19. 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 19.1.** A norm on $\mathcal{X}$ is a function $\| \cdot \| : \mathcal{X} \to \mathbb{K}$ 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 $x, y \in \mathcal{X}$.

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 called 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 19.2 (Continuity of vector space operations).** Let $\mathcal{X}$ be a normed vector space over $\mathbb{K}$.

a) If $x_n \to x$ in $\mathcal{X}$, then $\| x_n \| \to \| x \|$ in $\mathbb{R}$.

b) If $k_n \to k$ in $\mathbb{K}$ and $x_n \to x$ in $\mathcal{X}$, then $k_n x_n \to kx$ in $\mathcal{X}$.

c) If $x_n \to x$ and $y_n \to y$ in $\mathcal{X}$, then $x_n + y_n \to x + y$ in $\mathcal{X}$.

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

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

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

(19.2)

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

We will want to prove the completeness of the first examples we consider, so we begin with a useful proposition. Say a series $\sum_{n=1}^\infty x_n$ in $\mathcal{X}$ is absolutely convergent if $\sum_{n=1}^\infty \| x_n \| < \infty$. Say that 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, we say that the series $\sum_{n=1}^\infty x_n$ converges to $x \in \mathcal{X}$ if $\lim_{N \to \infty} \| x - \sum_{n=1}^N x_n \| = 0$.)

**Proposition 19.3.** 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 let $\sum_{n=1}^{\infty} x_n$ be absolutely convergent. Write $s_N = \sum_{n=1}^{N} x_n$ for the $N^{th}$ partial sum and let $\epsilon > 0$. Since $\sum_{n=1}^{\infty} \|x_n\|$ is convergent, we can choose $L$ such that $\sum_{n=L}^{\infty} \|x_n\| < \epsilon$. Then for all $N > M \geq L$,

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

(19.3)

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

Conversely, suppose every absolutely convergent series in $\mathcal{X}$ is convergent. Let $(x_n)$ be a Cauchy sequence in $\mathcal{X}$. The idea is to arrange an absolutely convergent series whose partial sums form a subsequence of $(x_n)$. To do this, first choose $N_1$ such that $\|x_n - x_m\| < 2^{-1}$ for all $n, m \geq N_1$ and put $y_1 = x_{N_1}$. Next choose $N_2 > N_1$ such that $\|x_n - x_m\| < 2^{-2}$ for all $n, m \geq N_2$ and put $y_2 = x_{N_2} - x_{N_1}$. Continuing, we may choose inductively an increasing sequence of integers $(N_k)_{k=1}^{\infty}$ such that $\|x_n - x_m\| < 2^{-k}$ for all $n, m \geq N_k$ and define $y_k = x_{N_k} - x_{N_{k-1}}$. We have $\sum_{k=1}^{\infty} \|y_k\| < \sum_{k=1}^{\infty} 2^{-(k-1)} < \infty$ by construction, and the $K^{th}$ partial sum of $\sum_{k=1}^{\infty} y_k$ is $x_{N_K}$. Thus by hypothesis, the series $\sum_{k=1}^{\infty} y_k$ is convergent in $\mathcal{X}$, which means that the subsequence $(x_{N_k})$ of $(x_n)$ is convergent in $\mathcal{X}$.

The proof is finished by invoking a standard fact about convergence in metric spaces: if $(x_n)$ is a Cauchy sequence which has a subsequence converging to $x$, then the full sequence converges to $x$. □

19.1. Examples.

a) Of course, $\mathbb{R}^n$ with the usual Euclidean norm $\|(x_1, \ldots, x_n)\| = (\sum_{k=1}^{n} |x_k|^2)^{1/2}$ is a Banach space. (Likewise $\mathbb{C}^n$ with the Euclidean norm.) Besides these, $\mathbb{K}^n$ can be equipped with the $\ell^p$-norms

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

(19.4)

for $1 \leq p < \infty$, and the $\ell^\infty$-norm

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

(19.5)

It is not too hard to show that all of the $\ell^p$ norms ($1 \leq p \leq \infty$) are equivalent on $\mathbb{K}^n$ (though the constants $c, C$ must 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

$$c_0 := \{ f : \mathbb{N} \to \mathbb{K} | \lim_{n \to \infty} |f(n)| = 0 \}$$

$$\ell^\infty := \{ f : \mathbb{N} \to \mathbb{K} | \sup_{n \in \mathbb{N}} |f(n)| < \infty \}$$

$$\ell^1 := \{ f : \mathbb{N} \to \mathbb{K} | \sum_{n=0}^{\infty} |f(n)| < \infty \}$$
It is a simple exercise to check that each of these is a vector space. Define for functions \( f : \mathbb{N} \to K \):

\[
\|f\|_\infty := \sup_n |f(n)|
\]

\[
\|f\|_1 := \sum_{n=1}^{\infty} |f(n)|.
\]

Then \( \|f\|_\infty \) defines a norm on both \( c_0 \) and \( \ell^\infty \), and \( \|f\|_1 \) is a norm on \( \ell^1 \). Equipped with these respective norms, each is a Banach space. We sketch the proof for \( c_0 \), the others are left as exercises (Problem 20.4).

The key observation is that \( f_n \to f \) in the \( \|\cdot\|_\infty \) norm if and only if \( f_n \to f \) 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 \( f \in c_0 \), but this is a straightforward consequence of uniform convergence on \( \mathbb{N} \). The details are left as an exercise.

Along with these spaces it is also helpful to consider the vector space \( c_{00} := \{ f : \mathbb{N} \to K | f(n) = 0 \text{ for all but finitely many } n \} \) (19.6)

Notice that \( c_{00} \) is a vector subspace of each of \( c_0 \), \( \ell^1 \) and \( \ell^\infty \). 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 \( \ell^1 \) (but not in \( \ell^\infty \)). (See Problem 20.9).

c) (\( L^1 \) spaces) Let \((X, \mathcal{M}, m)\) 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 \( \ell^1 \) from the last example is a special case of this (what is the measure space?)�)

**Proposition 19.4.** \( L^1(m) \) is a Banach space.

**Proof.** We use Proposition 19.3. Suppose \( \sum_{n=1}^{\infty} f_n \) 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 series \( \sum_{n=1}^{\infty} f_n(x) \) is absolutely convergent in \( K \) for a.e. \( x \in X \). Define \( f(x) = \sum_{n=1}^{\infty} f_n(x) \). Then \( f \) is an a.e.-defined measurable function, and belongs to \( L^1(m) \) since \( |f| \leq g \). We claim that the partial sums \( \sum_{n=1}^{N} f_n \) converge to \( f \) in the \( L^1(m) \) norm: indeed,

\[
\left\| f - \sum_{n=1}^{N} f_n \right\|_1 \leq \int_X \left| f - \sum_{n=1}^{N} f_n \right| \, dm \leq \int_X \sum_{n=N+1}^{\infty} |f_n| \, dm \leq \sum_{n=N+1}^{\infty} \|f_n\|_1 \to 0
\]

(19.8)

(19.9)

(19.10)
as $N \to \infty$. (What justifies the equality in the last line?) □

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

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

(19.11)

(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 $\ell^p$ is defined analogously: it is the set of $f$:

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

(19.12)

and this quantity is a norm making $\ell^p$ 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

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

(19.13)

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 (19.13) 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

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

(19.14)
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
\[
\| (x, y) \|_\infty \leq \| (x, y) \|_1 \leq 2 \| (x, y) \|_\infty. \tag{19.15}
\]
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. Shortly we will construct further examples from linear transformations \( T : \mathcal{X} \to \mathcal{Y} \); to do this we first need to build up a few facts.

\subsection*{19.2. Linear transformations between normed spaces.}

**Definition 19.5.** Let \( \mathcal{X}, \mathcal{Y} \) be normed vector spaces. A linear transformation \( T : \mathcal{X} \to \mathcal{Y} \) is called **bounded** if there exists a constant \( C > 0 \) such that \( \| Tx \|_\mathcal{Y} \leq C \| x \|_\mathcal{X} \) for all \( x \in \mathcal{X} \).

**Remark:** 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?)

The importance of boundedness is hard to overstate; the following proposition explains its importance.

**Proposition 19.6.** Let \( T : \mathcal{X} \to \mathcal{Y} \) be a linear transformation between normed spaces. Then the following are equivalent:

(i) \( T \) is bounded.

(ii) \( T \) is continuous.

(iii) \( T \) is continuous at 0.

**Proof.** Suppose \( T \) is bounded and \( \| Tx \| \leq C \| x \|_\mathcal{X} \) for all \( x \in \mathcal{X} \); let \( \epsilon > 0 \). Then if \( \| x_1 - x_2 \| < \delta := \epsilon/C \), we have
\[
\| Tx_1 - Tx_2 \| = \| T(x_1 - x_2) \| \leq C \| x_1 - x_2 \| < \epsilon
\]
so \( T \) is continuous, and (i) implies (ii).

(ii) implies (iii) is trivial. For (iii) implies (i), we exploit homogeneity of the norm and the linearity of \( T \). By hypothesis, given \( \epsilon > 0 \) there exists \( \delta > 0 \) such that if \( \| x \| < \delta \), then
\[ \|Tx\| < \epsilon. \] Fix a nonzero vector \( x \in \mathcal{X} \) and a real number \( 0 < \lambda < \delta \). The vector \( \lambda x/\|x\| \) has norm less than \( \delta \), so
\[
\left\| T \left( \frac{\lambda x}{\|x\|} \right) \right\| = \lambda \frac{\|Tx\|}{\|x\|} < \epsilon. \tag{19.18}
\]
Rearranging this we find \( \|Tx\| \leq (\epsilon/\lambda)\|x\| \) for all \( x \neq 0 \), which shows \( T \) is bounded; in fact we can take \( C = \epsilon/\delta \).

**Definition 19.7** (The operator norm). Let \( \mathcal{X}, \mathcal{Y} \) be normed vector spaces and \( T : \mathcal{X} \to \mathcal{Y} \) a bounded linear transformation. Define
\[
\|T\| := \sup\{\|Tx\| : \|x\| = 1\} \tag{19.19}
\]
\[
= \sup\{\|Tx\| : x \neq 0, \|x\| = 1\} \tag{19.20}
\]
\[
= \inf\{C : \|Tx\| \leq C\|x\| \text{ for all } x \in \mathcal{X}\}. \tag{19.21}
\]
(As an exercise, verify that the three quantities on the right-hand side are equal.) The quantity \( \|T\| \) is called the **operator norm** of \( T \). We write \( B(\mathcal{X}, \mathcal{Y}) \) for the set of all bounded linear operators from \( \mathcal{X} \) to \( \mathcal{Y} \). Note that immediately we have the inequality
\[
\|Tx\| \leq \|T\|\|x\| \tag{19.22}
\]
for all \( x \in \mathcal{X} \). If \( T \in B(\mathcal{X}, \mathcal{Y}) \) and \( S \in B(\mathcal{Y}, \mathcal{Z}) \) then from two applications of the the inequality (19.22) we have for all \( x \in \mathcal{X} \)
\[
\|STx\| \leq \|S\|\|Tx\| \leq \|S\|\|T\|\|x\| \tag{19.23}
\]
and it follows that \( ST \in B(\mathcal{X}, \mathcal{Z}) \) and \( \|ST\| \leq \|S\|\|T\| \). If \( \mathcal{X} \) is a Banach space, then by the next proposition \( B(\mathcal{X}, \mathcal{X}) \) is complete, and this inequality says that \( B(\mathcal{X}, \mathcal{X}) \) is a **Banach algebra**.

**Proposition 19.8.** 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
\[
\|T_n x - T_m x\| = \|(T_n - T_m)x\| \leq \|T_n - T_m\|\|x\| \tag{19.24}
\]
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. We must show that \( T \) is bounded and \( \|T_n - T\| \to 0 \).

To see that \( T \) is bounded, note first that since \( (T_n) \) is a Cauchy sequence in the metric space \( B(\mathcal{X}, \mathcal{Y}) \), it is bounded as a subset of \( B(\mathcal{X}, \mathcal{Y}) \); that is, \( M := \sup_n \|T_n\| = \sup_n d(0, T_n) < \infty \). But then \( \|T_n x\| \leq M\|x\| \) for every \( x \), and since \( Tx = \lim T_n x \), we have \( \|Tx\| \leq M\|x\| \) for every \( x \) also.

Finally, given \( \epsilon > 0 \) choose \( N \) so that \( \|T_n - T_m\| < \epsilon \) for all \( n, m \geq N \). Then for all \( x \in \mathcal{X} \) with \( \|x\| = 1 \), we have \( \|T_n x - T_m x\| < \epsilon \). Taking the limit as \( m \to \infty \), we see that \( \|T_n x - Tx\| \leq \epsilon \) for all \( \|x\| = 1 \); thus \( \|T_n - T\| \leq \epsilon \) for \( n \geq N \). \( \square \)
The following proposition is very useful in constructing bounded operators—at least when the codomain is complete, it suffices to define the operator (and show that it is bounded) on a dense subspace:

**Proposition 19.9 (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 $\tilde{T} : \mathcal{X} \to \mathcal{Y}$ extending $T$ (so $\tilde{T}|_{\mathcal{E}} = T$), and $\|\tilde{T}\| = \|T\|$. 

**Proof.** Let $x \in \mathcal{X}$. By hypothesis there is a sequence $(x_n)$ in $\mathcal{E}$ converging to $x$; in particular this sequence is Cauchy. Since $T$ is bounded on $\mathcal{E}$, the sequence $(Tx_n)$ is Cauchy in $\mathcal{Y}$, hence convergent by hypothesis. Define $\tilde{T}x = \lim_n Tx_n$; using the fact that $T$ is bounded on $\mathcal{E}$ it is not hard to see that $\tilde{T}$ is well-defined and agrees with $T$ on $\mathcal{E}$ (the proof is an exercise). It also follows readily that $\tilde{T}$ is linear. To see that $\tilde{T}$ is bounded (and prove the equality of norms), fix $x \in \mathcal{X}$ with $\|x\| = 1$ and let $(x_n)$ be a sequence in $\mathcal{E}$ converging to $x$. Then $\|x_n\| \to \|x\|$; in particular for any $\epsilon$ we have $\|x_n\| < 1 + \epsilon$ for all $n$ sufficiently large. By enlarging $n$ if necessary, we may also assume that $\|\tilde{T}x_n - Tx_n\| < \epsilon$. But then using the triangle inequality,

$$\|\tilde{T}x\| < \epsilon + \|Tx_n\| < \epsilon + (1 + \epsilon)\|T\|.$$  

(19.25)

Since $\epsilon$ was arbitrary, we have $\|\tilde{T}\| \leq \|T\|$. As the reverse inequality is trivial, the proof is finished.  

□

**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 (so $T^{-1}$ exists as a linear transformation) and $T^{-1}$ is bounded from $\mathcal{Y}$ to $\mathcal{X}$. We can now define two useful notions of equivalence for normed vector spaces. Two normed spaces $\mathcal{X}$, $\mathcal{Y}$ are said to be (boundedly) isomorphic if there exists an invertible linear transformation $T : \mathcal{X} \to \mathcal{Y}$. The spaces are isometrically isomorphic if additionally $\|Tx\| = \|x\|$ for all $x$. A $T$ with this property is called an isometry; note that an isometry automatically injective; if it is also surjective then it is automatically invertible, in which case $T^{-1}$ is also an isometry. An isometry need not be surjective, however.

**19.3. 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 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

$$T_a f(n) = a(n) f(n).$$  

(19.26)

Then $T_a$ is bounded if and only if $M = \sup_{n \in \mathbb{N}} |a(n)| < \infty$, in which case $\|T_a\| = M$. When this happens, $T_a$ extends uniquely to a bounded operator from $c_0$ to $c_0$, and one may check that the formula (19.26) defines the extension. All of these claims...
remain true if we use the $\| \cdot \|_1$ norm instead of the $\| \cdot \|_\infty$ norm; we then get a bounded operator from $\ell^1$ to itself.

d) For $f \in \ell^1$, define the shift operator by $Sf(1) = 0$ and $Sf(n) = f(n-1)$ for $n \geq 1$. (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 $X$ is finite-dimensional, then the rank-nullity theorem from linear algebra guarantees that every injective linear map $T : X \to X$ 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. Problems


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

Problem 20.3. a) Prove that all norms on a finite dimensional vector space $X$ are equivalent. (Hint: fix a basis $e_1, \ldots, e_n$ for $X$ 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” $S = \{ x \in X : \|x\|_1 = 1 \}$ 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 $X$ be a normed vector space and $M \subset X$ a finite-dimensional subspace. Prove that $M$ is closed in $X$.

Problem 20.4. Finish the proofs from Example 19.1(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 $\text{Lip}[0,1]$ denote the set of all Lipschitz continuous functions on $[0,1]$. Prove that $\|f\| := |f(0)| + \|f\|_{Lip}$ is a norm on $\text{Lip}[0,1]$, and that $\text{Lip}[0,1]$ 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 that
$$\|f\| := \|f\|_\infty + \|f'\|_\infty$$
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 $(X, \mathcal{M})$ be a measurable space. Let $M(X)$ denote the (real) vector space of all signed measures on $(X, \mathcal{M})$. Prove that 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 that if $X, Y$ are normed spaces, then the operator norm is a norm on $B(X, Y)$.

Problem 20.9. Prove that $c_{00}$ is dense in $c_0$ and $\ell^1$. (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 $\ell^1$.) Using these facts, or otherwise, prove that $c_{00}$ is not dense in $\ell^\infty$. (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 that $c_{00}$ 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 that 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 19.9 can fail if $Y$ is not complete.)

Problem 20.12. Prove that 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 that $f \in C_0(\mathbb{R}^n)$ 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 that $C_c(\mathbb{R}^n)$ is dense in $C_0(\mathbb{R}^n)$ (where $C_0(\mathbb{R}^n)$ is equipped with sup norm).

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

Problem 20.15. Prove the claims in Example 19.3(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 that $Mg$ is bounded from $L^1(\mathbb{R})$ to itself if and only if $g \in L^\infty(\mathbb{R})$, in which case $\|Mg\| = \|g\|_\infty$.

Problem 20.17. Prove the claims about direct sums in Example 19.1(f).

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

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

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

a) Prove that the quotient norm is a norm (see Example 19.1(g)).

b) Show that the quotient map $x \to x + M$ has norm 1. (Use Problem 20.18.)

c) Prove that if $X$ is complete, so is $X/M$. 

9
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 that $c_0$ and $\ell^1$ are separable, but $\ell^\infty$ is not. (Hint: for $\ell^\infty$, show that there is an uncountable collection of elements $\{f_\alpha\}$ such that $\|f_\alpha - f_\beta\| = 1$ for $\alpha \neq \beta$.)

21. LINEAR FUNCTIONALS AND THE HAHN-BANACH THEOREM

If there is a “fundamental theorem of functional analysis,” it is the Hahn-Banach theorem. The particular version of it we will prove 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 normed vector space, and is always complete (even if $\mathcal{X}$ is not). This space is called the dual space of $\mathcal{X}$ and is denoted $\mathcal{X}^\ast$. It is not yet obvious that $\mathcal{X}^\ast$ 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 $\mathcal{X}$).


a) For each of the sequence spaces $c_0, \ell^1, \ell^\infty$, 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

$$L_g(f) := \sum_{n=0}^\infty f(n)g(n) \quad (21.1)$$

is bounded, since

$$|L_g(f)| \leq \sum_{n=0}^\infty |f(n)g(n)| \leq \|f\|_\infty \sum_{n=0}^\infty |g(n)| = \|g\|_1 \|f\|_\infty. \quad (21.2)$$

This 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 $g \to L_g$ is an isometric isomorphism from $\ell^1$ onto the dual space $c_0^\ast$.

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

$$L_g(f) := \sum_{n=0}^\infty g(n)f(n) \quad (21.3)$$

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

Let $L \in c_0^\ast$, 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

$$e_n(m) = \delta_{nm}. \quad (21.4)$$

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

$$g(n) = L(e_n). \quad (21.5)$$
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

$$h(n) = \begin{cases} \frac{g(n)}{|g(n)|} & \text{if } n \leq N \text{ and } g(n) \neq 0 \\ 0 & \text{otherwise} \end{cases}. \quad (21.6)$$

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

$$\sum_{n=0}^N |g(n)| = \sum_{n=0}^N h(n)g(n) = L(h) = |L(h)| \leq \|L\|\|h\| \leq \|L\|. \quad (21.7)$$

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, $L = L_g$. Uniqueness of $g$ is clear from its construction, since if $L = L_g$, we must have $g(n) = L_g(e_n)$. Thus the map $g \rightarrow L_g$ is onto and $\|L_g\|_{c_0^*} = \|g\|_1$. □

**Proposition 21.2.** $(\ell^1)^* \text{ is isometrically isomorphic to } \ell^\infty$.

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

The same mapping $g \rightarrow L_g$ also shows that every $g \in \ell^1$ gives a bounded linear functional on $\ell^\infty$, but it turns out these do not exhaust $(\ell^\infty)^*$ (see Problem 22.7).

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

$$L_g(f) := \int_X fg \, dm \quad (21.8)$$

is a bounded linear functional of norm at most $M$. We will prove in Section 27 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 $\sigma$-finite).

c) If $X$ is a compact metric space and $\mu$ is a finite, signed Borel measure on $X$, then

$$L_\mu(f) := \int_X f \, d\mu \quad (21.9)$$

is a bounded linear functional on $C_\mathbb{R}(X)$ of norm at most $\|\mu\|$.

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

**Definition 21.3.** Let $\mathcal{X}$ be a real vector space. A **Minkowski functional** is a function $p : \mathcal{X} \rightarrow \mathbb{R}$ 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 $L : \mathcal{X} \rightarrow \mathbb{R}$ 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 21.4 (The Hahn-Banach Theorem, real version). Let $\mathcal{X}$ be a normed vector space over $\mathbb{R}$, $p$ a Minkowski functional on $\mathcal{X}$, $\mathcal{M}$ a subspace of $\mathcal{X}$, and $L$ 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 $L'(x) \leq p(x)$ for all $x \in \mathcal{X}$ and $L'|_{\mathcal{M}} = L$.

Proof. The idea is to show that the extension can be done one dimension at a time, and we then infer the existence of an extension to the whole space by appeal to Zorn’s lemma. 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 $m_1, m_2 \in \mathcal{M}$, we have by hypothesis

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

(21.10)

which rearranges to

$$L(m_1) - p(m_1 - x) \leq p(m_2 + x) - L(m_2) \quad \text{for all } m_1, m_2 \in \mathcal{M}$$

(21.11)

so

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

(21.12)

Now choose any real number $\lambda$ satisfying

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

(21.13)

and define $L'(m + tx) = L(m) + t\lambda$. This $L'$ is linear by definition and agrees with $L$ on $\mathcal{M}$. We now check that $L'(y) \leq p(y)$ for all $y \in \mathcal{M} + \mathbb{R}x$. This is immediate if $y \in \mathcal{M}$. In general let $y = m + tx$ with $m \in \mathcal{M}$ and $t > 0$. Then

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

(21.14)

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 $(L'_1, \mathcal{N}_1) \leq (L'_2, \mathcal{N}_2)$ if $\mathcal{N}_1 \subset \mathcal{N}_2$ and $L'_2|_{\mathcal{N}_1} = L'_1$. This is a partial order on $\mathcal{L}$. Given any increasing chain $(L'_\alpha, \mathcal{N}_\alpha)$ in $\mathcal{L}$, it has 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, then, the collection $\mathcal{L}$ has a maximal element $(L', \mathcal{N})$ with respect to the order $\leq$. 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 21.5. Let $\mathcal{X}$ be a normed vector space over $\mathbb{R}$, $\mathcal{M}$ a subspace, and $L$ a bounded linear functional on $\mathcal{M}$ satisfying $|L(x)| \leq C\|x\|$ for all $x \in \mathcal{M}$. Then there exists a bounded linear functional $L'$ on $\mathcal{X}$ extending $L$, with $\|L'\| \leq C$.

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 21.6. Let \( \mathcal{X} \) be a vector space over \( \mathbb{C} \). If \( L : \mathcal{X} \to \mathbb{C} \) is a \( \mathbb{C} \)-linear functional, then \( u(x) = \text{Re} L(x) \) 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 \( \mathcal{X} \) is normed, then \( \|u\| = \|L\| \).

Proof. Problem ?? \( \square \)

Theorem 21.7 (The Hahn-Banach Theorem, complex version). Let \( \mathcal{X} \) be a normed vector space over \( \mathbb{C} \), \( p \) a seminorm on \( \mathcal{X} \), \( \mathcal{M} \) a subspace of \( \mathcal{X} \), and \( L : \mathcal{M} \to \mathbb{C} \) a \( \mathbb{C} \)-linear functional satisfying \( |L(x)| \leq p(x) \) for all \( x \in \mathcal{M} \). Then there exists a linear functional \( L' : \mathcal{X} \to \mathbb{C} \) satisfying \( |L'(x)| \leq p(x) \) for all \( x \in \mathcal{X} \).

Proof. The proof consists of applying the real Hahn-Banach theorem to extend the \( \mathbb{R} \)-linear functional \( u = \text{Re} L \) to a functional \( u' : \mathcal{X} \to \mathbb{R} \) and then defining \( L' \) from \( u' \) as in Proposition 21.6. The details are left as an exercise. \( \square \)

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

Corollary 21.8. Let \( \mathcal{X} \) be a normed vector space.

(i) (Linear functionals detect norms) If \( x \in \mathcal{X} \) is nonzero, there exists \( L \in \mathcal{X}^* \) with \( \|L\| = 1 \) such that \( L(x) = \|x\| \).

(ii) (Linear functionals separate points) If \( x \neq y \) in \( \mathcal{X} \), there exists \( L \in \mathcal{X}^* \) such that \( L(x) \neq L(y) \).

(iii) (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 \( L |_{\mathcal{M}} = 0 \) and \( L(x) = \text{dist}(x, \mathcal{M}) = \inf_{y \in \mathcal{M}} \|x - y\| > 0 \).

Proof. (i): Let \( \mathcal{M} \) be the one-dimensional subspace of \( \mathcal{X} \) spanned by \( x \). Define a functional \( L : \mathcal{M} \to \mathbb{K} \) by \( L(t \frac{x}{\|x\|}) = t \). For \( p(x) = \|x\| \), we have \( |L(x)| \leq p(x) \) on \( \mathcal{M} \), and \( L(x) = \|x\| \).

By the Hahn-Banach theorem the functional \( L \) extends to a functional (still denoted \( L \)) on \( \mathcal{X} \) and satisfies \( |L(y)| \leq \|y\| \) for all \( y \in \mathcal{X} \); thus \( \|L\| \leq 1 \); since \( L(x) = \|x\| \) by construction we conclude \( \|L\| = 1 \).

(ii): Apply (i) to the vector \( x - y \).

(iii): Let \( \delta = \text{dist}(x, \mathcal{M}) \). Define a functional \( L : \mathcal{M} + \mathbb{K}x \to \mathbb{K} \) by \( L(y + tx) = t\delta \). Since for \( t \neq 0 \)

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

(21.15)

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

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 \( T : \mathcal{Y} \to \mathcal{X}^* \).

A final corollary, which is also quite important. Note that since \( \mathcal{X}^* \) is a normed space, we can form its dual, denoted \( \mathcal{X}^{**} \) and called the bipedal of \( \mathcal{X} \). There is a canonical relationship between \( \mathcal{X} \) and \( \mathcal{X}^{**} \). First observe that if \( L \in \mathcal{X}^* \), each fixed \( x \in \mathcal{X} \) gives rise to a linear functional \( \hat{x} : \mathcal{X}^* \to \mathbb{K} \) via evaluation:

\[
\hat{x}(L) := L(x).
\]

(21.16)
**Corollary 21.9.** *(Embedding in the bidual)* The map \( x \rightarrow \hat{x} \) is an isometric linear map from \( \mathcal{X} \) into \( \mathcal{X}^{**} \).

**Proof.** First, from the definition we see that

\[
|\hat{x}(L)| = |L(x)| \leq \|L\| \|x\| \tag{21.17}
\]

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 \( x \rightarrow \hat{x} \) is linear. Finally, to show that \( \|\hat{x}\| = \|x\| \), fix a nonzero \( x \in \mathcal{X} \). From Corollary 21.8(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. \( \square \)

**Definition 21.10.** A Banach space \( \mathcal{X} \) is called reflexive if the map \( ^\circ : \mathcal{X} \rightarrow \mathcal{X}^{**} \) is surjective.

In other words, \( \mathcal{X} \) is reflexive if the map \( ^\circ \) is an isomorphism of \( \mathcal{X} \) with \( \mathcal{X}^{**} \). For example, every finite dimensional Banach space is reflexive (Problem 21.11). On the other hand, we will see below that \( c_0^* = \ell^\infty \), so \( c_0 \) is not reflexive. After we have studied the \( L^p \) and \( \ell^p \) 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 21.11** *(Completion of normed spaces).* If \( \mathcal{X} \) is a normed vector space, there is a Banach space \( \overline{\mathcal{X}} \) and an isometric map \( \iota : \mathcal{X} \rightarrow \overline{\mathcal{X}} \) such that the image \( \iota(\mathcal{X}) \) is dense in \( \overline{\mathcal{X}} \).

**Proof.** Embed \( \mathcal{X} \) into \( \mathcal{X}^{**} \) via the map \( x \rightarrow \hat{x} \) 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. \( \square \)

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 \( j : \mathcal{X} \rightarrow \mathcal{Y} \) 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.

### 21.2. Dual spaces and adjoint operators.

Let \( X, Y \) be normed spaces with duals \( X^*, Y^* \). If \( T : X \rightarrow Y \) is a linear transformation, then given any linear map \( f : Y \rightarrow \mathbb{K} \) we can define another linear map \( T^*f : X \rightarrow \mathbb{K} \) by the formula

\[
(T^*f)(x) = f(Tx). \tag{21.18}
\]

In fact, if \( f \) is bounded then so is \( T^*f \), and more is true:

**Theorem 21.12.** The formula (21.18) defines a bounded linear transformation \( T^* : Y^* \rightarrow X^* \), and in fact \( \|T^*\| = \|T\| \).

**Proof.** Let \( f \in Y^* \) and \( x \in X \). Then

\[
|T^*(f)(x)| = |f(Tx)| \leq \|f\| \|Tx\| \leq \|f\| \|T\| \|x\|. \tag{21.19}
\]

Taking the supremum over \( \|x\| = 1 \), we find that \( T^*f \) is bounded and \( \|T^*f\| \leq \|T\| \|f\| \), so \( \|T^*\| \leq \|T\| \). For the reverse inequality, let \( 0 < \epsilon < 1 \) be given and choose \( x \in X \) with \( \|x\| = 1 \) and \( \|Tx\| > (1 - \epsilon)\|T\| \). Now consider \( Tx \). By the Hahn-Banach theorem, there exists \( f \in Y^* \) such that \( \|f\| = 1 \) and \( f(Tx) = \|Tx\| \). For this \( f \), we have

\[
\|T^*f\| \geq |T^*(f)(x)| = |f(Tx)| = \|Tx\| > (1 - \epsilon)\|T\|. \tag{21.20}
\]

Since \( \epsilon \) was arbitrary, we conclude that \( \|T^*\| := \sup_{\|f\| = 1} \|T^*f\| \geq \|T\| \). \( \square \)
22. Problems

Problem 22.1. a) Prove that if $\mathcal{X}$ is a finite-dimensional normed space, then every linear functional $f : \mathcal{X} \to \mathbb{K}$ is bounded.

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

Problem 22.2. Let $\mathcal{X}, \mathcal{Y}$ be normed spaces and $T : \mathcal{X} \to \mathcal{Y}$ a linear transformation. Prove that $T$ is bounded if and only if there exists a constant $C$ such that for all $x \in \mathcal{X}$ and $f \in \mathcal{Y}^*$,

$$|f(Tx)| \leq C\|f\|\|x\|; \quad (22.1)$$

in which case $\|T\|$ is equal to the best possible $C$ in (22.1).

Problem 22.3. 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 $\mathcal{M} + \mathbb{C}x$ is closed. Use this to give another proof that every finite-dimensional subspace of $\mathcal{X}$ is closed.

Problem 22.4. Prove that 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 $x_1, \ldots, x_n$ for $\mathcal{M}$ and construct bounded linear functionals $f_1, \ldots, f_n$ 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.)

Problem 22.5. 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 that $T^*$ 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 22.6. Prove that 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 $|f_n(x_n)| \geq \frac{1}{2}\|f_n\|$. Show that the set of linear combinations of $\{x_n\}$ is dense in $\mathcal{X}$.

Problem 22.7. a) Prove that 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 closed 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 22.

23. The Baire Category Theorem and Applications

A topological space $X$ is called a Baire space if it has the following property: whenever $\{U_n\}_{n=1}^\infty$ is a countable sequence of open, dense subsets of $X$, the intersection $\bigcap_{n=1}^\infty U_n$ is dense in $X$. An equivalent formulation can be given in terms of nowhere dense sets: $X$ is Baire if and only if it is not a countable union of nowhere dense sets. (Recall that a set $E \subset X$
is called nowhere dense if its closure has empty interior.) In particular, note that if $E$ is nowhere dense, then $X \setminus E$ is open and dense. 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.

**Theorem 23.1** (The Baire Category Theorem). Every complete metric space $X$ is a Baire space.

**Proof.** Let $U_n$ be a sequence of open dense sets in $X$. Note that a set is dense if and only if it has nonempty intersection with every nonempty open set $W \subset X$. Fix such a $W$. Since $U_1$ is open and dense, there is a point $x_1 \in W \cap U_1$ and a radius $0 < r_1 < 1$ such that the closure $\overline{B(x_1, r_1)}$ is contained in $W \cap U_1$. Similarly, since $U_2$ is dense 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

$$\overline{B(x_2, r_2)} \subset B(x_1, r_1) \cap U_2. \quad (23.1)$$

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

$$\overline{B(x_n, r_n)} \subset B(x_{n-1}, r_{n-1}) \cap U_n. \quad (23.2)$$

Since $x_n \in B(x_n, r_n)$ for all $m \geq n$ and $r_n \to 0$, the sequence $(x_n)$ is Cauchy, and thus convergent since $X$ is complete. Moreover, since each the closed sets $\overline{B(x_n, r_n)}$ contains a tail of the sequence, each contains the limit $x$ as well. Thus $x \in W \cap U_1 \cap \cdots \cap U_n$ for all $n$, and the proof is complete. $\square$

**Remark:** In some books a countable intersection of open dense sets is called residual or second category and a countable union of nowhere dense sets is called meager or first category. In $\mathbb{R}$ with the usual topology, $\mathbb{Q}$ is first category and $\mathbb{R} \setminus \mathbb{Q}$ is second category.

We now give three important applications of the Baire category theorem in functional analysis: 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 23.2** (The Principle of Uniform Boundedness (PUB)). Let $\mathcal{X}, \mathcal{Y}$ be normed spaces with $\mathcal{X}$ complete, and let $\{T_\alpha : \alpha \in A\} \subset B(\mathcal{X}, \mathcal{Y})$ a collection of bounded linear transformations from $\mathcal{X}$ to $\mathcal{Y}$. If

$$M(x) := \sup_\alpha \|T_\alpha x\| < \infty \quad (23.3)$$

for each $x \in \mathcal{X}$, then $\sup_\alpha \|T_\alpha\| < \infty$. (In other words, a pointwise bounded family of linear operators is uniformly bounded.)

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

$$V_n := \{x \in \mathcal{X} : M(x) > n\}. \quad (23.4)$$

Since each $T_\alpha$ is bounded, the sets $V_n$ are open. (Indeed, for each $\alpha$ the map $x \to \|T_\alpha x\|$ is continuous from $\mathcal{X}$ to $\mathbb{R}$, so if $\|T_\alpha x\| > n$ for some $\alpha$ then also $\|T_\alpha y\| > n$ for all $y$ sufficiently close to $x$.) If each $V_n$ were dense, then since $\mathcal{X}$ is complete we would conclude from the Baire Category theorem that the intersection $V = \cap_{n=1}^\infty V_n$ is nonempty. But then for any
Choose \(x_0 \in \mathcal{X} \setminus \overline{V_N}\) and \(r > 0\) so that \(x_0 - x \in \mathcal{X} \setminus \overline{V_N}\) for all \(\|x\| < r\). Then for every \(\alpha\) and every \(\|x\| < r\) we have
\[
\|T_\alpha x\| \leq \|T_\alpha (x - x_0)\| + \|T_\alpha x_0\| \leq N + M(x_0) := R. \tag{23.5}
\]
That is, \(\|T_\alpha x\| \leq R\) for all \(\|x\| < r\), so by rescaling \(x\) we conclude that \(\|T_\alpha x\| \leq R/r\) for all \(\|x\| < 1\), so \(\sup_\alpha \|T_\alpha\| \leq R/r < \infty\). \hfill \Box

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 the case of normed linear spaces the condition that a linear map be open can be refined somewhat:

**Lemma 23.3** (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. Then the map \(T\) is open if and only if \(T(B)\) contains an open ball centered at 0.

**Proof.** This is more or less immediate from the fact that the translation map \(z \mapsto z + z_0\), for fixed \(z_0\), and dilation map \(z \mapsto rz, r \in \mathbb{K}\) are continuous in any normed vector space. The details are left as an exercise. \hfill \Box

**Theorem 23.4** (The Open Mapping Theorem). Suppose that \(\mathcal{X}, \mathcal{Y}\) are Banach spaces and \(T : \mathcal{X} \to \mathcal{Y}\) is a bounded, surjective linear map. Then \(T\) is open.

**Proof.** Let \(B_r\) denote the open ball of radius \(r\) centered at 0 in \(\mathcal{X}\). Trivially \(\mathcal{X} = \bigcup_{n=1}^\infty B_n\) and since \(\mathcal{Y}\) is surjective, we have \(\mathcal{Y} = \bigcup_{n=1}^\infty T(B_n)\). Since \(\mathcal{Y}\) is complete, from the Baire category theorem we have that \(T(B_N)\) is somewhere dense, for some \(N\). That is, \(\overline{T(B_N)}\) has nonempty interior; by rescaling we conclude that \(\overline{T(B_1)}\) has nonempty interior. We must deduce from this that in fact \(T(B_1)\) contains an open ball centered at 0 in \(\mathcal{Y}\). First we show that \(\overline{T(B_1)}\) contains an open ball centered at 0.

There exists \(y_0 \in \mathcal{Y}\) and \(r > 0\) such that \(\overline{T(B_1)}\) contains \(B_r (y_0)\). Choose \(y_1 = Tx_1\) in \(T(B_1)\) such that \(\|y_1 - y_0\| < r/2\), then by the triangle inequality
\[
B_{r/2} (y_1) \subset B_r (y_0) \subset \overline{T(B_1)} \tag{23.6}
\]
It follows that for all \(\|y\| < r/2\),
\[
y = -y_1 + (y + y_1) = -Tx_1 + (y + y_1) \in \overline{(-x_1 + B_1)} \subset \overline{T(B_2)} \tag{23.7}
\]
Halving the radius again, we conclude that if \(\|y\| < r/4\), then \(y \in \overline{T(B_1)}\).

Finally, we show that by shrinking the radius further we can find an open ball contained in \(T(B_1)\). From the previous paragraph, by scaling again for each \(n\), we see that if \(\|y\| < r/2^{n+2}\) then \(y \in \overline{T(B_{1/2}^n)}\). If \(\|y\| < r/8\), then there exists \(x_1 \in B_{1/2}\) such that \(\|y - Tx_1\| < r/16\). Inductively there exist \(x_n \in B_{1/2^n}\), with \(\|y - T \sum_{k=1}^n x_j\| < r2^{-n-3}\). The series \(\sum_{k=1}^\infty x_k\) is thus absolutely convergent, and hence convergent, since \(\mathcal{X}\) is complete. Call its sum \(x\). By construction we have the estimate \(\|x\| < \sum_{k=1}^\infty 2^{-k}\), so \(x \in B_1\) and \(y = Tx\). In conclusion, we have shown that \(y \in \overline{T(B_1)}\) whenever \(\|y\| < r/8\), so the proof is finished. \hfill \Box

**Corollary 23.5** (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).
Proof. Note that when \( T \) is bijection, \( T \) is open if and only if \( T^{-1} \) is continuous. The result then follows from the Open Mapping Theorem and Proposition 19.6.

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 then a topological space in the product topology. (In fact the product topology can be realized by norming \( \mathcal{X} \times \mathcal{Y} \), e.g. with the norm \( \|(x, y)\| := \max(\|x\|, \|y\|) \).) 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 that these maps are continuous. Given a linear map \( T : \mathcal{X} \to \mathcal{Y} \), its graph is the set
\[
G(T) := \{(x, y) \in \mathcal{X} \times \mathcal{Y} : y = Tx\} \tag{23.8}
\]
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 \( \mathcal{X} \times \mathcal{Y} \). This criterion is usually expressed in the following equivalent formulation: whenever \( x_n \to x \) in \( \mathcal{X} \) and \( Tx_n \to y \) in \( \mathcal{Y} \), it holds that \( Tx = y \). Note that this is not the same thing as saying that \( T \) is continuous: \( T \) is continuous if and only if whenever \( x_n \to x \) in \( \mathcal{X} \), then \( Tx_n \to Tx \) in \( \mathcal{Y} \). Evidently, if \( T \) is continuous then \( T \) is closed. The converse is false in general (see Problem 24.2). However, the converse does hold when \( \mathcal{X} \) and \( \mathcal{Y} \) are complete:

**Theorem 23.6** (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 \( \pi_1, \pi_2 \) 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 \( \pi_1 \) is a bijection between \( G(T) \) and \( \mathcal{X} \). Moreover, as maps between sets we have \( T = \pi_2 \circ \pi_1^{-1} \). So, it suffices to prove that \( \pi_1^{-1} \) is bounded.

By the remarks above \( \pi_1 \) and \( \pi_2 \) are bounded (from \( G(T) \) to \( \mathcal{X}, \mathcal{Y} \) respectively). Since \( \mathcal{X} \) and \( \mathcal{Y} \) are complete, so is \( \mathcal{X} \times \mathcal{Y} \), and therefore \( G(T) \) is complete since it is assumed closed. Since \( \pi_1 \) is a bijection between \( G(T) \) and \( \mathcal{X} \), its inverse is also bounded (Corollary 23.5).

\[ \square \]

## 24. Problems

**Problem 24.1.** 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 24.2.** Consider the linear subspace \( \mathcal{D} \subset c_0 \) defined by
\[
\mathcal{D} = \{f \in c_0 : \lim_{n \to \infty} |nf(n)| = 0\} \tag{24.1}
\]
and the linear transformation \( T : \mathcal{D} \to c_0 \) defined by \( (Tf)(n) = nf(n) \).

a) Prove that \( T \) is closed, but not bounded. b) Prove that \( T^{-1} : c_0 \to \mathcal{D} \) is bounded and surjective, but not open.

**Problem 24.3.** Suppose \( \mathcal{X} \) is a vector space equipped with two norms \( \| \cdot \|_1, \| \cdot \|_2 \) such that \( \| \cdot \|_1 \leq \| \cdot \|_2 \). Prove that if \( \mathcal{X} \) is complete in both norms, then the two norms are equivalent.

**Problem 24.4.** 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 that if \( T \) is weakly bounded, then \( T \) is bounded.
Problem 24.5. 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 that if $T$ is defined by $Tx = \lim_n T_n x$, then $T$ is bounded.

Problem 24.6. 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 that there is no norm on $\mathcal{X}$ under which it is complete. (Hint: consider the finite-dimensional subspaces $\mathcal{X}_n := \text{span}\{x_1, \ldots, x_n\}$.)

Problem 24.7. 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 $F_n$ 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 $|f(x) - f(x_0)| \leq n|x - x_0|$ for all $x \in [0, 1]$. Prove that 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 residual. (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 $E_n$.)

25. Hilbert spaces

25.1. Inner products.

Definition 25.1. 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: 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.

25.2. Examples.

a) ($\mathbb{R}^n$ and $\mathbb{C}^n$) It is easy to check that the standard scalar product on $\mathbb{R}^n$ is an inner product; it is defined as usual by

$$\langle x, y \rangle = \sum_{j=1}^{n} x_j y_j \quad \text{(25.1)}$$

where we have written $x = (x_1, \ldots, x_n)$; $y = (y_1, \ldots, y_n)$. Similarly, the standard inner product of vectors $z = (z_1, \ldots, z_n)$, $w = (w_1, \ldots, w_n)$ in $\mathbb{C}^n$ is given by

$$\langle z, w \rangle = \sum_{j=1}^{n} z_j \overline{w_j}. \quad \text{(25.2)}$$
(Note that it is necessary to take complex conjugates of the \(w\)'s to obtain positive definiteness.)

b) \(\ell^2(\mathbb{N})\) is defined to be the vector space of square-summable sequences of numbers in \(\mathbb{K}\):

\[
\ell^2(\mathbb{N}) = \{(a_1, a_2, \ldots a_n, \ldots) \mid a_n \in \mathbb{K}, \sum_{j=1}^{\infty} |a_n|^2 < \infty\} \quad (25.3)
\]

(Exercise: prove that \(\ell^2(\mathbb{N})\) is a vector space. The Cauchy-Schwarz inequality, proved below, will be helpful). The standard inner product on \(\ell^2(\mathbb{N})\) is, for sequences \(a = (a_1, a_2, \ldots)\) and \(b = (b_1, b_2, \ldots)\),

\[
\langle a, b \rangle = \sum_{n=1}^{\infty} a_n b_n. \quad (25.4)
\]

It is not immediate that this series is absolutely convergent; this will follow easily, however, once we have proven the Cauchy-Schwarz inequality. Assuming convergence for now, it is straightforward to prove that this defines an inner product.

c) \(L^2(\mu)\): Let \((X, \mathcal{M}, \mu)\) be a measure space. Consider the set of all measurable functions \(f : X \to \mathbb{K}\) such that

\[
\int_X |f|^2 \, d\mu < \infty \quad (25.5)
\]

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. The resulting object is a vector space (Exercise: prove this), and as one would expect the inner product is given by

\[
\langle f, g \rangle = \int_X f \overline{g} \, d\mu. \quad (25.6)
\]

(Again we must check that the integral is convergent, and again this will follow from Cauchy-Schwarz, or more precisely by imitating the proof of Cauchy-Schwarz.) Notice that the previous examples are special cases of this one, with \(\mu\) equal to counting measure on \(\{1, \ldots n\}\) and \(\mathbb{N}\) respectively.)

**Theorem 25.2** (The Cauchy-Schwarz inequality). If \(u\) is a pre-inner product on the vector space \(X\), then for all \(x, y \in X\),

\[
|u(x, y)|^2 \leq u(x, x)u(y, y). \quad (25.7)
\]

**Proof.** For \(\alpha \in \mathbb{K}\) and \(x, y \in X\), we have

\[
0 \leq u(x - \alpha y, x - \alpha y) \quad (25.8)
\]

\[
= u(x, x) - \alpha u(y, x) - \overline{\alpha} u(x, y) + u(y, y). \quad (25.9)
\]

Write \(u(x, y)\) in polar form as \(re^{i\theta}\). Fix \(t\) for now (we will make an optimal choice later) and let \(\alpha = te^{-i\theta}\). Then (25.9) becomes

\[
0 \leq u(x, x) - e^{-i\theta}tre^{i\theta} - e^{i\theta}rte^{-i\theta} + t^2 u(y, y) \quad (25.10)
\]

\[
= u(x, x) - 2rt + t^2 u(y, y) \quad (25.11)
\]
The polynomial \( q(t) = u(y, y)t^2 - 2rt + u(x, x) \) has real coefficients and is nonnegative for all \( t \), so it has either a repeated real root or complex roots. This means that the discriminant \( b^2 - 4ac \) is nonpositive, in other words
\[
4r^2 - 4u(x, x)u(y, y) \leq 0 \tag{25.12}
\]
which proves (25.7).

\[\square\]

**Remark:** Here is a second proof; the “arbitrage” argument given here may help to clarify the optimization trick used in the proof of Cauchy-Schwarz. Expanding out \( u(x - y, x - y) \) and rearranging, we get
\[
2\text{Re } u(x, y) \leq u(x, x) + u(y, y) \tag{25.13}
\]
Notice that \( \text{Re } u(e^{-i\theta}x, y) = |u(x, y)| \), so replacing \( x \) by \( e^{-i\theta}x \) (which leaves the right hand side unchanged) we get
\[
2|u(x, y)| \leq u(x, x) + u(y, y) \tag{25.14}
\]
which says that \( |u(x, y)| \) is dominated by the arithmetic mean of \( u(x, x) \) and \( u(y, y) \). But the inequality we are after says that \( |u(x, y)| \) is in fact controlled by the geometric mean—how can the stronger inequality follow from the weaker one? Notice that for \( \lambda \) nonzero and real, the transformation \( x \to \lambda x, \ y \to \frac{1}{\lambda}y \) leaves the left hand side unchanged, so in fact
\[
2|u(x, y)| \leq \lambda^2u(x, x) + \frac{1}{\lambda^2}u(y, y) \quad \text{for all } \lambda \in \mathbb{R} \setminus \{0\}. \tag{25.15}
\]
So now we can minimize the right hand side as a function of \( \lambda \); doing a little calculus (and taking care of the cases where \( u(x, x) \) or \( u(y, y) \) is zero) finishes. Moral: inequalities can “self-improve” if there are transformations that leave one side invariant but not the other; apply the transformation and optimize.

**Example:** Let us go back and verify that the series defining the inner product in \( \ell^2(\mathbb{N}) \) is absolutely convergent. Let \( a, b \in \ell^2(\mathbb{N}) \); by assumption we have
\[
\sum_{n=1}^{\infty} |a_n|^2 = A < \infty, \quad \sum_{n=1}^{\infty} |b_n|^2 = B < \infty \tag{25.16}
\]
Fix \( N \); by applying the Cauchy-Schwarz inequality to the standard inner product (25.2) on \( \mathbb{C}^N \), we have
\[
\sum_{n=1}^{N} |a_n b_n| \leq \left( \sum_{n=1}^{N} |a_n|^2 \right)^{1/2} \left( \sum_{n=1}^{N} |b_n|^2 \right)^{1/2} \leq \sqrt{AB}. \tag{25.17}
\]
\[
\leq \sqrt{AB}. \tag{25.18}
\]
So, the partial sums of the series \( \sum_{n=1}^{\infty} |a_n b_n| \) are increasing and bounded above, so the series converges (25.4) converges absolutely.
25.3. Norms. Given a vector space $X$ over $\mathbb{K}$ and a semi-inner product $\langle \cdot , \cdot \rangle$, define for each $x \in X$
\[
\| x \| := \sqrt{\langle x, x \rangle}.
\] (25.19)
This quantity should act something like a “length” of the vector $x$. Clearly $\| x \| \geq 0$ for all $x$, and moreover we have:

**Theorem 25.3.** Let $X$ be a semi-inner product space over $\mathbb{K}$, with $\| \cdot \|$ defined by equation (25.19). Then for all $x, y \in X$ and $\alpha \in \mathbb{K}$,

1) $\| x + y \| \leq \| x \| + \| y \|$
2) $\| \alpha x \| = |\alpha| \| x \|

If $\langle \cdot , \cdot \rangle$ is an inner product, then also
3) $\| x \| = 0$ if and only if $x = 0$.

**Proof.** We prove 1) and leave 2) and 3) as exercises. For all $x, y \in X$ we have
\[
\| x + y \|^2 = \langle x + y, x + y \rangle
\]
\[
= \| x \|^2 + 2 \text{Re} \langle x, y \rangle + \| y \|^2
\]
\[
\leq \| x \|^2 + 2 |\langle x, y \rangle| + \| y \|^2
\]
\[
\leq \| x \|^2 + 2 \| x \| \| y \| + \| y \|^2
\]
\[
= (\| x \| + \| y \|)^2
\]
where we have used the Cauchy-Schwarz inequality in (25.23). Taking square roots finishes.

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 $\mathbb{R}^n$, we get
\[
\| x \| = (x_1^2 + \cdots + x_n^2)^{1/2}
\]
which is of course the usual Euclidean norm.

We are now ready to define Hilbert spaces.

**Definition 25.4.** A **Hilbert space** over $\mathbb{K}$ is a vector space $X$ over $\mathbb{K}$, equipped with an inner product, and such that $X$ is complete in the metric $d(x, y) = \| x - y \|$.

25.4. Completeness. Our next task is to show that the examples of inner product spaces considered thus far are complete with respect to their norms, and are therefore Hilbert spaces. That $\mathbb{R}^n$ and $\mathbb{C}^n$ are complete 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 25.5.** $L^2(\mu)$ is complete.

**Proof.** We use Proposition 19.3. Suppose $f_k$ is a sequence in $L^2(\mu)$ and $\sum_{k=1}^{\infty} \| f_k \| = B < \infty$. Define
\[
G_n = \sum_{k=1}^{n} |f_k| \quad \text{and} \quad G = \sum_{k=1}^{\infty} |f_k|.
\] (25.26)
By the triangle inequality, \( \|G_n\| \leq \sum_{k=1}^{n} \|f_k\| \leq B \) for all \( n \), and so by the Monotone Convergence Theorem,
\[
\int_X G^2 \, d\mu = \lim_{n \to \infty} \int_X G_n^2 \, d\mu \leq B^2,
\]
so \( G \) belongs to \( L^2(\mu) \) and in particular \( G(x) < \infty \) for almost every \( x \). By the definition of \( G \), this implies that the sum
\[
\sum_{k=1}^{\infty} f_k(x)
\]
converges absolutely for almost every \( x \); call this (a.e. defined) sum \( f \). By construction, \( |f| \leq G \) and thus \( f \in L^2(\mu) \). Moreover, for all \( n \) we have
\[
\left| f - \sum_{k=1}^{n} f_k \right|^2 \leq (2G)^2.
\]
Equation (25.27) says that \( G^2 \) is integrable, so we can apply the Dominated Convergence Theorem to obtain
\[
\lim_{n \to \infty} \left\| f - \sum_{k=1}^{n} f_k \right\|^2 = \lim_{n \to \infty} \int_X \left| f - \sum_{k=1}^{n} f_k \right|^2 \, d\mu = 0.
\]
□

25.5. Orthogonality. In this section we show that many of the basic features of the Euclidean geometry of \( \mathbb{R}^n \) can be extended to the Hilbert space setting.

**Definition 25.6.** Let \( H \) be an inner product space. Two vectors \( x, y \in X \) are **orthogonal** if \( \langle x, y \rangle = 0 \).

When \( x \) and \( y \) are orthogonal we write \( x \perp y \). More generally, if \( A \) and \( B \) are any two subsets of \( H \), we say \( A \perp B \) if \( x \perp y \) for all \( x \in A, y \in B \). Also, we let \( A^\perp \) denote the set \( \{x \in H \mid x \perp A\} \).

**Theorem 25.7** (The Pythagorean Theorem). *If \( H \) is an inner product space and \( f_1, \ldots, f_n \) 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
\[
\|f_1 + f_2\|^2 = \|f_1\|^2 + 2\text{Re} \langle f_1, f_2 \rangle + \|f_2\|^2 = \|f_1\|^2 + \|f_2\|^2.
\]
The general case follows by induction. □

**Theorem 25.8** (The Parallelogram Law). *If \( H \) is an inner product space and \( f, g \in H \), then*
\[
\|f + g\|^2 + \|f - g\|^2 = 2(\|f\|^2 + \|g\|^2).
\]

**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. □
If we look at the proof of the Parallelogram Law and subtract instead of adding, we find
\[ \| f + g \|^2 - \| f - g \|^2 = 4 \text{Re} \langle f, g \rangle. \]  
(25.36)
This proves

**Theorem 25.9** (The Polarization identity). If \( H \) is an inner product space over \( \mathbb{R} \), then
\[ \langle f, g \rangle = \frac{1}{4} (\| f + g \|^2 - \| f - g \|^2) \]  
(25.37)

**Remark:** There is also a version of the polarization identity in the complex case; its statement and proof are left as an exercise. 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 (25.34) holds for all \( f, g \in H \), then \( H \) is an inner product space; the inner product is then given by the formula (25.37). (The proof is simply to define the inner product by equation (25.37), and check that it is indeed an inner product.) There is of course a complex version as well.

25.6. **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 \( \mathbb{F} \), a subset \( K \subseteq X \) is called **convex** if whenever \( a, b \in K \) and \( 0 \leq t \leq 1 \), 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 25.10.** Suppose \( H \) is a Hilbert space, \( K \subseteq H \) is a closed, convex, nonempty set, and \( h \in H \). Then there exists a unique vector \( k_0 \in K \) such that
\[ \| h - k_0 \| = \text{dist}(h, K) := \inf \{ \| h - k \| : k \in K \}. \]  
(25.38)

**Proof.** First, let's replace \( K \) by \( K - h = \{ k - h \mid k \in K \} \). Then \( \text{dist}(h, K) = \text{dist}(0, K - h) \), so we may assume \( h = 0 \). (Is there anything we need to check here?) Let \( d = \text{dist}(0, K) = \inf_{k \in K} \| k \| \). By assumption, there exists a sequence \( (k_n) \) in \( K \) so that \( \| k_n \| \to d \). It follows from the parallelogram law that for all \( m, n \),
\[ \left\| \frac{k_m - k_n}{2} \right\|^2 = \frac{1}{2} (\| k_m \|^2 + \| k_n \|^2) - \left\| \frac{k_m + k_n}{2} \right\|^2 \]  
(25.39)
Since \( K \) is convex, \( \frac{1}{2}(k_m + k_n) \in K \), and therefore \( \left\| \frac{1}{2}(k_m + k_n) \right\|^2 \geq d^2 \). Now let \( \epsilon > 0 \) and choose \( N \) such that for all \( n \geq N \), \( \| k_n \|^2 < d^2 + \frac{1}{4} \epsilon^2 \). By (25.39), if \( m, n \geq N \) then
\[ \left\| \frac{k_m - k_n}{2} \right\|^2 < \frac{1}{2} (2d^2 + \frac{1}{2} \epsilon^2) - d^2 = \frac{1}{4} \epsilon^2. \]  
(25.40)
This says that \( \| k_m - k_n \| < \epsilon \) for \( m, n \geq N \); that is, \( (k_n) \) is a Cauchy sequence. Since \( H \) is complete, \( k_n \) converges to a limit \( k_0 \), and since \( K \) is closed, \( k_0 \in K \). Moreover, for all \( k_n \),
\[ d \leq \| k_0 \| = \| k_0 - k_n \| + \| k_n \| \leq \| k_0 - k_n \| + \| k_n \| \to d \quad \text{as} \quad n \to \infty. \]  
(25.41)
This means \( \|k_0\| = d \). It remains to show that \( k_0 \) is the unique element of \( K \) with this property. So, suppose also \( k' \in K \) and \( \|k'\| = d \). Then \( (k_0 + k')/2 \) belongs to \( K \), and has norm at least \( d \). By the parallelogram law again,

\[
0 \leq \left\| \frac{k_0 - k'}{2} \right\|^2 = \frac{1}{2} (\|k_0\|^2 + \|k'\|^2) - \left\| \frac{k_0 + k'}{2} \right\|^2 \tag{25.43}
\]

\[
\leq \frac{1}{2} (d^2 + d^2) - d^2 = 0 \tag{25.44}
\]

so \( \|k_0 - k'\| = 0 \), which means \( k_0 = k' \). \( \square \)

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 25.11.** Let \( H \) be a Hilbert space, \( M \leq H \), and \( h \in H \). If \( f_0 \) is the unique element of \( M \) such that \( \|h - f_0\| = \text{dist}(h, M) \), then \( (h - f_0) \perp M \). Conversely, if \( f_0 \in M \) and \( (h - f_0) \perp M \), then \( \|h - f_0\| = \text{dist}(h, M) \).

**Proof.** Let \( f_0 \in M \) with \( \|h - f_0\| = \text{dist}(h, M) \). If \( f \) is any vector in \( M \), then \( f_0 + f \) belongs to \( M \) and so

\[
0 \leq \|h - f_0\|^2 \leq \|h - (f_0 + f)\|^2 \tag{25.45}
\]

\[
= \|(h - f_0) + f\|^2 \tag{25.46}
\]

\[
= \|h - f_0\|^2 - 2 \Re \langle h - f_0, f \rangle + \|f\|^2. \tag{25.47}
\]

This implies that

\[
2 \Re \langle h - f_0, f \rangle \leq \|f\|^2 \quad \text{for all } f \in M. \tag{25.48}
\]

As in the proof of Cauchy-Schwarz, we optimize over \( f \): first, let \( re^{i\theta} \) be the polar form of \( \langle h - f_0, f \rangle \). Let \( t > 0 \). If we replace \( f \) in (25.48) by \( te^{i\theta}f \), we get

\[
2t|\langle h - f_0, f \rangle| \leq t^2\|f\|^2. \tag{25.49}
\]

Dividing by \( t \) and taking the limit as \( t \to 0 \), we see that \( \langle h - f_0, f \rangle = 0 \), as desired.

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 \( f \in M \)

\[
\|h - f\|^2 = \|(h - f_0) + (f_0 - f)\|^2 \tag{25.50}
\]

\[
= \|h - f_0\|^2 + \|f_0 - f\|^2 \quad \text{(why?)} \tag{25.51}
\]

\[
\geq \|h - f_0\|^2. \tag{25.52}
\]

Thus \( \|h - f_0\| = \text{dist}(h, M) \). \( \square \)

To reiterate: 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

\[
P(h) = f_0. \tag{25.53}
\]
Henceforth we omit the parentheses and just write $Ph = f_0$. If the space $M$ needs to be emphasized we will write $P_M$ for $P$. The function $P$ is called the orthogonal projection of $H$ on $M$. The following theorem describes the basic properties of $P$.

**Theorem 25.12.** Let $H$ be a Hilbert space, $M \leq H$, and $P$ the orthogonal projection on $M$. Then:

a) $P$ is a linear transformation from $H$ into itself.

b) $\|Ph\| \leq \|h\|$ for all $h \in H$.

c) $P^2 = P$.

d) $\ker P = M^\perp$ and $\text{ran } P = M$.

**Proof.**

a) Let $\alpha \in \mathbb{F}$ and $h_1, h_2 \in H$. We must prove that $P(\alpha h_1 + h_2) = \alpha Ph_1 + Ph_2$. To do this we prove that $\alpha Ph_1 + Ph_2$ has the property uniquely characterizing $P(\alpha h_1 + h_2)$, namely, that it is the unique vector $f_0 \in H$ such that $\alpha h_1 + h_2 - f_0$ is orthogonal to $M$. So, let $f \in M$ and compute:

\[
\langle (\alpha h_1 + h_2) - (\alpha Ph_1 + Ph_2), f \rangle = \alpha \langle h_1 - Ph_1, f \rangle + \langle h_2 - Ph_2, f \rangle = 0
\]

by the definition of $Ph_1$ and $Ph_2$, and Theorem 25.11.

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

\[
\|h\|^2 = \|h - Ph\|^2 + \|Ph\|^2,
\]

so $\|Ph\| \leq \|h\|$.

c) If $f \in M$, then evidently $Pf = f$. But $Ph \in M$ always, so $P^2h = P(Ph) = Ph$.

d) If $Ph = 0$, then $h - Ph = h \in M^\perp$, so $\ker P \subseteq M^\perp$. On the other hand, if $h \in M^\perp$, then $h - 0 \in M^\perp$, so by Theorem 25.11 $Ph = 0$. That $\text{ran } P = M$ is evident. \qed

**Corollary 25.13.** If $M \leq H$, then $(M^\perp)^\perp = M$.

**Corollary 25.14.** If $E$ is any subset of $H$, then $(E^\perp)^\perp$ is equal to the smallest closed subspace of $H$ containing $E$.

25.7. 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: fix a vector $h_0 \in H$ and define

\[
L(h) = \langle h, h_0 \rangle.
\]

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:

\[
|L(h)| = |\langle h, h_0 \rangle| \leq \|h_0\| \|h\|
\]

So $\|L\| \leq \|h_0\|$, but in fact it is easy to see that $\|L\| = \|h_0\|$; just apply $L$ to the unit vector $h_0/\|h_0\|$.

In fact, it is clear from linear algebra that every linear functional on $\mathbb{K}^n$ is of this form. More generally, every bounded linear functional on a Hilbert space has the form just described.
**Theorem 25.15** (The Riesz Representation Theorem). *If $H$ is a Hilbert space and $L : H \to \mathbb{F}$ is a bounded linear functional, then there exists a unique vector $h_0 \in H$ such that*

$$L(h) = \langle h, h_0 \rangle$$  \hfill (25.59)

*for all $h \in H$. Moreover, $\|L\| = \|h_0\|$.*

**Proof.** If $L = 0$ we of course have $h_0 = 0$. So, assume $L \neq 0$. Notice that if (25.59) holds, then necessarily $h_0 \perp \ker L$, so this tells us where to look for $h_0$. In particular, since we are assuming $L$ is bounded, it is continuous by Proposition 19.6, and therefore $\ker L = L^{-1}(\{0\})$ is a proper, closed subspace of $H$. We can then find a nonzero vector $f_0 \in (\ker L) \perp$ (why?) and by rescaling we may assume $L(f_0) = 1$.

Let $h \in H$ and $\alpha = L(h)$. Then

$$L(h - \alpha f_0) = L(h) - \alpha = 0,$$  \hfill (25.60)

so $h - \alpha f_0 \in \ker L$. Therefore

$$0 = \langle h - \alpha f_0, f_0 \rangle = \langle h, f_0 \rangle - \alpha \|f_0\|^2.$$  \hfill (25.61)

Solving for $\alpha$, we find

$$\alpha = L(h) = \langle h, \frac{f_0}{\|f_0\|^2} \rangle.$$  \hfill (25.63)

Thus, $h_0 = f_0/\|f_0\|^2$ works. Uniqueness is immediate, since if $\langle h, h_0 \rangle = \langle h, h_1 \rangle$ for all $h$, then $h_0 = h_1$. Finally, the expression for the norm was proved in the discussion preceding the theorem. \hfill $\square$

25.8. Orthonormal Sets and Bases.

**Definition 25.16.** 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 $e \neq f$, 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 $H$.

Observe that an orthonormal set $E$ is maximal if and only if the only vector orthogonal to $E$ is the zero vector.

**Remark:** It must be stressed that a basis in the above sense need not be a basis in the sense of linear algebra. In particular, it is always true that an orthonormal set is linearly independent (Exercise: prove this), but in general an orthonormal basis need not span $H$.

25.9. Examples.

a) Of course the standard basis $\{e_1, \ldots, e_n\}$ 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

$$e_n(k) = \begin{cases} 
1 & \text{if } k = n \\
0 & \text{if } k \neq n
\end{cases}$$  \hfill (25.64)

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 $h : \mathbb{N} \to \mathbb{F}$ 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 $h = 0$. 

27
c) Let \( H = L^2[0,1] \). Consider for \( n \in \mathbb{Z} \) the functions \( e_n(x) = e^{2\pi nx} \). It is a calculus exercise to check that this set is orthonormal. It is in fact a basis, but this is not obvious; its proof relies on facts from the theory of Fourier series.

d) More generally, given any linearly independent sequence \( f_1, f_2, \ldots \), we can apply the Gram-Schmidt process to obtain an orthonormal set. More precisely:

**Theorem 25.17** (Gram-Schmidt process). If \( (f_n)_{n=1}^{\infty} \) is a linearly independent sequence in \( H \), there exists an orthonormal sequence \( (e_n)_{n=1}^{\infty} \) such that for each \( n \),

\[
\text{span} \{f_1, \ldots, f_n\} = \text{span} \{e_1, \ldots, e_n\}.
\]

**Proof.** Put \( e_1 = f_1/\|f_1\| \). Assuming \( e_1, \ldots, e_n \) have been constructed satisfying the conditions of the theorem, define

\[
e_{n+1} = \frac{f_{n+1} - \sum_{j=1}^{n} \langle f_{n+1}, e_j \rangle e_j}{\|f_{n+1} - \sum_{j=1}^{n} \langle f_{n+1}, e_j \rangle e_j\|}.
\]

(25.65)

The rest of the proof is left as an exercise. \( \square \)

25.10. 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. The first step in this direction is Bessel’s inequality, of fundamental importance in its own right. Before doing this, we record a useful computation regarding orthogonal projections.

**Theorem 25.18.** Let \( \{e_1, \ldots, e_n\} \) be an orthonormal set in \( H \), and let \( M = \text{span}\{e_1, \ldots, e_n\} \). Then \( M \) is closed, and the orthogonal projection onto \( M \) is given by

\[
Ph = \sum_{j=1}^{n} \langle h, e_j \rangle e_j.
\]

(25.66)

**Proof.** The proof that \( M \) is closed is left as an exercise. Now the right-hand side of (25.66) belongs to \( M \) by definition, so to compute \( P \), it is enough to show that if \( Ph \) is given by the formula (25.66), then \( (h - Ph) \perp M \). But this is easy: for \( 1 \leq m \leq n \),

\[
\langle h - Ph, e_m \rangle = \langle h, e_m \rangle - \sum_{j=1}^{n} \langle h, e_j \rangle \langle e_j, e_m \rangle
\]

(25.67)

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

(25.68)

\[
= \langle h, e_m \rangle - \langle h, e_m \rangle = 0.
\]

(25.69)

\( \square \)

**Theorem 25.19** (Bessel’s inequality). If \( \{e_1, e_2, \ldots\} \) is an orthonormal sequence in \( H \), then for all \( h \in H \)

\[
\sum_{n=1}^{\infty} |\langle h, e_n \rangle|^2 \leq \|h\|^2.
\]

(25.70)

**Proof.** Fix a vector \( h \) and a positive integer \( N \). Define

\[
h_N = h - \sum_{j=1}^{N} \langle h, e_j \rangle e_j.
\]

(25.71)
As we computed in the proof of Theorem 25.18, $h_N \perp e_j$ for all $j = 1, \ldots, N$. Then by the Pythagorean theorem (twice),

\[
\|h\|^2 = \|h_N\|^2 + \left\| \sum_{j=1}^{N} \langle h, e_j \rangle e_j \right\|^2
\]

(25.72)

\[
= \|h_N\|^2 + \sum_{j=1}^{N} \left| \langle h, e_j \rangle \right|^2
\]

(25.73)

\[
\geq \sum_{j=1}^{N} \left| \langle h, e_j \rangle \right|^2.
\]

(25.74)

Since this holds for all $N$, we are done. \hfill \Box

**Corollary 25.20.** 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

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

(25.75)

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

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

(25.76)

a contradiction. But now,

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

(25.77)

which is a countable union of finite sets, and therefore countable. \hfill \Box

We can now extend Bessel’s inequality to arbitrary (i.e., not necessarily countable) orthonormal sets, if we’re careful.

**Theorem 25.21.** If $E \subset H$ is an orthonormal set and $h \in H$, then

\[
\sum_{e \in E} |\langle h, e \rangle|^2 \leq \|h\|^2.
\]

(25.78)

**Remark:** The (possibly uncountable) sum in (25.78) is interpreted as follows: we know from Corollary 25.20 that at only countably many of the terms $\langle h, e \rangle$ are nonzero. Fix an enumeration of these terms, and sum the ordinary infinite series that results. Since terms of the series are positive, if it is convergent then it is absolutely convergent, and hence its sum is independent of the ordering, that is, independent of the enumeration we chose.

**Proof of Theorem 25.21.** The proof consists of combining Bessel’s inequality and the foregoing remark; the details are left as an exercise. \hfill \Box
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. Say the series converges absolutely if \( \sum_{n=1}^{\infty} \| h_n \| < \infty \).

Now, say that the series \( \sum_{n=1}^{\infty} h_n \) is unconditionally convergent if it converges to a sum \( h \), and if \( \varphi : \mathbb{N} \to \mathbb{N} \) is a bijection of \( \mathbb{N} \), then \( \sum_{n=1}^{\infty} h_{\varphi(n)} \) also converges to \( h \). (In other words, every reordering of the series converges, and to the same sum.) Note that for ordinary scalar series, unconditional convergence is equivalent to absolute convergence, but this is not so in Hilbert space.

**Theorem 25.22.** If \( E \subset H \) is an orthonormal set and \( h \in H \), then the series

\[
\sum_{e \in E} \langle h, e \rangle e
\]

is unconditionally convergent.

**Proof.** Let \( F = \{ e \in E : \langle h, e \rangle \neq 0 \} \). By Corollary 25.20, we know \( F \) is countable. We divide the proof into two parts: first, we fix an enumeration \( e_1, e_2, \ldots \) of \( F \), and show that the resulting series converges. We then show that the sum was independent of the choice of enumeration.

**Part I:** Let \( \{e_1, e_2, \ldots \} \) be an enumeration of \( F \) and consider the series

\[
\sum_{n=1}^{\infty} \langle h, e_n \rangle e_n.
\]

Let \( \epsilon > 0 \). By Bessel’s inequality, there exists an integer \( N \) so that

\[
\sum_{n=N+1}^{\infty} |\langle h, e_n \rangle|^2 < \epsilon.
\]

Let \( s_j \) denote the partial sum \( \sum_{n=1}^{j} \langle h, e_n \rangle e_n \). Then for \( k > j \geq N \), we have by the Pythagorean theorem

\[
\| s_k - s_j \|^2 = \left\| \sum_{n=j+1}^{k} \langle h, e_n \rangle e_n \right\|^2
\]

\[
= \sum_{n=j+1}^{k} |\langle h, e_n \rangle|^2
\]

\[
\leq \sum_{n=N+1}^{\infty} |\langle h, e_n \rangle|^2
\]

\[
< \epsilon.
\]

Thus \( (s_j) \) is a Cauchy sequence, which converges since \( H \) is complete. So, the series (25.80) converges.

**Part II:** Let \( e_1, e_2, \ldots \) be the enumeration used in Part I, and let \( g \) denote the sum of the series (25.80). Now let \( \varphi : \mathbb{N} \to \mathbb{N} \) be any bijection, and consider the corresponding
The proof in Part I shows that the series
\[ \sum_{n=1}^{\infty} \langle h, e_{\varphi(n)} \rangle e_{\varphi(n)} \]  
also converges; call its sum \( g' \). Let \( t_j \) denote the partial sum \( \sum_{n=1}^{j} \langle h, e_{\varphi(n)} \rangle e_{\varphi(n)} \). To show that \( g' = g \), it suffices to show that given \( \epsilon > 0 \), there exists an integer \( M \) so that if \( j \geq M \), then \( \| s_j - t_j \| < \epsilon \) (why?).

Given \( \epsilon > 0 \), choose \( N \) so that (25.81) holds. Now choose \( M \geq N \) so that \( \{ 1, 2, \ldots, N \} \subseteq \{ \varphi(1), \varphi(2), \ldots, \varphi(M) \} \). (25.87) For each \( j \geq M \), let \( G_j \) be the symmetric difference of the sets \( \{ 1, 2, \ldots, j \} \) and \( \{ \varphi(1), \varphi(2), \ldots, \varphi(j) \} \) (that is, their union minus their intersection). Since \( j \geq M \), the set \( G_j \) is disjoint from \( \{ 1, 2, \ldots, N \} \). It follows that
\[
\| s_j - t_j \| ^2 = \sum_{n \in G_j} | \langle h, e_n \rangle | ^2 
\leq \sum_{N+1}^{\infty} | \langle h, e_n \rangle | ^2 
< \epsilon, 
\]  
as desired. This finishes the proof. □

We are almost ready to prove the main theorem of this section. Given a set of vectors \( S \subset H \), the closed linear span of \( S \), denoted \( \vee S \), is the closure in \( H \) of the vector subspace \( \text{span}(S) \). (As an exercise, prove that the closure of a subspace is again a subspace).

**Theorem 25.23.** If \( E \subset H \) is an orthonormal set, then the following are equivalent:

a) \( E \) is a maximal orthonormal set in \( H \).

b) If \( h \perp E \), then \( h = 0 \).

c) \( \vee E = H \).

d) \( h = \sum_{e \in E} \langle h, e \rangle e \).

e) (Plancherel theorem) For all \( g, h \in H \), \( \langle g, h \rangle = \sum_{e \in E} \langle g, e \rangle \langle e, h \rangle \).

f) (Parseval identity) For all \( h \in H \), \( \| h \| ^2 = \sum_{e \in E} | \langle h, e \rangle | ^2 \).

**Proof.** The theorem says that any of these six equivalent conditions can be taken to be the definition of an orthonormal basis. a) \( \implies \) b): This is immediate from the maximality of \( E \).

b) \( \implies \) c): Let \( M = \vee E \). Since \( M \) is a closed subspace, \( M^\perp = M \). By b), \( M^\perp = \{ 0 \} \), so \( M^\perp = H \).

c) \( \implies \) b): If \( h \perp E \), then \( h \perp \vee E \), hence \( h = 0 \).

b) \( \implies \) d): Let \( h' = \sum_{e \in E} \langle h, e \rangle e \). Then \( (h - h') \perp E \) (check this carefully!), so \( h = h' \).

d) \( \implies \) e): Exercise.

e) \( \implies \) f): Immediate; put \( g = h \).

f) \( \implies \) a): If \( E \) is not a basis, choose \( h \perp E \) with \( \| h \| = 1 \). Then
\[
1 = \| h \| ^2 = \sum_{e \in E} | \langle h, e \rangle | ^2 = 0, 
\]  
a contradiction. □
Definition 25.24. An orthonormal basis for a Hilbert space is a maximal orthonormal set.

Theorem 25.25. Every Hilbert space 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. The Gram-Schmidt process shows that \( \mathcal{E} \) is nonempty. If \( (E_a) \) is an ascending chain in \( \mathcal{E} \), then it is straightforward to verify that \( \cup_a E_a \) is an orthonormal set, and is an upper bound for \( (E_a) \). Thus by Zorn’s lemma, \( \mathcal{E} \) has a maximal element. □

Theorem 25.26. Any two bases of a Hilbert space \( H \) have the same cardinality.

Proof. Let \( E, F \) be two orthonormal bases. The case where either \( E \) or \( F \) is finite is left as an exercise. So, assume both \( E \) and \( F \) are infinite. Fix \( e \in E \) and consider the set

\[
F_e = \{ f \in F \mid \langle f, e \rangle = 0 \}.
\]

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. By symmetry, \( |F| \leq |E| \) and the proof is complete. □

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

Definition 25.27. The (orthogonal) dimension of a Hilbert space \( H \) is the cardinality of any orthonormal basis, and is denoted \( \dim H \).

It can be shown that if \( H \) has finite dimension as a vector space, then it also has finite orthogonal dimension, and the two dimensions are equal. (Exercise). This is false in infinite dimensions however.

25.11. 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 separable case and just study weakly convergent sequences.

Definition 25.28. Let \( H \) be a separable Hilbert space and \((h_n)\) a sequence in \( H \). Say that \( h_n \) converges weakly to \( h \in H \) if for all \( g \in H \),

\[
\langle h_n, g \rangle \to \langle h, g \rangle.
\]

(25.94)

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

Proposition 25.29. 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) \( \langle h_n, g \rangle \to \langle h, g \rangle \) for all \( g \in H \).
ii) \( \lim_{n} \langle h_n, e_j \rangle \) exists for each \( j \).

**Proof.** Suppose \( h_n \to h \) weakly. Then for each \( n \) we have a linear functional

\[
L_n(g) = \langle g, h_n \rangle.
\]

(25.95)

It follows that for each fixed \( g \), the sequence \( |L_n(g)| \) is uniformly bounded. Thus, the family of linear functionals \( (L_n) \) is pointwise bounded, so by the Principle of Uniform boundedness, we have \( \sup \|h_n\| = \sup \|L_n\| < \infty \). This proves (i), and (ii) is immediate from the definition of weak convergence.

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

\[
h_j = \lim_{n} \langle h_n, e_j \rangle.
\]

(25.96)

We will show that \( \sum_j |h_j|^2 \leq M \) (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.

Fix a positive integer \( J \). Then

\[
\sum_{j=1}^{J} |h_j|^2 = \sum_{j=1}^{J} \lim_{n} |\langle h_n, e_j \rangle|^2 = \lim_{n} \sum_{j=1}^{J} |\langle h_n, e_j \rangle|^2 \leq \lim_{n} \sup \|h_n\|^2 \leq M^2
\]

(25.97)

using Bessel’s inequality. This proves that \( \sum_j |h_j|^2 \leq M \) and thus the series \( \sum_j h_j e_j \) is norm convergent; call its sum \( h \) and observe that \( \langle h, e_j \rangle = h_j \) for each \( j \). Also note that \( \|h\| \leq M \).

Now we prove that \( h_n \to h \) weakly. Fix \( g \in H \) and let \( \epsilon > 0 \). Since \( g = \sum_{j} \langle g, e_j \rangle e_j \) (where the series is norm convergent) we can choose an integer \( J \) large enough so that

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

(25.98)

Let \( g_0 = \sum_{j=1}^{J} \langle g, e_j \rangle e_j \) and write \( g = g_0 + g_1 \); so that \( \|g_1\| < \epsilon \). Then

\[
|\langle h_n - h, g \rangle| \leq |\langle h_n - h, g_0 \rangle| + |\langle h_n - h, g_1 \rangle| \leq 2M \epsilon.
\]

(25.99)

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

**Remark:** The above proof shows that if \( h_n \to h \) weakly, then \( \|h\| \leq \limsup \|h_n\| \), however it will not be the case that \( \|h\| = \lim \|h_n\| \) in general. This only occurs if in fact \( h_n \to h \) in norm; see Problem ??.

**Theorem 25.30** (Weak compactness of the unit ball in Hilbert space). If \( (h_n) \) is a bounded sequence in a Hilbert space \( H \), then \( (h_n) \) has a weakly convergent subsequence.

**Proof.** Using the previous proposition, it suffices to fix an orthonormal basis \( (e_j) \) and produce a subsequence \( h_{n_k} \) 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 26.12)

33
26. Problems

Problem 26.1. Prove the complex form of the polarization identity: if $H$ is a Hilbert space over $\mathbb{C}$, then for all $g, h \in H$

$$\langle g, h \rangle = \frac{1}{4} \left( \| g + h \|^2 - \| g - h \|^2 + i \| g + ih \|^2 - i \| g - ih \|^2 \right)$$  \hspace{1cm} (26.1)

Problem 26.2. (Adjoint operators) Let $H$ be a Hilbert space and $T : H \to H$ a bounded linear operator.

a) Prove that there is a unique bounded operator $T^* : H \to H$ satisfying $\langle Tg, h \rangle = \langle g, Th \rangle$ for all $g, h \in H$, and $\| T^* \| = \| T \|$.  

b) Prove that 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 that $\| T^*T \| = \| T \|^2$.

d) Prove that $\ker T$ is a closed subspace of $H$, $(\text{ran} T)^\perp = (\ker T^*)^\perp$ and $\ker T^* = (\text{ran} T)^\perp$.

Problem 26.3. 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 that if $E \subset H$ is an orthonormal set and $T$ is an isometry, then $T(E)$ is an orthonormal set in $K$.

d) Prove that 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

$$S(a_0, a_1, a_2, \ldots) = (0, a_0, a_1, \ldots)$$  \hspace{1cm} (26.2)

Prove that $S$ is an isometry, but not unitary. Compute $S^*$ and $SS^*$.

Problem 26.4. 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 that $\ell^2(I)$ is isometrically isomorphic to $\ell^2(J)$ if and only if $I$ and $J$ have the same cardinality. (Hint: use Problem 26.3(c).)

b) Prove that if $H$ is any Hilbert space, then $H$ is isometrically isomorphic to $\ell^2(J)$ for some set $J$.

Problem 26.5. Let $(X, \mathcal{M}, \mu)$ be a $\sigma$-finite measure space. Prove that the simple functions that belong to $L^2(\mu)$ are dense in $L^2(\mu)$.

Problem 26.6. (The Fourier basis) Prove that the set $E = \{ e_n(t) := e^{2\pi int} | n \in \mathbb{Z} \}$ is an orthonormal basis for $L^2[0, 1]$. (Hint: use the Stone-Weierstrass theorem to prove that the set of trigonometric polynomials $P = \{ \sum_{n=-M}^N c_n e^{2\pi int} \}$ is uniformly dense in the space of continuous functions $f$ on $[0, 1]$ which 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 26.7. (The Haar basis) Let $f_0 := 1_{[0,1)}$ and consider the function $f_1 := 1_{[0, \frac{1}{2})} - 1_{[\frac{1}{2}, 1)}$ on $\mathbb{R}$. Inductively define a sequence of functions $f_{n+1}(x) = f_n(2x) + f_n(2x - 1)$. (Draw a
picture of the first few of these to see what is going on). Note that each \( f_n \) is supported on \([0, 1]\). Prove that \((f_n)_{n=0}^\infty\) is an orthonormal basis for \(L^2[0, 1]\).

**Problem 26.8.** 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 that the functions \(h_{mn}(x) := 1_{[m,m+1]}(x)g_n(x-m)\), \(n \in \mathbb{N}, m \in \mathbb{Z}\) form an orthonormal basis of \(L^2(L)\). (Thus \(L^2(\mathbb{R})\) is separable.)

**Problem 26.9.** Let \((X, \mathcal{M}, \mu), (Y, \mathcal{N}, \nu)\) be \(\sigma\)-finite measure spaces, and let \(\mu \times \nu\) denote the product measure. Prove that 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 orthonormal basis for \(L^2(\mu \times \nu)\). (Problem 26.5 may be helpful.) Use this to construct an orthonormal basis for \(L^2(\mathbb{R}^n)\), and conclude that \(L^2(\mathbb{R}^n)\) is separable.

**Problem 26.10.** (Weak Convergence)

a) Prove that if \(h_n \to h\) in norm, then also \(h_n \to h\) weakly. (Hint: Cauchy-Schwarz.)

b) Prove that 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 that \(h_n \to h\) in norm if and only if \(h_n \to h\) weakly and \(\|h_n\| \to \|h\|\).

**Problem 26.11.** Suppose \(H\) is infinite-dimensional. Prove that if \(h \in H\) and \(\|h\| \leq 1\), then there is a sequence \(h_n\) in \(H\) with \(\|h_n\| = 1\) for all \(n\), and \(h_n \to h\) weakly.

**Problem 26.12.** Prove Theorem 27.1.

### 27. \(L^p\) Spaces

**Definition 27.1.** Let \((X, \mathcal{M}, \mu)\) 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

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

(27.1)

where we identify \(f\) and \(g\) when \(f = g\) a.e.

First by the inequality

\[
|f + g|^p \leq (2 \max(|f|, |g|))^p \leq 2^p (|f|^p + |g|^p)
\]

(27.2)

and the monotonicity of the integral, we see that \(L^p\) is a vector space for all \(0 < p < \infty\).

**Proposition 27.2.** If \(1 \leq p < \infty\), then \(\|f\|_p\) is a norm on \(L^p\).

**Proof.** Trivially \(\|f\|_p \geq 0\), and if \(\|f\|_p = 0\) then \(f = 0\) a.e., which means \(f = 0\) since we have identified functions that agree almost everywhere. The homogeneity \(\|cf\|_p = |c|\|f\|_p\) is evident from the definition. To prove the triangle inequality, let \(f, g \in L^p\). To prove the triangle inequality we make several reductions. First, since \(|f + g|^p \leq (|f| + |g|)^p\), we can assume \(f, g \geq 0\). Next, we can scale \(f\) and \(g\) by the same constant factor so that \(\|f\|_p + \|g\|_p = 1\) and using homogeneity again we can write \(f = tF, g = (1-t)G\) with \(F, G \in L^p\) and \(\|F\|_p = \|G\|_p = 1\). With this setup proving the triangle inequality amounts to proving

\[
\int_X |tF + (1-t)G|^p \, d\mu \leq 1.
\]

(27.3)
But now note that the function $x \to |x|^p$ is convex on $[0, +\infty)$, that is,
\[ |tx_1 + (1-t)x_2|^p \leq t|x_1|^p + (1-t)|x_2|^p \] (27.4)
for all $x_1, x_2 \geq 0$. Applying this in our case we have
\[ \int_X |tF + (1-t)G|^p \, d\mu \leq \int_X t|F|^p + (1-t)|G|^p \, d\mu = 1 \] (27.5)
as desired.

\textbf{Remark:} 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, we have that $a^p + b^p > (a+b)^p$ for all $a, b > 0$. Now take two disjoint sets $E, F$ of finite, positive measure. We have
\[
\|1_E + 1_F\|_p = (\mu(E) + \mu(F))^{1/p} > \mu(E)^{1/p} + \mu(F)^{1/p} = \|1_E\|_p + \|1_F\|_p. \tag{27.6}
\]

When $p = \infty$, we define the space $L^\infty(\mu)$ to be the space of all essentially bounded measurable functions, again identifying functions that agree almost everywhere. Recall that a function is \textit{essentially bounded} if there exists a number $M < \infty$ such that $|f(x)| \leq M$ for a.e. $x \in X$. We can then define $\|f\|_\infty$ to be the smallest $M$ with this property. In particular we have
\[
\|f\|_\infty = \inf\{M : |f(x)| \leq M \text{ a.e.} \} = \sup\{M : \mu(\{x : |f(x)| > M\}) > 0\}. \tag{27.7}
\]
It is straightforward to check that $\|f\|_\infty$ is a norm on $L^\infty(\mu)$, and that $f_n \to f$ in $L^\infty$ if and only if $f_n$ converges to $f$ essentially uniformly; and it follows from this that $L^\infty$ is complete. (Problem 28.5.)

\textbf{Example 27.3.} Suppose $f = A1_E$ 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 $\|f\|_\infty = \lim_{p \to \infty} \|f\|_p$. In fact, this is true generically (see Problem ??).

We next show that $L^p$ is complete for $1 \leq p < \infty$; the proof is essentially the same as the ones already given in the case of $L^1$ and $L^2$:

\textbf{Theorem 27.4.} For $1 \leq p \leq \infty$, $L^p$ is a Banach space.

\textit{Proof.} The $p = \infty$ case is Problem 28.5. For $1 \leq p < \infty$, as before we use Proposition 19.3 and the monotone convergence theorem. Suppose $f_n$ is a sequence in $L^p$ and $M = \sum_{n=1}^\infty \|f_n\|_p < \infty$. Then $\|\sum_{n=1}^N f_n\|_p \leq M$ for all $N$, so by MCT the series $\sum_{n=1}^\infty f_n(x)$ is absolutely convergent for almost every $x$. Let $f$ denote its sum; then $\|f\|_p \leq M < \infty$, so $f \in L^p$. Finally we need to check that the sum converges to $f$ in the $L^p$ norm. But
\[
\left\| f - \sum_{n=1}^N f_n \right\|_p = \left\| \sum_{n=N+1}^\infty f_n \right\|_p \leq \sum_{n=N+1}^\infty \|f_n\|_p \to 0 \text{ as } N \to \infty \tag{27.8}
\]
and we are done. \qed

Suppose $p < \infty$. Then $f \in L^p$ if and only if $|f|^p \in L^1$, so by applying Markov’s inequality to $|f|^p$ we obtain
Proposition 27.5 (Chebyshev’s inequality). Suppose $f \in L^p(\mu)$ and $t > 0$. Then
\[
\mu(\{x : |f| > t\}) \leq \frac{1}{t^p} \int_X |f|^p d\mu = \left(\frac{\|f\|_p}{t}\right)^p.
\] (27.9)

One consequence of Chebyshev’s inequality is that if $f \in L^p$ (for finite $p$) then $f$ is essentially supported on a $\sigma$-finite set. (Say $f$ is essentially supported on $E$ if $f = 0$ a.e. on the complement of $E$.) Indeed, if we let $E_n = \{|f| > \frac{1}{n}\}$ then by Chebyshev each $E_n$ has finite measure, and $f$ is essentially supported on $\bigcup_{n=1}^\infty E_n$. Thus many questions about $L^p$ functions can be reduced to the $\sigma$-finite case. In particular this is usually possible when one is dealing with at most countably many functions at a time.

As in the case of $L^1$, we have the following density result:

Proposition 27.6. Simple functions are dense in $L^p$ for all $1 \leq p \leq \infty$.

Proof. We prove the $p < \infty$ case and leave $p = \infty$ as an exercise. Let $f \in L^p$. It is straightforward to check that $\text{Re} f, \text{Im} f$ are in $L^p$, as are the positive and negative parts when $f$ is real valued. Thus we may assume $f$ is unsigned. Next, by monotone convergence we see that $f$ can be approximated in $L^p$ by bounded functions (take $f_n = \min(f, n)$), so we can assume $f$ is bounded; and again by monotone convergence we can approximate $f$ in $L^p$ by functions supported on sets of finite measure (take $f_n$ to be $1_{E_n} f$, 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 (see ?? from last semester’s notes) and it is easy to verify that if $f_n \to f$ essentially uniformly on a finite measure space, then $f_n \to f$ in $L^p$ also (the proof is the same as in the $L^1$ case). \ \qed

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 and is left as an exercise.

Proposition 27.7. 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?)

There is an important algebraic relation among the $L^p$ spaces, expressed by the following inequality:

Theorem 27.8 (Hölder’s inequality). Suppose $1 \leq p, q \leq \infty$ 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
\[
\|fg\|_1 \leq \|f\|_p \|g\|_q.
\] (27.10)

Proof. The proof is very simple in the case $p = \infty$ or $q = \infty$; (in which case the other exponent is 1), so we assume $1 < p, q < \infty$. We may assume $f, g$ are both nonzero, and by homogeneity we may normalize so that $\|f\| = \|g\|_q = 1$. We must now show that
\[
\int_X |fg| \, d\mu \leq 1.
\] (27.11)

We again use convexity, this time of the exponential function $t \to e^t$. This implies the convexity of the function $H(t) = |f(x)|^p(1-t)|g(x)|^q t$ for each fixed $x$. (To see this, assume $f(x), g(x) \neq 0$ and rearrange the function as $H(t) = |f(x)|^p \cdot \exp(ct)$, where $c =
log(|f(x)|^p/|g(x)|^q).) In particular, the convexity of $H$ means that (using the fact that $\frac{1}{p} + \frac{1}{q} = 1$)
\[
H\left(\frac{1}{q}\right) = H\left(\frac{1}{p} \cdot 0 + \frac{1}{q} \cdot 1\right) \leq \frac{1}{p}H(0) + \frac{1}{q}H(1) \tag{27.12}
\]
which says
\[
|f(x)g(x)| \leq \frac{1}{p}|f(x)|^p + \frac{1}{q}|g(x)|^q \tag{27.13}
\]
Integrating this last expression $d\mu$ and applying the normalizations on $p, q, f, g$ gives (27.11).

**Remark:** 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 $f = 1_E, g = 1_F$. Then $\|fg\|_1 = \mu(E \cap F)$ and
\[
\|f\|_p \|g\|_q = \mu(E)^{1/p}\mu(F)^{1/q}, \tag{27.14}
\]
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:** A more general version of Hölder’s inequality says the following: let $1 \leq p, q \leq \infty$ and $\frac{1}{p} = \frac{1}{p} + \frac{1}{q}$. If $f \in L^p$ and $g \in L^q$, then $fg \in L^r$, and
\[
\|fg\|_r \leq \|f\|_p \|g\|_q \tag{27.15}
\]
(The proof is Problem 28.1.

One important consequence of Hölder’s inequality is the following: given $1 \leq p \leq \infty$, we let $p'$ denote the conjugate exponent to $p$, namely the $p'$ for which $\frac{1}{p} + \frac{1}{p'} = 1$. Then for any fixed $g \in L^q$, the linear functional $\lambda_g$ defined by
\[
\lambda_g(f) = \int_X fg \, d\mu \tag{27.16}
\]
is bounded on $L^p$, and $\|\lambda_g\| \leq \|g\|_q$. One of the most important facts about $L^p$ spaces is that, in most cases, the converse is true:

**Theorem 27.9.** Let $1 \leq p < \infty$ and suppose $\mu$ is a $\sigma$-finite measure. Then if $\lambda : L^p(\mu) \to \mathbb{C}$ is a bounded linear functional, there exists a unique $g \in L^{p'}(\mu)$ such that $\lambda = \lambda_g$; moreover $\|\lambda\| = \|g\|_{p'}$.

**Proof.** The idea of the proof is to use the linear functional $\lambda$ to define a set function $E \to \lambda(1_E)$. One proves that this defines a measure absolutely continuous to $\mu$, from which we obtain $g$ via the Radon-Nikodym theorem. Finally we must show that this $g$ belongs to $L^{p'}$. We now turn to the details.

We first prove the theorem under the assumption that $\mu$ is finite. Let $u = \text{Re}(\lambda)$. Define
\[
\nu(E) := u(1_E) \tag{27.17}
\]
Since $\mu$ is finite, $1_E \in L^p$ for all measurable $E$, so $u$ takes only finite, real values. Next, since $\lambda$ is bounded, $u$ is also bounded, hence continuous. If $(E_n)$ is a sequence of disjoint measurable sets, then the monotone convergence theorem shows that the series $\sum_{n=1}^\infty 1_{E_n}$ is convergent in $L^p$, so by the linearity and continuity of $\nu$ we have $\nu(\bigcup_{n=1}^\infty E_n) = \sum_{n=1}^\infty \nu(E_n)$. Thus $\nu$ is countably additive, and hence defines a signed measure on $(X, \mathcal{M})$. Moreover, if $\mu(E) = 0$, then $1_E = 0$ in $L^p$ so $\nu(E) = 0$. Thus, the measure $\nu$ is absolutely continuous to
µ. Let \( g_1 \) denote the Radon-Nikodym derivative \( d\nu/d\mu \). Applying the same arguments to \( \text{Im}\lambda \), we define \( g_2 \) similarly, and put \( g = g_1 + ig_2 \). Moreover, the finiteness of the measure \( \nu \) shows that \( g \in L^1(\mu) \).

Next we prove that \( g \in L^{p'} \). For this it suffices to treat the real and imaginary parts separately; and in turn considering positive and negative parts we may assume \( g \geq 0 \). First assume \( p > 1 \). If we could choose \( f \in L^p \) so that \( fg = g^{p'} \) (that is, if we knew that \( g^{p'-1} \) belonged to \( L^p \)) we would be done. However from the identity \( pp' - p = p' \) it follows that \( g^{p'-1}p = g^{p'} \), so this argument is circular. However, since we have already shown that \( g \in L^1 \), by vertical truncation (Problem 28.3) it follows that \( \min(g, N) \) belongs to \( L^q \) for all \( q \geq 1 \) and all \( N \geq 0 \). If we now put \( f_N = \min(g, N)^{p'-1} \), then \( f_N \in L^p \). Thus, testing \( g \) against \( f_N \), we have

\[
\|f_N\|\|\lambda_g\| \geq \lambda_g(f_N) = \int_X \min(g, N)^{p'-1} g \, d\mu \geq \int_X \min(g, N)^{p'} \geq \|\min(g, N)\|^{p'}. \tag{27.18}
\]

On the other hand, \( \|f_N\| = \|\min(g, N)\|^{p'-1} \), and combining this with (27.18) we see that \( \|\min(g, N)\|^{p'} \leq \|\lambda_g\| \). Letting \( N \to \infty \), by monotone convergence we conclude that \( \|g\|^{p'} \leq \|\lambda_g\| < \infty \), and since Hölder gives the reverse inequality, we have \( \|\lambda_g\| = \|g\|^{p'} \).

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

\[
\|\lambda_g\| = \|\lambda_g\| f_t \geq \int_X f_t g \, d\mu = \frac{1}{\mu(E_t)} \int_{E_t} g \, d\mu \geq t \tag{27.19}
\]

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 \( \|g\|_{\infty} \leq \|\lambda_g\| \).

The proof that \( g \) is unique is similar to the uniqueness part of the proof of the Lebesgue-Radon-Nikodym theorem and is left as an exercise.

Finally, we move from the finite to the \( \sigma \)-finite case. Let \((X, \mathcal{M}, \mu)\) be a \( \sigma \)-finite measure space. We can write \( X \) as an increasing union of sets \( X_n \) of finite measure. Let \( \lambda \) be a bounded linear functional on \( L^p(\mu) \) and let \( \mu_n \) be the restriction of \( \mu \) to \( X_n \). By restriction we get bounded linear functionals \( \lambda_n \) on each \( L^p(\mu_n) \), and \( \|\lambda_n\| \leq \|\lambda\| \). By the arguments above, on each \( X_n \) there is a function \( g_n \in L^{p'}(\mu_n) \) such that \( \lambda_n = \lambda g_n \). By uniqueness, when \( n \geq m \) we must have \( g_n|_{X_m} = g_m \) almost everywhere on \( X_m \). The \( g_n \) then converge pointwise a.e. on \( X \) to a function \( g \), and by monotone convergence we have \( g \in L^{p'}(\mu) \), since \( \int_X |g_n|^{p'} \, d\mu = \|\lambda g_n\| \leq \|\lambda g\| \).

\[\square\]

**Corollary 27.10.** For \( 1 < p < \infty \), \( L^p \) is reflexive.

The theorem above fails in general when \( p = \infty \). Certainly every \( g \in L^1 \) defines a bounded linear functional on \( L^\infty \) by the formula \( \lambda_g(f) = \int_X f \overline{g} \, d\mu \), but unless \( \mu \) is a finite sum of atoms, there exist bounded linear functionals on \( L^\infty \) that are not of this form. An abstract way to see this is that in when \( \mu \) is \( \sigma \)-finite, \( L^1(\mu) \) is separable, but \( L^\infty(\mu) \) is not separable if \( \mu \) is not a finite sum of atoms. Thus if it were the case that \((L^\infty)^* \cong L^1 \), then \( L^\infty \) would be separable by the result in Problem 22.6, a contradiction. Problem 28.10 gives a somewhat more explicit argument in the case of \( L^\infty(\mathbb{R}) \).
27.1. Distribution functions and weak $L^p$. Let $(X, \mathcal{M}, \mu)$ 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

$$\lambda_f(t) = \mu(\{x : |f(x)| > t\}). \quad (27.20)$$

To begin building an intuition about $\lambda_f$, we have the following lemma.

**Lemma 27.11.** Let $f = \sum_{j=1}^n c_j \mathbb{1}_{E_j}$ 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

$$\lambda_f(t) = \begin{cases} 
\mu(E_n) + \cdots + \mu(E_2) + \mu(E_1), & 0 \leq t < c_1, \\
\mu(E_n) + \cdots + \mu(E_2), & c_1 \leq t < c_2, \\
\cdots & \cdots \\
\mu(E_n), & c_{n-1} \leq t < c_n, \\
0 & t \geq c_n.
\end{cases} \quad (27.21)$$

**Proof.** Problem 28.11.

The basic properties of $\lambda_f$ are collected in the following proposition:

**Proposition 27.12.**

a) $\lambda_f$ is decreasing and right continuous.

b) [Monotonicity] If $|f| \leq |g|$ a.e., then $\lambda_f \leq \lambda_g$ everywhere.

c) [Monotone convergence] If $|f_n|$ increases to $|f|$ pointwise a.e., then $\lambda_{f_n}$ increases to $\lambda_f$.

d) [Subadditivity] If $f = g + h$, then $\lambda_f(t) \leq \lambda_g(t/2) + \lambda_h(t/2)$.

**Proof.** Let $E(t, f) = \{x : |f(x)| > t\}$. $\lambda_f$ is decreasing since $E(t, f) \subset E(s, f)$ when $s < t$, and right continuous since $E(t, f)$ is the increasing union of $E(t, f_n)$. (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 $E(t_n, f)$. Finally (d) is a pigeonhole argument: if $|f(x)| > t$, then either $|g(x)| > t/2$ or $|h(x)| > t/2$.

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

$$\nu((a, b]) := \lambda_f(a) - \lambda_f(b) \quad (27.22)$$

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 $\mu$.

**Proposition 27.13.** Suppose $\lambda_f(t) < \infty$ for all $t > 0$ and $\varphi \geq 0$ is an unsigned Borel function. Then

$$\int_X \varphi(|f|) \ d\mu = -\int_0^\infty \varphi(t) \ d\lambda_f(t). \quad (27.23)$$

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

$$\int_X \varphi(|f|) \ d\mu = \mu(a < |f| \leq b) = \lambda_f(b) - \lambda_f(a); \quad (27.24)$$
on the other hand
\[-\int_0^\infty \varphi(t) d\lambda_f(t) = -\nu((a,b)) = -(\lambda_f(a) - \lambda_f(b)) \tag{27.25}\]
so (27.23) holds when \(\varphi = 1_{(a,b)}\). Since both sides are linear in \(\varphi\), 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 (27.23) itself but its “integrated-by-parts” form:

**Proposition 27.14.** If \(0 < p < \infty\) then
\[
\int_X |f|^p d\mu = p \int_0^\infty t^{p-1} \lambda_f(t) dt. \tag{27.26}
\]

**Proof.** This can be proved using the previous proposition and integration by parts for Lebesgue-Stieltjes measures, or directly as follows: first, 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 27.11. For general \(f\), take a sequence of simple functions \(f_n\) increasing to \(|f|\); then the formula holds by Lemma 27.11, Proposition 27.12(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
\[
\mu(\{|f| > t\}) \leq \frac{1}{t^p} \int_X |f|^p d\mu \tag{27.27}
\]
or rearranging
\[
t^p \lambda_f(t) \leq \|f\|_p^p \quad \text{for all } t > 0. \tag{27.28}
\]
For general \(f\), say that \(f\) belongs to weak \(L^p\) if
\[
(\sup_{t>0} t^p \lambda_f(t))^{1/p} := [f]_p < \infty. \tag{27.29}
\]
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 27.14 we find
\[
\int_X |f|^p d\mu = p \int_0^\infty t^{p-1} \lambda_f(t) dt \leq [f]_p \int_0^\infty t^{-1} dt \tag{27.30}
\]
and the integral is just barely divergent.

### 27.2. 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
\[
(Mf)(x) := \sup_{r>0} \frac{1}{m(B(r,x))} \int_{B(r,x)} |f(y)| dy \tag{27.31}
\]
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 27.15.** If \( f \in L^p(\mathbb{R}^n) \), \( 1 < p < \infty \), then \( Mf \in L^p \) and

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

(27.32)

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

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

(27.33)

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

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

(27.34)

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

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

(27.35)

Now write \( g = Mf \), then \( m(E_t) \) is just the distribution function \( \lambda_g(t) \). By Proposition 27.14 we have

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

(27.36)

Using (27.35) we obtain

\[
\|Mf\|_p^p = p \int_0^\infty t^{p-1} m(E_t) \, dt \leq p \int_0^\infty t^{p-1} \left( \frac{2\cdot 3^n}{t} \int_{|f| > t/2} |f(y)| \, dy \right) \, dt
\]  

(27.37)

We apply Fubini to the double integral and integrate \( dt \) first. This gives

\[
p \int_0^\infty t^{p-1} \left( 2 \cdot 3^n \int_{|f| > t/2} |f(y)| \, dy \right) \, dt = \int_{\mathbb{R}^n} |f(y)| \left( \int_0^{2|f(y)|} t^{p-2} \, dt \right) \, dy
\]  

(27.38)

The inner integral is

\[
\int_0^{2|f(y)|} t^{p-2} \, dt = \frac{1}{p - 1} 2^{p-1} |f(y)|^{p-1}
\]  

(27.39)

so we have finally

\[
\|Mf\|_p \leq \frac{2p \cdot 3^n p}{p - 1} \int_{\mathbb{R}^n} |f(y)|^p \, dy
\]  

(27.40)

which finishes the proof. \( \square \)
27.3. 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 \((\mathcal{X}, \mathcal{M}, \mu)\); \( L^p \) always refers to \( L^p(\mu) \).

We need a few definitions. Let \( 1 \leq p, q \leq \infty \). A mapping \( T : L^p \to L^q \) is of type \((p, q)\) if there is a constant \( A > 0 \) so that

\[
\|Tf\|_q \leq A\|f\|_p
\]  

(27.41)

for all \( f \in L^p \). The mapping is of weak type \((p, q)\) if

\[
\mu(\{x : |Tf(x)| > t\} \leq \left( \frac{A\|f\|_p}{t} \right)^q, \quad \text{for } q < \infty
\]  

(27.42)

where the constant \( A \) does not depend on \( f \) or \( t \). (In other words, \( T \) maps \( L^p \) into weak \( L^q \), with \( \|Tf\|_q \leq A\|f\|_p \).)

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 sublinear if \( |Tf(x)| \leq |Tg(x)| + |Th(x)| \) 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 27.16** (Marcinkiewicz interpolation theorem (special case)). Suppose that \( 1 < r \leq \infty \). 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

\[
f_1(x) := \begin{cases} f(x) & \text{if } |f(x)| > t \\ 0 & \text{if } |f(x)| \leq t \end{cases}
\]  

(27.43)

\[
f_2(x) := \begin{cases} f(x) & \text{if } |f(x)| \leq t \\ 0 & \text{if } |f(x)| > t \end{cases}
\]  

(27.44)

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 subadditivity of \( \lambda \)

\[
\lambda_{Tf}(t) \leq \lambda_{Tf_1}(t/2) + \lambda_{Tf_2}(t/2)
\]  

(27.45)

and so by the assumption that \( T \) is of weak type \((1, 1)\) and \((r, r)\),

\[
\lambda_{Tf}(t) \leq \frac{A_1}{t/2} \int |f_1(x)| \, d\mu(x) + \frac{A_r}{(t/2)^r} \int |f_2(x)|^r \, d\mu(x)
\]  

(27.46)
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

$$
\lambda_{Tf}(t) \leq \frac{2A_1}{t} \int_{|f| > t} |f(x)| \, d\mu(x) + \frac{(2A_r)^r}{t^r} \int_{|f| \leq t} |f(x)|^r \, d\mu(x) \quad (27.47)
$$

Using the formula $\|Tf\|_p = p \int_0^\infty t^{p-1} \lambda_{Tf}(t) \, dt$ we multiply both sides of (27.47) by $pt^{p-1}$ and integrate $dt$. To handle first integral in (27.47) we observe

$$
\int_0^\infty t^{p-1} t^{-1} \int_{|f| > t} |f(x)| \, d\mu(x) \, dt = \int_X |f| \int_0^{|f|} t^{p-2} \, dt \, d\mu(x) \quad (27.48)
$$

$$
= \frac{1}{p-1} \int_X |f| |f|^{p-1} \, d\mu \quad (27.49)
$$

since $p > 1$, similarly for the second integral

$$
\int_0^\infty t^{p-1} t^{-r} \int_{|f| \leq t} |f(x)|^r \, d\mu(x) \, dt = \int_X |f|^r \int_0^{|f|} t^{p-1-r} \, dt \, d\mu(x) \quad (27.50)
$$

$$
= \frac{1}{r-p} \int_X |f|^r |f|^{p-r} \, d\mu \quad (27.51)
$$

Putting these together we find that

$$
\|Tf\|_p \leq A_p \|f\|_p, \quad \text{with} \quad A_p = \left( \frac{2^r A_1}{p-1} + \frac{(2A_r)^r}{r-p} \right) p. \quad (27.52)
$$

\[ \square \]

### 27.4. 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 27.17** (Minkowski’s inequality for integrals). Let $(X, \mathcal{M}, \mu)$ and $(Y, \mathcal{N}, \nu)$ be $\sigma$-finite measure spaces and let $f$ be a jointly measurable function on $X \times Y$.

a) If $f$ is unsigned and $1 \leq p < \infty$, then

$$
\left[ \int_X \left( \int_Y |f(x, y)|^p \, d\nu(y) \right)^{\frac{1}{p}} \, d\mu(x) \right]^{1/p} \leq \int_Y \left[ \int_X |f(x, y)|^p \, d\mu(x) \right]^{\frac{1}{p}} \, d\nu(y) \quad (27.53)
$$

b) Suppose $1 \leq p \leq \infty$, $f(\cdot, y) \in L^p(\mu)$ 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

$$
\left\| \int f(\cdot, y) \, d\nu(y) \right\|_p \leq \int \|f(\cdot, y)\|_p \, d\nu(y). \quad (27.54)
$$

**Proof.** (b) follows immediately from (a) (replacing $f$ with $|f|$ and applying Fubini’s theorem). To prove (a), note that this is just Tonelli’s theorem when $p = 1$. For the case $1 < p < \infty$, let $h(x, y) = \int f(x, y) \, d\nu(y)$ and fix $g \in L^q(\mu)$ where $q$ is the conjugate index to $p$. If the
right-hand side of (27.53) is infinite, there is nothing to prove. If it is finite, then the function 
\( f(\cdot, y) \) belongs to \( L^p(\mu) \) for almost every \( y \), so by Fubini and Hölder, 
\[
\int h(x)|g(x)|\,d\mu(x) = \int \int f(x,y)|g(x)|\,d\mu(x)d\nu(y) 
\leq \int \left[ \int f(x,y)^p\,d\mu(x) \right]^{1/p} d\nu(y)\|g\|_q. 
\]  
(27.56)
Thus the linear functional \( g \to \int hg\,d\mu \) is bounded on \( L^q(\mu) \) with norm at most 
\[
\int \left[ \int f(x,y)^p\,d\mu(x) \right]^{1/p} d\nu(y) 
\]  
(27.57)
so by Theorem 27.9, we see that \( h \in L^p(\mu) \) and obeys the estimate (27.53). \( \square \)

There are many important linear operators on \( L^p \) spaces that are expressible in the form 
\[
Tf(x) = \int K(x,y)f(y)\,d\nu(y). 
\]  
(27.58)
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 27.18** (Young's inequality). Let \( \mu, \nu \) be \( \sigma \)-finite positive measures on spaces \( X,Y \) respectively and let \( K(x,y) \) be a jointly measurable function on \( X \times Y \). Suppose there is a constant \( C \) so that 
\[
\sup_{x \in X} \int_Y |K(x,y)|\,d\nu(y) \leq C, \quad \sup_{y \in Y} \int_X |K(x,y)|\,d\mu(x) \leq C. 
\]  
(27.59)
If \( f \in L^p(\nu), 1 \leq p \leq \infty \) and \( Tf \) is defined by (27.58), then \( Tf \in L^p(\mu) \) and 
\[
\|Tf\|_p \leq C\|f\|_p. 
\]  
(27.60)
**Proof.** If either \( p = 1 \) or \( p = \infty \), the proof is straightforward and left as an exercise. Suppose \( 1 < p < \infty \). The idea is to split \( K \) as \( K = K^{1/q}K^{1/p} \) and apply Hölder’s inequality. So, let \( \frac{1}{p} + \frac{1}{q} = 1 \), then 
\[
|Tf(x)| \leq \int_Y |K(x,y)||f(y)|\,d\nu(y) \leq \left( \int_Y |K(x,y)|\,d\nu(y) \right)^{1/q} \left( \int_Y |K(x,y)||f(y)|^p\,d\nu(y) \right)^{1/p} 
\leq C^{1/q} \left( \int_Y |K(x,y)||f(y)|^p\,d\nu(y) \right)^{1/p} 
\]  
so using Tonelli’s theorem 
\[
\int_X |Tf(x)|^p\,d\mu(x) \leq C^{p/q} \int_X \left( \int_Y |K(x,y)||f(y)|^p\,d\nu(y) \right)\,d\mu(x) 
= C^{p/q} \int_Y |f(y)|^p \left( \int_X |K(x,y)|\,d\mu \right)\,d\nu(y) 
\leq C^{1+p/q} \int_Y |f(y)|^p\,d\nu(y) 
= C^p\|f\|_p^p 
\]  
45
and taking $p^{th}$ roots finishes.

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

$$Tf(x) := \int_{\mathbb{R}^n} g(x - y)f(y) \, dy$$  \hfill (27.61)

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, we have

$$\sup_{x \in \mathbb{R}^n} \int |g(x - y)| \, dy = \sup_{y \in \mathbb{R}^n} \int |g(x - y)| \, dx = \|g\|_1$$  \hfill (27.62)

so the hypotheses of Proposition 27.18 are satisfied with $C = \|g\|_1$. We conclude

**Corollary 27.19** (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 $g \ast f \in L^p(\mathbb{R}^n)$ and

$$\|g \ast f\|_p \leq \|g\|_1 \|f\|_p.$$  \hfill (27.63)

28. Problems

**Problem 28.1.** Prove the generalized Hölder inequality (27.15). (Hint: consider $F = \|f\|^r, G = \|g\|^r$.)

**Problem 28.2.** 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 28.3** (Truncation of $L^p$ functions). Suppose $f$ is an unsigned function in $L^p(\mu)$, $1 < p < \infty$. For $t > 0$ let

$$E_t = \{x : f(x) > t\}.$$  \hfill (28.1)

a) Show that for each real number $t > 0$, the **horizontal truncation** $1_{E_t} f$ 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 28.4.** Suppose $f \in L^{p_0} \cap L^\infty$ for some $p_0 < \infty$. Prove that $f \in L^p$ for all $p_0 \leq p \leq \infty$, and $\lim_{p_0 \to \infty} \|f\|_p = \|f\|_\infty$.

**Problem 28.5.** Prove that $f_n \to f$ in the $L^\infty$ norm if and only if $f_n \to f$ essentially uniformly, and that $L^\infty$ is complete.

**Problem 28.6.** Suppose $p_0 < p < p_1$ and $f \in L^{p_0} \cap L^{p_1}$. Prove that $f \in L^p$ and $\|f\|_p \leq \|f\|^{1-p}_{p_0} \|f\|^p_{p_1}$, 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 28.7 (Containments of $L^p$ spaces). a) Show that if $\mu$ 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 $\mu$ does not admit sets of arbitrarily large finite measure, and $L^p \supset L^q$ for all $p \geq q$ if and only if $\mu$ does not admit sets of arbitrarily small positive measure.

Problem 28.8. Show that $L^p(\mathbb{R}^n) \not\subset L^q(\mathbb{R}^n)$ for any pair $p, q$.

Problem 28.9 (Convergence in $L^p$ norm). Prove that 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 28.10. 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 $L^\infty$ functions that are a.e. equal to a $C_0$ function is closed in $L^\infty$). Prove that there is a bounded linear functional $\lambda : L^\infty \to \mathbb{K}$ such that $\lambda|_{\mathcal{M}} = 0$ and $\lambda(1_\mathbb{R}) = 1$.

b) Prove that 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 $\lambda$ to $C_0(\mathbb{R})$.)

Problem 28.11. Prove Lemma 27.11, and use it to carry out the calculation omitted in the proof of Proposition 27.14.

Problem 28.12. Prove that if $f \in L^p$ then
\[
\lim_{t \to 0} t^p \lambda_f(t) = \lim_{t \to \infty} t^p \lambda_f(t) = 0
\]
(Hint: first suppose $f$ is a simple function.)

Problem 28.13. Prove the $r = \infty$ case of the Marcinkiewicz interpolation theorem.

Problem 28.14. 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 that 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 $|K| = |K|^{1/2}|K|^{\beta}$, 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 $g \ast f \in L^q(\mathbb{R}^n)$ and
\[
\|g \ast f\|_q \leq \|g\|_r \|f\|_p.
\]

Problem 28.15. Let $f$ be an unsigned measurable function on a $\sigma$-finite measure space. Prove that for $1 < p < \infty$, $f$ belongs to weak $L^p$ if and only if there exists a constant $C > 0$ such that for every set $E$ of finite measure,
\[
\frac{1}{\mu(E)} \int_E f \, d\mu \leq \frac{C}{\mu(E)^{1/p}}.
\]
(28.2)

Are these conditions still equivalent when $p = 1$?
29. The Fourier transform

We assume all functions are complex-valued unless stated otherwise.

**Definition 29.1** (The Fourier transform). Let \( f \in L^1(\mathbb{R}) \). The Fourier transform of \( f \) is the function \( \hat{f} \) defined at each \( t \in \mathbb{R} \) by

\[
\hat{f}(t) := \int_{-\infty}^{\infty} f(x) e^{-2\pi itx} \, dx \tag{29.1}
\]

Note that \( \hat{f} \) 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 \( \hat{f} \). 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 \( \chi_t : x \mapsto \exp(2\pi itx) \) (29.2)

is a character of the additive group \((\mathbb{R}, +)\). This means that \( \chi_t \) is a homomorphism from \( \mathbb{R} \) into the multiplicative group of unimodular complex numbers, explicitly for all \( s, t \in \mathbb{R} \)

\[
\chi_t(x + y) = \chi_t(x) \chi_t(y). \tag{29.3}
\]

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 29.2** (Basic properties of the Fourier transform). Let \( f, g \in L^1(\mathbb{R}) \) and let \( \alpha \in \mathbb{R} \).

a) **(Linearity)** \( c f + g = c \hat{f} + \hat{g} \)

b) **(Translation)** \( \hat{f_y}(t) = e^{-2\pi i t y} \hat{f}(t) \)

c) **(Modulation)** If \( g(x) = e^{2\pi i \alpha x} f(x) \), then \( \hat{g}(t) = \hat{f}(t - \alpha) \)

d) **(Reflection)** If \( g(x) = f(-x) \), then \( \hat{g}(t) = \overline{\hat{f}(t)} \).

e) **(Scaling)** If \( \lambda > 0 \) and \( g(x) = f(x/\lambda) \) then \( \hat{g}(t) = \lambda \hat{f}(\lambda t) \).

**Proof.** Each of these properties is verified by elementary transformations of the integral defining \( \hat{f} \); the details are left as an exercise. \( \square \)

It is immediate from the definition that \( \hat{f} \) is always a bounded function; indeed \( |\hat{f}(t)| \leq \|f\|_1 \) for all \( t \). Our next observation is:

**Proposition 29.3.** If \( f \in L^1(\mathbb{R}) \), then \( \hat{f} \) is continuous. Moreover if \( f_n \) is sequence in \( L^1 \) and \( f_n \to f \) in the \( L^1 \) norm, then \( \hat{f_n} \to \hat{f} \) uniformly.

**Proof.** Fix \( t \in \mathbb{R} \) and a sequence \( t_n \to t \). Then \( f(x) e^{2\pi i t_n x} \to f(x) e^{2\pi i t x} \) pointwise on \( \mathbb{R} \), and since trivially \( |f(x) e^{2\pi i t_n x}| \leq |f(x)| \) for all \( n \), we have by dominated convergence
\[
\hat{f}(t) = \int_{-\infty}^{\infty} f(x)e^{2\pi itx} \, dx
\]
\[
= \int_{-\infty}^{\infty} \lim_{n \to \infty} [f(x)e^{2\pi n \pi itx}] \, dx
\]
\[
= \lim_{n \to \infty} \int_{-\infty}^{\infty} f(x)e^{2\pi n \pi itx} \, dx
\]
\[
= \lim_{n \to \infty} \hat{f}(t_n).
\]

The second statement of the theorem follows immediately from the estimate \( \sup_{t \in \mathbb{R}} |\hat{f}(t)| \leq \|f\|_1 \). □

In fact, \( \hat{f} \) 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 29.4** (Translation is continuous on \( L^p \)). If \( 1 \leq p < \infty \) and \( f \in L^p(\mathbb{R}) \), then \( f_y \to f \) in \( L^p \) as \( y \to 0 \) in \( \mathbb{R} \).

**Proof.** We sketch the proof, the details are left as an exercise. The proof is accomplished by an approximation argument, in the following form: let \( X \subset L^p \) denote the set of all \( f \) for which the conclusion of the theorem is true. We show that \( X \) is a closed subspace of \( L^p \) which contains the indicator functions of intervals; but since the simple functions are dense in \( L^p \) the theorem follows since we can approximate simple functions in \( L^p \) by linear combinations of indicator functions of intervals (Littlewood’s principle).

We leave it as an exercise to check that \( X \) is a vector space and contains \( 1_I \) for all finite intervals \( I \). To see that \( X \) is closed, note that \( f, g \in L^p \) and \( \|f-g\|_p < \epsilon \), then \( \|f_y-g_y\|_p < \epsilon \) for all \( y \in \mathbb{R} \), by the translation invariance of Lebesgue measure. The proof finishes with an \( \epsilon/3 \) argument: suppose that \( g \) is in the closure of \( X \) and let \( \epsilon > 0 \) be given. Choose \( f \in X \) with \( \|f-g\|_p < \epsilon \), and choose \( \delta > 0 \) so that \( \|f_y-f\|_p < \epsilon \) for all \( |y| < \delta \). Then for all \( |y| < \delta \),

\[
\|g_y-g\|_p < \|g_y-f_y\|_p + \|f_y-f\|_p + \|f-g\|_p < 3\epsilon.
\]

Thus \( g \in X \) as well, and the proof is finished. □

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 29.5** (The Riemann-Lebesgue Lemma). Let \( f \in L^1(\mathbb{R}) \). Then \( \hat{f} \in C_0(\mathbb{R}) \).

**Proof.** The proof is accomplished using the continuity of translation in \( L^1 \), and a simple trick: first, since \( e^{-\pi i} = -1 \), we can write

\[
\hat{f}(t) = - \int_{-\infty}^{\infty} f(x)e^{-2\pi it(x+(1/2t))} \, dx = - \int_{-\infty}^{\infty} f \left( x - \frac{1}{2t} \right) e^{-2\pi ixt} \, dx \quad (29.4)
\]

Combining this with the usual definition of \( \hat{f} \), we have

\[
\hat{f}(t) = \frac{1}{2} \int_{-\infty}^{\infty} \left( f(x) - f \left( x - \frac{1}{2t} \right) \right) e^{-2\pi ixt} \, dx \quad (29.5)
\]
so

\[ |\hat{f}(t)| \leq \frac{1}{2} \| f - f_{1/2t} \|_1 \]  

(29.6)

But by Lemma 29.4, we have \( \| f - f_{1/2t} \|_1 \rightarrow 0 \) as \( t \rightarrow \pm \infty \).

Continuing our catalog of basic properties, we see that the Fourier transform also interacts nicely with differentiation:

**Proposition 29.6** (Multiplication becomes differentiation). Let \( f \in L^1(\mathbb{R}) \). If \( g(x) := xf(x) \) belongs to \( L^1 \), then \( \hat{f} \) is differentiable for all \( t \in \mathbb{R} \), and \( \hat{g}(t) = \frac{-1}{2\pi i} \frac{d}{dt} \hat{f}(t) \).

**Proof.** This is another application of dominated convergence. Let \( s \neq t \) be real numbers; from the definition of \( \hat{f} \) we have

\[
\frac{\hat{f}(s) - \hat{f}(t)}{s-t} = \int_{-\infty}^{\infty} \frac{e^{-2\pi i sx} - e^{-2\pi itx}}{s-t} f(x) \, dx
\]

and it follows that

\[
\left| \frac{\hat{f}(s) - \hat{f}(t)}{s-t} \right| \leq \int_{-\infty}^{\infty} \left| \frac{e^{-2\pi i sx} - e^{-2\pi itx}}{s-t} \right| |f(x)| \, dx
\]

(29.7)

Now the estimate

\[
\left| \frac{e^{-2\pi i sx} - e^{-2\pi itx}}{s-t} \right| \leq 2\pi |x|
\]

holds for all \( s \neq t \), so by the assumption \( xf(x) \in L^1 \) we can apply dominated convergence in (29.8) to take the limit as \( s \rightarrow t \):

\[
\lim_{s \rightarrow t} \frac{\hat{f}(s) - \hat{f}(t)}{s-t} = \lim_{s \rightarrow t} \int_{-\infty}^{\infty} \frac{e^{-2\pi i sx} - e^{-2\pi itx}}{s-t} f(x) \, dx
\]

\[
= \int_{-\infty}^{\infty} (-2\pi i) e^{-2\pi itx} xf(x) \, dx
\]

\[
= -2\pi i \hat{g}(t)
\]

Thus \( \hat{f} \) 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 29.7.** If \( f \in L^1 \) and \( g := x^n f \in L^1 \), then \( \hat{f} \) is \( n \) times differentiable, and

\[
\hat{x^k f} = \left( \frac{-1}{2\pi i} \right)^k \hat{f}^{(k)} \text{ for each } 0 \leq k \leq n.
\]

(29.10)

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 29.8. Let \( k \geq 1 \) be an integer and \( f \in L^1(\mathbb{R}) \). Suppose \( f^{(j)} \) exists a.e. for each \( j = 1, \ldots, k \), that \( f^{(j)} \in L^1 \) for \( j = 1, \ldots, k \), and that \( f, f', \ldots, f^{(k-1)} \in C_0(\mathbb{R}) \). Then
\[
\hat{f}^{(j)}(t) = (2\pi i)^j \hat{f}(t) \quad \text{for } j = 1, \ldots, k.
\]
(29.11)

Proof. We prove the case \( k = 1 \), the general case follows by induction. Since \( f' \in L^1 \) and \( f \in C_0(\mathbb{R}) \) by hypothesis, we can integrate by parts:
\[
\hat{f}'(t) = \int_{-\infty}^{\infty} f'(x)e^{-2\pi itx} \, dx
\]
(29.12)
\[
= \lim_{b \to \infty} \int_{-b}^{b} f'(x)e^{-2\pi itx} \, dx
\]
(29.13)
\[
= \lim_{b \to \infty} \left[ f(b)e^{-2\pi ib} - f(-b)e^{2\pi ib} \right] + 2\pi it \int_{-\infty}^{\infty} f(x)e^{-2\pi itx} \, dx
\]
(29.14)
\[
= 2\pi it \hat{f}(t).
\]
(29.15)

□

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
\[
(f \ast g)(x) := \int_{-\infty}^{\infty} f(x-y)g(y) \, dy
\]
(29.16)
defined at each \( x \) for which the integral makes sense. The most basic fact about convolution is:

Proposition 29.9. If \( f, g \in L^1(\mathbb{R}) \) then \( f \ast g \) is defined for almost every \( x \in \mathbb{R} \), \( f \ast g \) is measurable, and \( f \ast g \in L^1(\mathbb{R}) \).

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 \( y \). We may then Fubinate:
\[
\iint |H(x, y)| \, dx \, dy = \int_{-\infty}^{\infty} |g(y)| \left( \int_{-\infty}^{\infty} |f(x-y)| \, dx \right) \, dy
\]
\[
= \|f\|_1 \int_{-\infty}^{\infty} |g(y)| \, dy
\]
\[
= \|f\|_1 \|g\|_1
\]

where we have used the translation invariance of Lebesgue measure in the second equality. 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 \( f \ast g \) is defined almost everywhere. Now by Fubini again
\[
\int_{-\infty}^{\infty} \left| \int_{-\infty}^{\infty} f(x-y)g(y) \, dy \right| \, dx \leq \iint |H(x, y)| \, dy \, dx = \|f\|_1 \|g\|_1.
\]

□

The following basic properties of convolution are immediate, the proof is left as an exercise.

Proposition 29.10. Let \( f, g, h \in L^1(\mathbb{R}) \).
a) (Commutativity) \( f \ast g = g \ast f \).
b) (Associativity) \((f \ast g) \ast h = f \ast (g \ast h)\).
c) (Distributivity) \((f + g) \ast h = f \ast g + f \ast h\).
d) (Scalar multiplication) If \( c \in \mathbb{C} \), then \((cf) \ast g = c(f \ast g)\).

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 29.11** (Convolution becomes multiplication). Let \( f, g \in L^1(\mathbb{R}) \). Then \( \hat{f \ast g}(t) = \hat{f}(t) \hat{g}(t) \).

**Proof.** By virtue of Proposition 29.9, we can use Fubini’s theorem to compute \( \hat{f \ast g}(t) \):

\[
\hat{f \ast g}(t) = \int_{-\infty}^{\infty} \left( \int_{-\infty}^{\infty} f(x - y)g(y) \, dy \right) e^{-2\pi ixt} \, dx
\]

\[
= \int_{-\infty}^{\infty} g(y) \left( \int_{-\infty}^{\infty} f(x - y) e^{-2\pi ixt} \, dx \right) \, dy
\]

\[
= \int_{-\infty}^{\infty} \hat{f}(t) e^{-2\pi iyt} g(y) \, dy
\]

\[
= \hat{f}(t) \hat{g}(t)
\]

where we have used Proposition 29.2(b). \( \square \)

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

\[
g_a(x) := e^{-\pi ax^2}.
\]

(29.17)

(The factor of \( \pi \) will be convenient given our choice of normalization in the definition of the Fourier transform.) It should be clear that \( x^n g_a(x) \in L^1 \) for all \( a > 0 \) and \( n \geq 0 \).

**Lemma 29.12.** \( \hat{g}_a(t) = \frac{1}{\sqrt{a}} e^{-\pi t^2/a} \).

**Proof.** Rather than computing the integral directly, we exploit Propositions 29.6 and 29.8. We may also assume \( a = 1 \); the general case follows from this by scaling (Proposition 29.2(e)). Write \( g = g_1 \). Since \( xg \in L^1 \), we have

\[
(g)'(t) = (-2\pi ixe^{-\pi x^2})\hat{g}(t)
\]

\[
= i((e^{-\pi x^2})'\hat{g})(t)
\]

\[
= i(2\pi it)\hat{g}(t)
\]

\[
= -2\pi t\hat{g}(t).
\]
It follows from this computation and the product rule that
\[
\frac{d}{dt}(e^{\pi t^2}\hat{g}(t)) = 0,
\] (29.18)
so the function \(e^{\pi t^2}\hat{g}(t)\) is constant. To evaluate the constant, we set \(t = 0\) and use the well-known Gaussian integral
\[
\hat{g}(0) = \int_{-\infty}^{\infty} e^{-\pi x^2} \, dx = 1.
\] (29.19)
\[\square\]

30. 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 30.1.** Suppose \(f \in L^\infty\), \(g \in L^1\), and \(f\) is continuous. Then \(f \ast g\) is bounded and continuous.

**Proof.** Since \(f\) is bounded,
\[
|(f \ast g)(x)| \leq \int_{-\infty}^{\infty} |f(x - y)||g(y)| \, dy \leq \|f\|_\infty \|g\|_1.
\] (30.1)
To see that \(f \ast g\) is continuous, fix \(x \in \mathbb{R}\) and let \(x_n \to x\). The functions \(h_n(y) := f(x_n - y)g(y)\) are dominated by \(\|f\|_\infty |g(y)| \in L^1\), and converge pointwise to \(f(x - y)g(y)\), so by the dominated convergence theorem \(\lim_{n \to \infty}(f \ast g)(x_n) = (f \ast g)(x)\). \(\square\)

In fact, the continuity assumption on \(f\) isn’t needed; we can conclude that \(f \ast g\) is continuous assuming only \(f\) is bounded, by using the continuity of translation in \(L^1\). See Problem 35.4.

**Proposition 30.2.** If \(f, g\) are \(L^1\) functions and are both compactly supported, then so if \(f \ast g\).

**Proof.** Problem 35.3. \(\square\)

The dominated convergence argument can be extended to apply to differentiability:

**Proposition 30.3.** Suppose \(f\) is a compactly supported \(C^k\) function \((1 \leq k \leq \infty)\) and \(g \in L^1\). Then \(f \ast g \in C^k\).

**Proof.** First assume \(k = 1\). Then one can differentiate under the integral sign as in the proof of Proposition 29.6. The general case is proved by induction; the details are left as an exercise. \(\square\)

**Definition 30.4.** An \(L^1\) approximate unit is a collection of functions \(\phi_\lambda \in L^1(\mathbb{R})\) indexed by \(\lambda > 0\) such that:

a) \(\phi_\lambda(t) \geq 0\) almost everywhere, for each \(\lambda\),
b) \( \int_{-\infty}^{\infty} \phi_\lambda(t) \, dt = 1 \) for all \( \lambda \), and
c) For each fixed \( \delta > 0 \), we have \( \| 1_{|t|>\delta} \phi_\lambda \|_1 \to 0 \) as \( \lambda \to 0 \).

Approximate units are easy to construct: Let \( \phi \) 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 (a)). The simplest example comes from taking \( \phi(x) = 1_{[-1/2,1/2]} \); the resulting \( \phi_\lambda \) is known as the box kernel. (Draw a few of these for different values of \( \lambda \) 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 30.5.** Let \( \phi_\lambda \) be an \( L^1 \) approximate unit and let \( 1 \leq p < \infty \). Then for all \( f \in L^p(\mathbb{R}) \), we have \( \| \phi_\lambda * f - f \|_p \to 0 \) as \( \lambda \to 0 \).

To prove the theorem we need two facts. The first is Minkowski’s integral inequality (Theorem 27.17). 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 30.6.** Let \( \phi_\lambda \) be an \( L^1 \) approximate unit and \( g \in L^\infty \). If \( g \) is continuous at a point \( x \in \mathbb{R} \), then

\[
\lim_{\lambda \to 0} (g * \phi_\lambda)(x) = g(x). \tag{30.2}
\]

**Proof.** Using the trick \( g(x) = \int_{-\infty}^{\infty} \phi_\lambda(y)g(x) \, dy \), we have

\[
(g * \phi_\lambda)(x) - g(x) = \int_{-\infty}^{\infty} (g(x-y) - g(x)) \phi_\lambda(y) \, dy, \tag{30.3}
\]

so using the positivity of \( \phi_\lambda \)

\[
|(g * \phi_\lambda)(x) - g(x)| \leq \int_{-\infty}^{\infty} |g(x-y) - g(x)| \phi_\lambda(y) \, dy \tag{30.4}
\]

To estimate the right-hand side, let \( \epsilon > 0 \). 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 (30.4) into two integrals, over the regions \( |y| \leq \delta \) and \( |y| > \delta \):

\[
\int_{-\infty}^{\infty} |g(x-y) - g(x)| \phi_\lambda(y) \, dy = \int_{\{|y| \leq \delta\}} |g(x-y) - g(x)| \phi_\lambda(y) \, dy + \int_{\{|y| > \delta\}} |g(x-y) - g(x)| \phi_\lambda(y) \, dy \tag{30.5}
\]

The first integrand is bounded by \( \epsilon \phi_\lambda \), so

\[
\int_{\{|y| \leq \delta\}} |g(x-y) - g(x)| \phi_\lambda(y) \, dy \leq \epsilon \int_{\{|y| \leq \delta\}} \phi_\lambda(y) \, dy \leq \epsilon \tag{30.6}
\]

since \( \int_{-\infty}^{\infty} \phi_\lambda(y) \, dy = 1 \). The second integrand is bounded by \( 2\|g\|_\infty 1_{\{|y| > \delta\}} \phi_\lambda(y) \), so goes to 0 as \( \lambda \to 0 \) by property (c) in the definition of approximate unit. This proves the lemma. \( \square \)

**Remark:** If one assumes that the approximate unit \( \phi_\lambda \) was constructed as \( \phi_\lambda(y) = \frac{1}{\lambda} \phi \left( \frac{y}{\lambda} \right) \) for some \( \phi \) satisfying \( \phi \geq 0 \) and \( \int \phi = 1 \), then the lemma has an easier proof; see Problem 35.5(b).
Proof of Theorem 30.5. Let \( f \in L^p \). Write
\[
(f * \phi_\lambda)(x) - f(x) = \int_{-\infty}^{\infty} (f(x-y) - f(x)) \phi_\lambda(y) dy.
\]
(30.7)
Then by Theorem 27.17(b) (with \( \mu \) being Lebesgue measure and \( \nu \) the measure \( \phi_\lambda(y) dy \)), we have
\[
\|f * \phi_\lambda - f\|_p \leq \int_{-\infty}^{\infty} \|f_y - f\|_p \phi_\lambda(y) dy
\]
(30.8)
Consider the function \( g(y) := \|f_y - f\|_p \). This function belongs to \( L^\infty \) (indeed \( \|g\|_\infty \leq 2\|f\|_1 \)) and continuous (by the continuity of translation in \( L^p \)) and \( g(0) = 0 \). Thus by Lemma 30.6, 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 \( C^\infty \) functions with compact support. This is accomplished via bump functions:

Definition 30.7. A bump function is a function \( \psi : \mathbb{R} \to \mathbb{R} \) such that:
\begin{itemize}
  \item[a)] \( \psi \in C^\infty(\mathbb{R}) \),
  \item[b)] \( \psi \) is compactly supported,
  \item[c)] \( \psi \geq 0 \), and
  \item[d)] \( \int_{-\infty}^{\infty} \psi(x) dx = 1 \).
\end{itemize}

Lemma 30.8. Bump functions exist.

Proof. The main issue is to construct a \( C^\infty \) function with compact support. Consider the function
\[
h(x) = \begin{cases} 
  e^{-1/x} & \text{if } x > 0 \\
  0 & \text{if } x \leq 0 
\end{cases}
\]
(30.9)
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 \( \psi \) 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 \( C^\infty_c(\mathbb{R}) \) is dense in \( L^p(\mathbb{R}) \) for all \( 1 \leq p < \infty \). A proof is outlined in Problem 35.6.

31. Inversion and uniqueness

31.1. Fourier inversion—the problem. In this section we study the problem of recovering \( f \) from \( \hat{f} \). 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 \( \hat{f}(t) \) measures the “amplitude” of \( f \) in the “frequency” \( t \). This suggests that a formula like
\[
f(x) = \int_{-\infty}^{\infty} \hat{f}(t)e^{2\pi itx} dt
\]
(31.1)
ought to hold, at least if \( \hat{f} \in L^1 \). If we formally substitute the definition of \( \hat{f} \) 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 \( \hat{f} \). In fact (31.1) does hold when \( \hat{f} \in L^1 \), but a more delicate argument is necessary. So, the goal of this section will be to prove:

**Theorem 31.1** (Fourier inversion, \( L^1 \) case). Suppose \( f \) and \( \hat{f} \) belong to \( L^1 \). Then

\[
f(x) = \int_{-\infty}^{\infty} \hat{f}(t) e^{2\pi ixt} \, dt
\]

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 31.2.** Suppose \( f, g \in L^1 \). If \( \hat{f} = \hat{g} \), then \( f = g \) a.e.

**Proof.** From the inversion theorem, if \( f \in L^1 \) and \( \hat{f} = 0 \), then \( f = 0 \). By the linearity of the Fourier transform, \( \hat{f-g} = \hat{f} - \hat{g} \), and the corollary follows. \( \square \)

So, in principle, \( f \) is fully determined by \( \hat{f} \), even if \( \hat{f} \notin L^1 \); and this is often the case. For example, suppose \( f = 1_{[0,1]} \). Then

\[
\hat{1_{[0,1]}}(t) = \int_{0}^{1} e^{-2\pi ixt} \, dx = \frac{1 - e^{-2\pi it}}{2\pi it}
\]

which does not belong to \( L^1 \) (check this). To recover \( f \) from \( \hat{f} \) in these cases, we turn to the summability methods of Section ???. 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** \( \psi_a(t) := 1_{[-a,a]} \), which is positive, integrable, and increases to 1 pointwise as \( a \to \infty \). Given any family of functions \( \psi_a \) with these three properties, we can consider the integrals

\[
\int_{-\infty}^{\infty} h(t) \psi_a(t) \, dt
\]

It will turn out that the “square” cutoff \( 1_{[-a,a]} \) 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 \( e^{-a|t|} \) (the extra factor of \( \pi \) turns out to be a convenient normalization).

**Lemma 31.3.** For all \( a > 0 \),
\[
\int_{-\infty}^{\infty} e^{-a|t|} e^{2\pi itx} \, dt = \frac{1}{\pi} \frac{a}{a^2 + x^2} \tag{31.8}
\]

**Proof.** Problem 35.7. \( \square \)

Let us fix the notation
\[
P_a(x) := \frac{1}{\pi} \frac{a}{a^2 + x^2} \tag{31.9}
\]
Notice that \( P_1(x) \) is nonnegative and \( \int_{-\infty}^{\infty} P_1(x) \, dx = 1 \). Moreover, \( P_a(x) = \frac{1}{a} P_1\left(\frac{x}{a}\right) \).

**Lemma 31.4.** \( \{P_a\}_{a>0} \) is an \( L^1 \) approximate unit.

The function \( P_a(x) \) (viewed as a function of the two arguments \( a, x \)) is known as the Poisson kernel. We are now able to compute the integral (31.1) modified by the cutoff function \( e^{-a|t|} \):

**Proposition 31.5.** If \( f \in L^1 \), then for all \( a > 0 \) and all \( x \in \mathbb{R} \)
\[
(f * P_a)(x) = \int_{-\infty}^{\infty} e^{-a|t|} \hat{f}(t) e^{2\pi itx} \, dt. \tag{31.10}
\]

**Proof.** Because we have introduced the cutoff function, Fubini’s theorem can be applied:
\[
\int_{-\infty}^{\infty} e^{-a|t|} \hat{f}(t) e^{2\pi itx} \, dt = \int_{-\infty}^{\infty} e^{-a|t|} \int_{-\infty}^{\infty} f(y) e^{2\pi i(x-y)t} \, dy \, dt \tag{31.11}
\]
\[
= \int_{-\infty}^{\infty} e^{-a|t|} \int_{-\infty}^{\infty} f(x-y) e^{2\pi iyt} \, dy \, dt \tag{31.12}
\]
\[
= \int_{-\infty}^{\infty} f(x-y) \int_{-\infty}^{\infty} e^{-a|t|} e^{2\pi iyt} \, dt \, dy \tag{31.13}
\]
\[
= (f * P_a)(x) \tag{31.14}
\]

**Proof of Theorem 31.1.** Assume \( f, \hat{f} \in L^1(\mathbb{R}) \). Define
\[
g(x) = \int_{-\infty}^{\infty} \hat{f}(t) e^{2\pi itx} \, dt \tag{31.15}
\]
By the Riemann-Lebesgue lemma, \( g \in C_0(\mathbb{R}) \). We want to show \( g = f \) a.e. From Proposition 31.5 we have for all \( a > 0 \)
\[
(f * P_a)(x) = \int_{-\infty}^{\infty} e^{-a|t|} \hat{f}(t) e^{2\pi itx} \, dt \tag{31.16}
\]
Fix a sequence \( a_n \to 0 \). Since \( \hat{f} \in L^1 \), we can apply dominated convergence to show that the integral in the right-hand side of (31.16) converges to \( g(x) \) as \( a_n \to 0 \), for all \( x \). On the other hand, since \( P_a \) is an \( L^1 \) approximate unit, we know from Theorem 30.5 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 (31.16) we have $f * P_{a_n} \to g$ a.e., so $f = g$ a.e. and the theorem is proved.

**Remark:** Observe that the above proof did not really use the explicit form of $P_a$; rather the point was that $\{e^{-a\pi|t|}\}_{a>0}$ was a cutoff function 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 31.5 is that we can recover $f$ from $\hat{f}$ in a weaker sense for any $f \in L^1$ (that is, not assuming $\hat{f} \in L^1$). Indeed, combining Proposition 31.5 and Theorem 30.5 we have immediately:

**Corollary 31.6** (Fourier inversion in the $L^1$ norm). If $f \in L^1$, then

$$\int_{-\infty}^{\infty} e^{-a\pi|t|} \hat{f}(t) e^{2\pi i xt} dt$$

(31.17)

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 regularized integrals. In fact, it is true that the integrals (31.17) converge to $f$ a.e., but this requires a more delicate argument which we take up in Section ??.

### 32. 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 28.8 $L^2 \not\subset L^1$, so the integral (29.1) need not be defined. However, we can observe that $L^1 \cap L^2$ is dense in $L^2$ (why?), and start there.

**Lemma 32.1.** If $f \in L^1 \cap L^2$, then $\hat{f}$ belongs to $L^2$ and $\|\hat{f}\|_2 = \|f\|_2$.

**Proof.** Let $\tilde{f}(x) := \overline{f(-x)}$ and define $g = f * \tilde{f}$. Since $f, \tilde{f} \in L^1$ we have $g \in L^1$ by Proposition 29.9. Now

$$g(x) = \int_{-\infty}^{\infty} f(x-y) \overline{f(-y)} dy = \int_{-\infty}^{\infty} f(x+y) f(y) dy$$

(32.1)

so we can write $g(x) = \langle f_{-x}, f \rangle_{L^2}$. By Lemma 29.4, the map $x \to f_{-x}$ is continuous from $\mathbb{R}$ into $L^2$, so by Cauchy-Schwarz we see that $g$ is a continuous function of $x$, and $g(0) = \|f\|^2_2$. By Cauchy-Schwarz again,

$$|g(x)| \leq \|f_{-x}\|_2 \|f\|_2 = \|f\|^2_2,$$

(32.2)

so $g$ is bounded.

Now, since $g \in L^1$ we can apply Proposition 31.5 to compute

$$(g * P_a)(0) = \int_{-\infty}^{\infty} e^{-a\pi|t|} \hat{g}(t) dt$$

(32.3)

As $g$ is continuous, we have by Lemma 30.6

$$\|f\|^2_2 = g(0) = \lim_{a \to 0} (g * P_a)(0) = \lim_{a \to 0} \int_{-\infty}^{\infty} e^{-a\pi|t|} \hat{g}(t) dt.$$  

(32.4)
Let us compute the limit of this last integral in a different way: recall that by definition $g = f \ast \hat{f}$, so by Propositions 29.11 and 29.2(d),

$$\hat{g}(t) = |\hat{f}(t)|^2.$$  

(32.5)

Making this substitution in the integral in (32.4) and applying the monotone convergence theorem, we have

$$\|f\|_2^2 = \lim_{a \to 0} \int_{-\infty}^{\infty} e^{-a|t|} \hat{g}(t) \, dt = \lim_{a \to 0} \int_{-\infty}^{\infty} e^{-a|t|} |\hat{f}(t)|^2 \, dt = \|\hat{f}\|_2^2.$$  

(32.6)

which proves the lemma. □

**Theorem 32.2 (The Fourier transform on $L^2$).** There is a unique bounded linear transformation $\mathcal{F} : L^2 \to L^2$ satisfying the following conditions:

1. For all $f \in L^1 \cap L^2$, $\mathcal{F}f = \hat{f}$.
2. (The Plancherel theorem) $\|\mathcal{F}f\|_2 = \|f\|_2$ for all $f \in L^2$.
3. The mapping $f \to \mathcal{F}f$ is an Hilbert space isomorphism of $L^2$ onto $L^2$.
4. (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 $f \in L^1 \cap L^2$, the Parseval identity reads

$$\int_{-\infty}^{\infty} f(x)\overline{g(x)} \, dx = \int_{-\infty}^{\infty} \hat{f}(t)\overline{\hat{g}(t)} \, dt.$$  

(32.7)

**Proof of Theorem 32.2.** By Lemma 32.1, the map $f \to \hat{f}$ is bounded linear transformation from a dense subspace of $L^2$ into $L^2$. Thus, since the codomain $L^2$ is complete, by Proposition 19.9 the map $f \to \hat{f}$ 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 26.3(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, this shows that $\mathcal{F}$ is in fact onto. (The proof of this is left as an exercise). Let $M$ denote the set of all functions $g \in L^2$ such that $g = \hat{f}$ 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 $e^{-a|x|}$, $a > 0$. The functions $e^{2\pi ibx}e^{-a\pi|x|}$ belong to $L^1 \cap L^2$ for all $a > 0$ and $b \in \mathbb{R}$, so their Fourier transforms

$$P_a(t - b) = \int_{-\infty}^{\infty} e^{2\pi ibx}e^{-a\pi|x|}e^{-2\pi itx} \, dx$$  

(32.8)

belong to $M$. So, let $h \in M$. Then

$$(P_a * \overline{h})(b) = \int_{-\infty}^{\infty} P_a(b - t)\overline{h(t)} \, dt = 0$$  

(32.9)

for all $b$, and therefore $h = 0$ by Theorem 30.5. Thus $M$ is dense in $L^2$ and the proof is finished. □

**Theorem 32.3 ($L^2$ inversion).** Let $f \in L^2$. Define

$$\phi_N(t) = \int_{-N}^{N} f(x)e^{-2\pi ixt} \, dx, \quad \psi_N(t) = \int_{-N}^{N} (\mathcal{F}f)(t)e^{2\pi ixt} \, dt.$$  

(32.10)
Then \( \| \phi_N - \mathcal{F}f \|_2 \to 0 \) and \( \| \psi_N(t) - f \|_2 \to 0 \) as \( N \to \infty \).

**Proof.** Let \( f_N := 1_{[-N,N]} f \). Then \( f_N \in L^1 \cap L^2 \), and \( \phi_N = \hat{f}_N \). 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 is an isometry, we have also that \( \phi_N \) is a Cauchy sequence in \( L^2 \), and hence converges to some \( \phi \in L^2 \). Now if \( f \in L^1 \cap L^2 \), we have that \( \phi_N \) converges pointwise to \( \hat{f} \), but since also a subsequence of \( \phi_N \) converges pointwise to \( \phi \), we have \( \phi = \hat{f} \). This means that the linear map taking \( f \) to \( \lim \phi_N \) is a bounded linear transformation on \( L^2 \) that agrees with \( f \to \hat{f} \) on \( L^1 \cap L^2 \), so it follows that \( \lim \phi_N = \mathcal{F}f \).

The statement for \( \psi_N \) is proved by similar methods and is left as an exercise (Problem 35.10).

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 \( \hat{f} \) for \( \mathcal{F}f \) when \( f \in L^2 \), with the understanding that the integral definition is only valid when \( f \in L^1 \cap L^2 \).

### 33. Fourier Series

We now replace the group \((\mathbb{R}, +)\) with the multiplicative group of unimodular complex numbers

\[ T = \{ e^{i\theta} : -\pi \leq \theta < \pi \} \subset \mathbb{C}. \tag{33.1} \]

By the properties of the exponential, this group isomorphic to the group \([-\pi, \pi)\) with addition mod \( 2\pi \). (This amounts to the isomorphism of the quotient group \( \mathbb{R}/2\pi \mathbb{Z} \cong \mathbb{T} \).) We will typically use \( \theta \) to identify elements of \( T \). We will treat functions \( f : T \to \mathbb{C} \) either as functions on the unit circle in \( \mathbb{C} \), or, when convenient, as functions on the interval \([-\pi, \pi)\) extended \( 2\pi \)-periodically to \( \mathbb{R} \). For each integer \( n \), the map

\[ \chi_n : \theta \to e^{in\theta} \tag{33.2} \]

is a homomorphism from \( T \) to \( \mathbb{T} \). We equip \( T \) with normalized arc length measure, or equivalently normalized Lebesgue measure on \([-\pi, \pi)\). We write \( dm \) for this measure. Note that \( T \) acts on itself by translation: for fixed \( \theta \), we have a map \( \tau_\theta : T \to T \) given by

\[ \tau_\theta(e^{i\psi}) = e^{i(\psi + \theta)}. \tag{33.3} \]

This amounts to rotation of the circle through the angle \( \theta \), and \( m \) is invariant under \( \tau_\theta \) for each \( \theta \). Arguing exactly as on the line, one can prove the continuity of translation in \( L^p(T) \): for fixed \( \varphi \in [-\pi, \pi) \), let \( f_\varphi(\theta) = f(\theta - \varphi) \).

**Proposition 33.1.** Let \( 1 \leq p < \infty \). If \( f \in L^p(T) \), then \( f_\varphi \to f \) in \( L^p(T) \) as \( \varphi \to 0 \).

Now that we have a translation invariant measure and the characters \( \chi_n \), we can define a Fourier transform.

**Definition 33.2.** Let \( f \in L^1(T) \). The **Fourier coefficients** of \( f \) are the numbers

\[ \hat{f}(n) := \int_\pi^{-\pi} f(\theta) e^{-inx} \, dm(\theta) = \frac{1}{2\pi} \int_\pi^{-\pi} f(x) e^{-inx} \, dx, \quad n \in \mathbb{Z}. \tag{33.4} \]
The Fourier series of $f$ is the series

$$f \sim \sum_{n=-\infty}^{\infty} \hat{f}(n)e^{in\theta}. \quad (33.5)$$

The Fourier coefficients $\hat{f}(n)$ are the analogs of the pointwise values of $\hat{f}(t)$ in the case of $\mathbb{R}$, and the (possibly divergent) Fourier series is the analog of the (possibly divergent) integral

$$\int_{-\infty}^{\infty} \hat{f}(t)e^{2\pi itx} dt. \quad (33.6)$$

The function $\hat{f} : \mathbb{Z} \to \mathbb{C}$ is called the Fourier transform of $f$. 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 $\hat{f}$ is always a bounded function (that is, $\hat{f} \in \ell^\infty(\mathbb{Z})$, and in fact the Riemann-Lebesgue lemma holds:

**Theorem 33.3** (Riemann-Lebesgue lemma on the circle). If $f \in L^1(\mathbb{T})$, then

$$\lim_{n \to \pm\infty} |\hat{f}(n)| = 0. \quad (33.7)$$

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 $\hat{f} \notin \ell^1(\mathbb{Z})$. Thus, there is an immediate difficulty in interpreting the series (33.5)

As before, the basic problem is to recover $f$ from $\hat{f}$, 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

$$s_N(\theta) = \sum_{n=-N}^{N} \hat{f}(n)e^{in\theta}. \quad (33.8)$$

Since this is a finite sum, we can expand $\hat{f}$ as an integral and pull the sum inside:

$$s_N(\theta) = \int_{\mathbb{T}} f(\phi) \left\{ \sum_{n=-N}^{N} e^{-in\phi}e^{in\theta} \right\} dm(\phi) \quad (33.9)$$

Working with the inner sum, we consider the expression

$$D_N(t) := \sum_{n=-N}^{N} e^{int} = \sin \left( N + \frac{1}{2} \right) \frac{t}{\sin \frac{t}{2}}. \quad (33.10)$$

Then (33.9) can be written

$$s_N(\theta) = \int_{\mathbb{T}} f(\phi)D_N(\theta - \phi) dm = (f * D_N)(\theta) \quad (33.11)$$
where we have introduced convolution on the circle group \( T \). (Note that the difference \( \theta - \phi \) is interpreted in the group \( T \), that is, is carried out mod \( 2\pi \).) 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 \( L^1(T) \) approximate unit, the partial sums \( s_N \) can be badly behaved, even for nice \( f \). For example we have the following:

**Theorem 33.4.** There exists a continuous function \( f \) on \( T \) 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 \( \theta \) is given by \( (f * D_N)(\theta) \). We suppose that \( (f * D_N)(0) \to f(0) \) for every \( f \in C(T) \) and derive a contradiction. Now

\[
s_N(0) = (f * D_N)(0) = \int_T f(\phi)D_N(\phi) \, dm(\phi). \tag{33.12}
\]

By the construction of \( D_N \), it is clear that \( D_N \in L^1(T) \) for each \( N \). Thus for each \( N \) the map \( L_N : f \to \int_T fD_N \, dm \) is a bounded linear functional on \( C(T) \), 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 \( g_n \to \text{sgn}D_N \) 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 \( f \in C(T) \), then the family of linear functionals \( L_N \) would be pointwise bounded on \( C(T) \), hence uniformly bounded, which is a contradiction. Problem 35.17 gives some hints on filling the details. \( \square \)

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 33.5.** Suppose \( f \in L^1(T) \) and \( f \) is differentiable at a point \( \theta_0 \). 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

\[
s_N(0) = (f * D_n)(0) = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(\phi)D_N(\phi) \, d\phi \tag{33.13}
\]

\[
= \frac{1}{2\pi} \int_{-\pi}^{\pi} f(\phi) \frac{\sin (N + \frac{1}{2}) \phi}{\sin \frac{\phi}{2}} \, d\phi, \tag{33.14}
\]

and we wish to prove \( s_N(0) \to 0 \) as \( N \to \infty \). The key observation is that the function

\[
g(\phi) = \frac{f(\phi)}{\sin \frac{\phi}{2}} \tag{33.15}
\]

belongs to \( L^1(T) \). To see this, first note the elementary estimate

\[
\frac{|\phi|}{\pi} \leq \left| \sin \frac{\phi}{2} \right| \leq \frac{|\phi|}{2} \tag{33.16}
\]
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
\[
|g(\phi)| = \left| \frac{f(\phi)}{\phi} \right| \frac{\phi}{\sin \frac{\phi}{2}} \leq \pi M.
\]
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 (33.13), we have
\[
s_N(0) = \frac{1}{2\pi} \int_{-\pi}^{\pi} g(\phi) \sin \left( N + \frac{1}{2} \right) \frac{\phi}{\sin \frac{\phi}{2}} d\phi
\]
Making the change of variable $\theta = \phi/2$ this becomes
\[
s_N(0) = \frac{1}{\pi} \int_{-\pi/2}^{\pi/2} g(2\theta) \sin((2N + 1)\theta) d\theta
\]
If we now put
\[
h(\theta) = 21_{[-\pi/2,\pi/2]}(\theta)g(2\theta),
\]
then $h \in L^1$ and the integral in (33.21) is nothing but the imaginary part of $\hat{h}(2N + 1)$, 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 $1_{[-N,N]}$ (corresponding to ordinary partial sums) is badly behaved, we choose a smoother cutoff. It turns out that the functions
\[
n \to r^{\lfloor n \rfloor}
\]
for $0 \leq r < 1$ are a good choice (analogous to $e^{-a|t|}$ on the line). Thus we consider the Abel means of the Fourier series
\[
A(r, \theta) := \sum_{n=-\infty}^{\infty} \hat{f}(n)r^{\lfloor n \rfloor}e^{in\theta}.
\]
Since $\hat{f}$ is bounded and $r < 1$, this series is absolutely convergent for all $\theta$, and uniformly convergent on $\mathbb{T}$ for each fixed $r$. Thus, we can again expand $\hat{f}$ as an integral, and interchange the sum and integral:
\[
A(r, \theta) = \int_{\mathbb{T}} f(\phi) \left\{ \sum_{n=-\infty}^{\infty} r^{\lfloor n \rfloor}e^{in(\theta - \phi)} \right\} dm(\phi) := (f * P_r)(\theta)
\]
where, by summing the geometric series, $P_r(\theta)$ is given by
\[
P_r(\theta) := \sum_{n=-\infty}^{\infty} r^{\lfloor n \rfloor}e^{in\theta} = \frac{1 - r^2}{1 - 2r \cos \theta + r^2}
\]
We now observe that the family of functions $P_r$ has the following properties:

**Lemma 33.6.**

i) $P_r(\theta) \geq 0$ for all $\theta \in [-\pi, \pi]$ and all $0 \leq r < 1$,

ii) For each $r$, \( \int_{-\pi}^{\pi} P_r(\theta) \, d\theta = 1 \),

iii) For each fixed $0 < \delta < \pi$,
\[
\frac{1}{2\pi} \int_{\delta \leq |\theta| \leq \pi} P_r(\theta) \, d\theta \to 0 \quad \text{as} \, r \to 1. \tag{33.27}
\]

**Proof.** Exercise. \qed

In other words, $\{P_r\}_{r<1}$ is an $L^1(\mathbb{T})$ approximate unit. Just as on the line, we have

**Theorem 33.7.** Let $1 \leq p < \infty$. If $f \in L^p(\mathbb{T})$, 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. \qed

With these results in hand, we can obtain Abel summability of Fourier series on the circle:

**Corollary 33.8.** 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 33.6:

**Lemma 33.9.** 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 $\delta$. For $\delta < |\theta| \leq \pi$, we have $-1 \leq \cos\theta \leq \cos\delta < 1$. Thus for such $\theta$
\[
P_r(\theta) = \frac{1 - r^2}{1 - 2r \cos \theta + r^2} \leq \frac{1 - r^2}{1 - 2r \cos \delta + r^2}. \tag{33.28}
\]

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. \qed

**Theorem 33.10.** 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
\[
|(f * P_r)(\theta) - f(\theta)| \leq \frac{1}{2\pi} \int_{-\pi}^{\pi} |f(\theta - \phi) - f(\theta)|P_r(\phi) \, d\phi. \tag{33.29}
\]

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
\[
\frac{1}{2\pi} \int_{|\phi|<\delta} |f(\theta - \phi) - f(\theta)|P_r(\phi) \, d\phi \leq \frac{1}{2\pi} \int_{-\pi}^{\pi} P_r(\theta) \, d\phi = \epsilon. \tag{33.30}
\]

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
\[
\frac{1}{2\pi} \int_{\delta<|\phi|\leq\pi} |f(\theta - \phi) - f(\theta)|P_r(\phi) \, d\phi \leq 2\epsilon\|f\|_{\infty}. \tag{33.31}
\]
Thus for \( r \) sufficiently close to 1, we get \(|f * P_r(\theta) - f(\theta)| \leq (1 + 2\|f\|_\infty)\epsilon\), so \( f * P_r \to f \) uniformly on \( \mathbb{T} \).

From the smoothing properties of convolution, we see that if \( f \in L^1 \) then \( f * P_r \) is continuous in \( \theta \) for each \( r \). Thus there is no hope that \( f * P_r \to f \) in the \( L^\infty \) norm when \( f \in L^\infty \). However we do have the following weaker form of convergence:

**Proposition 33.11.** Let \( f \in L^\infty(\mathbb{T}) \). Then for each \( g \in L^1(\mathbb{T}) \),

\[
\lim_{r \to 1} \frac{1}{2\pi} \int_{-\pi}^{\pi} (f * P_r)(\theta)g(\theta) \, d\theta = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(\theta)g(\theta) \, d\theta.
\]  

**Proof.** Problem 35.18. \( \square \)

As for the line, the Fourier transform is especially 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 is left as an exercise.

**Theorem 33.12.** The Fourier transform is a unitary transformation from \( L^2(\mathbb{T}) \) onto \( \ell^2(\mathbb{Z}) \). In particular,

\begin{enumerate}
  \item (Plancherel theorem) For all \( f \in L^2(\mathbb{T}) \), we have \( \|f\|_2^2 = \sum_{n=-\infty}^{\infty} |\hat{f}(n)|^2 \).
  \item (Parseval identity) For all \( f, g \in L^2(\mathbb{T}) \),

\[
\int_{\mathbb{T}} f(\theta)\overline{g(\theta)} \, dm(\theta) = \sum_{n=-\infty}^{\infty} \hat{f}(n)\overline{\hat{g}(n)}.
\]  
\end{enumerate}

In particular, note that the Fourier transform takes the orthonormal basis \( E = \{e^{i\pi n \theta}\}_{n \in \mathbb{Z}} \) 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 \( e_n(\theta) = e^{i\pi n \theta} \), then

\[
\hat{f}(n) := \int_{\mathbb{T}} f(\theta)e^{-i\pi n \theta} \, dm(\theta) = \langle f, e_n \rangle_{L^2(\mathbb{T})}.
\]  

Thus, if one has already proved that the functions \( \{e_n\} \) are an orthonormal basis for \( L^2(\mathbb{T}) \), the proof of Theorem 33.12 becomes quite simple. (Indeed, the theorem is essentially equivalent to the assertion that the characters \( \{e_n\} \) form an orthonormal basis. Why?)

### 34. Schwartz functions and distributions

In this section we introduce the Schwartz space \( \mathcal{S} \) and the space of tempered distributions \( \mathcal{S}' \). 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 \( \psi \) is a very smooth function (say \( C^\infty \)) and vanishes at infinity, then if \( f \) is differentiable we have the integration by parts formula

\[
\int_{\mathbb{R}} f'(x)\psi(x) \, dx = -\int_{\mathbb{R}} f(x)\psi'(x) \, dx.
\]  

(34.1)
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
\[
\psi \mapsto \int_{\mathbb{R}} f \psi, \tag{34.2}
\]
then the above calculation suggests that we can interpret “\( f' \)” as the linear functional
\[
\psi \mapsto -\int f' \psi \tag{34.3}
\]
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 \( C^\infty \) functions on \( \mathbb{R} \).

**Definition 34.1.** The Schwartz space \( \mathcal{S} \) consists of all functions \( \psi \in C_b^\infty(\mathbb{R}) \) such that \( x^\alpha \psi^{(\beta)}(x) \) is bounded for all integers \( \alpha, \beta \geq 0 \).

We say that a function \( f \) is rapidly decreasing if \( x^\alpha \psi(x) \) is bounded for all \( \alpha \geq 0 \). So \( \mathcal{S} \) consists of those \( \psi \) such that \( \psi \) and all of its derivatives are rapidly decreasing. For example, \( \psi(x) = e^{-x^2} \) belongs to \( \mathcal{S} \). It is an important fact that \( \mathcal{S} \) is closed under differentiation, and under multiplication by polynomials:

**Lemma 34.2.** \( \mathcal{S} \) is a vector space, and if \( \psi \in \mathcal{S} \) then \( x \psi(x) \) and \( \psi'(x) \) belong to \( \mathcal{S} \), and in fact \( x^\alpha \psi^{(\beta)} \in \mathcal{S} \) for all \( \alpha, \beta \geq 0 \).

*Proof.* Exercise. \( \square \)

**Definition 34.3.** For integers \( \alpha, \beta \geq 0 \), define for \( \psi \in \mathcal{S} \)
\[
\| \psi \|_{\alpha, \beta} := \| x^\alpha \psi^{(\beta)} \|_{\infty} \tag{34.4}
\]

**Lemma 34.4.** Each \( \| \cdot \|_{\alpha, \beta} \) is a norm on \( \mathcal{S} \).

It turns out that it is appropriate to topologize \( \mathcal{S} \) not with a single norm, but with the whole family of norms \( \| \cdot \|_{\alpha, \beta} \) simultaneously.

**Definition 34.5.** Say that a sequence \( \psi_n \subset \mathcal{S} \) is Cauchy if it is Cauchy in each of the norms \( \| \cdot \|_{\alpha, \beta} \), and say that a sequence \( \psi_n \subset \mathcal{S} \) converges if there exists \( \psi \in \mathcal{S} \) such that \( \| \psi_n - \psi \|_{\alpha, \beta} \to 0 \) for all \( \alpha, \beta \geq 0 \).

Of course, if \( \psi_n \) converges in \( \mathcal{S} \) then the limit \( \psi \) is unique. (Check this.) We also have:

**Proposition 34.6.** \( \mathcal{S} \) is complete. That is, if \( \psi_n \) is Cauchy in \( \mathcal{S} \), then there exists \( \psi \in \mathcal{S} \) such that \( \| \psi_n - \psi \|_{\alpha, \beta} \to 0 \) for all \( \alpha, \beta \geq 0 \).

*Proof.* Since \( C_b^\infty(\mathbb{R}) \) is complete with respect to the \( \| \cdot \|_{\infty} \) norm, by the definition of \( \mathcal{S} \) and the \( \| \cdot \|_{\alpha, \beta} \) norms we have that, for each \( \alpha, \beta \geq 0 \), there is a function \( \psi_{\alpha, \beta} \in C_b^\infty(\mathbb{R}) \) 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 \( \alpha, \beta \).

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 \( \psi_{0,1} = \psi'_0 \), and applying it inductively we have \( \psi_{0,\beta} = \psi^{(\beta)} \) for all \( \beta \geq 0 \). (In particular,
ψ^{(\beta)} \to \psi^{(\beta)}$ uniformly for all β.) From this it follows that $x^\alpha \psi^{(\beta)}_n \to x^\alpha \psi^{(\beta)}$ pointwise for all $\alpha, \beta$, but since this sequence also converges to $\psi_{\alpha, \beta}$ uniformly we conclude that $\psi_{\alpha, \beta} = x^\alpha \psi^{(\beta)}$ for all $\alpha, \beta$ as desired. □

We observed earlier that $\mathcal{S}$ is closed under differentiation and multiplication by polynomials; we now see that these operations are continuous:

**Lemma 34.7.** If $\psi_n \to \psi$ in $\mathcal{S}$, then $x^\alpha \psi_n \to x^\alpha \psi$ and $\psi^{(\beta)}_n \to \psi^{(\beta)}$ in $\mathcal{S}$.

*Proof.* Exercise. □

It is useful to observe (in connection with our discussion of the Fourier transform later) that convergence in the family of norms $\| \cdot \|_{\alpha, \beta}$ controls $L^p$ convergence:

**Proposition 34.8.** $\mathcal{S} \subset L^p$ for all $1 \leq p \leq \infty$, and if $\psi_n \to \psi$ in $\mathcal{S}$ then also $\psi_n \to \psi$ in $L^p(\mathbb{R})$.

*Proof.* By the previous lemma, we have $(1 + x^2)\psi \in \mathcal{S}$ for all $\psi \in \mathcal{S}$, so in particular for all $x \in \mathbb{R}$

$$|\psi(x)| \leq \frac{|\psi(x)| + |x^2 \psi(x)|}{1 + x^2} \leq \frac{\|\psi\|_{0,0} + \|\psi\|_{2,0}}{1 + x^2}. \tag{34.5}$$

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

$$\|\psi_n - \psi\|_p \leq (\|\psi_n - \psi\|_{0,0} + \|\psi_n - \psi\|_{2,0}) \|(1 + x^2)^{-1}\|_p \tag{34.6}$$

and the right hand side goes to 0 as $\psi_n \to \psi$ in $\mathcal{S}$. □

**Definition 34.9.** The space of *tempered distributions* $\mathcal{S}'$ consists of all continuous linear maps $F : \mathcal{S} \to \mathbb{C}$. That is, a map $F : \mathcal{S} \to \mathbb{C}$ belongs to $\mathcal{S}'$ if and only if it is linear and $F(\psi_n) \to F(\psi)$ whenever $\psi_n \to \psi$ in $\mathcal{S}$.

It is straightforward to check that $\mathcal{S}'$ is a vector space. To emphasize the role of $\mathcal{S}'$ as the dual space of $\mathcal{S}$, we will write $\langle F, \psi \rangle$ for $F(\psi)$. Tempered distributions $F$ are sometimes called *generalized functions*. We will topologize $\mathcal{S}'$ as follows: say $F_n \to F$ in $\mathcal{S}'$ if $\langle F_n, \psi \rangle \to \langle F, \psi \rangle$ for all $\psi \in \mathcal{S}$.

The following examples are fundamental; the unproved claims are left as exercises.

**Examples 34.10.**

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

$$\langle f, \psi \rangle = \int_\mathbb{R} f \psi. \tag{34.7}$$

(To see that $f \psi \in L^1$ for every $\psi \in \mathcal{S}$, 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.”

67
b) (Tempered measures) A (positive, signed, or complex) Borel measure $\mu$ 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

$$\langle \mu, \psi \rangle = \int_{\mathbb{R}} \psi \, d\mu.$$  

(34.8)

If $\mu$ is absolutely continuous with respect to Lebesgue measure $m$, with Radon-Nikodym derivative $f = \frac{d\mu}{dm}$, then $\mu$ 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 34.11.** If $F \in \mathcal{S}'$ and $T : \mathcal{S} \to \mathcal{S}$ is a continuous linear map, then

$$\langle T'F, \psi \rangle := \langle F, T\psi \rangle$$

(34.9)

defines a tempered distribution.

**Proof.** If $\psi_n \to \psi$ in $\mathcal{S}$, then

$$\langle F, T\psi_n \rangle \to \langle F, T\psi \rangle$$

(34.10)

so $T'F$ defines a distribution. \hfill \Box

Before moving on to more general classes of distributions, we consider one more special example:

**Proposition 34.12** (The Principal Value integral). For each $\psi \in \mathcal{S}$, the limit

$$\langle P_{1/x}, \psi \rangle := \lim_{\epsilon \to 0} \int_{|x|\geq \epsilon} \frac{1}{x} \psi(x) \, dx$$

(34.11)

exists, and defines a tempered distribution.

**Proof.** We first show that (34.11) is well-defined on $\mathcal{S}$. Let $\psi \in \mathcal{S}$, then by changing variables in the integral on the negative half-line we get for each $\epsilon > 0$

$$\int_{|x|\geq \epsilon} \frac{1}{x} \psi(x) \, dx = \int_{\epsilon}^{\infty} \frac{\psi(x) - \psi(-x)}{x} \, dx$$

(34.12)

Since $\psi$ 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

$$\int_{0}^{\infty} \frac{\psi(x) - \psi(-x)}{x} \, dx.$$  

(34.13)

To see that $P_{1/x}$ is continuous on $\mathcal{S}$, first observe that for $x > 0$

$$\left| \frac{\psi(x) - \psi(-x)}{x} \right| = \left| \frac{1}{x} \int_{-x}^{x} \psi'(t) \, dt \right|$$

(34.14)

$$\leq \frac{1}{x} \int_{-x}^{x} |\psi'(t)| \, dt$$

(34.15)

$$\leq 2\|\psi'\|_{\infty}$$

(34.16)
It follows that
\[ |\langle P_{1/x}, \psi \rangle| \leq \int_{0}^{1} \left| \frac{\psi(x) - \psi(-x)}{x} \right| + \int_{1}^{\infty} \left| \frac{\psi(x) - \psi(-x)}{x} \right| \]
\[ \leq 2 \| \psi' \|_{\infty} + \int_{1}^{\infty} (|x\psi(x)| + |x\psi(-x)|) \frac{dx}{x^2} \]
\[ \leq 2 \| \psi' \|_{\infty} + 2 \| \psi(x) \|_{\infty} \]
\[ = 2 \| \psi \|_{0,1} + 2 \| \psi \|_{1,0} \]  
(34.17)
\[ \leq 2 \| \psi \|_{0,1} + 2 \| \psi \|_{1,0} \]  
(34.18)
\[ = 2 \| \psi \|_{0,1} + 2 \| \psi \|_{1,0} \]  
(34.19)
\[ = 2 \| \psi \|_{0,1} + 2 \| \psi \|_{1,0} \]  
(34.20)

If we consider now \( \psi_n \to \psi \) in \( \mathcal{S} \), 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. \( \square \)

**Proposition 34.13** (Differentiation of tempered distributions). For any integer \( \beta \geq 0 \) and any tempered distribution \( F \in \mathcal{S}' \), the map
\[ \psi \to \langle F, (-1)^{\beta} \psi^{(\beta)} \rangle \]  
(34.21)
defines a tempered distribution, called the \( \beta^{th} \) distributional derivative of \( F \), denoted \( F^{(\beta)} \).

**Proof.** By Lemma ??, the map \( T \psi = (-1)^{\beta} \psi^{(\beta)} \) is continuous on \( \mathcal{S} \), and the result follows by Proposition 34.11. \( \square \)

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 34.14** (The Heaviside function). Let \( H \) be the Heaviside function \( H(x) = 1_{[0,\infty)} \). Then \( H' = \delta \) in the sense of distributions.

**Proof.** Let \( \psi \in \mathcal{S} \). \( H \) is a tempered function, so \( H' \) is given by
\[ \langle H', \psi \rangle = -\langle H, \psi' \rangle \]
\[ = -\int_{-\infty}^{\infty} H(x) \frac{d\psi}{dx} \, dx \]
\[ = -\int_{0}^{\infty} \frac{d\psi}{dx} \, dx \]
\[ = \psi(0) \]
\[ = \langle \delta, \psi \rangle. \]  
(34.22)

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 \( \langle \delta', \psi \rangle = -\psi'(0) \). It can be shown that \( \delta' \) is not given by any tempered measure (see Problem 35.22).

Our next use of Proposition 34.11 will allow us to define the convolution of a distribution with a Schwartz function \( \phi \).

**Proposition 34.15.** Let \( \phi \in \mathcal{S} \) and \( F \in \mathcal{S}' \). Then the map
\[ \langle \phi * F, \psi \rangle := \langle F, \tilde{\phi} * \psi \rangle \]  
(34.22)
defines a tempered distribution, called the convolution of \( F \) and \( \phi \). (Here \( \tilde{\phi}(x) = \phi(-x) \).)
Proof. Again it suffices to verify that the map $\psi \to \tilde{\phi} \ast \psi$ is continuous on $\mathcal{S}$. The proof is left as Problem 35.24.

If $f \in L^1$, one can also verify that $f \ast \phi$, viewed as a distribution, agrees with the distribution induced by the $L^1$ function $f \ast \phi$ defined by ordinary convolution (Problem 35.24).

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 $\phi_\lambda$ is a tempered function and hence defines a distribution.

**Proposition 34.16.** If $\phi_\lambda$ is an $L^1$ approximate unit, then $\phi_\lambda \to \delta$ in $\mathcal{S}'$.

**Proof.** By definition, for any $\psi \in \mathcal{S}$

$$\langle \phi_\lambda, \psi \rangle = \int_{-\infty}^{\infty} \phi_\lambda(y)\psi(y) \, dy = (\phi_\lambda \ast \tilde{\psi})(0)$$  \hspace{1cm} (34.23)

where $\tilde{\psi}(x) = \psi(-x)$. As $\lambda \to 0$, by Lemma 30.6 we have $(\phi_\lambda \ast \tilde{\psi})(0) \to \psi(0) = \langle \delta, \psi \rangle$. \hfill \Box

Finally we consider the Fourier transform. The key fact is the following:

**Lemma 34.17.** If $\psi \in \mathcal{S}$, then $\hat{\psi} \in \mathcal{S}$, and the map $\hat{\cdot}: \mathcal{S} \to \mathcal{S}$ is continuous.

**Proof.** The fact that $\hat{\psi}$ belongs to $\mathcal{S}$ follows from repeated application of Propositions 29.6 and 29.8. Continuity follows from the fact that if $\psi_n \to \psi$ in $\mathcal{S}$, then also $\|x^\alpha \frac{d^\beta}{dx^\beta}(\psi_n - \psi)\|_1 \to 0$ for all $\alpha, \beta$. See Problem 35.23. \hfill \Box

**Proposition 34.18** (Fourier transforms of tempered distributions). If $\psi \in \mathcal{S}$, then $\hat{\psi} \in \mathcal{S}$, and for any $F \in \mathcal{S}'$ the formula

$$\langle \hat{F}, \psi \rangle := \langle F, \hat{\psi} \rangle$$  \hspace{1cm} (34.24)

defines a tempered distribution, called the Fourier transform of $F$.

**Examples 34.19.**

a) Let $\delta_t$ be the point mass at $t \in \mathbb{R}$. We can compute $\hat{\delta}_t$: for $\psi \in \mathcal{S}$ we have

$$\langle \hat{\delta}_t, \psi \rangle = \langle \delta_t, \hat{\psi} \rangle$$  \hspace{1cm} (34.25)

$$= \hat{\psi}(t)$$  \hspace{1cm} (34.26)

$$= \int_{-\infty}^{\infty} e^{-2\pi ixt} \psi(x) \, dx$$  \hspace{1cm} (34.27)

$$= \langle e^{-2\pi ixt}, \psi \rangle$$  \hspace{1cm} (34.28)

so $\hat{\delta}_t = e^{-2\pi ixt}$.

The expected inversion $(e^{-2\pi ixt}) = \delta_t$ also holds; the proof is left as an exercise.

b) Consider the distribution $P_{1/x}$ of Proposition 34.12. One can show that $P_{1/x}$ is the tempered distribution given by the tempered function

$$F(t) = -\pi i \text{sgn}(t).$$  \hspace{1cm} (34.29)
35. Problems

Problem 35.1. Prove Proposition 29.2

Problem 35.2. Complete the proof of Lemma 29.4.

Problem 35.3. Prove Proposition 30.2

Problem 35.4.  
(a) Prove that if $f \in L^\infty$ and $g \in L^1$, then $f \ast g$ is continuous.
(b) Prove that 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 $h(x) = 1_{-E} \ast 1_E$.)

Problem 35.5. Suppose $\phi$ 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 that $\{\phi_\lambda\}_{\lambda > 0}$ is an $L^1$ approximate unit.
(b) Give a simpler proof of Lemma 30.6 by making a change of variables in equation (30.4).

Problem 35.6.  
(a) Prove that $f \in C^1_c(\mathbb{R})$ and $g$ is a compactly supported $L^1$ function, then $f \ast g$ is $C^1$ with compact support. (Hint: justify differentiation under the integral sign.)
(b) By induction, conclude that if $f \in C^\infty_c(\mathbb{R})$ and $g \in L^1$ is compactly supported, then $f \ast g \in C^\infty_c(\mathbb{R})$.
(c) Conclude that $C^\infty_c(\mathbb{R})$ is dense in $L^p$ for all $1 \leq p < \infty$. (Apply Theorem 30.5 with $\phi$ a bump function.)
(d) Construct a bump function on $\mathbb{R}^n$ and extend the above results to $n > 1$.

Problem 35.7. Compute the integral in Lemma 31.3.

Problem 35.8. This problem gives a proof that the Fourier transform $\hat{\cdot} : L^1 \to C_0(\mathbb{R})$ is not surjective.

(a) Compute $h_n := 1_{[-n,n]} \ast 1_{[-1,1]}$ explicitly.
(b) Show that $h_n$ is, up to a multiplicative constant, the Fourier transform of the $L^1$ function

$$f_n := \frac{\sin 2\pi x \sin 2\pi nx}{x^2}.$$ 

(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 35.9. 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 that $\hat{f}' = 2\pi i t \hat{f}(t)$.

Problem 35.10. Complete the proof of Theorem 32.3.

Problem 35.11. Let $\varphi_\lambda$ be an $L^1(\mathbb{T})$ approximate unit. Prove that if $f \in C(\mathbb{T})$, then $f \ast \varphi_\lambda \to f$ uniformly as $\lambda \to 0$.

Problem 35.13. Prove Theorem 33.3.

Problem 35.14. Prove Theorem 33.12. Also prove that $L^2$ inversion is possible in the following sense: for all $f \in L^2(\mathbb{T})$, we have

$$\lim_{N \to \infty} \frac{1}{2\pi} \int_{-\pi}^{\pi} \left| f(\theta) - \sum_{n=-N}^{N} \hat{f}(n)e^{in\theta} \right|^2 d\theta = 0. \quad (35.2)$$

(In other words, the partial sums $s_N$ of the Fourier series converge to $f$ in the $L^2$ norm.)

Problem 35.15. Let $A(\mathbb{T})$ denote the set of functions $f \in L^1(\mathbb{T})$ such that the Fourier transform $\hat{f}$ belongs to $\ell^1(\mathbb{Z})$.

a) Prove that if $f, g \in A(\mathbb{T})$, then their product $fg$ also belongs to $A(\mathbb{T})$. (Hint: use the $\ell^1$ 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, $A(\mathbb{T})$ is a ring.

b) Prove that the Fourier transform is a ring isomorphism from $A(\mathbb{T})$ onto $\ell^1(\mathbb{Z})$ (where the multiplication on $\ell^1(\mathbb{Z})$ is convolution).

Problem 35.16. Prove that if $f \in C^k(\mathbb{T})$, $k \geq 1$, then the Fourier coefficients of $f$ satisfy

$$\lim_{n \to \pm\infty} |n|^k |\hat{f}(n)| = 0. \quad (35.3)$$

(Hint: first compute the Fourier transform of $f'$ explicitly.)

Problem 35.17. Fill in the details in the proof of Theorem 33.4. To show that $\|D_N\|_1 \to \infty$, fix $N$, and for each $0 \leq k \leq 2N$ let $I_k$ denote the interval

$$\left[ \frac{1}{2} \left( k\pi + \frac{1}{2} \right), \frac{1}{2} \left( (k+1)\pi + \frac{1}{2} \right) \right]$$

(These are the intervals on which $D_N$ has constant sign.) Then

$$\int_{-\pi}^{\pi} |D_N(t)| \, dt = 2 \sum_{k=0}^{2N} \int_{I_k} \left| \frac{\sin(N + \frac{1}{2})t}{\sin \frac{t}{2}} \right| \, dt \quad (35.4)$$

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 35.18. Prove Proposition 33.11.

Problem 35.19 (Fourier transforms of measures). Let $\mu$ be a finite (signed or complex) Borel measure on $\mathbb{T}$. The Fourier transform of $\mu$ is the function $\hat{\mu} : \mathbb{Z} \to \mathbb{C}$ defined by

$$\hat{\mu}(n) := \int_{\mathbb{T}} e^{-in\theta} \, d\mu(\theta). \quad (35.5)$$

a) Prove that $\hat{\mu}$ is bounded. Give an example of a measure $\mu$ such that $\hat{\mu} \notin c_0(\mathbb{Z})$. 
b) For fixed \( \mu \), define for each \( 0 \leq r < 1 \)

\[
A(r, \theta) := \sum_{n=-\infty}^{\infty} \hat{\mu}(n) r^{|n|} e^{in\theta}.
\]  

(35.6)

Prove that the measures \( \mu_r := A(r, \theta) dm(\theta) \) converge to \( \mu \) as \( r \to 1 \), in the following sense:

for every continuous function \( f \) on \( \mathbb{T} \),

\[
\lim_{r \to 1} \int_{\mathbb{T}} f(\theta) A(r, \theta) dm(\theta) = \int_{\mathbb{T}} f(\theta) d\mu(\theta).
\]  

(35.7)

**Problem 35.20** (The Fejér kernel). Consider the cutoff function on \( \mathbb{Z} \)

\[
\psi_N(k) = \begin{cases} 
0 & \text{if } |k| > N \\
1 - \frac{|k|}{N} & \text{if } |k| \leq N 
\end{cases}
\]  

(35.8)

Find a closed form expression for

\[
F_N(\theta) = \sum_{k=-N}^{N} \psi_N(k) e^{ik\theta},
\]  

(35.9)

and show that the family \( \{F_N(\theta)\}_{N \geq 1} \) is an \( L^1(\mathbb{T}) \) approximate unit.

**Problem 35.21.** Prove that \( \mathcal{S} \) is dense in \( L^p(\mathbb{R}) \) for \( 1 \leq p < \infty \).

**Problem 35.22.** Prove that there is no finite Borel measure \( \mu \) on \([ -1, 1 ]\) such that \( \int_{-1}^{1} f d\mu = f'(0) \) for all \( f \in C^1[-1,1] \).

**Problem 35.23.**

a) Complete the proof of Lemma 34.17.

c) Prove that if \( \phi, \psi \in \mathcal{S} \) then

\[
\int_{\mathbb{R}} \hat{\phi} \psi = \int_{\mathbb{R}} \phi \hat{\psi}.
\]  

(35.10)

(This justifies the definition of \( \hat{\mathcal{F}} \).)

b) Prove that Fourier inversion holds in \( \mathcal{S} \); that is, \( (\hat{\psi}) = \psi \).

c) State and prove a Fourier inversion theorem for the Fourier transform \( F \to \hat{F} \) on \( \mathcal{S} \).

**Problem 35.24.**

a) Prove that if \( \phi \in \mathcal{S} \) and \( \psi_n \to \psi \) in \( \mathcal{S} \), then \( \phi \ast \psi_n \to \phi \ast \psi \) in \( \mathcal{S} \). (Here convolution means ordinary convolution of functions.)

b) Let \( f \in L^1 \) and \( \phi \in \mathcal{S} \). Prove that the tempered distribution \( f \ast \phi \) coincides with the distribution defined by the tempered function \( f \ast \phi \).

c) Show that \( \delta \ast \phi = \phi \) in the sense of distributions.

d) Let \( \phi \in \mathcal{S} \) be a nonnegative function with \( \int \phi = 1 \), and let \( \phi_{\lambda}(x) := \frac{x}{\lambda^2} \phi \left( \frac{x}{\lambda} \right) \) the corresponding \( L^1 \) approximate unit. Prove that for any \( F \in \mathcal{S}' \), \( F \ast \phi_{\lambda} \to F \) in \( \mathcal{S}' \) as \( \lambda \to 0 \).
36. **Some probabilistic concepts**

In this section we sketch out the fundamentals of the measure-theoretic approach to probability, emphasizing what distinguishes probability from analysis. We begin with the following “dictionary”:

<table>
<thead>
<tr>
<th><strong>analysis</strong></th>
<th><strong>probability</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>(\sigma)-algebra (\mathcal{M})</td>
<td>(\sigma)-field (\mathcal{E})</td>
</tr>
<tr>
<td>measurable space ((X, \mathcal{M}))</td>
<td>sample space ((\Omega, \mathcal{E}))</td>
</tr>
<tr>
<td>measure space ((X, \mathcal{M}, \mu)) with (\mu(X) = 1)</td>
<td>probability space ((\Omega, \mathcal{E}, P))</td>
</tr>
<tr>
<td>measurable function (f : X \to \mathbb{R})</td>
<td>random variable (X : \Omega \to \mathbb{R})</td>
</tr>
<tr>
<td>integral of (f), (\int_X f , d\mu)</td>
<td>expectation of (X), (\mathbb{E}(X) = \int_{\Omega} X(\omega) , dP(\omega))</td>
</tr>
<tr>
<td>(L^p) function, (\int</td>
<td>f</td>
</tr>
<tr>
<td>push-forward measure, (f_\ast \mu(E) = \mu(f^{-1}(E)))</td>
<td>distribution (P_X(B) = P(X^{-1}(B)) = P(X \in B))</td>
</tr>
</tbody>
</table>

For the remainder of this section we will use probabilistic language. There are a few important probabilistic notions that have no counterpart in the analysis we have studied so far.

**Definition 36.1.** Let \(X\) be a real-valued random variable with \(\mathbb{E}(|X|^2) < \infty\). The **variance** of \(X\) is the quantity

\[
\text{var}(X) = \sigma^2(X) := \mathbb{E}((X - \mathbb{E}(X))^2)
\]  

and \(\sigma(X)\) is called the **standard deviation** of \(X\).

The concept which most clearly distinguishes probability from analysis is that of **independence**. To introduce it we first look more closely at the distribution of a random variable. We recall the push-forward measure construction: if \((\Omega, \mathcal{E}, P)\) is a probability space, a random variable \(X : \Omega \to \mathbb{R}\) induces a Borel measure \(P_X\) on \(\mathbb{R}\) via the formula

\[
P_X(B) = P(X^{-1}(B))
\]

for each Borel set \(B \subset \mathbb{R}\). The set \(X^{-1}(B) \in \mathcal{E}\) is interpreted as the event “\(X\) lies in \(B\).”

We then have:

**Proposition 36.2.** Let \(f : \mathbb{R} \to \mathbb{R}\) be a Borel measurable function and \(X\) a random variable. Suppose that either \(f\) is unsigned, or \(\mathbb{E}(|f(X)|) < \infty\). Then

\[
\mathbb{E}(f(X)) := \int_{\Omega} f(X(\omega)) \, dP(\omega) = \int_{\mathbb{R}} f(t) \, dP_X(t).
\]

**Proof.** The equation is true by definition when \(f = 1_B\), and follows for general \(f\) by the usual approximation arguments. \(\square\)

Likewise, if \(X_1, \ldots, X_n\) are random variables defined on the same probability space, we can assemble them in to a single \(\mathbb{R}^n\)-valued random variable, and define the **joint distribution** of
Let us remark that it is easy to construct random variables with a given joint distribution $\mu$. Indeed, given a probability measure $\mu$ on $\mathbb{R}^n$, we form the probability space $\mathbb{R}^n, \mathcal{B}_{\mathbb{R}^n}, \mu);$ the random variables $X_j(t_1, \ldots, t_n) = t_j$ then have joint distribution $\mu$. It is also important to observe that the joint distribution is not determined by the individual distributions. There is an important exception, though, in case the random variables are independent, which we now discuss.

Without discussing it formally, let us recall that two events $E, F$ should be called independent if the probability of $E$ occurring, given that $F$ has occurred, is the same as the probability of $E$ occurring. That is, $P(E \cap F)/P(F) = P(E)$, or $P(E \cap F) = P(E) \cap P(F)$. We may then define:

**Definition 36.3.** A set of events $E_1, \ldots, E_n \in \mathcal{E}$ is called independent if

$$P(E_1 \cap \cdots \cap E_n) = \prod_{j=1}^{n} P(E_j). \quad (36.6)$$

Now, each random variable $X$ generates a family of events, namely $X^{-1}(B)$ for each Borel set $B \subset \mathbb{R}$. So we will say that a set of random variables $X_1, \ldots, X_n$ is independent if they generate independent events:

**Definition 36.4.** A set of random variables $X_1, \ldots, X_n$ is independent if for all Borel sets $B_1, \ldots, B_n \subset \mathbb{R}$, the events $X_1^{-1}(B_1), \ldots, X_n^{-1}(B_n)$ are independent.

As noted above, when random variables are independent, their joint distribution can be read off from the individual distributions:

**Proposition 36.5.** Let $X_1, \ldots, X_n$ be independent random variables. Then their joint distribution is the product of the individual distributions:

$$P_{(X_1, \ldots, X_n)} = P_{X_1} \times \cdots \times P_{X_n} \quad (36.7)$$

**Proof.** To determine $P_{(X_1, \ldots, X_n)}$, we must determine $P_{(X_1, \ldots, X_n)}(B)$ for every Borel set $B$, but since the Borel $\sigma$-algebra is generated by measurable rectangles $B_1 \times \cdots \times B_n$, we may assume
\( B \) has this form. Using the definition of distribution and independence, we have
\[
P_{(X_1, \ldots, X_n)}(B_1 \times \cdots \times B_n) = P((X_1, \ldots, X_n)^{-1}(B_1 \times \cdots \times B_n))
\]
\[
= \prod_{j=1}^{n} P(X_j^{-1}(B_j))
\]
\[
= \prod_{j=1}^{n} P_X(j)(B_j)
\]
\[
= \left( \prod_{j=1}^{n} P_X(j) \right)(B_1 \times \cdots \times B_n).
\]

\[\square\]

**Corollary 36.6.** Suppose that \( X_1, \ldots, X_n \) are independent random variables. Let \( f_1, \ldots, f_n : \mathbb{R} \to \mathbb{R} \) be measurable functions and suppose that either each \( f_j \) is unsigned, or \( \mathbb{E}(|f_j(X_j)|) < \infty \) for each \( j \). Then
\[
\mathbb{E}(f_1(X_1) \cdots f_n(X_n)) = \prod_{j=1}^{n} \mathbb{E}(f_j(X_j)).
\]

**Proof.** Equation 36.5, Proposition 36.5, and Fubini’s theorem. \( \square \)

**Corollary 36.7.** If \( X_1, \ldots, X_n \) are independent and \( \mathbb{E}(|X_j|^2) < \infty \) for each \( j \), then
\[
\text{var}(X_1 + \cdots + X_n) = \sum_{j=1}^{n} \text{var}(X_j).
\]

**Proof.** We first replace each \( X_j \) by \( Y_j := X_j - \mathbb{E}(X_j) \), so that \( \mathbb{E}(Y_j) = 0 \) for each \( j \). Then the \( Y_j \) are independent (check this), and we have
\[
\mathbb{E}(Y_j Y_k) = \begin{cases} 
\text{var}(X_j) & \text{if } j = k, \\
0 & \text{otherwise} \end{cases}
\]

Thus
\[
\text{var}\left( \sum_{j=1}^{n} X_j \right) = \text{var}\left( \sum_{j=1}^{n} Y_j \right) = \sum_{j,k} \mathbb{E}(Y_j Y_k) = \sum_{j=1}^{n} \text{var}(X_j).
\]

\[\square\]

36.1. **The Law of Large Numbers.** We will be interested in an infinite sequence of iid random variables; in other words, infinitely many independent copies of a single random variable \( X \). We introduce the notation
\[
\bar{X}_n := \frac{1}{n} \sum_{j=1}^{n} X_j
\]

Loosely speaking, a law of large numbers asserts that the \( \bar{X}_n \) converge to \( \mathbb{E}(X) \) in some sense. We will sketch a proof of the following:
Theorem 36.8 (The Strong Law of Large Numbers). If \(X_n\) is a sequence of iid \(L^2\) random variables, then \(\bar{X}_n \to \mathbb{E}(X)\) almost surely.

In this section we give only an outline of the proof; the exercises for this section are to fill in all the details. We begin with an elementary but very frequently used result. Recall that if \((E_n)\) is sequence of sets, their lim sup is defined by

\[
\limsup_{n} E_n = \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} E_k.
\]

(36.13)

In other words, a point belongs to \(\limsup E_n\) if and only if it belongs to infinitely many of the \(E_n\).

Lemma 36.9 (The Borel-Cantelli Lemma).

a) If \((E_j)\) is a sequence of events, and \(\sum_n P(E_n) < \infty\), then \(P(\limsup E_n) = 0\).

b) If \((E_j)\) is a sequence of independent events, and \(\sum_n P(E_n) = +\infty\), then \(P(\limsup E_n) = 1\).

Proof. □

The next lemmas also represent frequently used techniques in probability (the zeroth moment method (or union bound), first moment method, and second moment method respectively).

Lemma 36.10 (Zeroth moment tail bound). \(P(X_n \neq 0) \leq nP(X \neq 0)\).

Proof. □

Lemma 36.11 (First moment tail bound). If \(E(|X|) < \infty\), then for all \(t > 0\)

\[
P(|X_n| \geq t) \leq \frac{E(|X|)}{t}.
\]

(36.14)

Proof. □

A stronger bound is available if we assume that \(X\) also has finite second moment:

Lemma 36.12 (Second moment tail bound). If \(E(|X|^2) < \infty\), then

\[
P(|X_n - \mathbb{E}(X)| \geq t) \leq \frac{\text{var}(X)}{nt^2}.
\]

(36.15)

Proof. □

The final ingredients in the proof of the strong law will be some elementary results on truncation of random variables. As we have already seen in the analysis context (e.g. in the proof of the Marcinkiewicz interpolation theorem), it is often useful to split random variables into their small and large parts.

Lemma 36.13. Let \(X\) be a random variable and \(N > 0\) a real number. Let

\[
X_{>N} := 1_{|X| > N}X, \quad X_{\leq N} := 1_{|X| \leq N}X.
\]

(36.16)

Then \(X = X_{>N} + X_{\leq N}\) and the following hold:

a) \(E(|X_{\leq N}|^2) \leq N E(|X|)\),

b) \(\text{var}(X_{\leq N}) \leq N E(|X|)\), and

77
c) \( \mathbb{E}(|X_{>N}|) \to 0 \) as \( N \to \infty \).

Proof. □

Proof of Theorem 36.8. By considering positive and negative parts separately, we may assume that \( X \geq 0 \) almost surely. Since all the \( X_n \) are now nonnegative, we observe that the means \( \overline{X}_n \) obey a weak monotonicity: if \( n \geq m \), we have
\[
\overline{X}_n \geq \frac{m \overline{X}_m}{n}
\]
and thus if \( (1 - \epsilon)n \leq m \leq n \), we have
\[
\overline{X}_n \geq (1 - \epsilon)\overline{X}_m.
\]
Thus the \( \overline{X}_n \) cannot decrease too quickly in \( n \). This weak monotonicity allows us to reduce to proving almost sure convergence along “sparse” subsequences of \( \overline{X}_n \). More precisely, say that a sequence of integers \( 1 \leq n_1 \leq n_2 \leq \cdots \leq n_j \leq \cdots \) is lacunary if there exists a constant \( c > 1 \) such that \( n_j + 1/n_j \geq c \) eventually. Then we have:

Claim: Let \( X_n \) be a sequence of unsigned, iid random variables with \( E(|X|) < \infty \). If \( \overline{X}_{n_j} \to \mathbb{E}(X) \) almost surely along every lacunary sequence \( (n_j) \), then \( \overline{X}_n \to \mathbb{E}(X) \) almost surely.

Proof of claim: Assume the claim holds and let \( 0 < \epsilon < 1 \). Applying the claim to the lacunary sequence \( n_j = \lceil (1 + \epsilon)^j \rceil \), and using the weak monotonicity (36.18), we see that \( \limsup |\overline{X}_n - \mathbb{E}(X)| \leq \epsilon |\mathbb{E}(X)| \) almost surely. Letting \( \epsilon \to 0 \) along a countable sequence of values proves the claim. □

Thus, it suffices to prove the “lacunary” version of the strong law. By repeated application of Borel-Cantelli, it suffices to prove that

\[
\sum_{j=1}^{\infty} P(|\overline{X}_{n_j} - \mathbb{E}(X)| \geq \epsilon) < \infty
\]

(36.19)

for every lacunary sequence \( (n_j) \) and every \( \epsilon > 0 \). To control the probabilities \( P(|\overline{X}_{n_j} - \mathbb{E}(X)| \geq \epsilon) \), we truncate the \( X_{n_j} \) and apply the second and zeroth tail bounds. To estimate the average \( \overline{X}_{n_j} \), we truncate each copy of \( X \) at a cutoff \( N_j \); we will leave the values of \( N_j \) unspecified and optimize them later. At this point we only insist that \( N_j \) is chosen large enough so that \( |\mathbb{E}(X_{\leq N_j}) - \mathbb{E}(X)| < \epsilon \).

So, let us fix \( \epsilon \) (we may assume \( \epsilon < \mathbb{E}(X) \)) and the lacunary sequence \( n_j \). For each \( j \), from the second moment tail bound we have
\[
P(|\overline{X}_{n_j} \leq N_j| - \mathbb{E}(X)| \geq \epsilon/2) \leq \left( \frac{2}{\epsilon n_j} \right)^2 \mathbb{E}(|X_{\leq N_j}|^2)
\]

(36.20)

We control the large parts \( X_{>N_j} \) using the zeroth moment bound, which is a useful estimate here since \( X_{>N_j} \) is mostly zero. We have
\[
P(|\overline{X}_{n_j} \leq N_j| - \mathbb{E}(X)| \geq \epsilon/2) \leq P(|\overline{X}_{n_j} \leq N_j| \neq 0) \leq n_j P(X > N_j).
\]

(36.21)

Putting these together, we have
\[
P(|\overline{X}_{n_j} - \mathbb{E}(X)| \geq \epsilon) \leq \left( \frac{2}{\epsilon n_j} \right)^2 \mathbb{E}(|X_{\leq N_j}|^2) + n_j P(X > N_j).
\]

(36.22)
Now, it turns out that the simple guess $N_j = n_j$ works: indeed, since the sequence $n_j$ is lacunary one can show that for some absolute constant $C > 0$ (depending only on the lacunary parameter $c$) we have
\[
\sum_j \frac{1}{n_j} |X_{\leq n_j}(\omega)|^2 \leq CX(\omega) \quad \text{almost surely,} \tag{36.23}
\]
and
\[
\sum_j n_j 1_{X_{>N_j}}(\omega) \leq CX(\omega) \quad \text{almost surely.} \tag{36.24}
\]
Taking expectations of (36.23) and (36.24), and substituting into (36.22), we conclude that (36.19) holds, and the proof is finished. \[\square\]

36.2. **The Central Limit Theorem.** Suppose $(X_n)$ is a sequence of iid random variables with mean 0 and variance $\sigma^2$. Then the empirical means $X_n$ are random variables with mean 0 and variance $\sigma^2/n$. Since the variance goes to 0 as $n \to \infty$, we expect the $X_n$ to become tightly centered around their mean, and this is the content of the Law of Large Numbers. However, if we instead consider the random variables
\[
Z_n = \frac{1}{\sqrt{n}} \sum_{j=1}^n X_j, \tag{36.25}
\]
then the $Z_n$ all have mean 0 and variance $\sigma^2$, so we might ask if the $Z_n$ have some limiting distribution. The (surprising) answer is “yes,” and in fact the limiting distribution does not depend on the distribution of $X$. This is the content of the Central Limit Theorem. To state it we first introduce one more mode of convergence:

**Definition 36.14.** A sequence of $\mathbb{R}$-valued random variables $X_n$ *converges in distribution* to a random variable $X$ if for every bounded, continuous function $f : \mathbb{R} \to \mathbb{R}$, we have
\[
\mathbb{E}(f(X_n)) \to \mathbb{E}(f(X)). \tag{36.26}
\]

From the definition of distribution, this is equivalent to the condition
\[
\int_{\mathbb{R}} f(t) dP_{X_n}(t) \to \int_{\mathbb{R}} f(t) dP_X(t) \tag{36.27}
\]
for all $f \in C_b(\mathbb{R})$. Since the $P_X$ are all probability measures, by standard approximation arguments it suffices to check smaller classes of continuous functions:

**Lemma 36.15.** Let $(\mu_n)_{n=1}^\infty, \mu$ be probability measures on $\mathbb{R}$. Then $\int_{\mathbb{R}} f d\mu_n \to \int_{\mathbb{R}} f d\mu$ for all $f \in C_b(\mathbb{R})$ if and only if the convergence holds for all $f \in C_c^\infty(\mathbb{R})$.

*Proof. Exercise.* \[\square\]

If $\mu_n, \mu$ are all probability measures on $\mathbb{R}$, we say $\mu_n \to \mu$ *vaguely* if $\int_{\mathbb{R}} f d\mu_n \to \int_{\mathbb{R}} f d\mu$ for all $f \in C_b(\mathbb{R})$. Thus $X_n \to X$ in distribution if and only if $P_{X_n} \to P_X$ vaguely.

The *Gaussian distribution of mean $\mu$ and variance $\sigma^2$* (or standard normal distribution $\mathcal{N}(\mu, \sigma)$) is the measure on $\mathbb{R}$ given by the density
\[
\nu^\mu_{\sigma}(E) = \frac{1}{\sigma \sqrt{2\pi}} \int_{\mathbb{R}} \exp\left(-\frac{(t - \mu)^2}{2\sigma^2}\right) dt. \tag{36.28}
\]
One may check that
\[ \int_{\mathbb{R}} t \, d\nu_{\mu} = \mu, \quad \int_{\mathbb{R}} (t - \mu)^2 \, d\nu_{\mu} = \sigma^2. \]  
(36.29)

**Theorem 36.16** (The Central Limit Theorem). Let \((X_n)\) be a sequence of iid \(L^2\) random variables with mean \(\mu\) and variance \(\sigma^2\). Let
\[ Z_n := \frac{1}{\sqrt{n}} \sum_{j=1}^{n} X_j. \]  
(36.30)
Then \(Z_n \to \mathcal{N}(\mu, \sigma)\) in distribution.

There are several different proofs of the Central Limit Theorem available; since we have already studied the Fourier transform we will give the Fourier-analytic proof. Given a random variable \(X\), we define its **Fourier transform** (or, as a probabilist would say, its **characteristic function**) by
\[ F_X(t) := \mathbb{E}(e^{2\pi itX}) = \int_{\mathbb{R}} e^{2\pi iyt} dP_X(y). \]  
(36.31)
In other words, \(F_X\) is just the Fourier transform of the measure \(P_X\). As such, it is defined for any random variable \(X\); we do not need any integrability assumptions. Also, as we have already seen, \(F_X\) is a bounded, continuous function on \(\mathbb{R}\). The Fourier-analytic proof of CLT boils down to two facts: first, that the Gaussian is (up to some constants) its own Fourier transform, and second (which allows the first fact to be used), that convergence in distribution is governed by convergence of Fourier transforms. Specifically, we have:

**Theorem 36.17** (Levy continuity theorem, special case). Let \(X_n\) be a sequence of \(\mathbb{R}\)-valued random variables and \(X\) another \(\mathbb{R}\)-valued random variable. The following are equivalent:

a) \(X_n \to X\) in distribution.

b) \(F_{X_n} \to F_X\) pointwise.

**Proof.** (a) implies (b) is immediate from the definition of convergence in distribution, applied to the bounded continuous functions \(f_t(x) = e^{-2\pi itx}\). For the converse, we use Lemma 36.15. By the lemma, it suffices to show that
\[ \mathbb{E}(\psi(X_n)) \to \mathbb{E}(\psi(X)) \]  
(36.32)
for every Schwartz function \(\psi \in \mathcal{S}\). Since \(\psi \in \mathcal{S}\), we know that \(\hat{\psi} \in L^1\) and Fourier inversion holds:
\[ \psi(x) = \int_{-\infty}^{\infty} \hat{\psi}(t) e^{2\pi ixt} \, dt. \]  
(36.33)
For \(X_n\) defined on the probability space \((\Omega, \mathcal{E}, P)\) we then have for each \(\omega \in \Omega\)
\[ \psi(X_n(\omega)) = \int_{-\infty}^{\infty} \hat{\psi}(t) e^{2\pi iX_n(\omega)t} \, dt. \]  
(36.34)
Taking expectations (that is, integrating \(dP(\omega)\)) and using Fubini’s theorem, we have
\[ \mathbb{E}(\psi(X_n)) = \int_{-\infty}^{\infty} \hat{\psi}(t) F_{X_n}(t) \, dt. \]  
(36.35)
and the theorem follows by dominated convergence. \(\square\)
Proof of Theorem 36.16. Replacing each $X_n$ by $\frac{1}{\sigma}(X_n - \mu)$, we may assume that $\mu = 0$ and $\sigma = 1$. For brevity let $\lambda = P_X$ denote the common distribution of the $X_n$. By independence, the distribution of $Z_n$ is given by

$$P_{Z_n}(E) = \lambda \ast \lambda \ast \cdots \ast \lambda(\sqrt{n}E).$$  \hfill (36.36)

Write $\lambda_n$ for this measure. We are done if we show that $\lambda_n \rightarrow \nu_1^0$ vaguely; by the Levy continuity theorem it suffices to prove the pointwise convergence of the Fourier transforms.

By Lemma 29.12, we have $\hat{\nu}_1^0(t) = e^{-2\pi^2 t^2}$. To investigate the Fourier transform of $\lambda_n$, first observe from our normalizations we have $X \in L^2$ and

$$\int \mathbb{R} y d\lambda (y) = 0, \quad \int \mathbb{R} y^2 d\lambda (y) = 1.$$  \hfill (36.37)

Since $y, y^2$ both lie in $L^1(\lambda)$, it follows (by exactly the same arguments as for the Fourier transform of functions) that $\hat{\lambda}$ belongs to $C^2$, and

$$\hat{\lambda}(0) = 1, \quad \hat{\lambda}'(0) = 0, \quad \hat{\lambda}''(0) = -4\pi^2.$$  \hfill (36.38)

(See Corollary 29.7.) Taylor expanding $\hat{\lambda}$ we have

$$\hat{\lambda}(t) = 1 - 2\pi^2 t^2 + o(t^2)$$  \hfill (36.39)

where $o(x)$ denotes a quantity that satisfies $o(x)/x \rightarrow 0$ as $x \rightarrow 0$. Since the Fourier transform converts convolution to multiplication, and noting the behavior of the Fourier transform under rescaling, we have

$$\hat{\lambda}_n(t) = \left(1 - \frac{2\pi^2 t^2}{n} + o\left(\frac{t^2}{n}\right)\right)^n.$$  \hfill (36.40)

Taking logs, and using the fact that $\log(1 + x) = x + o(x)$, we have

$$\log \hat{\lambda}_n(t) = n \log \left(1 - \frac{2\pi^2 t^2}{n} + o\left(\frac{t^2}{n}\right)\right) = -2\pi^2 t^2 + n \cdot o\left(\frac{t^2}{n}\right)$$  \hfill (36.41)

which goes to $-2\pi^2 t^2$ as $n \rightarrow \infty$. Thus, $\hat{\lambda}_n \rightarrow e^{-2\pi^2 t^2} = \hat{\nu}_1^0$ pointwise, and we are done. \hfill \Box

We can improve the conclusion of the central limit theorem somewhat by exploiting the following characterizaition of vague convergence:

**Proposition 36.18.** Suppose $\mu_n, \mu$ are all Borel probability measures on $\mathbb{R}$. For any Borel measure $\nu$ on $\mathbb{R}$, let

$$F_{\nu}(a) = \nu((-\infty, a]).$$  \hfill (36.42)

Then $\mu_n \rightarrow \mu$ vaguely if and only if $F_{\mu_n}(a) \rightarrow F_{\mu}(a)$ at every point where $F_{\mu}$ is continuous.

**Proof.** \hfill \Box

**Corollary 36.19** (Central Limit Theorem, second version). Let $X_n$ be a sequence of iid $L^2$ random variables with mean $\mu$ and variance $\sigma^2$. Then for every real number $a$, we have

$$P\left(\frac{1}{\sigma \sqrt{n}} \sum_{j=1}^{n} (X_j - \mu) \geq a\right) \rightarrow \frac{1}{\sqrt{2\pi}} \int_{a}^{\infty} e^{-x^2/2} \, dx.$$  \hfill (36.43)

**Proof.** This is immediate by combining CLT and Proposition 36.18. \hfill \Box