(x)| + t/2$ and $(Mf)(x) \leq (Mf_1)(x) + t/2$. It follows that

$$\{x : Mf(x) > t\} \subset \{x : Mf_1(x) > t/2\}$$

so by the Hardy-Littlewood maximal theorem applied to the L^1 function f_1, writing $E_t = \{x : Mf(x) > t\}$ we have

$$m(E_t) \leq \frac{2 \cdot 3^n}{t} \|f_1\|_1 = \frac{2 \cdot 3^n}{t} \int_{|f|>t/2} |f(y)| \, dy. \tag{44}$$

Now write $g = Mf$, then $m(E_t)$ is just the distribution function $\lambda_g(t)$. By Proposition 24.21 we have

$$\int_{\mathbb{R}^n} |Mf(x)|^p \, dx = p \int_0^\infty t^{p-1} \lambda_g(t) \, dt.$$

Using (44) we obtain

$$\|Mf\|_p^p = p\int_0^\infty t^{p-1} m(E_t)\,dt \le p\int_0^\infty t^{p-1}\left(\frac{2\cdot 3^n}{t}\int_{|f|>t/2}|f(y)|\,dy\right)\,dt$$

We apply Fubini to the double integral and inegrate dt first. This gives

$$p\int_0^\infty t^{p-1}\left(2\cdot 3^n\int_{|f|>t/2}|f(y)|\,dy\right)\,dt = \int_{\mathbb{R}^n}|f(y)|\left(\int_0^{2|f(y)|} t^{p-2}\,dt\right)\,dy$$

The inner integral is

$$\int_0^{2|f(y)|} t^{p-2}\,dt = \frac{1}{p-1}2^{p-1}|f(y)|^{p-1}$$

so we have finally

$$\|Mf\|_p^p \le \frac{2^p\cdot 3^n p}{p-1}\int_{\mathbb{R}^n}|f(y)|^p\,dy$$

which finishes the proof. □

24.4. **The Marcinkiewicz interpolation theorem.** The idea used in the proof of the L^p boundedness of the Hardy-Littlewood maximal operator can be extended to prove a more general result, called the *Marcinkiewicz interpolation theorem.* We will not consider the most general version of the theorem here (it may be found in Folland) but prove a special case that is adequate for many purposes. We fix a measureable space $(X, \mathscr{M}, \mu)$; L^p always refers to $L^p(\mu)$.

We need a few definitions. Let $1 \le p, q \le \infty$. A mapping $T : L^p \to L^q$ is of *type* (p, q) if there is a constant $A > 0$ so that

$$\|Tf\|_q \le A\|f\|_p$$

for all $f \in L^p$. The mapping is of *weak type* (p, q) if

$$\mu(\{x : |Tf(x)| > t\} \le \left(\frac{A\|f\|_p}{t}\right)^q, \quad \text{for } q < \infty$$

where the constant A does not depend on f or t. (In other words, T maps L^p into weak L^q, with $[Tf]_q \le A\|f\|_p$.)

When $q = \infty$ weak type (p, q) simply means type (p, q).

We say that a transformation T (defined on some space of measurable functions) is *sub-linear* if $|Tf(x)| \le |Tg(x)| + |Th(x)|$ for $f = g + h$, and $|T(cf)| = |c||Tf|$ for all scalars c.

Finally we let $L^p + L^q$ denote the vector space of functions of the form $f = g + h$ where $g \in L^p, h \in L^q$. Note that if $p < r < q$, then by the truncation lemmas we see that $L^r \subset L^p + L^q$.

Theorem 24.23 (Marcinkiewicz interpolation theorem (special case))**.** *Suppose that* $1 < r \le \infty$*. If* T *is a sub-linear transformation from* $L^1 + L^r$ *to the vector space of measurable functions, and* T *is of weak type* $(1, 1)$ *and weak type* (r, r)*, then* T *is of type* (p, p) *for all* $1 < p < r$*.*

Proof. The case $r = \infty$ closely parallels the proof given for the Hardy-Littlewood maximal function and is left as an exercise, so we assume $1 < r < \infty$. Fix $f \in L^p$, we wish to estimate its distribution function $\lambda_f(t)$. We fix $t > 0$ for the moment and use this to cut off f: define

$$f_1(x) := \begin{cases} f(x) & \text{if } |f(x)| > t \\ 0 & \text{if } |f(x)| \leq t \end{cases}$$

$$f_2(x) := \begin{cases} f(x) & \text{if } |f(x)| \leq t \\ 0 & \text{if } |f(x)| > t \end{cases}$$

so that $f = f_1 + f_2$, $f_1 \in L^1$ and $f_2 \in L^r$. Since $|Tf| \leq |Tf_1| + |Tf_2|$, we have by the subadditvity of λ

$$\lambda_{Tf}(t) \leq \lambda_{Tf_1}(t/2) + \lambda_{Tf_2}(t/2)$$

and so by the assumption that T is of weak type $(1,1)$ and (r,r),

$$\lambda_{Tf}(t) \leq \frac{A_1}{t/2} \int |f_1(x)|\, d\mu(x) + \frac{A_r^r}{(t/2)^r} \int |f_2(x)|^r\, d\mu(x)$$

for some fixed constants A_1, A_r as in the definition of weak type. Because of the choice of splitting $f = f_1 + f_2$, we have

$$\lambda_{Tf}(t) \leq \frac{2A_1}{t} \int_{|f|>t} |f(x)|\, d\mu(x) + \frac{(2A_r)^r}{t^r} \int_{|f|\leq t} |f(x)|^r\, d\mu(x) \tag{45}$$

Using the formula $\|Tf\|_p^p = p\int_0^\infty t^{p-1}\lambda_{Tf}(t)\, dt$ we multiply both sides of (45) by pt^{p-1} and integrate dt. To handle first integral in (45) we observe

$$\int_0^\infty t^{p-1}t^{-1} \int_{|f|>t} |f(x)|\, d\mu(x)dt = \int_X |f| \int_0^{|f|} t^{p-2}\, dt\, d\mu(x) \tag{46}$$

$$= \frac{1}{p-1} \int_X |f||f|^{p-1}\, d\mu \tag{47}$$

since $p > 1$, similarly for the second integral

$$\int_0^\infty t^{p-1}t^{-r} \int_{|f|\leq t} |f(x)|^r\, d\mu(x)dt = \int_X |f|^r \int_{|f|}^\infty t^{p-1-r}\, dt\, d\mu(x) \tag{48}$$

$$= \frac{1}{r-p} \int_X |f|^r|f|^{p-r}\, d\mu \tag{49}$$

Putting these together we find that

$$\|Tf\|_p \leq A_p\|f\|_p, \quad \text{with } A_p^p = \left(\frac{2^r A_1}{p-1} + \frac{(2A_r)^r}{r-p}\right) p.$$

$\square$

24.5. **Some inequalities.** The following should be viewed as a continuous analog of the triangle inequality for the L^p norm. It says that "the L^p norm of the integral is less than the integral of the L^p norms."

Theorem 24.24 (Minkowski's inequality for integrals)**.** *Let* $(X, \mathscr{M}, \mu)$ *and* $(Y, \mathscr{N}, \nu)$ *be* σ*-finite measure spaces and let* f *be a jointly measurable function on* $X \times Y$*.*

a) *If* f *is unsigned and* $1 \le p < \infty$*, then*

$$\left[\int_X \left(\int_Y f(x,y)\, d\nu(y)\right)^p d\mu(x)\right]^{1/p} \le \int_Y \left[\int_X f(x,y)^p\, d\mu(x)\right]^{1/p} d\nu(y) \tag{50}$$

b) *If* $1 \le p \le \infty$*,* $f(\cdot, y) \in L^p(\mu)$ *for a.e.* y*, and the function* $\|f(\cdot,y)\|_p$ *is in* $L^1(\nu)$*, then* $f(x,\cdot) \in L^1(\nu)$ *for a.e.* x*, the function* $\int f(x,y)\, d\nu(y)$ *is in* $L^p(\mu)$*, and*

$$\left\| \int f(\cdot, y)\, d\nu(y) \right\|_p \le \int \|f(\cdot,y)\|_p\, d\nu(y).$$

The σ-finite hypothesis is needed to invoke Tonelli's Theorem.

Proof. For $1 \le p < \infty$, item (b) follows easily from item (a). For $p = \infty$, item (b) is immediate. To prove (a), note that the result is just Tonelli's theorem when $p = 1$. For the case $1 < p < \infty$, if the right-hand side of (50) is infinite, there is nothing to prove. Assuming this right hand side is finite, let $h(x) = \int f(x,y)\, d\nu(y)$ and fix $g \in L^q(\mu)$ where q is the conjugate index to p. By Tonelli and Hölder,

$$\begin{aligned} \int h(x)|g(x)|\, d\mu(x) &= \iint f(x,y)|g(x)|\, d\mu(x) d\nu(y) \\ &\le \left[\int \left[\int f(x,y)^p\, d\mu(x)\right]^{1/p} d\nu(y)\right] \|g\|_q. \end{aligned}$$

Thus the linear functional $\lambda : L^q(\mu) \to \mathbb{C}$ given by $\lambda(g) = \int hg\, d\mu$ is bounded on $L^q(\mu)$ with norm at most

$$\int \left[\int f(x,y)^p\, d\mu(x)\right]^{1/p} d\nu(y).$$

By Theorem 24.13, it follows that $h \in L^p(\mu)$ and obeys the estimate (50). $\square$

There are many important linear operators on L^p spaces arise as *integral operators* in the sense that they are expressible in the form

$$Tf(x) = \int K(x,y) f(y)\, d\nu(y). \tag{51}$$

The next proposition gives a simple sufficient condition for the boundedness of such an operator on L^p. It is a special case of a more general criterion known as the *Schur test.*

Proposition 24.25 (Young's inequality)**.** *Let μ, ν be σ-finite positive measures on spaces X, Y respectively and let $K(x,y)$ be a jointly measurable funciton on $X \times Y$. Suppose there is a constant C so that*

$$\sup_{x\in X} \int_Y |K(x,y)|\, d\nu(y) \le C, \quad \sup_{y \in Y} \int_X |K(x,y)|\, d\mu(x) \le C.$$

If $f \in L^p(\nu)$, $1 \le p \le \infty$, then, for almost every x, $K(x,y)f(y) \in L^1(\nu)$, the function Tf defined by (51) is in $L^p(\mu)$ and

$$\|Tf\|_p \le C\|f\|_p.$$

Once again, the σ-finite hypothesis is used to invoke Tonelli's Theorem.

Proof. For $p = 1$ or $p = \infty$, the proof is straightforward and left as an exercise. Suppose $1 < p < \infty$ and let q denote the conjugate index. The idea is to split K as $K = K^{1/q}K^{1/p}$ and apply Hölder's inequality. Indeed,

$$\begin{aligned}\int_Y |K(x,y)||f(y)|\, d\nu(y) &\le \left(\int_Y |K(x,y)|\, d\nu(y)\right)^{1/q} \left(\int_Y |K(x,y)||f(y)|^p\, d\nu(y)\right)^{1/p} \\ &\le C^{1/q}\left(\int_Y |K(x,y)||f(y)|^p\, d\nu(y)\right)^{1/p}.\end{aligned}$$

Thus, by Tonelli there is a measurable function $g(x)$ such that $g(x) = \int K(x,y)f(y)\, d\nu(y)$ almost everywhere ν. Further, by Tonelli's theorem,

$$\begin{aligned}\int_X |g(x)|^p\, d\mu(x) &\le C^{p/q}\int_X \left(\int_Y |K(x,y)||f(y)|^p\, d\nu(y)\right) d\mu(x) \\ &= C^{p/q}\int_Y |f(y)|^p \left(\int_X |K(x,y)|\, d\mu\right) d\nu(y) \\ &\le C^{1+p/q}\int_Y |f(y)|^p\, d\nu(y) \\ &= C^p\|f\|_p^p.\end{aligned}$$

Hence $g \in L^p(\mu)$ and and taking p^{th} roots finishes the proof. □

One context in which Young's inequality is often used is the following: let $X = Y = \mathbb{R}^n$ with Lebesgue measure, and fix a function $g \in L^1(\mathbb{R}^n)$. If we put $K(x,y) = g(x-y)$, then for a measurable function f on $\mathbb{R}^n$ the function

$$Tf(x) := \int_{\mathbb{R}^n} g(x-y)f(y)\, dy$$

is called the *convolution* of f and g, defined at each x where the integrand is L^1. Usually we write $Tf = g * f$. From the translation invariance of Lebesgue measure,

$$\sup_{x\in\mathbb{R}^n} \int |g(x-y)|\, dy = \sup_{y\in\mathbb{R}^n} \int |g(x-y)|\, dx = \|g\|_1$$

so the hypotheses of Proposition 24.25 are satisfied with $C = \|g\|_1$. We conclude

Corollary 24.26 (Young's inequality for convolutions)**.** *If $g \in L^1(\mathbb{R}^n)$, $f \in L^p(\mathbb{R}^n)$, $1 \leq p \leq \infty$, then $g * f \in L^p(\mathbb{R}^n)$ and*

$$\|g * f\|_p \leq \|g\|_1 \|f\|_p.$$

24.6. **Problems.**

Problem 24.1. Prove the generalized Hölder inequality (39). (Hint: consider $F = |f|^r, G = |g|^r$.)

Problem 24.2. Suppose $f, g \geq 0$ with $f \in L^p, g \in L^q$, $0 < p, q < \infty$. Show that equality holds in Hölder's inequality (and the generalized Hölder inequality) if and only if one the functions f^p, g^q is a scalar multiple of the other.

Problem 24.3. [Truncation of L^p functions] Suppose f is an unsigned function in $L^p(\mu)$, $1 < p < \infty$. For $t > 0$ let

$$E_t = \{x : f(x) > t\}.$$

a) Show that for each real number $t > 0$, the *horizontal truncation* $\mathbf{1}_{E_t} f$ belongs to L^q for all $1 \leq q \leq p$.

b) Show that for each real number $t > 0$, the *vertical truncation* $f_t := \min(f, t)$ belongs to L^q for all $p \leq q \leq \infty$.

c) As a corollary, show that every $f \in L^p$, $1 < p < \infty$, can be decomposed as $f = g + h$ where $g \in L^1$ and $h \in L^\infty$.

Problem 24.4. Suppose $f \in L^{p_0} \cap L^\infty$ for some $p_0 < \infty$. Prove $f \in L^p$ for all $p_0 \leq p \leq \infty$, and $\lim_{p \to \infty} \|f\|_p = \|f\|_\infty$.

Problem 24.5. Prove $f_n \to f$ in the L^∞ norm if and only if $f_n \to f$ essentially uniformly, and that L^∞ is complete.

Problem 24.6. Suppose $p_0 < p < p_1$ and $f \in L^{p_0} \cap L^{p_1}$. Prove $f \in L^p$ and $\|f\|_p \leq \|f\|_{p_0}^{1-\theta} \|f\|_{p_1}^{\theta}$, where $0 < \theta < 1$ is chosen so that $\frac{1}{p} = \frac{1-\theta}{p_0} + \frac{\theta}{p_1}$. When does equality hold?

Problem 24.7. [Containments of L^p spaces] a) Show that if μ is a finite measure, then $L^p \subset L^q$ for all $p \geq q$. b) Show that $\ell^p \supset \ell^q$ for all $p \geq q$. c) More generally, show that $L^p \subset L^q$ for all $p \geq q$ if and only if μ does not admit sets of arbitrarily large finite measure, and $L^p \supset L^q$ for all $p \geq q$ if and only if μ does not admit sets of arbitrarily small positive measure.

Problem 24.8. Show that $L^p(\mathbb{R}^n) \not\subset L^q(\mathbb{R}^n)$ for any pair p, q.

Problem 24.9. [Convergence in L^p norm] Prove, if $f_n \to f$ in the L^p norm, then $f_n \to f$ in measure, and hence a subsequence converges to f a.e. Conversely, show that if $f_n \to f$ in measure and there exists a $g \in L^p$ such that $|f_n| \leq g$ for all n, then $f_n \to f$ in the L^p norm. (Hint: go back and look at the results in Section 12 in last semester's notes, especially Remark 12.18 and Corollary 12.19.)

Problem 24.10. Consider $L^\infty(\mathbb{R})$.

a) Show that $\mathcal{M} := C_0(\mathbb{R})$ is a closed subspace of $L^\infty(\mathbb{R})$ (more precisely, that the set of L^∞ functions that are a.e. equal to a C_0 function is closed in L^∞). Prove there is a bounded linear functional $\lambda : L^\infty \to \mathbb{K}$ such that $\lambda|_{\mathcal{M}} = 0$ and $\lambda(\mathbf{1}_{\mathbb{R}}) = 1$.
b) Prove there is no function $g \in L^1(\mathbb{R})$ such that $\lambda(f) = \int_{\mathbb{R}} fg\, dm$ for all $f \in L^\infty$. (Hint: look at the restriction of λ to $C_0(\mathbb{R})$.)

Problem 24.11. Prove Lemma 24.18, and use it to carry out the calculation omitted in the proof of Proposition 24.21.

Problem 24.12. Prove, if $f \in L^p$ then

$$\lim_{t\to 0} t^p \lambda_f(t) = \lim_{t\to\infty} t^p \lambda_f(t) = 0$$

(One way to proceed is to first suppose f is a simple function. Another is to consider the integrals $\int_{\frac{s}{2}}^{s} t^{p-1}\lambda_f(t)dt$.)

Problem 24.13. Prove the $r = \infty$ case of the Marcinkiewicz interpolation theorem.

Problem 24.14. Prove the following more general form of Young's inequality: suppose $p, q, r \geq 1$ satisfy $\frac{1}{q} = \frac{1}{p} + \frac{1}{r} - 1$. Suppose $K(x, y)$ satisfies

$$\sup_{x\in X} \|K(x, \cdot)\|_{L^r(\nu)} \leq C, \quad \sup_{y\in Y} \|K(\cdot, y)\|_{L^r(\mu)} \leq C$$

for some constant C. Prove, if $f \in L^p(\nu)$, then $Tf = \int_Y K(x, y)f(y)\, d\nu(y) \in L^q(\mu)$, and

$$\|Tf\|_q \leq C\|f\|_p.$$

(Hint: use the same strategy of splitting $|K| = |K|^\alpha |K|^\beta$, for a suitable choice of $\alpha+\beta = 1$.)

Deduce the following corollary for convolutions on $\mathbb{R}^n$: with p, q, r as above, if $f \in L^p(\mathbb{R}^n)$ and $g \in L^r(\mathbb{R}^n)$, then $g * f \in L^q(\mathbb{R}^n)$ and

$$\|g * f\|_q \leq \|g\|_r \|f\|_p.$$

25. The Fourier transform

We assume all functions are complex-valued unless stated otherwise.

Definition 25.1. [The Fourier transform] Let $f \in L^1(\mathbb{R})$. The *Fourier transform* of f is the function $\widehat{f}$ defined at each $t \in \mathbb{R}$ by

$$\widehat{f}(t) := \int_{-\infty}^{\infty} f(x)e^{-2\pi itx}\,dx \tag{52}$$

◁

Note that $\widehat{f}$ makes sense, since the integrand belongs to L^1 for each $t \in \mathbb{R}$. We sometimes also use the phrase *Fourier transform* for the mapping that sends f to $\widehat{f}$. The basic properties of the Fourier transform listed in the following proposition stem from two basic facts: first, that Lebesgue measure is translation invariant, and second that, for each $t \in \mathbb{R}$, the function

$$\chi_t : x \to \exp(2\pi itx)$$

is a *character* of the additive group $(\mathbb{R}, +)$. This means that χ_t is a homomorphism from $\mathbb{R}$ into the mulitplicative group of unimodular complex numbers, explicitly for all $s, t \in \mathbb{R}$

$$\chi_t(x+y) = \chi_t(x)\chi_t(y).$$

Before going further we introduce some notation: for fixed $y \in \mathbb{R}$ and a function $f : \mathbb{R} \to \mathbb{C}$, define $f_y(x) := f(x-y)$.

Proposition 25.2 (Basic properties of the Fourier transform)**.** *Let $f, g \in L^1(\mathbb{R})$ and let $\alpha \in \mathbb{R}$.*

(a) (Linearity) $\widehat{cf+g} = c\widehat{f} + \widehat{g}$
(b) (Translation) $\widehat{f_y}(t) = e^{-2\pi ity}\widehat{f}(t)$
(c) (Modulation) If $g(x) = e^{2\pi i\alpha x}f(x)$, then $\widehat{g}(t) = \widehat{f}(t-\alpha)$
(d) (Reflection) If $g(x) = \overline{f(-x)}$, then $\widehat{g}(t) = \overline{\widehat{f}(t)}$.
e) (Scaling) If $\lambda > 0$ and $g(x) = f(x/\lambda)$ then $\widehat{g}(t) = \lambda\widehat{f}(\lambda t)$.

Proof. Each of these properties is verified by elementary transformations of the integral defining $\widehat{f}$; the details are left as an exercise. □

It is immediate from the definition that $\widehat{f}$ is always a bounded function; indeed $|\widehat{f}(t)| \leq \|f\|_1$ for all t. Our next observation is:

Proposition 25.3. *If $f \in L^1(\mathbb{R})$, then $\widehat{f}$ is continuous and bounded ($\widehat{f} \in C_b(\mathbb{R})$) and $\|\widehat{f}\|_\infty \leq \|f\|_1$. In particular, the mapping $L^1(\mathbb{R}) \to L^\infty(\mathbb{R})$ is a bounded linear map. Moreover if f_n is sequence in L^1 and $f_n \to f$ in the L^1 norm, then $\widehat{f_n} \to \widehat{f}$ uniformly.*

Proof. Fix $t \in \mathbb{R}$ and a sequence $t_n \to t$. The sequence $f(x)e^{-2\pi it_nx}$ converges to $f(x)e^{-2\pi itx}$ pointwise on $\mathbb{R}$, and since trivially $|f(x)e^{-2\pi it_nx}| \leq |f(x)|$ for all n, we have by dominated convergence

$$\begin{aligned}\widehat{f}(t) &= \int_{-\infty}^{\infty} f(x)e^{-2\pi itx}\,dx \\ &= \int_{-\infty}^{\infty} \lim_{n\to\infty}[f(x)e^{-2\pi it_n x}]\,dx \\ &= \lim_{n\to\infty}\int_{-\infty}^{\infty} f(x)e^{-2\pi it_n x}\,dx \\ &= \lim_{n\to\infty}\widehat{f}(t_n).\end{aligned}$$

The second statement of the theorem follows immediately from the estimate $\sup_{t\in\mathbb{R}}|\widehat{f}(t)| \le \|f\|_1$. □

In fact, $\widehat{f}$ always belongs to $C_0(\mathbb{R})$, this is known as the *Riemann-Lebesgue Lemma.* To prove it we first need the following result, which we will apply often (recall the notation $f_y(x) := f(x-y)$):

Lemma 25.4 (Translation is continuous on L^p). *If $1 \le p < \infty$ and $f \in L^p(\mathbb{R})$, then $f_y \to f$ in L^p as $y \to 0$ in $\mathbb{R}$.*

Proof. First observe, if $f \in L^p$, the $\|f_y\|_p = \|f\|_p$ by translation invariance of Lebesgue measure. Next, if g is continuous with compact support, then $g \in L^p$ and is uniformly continuous from which it readily follows that $\lim_{y\to 0}\|g-g_y\|_p = 0$. Continuous functions with compact support are dense in L^p. Thus, given $f \in L^p$ and $\epsilon > 0$ there is a continuous function g with compact support such that $\|f-g\|_p < \epsilon$. Hence,

$$\|f-f_y\|_p \le \|f-g\|_p + \|g-g_y\|_p + \|(g-f)_y\|_p < 2\epsilon + \|g-g_y\|_p$$

and the result follows. □

Note that the translation invariance of Lebesgue measure shows that the above proposition implies a more general version of itself: if $y_n \to y$ in $\mathbb{R}$, then $f_{y_n} \to f_y$ in L^p.

Lemma 25.5 (The Riemann-Lebesgue Lemma). *If $f \in L^1(\mathbb{R})$, then $\widehat{f} \in C_0(\mathbb{R})$.*

Proof. The proof is accomplished using the continuity of translation in L^1 (Lemma 25.4), and a simple trick: first, since $e^{-\pi i} = -1$, we can write

$$\widehat{f}(t) = -\int_{-\infty}^{\infty} f(x)e^{-2\pi it(x+(1/2t))}\,dx = -\int_{-\infty}^{\infty} f\left(x-\frac{1}{2t}\right)e^{-2\pi ixt}\,dx.$$

Combining this identity with the usual definition of $\widehat{f}$, we have

$$\widehat{f}(t) = \frac{1}{2}\int_{-\infty}^{\infty}\left(f(x) - f\left(x-\frac{1}{2t}\right)\right)e^{-2\pi ixt}\,dx$$

so

$$|\widehat{f}(t)| \le \frac{1}{2}\|f - f_{\frac{1}{2t}}\|_1.$$

But by Lemma 25.4, we have $\|f-f_{\frac{1}{2t}}\|_1 \to 0$ as $t \to \pm\infty$. □

Continuing our catalog of basic properties, we see that the Fourier transform also interacts nicely with differentiation.

Proposition 25.6 (Multiplication becomes differentiation)**.** *Suppose $f \in L^1(\mathbb{R})$. If $g(x) := xf(x)$ belongs to L^1, then $\widehat{f}$ is differentiable for all $t \in \mathbb{R}$, and $\widehat{g}(t) = \frac{-1}{2\pi i}\frac{d}{dt}\widehat{f}(t)$.*

The proof uses the standard estimate,

$$|1 - e^{it}| \leq |t|$$

for t real and dominated convergence.

Proof. Let $s \neq t$ be real numbers; from the definition of $\widehat{f}$ we have

$$\frac{\widehat{f}(s) - \widehat{f}(t)}{s - t} = \int_{-\infty}^{\infty} \frac{e^{-2\pi isx} - e^{-2\pi itx}}{s - t} f(x)\, dx. \tag{53}$$

Now the estimate

$$\left| \frac{e^{-2\pi isx} - e^{-2\pi itx}}{s - t} \right| \leq 2\pi|x|$$

holds for all $s \neq t$, so by the assumption $xf(x) \in L^1$ we can apply dominated convergence in (53) to take the limit as $s \to t$ to obtain

$$\begin{aligned} \lim_{s\to t} \frac{\widehat{f}(s) - \widehat{f}(t)}{s - t} &= \lim_{s\to t} \int_{-\infty}^{\infty} \frac{e^{-2\pi isx} - e^{-2\pi itx}}{s - t} f(x)\, dx \\ &= \int_{\infty}^{\infty} (-2\pi i)e^{-2\pi itx} xf(x)\, dx \\ &= -2\pi i\widehat{g}(t). \end{aligned}$$

Thus $\widehat{f}$ is differentiable and the claimed formula holds. □

Note that if $f \in L^1$ and also $g(x) := x^n f(x) \in L^1$ for some integer $n \geq 1$, then $x^k f(x)$ belongs to L^1 for all $0 \leq k \leq n$. The previous proposition can then be applied inductively to conclude.

Corollary 25.7. *If $f \in L^1$ and $g := x^n f \in L^1$, then $\widehat{f}$ is n times differentiable, and*

$$\widehat{x^k f} = \left(\frac{-1}{2\pi i}\right)^k \widehat{f}^{(k)} \quad \textit{for each } 0 \leq k \leq n.$$

One would also expect a theorem in the opposite direction: the Fourier transform should convert differentiation to multiplication by the independent variable. Under reasonable hypotheses, this is the case.

Proposition 25.8. *If $f \in C_0(\mathbb{R})$ and f' is continuous and in L^1, then*

$$\widehat{f'}(t) = 2\pi it\hat{f}(t).$$

Proof. Compute

$$\begin{aligned}
\widehat{f'}(t) &= \int_{-\infty}^{\infty} f'(x)e^{-2\pi itx}\,dx && (54)\\
&= \lim_{b\to\infty}\int_{-b}^{b} f'(x)e^{-2\pi itx}\,dx && (55)\\
&= \lim_{b\to\infty}\left([f(b)e^{-2\pi ibt} - f(-b)e^{2\pi ibt}] + 2\pi it\int_{-\infty}^{\infty} f(x)e^{-2\pi ixt}\,dx\right) && (56)\\
&= 2\pi it\widehat{f}(t), && (57)
\end{aligned}$$

where the second equality follows from the Dominated Convergence Theorem, the third using integration by parts, and the fourth from the $C_0(\mathbb{R})$ assumption on f. □

The last set of basic properties of the Fourier transform concern its interaction with convolution, which we now introduce. If f, g are measurable functions on $\mathbb{R}$, the *convolution* of f and g is the function

$$(f * g)(x) := \int_{-\infty}^{\infty} f(x-y)g(y)\,dy \tag{58}$$

defined at each x for which the integral makes sense. In particular, if $f \in L^\infty$ and $g \in L^1$, then $f * g$ is defined on all of $\mathbb{R}$. Observe, using the invariance of Lebesgue measure with respect to $x \to -x$ and a simple change of variable,

$$f * g(x) = \int_{-\infty}^{\infty} g(x-y)f(y)\,dy = g * f(x). \tag{59}$$

The next most basic fact about convolution is the following. (See Corollary 24.26.)

Proposition 25.9. *If $f, g \in L^1(\mathbb{R})$ then $f * g$ is defined for almost every $x \in \mathbb{R}$, $f * g$ is measurable, and $f * g \in L^1(\mathbb{R})$ and $\|f * g\|_1 \le \|f\|_1\,\|g\|_1$.*

Proof. Let $H(x, y) = f(x-y)g(y)$. One may check (exercise) that H is jointly measurable as a function of x and y. By Tonelli

$$\begin{aligned}
\iint |H(x,y)|\,dxdy &= \int_{-\infty}^{\infty} |g(y)|\left(\int_{-\infty}^{\infty} |f(x-y)|\,dx\right)\,dy\\
&= \|f\|_1 \int_{-\infty}^{\infty} |g(y)|\,dy\\
&= \|f\|_1\|g\|_1
\end{aligned}$$

where we have used the translation invariance of Lebesgue measure in the second equality. Hence H is in $L^1(\mathbb{R}^2)$. Thus by Fubini, $\int_{-\infty}^{\infty} |f(x-y)g(y)|\,dy = \int_{-\infty}^{\infty} |H(x,y)|\,dy$ is finite for almost every $x \in \mathbb{R}$, so $f * g$ is defined almost everywhere. Now by Fubini again

$$\int_{-\infty}^{\infty}\left|\int_{-\infty}^{\infty} f(x-y)g(y)\,dy\right|\,dx \le \iint |H(x,y)|\,dydx = \|f\|_1\|g\|_1.$$

□

The following basic properties of convolution are immediate, the proof is left as an exercise.

Proposition 25.10. *Let $f, g, h \in L^1(\mathbb{R})$.*

a) *(Commutativity)* $f * g = g * f$.
b) *(Associativity)* $(f * g) * h = f * (g * h)$.
c) *(Distributivity)* $(f + g) * h = f * g + f * h$.
d) *(Scalar multiplication) If $c \in \mathbb{C}$, then* $(cf) * g = c(f * g)$.

Notice that these properties together say that, if we equip $L^1(\mathbb{R})$ with the usual addition of functions and treat convolution as multiplication, then $L^1(\mathbb{R})$ becomes a commutative ring. (In fact it has even more structure, that of a *Banach algebra*, but we will not pursue this direction in this course). We can now describe how convolution behaves under the Fourier transform.

Proposition 25.11 (Convolution becomes multiplication)**.** *Let $f, g \in L^1(\mathbb{R})$. Then* $\widehat{f * g}(t) = \widehat{f}(t)\widehat{g}(t)$.

Proof. By virtue of Proposition 25.9 the function $G(x, y) = f(x - y)g(y)e^{-2\pi xt}$ is in $L^1(\mathbb{R}^2)$ (the product of Lebesgue measure with itself) and thus we can use Fubini's theorem to compute $\widehat{f * g}(t)$:

$$\begin{aligned}
\widehat{f * g}(t) &= \int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} f(x - y)g(y)\, dy \right) e^{-2\pi i x t}\, dx \\
&= \int_{-\infty}^{\infty} g(y) \left(\int_{-\infty}^{\infty} f(x - y)e^{-2\pi i x t}\, dx \right) dy \\
&= \int_{-\infty}^{\infty} \widehat{f}(t)e^{-2\pi i y t} g(y)\, dy \\
&= \widehat{f}(t)\widehat{g}(t)
\end{aligned}$$

where we have used Proposition 25.2(b). □

Given what we have proved so far, it follows that the Fourier transform is a ring homomorphism from $L^1(\mathbb{R})$ (with addition and convolution) to $C_0(\mathbb{R})$ (with pointwise addition and multiplication). We will see later that the Fourier transform is injective. It turns out that it is not surjective, however.

Let us finish this section by computing an important example. Let $a > 0$ and consider the *Gaussian*

$$g(x) := e^{-\pi a x^2}.$$

(The factor of π will be convenient given our choice of normalization in the definition of the Fourier transform.) It should be clear that $x^n g(x) \in L^1$ for all $a > 0$ and $n \geq 0$.

Lemma 25.12. $\widehat{g}(t) = \frac{1}{\sqrt{a}} e^{-\pi t^2/a}$.

Proof. Rather than computing the integral directly, we exploit Propositions 25.6 and 25.8. We may also assume $a = 1$ since the general case follows from this by scaling (Proposition 25.2(e)). Let $h = xg$ and note that $g' = -2\pi h$. Since $h \in L^1$,

$$\begin{aligned}(\widehat{g})'(t) &= -2\pi i\hat{h}(t)\\ &= -2\pi i\widehat{(-\frac{1}{2\pi}g)'}\\ &= i2\pi it\hat{g}(t)\\ &= -2\pi t\hat{g}(t).\end{aligned}$$

It follows from this computation and the product rule that

$$\frac{d}{dt}(e^{\pi t^2}\widehat{g}(t)) = 0,$$

so the function $e^{\pi t^2}\widehat{g}(t)$ is constant. To evaluate the constant, we set $t = 0$ and use the well-known Gaussian integral

$$\widehat{g}(0) = \int_{-\infty}^{\infty} e^{-\pi x^2}\, dx = 1.$$

□

25.1. **Digression - convolution and approximate units.** In this section we pause to develop further properties of convolutions; these will be necessary in order to study the Fourier inversion problem later.

A general principle is that the convolution of two functions inherits the best properties of both. We will see a number of instances of this phenomenon. A simple expression of this principle is the following:

Proposition 25.13. *If $f \in L^\infty$ and $g \in L^1$, then then $f * g$ is bounded and continuous.*

Proof. Since f is bounded,

$$|(f * g)(x)| \leq \int_{-\infty}^{\infty} |f(x-y)||g(y)|\, dy \leq \|f\|_\infty \|g\|_1.$$

To see that $f * g$ is continuous, fix $x \in \mathbb{R}$ and let $x_n \to x$. Let $h_n(y) = [g(x_n - y) - g(x-y)]\, f(y)$. Since $g \in L^1$ and $f \in L^\infty$, each $h_n \in L^1$ and

$$|g * f(x_n) - g * f(x)| \leq \|h_n\|_1 \leq \|g_{x_n} - g_x\|_1\, \|f\|_\infty,$$

where $g_x(y) = g(y-x)$. The result now follow from continuity of translation in L^1 (Lemma 25.4) (and commutativity of the convolution). □

Proposition 25.14. *If f, g are L^1 functions and are both compactly supported, then so if $f * g$.*

Proof. Problem 25.3.

□

The dominated convergence argument can be extended to apply to differentiability:

Proposition 25.15. *If f is a compactly supported C^k function ($1 \le k \le \infty$) and $g \in L^1$, then $f * g \in C^k$.*

Proof. First assume $k = 1$. Then one can differentiate under the integral sign as in the proof of Proposition 25.6 using *uniform differentiability of f*. The general case is proved by induction; the details are left as an exercise. □

Definition 25.16. An L^1 *approximate unit* is a collection of functions $\phi_\lambda \in L^1(\mathbb{R})$ indexed by $\lambda > 0$ such that:

a) $\phi_\lambda(t) \ge 0$ almost everywhere, for each λ,
b) $\int_{-\infty}^{\infty} \phi_\lambda(t)\,dt = 1$ for all λ, and
c) For each fixed $\delta > 0$, we have $\|\mathbf{1}_{|t|>\delta}\phi_\lambda\|_1 \to 0$ as $\lambda \to 0$.

◁

Approximate units are easy to construct: Let ϕ be any nonnegative, L^1 function with $\int_{-\infty}^{\infty} \phi(x)\,dx = 1$; then the functions $\phi_\lambda(x) := \frac{1}{\lambda}\phi\left(\frac{x}{\lambda}\right)$ form an L^1 approximate unit (Problem 25.5(a)). The simplest example comes from taking $\phi(x) = \mathbf{1}_{[-1/2,1/2]}$; the resulting ϕ_λ is known as the *box kernel*. (Draw a few of these for different values of λ to see what is going on.) Of course approximate units need not be compactly supported; we will see a very important example later (the Poisson kernel). The significance of approximate units (and their name) is explained by the following theorem.

Theorem 25.17. *Let ϕ_λ be an L^1 approximate identity. If $1 \le p < \infty$ and $f \in L^p(\mathbb{R})$, then for each λ the convolution $f * \phi_\lambda$ is defined almost everywhere, $f * \phi_\lambda \in L^p(\mathbb{R})$ and $\|\phi_\lambda * f - f\|_p \to 0$ as $\lambda \to 0$.*

To prove the theorem we need two facts. The first is Minkowski's integral inequality (Theorem 24.24). The second is the following lemma (which is really the heart of the matter); note where each of the properties the approximate unit is used.

Lemma 25.18. *Suppose ϕ_λ is an L^1 approximate unit and $g \in L^\infty$. If g is continuous at a point $x \in \mathbb{R}$, then*

$$\lim_{\lambda \to 0} (g * \phi_\lambda)(x) = g(x).$$

Proof. Using the trick $g(x) = \int_{-\infty}^{\infty} \phi_\lambda(y) g(x)\,dy$,

$$(g * \phi_\lambda)(x) - g(x) = \int_{-\infty}^{\infty} (g(x-y) - g(x))\,\phi_\lambda(y)\,dy,$$

so using the positivity of ϕ_λ

$$|(g * \phi_\lambda)(x) - g(x)| \le \int_{-\infty}^{\infty} |g(x-y) - g(x)|\phi_\lambda(y)\,dy \tag{60}$$

To estimate the right-hand side, let $\epsilon > 0$ be given. By the continuity of g at x, choose $\delta > 0$ so that $|g(x-y) - g(x)| < \epsilon$ when $|y| < \delta$. We then split the integral in (60) into two integrals, over the regions $|y| \le \delta$ and $|y| > \delta$:

$$\int_{-\infty}^{\infty} |g(x-y)-g(x)|\phi_\lambda(y)\,dy = \int_{\{|y|\le\delta\}} |g(x-y)-g(x)|\phi_\lambda(y)\,dy + \int_{\{|y|>\delta\}} |g(x-y)-g(x)|\phi_\lambda(y)\,dy.$$

The first integrand is bounded by $\epsilon\phi_\lambda$, so

$$\int_{\{|y|\le\delta\}} |g(x-y) - g(x)|\phi_\lambda(y)\,dy \le \epsilon \int_{\{|y|\le\delta\}} \phi_\lambda(y)\,dy \le \epsilon$$

since $\int_{-\infty}^{\infty} \phi_\lambda(y)\,dy = 1$. The second integrand is bounded by $2\|g\|_\infty \mathbf{1}_{\{|y|>\delta\}}\phi_\lambda(y)$, so goes to 0 as $\lambda \to 0$ by property (c) in the definition of approximate unit. This proves the lemma. □

Remark: If one assumes that the approximate unit ϕ_λ was constructed as $\phi_\lambda(y) = \frac{1}{\lambda}\phi\left(\frac{y}{\lambda}\right)$ for some ϕ satisfying $\phi \ge 0$ and $\int \phi = 1$, then the lemma has an easier proof (see Problem 25.5(b)).

Proof of Theorem 25.17. Let $f \in L^p$ be given. By Corollary 24.26 (or a routine argument), $f * \phi_\lambda$ is defined for almost every x and is in L^p. Let $d\mu = dx$ denote Lebesgue measure and $d\nu$ the measure $\phi_\lambda(y)dy$. Define $F(x,y) = (f_y - f)(x) = f(x-y) - f(x)$. In particular, $F(\cdot, y) \in L^p(\mu)$ and

$$\|F(\cdot, y)\|_p \le 2\|f\|_p$$

for each x. Since ν is a finite measure, $H(y) = \|F(\cdot, y)\|_p$ is in $L^1(\mu)$.

It follows from Theorem 24.24 that for ν-almost every x the function $G_x(y) = F(x,y) = f(x-y) - f(x)$ is in $L^1(\nu)$,

$$g(x) = \int_{-\infty}^{\infty} F(x,y)d\nu(y) = \int_{-\infty}^{\infty} F(x,y)\phi_\lambda(y)\,dy = (f_y * \phi_\lambda - f)(x)$$

is in $L^p(\mu)$ with

$$\|g\|_p \le \int_{-\infty}^{\infty} \|F(\cdot, y)\|_p\,d\nu = \int_{-\infty}^{\infty} \|f_y - f\|_p \phi_\lambda(y)\,dy.$$

Thus,

$$\|f * \phi_\lambda - f\|_p \le h * \phi_\lambda(0),$$

where $h(y) = \|f_{-y} - f\|_p$. The function h belongs to L^∞ (indeed $\|h\|_\infty \le 2\|f\|_p$) and is continuous (by the continuity of translation in L^p) and $h(0) = 0$. Thus by Lemma 25.18, the right hand side goes to 0 as $\lambda \to 0$, which proves the theorem. □

It is often useful to have approximate units with additional properties, such as smoothness or compact support. In fact it is possible to construct an L^1 approximate unit $\{\psi_\lambda\}$ consisting of C^∞ functions with compact support. This is accomplished via *bump functions*:

Definition 25.19. A *bump function* is a function $\psi : \mathbb{R} \to \mathbb{R}$ such that:

a) $\psi \in C^\infty(\mathbb{R})$,
b) ψ is compactly supported,
c) $\psi \geq 0$, and
d) $\int_{-\infty}^{\infty} \psi(x)\, dx = 1$.

◁

Lemma 25.20. *Bump functions exist.*

Proof. The main issue is to construct a C^∞ function with compact support. Consider the function

$$h(x) = \begin{cases} e^{-1/x} & \text{if } x > 0 \\ 0 & \text{if } x \leq 0 \end{cases}$$

Clearly $h \geq 0$ and h is differentiable for all $x \neq 0$; it is a calculus exercise to verify that h is infinitely differentiable at 0 and $h^{(n)} = 0$ for all n. It is then straightforward to verify that $\psi(x) = c \cdot h(x+1)h(1-x)$ is a bump function (for a suitable normalizing constant c), supported on $[-1, 1]$. □

Note that if ψ is a bump function supported in $[-1, 1]$, then the functions $\psi_\lambda(x) := \frac{1}{\lambda}\psi\left(\frac{x}{\lambda}\right)$ are also bump functions, supported in $[-\lambda, \lambda]$. (Draw a picture of what these functions look like as $\lambda \to 0$). Thus there exist approximate units consisting of smooth, compactly supported functions. As an application, we can prove that $C_c^\infty(\mathbb{R})$ is dense in $L^p(\mathbb{R})$ for all $1 \leq p < \infty$. A proof is outlined in Problem 25.6.

25.2. **Inversion and uniqueness.** In this section we study the problem of recovering f from $\widehat{f}$. Loosely, the Fourier transform can be thought of as a resolution of f as a superposition of sinusoidal functions $e^{2\pi itx}$; the value of $\widehat{f}(t)$ measures the "amplitude" of f in the "frequency" t. This suggests that a formula like

$$f(x) = \int_{-\infty}^{\infty} \widehat{f}(t) e^{2\pi itx}\, dt \tag{61}$$

ought to hold, at least if $\widehat{f} \in L^1$. If we formally substitute the definition of $\widehat{f}$ and switch the order of integration, we are confronted with

$$\int_{-\infty}^{\infty} f(u) \left(\int_{-\infty}^{\infty} e^{2\pi i(x-u)t}\, dt \right) du$$

and the inner integral is not convergent, regardless of any assumption on $\widehat{f}$. In fact (61) *does* hold when $\widehat{f} \in L^1$, but a more delicate argument is necessary. So, the goal of this section will be to prove:

Theorem 25.21 (Fourier inversion, L^1 case)**.** *If f and $\widehat{f}$ belong to L^1, then*

$$f(x) = \int_{-\infty}^{\infty} \widehat{f}(t) e^{2\pi ixt}\, dt \tag{62}$$

for almost every $x \in \mathbb{R}$.

Once we have the inversion formula, we see that L^1 functions are determined by their Fourier transforms:

Corollary 25.22. *Suppose $f, g \in L^1$. If $\widehat{f} = \widehat{g}$, then $f = g$ a.e.*

Proof. From the inversion theorem, if $f \in L^1$ and $\widehat{f} = 0$, then $f = 0$. By the linearity of the Fourier transform, $\widehat{f - g} = \widehat{f} - \widehat{g}$, and the corollary follows. □

So, in principle, f is fully determined by $\widehat{f}$, even if $\widehat{f} \notin L^1$; and this is often the case. For example, suppose $f = \mathbf{1}_{[0,1]}$. Then

$$\widehat{\mathbf{1}_{[0,1]}}(t) = \int_0^1 e^{-2\pi ixt}\, dx = \frac{1 - e^{-2\pi it}}{2\pi it}$$

which does not belong to L^1 (check this). To recover f from $\widehat{f}$ in these cases, we turn to *summability methods*; in fact summability methods will already be of use in proving the inversion theorem. The idea is this: suppose we have a divergent integral

$$\int_{-\infty}^{\infty} h(t)\, dt$$

where the function h is, say, locally L^1, but not L^1. We might try to make sense of the integral as

$$\lim_{a \to +\infty} \int_{-a}^{a} h(t)\, dt,$$

effectively we have introduced the *cutoff function* $\psi_a(t) := \mathbf{1}_{[-a,a]}$, which is positive, integrable, and increases to 1 pointwise as $a \to \infty$. Given any family of functions ψ_a with these three propertes, we can consider the integrals

$$\int_{-\infty}^{\infty} h(t)\psi_a(t)\, dt.$$

It will turn out that the "square" cutoff $\mathbf{1}_{[-a,a]}$ has some undesirable properties; for example its Fourier transform is not L^1 (and not of constant sign). We will work first with smoother cutoff functions, in particular the functions $t \to \exp(-a|t|)$ (here we consider $a \to 0$ rather than $a \to \infty$, but this is not important).

The first step is to compute the (inverse) Fourier transform of

$$Q_a(t) = e^{-2a\pi|t|} \tag{63}$$

(the extra factor of 2π turns out to be a convenient normalization).

Lemma 25.23. *For all $a > 0$,*

$$\int_{-\infty}^{\infty} Q_a(t) e^{2\pi itx}\, dt = \frac{1}{\pi}\frac{a}{a^2 + x^2}.$$

Proof. Problem 25.7. □

Let us fix the notation

$$P_a(x) := \frac{1}{\pi}\frac{a}{a^2+x^2} \tag{64}$$

Notice that $P_1(x)$ is nonnegative and $\int_{-\infty}^{\infty} P_1(x)\,dx = 1$. Moreover, $P_a(x) = \frac{1}{a}P_1(\frac{x}{a})$. By the remarks following Definition 25.16, we have:

Lemma 25.24. *$\{P_a\}_{a>0}$ is an L^1 approximate unit.*

The function $P_a(x)$ (viewed as a function of the two arguments (a and x) is known as the *Poisson kernel*. We are now able to compute the integral (61) modified by the cutoff function $Q_a(t)$:

Proposition 25.25. *If $f \in L^1$, then for all $a > 0$ and all $x \in \mathbb{R}$*

$$(f * P_a)(x) = \int_{-\infty}^{\infty} Q_a(t)\widehat{f}(t)e^{2\pi itx}\,dt.$$

Proof. Because we have introduced the cutoff function, Fubini's theorem can be applied:

$$\begin{aligned}
\int_{-\infty}^{\infty} Q_a(t)\widehat{f}(t)e^{2\pi ix}\,dt &= \int_{-\infty}^{\infty} Q_a(t)\int_{-\infty}^{\infty} f(y)e^{2\pi i(x-y)t}\,dy\,dt && (65)\\
&= \int_{-\infty}^{\infty} Q_a(t)\int_{-\infty}^{\infty} f(x-y)e^{2\pi iyt}\,dy\,dt && (66)\\
&= \int_{-\infty}^{\infty} f(x-y)\int_{-\infty}^{\infty} Q_a(t)e^{2\pi iyt}\,dt\,dy && (67)\\
&= (f * P_a)(x). && (68)
\end{aligned}$$

□

Proof of Theorem 25.21. Assume $f, \widehat{f} \in L^1(\mathbb{R})$. Define

$$g(x) = \int_{-\infty}^{\infty} \widehat{f}(t)e^{2\pi itx}\,dt.$$

By the Riemann-Lebesgue lemma (Lemma 25.5), $g \in C_0(\mathbb{R})$. We want to show $g = f$ a.e. From Proposition 25.25 we have for all $a > 0$

$$(f * P_a)(x) = \int_{-\infty}^{\infty} Q_a(t)\widehat{f}(t)e^{2\pi itx}\,dt. \tag{69}$$

Fix a sequence $a_n \to 0$. Using the hypothesis $\widehat{f} \in L^1$, an application of dominated convergence shows, for all x, the integral in the right-hand side of (69) converges to $g(x)$ as $a_n \to 0$. On the other hand, since P_a is an L^1 approximate unit, we know from Theorem 25.17 that $f * P_{a_n} \to f$ in L^1. Passing to a subsequence, we may assume that $f * P_{a_n} \to f$ almost everywhere, but then by (69) we have $f * P_{a_n} \to g$ a.e., so $f = g$ a.e. and the theorem is proved. □

Remark 25.26. Observe that the above proof did not really use the explicit form of P_a; rather the point was that $Q_a(t) = \{e^{-2a\pi|t|}\}_{a>0}$ was a cutoff function (uniformly bounded and converging pointwise to the constant function 1) whose Fourier transform

$\{P_a\}$ was an L^1 approximate unit. Any other cutoff function with this property could have been used. ◇

Another corollary of Proposition 25.25 is that we can recover f from $\widehat{f}$ in a weaker sense for any $f \in L^1$ (that is, not assuming $\widehat{f} \in L^1$). Indeed, combining Proposition 25.25 and Theorem 25.17 we have immediately:

Corollary 25.27 (Fourier inversion in the L^1 norm)**.** *If $f \in L^1$, then*

$$\int_{-\infty}^{\infty} Q_a(t)\widehat{f}(t)e^{2\pi ixt}\,dt \tag{70}$$

converges to f in the L^1 norm as $a \to 0+$.

So, we can recover f but only in the L^1 norm; the corollary does not say anything about the pointwise covergence of the reguralized integrals. In fact, it is true that the integrals (70) converge to f a.e., but this requires a more delicate argument.

25.3. **The L^2 theory.** In this section we study the Fourier transform on L^2. There is an immediate problem, of course, since by Problem 24.8 $L^2 \not\subset L^1$, so the integral (52) need not be defined. However, we can observe that $L^1 \cap L^2$ is dense in L^2 (why?), and start there.

Lemma 25.28. *If $f \in L^1 \cap L^2$, then $\widehat{f}$ belongs to L^2 and $\|\widehat{f}\|_2 = \|f\|_2$.*

Proof. Let $\widetilde{f}(x) := \overline{f(-x)}$ and define $g = f * \widetilde{f}$. Since $f, \widetilde{f} \in L^1$ we have $g \in L^1$ by Proposition 25.9. Now

$$g(x) = \int_{-\infty}^{\infty} f(x-y)\overline{f(-y)}\,dy = \int_{-\infty}^{\infty} f(x+y)\overline{f(y)}\,dy.$$

so we can write $g(x) = \langle f_{-x}, f\rangle_{L^2}$. By Lemma 25.4, the map $x \to f_{-x}$ is continuous from $\mathbb{R}$ into L^2 and of course the vector f determines a continuous linear functional. Thus g is a continuous function of x, and $g(0) = \|f\|_2^2$. By Cauchy-Schwarz again,

$$|g(x)| \le \|f_{-x}\|_2\,\|f\|_2 = \|f\|_2^2,$$

so g is bounded.

Let, as before $Q_a(t) = \exp(-2a\pi|t|)$. Since $g \in L^1$ we can apply Proposition 25.25 to compute

$$(g * P_a)(0) = \int_{-\infty}^{\infty} Q_a(t)\widehat{g}(t)\,dt.$$

As g is continuous, by Lemma 25.18

$$\|f\|_2^2 = g(0) = \lim_{a\to 0}(g * P_a)(0) = \lim_{a\to 0}\int_{-\infty}^{\infty} Q_a(t)\widehat{g}(t)\,dt. \tag{71}$$

Let us compute the limit of this last integral in a different way. Recall that by definition $g = f * \widetilde{f}$, so by Propositions 25.11 and 25.2(d),

$$\widehat{g}(t) = |\widehat{f}(t)|^2.$$

Making this substitution in the integral in (71) and applying the monotone convergence theorem, we have

$$\|f\|_2^2 = \lim_{a\to 0}\int_{-\infty}^{\infty} Q_a(t)\widehat{g}(t)\,dt = \lim_{a\to 0}\int_{-\infty}^{\infty} Q_a(t)|\widehat{f}(t)|^2\,dt = \|\widehat{f}\|_2^2.$$

which proves the lemma. □

Theorem 25.29 (The Fourier transform on L^2)**.** *There is a unique bounded linear transformation $\mathcal{F}: L^2 \to L^2$ satisfying the following conditions:*

a) *For all $f \in L^1 \cap L^2$, $\mathcal{F}f = \widehat{f}$.*
b) *(The Plancherel theorem) $\|\mathcal{F}f\|_2 = \|f\|_2$ for all $f \in L^2$.*
c) *The mapping $f \to \mathcal{F}f$ is an Hilbert space isomorphism of L^2 onto L^2.*
d) *(The Parseval identity) $\langle f, g\rangle = \langle \mathcal{F}f, \mathcal{F}g\rangle$ for all $f, g \in L^2$.*

Remark: Note that when $f, g \in L^1 \cap L^2$, the Parseval identity reads

$$\int_{-\infty}^{\infty} f(x)\overline{g(x)}\,dx = \int_{-\infty}^{\infty} \widehat{f}(t)\overline{\widehat{g}(t)}\,dt.$$

Proof of Theorem 25.29. By Lemma 25.28, the map $f \to \widehat{f}$ is bounded linear transformation from a dense subspace of L^2 into L^2. Thus, since the codomain L^2 is complete, by Proposition 20.9 the map $f \to \widehat{f}$ has a unique bounded linear extension to a map $\mathcal{F}: L^2 \to L^2$. This proves (a), and (b) follows since $\|f\|_2 = \|\mathcal{F}f\|_2$ on a dense set (namely $L^1 \cap L^2$). (d) is also an immediate consequence of (b), by Problem 23.3(a). It remains to prove (c); what we must show is that $\mathcal{F}$ is onto.

We show that $\mathcal{F}$ has dense range; combined with the fact that $\mathcal{F}$ is an isometry, it follws that $\mathcal{F}$ is in fact onto. (The proof of this last assertion is left as an exercise). Let M denote the set of all functions $g \in L^2$ such that $g = \widehat{f}$ for some $f \in L^1 \cap L^2$. Clearly the range of $\mathcal{F}$ contains M, so it will suffice to prove that M is dense, or equivalently, that $M^\perp = \{0\}$.

Recall the cutoff functions $Q_a(x) = e^{-2a\pi|x|}$, $a > 0$ introduced in equation (63). The functions $e^{2\pi ibx}e^{-2a\pi|x|}$ belong to $L^1 \cap L^2$ for all $a > 0$ and $b \in \mathbb{R}$, so their Fourier transforms

$$P_a(t-b) = \int_{-\infty}^{\infty} e^{2\pi ibx}\,Q_a(x)\,e^{-2\pi itx}\,dx$$

belong to M. So, let $h \in M^\perp$ be given and let $H(x) = h(-x)$. Thus,

$$(P_a * \overline{H})(-b) = \int_{-\infty}^{\infty} P_a(-b-t)\overline{h(-t)}\,dt = \int_{-\infty}^{\infty} P_a(t-b)\overline{h(t)}\,dt = 0$$

for all b, and therefore $H = 0$ by Lemma 25.24 and Theorem 25.17. Thus M is dense in L^2 and the proof is finished. □

Theorem 25.30 (L^2 inversion)**.** *Let $f \in L^2$. Define*

$$\phi_N(t) = \int_{-N}^{N} f(x)e^{-2\pi ixt}\,dx, \quad \psi_N(t) = \int_{-N}^{N} (\mathcal{F}f)(t)e^{2\pi ixt}\,dt.$$

Then $\|\phi_N - \mathcal{F}f\|_2 \to 0$ *and* $\|\psi_N(t) - f\|_2 \to 0$ *as* $N \to \infty$.

Proof. For $f \in L^2$, let $f_N := \mathbf{1}_{[-N,N]}f$. In particular, $f_N \in L^2$ and $\|f_N\| \le \|f\|$ (in L^2). Define $\mathcal{G}_N f = \widehat{f_N} = \phi_N$ and note that $\|\mathcal{G}_N f\| = \|f_N\|$ by Theorem 25.29 and thus G_N is a bounded linear map from L^2 to L^2. An application of dominated convergence shows that $f_N \to f$ in the L^2 norm, and in particular is a Cauchy sequence in L^2. Since the Fourier transform $\mathcal{F}$ is an isometry, it follows that $(\phi_N) = (\widehat{f_N})$ is a Cauchy sequence in L^2, and hence converges to some $\phi \in L^2$. Thus, for each $f \in L^2$, the limit $(\mathcal{G}_N f)_N$ exists. It follows from Problem 22.5, that $\mathcal{G}f = \lim \mathcal{G}_N f$ defines a bounded linear operator on L^2.

Now if $f \in L^1 \cap L^2$, then $\mathcal{G}_N f = \phi_N$ converges pointwise to $\mathcal{F}f$ by Proposition 25.3. Since also a subsequence of $\mathcal{G}_N f$ converges pointwise a.e. to $\mathcal{G}f$, we have $\mathcal{G}f = \mathcal{F}f$. Consequently, $\mathcal{G} = \mathcal{F}$ on the dense set $L^1 \cap L^2$ of L^2. As they are both bounded operators, they agree on all of L^2.

The statement for ψ_N is proved by similar methods and is left as an exercise (Problem 25.10). □

It is important to note that, for a general function $f \in L^2$, its Fourier transform is defined only as an element of L^2. In particular it is defined only a.e., and cannot be evaluated at points. From now on we just write $\widehat{f}$ for $\mathcal{F}f$ when $f \in L^2$, with the understanding that the integral definition is only valid when $f \in L^1 \cap L^2$.

25.4. **Fourier Series.** We now replace the group $(\mathbb{R}, +)$ with the multiplicative group of unimodular complex numbers

$$\mathbb{T} = \{e^{i\theta} : -\pi \le \theta < \pi\} \subset \mathbb{C}.$$

By the properties of the exponential, this group isomorphic to the group $[-\pi, \pi)$ with addition mod 2π. (This amounts to the isomorphism of the quotient group $\mathbb{R}/2\pi\mathbb{Z} \cong \mathbb{T}$.) We will typically use θ to identify elements of $\mathbb{T}$. We will treat functions $f : \mathbb{T} \to \mathbb{C}$ either as functions on the unit circle in $\mathbb{C}$, or, when convenient, as functions on the interval $[-\pi, \pi)$ extended 2π-peridoically to $\mathbb{R}$. For each integer n, the map

$$\chi_n : e^{i\theta} \to e^{in\theta}$$

is a homomorphism from $\mathbb{T}$ to $\mathbb{T}$. In fact, any continuous homomorphism $h : \mathbb{T} \to \mathbb{C}$ is χ_n for some $n \in \mathbb{Z}$.

We equip $\mathbb{T}$ with normalized arc length measure, or equivalently normalized Lebesgue measure on $[-\pi, \pi)$. We write dm for this measure. Note that $\mathbb{T}$ acts on itself by rotation. For fixed θ, the map $\tau_\theta : \mathbb{T} \to \mathbb{T}$ given by

$$\tau_\theta(e^{i\psi}) = e^{i(\psi+\theta)}$$

amounts to rotation of the circle through the angle θ. Moreover, m is invariant under τ_θ for each θ. Arguing exactly as on the line, one can prove the continuity of translation in $L^p(\mathbb{T})$. For fixed $\varphi \in [-\pi, \pi)$, let $f_\varphi(\theta) = f(\theta - \varphi)$.

Proposition 25.31. *If $1 \le p < \infty$ and $f \in L^p(\mathbb{T})$, then the mapping $\mathbb{R} \to L^p(\mathbb{T})$ defined by $\varphi \mapsto f_\varphi$ is continuous.*

Now that we have a translation invariant measure and the characters χ_n, we can define a Fourier transform.

Definition 25.32. Let $f \in L^1(\mathbb{T})$. The *Fourier coefficients* of f are the numbers

$$\widehat{f}(n) := \int_{\mathbb{T}} f(\theta) e^{-in\theta}\, dm(\theta) = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(x) e^{-inx}\, dx, \quad n \in \mathbb{Z}.$$

The *Fourier series* of f is the series

$$f \sim \sum_{n=-\infty}^{\infty} \widehat{f}(n) e^{in\theta}. \tag{72}$$

◁

The function $\widehat{f} : \mathbb{Z} \to \mathbb{C}$ is called the *Fourier transform* of f. The Fourier coefficients $\widehat{f}(n)$ are the analogs of the pointwise values of $\widehat{f}(t)$ in the case of $\mathbb{R}$. Thus, in each case the Fourier transform of a function f is a function defined on the (group of) characters. Similarly, the (possibly divergent) Fourier series is the analog of the (possibly divergent) integral

$$\int_{-\infty}^{\infty} \widehat{f}(t) e^{2\pi i t x}\, dt.$$

Just as in the case of the real line, the Fourier transform behaves predictably under translation, modulation, reflection, and scaling; we leave the statements and proofs of these facts as exercises. Moreover the same kind of integral estimates show that $\widehat{f}$ is always a bounded function.

Theorem 25.33 (Riemann-Lebesgue lemma on the circle)**.** *If $f \in L^1(\mathbb{T})$, then*

$$\lim_{n \to \pm\infty} |\widehat{f}(n)| = 0.$$

This version of the Riemann-Lebesgue Lemma follows readily from the version on $\mathbb{R}$ (Lemma 25.5). In other words, the Fourier transform takes $L^1(\mathbb{T})$ into $c_0(\mathbb{Z})$. As in the case of the line, this map turns out to be injective (which will follow from an appropriate inversion theorem), but not surjective. Also, as we found on the line, typically $\widehat{f} \notin \ell^1(\mathbb{Z})$. Thus, there is an immediate difficulty in interpreting the series (72)

As before, the basic problem is to recover f from $\widehat{f}$, which in turn means finding a way to attach meaning to the (in general divergent) Fourier series. Broadly, the method is the same as in the case of $\mathbb{R}$: we introduce a cutoff function whose Fourier series is nicely convergent to an approximate unit for $L^1(\mathbb{T})$. Before doing this we have a look at what can go wrong, even for nice f. Let us try to naively sum the Fourier series: fix $f \in L^1(\mathbb{T})$ and consider the partial sums

$$s_N(\theta) = \sum_{n=-N}^{N} \widehat{f}(n) e^{in\theta}.$$

Since this is a finite sum, we can expand $\widehat{f}$ as an integral and pull the sum inside:

$$s_N(\theta) = \int_{\mathbb{T}} f(\phi) \left\{ \sum_{n=-N}^{N} e^{-in\phi} e^{in\theta} \right\} dm(\phi). \tag{73}$$

Working with the inner sum, we consider the expression

$$D_N(t) := \sum_{n=-N}^{N} e^{int} = \frac{\sin\left(N+\frac{1}{2}\right)t}{\sin\frac{t}{2}}$$

Then (73) can be written

$$s_N(\theta) = \int_{\mathbb{T}} f(\phi) D_N(\theta-\phi)\, dm = (f * D_N)(\theta)$$

where we have introduced convolution on the circle group $\mathbb{T}$. (Note that the difference $\theta - \phi$ is interpreted in the group $\mathbb{T}$, that is, is carried out mod 2π.) Thus, the question of whether the partial sums s_N converge to f in some sense (pointwise a.e., or in L^1, etc.) reduces to the question of whether $f * D_N$ converges to f in the same sense. Unfortunately, since D_N is not an $L^1(\mathbb{T})$ approximate unit, the partial sums s_N can be badly behaved, even for nice f. For example we have the following:

Theorem 25.34. *There exists a continuous function f on $\mathbb{T}$ such that the Fourier series for f diverges at $\theta = 0$.*

Proof. We present an outline of the proof; the details are left as an exercise. As noted above, the N^{th} partial sum of the Fourier series of f at a point θ is given by $(f*D_N)(\theta)$. We suppose that $(f * D_N)(0) \to f(0)$ for every $f \in C(\mathbb{T})$ and derive a contradiction. Now

$$s_N(0) = (f * D_N)(0) = \int_{\mathbb{T}} f(\phi) D_N(\phi)\, dm(\phi).$$

By the construction of D_N, it is clear that $D_N \in L^1(\mathbb{T})$ for each N. Thus for each N the map $L_N : f \to \int_{\mathbb{T}} f D_N\, dm$ is a bounded linear functional on $C(\mathbb{T})$, and one can show that the norm of this functional is equal to $\|D_N\|_1$. (To see this, find a sequence of continuous functions g_n such that $\|g_n\|_\infty \leq 1$ for all n and $g_n \to \mathrm{sgn} D_N$ pointwise. Then $|L_N(g_n)| \to \|D_N\|_1$.) Next, one can show by direct estimates of the integral that $\|D_N\|_1 \to \infty$ as $N \to \infty$. The proof finishes by appeal to the Principle of Uniform Boundedness: if it were the case that $L_N(f) = s_N(0) \to f(0)$ for all $f \in C(\mathbb{T})$, then the family of linear functionals L_N would be pointwise bounded on $C(\mathbb{T})$, hence uniformly bounded, which is a contradiction. Problem 25.17 gives some hints on filling the details. □

Before going further, let us observe that if we assume f has a certain amount of smoothness at a point, then the Fourier series for f will converge to f at that point. A simple result of this type is the following:

Proposition 25.35. *If $f \in L^1(\mathbb{T})$ and f is differentiable at a point θ_0, then $s_N(\theta_0) \to f(\theta_0)$.*

Proof. By considering real and imaginary parts, we may assume f is real-valued, and by replacing $f(\theta)$ by $f(\theta+\theta_0)-f(\theta_0)$ we may assume that $\theta_0=0$ and $f(0)=0$. As we have already observed, we have

$$s_N(0) = (f * D_n)(0) = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(\phi)D_N(\phi)\,d\phi \tag{74}$$

$$= \frac{1}{2\pi}\int_{-\pi}^{\pi} f(\phi)\frac{\sin\left(N+\frac{1}{2}\right)\phi}{\sin\frac{\phi}{2}}\,d\phi, \tag{75}$$

and we wish to prove $s_N(0)\to 0$ as $N\to\infty$. The key observation is that the function

$$g(\phi) = \frac{f(\phi)}{\sin\frac{\phi}{2}}$$

belongs to $L^1(\mathbb{T})$. To see this, first note the elementary estimate

$$\frac{|\phi|}{\pi} \leq |\sin\frac{\phi}{2}| \leq \frac{|\phi|}{2}$$

for $|\phi|\leq\frac{\pi}{2}$.

Now, since f is differentiable at 0 and $f(0)=0$, there exist $M>0$ and $0<\delta<\frac{\pi}{2}$ such that

$$\sup_{|\phi|<\delta}\left|\frac{f(\phi)}{\phi}\right| \leq M,$$

so

$$|g(\phi)| = \left|\frac{f(\phi)}{\phi}\right|\left|\frac{\phi}{\sin\frac{\phi}{2}}\right| \leq \pi M.$$

for $|\phi|\leq\delta$. On the other hand, for $\delta<|\phi|\leq\pi$,

$$\left|\frac{f(\phi)}{\sin\frac{\phi}{2}}\right| \leq \frac{\pi}{\delta}|f(\phi)|.$$

Thus, g is bounded near 0 and dominated by f away from 0, hence $g\in L^1$. Returning to (74), we have

$$s_N(0) = \frac{1}{2\pi}\int_{-\pi}^{\pi} g(\phi)\sin\left(N+\frac{1}{2}\right)\phi\,d\phi$$

Making the change of variable $\theta=\phi/2$ this becomes

$$s_N(0) = \frac{1}{\pi}\int_{-\pi/2}^{\pi/2} g(2\theta)\sin((2N+1)\theta)\,d\theta \tag{76}$$

If we now put

$$h(\theta) = 2\mathbf{1}_{[-\pi/2,\pi/2]}(\theta)g(2\theta),$$

then $h\in L^1$ and the integral in (76) is nothing but the imaginary part of $\widehat{h}(2N+1)$, which goes to 0 as $N\to\infty$ by the Riemann-Lebesgue lemma. This finishes the proof. □

So, to recover f from its Fourier series, as before we need to introduce a cutoff function, but since the "square" cutoff $\mathbf{1}_{[-N,N]}$ (corresponding to ordinary partial sums) is badly behaved, we choose a smoother cutoff. It turns out that the functions

$$n \to r^{|n|}$$

for $0 \leq r < 1$ are a good choice (analgous to $e^{-a\pi|t|}$ on the line). Thus we consider the *Abel means* of the Fourier series

$$A(r,\theta) := \sum_{n=-\infty}^{\infty} \widehat{f}(n) r^{|n|} e^{in\theta}.$$

Since $\widehat{f}$ is bounded and $r < 1$, this series is absolutely convergent for all θ, and unifomrly convergent on $\mathbb{T}$ for each fixed r. Thus, we can again expand $\widehat{f}$ as an integral, and interchange the sum and integral:

$$A(r,\theta) = \int_{\mathbb{T}} f(\phi) \left\{ \sum_{n=-\infty}^{\infty} r^{|n|} e^{in(\theta-\phi)} \right\} dm(\phi) := (f * P_r)(\theta)$$

where, by summing the geometric series, $P_r(\theta)$ is given by

$$P_r(\theta) := \sum_{n=-\infty}^{\infty} r^{|n|} e^{in\theta} = \frac{1-r^2}{1-2r\cos\theta + r^2}.$$

The function $P_r(\theta)$ is the *Poisson kernel.*

Lemma 25.36. *The family of function P_r has the following properties.*

i) *$P_r(\theta) \geq 0$ for all $\theta \in [-\pi, \pi]$ and all $0 \leq r < 1$,*
ii) *For each r, $\int_{\mathbb{T}} P_r(\theta)\, dm(\theta) = 1$,*
iii) *For each fixed $0 < \delta < \pi$,*

$$\frac{1}{2\pi} \int_{\delta \leq |\theta| \leq \pi} P_r(\theta)\, d\theta \to 0 \quad \text{as } r \to 1.$$

Proof. Exercise. □

In other words, $\{P_r\}_{r<1}$ is an $L^1(\mathbb{T})$ approximate unit. Just as on the line, we have

Theorem 25.37. *If $1 \leq p < \infty$ and $f \in L^p(\mathbb{T})$, then $\|f * P_r - f\|_p \to 0$ as $r \to 1$.*

Proof. This uses the properties of approximate units in the same way as on $\mathbb{R}$; the proof is an exercise. □

With these results in hand, we can obtain Abel summability of Fourier series on the circle:

Corollary 25.38. *If $f \in L^p(\mathbb{T})$, $1 \leq p < \infty$, then the Abel means $A(r,\theta)$ of the Fourier series for f converge to f in L^p as $r \to 1$.*

What happens when $p = \infty$? We have already seen that the Fourier series of a continuous function can diverge at a given point; however if we use the Abel means $A(r, \theta)$ we can do better. The reason is the following lemma, which says that the Poisson kernel $P_r(\theta)$ obeys a stronger condition than that of Lemma 25.36:

Lemma 25.39. *For each $0 < \delta < \pi$, we have $P_r(\theta) \to 0$ uniformly on $\delta \leq |\theta| \leq \pi$ as $r \to 1$.*

Proof. Fix δ. For $\delta < |\theta| \leq \pi$, we have $-1 \leq cos\theta \leq cos\delta < 1$. Thus for such θ

$$P_r(\theta) = \frac{1-r^2}{1-2r\cos\theta + r^2} \leq \frac{1-r^2}{1-2r\cos\delta + r^2}.$$

As $r \to 1$, the numerator of this last expression goes to 0 while the denominator is bounded away from 0, which proves the lemma. □

Theorem 25.40. *If $f \in C(\mathbb{T})$, then $f * P_r = A(r, \theta) \to f(\theta)$ uniformly as $r \to 1$.*

Proof. Fix $f \in C(\mathbb{T})$ and $\epsilon > 0$. Since f is continuous and $\mathbb{T}$ is compact, f is uniformly continuous, so there exists $\delta > 0$ such that $|f(\theta) - f(\phi)| < \epsilon$ whenever $|\theta - \phi| < \delta$. Using our usual tricks with approximate units we write $f(\theta) = \int f(\theta)P_r(\phi)d\phi$ to obtain

$$|(f * P_r)(\theta) - f(\theta)| \leq \frac{1}{2\pi}\int_{-\pi}^{\pi} |f(\theta - \phi) - f(\theta)|P_r(\phi)\, d\phi.$$

We split the integral as $\int_{|\phi|<\delta} + \int_{\delta<|\phi|\leq\pi}$. For the first integral, we have $|f(\theta-\phi) - f(\theta)| < \epsilon$ for $|\phi| < \delta$ by uniform continuity, so

$$\frac{1}{2\pi}\int_{|\phi|<\delta} |f(\theta - \phi) - f(\theta)|P_r(\phi)\, d\phi < \epsilon\frac{1}{2\pi}\int_{-\pi}^{\pi} P_r(\theta) = \epsilon.$$

For the second integral, for all r sufficiently large we have $P_r(\theta) < \epsilon$ on $\delta < |\phi| \leq \pi$ by the lemma, while $|f(\theta - \phi) - f(\theta)| \leq 2\|f\|_\infty$, so

$$\frac{1}{2\pi}\int_{\delta<|\phi|\leq\pi} |f(\theta - \phi) - f(\theta)|P_r(\phi)\, d\phi \leq 2\epsilon\|f\|_\infty.$$

Thus for r sufficiently close to 1, we get $|f * P_r(\theta) - f(\theta)| \leq (1 + 2\|f\|_\infty)\epsilon$, so $f * P_r \to f$ unifomrly on $\mathbb{T}$. □

From the smoothing properties of convolution, we see that if $f \in L^1$ then $f * P_r$ is continuous in θ for each r. Thus there is no hope that $f * P_r \to f$ in the L^∞ norm when $f \in L^\infty$. However we do have the following weaker form of convergence:

Proposition 25.41. *If $f \in L^\infty(\mathbb{T})$ and $g \in L^1(\mathbb{T})$, then*

$$\lim_{r\to 1} \frac{1}{2\pi}\int_{-\pi}^{\pi} (f * P_r)(\theta)g(\theta)\, d\theta = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(\theta)g(\theta)\, d\theta.$$

Proof. Problem 25.18. □

As for the line, the Fourier transform is espcially well-behaved in the L^2 setting, though here things are somewhat simpler—we have $L^2(\mathbb{T}) \subset L^1(\mathbb{T})$ since the measure is finite. The proof of the following theorem is left as an exercise.

Theorem 25.42. *The Fourier transform is a unitary transformation from $L^2(\mathbb{T})$ onto $\ell^2(\mathbb{Z})$. In particular,*

i) *(Plancherel theorem) If $f \in L^2(\mathbb{T})$, then $\|f\|_2^2 = \sum_{n=-\infty}^{\infty} |\widehat{f}(n)|^2$.*
ii) *(Parseval identity) If $f, g \in L^2(\mathbb{T})$, then*

$$\int_{\mathbb{T}} f(\theta)\overline{g(\theta)}\, dm(\theta) = \sum_{n=-\infty}^{\infty} \widehat{f}(n)\overline{\widehat{g}(n)}.$$

Moreover, for $f \in L^2(\mathbb{T})$ the partial sums S_N of the Fourier Series for f converges to f in $L^2(\mathbb{T})$.

In particular, note that the Fourier transform takes the orthonromal basis $E = \{e^{in\theta}\}_{n\in\mathbb{Z}}$ of $L^2(\mathbb{T})$ onto the standard orthonormal basis of $\ell^2(\mathbb{Z})$, and the Fourier transform of a function $f \in L^2(\mathbb{T})$ is just its sequence of coefficients with respect to this orthonormal basis. Indeed if we write $e_n(\theta) = e^{in\theta}$, then

$$\widehat{f}(n) := \int_{\mathbb{T}} f(\theta)e^{-in\theta}\, dm(\theta) = \langle f, e_n\rangle_{L^2(\mathbb{T})}.$$

Thus, if one has already proved that the functions $\{e_n\}$ are an orthonormal basis for $L^2(\mathbb{T})$, the proof of Theorem 25.42 becomes quite simple. (Indeed, the theorem is essentially equivalent to the assertion that the characters $\{e_n\}$ form an orthonormal basis. Why?)

25.5. **Schwartz functions and distributions.** This section introduces the *Schwartz space* $\mathscr{S}$ and the space of *tempered distritbutions* $\mathscr{S}'$. The theory of distributions allows many of the important operations of analysis (such as differentiation and the Fourier transform) to be extended to objects more singular than functions (indeed distributions are sometimes known as *generalized functions*). The basic idea is this: if ψ is a very smooth function (say C^∞) and vanishes at infinity, then if f is differentiable we have the integration by parts formula

$$\int_{\mathbb{R}} f'(x)\psi(x)\, dx = -\int_{\mathbb{R}} f(x)\psi'(x)\, dx.$$

However, the second integral will make sense even if the first does not (that is, even if f' does not exist). If we identify f with the linear functional

$$\psi \to \int_{\mathbb{R}} f\psi,$$

then the above calculation suggests that we can interpret "f'" as the linear functional

$$\psi \to -\int f\psi'$$

even if f' does not exist in the usual sense. The theory of distributions makes this heuristic precise. The first step is to carefully identify the space of smooth functions we wish to use, and topologize it appropriately so that we can speak of continuous linear functionals.

Let $C_b^\infty(\mathbb{R})$ denote the vector space of bounded C^∞ functions on $\mathbb{R}$.

Definition 25.43. The *Schwartz space* $\mathscr{S}$ consists of all functions $\psi \in C_b^\infty(\mathbb{R})$ such that $x^\alpha \psi^{(\beta)}(x)$ is bounded for all integers $\alpha, \beta \geq 0$. $\triangleleft$

We say that a function f is *rapidly decreasing* if $x^\alpha \psi(x)$ is bounded for all $\alpha \geq 0$. So $\mathscr{S}$ consists of those ψ such that ψ and all of its derivatives are rapidly decreasing. For example, $\psi(x) = e^{-x^2}$ belongs to $\mathscr{S}$. It is an important fact that $\mathscr{S}$ is closed under differentiation, and under multiplication by polynomials:

Lemma 25.44. *$\mathscr{S}$ is a vector space, and if $\psi \in \mathscr{S}$ then $x\psi(x)$ and $\psi'(x)$ belong to $\mathscr{S}$, and in fact $x^\alpha \psi^{(\beta)} \in \mathscr{S}$ for all $\alpha, \beta \geq 0$.*

Proof. Exercise. $\square$

Definition 25.45. For integers $\alpha, \beta \geq 0$, define for $\psi \in \mathscr{S}$

$$\|\psi\|_{\alpha,\beta} := \|x^\alpha \psi^{(\beta)}\|_\infty$$

$\triangleleft$

Lemma 25.46. *Each $\|\cdot\|_{\alpha,\beta}$ is a norm on $\mathscr{S}$.*

It turns out that it is appropriate to topologize $\mathscr{S}$ not with a single norm, but with the whole family of norms $\|\cdot\|_{\alpha,\beta}$ simultaneously.

Definition 25.47. Say that a sequence $\psi_n \subset \mathscr{S}$ is *Cauchy* if it is Cauchy in each of the norms $\|\cdot\|_{\alpha,\beta}$, and say that a sequence $\psi_n \subset \mathscr{S}$ *converges* if there exists $\psi \in \mathscr{S}$ such that $\|\psi_n - \psi\|_{\alpha,\beta} \to 0$ for all α, β. $\triangleleft$

Of course, if ψ_n converges in $\mathscr{S}$ then the limit ψ is unique. (Check this.) Further:

Proposition 25.48. *$\mathscr{S}$ is complete. That is, if ψ_n is Cauchy in $\mathscr{S}$, then there exists $\psi \in \mathscr{S}$ such that $\|\psi_n - \psi\|_{\alpha,\beta} \to 0$ for all $\alpha, \beta \geq 0$.*

Proof. Since $C_b(\mathbb{R})$ is complete with respect to the $\|\cdot\|_\infty$ norm, by the definition of $\mathscr{S}$ and the $\|\cdot\|_{\alpha,\beta}$ norms we have that, for each $\alpha, \beta \geq 0$, there is a function $\psi_{\alpha,\beta} \in C_b(\mathbb{R})$ such that $x^\alpha \psi_n^{(\beta)} \to \psi_{\alpha,\beta}$ uniformly on $\mathbb{R}$. Put $\psi = \psi_{0,0}$, the proof is finished if we can show that $\psi_{\alpha,\beta} = x^\alpha \psi^{(\beta)}$ for all α, β.

From advanced calculus we know that if f_n converges uniformly to f and f_n' converges uniformly to g, then f is differentiable and $f' = g$. Applying this fact we conclude that $\psi_{0,1} = \psi_{0,0}'$, and applying it inductively we have $\psi_{0,\beta} = \psi^{(\beta)}$ for all $\beta \geq 0$. (In particular, $\psi_n^{(\beta)} \to \psi^{(\beta)}$ uniformly for all β.) From this it follows that $x^\alpha \psi_n^{(\beta)} \to x^\alpha \psi^{(\beta)}$ pointwise for all α, β, but since this sequence also converges to $\psi_{\alpha,\beta}$ uniformly we conclude that $\psi_{\alpha,\beta} = x^\alpha \psi^{(\beta)}$ for all α, β as desired. $\square$

We observed earlier that $\mathscr{S}$ is closed under differentiation and multiplication by polynomials; we now see that these operations are continuous:

Lemma 25.49. *If $\psi_n \to \psi$ in $\mathscr{S}$, then $x^\alpha \psi_n \to x^\alpha \psi$ and $\psi_n^{(\beta)} \to \psi^{(\beta)}$ in $\mathscr{S}$.*

Proof. Exercise. □

It is useful to observe (in connection with our discussion of the Fourier transfrom later) that convergence in the family of norms $\|\cdot\|_{\alpha,\beta}$ controls L^p convergence:

Proposition 25.50. *$\mathscr{S} \subset L^p$ for all $1 \leq p \leq \infty$, and if $\psi_n \to \psi$ in $\mathscr{S}$ then also $\psi_n \to \psi$ in $L^p(\mathbb{R})$.*

Proof. By the previous lemma, we have $(1+x^2)\psi \in \mathscr{S}$ for all $\psi \in \mathscr{S}$, so in particular for all $x \in \mathbb{R}$

$$|\psi(x)| \leq \frac{|\psi(x)| + |x^2\psi(x)|}{1+x^2} \leq \frac{\|\psi\|_{0,0} + \|\psi\|_{2,0}}{1+x^2}.$$

But $(1+x^2)^{-1}$ belongs to $L^p(\mathbb{R})$ for all $1 \leq p \leq \infty$, so $\psi \in L^p$. Applying this same estimate to $\psi_n - \psi$ we see that

$$\|\psi_n - \psi\|_p \leq (\|\psi_n - \psi\|_{0,0} + \|\psi_n - \psi\|_{2,0}) \|(1+x^2)^{-1}\|_p$$

and the right hand side goes to 0 as $\psi_n \to \psi$ in $\mathscr{S}$. □

Definition 25.51. The space of *tempered distributions* $\mathscr{S}'$ consists of all continuous linear maps $F : \mathscr{S} \to \mathbb{C}$. That is, a map $F : \mathscr{S} \to \mathbb{C}$ belongs to $\mathscr{S}'$ if and only if it is linear and $F(\psi_n) \to F(\psi)$ whenever $\psi_n \to \psi$ in $\mathscr{S}$. ◁

It is straightforward to check that $\mathscr{S}'$ is a vector space. To emphasize the role of $\mathscr{S}'$ as the dual space of $\mathscr{S}$, we will write $\langle F, \psi\rangle$ for $F(\psi)$. Tempered distributions F are sometimes called *generalized functions*. We will topologize $\mathscr{S}'$ as follows: say $F_n \to F$ in $\mathscr{S}'$ if $\langle F_n, \psi\rangle \to \langle F, \psi\rangle$ for all $\psi \in \mathscr{S}$.

The following examples are fundamental; the unproved claims are left as exercises.

Example 25.52. a) (Tempered functions) A measurable function $f : \mathbb{R} \to \mathbb{C}$ is called *tempered* if $(1+|x|)^{-N} f \in L^1$ for some integer $N \geq 0$.

Each tempered function f defines a tempered distribution by the formula

$$\langle f, \psi\rangle = \int_{\mathbb{R}} f\psi.$$

(To see that $f\psi \in L^1$ for every $\psi \in \mathscr{S}$, write $f\psi = (1+|x|)^{-N} f (1+|x|)^N \psi$.) The fact that $\psi \to \langle f, \psi\rangle$ is continuous follows from dominated convergence. For examples of tempered functions, note that every $f \in L^p$, $1 \leq p \leq \infty$ is tempered (apply Hölder's inequality to $(1+|x|)^{-2} f$). More generally, any polynomial times a tempered function is a tempered function. Let us also observe that if f, g are tempered functions, then the associated tempered distributions are equal if and only if $f = g$ a.e. This justifies the name "generalized functions."

b) (Tempered measures) A (positive, signed, or complex) Borel measure μ on $\mathbb{R}$ is called *tempered* if $\int_{\mathbb{R}}(1+|x|)^{-N}\,d|\mu|(x) < \infty$ for some integer $N \geq 0$. Every tempered measure gives rise to a tempered distribution via the pairing

$$\langle \mu, \psi\rangle = \int_{\mathbb{R}} \psi\, d\mu.$$

If μ is absolutely continuous with respect to Lebesgue measure m, with Radon-Nikodym derivative $f = \frac{d\mu}{dm}$, then μ is tempered if and only if f is a tempered function, and we are back to example (a).

△

To give more examples, we first look at ways to obtain new tempered distributions from old ones. One elementary but important way is the following:

Proposition 25.53. *If $F \in \mathscr{S}'$ and $T : \mathscr{S} \to \mathscr{S}$ is a continuous linear map, then*

$$\langle T'F, \psi\rangle := \langle F, T\psi\rangle$$

defines a tempered distribution.

Proof. If $\psi_n \to \psi$ in $\mathscr{S}$, then

$$\langle F, T\psi_n\rangle \to \langle F, T\psi\rangle$$

so $T'F$ defines a distribution. □

Before moving on to more general classes of distributions, we consider one more special example:

Proposition 25.54 (The Principal Value integral)**.** *For each $\psi \in \mathscr{S}$, the limit*

$$\langle P_{1/x}, \psi\rangle := \lim_{\epsilon\to 0} \int_{|x|\geq\epsilon} \frac{1}{x}\psi(x)\, dx \tag{77}$$

exists, and defines a tempered distribution.

Proof. We first show that (77) is well-defined on $\mathscr{S}$. Let $\psi \in \mathscr{S}$, then by changing variables in the integral on the negative half-line we get for each $\epsilon > 0$

$$\int_{|x|\geq\epsilon} \frac{1}{x}\psi(x)\, dx = \int_{\epsilon}^{\infty} \frac{\psi(x)-\psi(-x)}{x}\, dx$$

Since ψ is differentiable at 0, the integrand is bounded in a neighborhood of 0, and since $x\psi(x)$ is bounded, the integrand decays faster than $1/x^2$ near infinity, so the integral is convergent. Thus the limit exists as $\epsilon \to 0$, and equals

$$\int_0^{\infty} \frac{\psi(x)-\psi(-x)}{x}\, dx.$$

To see that $P_{1/x}$ is continuous on $\mathscr{S}$, first observe that for $x > 0$

$$\left|\frac{\psi(x)-\psi(-x)}{x}\right| = \left|\frac{1}{x}\int_{-x}^{x}\psi'(t)\,dt\right| \tag{78}$$
$$\leq \frac{1}{x}\int_{-x}^{x}|\psi'(t)|\,dt \tag{79}$$
$$\leq 2\|\psi'\|_\infty \tag{80}$$

It follows that

$$|\langle P_{1/x},\psi\rangle| \leq \int_0^1 \left|\frac{\psi(x)-\psi(-x)}{x}\right| + \int_1^\infty \left|\frac{\psi(x)-\psi(-x)}{x}\right| \tag{81}$$
$$\leq 2\|\psi'\|_\infty + \int_1^\infty (|x\psi(x)| + |x\psi(-x)|)\frac{dx}{x^2} \tag{82}$$
$$\leq 2\|\psi'\|_\infty + 2\|x\psi(x)\|_\infty \tag{83}$$
$$= 2\|\psi\|_{0,1} + 2\|\psi\|_{1,0} \tag{84}$$

If we consider now $\psi_n \to \psi$ in $\mathscr{S}$, then the above estimate applied to $\psi_n - \psi$ shows that $\langle P_{1/x},\psi_n\rangle \to \langle P_{1/x},\psi\rangle$ and the proof is finished. □

Proposition 25.55 (Differentiation of temepered distributions). *For any integer $\beta \geq 0$ and any tempered distribution $F \in \mathscr{S}'$, the map*

$$\psi \to \langle F, (-1)^\beta \psi^{(\beta)}\rangle$$

defines a tempered distribution, called the β^{th} distributional derivative of F, denoted $F^{(\beta)}$.

Proof. By Lemma 25.49, the map $T\psi = (-1)^{(\beta)}\psi^{(\beta)}$ is continuous on $\mathscr{S}$, and the result follows by Proposition 25.53. □

The reason for including the sign $(-1)^\beta$ in the definition of the distributional derivative is so that our definition is compatible with integration by parts. In particular, if f and f' are tempered functions, then the (formal) integration by parts calculation at the beginning of this section is valid, and shows that the distributional derivative $\psi \to \langle f,\psi\rangle$ is $\psi \to \langle f',\psi\rangle$.

Proposition 25.56 (The Heaviside function). *Let H be the Heaviside function $H(x) = \mathbf{1}_{[0,\infty)}$. Then $H' = \delta$ in the sense of distributions.*

Proof. Let $\psi \in \mathscr{S}$. H is a tempered function, so H' is given by

$$\begin{aligned}\langle H', \psi\rangle &= -\langle H, \psi'\rangle \\ &= -\int_{-\infty}^{\infty} H(x)\frac{d\psi}{dx}\,dx \\ &= -\int_{0}^{\infty} \frac{d\psi}{dx}\,dx \\ &= \psi(0) \\ &= \langle \delta, \psi\rangle.\end{aligned}$$

□

Notice that every distribution is infinitely differentiable in the sense of distributions. So, we can take another derivative to get $H'' = \delta'$. A quick computation shows that $\langle \delta', \psi\langle = -\psi'(0)$. It can be shown that δ' is not given by any tempered measure (see Problem 25.22.

Our next use of Proposition 25.53 will allow us to define the convolution of a distribution with a Schwartz function ϕ.

Proposition 25.57. *Let $\phi \in \mathscr{S}$ and $F \in \mathscr{S}'$. Then the map*

$$\langle \phi * F, \psi\rangle := \langle F, \tilde{\phi} * \psi\rangle$$

defines a tempered distribution, called the convolution of F and ϕ. (Here $\tilde{\phi}(x) = \phi(-x)$.)

Proof. Again it suffices to verify that the map $\psi \to \tilde{\phi} * \psi$ is continuous on $\mathscr{S}$. The proof is left as Problem 25.24. □

If $f \in L^1$, one can also verify that $f * \phi$, viewed as a distribution, agrees with the distribution induced by the L^1 function $f * \phi$ defined by ordinary convolution (Problem 25.24).

It is instructive to revisit L^1 approximate units in the context of distributions. If $\{\phi_\lambda\}_{\lambda>0}$ is an L^1 approximate unit, then each ϕ_λ is a tempered function and hence defines a distribution.

Proposition 25.58. *If ϕ_λ is an L^1 approximate unit, then $\phi_\lambda \to \delta$ in $\mathscr{S}'$.*

Proof. By definition, for any $\psi \in \mathscr{S}$

$$\langle \phi_\lambda, \psi\rangle = \int_{-\infty}^{\infty} \phi_\lambda(y)\psi(y)\,dy = (\phi_\lambda * \tilde{\psi})(0)$$

where $\tilde{\psi}(x) = \psi(-x)$. As $\lambda \to 0$, by Lemma 25.18 we have $(\phi_\lambda * \tilde{\psi})(0) \to \psi(0) = \langle \delta, \psi\rangle$. □

Finally we consider the Fourier transform. The key fact is the following:

Lemma 25.59. *If $\psi \in \mathscr{S}$, then $\widehat{\psi} \in \mathscr{S}$, and the map $\widehat{\ }: \mathscr{S} \to \mathscr{S}$ is continuous.*

Proof. The fact that $\widehat{\psi}$ belongs to $\mathscr{S}$ follows from repeated application of Propositions 25.6 and 25.8. Continuity follows from the fact that if $\psi_n \to \psi$ in $\mathscr{S}$, then also $\|x^\alpha \frac{d^\beta}{dx^\beta}(\psi_n - \psi)\|_1 \to 0$ for all α, β. See Problem 25.23. □

Proposition 25.60 (Fourier transforms of tempered distributions)**.** *If $\psi \in \mathscr{S}$, then $\widehat{\psi} \in \mathscr{S}$, and for any $F \in \mathscr{S}'$ the formula*

$$\langle \widehat{F}, \psi \rangle := \langle F, \widehat{\psi} \rangle$$

defines a tempered distribution, called the Fourier transform of F.

Example 25.61. a) Let δ_t be the point mass at $t \in \mathbb{R}$. We can compute $\widehat{\delta_t}$: for $\psi \in \mathscr{S}$ we have

$$\begin{aligned} \langle \widehat{\delta_t}, \psi \rangle &= \langle \delta_t, \widehat{\psi} \rangle && (85)\\ &= \widehat{\psi}(t) && (86)\\ &= \int_{-\infty}^{\infty} e^{-2\pi i x t} \psi(x)\, dx && (87)\\ &= \langle e^{-2\pi i x t}, \psi \rangle && (88) \end{aligned}$$

so $\widehat{\delta_t} = e^{-2\pi i x t}$.

The expected inverstion $\widehat{(e^{-2\pi i x t})} = \delta_t$ also holds; the proof is left as an exercise.

b) Consider the distribution $P_{1/x}$ of Proposition 25.54. One can show that $\widehat{P_{1/x}}$ is the tempered distribution given by the tempered function

$$F(t) = -\pi i \operatorname{sgn}(t).$$

△

25.6. **Problems.**

Problem 25.1. Prove Proposition 25.2

Problem 25.2. Complete the proof of Lemma 25.4.

Problem 25.3. Prove Proposition 25.14

Problem 25.4. Prove, if $E \subset [0,1]$ has positive Lebesgue measure, then the set

$$E - E = \{x - y : x, y \in E\}$$

contains an interval centered at the origin. (Hint: let $-E = \{-x : x \in E\}$ consider the function $h(x) = \mathbf{1}_{-E} * \mathbf{1}_E$.)

Problem 25.5. Suppose ϕ is an unsigned L^1 function with $\int \phi = 1$, and let $\phi_\lambda(x) = \frac{1}{\lambda}\phi\left(\frac{x}{\lambda}\right)$.

a) Prove $\{\phi_\lambda\}_{\lambda>0}$ is an L^1 approximate unit.
b) Give a simpler proof of Lemma 25.18 by making a change of variables in equation (60).

Problem 25.6. a) Prove, if $f \in C_c^1(\mathbb{R})$ and g is a compactly supported L^1 function, then $f * g$ is C^1 with compact support. (Hint: justify differentiation under the integral sign.)

b) By induction, conclude that if $f \in C_c^\infty(\mathbb{R})$ and $g \in L^1$ is compactly supported, then $f * g \in C_c^\infty(\mathbb{R})$.

c) Conclude that $C_c^\infty(\mathbb{R})$ is dense in L^p for all $1 \leq p < \infty$. (Apply Theorem 25.17 with ϕ a bump function.)

d) Construct a bump function on $\mathbb{R}^n$ and extend the above results to $n > 1$.

Problem 25.7. Compute the integral in Lemma 25.23.

Problem 25.8. This problem gives a proof that the Fourier transform $\widehat{\ }: L^1 \to C_0(\mathbb{R})$ is not surjective.

a) Draw a picture of $h_n := \mathbf{1}_{[-n,n]} * \mathbf{1}_{[-1,1]}$ and determine its norm.

b) Show that h_n is, up to a multiplicative constant independent of n, the Fourier transform of the L^1 function

$$f_n := \frac{\sin 2\pi x \sin 2\pi n x}{x^2}.$$

(Hint: you can compute integrals, or use the L^1 inversion theorem.)

c) Show that $\|f_n\|_1 \to \infty$ as $n \to \infty$. Conclude that the Fourier transform is not surjective. (Hint: if it were surjective...). Prove, however, that the Fourier transform does have dense range.

Problem 25.9. Suppose that $f \in L^1$, f is differentiable a.e., $f' \in L^1$, and $f(x) = \int_{-\infty}^x f'(y)\,dy$ for a.e. $x \in \mathbb{R}$. Prove $\widehat{f'} = 2\pi i t \widehat{f}(t)$.

Problem 25.10. Complete the proof of Theorem 25.30.

Problem 25.11. Let φ_λ be an $L^1(\mathbb{T})$ approximate unit. Prove, if $f \in C(\mathbb{T})$, then $f * \varphi_\lambda \to f$ uniformly as $\lambda \to 0$.

Problem 25.12. State and prove an analog of Proposition 25.2 for Fourier series.

Problem 25.13. Prove Theorem 25.33.

Problem 25.14. Prove Theorem 25.42. Also prove that $L^2(\mathbb{T})$ inversion is possible in the following sense: for all $f \in L^2(\mathbb{T})$, we have

$$\lim_{N \to \infty} \frac{1}{2\pi} \int_{-\pi}^{\pi} \left| f(\theta) - \sum_{n=-N}^{N} \widehat{f}(n) e^{in\theta} \right|^2 d\theta = 0.$$

(In other words, the partial sums s_N of the Fourier series converge to f in the L^2 norm.)

Problem 25.15. Let $\mathcal{A}(\mathbb{T})$ denote the set of functions $f \in L^1(\mathbb{T})$ such that the Fourier transform $\widehat{f}$ belongs to $\ell^1(\mathbb{Z})$.

a) Prove, if $f, g \in \mathcal{A}(\mathbb{T})$, then their product fg also belongs to $\mathcal{A}(\mathbb{T})$. (Hint: use the ℓ^1 inversion theorem to write f and g as the sums of their Fourier series, and express the Fourier coefficients of fg in terms of the coefficients of f and g.) Thus, $\mathcal{A}(\mathbb{T})$ is a ring.

b) Prove the Fourier transform is a ring isomorphism from $\mathcal{A}(\mathbb{T})$ *onto* $\ell^1(\mathbb{Z})$ (where the multiplication on $\ell^1(\mathbb{Z})$ is convolution).

Problem 25.16. Prove, if $f \in C^k(\mathbb{T})$, $k \geq 1$, then the Fourier coefficients of f satisfy

$$\lim_{n\to\pm\infty} |n|^k |\widehat{f}(n)| = 0.$$

(Hint: first compute the Fourier transform of f' explicitly.)

Problem 25.17. Fill in the details in the proof of Theorem 25.34. To show that $\|D_N\|_1 \to \infty$, fix N, and for each $0 \leq k \leq 2N$ let I_k denote the interval

$$\left[\frac{1}{2}\left(\frac{k\pi}{N+\frac{1}{2}}\right), \frac{1}{2}\left(\frac{(k+1)\pi}{N+\frac{1}{2}}\right)\right]$$

(These are intervals on which D_N has constant sign.) Then

$$\int_{-\pi}^{\pi} |D_N(t)|\,dt = 2\sum_{k=0}^{2N} \int_{I_k} \left|\frac{\sin(N+\frac{1}{2})t}{\sin\frac{t}{2}}\right| dt$$

To estimate the integral over I_k, first show that there is a universal constant $C > 0$ such that

$$\left|\frac{1}{\sin\frac{t}{2}}\right| \geq C\frac{n+\frac{1}{2}}{k} \quad \text{for all } t \in I_k$$

for each $k > 0$.

Problem 25.18. Prove Proposition 25.41.

Problem 25.19. [Fourier transforms of measures] Let μ be a finite (signed or complex) Borel measure on $\mathbb{T}$. The Fourier transform of μ is the function $\widehat{\mu} : \mathbb{Z} \to \mathbb{C}$ defined by

$$\widehat{\mu}(n) := \int_{\mathbb{T}} e^{-in\theta}\,d\mu(\theta).$$

a) Prove $\widehat{\mu}$ is bounded. Give an example of a measure μ such that $\widehat{\mu} \notin c_0(\mathbb{Z})$.

b) For fixed μ, define for each $0 \leq r < 1$

$$A(r,\theta) := \sum_{n=-\infty}^{\infty} \widehat{\mu}(n) r^{|n|} e^{in\theta}.$$

Prove the measures $\mu_r := A(r,\theta)dm(\theta)$ converge to μ as $r \to 1$, in the following sense: for every continuous function f on $\mathbb{T}$,

$$\lim_{r\to 1} \int_{\mathbb{T}} f(\theta)A(r,\theta)dm(\theta) = \int_{\mathbb{T}} f(\theta)\,d\mu(\theta).$$

Problem 25.20. [The Fejér kernel] Consider the cutoff function on $\mathbb{Z}$

$$\psi_N(k) = \begin{cases} 0 & \text{if } |k| > N \\ 1 - \frac{|k|}{N} & \text{if } |k| \leq N \end{cases}$$

Find a closed form expression for

$$F_N(\theta) = \sum_{k=-N}^{N} \psi_N(k) e^{ik\theta},$$

and show that the family $\{F_N(\theta)\}_{N\geq 1}$ is an $L^1(\mathbb{T})$ approximate unit.

Problem 25.21. Prove $\mathscr{S}$ is dense in $L^p(\mathbb{R})$ for $1 \leq p < \infty$.

Problem 25.22. Prove there is no finte Borel measure μ on $[-1,1]$ such that $\int_{-1}^{1} f\,d\mu = f'(0)$ for all $f \in C^1[-1,1]$.

Problem 25.23. a) Complete the proof of Lemma 25.59.

c) Prove, if $\phi, \psi \in \mathscr{S}$ then
$$\int_{\mathbb{R}} \widehat{\phi}\psi = \int_{\mathbb{R}} \phi\widehat{\psi}.$$
(This justifies the definition of $\widehat{F}$.)

b) Prove Fourier inversion holds in $\mathscr{S}$; that is, $(\widehat{\psi})\check{} = \psi$.

c) State and prove a Fourier inversion theorem for the Fourier transform $F \to \widehat{F}$ on $\mathscr{S}$.

Problem 25.24. a) Prove, if $\phi \in \mathscr{S}$ and $\psi_n \to \psi$ in $\mathscr{S}$, then $\phi * \psi_n \to \phi * \psi$ in $\mathscr{S}$. (Here convolution means oridnary convolution of functions.)

b) Let $f \in L^1$ and $\phi \in \mathscr{S}$. Prove the tempered distribution $f * \phi$ coincides with the distribution defined by the tempered function $f * \phi$.

c) Show that $\delta * \phi = \phi$ in the sense of distributions.

d) Let $\phi \in \mathscr{S}$ be a nonnegative function with $\int \phi = 1$, and let $\phi_\lambda(x) := \frac{x}{\lambda}\phi\left(\frac{x}{\lambda}\right)$ the corresponding L^1 approximate unit. Prove for any $F \in \mathscr{S}'$, $F * \phi_\lambda \to F$ in $\mathscr{S}'$ as $\lambda \to 0$.

26. Banach Algebras and the $\frac{1}{f}$ Theorem

A *Banach Algebra* $\mathscr{A}$ is a complete normed (associative) algebra over $\mathbb{C}$ such that

$$\|a\,b\| \le \|a\|, \|b\|$$

for $a, b \in \mathscr{A}$ and in the case $\mathscr{A}$ is unital (with unit 1), $\|1\| = 1$. (In the unital case there is always an equivalent norm in which the unit has norm 1.)

Example 26.1. Here are three examples of unital Banach algebras.

- (i) For a Hilbert space H, the algebra $B(H)$ is a unital Banach algebra.
- (ii) Let X be a compact metric space. The Banach space $C(X)$ is a unital Banach algebra.
- (iii) Given $f, g \in \ell^1(\mathbb{Z})$ and $n \in \mathbb{Z}$ observe that

$$\sum_{n=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} |f(j)g(n-j)| = \sum_j [\sum_n |g(n-j)|]|f(j)| = \|f\|_1 \, \|g\|_1.$$

Define $f * g : \mathbb{Z} \to \mathbb{C}$ by

$$f * g(n) = \sum_{j=-\infty}^{\infty} f(j)\, g(n-j).$$

The computation above shows this sum converges and further that $f * g \in \ell^1(\mathbb{Z})$. It also verifies that $\ell^1(\mathbb{Z})$ under the multiplication $*$ is a Banach algebra. The operation $*$ is the *convolution product* on $\ell^1(\mathbb{Z})$.

△

26.1. **Ideals and quotients.**

Proposition 26.2. *Suppose $\mathscr{A}$ is a Banach algebra with unit* 1*.*

- *(i) if $\|a\| < 1$, then $1 - a$ is invertible and moreover,*

$$(1-a)^{-1} = \sum_{j=0}^{\infty} a^j,$$

and

$$\|(1-a)^{-1}\| \le \frac{\|a\|}{1 - \|a\|};$$

- *(ii) the set $\mathscr{G}$ of invertible elements is open;*
- *(iii) the mapping $\mathscr{G} \to \mathscr{G}$ given by $a \to a^{-1}$ is continuous;*
- *(iv) if $\mathscr{I} \subset \mathscr{A}$ is a proper ideal, then so is its closure;*
- *(v) if $\mathscr{M} \subset \mathscr{A}$ is a maximal ideal, then $\mathscr{M}$ is closed;*
- *(vi) every proper ideal is contained in a maximal ideal;*
- *(vii) if $\mathscr{I}$ is a closed ideal, then the quotient $\mathscr{A}/\mathscr{I}$ is a Banach algebra with unit and is commutative if $\mathscr{A}$ is.*

Proof. If $\|a\| < 1$, then the partial sums

$$s_n = \sum_{j=0}^{n} a_j$$

converge (as $\mathscr{A}$ is complete) to some s. Since

$$(1-a)s_n = 1 - a_{n+1}$$

and the right hands side tends to 1,

$$(1-a)s = 1.$$

Moreover, $\|1-a\| < \frac{\|a\|}{1-\|a\|}$.

Suppose a is invertible. Given c such that $\|c-a\| < \frac{1}{2}\|a^{-1}\|^{-1}$,

$$\|ca^{-1} - 1\| = \|(c-a)a^{-1}\| < \frac{1}{2}$$

and therefore $1-(1-ca^{-1}) = ca^{-1}$ is invertible and, thus so is c. Further,

$$\|ac^{-1}\| \le \frac{\|1-ca^{-1}\|}{1-\|1-ca^{-1}\|} < \frac{3}{2}$$

Hence,

$$\|c^{-1}\| = \|a^{-1}ac^{-1}\| \le \frac{3}{2}\|a^{-1}\|.$$

As for continuity, if $0 < \delta < \frac{1}{2}$ and $\|c-a\| < \frac{1}{2}\|a^{-1}\|^{-1}$, then

$$\|a^{-1} - c^{-1}\| = \|a^{-1}(c-a)c^{-1}\| \le \|a^{-1}\|\,\|c^{-1}\|\,\|c-a\| < \frac{3}{2}\|a^{-1}\|\,\|c-a\|.$$

Let $\mathscr{G}$ denote the set of invertible elements of $\mathscr{A}$. Thus $\mathscr{G}$ is open. If $\mathscr{I}$ is a proper ideal, then $\mathscr{I} \subset \mathscr{G}^c$ and hence $\overline{\mathscr{I}} \subset \mathscr{G}^c$. An easy argument, based upon continuity of multiplication and addition, shows $\overline{\mathscr{I}}$ is an ideal and thus a proper ideal.

Let $\pi : \mathscr{A} \to \mathscr{A}/\mathscr{I}$ denote the quotient map. That $\mathscr{A}/\mathscr{I}$ is again a Banach space with the quotient norm

$$\|\pi(a)\| = \inf\{\|a-b\| : b \in \mathscr{I}\|\} = \inf\{\|x\| : \pi(x) = \pi(a)\}.$$

has long since been proved. What remains to show is that the quotient norm is a Banach algebra norm. To this end, let $a, b \in \mathscr{A}$ and $\epsilon > 0$ be given. There exists $a', b' \in \mathscr{A}$ such that $\pi(a') = a$, $\pi(b') = b$ and

$$\|a'\| < \|\pi(a)\| + \epsilon,\ \|b'\| < \pi(b) + \epsilon.$$

Thus, $\pi(a'b') = \pi(ab)$ and

$$\|\pi(ab)\| \le \|(a'b')\| \le (\|\pi(a)\| + \epsilon)(\|\pi(b)\| + \epsilon).$$

Thus multiplication in the quotient is sub-multiplicative. In particular, $0 \ne \|\pi(1)\| = \|\pi(1)^2\| \le \|\pi(1)\|^2$ and therefore $\|\pi(1)\| \ge 1$. On the other hand, by the definition of the quotient norm, $\|\pi(1)\| \le \|1\| = 1$. Alternately, note if $\|1-m\| < 1$, then $m = 1-(1-m)$ is invertible and hence m does not lie in a proper ideal. $\square$

26.2. **The spectrum.** Suppose $\mathscr{A}$ is a Banach algebra with unit. The *spectrum* of an element $a \in \mathscr{A}$, denoted $\sigma(a)$, is given by

$$\sigma(a) = \sigma_A(a) = \{z \in \mathbb{C} : a - z1 \text{ is not invertible}\}.$$

The complement of σ is the *resolvent set*.

Theorem 26.3. *The spectrum $\sigma(a)$ is not empty and compact.*

Proof. Arguing by contradiction, suppose $a - z1$ is invertible for all $z \in \mathbb{C}$. Define $f : \mathbb{C} \to \mathscr{A}$ by $f(z) = (a - z1)^{-1}$. An easy argument shows

$$\lim_{|z| \to \infty} |f(z)| = 0.$$

and therefore f is bounded. Given a linear function $\lambda \in sA^*$, let $g : \mathbb{C} \to \mathbb{C}$ denote the function

$$g(z) = \lambda(f(z)).$$

It follows that g is bounded. Further, for $w \in \mathbb{C}$ not zero,

$$f(w) - f(z) = [(a - w)^{-1} - (a - z)^{-1}] = (a - w)^{-1}[z - w](a - z)^{-1}.$$

Hence, for $w \neq z$,

$$\frac{g(w) - g(z)}{w - z} + \lambda((a - z)^{-2}) = \lambda([(a - w)^{-1} - (a - z)^{-1}](a - z)^{-1}).$$

Fixing z and using continuity of the maps $w \mapsto a - w$ and of $b \mapsto b^{-1}$, it follows that g is differentiable at z and

$$g'(z) = -\lambda((a - z)^{-2}).$$

Thus g is entire and bounded and therefore constant by Liouville's Theorem. Now g vanishes at ∞ and hence is identically zero. Choosing λ such that $\lambda(a^{-1}) \neq 0$ gives a contradiction since $g(0) = \lambda(f(0)) = \lambda(a^{-1})$. Thus $\sigma(a)$ is not empty.

For z sufficiently large $(a - z)$ is invertible. Thus $\sigma(a)$ is a bounded set. The set $\mathscr{G}$ of invertible elements is open and the mapping $h : \mathbb{C} \to \mathscr{A}$ defined by $h(z) = a - z$ is continuous. Hence $O = h^{-1}(\mathscr{G})$ is open and therefore $\sigma(a) = O^c$ is closed. □

26.3. **Commutative Banach algebras.**

Proposition 26.4. *Suppose $\mathscr{A}$ is a commutative Banach algebra with unit and $\mathscr{M} \subset \mathscr{A}$. If $\mathscr{M}$ is a maximal ideal, then $\mathscr{Q} = \mathscr{A}/\mathscr{M}$ has no proper (non-zero) ideals and thus each non-zero element of $\mathscr{Q}$ is invertible.*

Proof. Let $\pi : \mathscr{A} \to \mathscr{A}/\mathscr{M}$ denote the quotient map. If $\mathscr{I} \subset \mathscr{A}/\mathscr{M}$ is a proper ideal, then $\pi^{-1}(\mathscr{I}) \subset \mathscr{A}$ is an ideal containing $\mathscr{M}$. It is also proper as π is onto. By maximality, $\pi^{-1}(\mathscr{I}) = \mathscr{M}$ and hence, as π is onto, $I = \pi(\pi^{-1}(I)) = \pi(\mathscr{M}) = (0) \subset \mathscr{Q}$. Thus $\mathscr{Q}$ has no non-trivial proper ideals. If $q \in \mathscr{Q}$, then $q\mathscr{Q}$ is an ideal that is not proper; i.e., $q\mathscr{Q} = \mathscr{Q}$ and hence q is invertible. □

Proposition 26.5 (Banach-Mazur)**.** *Suppose $\mathscr{A}$ is a commutative Banach algebra with unit. If each non-zero element of $\mathscr{A}$ is invertible, then $\mathscr{A}$ is $\mathbb{C}$; i.e., letting 1 denote the unit, $\mathscr{A} = \{z1 : z \in \mathbb{C}\}$ isometrically isomorphically. In particular, if $\mathscr{M}$ is a maximal ideal, then $\mathscr{A}/\mathscr{M} = \mathbb{C}$.*

Proof. Let $u \in \mathscr{A}$ be given. Since $\sigma(a) \neq \emptyset$, there is a λ such that $a-\lambda$ is not invertible. Hence $a - \lambda = 0$; i.e., $a = \lambda 1$. □

Proposition 26.6. *Suppose $\mathscr{A}$ is a commutative Banach algebra with unit. If $a \in \mathscr{A}$ is not invertible, then there is a homomorphism $\varphi : \mathscr{A} \to \mathbb{C}$ such that $\varphi(a) = 0$. In particular, $a \in \mathscr{A}$ is invertible if and only if $\varphi(a) \neq 0$ for every homomorphism $\varphi : \mathscr{A} \to \mathbb{C}$.*

Proof. If a is not invertible, then $\mathscr{I} = a\mathscr{A}$ is a proper ideal (it contains no invertible elements). Thus a is contained in a maximal ideal $\mathscr{M}$. Let $\varphi : \mathscr{A} \to \mathscr{A}/\mathscr{M} = \mathbb{C}$ denote the quotient map. It is a homomorphism and $\varphi(a) = 0$. □

Proposition 26.7. *Suppose $\mathscr{A}$ is a commutative Banach algebra with unit. If $\mathscr{M}$ is a maximal ideal, then there is a uniquely homomorphism $\varphi : \mathscr{A} \to \mathbb{C}$ such that $\ker(\varphi) = \mathscr{M}$. Conversely, if $\varphi : \mathscr{A} \to \mathbb{C}$ is a homomorphism, then $\ker(\varphi)$ is a maximal ideal. In particular homomorphisms $\varphi : \mathscr{A} \to \mathbb{C}$ are continuous with $\|\varphi\| = 1$ and there is a bijective correspondence between homomorphisms and maximal ideals.*

Proof. Suppose $\varphi, \psi : \mathscr{A} \to \mathbb{C}$ are homomorphisms with the same kernel $\mathscr{M}$. If $a \in \mathscr{M}$ then of course $\varphi(a) = \psi(a) = 0$. Otherwise, $\varphi(a) \neq 0$ and $\varphi(\varphi(a)1 - a) = 0$. Hence $0 = \psi(\varphi(a)1 - a) = \varphi(a) - \psi(a)$.

Now suppose $\varphi : \mathscr{A} \to \mathbb{C}$ is a homomorphism and let $\mathscr{M}$ denote its kernel. Thus $\mathscr{M}$ is an ideal and there is a homomorphism $\rho : \mathscr{Q} = \mathscr{A}/\mathscr{M} \to \mathbb{C}$ such that $\varphi(a) = \rho(\pi(a))$ where π is the quotient map. It follows that $\mathscr{Q}$ has no ideals and therefore $\mathscr{M}$ is a maximal ideal. Finally $\varphi = \tau \circ \pi$ where τ is the isometric isomorphism identifying $\mathscr{A}/\mathscr{M}$ with $\mathbb{C}$ since both are homomorphisms with kernel $\mathscr{M}$. Hence $\|\varphi\| = 1$. □

26.4. **The $\frac{1}{f}$ theorem.** The *Wiener algebra* $\mathscr{W}(\mathbb{T})$ consists of those functions $f \in L^1(\mathbb{T})$ such that $\hat{f} \in \ell^1(\mathbb{Z})$.

Proposition 26.8. *$\mathscr{W}(\mathbb{T})$ is a subalgebra of $C(\mathbb{T})$ and, for $f, g \in \mathscr{W}(\mathbb{T})$,*

$$\widehat{fg} = \hat{f} * \hat{g}.$$

In particular, $\mathscr{W}(\mathbb{T})$ is a Banach algebra when given the norm $\|f\| = \|\hat{f}\|_1$ and this Banach algebra is isometrically isomorphic to $\ell^1(\mathbb{Z})$ under the mapping $f \mapsto \hat{f}$.

Proof. Suppose $f, g \in \mathscr{W}(\mathbb{T})$. Since $\hat{f} \in \ell^1(\mathbb{Z})$, the series

$$\tilde{f}(e^{it}) = \sum_{n=-\infty}^{\infty} \hat{f}(n) e^{int}$$

converges absolutely and uniformly to a continuous function $\tilde{f}$ on $\mathbb{T}$. Evidently $\hat{\tilde{f}}(n) = \hat{f}(n)$ and therefore $\tilde{f} = f$ as elements of $L^1(\mathbb{T})$ and we may assume f is continuous. Thus both f, g are continuous and

$$f = \sum_n \hat{f}(n)e^{int}, \quad g = \sum_n \hat{g}(n)e^{int}$$

with the series converging absolutely and uniformly. Hence, for each n the sum

$$\sum_{j=-\infty}^{\infty} \hat{f}(j)\hat{g}(n-j)$$

converges,

$$\sum_n \sum_{j=-\infty}^{\infty} |\hat{f}(j)\hat{g}(n-j)| = \|\hat{f}\|_1 \, \|\hat{g}\|_1.$$

and

$$fg = \sum_n \left[\sum_{j=-\infty}^{\infty} \hat{f}(j)\hat{g}(n-j) \right] e^{int}$$

with the sum converging absolutely and uniformly. Thus $\widehat{fg}(n) = \hat{f} * \hat{g}(n)$. □

Proposition 26.9. *If $\varphi : \mathscr{W}(\mathbb{T}) \to \mathbb{C}$ is a homomorphism, then there is a θ such that*

$$\varphi(f) = f(e^{i\theta}).$$

Proof. Let ι denote the identity function, $\iota(e^{ix}) = e^{ix}$. Let $w = \varphi(\iota)$. For $|z| < 1$, the function $(1 - z\iota)$ is invertible and hence so is $\varphi = 1 - zw$. It follows that $|w| \leq 1$. The same argument with ι^{-1} in place of ι shows $|w| = 1$. Hence $w = e^{i\theta}$ for some θ. For a trigonometric polynomial $p(e^{ix}) = \sum_{-N}^{N} p_j e^{ijx}$, we have $\varphi(p) = p(e^{i\theta})$ and since an exercise shows polynomials are dense in $\mathscr{W}(\mathbb{T})$, the proof is complete. □

Theorem 26.10. *If $f \in \mathscr{W}(\mathbb{T})$ and if f is never 0, then $\frac{1}{f} \in \mathscr{W}(\mathbb{T})$; i.e., an element $g \in \ell^1(\mathbb{Z})$ is invertible if and only if the function $f = \sum g_j e^{ijt}$ has no zeros.*

Proof. An element $f \in \mathscr{W}(\mathbb{T})$ is invertible if and only if $\varphi(f) \neq 0$ for every homomorphism $\varphi : \mathscr{W}(\mathbb{T}) \to \mathbb{C}$. Each such homomorphism is evaluation is of the form $e^{i\theta}$ (for some real θ); i.e., $\varphi(f) = f(e^{i\theta})$ and since, by assumption, this value is not 0, the proof is complete. □

26.5. **The Weak-star topology and Gelfand Transform.** Let $\mathcal{X}$ denote a Banach space with its dual $\mathcal{X}^*$. Given $x \in \mathcal{X}$, let $p_x : \mathcal{X}^* \to \mathbb{C}$ denote the (linear) mapping $p_x(\lambda) = \lambda(x)$. The *weak-star* topology on $\mathcal{X}^*$ is the topology generated by the family of functions $\{p_x : x \in \mathcal{X}\}$. The (closed) unit ball $\mathcal{X}_1^*$ of $\mathcal{X}^*$ is the set

$$\mathcal{X}_1^* = \{\lambda \in \mathcal{X}^* : \|\lambda\| \leq 1\}.$$

Theorem 26.11 (Banach-Alaoglu)**.** *The unit ball $\mathcal{X}_1^*$ is a compact in the weak-star topology.*

Proof sketch. For $x \in \mathcal{X}$, let $D_x = \{z \in \mathbb{C} : |z| \leq \|x\|\}$ and let $P = \prod_{x \in X} D_x$ in the product topology (the topology generated by the canonical projections). The Tychonoff's Theorem says P is compact. The space P can be identified with functions $g : X \to C$ satisfying

$$|g(x)| \leq \|x\|.$$

Define $\tau : \mathcal{X}_1^* \to P$ by

$$\tau(\lambda)(x) = \lambda(x).$$

(Thus τ is simply identifies, in the canonical fashion, $\mathcal{X}_1^*$ as a subset of P.)

Evidently τ is injective.

Let f be in the (product topology) closure of $\mathcal{X}_1^* \subset P$. Given $\epsilon > 0$ and $x, y \in \mathcal{X}$ the set

$$\{g \in P : |g(x) - f(x)|, \, |g(y) - f(y)| \, |g(x+y) - f(x+y)| < \epsilon\}$$

is a neighborhood of f in the product topology. Hence there is a $\lambda \in \mathcal{X}_1^*$ in this neighborhood. Since λ is linear, it follows that

$$|f(x+y) - f(x) - f(y)| < 3\epsilon.$$

Thus f preserves sums. Similarly f preserves scalar multiplication. Hence f is linear. A similar argument shows $\|f\| \leq 1$. Hence $f \in \mathcal{X}_1^*$. Thus $R = \tau(\mathcal{X}_1^*)$ is closed in the product topology and therefore compact. Now view $\tau : \mathcal{X}_1^* \to R$. A subbasic open set in $\mathcal{X}_1^*$ takes the form $\{\lambda \in \mathcal{X}_1^* : |\lambda(x) - \lambda_0(x)| < \epsilon\}$ for some $\lambda_0 \in \mathcal{X}_1^*$ and $x \in \mathcal{X}$. Its preimage under τ^{-1} (so image under τ) is the (relatively) open set

$$\tau(\mathcal{X}_1^*) \cap \{f \in P : |f(x) - \lambda_0(x)| < \epsilon\}.$$

Hence τ^{-1} is continuous and since the continuous image of a compact set is compact, $\tau^{-1}(R) = \mathcal{X}_1^*$ is compact. □

Let $\mathscr{A}$ denote a commutative Banach algebra with unit. The *maximal ideal space* $\Sigma = \Sigma_{\mathscr{A}}$ of $\mathscr{A}$ is

$$\Sigma = \{\varphi : \mathscr{A} \to \mathbb{C} : \varphi \text{ is a homomorphism}\} \subset \mathscr{A}_1^*$$

given the relative weak-star topology.

Proposition 26.12. *Σ is a compact Hausdorff space.*

Proof. It suffices to show that Σ is closed (in the weak-star topology); i.e., if f is in the closure of Σ, then f is multiplicative and unital. Given $a, b \in \mathscr{A}$ and $\epsilon > 0$, the set

$$\{h \in \mathscr{A}^* : |h(a) - f(a)|, \, |h(b) - f(b)|, \, |h(ab) - f(ab)| < \epsilon\}$$

is a neighborhood of f. Hence it contains some $h \in \Sigma$. It follows that

$$|f(ab) - f(a)f(b)| \leq |f(ab) - h(ab)| + |h(a)| \, |h(b) - f(b)| + |f(b)| \, |f(a) - h(a)| < (1 + \|a\| + \|b\|)\epsilon.$$

Thus f is multiplicative.

A proof that Σ is Hausdorff is left to the gentle reader. □

Proposition 26.13. *Given $a \in \mathscr{A}$ the function $\hat{a} : \Sigma \to \mathbb{C}$ defined by $\hat{a}(\varphi) = \varphi(a)$ is continuous. Moreover, the mapping $\hat{}: \mathscr{A} \to C(\Sigma)$ is a continuous unital homomorphism of norm one. In particular,*

$$\widehat{p(a)}(\varphi) = \varphi(p(a)) = p(\varphi(a))$$

for $\varphi \in \Sigma$.

The mapping $\hat{}$ is the *Gelfand Transform.* From here on it will be denoted by $\mathscr{G}$.

Proof. The weak-star topology is defined exactly to make the $\mathscr{G}(a)$ continuous. Since $\varphi \in \Sigma$ has norm one, $\|\mathscr{G}(a)\| \le \|a\|$. On the other hand, $\mathscr{G}(1)$ is the function that is identically equal to 1 and thus $\|\mathscr{G}(1)\| = 1$. Thus $\|\mathscr{G}\| = 1$. An easy exercise shows $\mathscr{G}$ is multiplicative and hence a unital homomorphism. □

Proposition 26.14. *For $a \in \mathscr{A}$,*

$$\sigma(a) = \{\varphi(a) : \varphi \in \Sigma\}.$$

Moreover, if $\mathscr{A}$ is generated by a, then $\mathscr{G}(a) : \Sigma \to \sigma(a)$ is a homeomorphism.

Proof. The function $\mathscr{G}(a)$ is continuous with compact domain. It is onto, since $\sigma(a) = \{\varphi(a) : \varphi \in \Sigma\}$. Thus $\mathscr{G}(a)$ is a homeomorphism. □

26.6. **The spectral radius formula.** Given $\mathscr{A}$ a unital Banach algebra and $a \in \mathscr{A}$, the spectral radius of a is

$$r(a) = \max\{|z| : z \in \sigma(a)\}.$$

Proposition 26.15. $r(a) = \lim_{n\to\infty} \|a^n\|^{\frac{1}{n}}$.

Proof. First observe, for $m \in \mathbb{N}^+$ and $z \in \mathbb{C}$,

$$z^{m+1} - a^{m+1} = (z-a)\sum_{j=0}^{m} z^j a^{m-j}.$$

Thus, if $z - a$ is not invertible, then neither is $z^{m+1} - a^{m+1}$. Consequently, $\|z^{m+1}\| \le \|a^{m+1}\|$ and therefore $|z| \le \liminf \|a^n\|^{\frac{1}{n}}$. Hence, if $z \in \sigma(a)$, then $|z| \le \liminf \|a^n\|^{\frac{1}{n}}$.

Given $\rho \in \mathscr{A}^*$, consider the function

$$g(z) = \rho((1 - za)^{-1})$$

analytic on the set $\{z \in \mathbb{C} : |z| < \frac{1}{r(a)}\}$. For $|z| < \frac{1}{\|a\|}$, the function g has the power series expansion,

$$g(z) = \sum \rho(a^n) z^n.$$

Hence this series has radius of convergence at least $\frac{1}{r(a)}$ and hence, if $|z| < \frac{1}{r(a)}$, then the set $\{\rho(a^n)z^n\}$ is bounded. Fix $|z| < \frac{1}{r(a)}$ and viewing $f_n = z^n a^n \in \mathscr{A}$ as elements of $\mathscr{A}^{**}$ so that $f_n(\rho) = \rho(f_n) = z^n \rho(a^n)$ for $\rho \in \mathscr{A}^*$. For each $\rho \in \mathscr{A}^*$ there is a constant C_ρ such that

$$|f_n(\rho)| = |\rho(f_n)| \le C_\rho.$$

Thus (f_n) is a pointwise bounded set of linear functionals on the Banach space $\mathscr{A}^*$. By the principle of uniform boundedness, there is a constant C such that $\|f_n\| \leq C$ independent of n. Thus $\|z^n a^n\| \leq C$ and therefore $\limsup \|a^n\|^{\frac{1}{n}} \leq \frac{1}{|z|}$. Hence, by choosing z such that $|z| < \frac{1}{r(a)}$ is close to $\frac{1}{r(a)}$,

$$\limsup \|a^n\|^{\frac{1}{n}} \leq r(a).$$

□

Corollary 26.16. *Suppose $\mathscr{A}$ is a commutative Banach algebra with unit and maximal ideal space Σ. Given $a \in \mathscr{A}$,*

$$\|\mathscr{G}(a)\| = \lim \|a_n\|^{\frac{1}{n}} = r(a).$$

Proof. Recall $\mathscr{G}(a)$ is the function on Σ defined by $\mathscr{G}(a)(h) = h(a)$. Hence $\|\mathscr{G}(a)\| = \max\{|h(a)| : h \in \Sigma\} = r(a)$. □

27. C-star algebras and the functional calculus

A C-star algebra $\mathscr{A}$ is Banach algebra with norm $\|\cdot\|$ and an involution $\mathscr{A} \to \mathscr{A}$, denoted $a \to a^*$, such that for $a, b \in \mathscr{A}$ and $c \in \mathbb{C}$,

(i) $(a^*)^* = a$ (involutive);
(ii) $(ab)^* = b^* a^*$ (anti-multiplicative);
(iii) $(ca + b)^* = \overline{c} a^* + b^* = c^* a^* + b^*$ (anti-linear); and
(iv)
$$\|a^* a\| = \|a\|^2$$

In the case $\mathscr{A}$ has a unit 1, as with Banach algebras, we assume $\|1\| = 1$. Note $1^* = 1^* 1 = (1^* 1)^{**} = (1^* 1)^* = (1^*)^* = 1 = 1$.

For X a compact metric space (or just a compact Hausdorff space), $C(X)$ is a C-star algebra given the usual norm and involution $f^*(z) = \overline{f(z)} = f(z)^*$. For H a (complex) Hilbert space, $B(H)$, the bounded operators on H, is a C-star algebra given the usual norm and involution $T \mapsto T^*$ given by the map taking an operator T to its adjoint,

$$\langle Tx, y \rangle = \langle x, T^* y \rangle, \quad x, y \in H.$$

An element $a \in \mathscr{A}$ is *hermitian* if $a^* = a$.

Proposition 27.1. *If a is hermitian, then $\|a\| = r(a)$, the spectral radius of a.*

Proof. Since $a = a^*$, we have $\|a^2\| = \|a^* a\| = \|a\|^2$. By induction $\|a^{2n}\| = \|a\|^{2n}$. Hence, by the spectral radius formula,

$$\|a\| = \lim \|a^{2n}\|^{\frac{1}{2n}} = r(a).$$

□

An element $u \in \mathscr{A}$ is an *isometry* if $u^* u = 1$.

Proposition 27.2. *Suppose $\mathscr{A}$ is a C-star algebra with unit and $h : \mathscr{A} \to \mathbb{C}$ is a unital homomorphism.*

(1) if $a = a^*$, then $h(a) \in \mathbb{R}$;
(2) $h(a^*) = h(a)^*$;
(3) $h(a^* a) \geq 0$;
(4) $|h(u)| = 1$ if u is an isometry.

In particular, if a is hermitian, then $\sigma(a) \subset \mathbb{R}$; and any homomorphism $h : \mathscr{A} \to \mathbb{C}$ is actually a $*$-homomorphism.

Proof. Recall the norm of a unital homomorphism of a Banach algebra is 1. (Identify it with the mapping $\pi : \mathscr{A} \to \mathscr{A} / \ker(h) = \mathbb{C}$.) Assuming $a = a^*$, for $t \in R$,

$$|h(a + it)|^2 \leq \|a + it\|^2 = \|(a + it)^*(a + it)\| = \|a^2 + t^2\| \leq \|a\|^2 + t^2.$$

Writing $h(a) = \alpha + i\beta$,

$$\alpha^2 + (\beta + t)^2 \leq \|a\|^2 + t^2.$$

Hence $\beta = 0$.

Write $a = x + iy$ where x, y are hermitian (the *cartesian decomposition*) and apply the first part of the proposition.

For the third part, $h(a^*a) = h(a^*)h(a) = h(a)^*h(a) = |h(a)|^2 \geq 0$.

Finally $1 = h(1) = h(u^*u) = h(u^*)h(u) = h(u)^*h(u) = 1$. Thus $|h(u)| = 1$. □

Theorem 27.3. *If $\mathscr{A}$ is a commutative C-star algebra with unit and maximal ideal space Σ, then the gelfand transform $\mathscr{G} : \mathscr{A} \to C(\Sigma)$ is an isometric $*$-isomorphism; i.e., $\mathscr{A}$ and $C(\Sigma)$ are equal as C-star algebras.*

Proof. If $a \in \mathscr{A}$ is hermitian, then $\|a\| = r(a) = \|\mathscr{G}(a)\|$.

For $a \in \mathscr{A}$ and $\varphi \in \Sigma$,

$$\mathscr{G}(a^*)(\varphi) = \varphi(a^*) = \varphi(a)^* = \overline{\mathscr{G}(a)(\varphi)}.$$

Thus $\mathscr{G}(a^*) = \mathscr{G}(a)^*$ and $\mathscr{G}$ is a $*$-homomorphism. Further, given $a \in \mathscr{A}$, since a^*a is hermitian and $\mathscr{G}$ is a $*$-homomorphism,

$$\|a\|^2 = \|a^*a\| = \|\mathscr{G}(a^*a)\| = \||\mathscr{G}(a)|^2\| = \|\mathscr{G}(a)\|^2.$$

Hence $\|a\| = \|\mathscr{G}(a)\|$ and $\mathscr{G}$ is an isometry and thus has closed range. Therefore to show $\mathscr{G}$ is onto, it suffices to show it has dense range. Now $\widehat{\mathscr{A}}$ is a subalgebra of $C(\Sigma)$ that is closed under complex conjugation (the involution), contains the constants and separates points. Hence by Stone-Weierstrass it is dense. Thus $\mathscr{G}$ is a bijective isometric $*$-isomorphism. □

27.1. **Normal elements and the functional calculus.** Given an element a in a C-star algebra $\mathscr{A}$, the C-star algebra generated by a, again denoted $C^*(a)$, is the smallest C-star subalgebra of $\mathscr{A}$ containing a.

Proposition 27.4 (Spectral Permanence)**.** *Let $\mathscr{A}$ be a unital C-star algebra. An $a \in \mathscr{A}$ is invertible in $\mathscr{A}$ if and only if it is invertible in $C^*(a)$. In particular, if $\mathscr{B} \subset \mathscr{A}$ is a unital C-star algebra and $a \in \mathscr{B}$, then*

$$\sigma_{\mathscr{B}}(a) = \sigma_{\mathscr{A}}(a).$$

Thus the spectrum of a does not depend upon the C-star algebra.

Proof. First suppose a is hermitian and invertible in $\mathscr{A}$. In this case a is invertible in the commutative C-star algebra $\mathscr{D}$ generated by a and a^{-1}, namely the closure of expressions of the form $r(a) = \sum_{j=-N}^{N} r_j a^{-j}$. The Gelfand transform $\mathscr{G}_{\mathscr{D}} : \mathscr{D} \to \Sigma_{\mathscr{D}}$ is an isometric $*$-isomorphism. Suppose $\varphi, \psi \in \Sigma_{\mathscr{D}}$ and $\varphi(a) = \psi(a)$. Since a is invertible this common value is not zero and moreover $\varphi(a^{-1}) = \varphi(a)^{-1} = \psi(a)^{-1} = \psi(a^{-1})$. Hence $\varphi = \psi$. Therefore the subalgebra $\mathscr{G}_{\mathscr{D}}(C^*(a)) \subset C(\Sigma_{\mathscr{D}})$ separates points. It is also contains the constants, is closed under pointwise conjugation and closed. Hence, by the Stone-Weierstrass Theorem, $\mathscr{G}_{\mathscr{D}}(C^*(a)) = C(\Sigma_{\mathscr{D}})$ and therefore $C^*(a) = \mathscr{D}$. Hence a is invertible in $C^*(a)$.

For the general case, if a is invertible in $\mathscr{A}$ with inverse $b \in \mathscr{A}$, then a^*a is invertible in $\mathscr{A}$ with inverse bb^*. Thus a^*a is invertible in $C^*(a^*a)$; i.e., $bb^* \in C^*(a^*a)$. Finally, $b = b1 = b(b^*a^*) = (bb^*)a \in C^*(a)$. ☐

An element a in a C-star algebra $\mathscr{A}$ is *normal* if $a^*a = aa^*$. Assuming a is normal, given a polynomial $p(z, \overline{z}) = \sum p_{j,k} z^j \overline{z}^k$, let

$$p(a, a^*) = \sum p_{j,k} a^j a^{*k}$$

(since a and a^* commute the order of the products are not important). In this case, the C-star algebra generated by a is the closure of $\{p(a, a*) : p = p(z, \overline{z}) \text{ a polynomial}\}$. It is commutative.

Proposition 27.5. *Let $\mathscr{A} = C^*(a)$ is a C-star algebra generated by a normal element a and let Σ denote its maximal ideal space. The mapping $\mathscr{G}(a) : \Sigma \to \sigma(a)$ is a homeomorphism. Moreover, if $p = p(z, \overline{z})$ is a polynomial, then $\mathscr{G}(p(a, a^*)) = p(\mathscr{G}(a), \mathscr{G}(a)^*)$.*

Lemma 27.6. *If X and Y are compact Hausdorff spaces and $\tau : X \to Y$ is a homeomorphism, then $\tau^* : C(Y) \to C(X)$ defined by $\tau^*(f)(x) = f(\tau(x))$ is an isometric $*$-isomorphism.*

Given $a \in \mathscr{A}$ normal, choosing $\tau = \mathscr{G}(a)$, note that $\tau^*(f)(\varphi) = f(\tau(\varphi)) = f(\varphi(a))$. Hence we obtain an isometric $*$-isomorphism $\rho = \mathscr{G}^{-1} \circ \tau^* : C(\sigma(a)) \to C^*(a) \subset \mathscr{A}$. This mapping is the *functional calculus* for a. Note that, for a polynomial $p = p(z, \overline{z})$ viewed as an element of $C(\sigma(a))$, that $\rho(p) = p(a, a^*)$ and moreover ρ is the unique isometric $*$-homomorphism from $C(\sigma(a)) \to \mathscr{A}$ sending p to $p(a, a^*)$ (since polynomials in z and $z^* = \overline{z}$ are dense in $C(\sigma(a))$ by Stone-Weierstrass). It is customary to write $f(a) = \rho(f)$.

Proposition 27.7 (Spectral mapping)**.** *Suppose $a \in \mathscr{A}$ is normal. For $f \in C(\sigma(a))$,*

$$\sigma(f(a)) = f(\sigma(a))$$

and $\|f(a)\| = \max\{|f(z)| : z \in \sigma(a)\}$.

Proof. Since ρ is a $*$-isomorphism, $f - \lambda$ is invertible if and only if $\rho(f - \lambda) = \rho(f) - \lambda$ is invertible. Hence $f(\sigma(a)) = \sigma(f) = \sigma(f(a))$. ☐

An element $u \in \mathscr{A}$ is *unitary* if $u^*u = uu^* = 1$. In particular u is invertible and normal.

Proposition 27.8. *If u is unitary, then $\sigma(u) \subset \mathbb{T} = \{|z| = 1\}$. In particular, if p is an analytic polynomial $p(z) = \sum p_j z^j$, then*

$$\|p(u)\| \leq \{|p(z)| : z \in \mathbb{T}\} = \sup\{|f(z)| : |z| < 1\}.$$

Proof. Let Σ denote the maximal ideal space of $C^*(u)$. In particular each $\varphi \in \Sigma$ is a $*$-homomorphism. Thus $1 = \varphi(1) = \varphi(u^*u) = |\varphi(u)|^2$. Hence $|\varphi(u)| = 1$ and the proof of the first part is complete. (Or appeal to the earlier result about isometric elements in a C-star algebra.)

The second part (the equality) is simply an instance of the maximum principle. ☐

27.2. **Positive and contractive elements in a C-star algebra.** An element a in a C-star algebra with unit $\mathscr{A}$ is *positive (semidefinite)*, denoted $a \succeq 0$ if $a = a^*$ and $\sigma(a) \subset [0, \infty)$. For instance, if X is a compact Hausdorff space, then an element $f \in C(X)$ is positive if and only if the range of f lies in $[0, \infty)$. Moreover, if $\rho : C(X) \to \mathscr{A}$ is a $*$-homomorphism and $f \succeq 0$, then $\rho(f) \succeq 0$; and if ρ is an isometric $*$-isomorphism, then $\rho(f) \succeq 0$ if and only if $f \succeq 0$.

Lemma 27.9. *Suppose $\mathscr{A}$ is a unital C-star algebra and $a \in \mathscr{A}$ is hermitian.*

(i) If $\sigma(a) = \{0\}$, then $a = 0$;
(ii) If $\pm a \succeq 0$, then $a = 0$;
(iii) If $a^2 = 0$, then $a = 0$;
(iv) a^2 is positive;
(v) There exists positive elements $p, q \in C^*(a)$ such that $pq = 0$ and $q = p^2 - q^2$. In particular, if a is positive, then $q = 0$;
(vi) If a is postive, then each $n \in \mathbb{N}^+$ there exists a positive b such that $b^n = a$;
(vii) If $\|1 - a\| \leq 1$, then a is positive;
(viii) If a is positive, then $\|t - a\| \leq t$ for all $t \geq \|a\|$.

Proof. Since a is hermitian, $\|a\| = r(a)$, the spectral radius. Hence, if $\sigma(a) = \{0\}$, then $a = 0$. If $\pm a \succeq 0$, then $\pm\sigma(a) \subset [0, \infty)$ in which case $\sigma(a) = \{0\}$ and hence $a = 0$. If $a^2 = 0$, then $\sigma(a^2) = \{0\}$ and hence, using spectral mapping $\sigma(a) = \sigma(a)^2 = \{0\}$. Thus $a = 0$. Since a is hermitian, $\sigma(a) \subset \mathbb{R}$. Thus using the functional calculus, $\sigma(a^2) = \sigma(a)^2 \subset [0, \infty)$. Hence a is positive.

Let $\rho : C(\sigma(a)) \mapsto C^*(a)$ denote the functional calculus. Consider the functions $u, v : \mathbb{R} \to \mathbb{R}$ defined by $u(t) = \max\{0, t\}$ and $v = -\min\{0, t\}$. Thus $u - v = \iota$ is the identity function $\iota(t) = t$ and both u, v are positive functions. Hence $a = \rho(\iota) = \rho(u) - \rho(v)$. Letting $u^{\frac{1}{2}}$ and $v^{\frac{1}{2}}$ denote the (pointwise) positive square roots and $p = \rho(u^{\frac{1}{2}})$ and $q = \rho(v^{\frac{1}{2}})$, it follows that $a = p^2 - q^2$ and $pq = 0$. If a is positive, then $v = 0$ and hence $q = 0$. In this case, for $n \in \mathbb{N}^+$, we have $a = \rho(iota^{\frac{1}{n}})^n$.

Using again the functional calculus, $\rho(\iota) = a$, observe that $\|1 - \iota\| \leq 1$ implies $|1 - \iota| \leq 1$ pointwise. Hence the range$(\iota) \subset [0, 2]$ and therefore ι, and hence a, is positive. Finally, if a is positive, then, for $t \geq \|a\|$, pointwise $|\iota| \leq \|a\| \leq t$ and therefore $0 \leq t - \iota \leq t$. Hence $\|t - \iota\| \leq t$. □

Let $\mathscr{A}_+$ denote the positive elements in the unital C-star algebra $\mathscr{A}$.

Proposition 27.10. *The set $\mathscr{A}_+$ is a closed cone; i.e., it is a closed set that is closed under addition and multiplication by nonnegative scalars.*

Proof. That $\mathscr{A}_+$ is closed under multiplication by a $\lambda \geq 0$ is evident. Now suppose $a, b \in \mathscr{A}_+$. By multiplying by a $\lambda > 0$, we may assume that $\|a\|, \|b\| \leq 1$. Thus,

$$\|2 - (a + b)\| \leq [\|1 - a\| + \|1 - b\|] \leq 2$$

by the last part of the lemma. Hence, by the next to last part of the lemma, $\frac{a+b}{2}$, and hence, a is positive.

To prove $\mathscr{A}_+$ is closed, use, $a = a^*$, the fact that $\| \|a\| - a\| \le \|a\|$ if and only if $a \succeq 0$. □

Proposition 27.11. *For $a \in \mathscr{A}$, a unital C-star algebra, the following are equivalent.*

(i) a is positive;
(ii) there is a positive $b \in \mathscr{A}$ such that $a = b^2$;
(iii) there is a a hermitian $b \in \mathscr{A}$ such that $a = b^2$;
(iv) there is an $x \in \mathscr{A}$ such that $a = x^*x$;

Proof. The previous lemma shows (i) and (ii) are equivalent. Evidently (ii) implies (iii) implies (iv).

If (iv) holds with $a = x^*x$, then $a = a^*$ and, from the lemma, there exists $p, q \in C^*(a)$ such that $p, q \succeq 0$, $pq = 0$ and $a = p^2 - q^2$. Let $b + ic$ denote the cartesian decomposition of xq. Thus,

$$b^2 + c^2 + i(bc - cb) = (b - ic)(b + ic) = qx^*xq = qaq = q(p^2 - q^2)q = -q^4.$$

It follows that $\pm q^4 \succ 0$ and therefore $q^4 = 0$. Hence $q = 0$. □

Proposition 27.12. *Let H be a Hilbert space. An element $P \in B(H)$ is positive if and only if $\langle Px, x\rangle \ge 0$ for all $x \in H$.*

(Caution: the condition $P = P^*$ needs to be added in the case of real Hilbert space.)

Proof. If P is positive, then $P = X^*X$ for some $X \in B(H)$. Hence, for $x \in H$,

$$\langle Px, x\rangle = \langle X^*Xx, x\rangle = \langle Xx, Xx\rangle \ge 0.$$

Conversely, suppose $\langle Px, x\rangle \ge 0$ for all $x \in H$. It follow that

$$\langle (P - P^*)x, x\rangle = \langle Px, x\rangle - \langle x, Px\rangle = 0$$

since both terms are real. Hence $P = P^*$ and in particular $\sigma(P) \subset \mathbb{R}$. Now suppose $\lambda > 0$. For $x \in X$,

$$\|(P + \lambda)x\|^2 = \|Px\|^2 + 2\langle Px, x\rangle + \lambda^2\|x\|^2 \ge \lambda^2\|x\|^2.$$

It follows that $P + \lambda$ has no kernel and the range of $P + \lambda$. Thus $\text{range}(P + \lambda) = \ker((P + \lambda)^*)^\perp = \ker(P + \lambda)^\perp = H$. By the Banach isomorphism theorem, $P + \lambda$ is invertible. Hence $\sigma(P) \subset [0, \infty)$. □

27.3. **Contractions on Hilbert space and the de Branges-Rovnyak construction.** Let H denote a Hilbert space (over $\mathbb{C}$). An operator $C \in B(H)$ is a *contraction* if $\|C\| \le 1$.

Proposition 27.13. *C is a contraction if and only if $I - C^*C \succeq 0$.*

Proof. We have $I - C^*C \succeq 0$ if and only if for all $x \in H$,

$$0 \le \langle (I - C^*C)x, x\rangle = \langle x, x\rangle - \langle Cx, Cx\rangle = \|x\|^2 - \|Cx\|^2$$

if and only if $\|C\| \le 1$. □

The operator $I - C^*C$ being positive (semidefinite) has a positive square root $(1 - C^*C)^{\frac{1}{2}}$ denoted D_C and referred to as the *defect* of T.

An operator T on a Hilbert space H *lifts* to an operator J on a Hilbert space K if there exist an isometry $V : H \to K$ such that

$$VT = JV.$$

Note, in this case, if $p \in \mathbb{C}[z]$ is a polynomial, then

$$Vp(T) = p(J)V.$$

Hence sometimes a lift is referred to as a $\mathbb{C}[z]$-lift. An operator T on a Hilbert space H is a C_0-contraction if $\|T\| \le 1$ and $\lim T^n h = 0$ for each $h \in H$. The adjoint S^* of the shift operator S is an example of a C_0-contraction.

Given a Hilbert space E, let $K_+(E) = \oplus_{j=0}^{\infty} E = \ell^2_E(\mathbb{N})$ consisting of those sequences $\{(e_j) = \oplus_{j=0}^{\infty} e_j : \sum \|e_j\|^2 < \infty\}$ with the inner product,

$$\langle e, f\rangle = \langle (e_j), (f_k)\rangle = \sum \langle e_j, f_j\rangle.$$

It is an exercise to verify that K_+ is a Hilbert space. The corresponding *shift operator* S_E is defined by $(S_E e)_j = 0$ if $j = 0$ and $(Se)_j = e_{j-1}$ for $j > 0$. The *multiplicity* of S_E is the dimension of E and any two shifts of the same multiplicity are unitarily equivalent.

Proposition 27.14. *The operator S_E is an isometry with $\sigma(S_E) = \overline{\mathbb{D}}$. Its adjoint is given by*

$$S_E^* \oplus e_j = \oplus e_{j+1}.$$

In particular, S_E^* has kernel $\{e : e_j = 0, j > 0\}$ (which we identify with E) and is a C_0-contraction.

Proof. We will prove the proposition for the case of the shift S of multiplicity one where $E = \mathbb{C}$. The general case is both similar and easily derived from this case. We already know S is an isometry and in particular, $\|S\| = 1$. Hence $\sigma(S) \subset \mathbb{D}$. Likewise we have already seen that S^* is the *backward shift*. Given $\lambda \in \mathbb{D}$, let $s_\lambda = (\overline{\lambda}^j)_{j=0}^{\infty} \in \ell^2(\mathbb{N})$ and observe that $S^* s_\lambda = (\overline{\lambda}^{j+1})_{j=0}^{\infty} = \overline{\lambda} s_\lambda$. Hence $S^* - \overline{\lambda}$ has a non-trivial kernel and is thus not invertible. Hence $\lambda \in \sigma(S)$ and therefore $\mathbb{D} \subset \sigma(S) \subset \overline{\mathbb{D}}$. Since the spectrum is closed, the result follows. □

An operator T is a *shift operator* if it is unitarily equivalent to S_E for some E.

Theorem 27.15. *An operator T is a C_0-contraction if and only if T lifts to the adjoint of a shift operator.*

Proof. It is straightforward to verify if T lifts to the adjoint of a shift operator, then T is a C_0-contraction.

The proof of the converse is constructive - the de Branges-Rovnyak construction (a variation of a construction of Rota). Suppose T is a C_0-contraction. Let E denote the closure of the range of D_T, the defect of T. Define $V : H \to \ell^2_E(\mathbb{N})$ by

$$Vh = \oplus D_T T^j h.$$

A computation shows V is an isometry. Moreover,

$$S_E^* V h = S_E^* \oplus_{j=0}^{\infty} D_T T^j h = \oplus D_T T^{j+1} h = VTh.$$

Hence S_E^* lifts T. $\square$

Given a Hilbert space E, let $K = \ell^2_E(\mathbb{Z}) = \oplus_{j=-\infty}^{\infty} E$ and define $U_E : K \to K$ by

$$U \oplus e_j = e_{j+1}.$$

A computation shows $U^* = U^{-1}$ is given by

$$U^* \oplus e_j = e_{j-1}.$$

In particular U is a unitary operator. An operator T is a *bilateral shift* if there is a Hilbert space E such that T is unitarily equivalent to U_E.

Proposition 27.16. *The shift operator S_E lifts to the bilateral shift operator U_E. In particular, a shift operator lifts to a bilateral shift (unitary operator).*

An operator $T \in B(H)$ dilates to an operator $J \in B(K)$ or has a $\mathbb{C}[z]$-*dilation* to J if there is an isometry $V : H \to K$ such that

$$p(T) = V^* p(J) V,$$

for all $p \in \mathbb{C}[z]$. Note that if T lifts to J, then T dilates to J and T^* dilates to J^*.

Proposition 27.17. *If $C \in B(H)$ is a C_0-contraction, then C dilates to a bilateral shift.*

Proof. There is a Hilbert space E and isometry $V : H \to \ell^2_E(\mathbb{N})$ such that $Vp(C) = p(S_E^*)V$ for $p \in \mathbb{C}[z]$. Hence $q(C^*) = V^* q(S_E) V$ for all $q \in \mathbb{C}[z]$. The inclusion $\iota : \ell^2_E(\mathbb{N}) \to \ell^2_E(\mathbb{Z})$ is an isometry and $q(S_E) = \iota^* q(U_E)\iota$ for all polynomials $q \in \mathbb{C}[z]$. Hence $q(C^*) = W^* q(U_E) W$ where $W = \iota V$. It now follows that $p(C) = W^* p(U_E^*) W$ for all polynomials $p \in \mathbb{C}[z]$. $\square$

Remark 27.18. The Sz.-Nagy dilation theorem is the assertion that a contraction operator dilates to a unitary operator. In particular, the hypothesis and conclusion of the Nagy dilation theorem are both slightly different than that of Proposition 27.17. It has a relatively simple geometric proof. $\diamond$

Theorem 27.19 (The von Neumann inequality)**.** *If $C \in B(H)$ is a contraction and $p(z) = \sum_{j=0}^N p_j z^j$ is a polynomial, then $\|p(C)\| \le \|p\|_{\infty,\mathbb{D}}$.*

Proof. By the Sz.-Nagy dilation theorem there is a Hilbert space K and isometry V such that for all polynomials $q \in \mathbb{C}[z]$,

$$q(C) = V^* q(U) V.$$

By Proposition 27.8, $\|p(U)\| \le \|p\|_{\infty,\mathbb{D}}$. Thus,

$$\|p(C)\| \le \|V^*\| \, \|p(U)\| \, \|V\| \le \|p\|_{\infty,\mathbb{D}}.$$

$\square$

27.4. **Spectral measures.** Let H denote a Hilbert space (over $\mathbb{C}$). For $h \in H$, let $r_h : B(H) \to [0, \infty)$ denote the function $r_h(T) = \|Th\|$. The *strong operator topology (SOT)* on H is the topology generated by the functions $\{r_h : h \in H\}$. In particular, a sequence (T_n) from $B(H)$ SOT converges to $T \in B(H)$ if and only if $\lim_{n\to\infty} \|(T - T_n)h\| = 0$ for all $h \in H$. In particular, T is a C_0-contraction if and only if $\|T\| \le 1$ and (T^n) SOT-converges to 0.

An $B(H)$-valued *spectral measure* on a measurable space (X, Σ) is a mapping $E : \Sigma \to B(H)$ such that

(i) $E(X) = I$, the identity mapping;
(ii) $E(\emptyset) = 0$;
(iii) E is projection valued meaning if $\sigma \in \Sigma$, then $E(\sigma)$ is a projection;
(iv) if $(\omega_j)_{j=1}^{\infty}$ is a disjoint sequence of sets from Σ, then the series

$$\sum_{j=1}^{\infty} E(\omega_j)$$

SOT-converges to $E(\cup\omega_j)$.

Given self-adjoint (hermitian) operators P, R on a Hilbert space H the notation $P \preceq R$ means $R - P \succeq 0$. This partial order on the self-adjoint elements of $B(H)$ is the *Loewner* ordering.

Lemma 27.20. *Suppose $P, Q \in B(H)$. If P, Q and $P + Q$ are all projections, then $PQ = QP = 0$.*

Proof. The equality $P + Q = (P + Q)^2$ implies $PQ + QP = 0$. Multiplying this last equation on both sides by Q gives, $0 = 2QPQ = 2QPPQ = 2Y^*Y$, where $Y = PQ$. Thus $PQ = Y = 0$. □

Proposition 27.21. *Suppose E is an $B(H)$-valued spectral measure on (X, Σ) and $\omega, \rho \in \Sigma$.*

(i) if $\omega \cap \rho = \emptyset$, then $E(\omega)\,E(\rho) = 0$;
(ii) if $\omega \subset \rho$, then

$$E(\rho) = E(\rho \setminus \omega) + E(\omega) \tag{89}$$

and therefore
(a) $E(\omega) \preceq E(\rho)$; and
(b) $E(\omega)E(\rho) = E(\omega)$;
(iii) $E(\omega \cap \rho) = E(\omega)E(\rho) = E(\rho)E(\omega)$.

Proof. Letting $P = E(\omega)$ and $Q = E(\rho)$ by additivity of E, all of P, Q and $P + Q$ are projections. Hence $PQ = 0$. Item (ii)(a) follows immediately from (i). For (ii)(b), multiply equation (89) on the right by $E(\omega)$ and apply (i).

To prove (iii) observe

$$E(\omega \cup \rho) + E(\omega \cap \rho) = E(\omega) + E(\rho), \tag{90}$$

an identity that follows from

$$\begin{aligned} A\cup B =& (A\setminus B)\cup(B\setminus A)\cup(A\cap B)\\ A =& (A\setminus B)\cup(A\cap B)\\ B =& (B\setminus A)\cup(A\cap B). \end{aligned}$$

Multiplying equation equation (90) on the right by $E(\omega)$ gives

$$E(\omega\cup\rho)E(\omega)+E(\omega\cap\rho)E(\omega)=E(\omega)^2+E(\rho)E(\omega).$$

Using (ii)(b) on each term on the left hand side, gives

$$E(\omega)+E(\omega\cap\rho)=E(\omega)+E(\rho)E(\omega)$$

from which the result follows. □

Example 27.22. Let N be a normal matrix of size n. From undergraduate linear algebra, N is unitarily equivalent to a diagonal matrix; i.e., there exists a unitary U and a diagonal matrix $D=\operatorname{diag}(\lambda_1,\dots,\lambda_n)$ such that $N=UDU^*$. Letting u_j denote the columns of u

$$N=\sum\lambda_j u_j u_j^*$$

and of course $\sigma(N)=\{\lambda_1,\dots,\lambda_n\}$. Letting P_j denote the rank one matrix $u_ju_j^*$. Since U is unitary, each $u_ju_j^*$ is a projection and moreover, $P_jP_k=0$ for $j\neq k$ Letting E denote the mapping on the Borel sets of $\mathbb{C}$ defined by

$$E(\omega)=\sum_{\lambda_j\in\omega}u_ju_j^*=\sum_{\lambda_j\in\omega}P_j$$

gives a spectral measure and

$$N=\sum\lambda_k E(\{\lambda_k\}).$$

Moreover, if f is a polynomial, then

$$f(N)=\sum f(\lambda_j)P_j.$$

△

Example 27.23. Let $(X,\mathscr{M},\mu)$ denote a measure space. Define a $L^2(\mu)$-valued spectral measure E on $\mathscr{M}$ as follows. Given $A\in\mathscr{M}$, let $E(A):L^2(\mu)\to L^2(\mu)$ denote the mapping defined by $E(A)f=\mathbf{1}_Af$. Given (A_j) a disjoint sequence from $\mathscr{M}$ and $f\in L^2(\mu)$, observe that

$$\sum_{j=1}^N E(A_j)f=E(\cup_{j=1}^N A_j)f=\sum\mathbf{1}_{\cup^N A_j}f$$

converges to $\mathbf{1}_{\cup A_j}f$ by the dominated convergence theorem. Hence $\sum E(A_j)$ converges SOT to $E(\cup A_j)$. △

Proposition 27.24. *If E is an $B(H)$-valued spectral measure on (X,Σ) and if $h,k\in H$, then $\mu_{h,k}:\Sigma\to\mathbb{C}$ defined by $\mu_{h,k}(\omega)=\langle E(\omega)h,k\rangle$ is a measure with total variation $|\mu_{h,k}|(X):=\|\mu_{h,k}\|\leq\|h\|\,\|k\|$.*

Part of proof. Let $\Pi = \{\omega_1, \dots, \omega_n\}$ be a given measurable partition of X. For each j there is a unimodular $\gamma_j \in \mathbb{C}$ such that

$$\gamma_j \langle E(\omega_j)h, k\rangle \geq 0.$$

Hence,

$$\begin{aligned}\sum |\langle E(\omega_j)h, k\rangle| =& \sum \gamma_j \langle E(\omega_j)h, k\rangle \\ =& \langle \sum E(\omega_j)\gamma_j h, k\rangle \\ =& |\langle \sum E(\omega_j)\gamma_j h, k\rangle| \\ \leq& \| \sum E(\omega_j)\gamma_j h\| \, \|k\|.\end{aligned}$$

Now the vectors $h'_j = E(\omega_j)\gamma_j h$ and $h''_j = E(\omega_j)h$ are pairwise orthogonal (since the ω_j are disjoint). Hence,

$$\| \sum E(\omega_j)\gamma_j h\|^2 = \sum \|E(\omega_j)\gamma_j h\|^2 = \sum \|E(\omega_j)h\|^2 = \| \sum E(\omega_j)h\|^2 = \|h\|^2.$$

Putting things together gives,

$$\sum |\mu_{h,k}(\omega_j)| = \sum |\langle E(\omega_j)h, k\rangle| \leq \|h\| \, \|k\|$$

and the result follows by the definition of total variation norm. □

Spectral measures share many properties of measures and we will use these properties without (or with little) comment in what follows. As an example, we can assume if $E(\rho) = 0$, then Σ contains all subsets of ρ. A measurable function $f : X \to \mathbb{C}$ is *essentially bounded* (with respect to E) if there is a C such that $E(\{|f(x)| > C\}) = 0$ and in this case $\|f\|_\infty$ is the infimum of such C. The collection of such function, $L^\infty(E)$, is a C-star algebra. Given a measurable space $(X, \mathscr{M})$, let $\mathbb{B}(X) = \mathbb{B}(X, \mathscr{M})$ denote the bounded ($\mathbb{C}$-valued) functions on X. It is a C-star algebra (under the usual pointwise operations and involution given by pointwise conjugation) and the supremum norm.

Proposition 27.25. *If E is an $B(H)$-valued spectral measure on $(X, \mathscr{M})$ and $f \in \mathbb{B}(X)$, then there is a unique operator $I(f) \in B(H)$ such that for every $\epsilon > 0$, each measurable partition $\{A_1, \dots, A_n\}$ of X with $\sup\{|f(x) - f(y)| : x, y \in A_j\} < \epsilon$ for each j, and each choice of $x_j \in A_j$,*

$$\|I(f) - \sum f(x_j)E(A_j)\| < \epsilon.$$

Moreover, if $A \in \mathscr{M}$, then $I(\mathbf{1}_A) = E(A)$.

A mapping $J : H \times H \to \mathbb{C}$ is *sesquilinear* if, for all $g, h, k \in H$ and $c \in \mathbb{C}$, $g, h, k \in H$ and $c \in \mathbb{C}$,

$$\begin{aligned}J(g, h + ck) =& J(g, h) + \overline{c}J(g, k) \\ J(g + ch, k) =& J(g, k) + cJ(h, k) \\ J(g, h) =& \overline{J(h, g)}.\end{aligned}$$

The form is *bounded* if there is a (real) constant C such that

$$|J(h,k)| \le C\,\|h\|\,\|k\|.$$

In this case the infimum over all such C is the norm of J. If $T \in B(H)$, then $J(h,k) = \langle Th, k\rangle$ is a bounded sesquilinear form with $\|J\| = \|T\|$.

Lemma 27.26. *If J is a bounded sesquilinear form on the Hilbert space H, then there is a uniquely determined bounded operator T on H such that*

$$\langle Th, k\rangle = J(h,k).$$

Moreover, $\|T\|$ is the norm of J.

Sketch of proof. Fix $k \in H$. The mapping $\lambda_k : H \to \mathbb{C}$ defined by $\lambda_k(h) = J(h,k)$ is linear and bounded with norm at most $C\|k\|$. Hence, by the Riesz representation theorem, there is a vector $k_* \in H$ such that $\|k_*\| \le C\|k\|$ and

$$J(h,k) = \langle h, k_*\rangle.$$

Since J is conjugate linear in the second variable, the mapping $k \to k_*$ is linear. Let A denote this mapping. In particular, $\|Ak\| = \|k_*\| \le C\|k\|$. Hence A is bounded and $\|A\| \le C$ and for $h, k \in H$,

$$J(h,k) = \langle h, Ak\rangle.$$

Choose $T = A^*$. □

Sketch of proof of Proposition 27.25. Fix $f \in \mathbb{B}(X)$. Define $J : H \times H \to \mathbb{C}$ by $J(h,k) = \int f\, d\mu_{h,k}$. Verify that J is sesquilinear and $|J(h,k)| \le \|f\|_\infty \|h\|\,\|k\|$. Thus J is a bounded sesquilinear form. Thus, there is a (uniquely determined) bounded linear operator $I(f)$ such that

$$\langle I(f)h, k\rangle = J(h,k) = \int f\, d\mu_{h,k}.$$

Now suppose $\epsilon > 0$, a partition and points x_j as specified in the statement of the proposition are given and estimate, for $h, k \in H$,

$$\begin{aligned}|\langle [I(f) - \sum f(x_j)E(A_j)]h, k\rangle| =& |\sum \int_{A_j} (f - f(x_j))\, d\mu_{h,k}| \\ \le& \sum \epsilon |\mu_{h,k}|(A_j) \\ =& \epsilon |\mu_{h,k}|(X) \le \epsilon \|h\|\,\|k\|.\end{aligned}$$

□

We write $\int f\, dE$ for $I(f)$ and call it the *spectral integral of f*.

Proposition 27.27. *Suppose $(X, \mathscr{M})$ is a measure space and E is an $B(H)$-valued spectral measure on X. The mapping $I : \mathbb{B}(X) \to B(H)$ is a contractive $*$-homomorphism.*

Sketch of proof. While not particularly difficult, the most challenging part is to show I is multiplicative. The remainder of the proof is left to the gentle reader.

Fix $f, g \in \mathbb{B}(X)$ and let $\epsilon > 0$ be given. Choose a measurable partition $\{A_1, \dots, A_n\}$ of X such that $\sup\{|F(x) - F(y)| : x, y \in A_j\} < \epsilon$ for each j and $F \in \{f, g, fg\}$. Given $x_j \in A_j$, note that

$$\sum_j f(x_j)g(x_j)E(A_j) = \sum_{j,k} f(x_j)g(x_k)E(A_j)E(A_k),$$

since $E(A_j)E(A_k) = E(A_j \cap A_k) = 0$ for $j \neq k$. From here the proof is messy, but standard. □

Proposition 27.28. *If X is a compact Hausorff space and $\rho : C(X) \to B(H)$ is an isometric $*$-homomorphism, then there is a uniquely determined spectral measure on the Borel sets of X such that, for $f \in C(X)$,*

$$\rho(f) = \int f\, dE.$$

Sketch of proof. The first step is to extend ρ to an contractive $*$-homomorphism τ : $\mathbb{B}(X) \to B(H)$. Given $h, k \in H$, let $\rho_{h,k} : C(X) \to \mathbb{C}$ denote the linear functional

$$\rho_{h,k}(f) = \langle \rho(f)h, k\rangle.$$

By the Riesz-Markov Theorem, for each $h, k \in H$ there is a unique Borel measure $\mu_{h,k}$ such that

$$\rho_{h,k}(f) = \int f\, d\mu_{h,k}$$

and moreover, $\|h\|\,\|k\| \geq \|\rho_{h,k}\| = |\mu_{h,k}|(X)$. One checks that the mapping $H \times H \to M(X)$ (the space of measures) given by $(h, k) \to \mu_{h,k}$ is sesquilinear.

Given $g \in \mathbb{B}(X)$ we obtain the sesquilinear form $J : H \times H \to \mathbb{C}$ given by

$$J(h, k) = \int g\, d\mu_{h,k}.$$

Evidently $|J(h, k)| \leq \|g\|_\infty |\mu_{h,k}|(X) \leq \|g\|_\infty \|h\|\,\|k\|$. Hence there is a uniquely determined bounded linear operator $\tau(g)$ such that

$$\langle \tau(g)h, k\rangle = J(h, k) = \int g\, d\mu_{h,k}.$$

In particular, if $g \in C(X)$, then $\tau(g) = \rho(g)$ and $\|\tau(g)\| \leq \|g\|_\infty$.

We will verify that τ is multiplicative, leaving the rest of the verification that τ is a $*$-homomorphism (and hence contractive) to the gentle reader. If $f, g \in C(X)$, then

$\tau(fg) = \rho(fg) = \rho(f)\rho(g)$. Hence,

$$\begin{aligned}\langle \tau(fg)h, k\rangle &= \int fg\, d\mu_{h,k} \\ &= \langle \rho(f)\rho(g)h, k\rangle \\ &= \int f\, d\mu_{\rho(g)h,k}.\end{aligned}$$

In particular, $g\, d\mu_{h,k} = d\mu_{\rho(g)h,k}$ as measures. Thus, for $f \in \mathbb{B}(X)$,

$$\begin{aligned}\langle \tau(fg)h, k]\rangle &= \int fg\, d\mu_{h,k} \\ &= \int f\, d\mu_{\rho(g)h,k} \\ &= \langle \tau(f)\rho(g)h, k\rangle \\ &= \langle \rho(g)h, \tau(f)^* k\rangle \\ &= \int g\, d\mu_{h,\tau(f)^*k}.\end{aligned}$$

Hence $\tau(fg) = \tau(f)\tau(g)$ for $f \in \mathbb{B}(X)$ and $g \in C(X)$. Moreover, $f\, d\mu_{h,k} = d\mu_{h,\tau(f)^*k}$. Thus, for $g \in \mathbb{B}(X)$,

$$\begin{aligned}\langle \tau(fg)h, k\rangle &= \int fgd\, \mu_{h,k} \\ &= \int g\, d\mu_{h,\tau(f)^*k} \\ &= \langle \tau(g)h, \tau(f)^* k\rangle \\ &= \langle \tau(f)\tau(g)h, k\rangle\end{aligned}$$

and hence $\tau(fg) = \tau(f)\tau(g)$ for all $f, g \in \mathbb{B}(X)$. That τ is contractive, additive, respects scalar multiplication and satisfies $\tau(f^*) = \tau(f)^*$ are left as exercises.

Given a measurable set ω, let $E(\omega) = \tau(\mathbf{1}_\omega)$. Now suppose (f_n) is a bounded sequence from $\mathbb{B}(X)$ that converges pointwise to f (automatically in $\mathbb{B}(X)$). Given $h \in H$,

$$\begin{aligned}\|\tau(f_n - f)h\|^2 &= \langle (\tau(f_n - f)^*\, \tau(f_n - f))h, h\rangle \\ &= \langle \tau(|f_n - f|^2)h, k\rangle \\ &= \int |f_n - f|^2\, d\mu_{h,k}\end{aligned}$$

Now $|f_n - f|^2$ is a (uniformly) bounded sequence converging to 0 pointwise. Thus, since $|\mu_{h,k}|$ is a finite measure, the sequence of integrals above converges to 0 by dominated convergence. Thus $(\tau(f_n))$ SOT-converges to $\tau(f)$. Now let $(\omega_j)_{j=1}^{\infty}$ be a sequence of pairwise disjoint measurable sets and let $f_n = \sum_{j=1}^{n} \mathbf{1}_{\omega_j}$. Hence (f_n) is a bounded sequence of measurable functions that converges pointwise to f, the characteristic function of $\omega = \cup \omega_j$. Hence $\tau(f_n)$ SOT-converges to f; i.e.,

$$\text{SOT}\lim \tau(f_n) = \sum_{j=1}^{\infty} E(\omega) = E(\omega) = \tau(f).$$

If $s = \sum c_j \mathbf{1}_{\omega_j}$ is a simple measurable function, then

$$\tau(s) = \sum c_j \tau(\mathbf{1}_{\omega_j}) = \sum c_j E(\omega_j) = \int s\, dE.$$

Since an $f \in \mathbb{B}(X)$ is a uniform limit of simple functions (s_n) and τ is contractive, $(\tau(s_n))$ converges in norm to $\tau(f)$. Likewise, since the mapping $I : \mathbb{B}(X) \to B(H)$ given by $I(h) = \int h\, dE$ is contractive, $(I(s_n))$ converges to $I(f)$. Using $I(s_n) = \int s_n\, dE$, it follows that

$$\tau(f) = \int f\, dE$$

for $f \in \mathbb{B}(X)$.

□

Theorem 27.29 (Spectral Theorem)**.** *Suppose $\mathscr{A} \subset B(H)$ is a commutative sub-C-star algebra with maximal ideal space Σ and let $\mathscr{G} : \mathscr{A} \to C(\Sigma)$ denote the Gelfand transform. There is an $B(H)$-valued spectral measure E such that*

(1) for $T \in \mathscr{A}$,

$$T = \int_\Sigma \mathscr{G}(T)\, dE,$$

(2) if $O \subset \Sigma$ is open, then $E(O) \neq 0$;
(3) $R \in B(H)$ commutes with $\mathscr{A}$ if and only if $RE(\omega) = E(\omega)R$ for all $\omega \in \Sigma$.

Sketch of proof. The inverse of the Gelfand transform $\mathscr{G}^{-1} : C(\Sigma) \to \mathscr{A} \subset B(H)$ is an isometric $*$-isomorphism. Hence there is an $B(H)$-valued spectral measure on the Borel sets of Σ such that

$$\mathscr{G}^{-1}(f) = \int f\, dE$$

for $f \in C(\Sigma)$. Choose $f = \mathscr{G}(T)$ to obtain the first item.

If $O \subset \Sigma$ is open, then there exists a non-zero continuous function f that is 0 on O^C. Hence, $\|\mathscr{G}^{-1}(f)\| > 0$ and hence $E(O) \neq 0$.

Suppose R commutes with each $E(\omega)$. Given $h, k \in H$ and ω a Borel set,

$$\langle E(\omega)Rh, k\rangle = \langle E(\omega)h, R^*k\rangle.$$

Thus, $d\mu_{Rh,k} = d\mu_{h,R^*k}$. Hence, for $f \in C(\Sigma)$,

$$\begin{aligned}\langle [\int f\,dE]\,R\,h,k\rangle =& \int f\,d\mu_{Rh,k}\\ =& \int f\,d\mu_{h,R^*k}\\ =&\langle \int f\,dE\,h, R^*k\rangle\\ =&\langle R\int f\,dE\,h,k\rangle\end{aligned}$$

Conversely, if R commutes with $\mathscr{A}$, then for each $f \in C(\Sigma)$ and pair of vectors h,k,

$$\begin{aligned}\int f\,d\mu_{Rh,k} =&\langle [\int f\,dE]Rh,k\rangle\\ =&\langle [\int f\,dE]h,R^*k\rangle\\ =&\int f\,d\mu_{h,R^*k}.\end{aligned}$$

Hence $\mu_{Rh,k} = \mu_{h,R^*k}$ and consequently $\langle E(\omega)Rh,k\rangle = \langle RE(\omega)h,k\rangle$ for all measurable sets ω and vectors $h,k \in H$. The result follows. □

Corollary 27.30 (Spectral Theorem)**.** *If $T \in B(H)$ is normal, then there exists an $B(H)$-valued spectral measure E on the Borel sets of $\sigma(T)$ such that*

$$f(T) = \int f\,dE;$$

i.e., the isometric $*$-homomorphism $I : L^\infty(E) \to B(H)$ given by $I(f) = \int f\,dE$ extends the functional calculus.

Proof. Let $C^*(T)$ denote the C-star algebra generated by T. Thus $C^*(T)$ is a commutative C-star algebra and its maximal ideal space Σ is identified with $\sigma(T) \subset \mathbb{C}$ via the homeomorphism $\psi : \Sigma \to \sigma(T)$ given by $\psi(h) = \mathscr{G}(T)(h)$. Hence there is a uniquely determined spectral measure E on $\sigma(T)$ such that, for polynomials $p(z,z^*)$ (viewed as elements of $C(\sigma(T))$,

$$p(T,T^*) = \int_{\sigma(T)} \mathscr{G}(p(T,T^*) \circ \psi^{-1}\,dE.$$

Unraveling the definitions, $\mathscr{G}(p(T,T^*) \circ \psi^{-1}(z,z^*) = p(z,z^*)$; i.e.,

$$\rho(p) = \int p\,dE,$$

where ρ is the functional calculus. By continuity, $\rho(f) = \int f\,dE$ for $f \in C(\sigma(T))$. □