For each \( f \in L^p(\Omega) \) we define the norm to be
\[
\|f\|_p = \left( \int_{\Omega} |f(x)|^p \mu(dx) \right)^{1/p}.
\] (2)

Sometimes we shall write this as \( \|f\|_{L^p(\Omega)} \) if there is possibility of confusion. This norm has the following three crucial properties that make it truly a norm:

(a) \( \|\lambda f\|_p = |\lambda|\|f\|_p \) for \( \lambda \in \mathbb{C} \).

(b) \( \|f\|_p = 0 \) if and only if \( f(x) = 0 \) for \( \mu \)-almost every point \( x \). (3)

(c) \( \|f + g\|_p \leq \|f\|_p + \|g\|_p \).

(technically, (2) only defines a semi-norm because of the ‘almost every’ caveat in (b), i.e., \( \|f\|_p \) can be zero without \( f \equiv 0 \). Later on, when we define equivalence classes, (2) will be an honest norm on these classes.) Property (a) is obvious and (b) follows from the definition of the integral. Less trivial is property (c) which is called the triangle inequality. It will follow immediately from theorem 2.4 (Minkowski’s inequality). The triangle inequality is the same thing as convexity of the norm, i.e., if \( 0 \leq \lambda \leq 1 \), then
\[
\|\lambda f + (1 - \lambda)g\|_p \leq \lambda\|f\|_p + (1 - \lambda)\|g\|_p.
\]

We can also define \( L^\infty(\Omega, d\mu) \) by
\[
L^\infty(\Omega, d\mu) = \{ f : f : \Omega \to \mathbb{C}, f \text{ is } \mu\text{-measurable and there exists a finite constant } K \text{ such that } |f(x)| \leq K \text{ for } \mu\text{-a.e. } x \in \Omega \}. \tag{4}
\]

For \( f \in L^\infty(\Omega) \) we define the norm
\[
\|f\|_\infty = \inf\{ K : |f(x)| \leq K \text{ for } \mu\text{-almost every } x \in \Omega \}. \tag{5}
\]

Note that the norm depends on \( \mu \). This quantity is also called the essential supremum of \( |f| \) and is denoted by \( \text{ess sup}_x |f(x)| \). (Do not confuse this with ess supp—which has one more p.) Unlike the usual supremum, ess sup ignores sets of \( \mu \)-measure zero. E.g., if \( \Omega = \mathbb{R} \) and \( f(x) = 1 \) if \( x \) is rational and \( f(x) = 0 \) otherwise, then (with respect to Lebesgue measure) \( \text{ess sup}_x |f(x)| = 0 \), while \( \text{sup}_x |f(x)| = 1 \).

One can easily verify that the \( L^\infty \) norm has the same properties (a), (b) and (c) as above. Note that property (b) would fail if ess sup is replaced by sup. Also note that \( |f(x)| \leq \|f\|_\infty \) for almost every \( x \).
We leave it as an exercise to the reader to prove that when \( f \in L^\infty(\Omega) \cap L^q(\Omega) \) for some \( q \) then \( f \in L^p(\Omega) \) for all \( p > q \) and
\[
\|f\|_\infty = \lim_{p \to \infty} \|f\|_p.
\] (6)

This equation is the reason for denoting the space defined in (4) by \( L^\infty(\Omega) \).

An important concept, whose meaning will become clear later, is the **dual index** to \( p \) (for \( 1 \leq p \leq \infty \), of course). This is often denoted by \( p' \), but we shall often use \( q \), and it is given by
\[
\frac{1}{p} + \frac{1}{p'} = 1.
\] (7)

Thus, 1 and \( \infty \) are dual, while the dual of 2 is 2.

Unfortunately, the norms we have defined do not serve to distinguish all different measurable functions, i.e., if \( \|f - g\|_p = 0 \) we can only conclude that \( f(x) = g(x) \) \( \mu \)-almost everywhere. To deal with this nuisance we can redefine \( L^p(\Omega, \mathrm{d}\mu) \) so that its elements are not functions but equivalence classes of functions. That is to say, if we pick an \( f \in L^p(\Omega) \) we can define \( \tilde{f} \) to be the set of all those functions that differ from \( f \) only on a set of \( \mu \)-measure zero. If \( h \) is such a function we write \( f \sim h \); moreover if \( f \sim h \) and \( h \sim g \), then \( f \sim g \). Consequently, two such sets \( \tilde{f} \) and \( \tilde{k} \) are either identical or disjoint. We can now define
\[
\|\tilde{f}\|_p := \|f\|_p
\]
for some \( f \in \tilde{f} \). The point is that this definition does not depend on the choice of \( f \in \tilde{f} \).

Thus we have two vector spaces. The first consists of functions while the second consists of equivalence classes of functions. (It is left to the reader to understand how to make the set of equivalence classes into a vector space.) For the first, \( \|f - g\|_p = 0 \) does not imply \( f = g \), but for the second space it does. Some authors distinguish these spaces by different symbols, but all agree that it is the second space that should be called \( L^p(\Omega) \). Nevertheless most authors will eventually slip into the tempting trap of saying 'let \( f \) be a function in \( L^p(\Omega) \)' which is technically nonsense in the context of the second definition. Let the reader be warned that we will generally commit this sin. Thus when we are talking about \( L^p \)-functions and we write \( f = g \) we really have in mind that \( f \) and \( g \) are two functions that agree \( \mu \) almost everywhere. If the context is changed to, say, continuous functions, then \( f = g \) means \( f(x) = g(x) \) for all \( x \). In particular, we note that it makes no sense to ask for the value \( f(0) \), say, if \( f \) is an \( L^p \)-function.
A convex set $K \subset \mathbb{R}^n$ is one for which $\lambda x + (1 - \lambda)y \in K$ for all $x, y \in K$ and all $0 \leq \lambda \leq 1$. A convex function, $f$, on a convex set $K \subset \mathbb{R}^n$ is a real-valued function satisfying

$$f(\lambda x + (1 - \lambda)y) \leq \lambda f(x) + (1 - \lambda)f(y)$$  \hspace{1cm} (8)

for all $x, y \in K$ and all $0 < \lambda \leq 1$. If equality never holds in (8) when $y \neq x$ and $0 < \lambda < 1$, then $f$ is strictly convex. More generally, we say that $f$ is strictly convex at a point $x \in K$ if $f(x) < \lambda f(y) + (1 - \lambda)f(z)$ whenever $x = \lambda y + (1 - \lambda)z$ for $0 < \lambda < 1$ and $y \neq z$. If the inequality (8) is reversed, $f$ is said to be concave (alternatively, $f$ is concave $\iff -f$ is convex). It is easy to prove that if $K$ is an open set, then a convex function is continuous.

A support plane to a graph of a function $f : K \to \mathbb{R}$ at a point $x \in K$ is a plane (in $\mathbb{R}^{n+1}$) that touches the graph at $(x, f(x))$ and that nowhere lies above the graph. In general, a support plane might not exist at $x$, but if $f$ is convex on $K$, its graph has at least one support plane at each point of the interior of $K$. Thus there exists a vector $V \in \mathbb{R}^n$ (which depends on $x$) such that

$$f(y) \geq f(x) + V \cdot (y - x)$$  \hspace{1cm} (9)

for all $y \in K$. If the support plane at $x$ is unique it is called a tangent plane. If $f$ is convex, the existence of a tangent plane at $x$ is equivalent to differentiability at $x$.

If $n - 1$ and if $f$ is convex, $f$ need not be differentiable at $x$. However, when $x$ is in the interior of the interval $K$, $f$ always has a right derivative, $f'_+(x)$, and a left derivative, $f'_-(x)$, at $x$, e.g.,

$$f'_+(x) := \lim_{\varepsilon \to 0^+} \frac{f(x + \varepsilon) - f(x)}{\varepsilon}.$$

See [Hardy–Littlewood–Pólya] and Exercise 18.

2.2 THEOREM (Jensen’s inequality)

Let $J : \mathbb{R} \to \mathbb{R}$ be a convex function. Let $f$ be a real-valued function on some set $\Omega$ that is measurable with respect to some $\Sigma$-algebra, and let $\mu$ be a measure on $\Sigma$. Since $J$ is convex, it is continuous and therefore $(J \circ f)(x) := J(f(x))$ is also a $\Sigma$-measurable function on $\Omega$. We assume that $\mu(\Omega) = \int_{\Omega} \mu(dx)$ is finite.

Suppose now that $f \in L^1(\Omega)$ and let $\langle f \rangle$ be the average of $f$, i.e.,

$$\langle f \rangle = \frac{1}{\mu(\Omega)} \int_{\Omega} f \, d\mu.$$
Then

(i) \([J \circ f]_-,\) the negative part of \([J \circ f],\) is in \(L^1(\Omega),\) whence
\[
\int_{\Omega} (J \circ f)(x) \mu(dx) \text{ is well defined although it might be } +\infty.
\]

(ii) \(\langle J \circ f \rangle \geq J(\langle f \rangle).\) \hspace{1cm} (1)

If \(J\) is strictly convex at \(\langle f \rangle\) there is equality in (1) if and only if \(f\) is a constant function.

PROOF. Since \(J\) is convex its graph has at least one support line at each point. Thus, there is a constant \(V \in \mathbb{R}\) such that

\[
J(t) \geq J(\langle f \rangle) + V(t - \langle f \rangle)
\]
for all \(t \in \mathbb{R}.\) From this we conclude that

\[
[J(f)]_-(x) \leq |J(\langle f \rangle)| + |V||\langle f \rangle| + |V||f(x)|,
\]
and hence, recalling that \(\mu(\Omega) < \infty,\) (i) is proved.

If we now substitute \(f(x)\) for \(t\) in (2) and integrate over \(\Omega\) we arrive at (1).

Assume now that \(J\) is strictly convex at \(\langle f \rangle\). Then (2) is a strict inequality either for all \(t > \langle f \rangle\) or for all \(t < \langle f \rangle\). If \(f\) is not a constant, then \(f(x) - \langle f \rangle\) takes on both positive and negative values on sets of positive measure. This implies the last assertion of the theorem.  

\[\blacksquare\]

• The importance of the next inequality can hardly be overrated. There are many proofs of it and the one we give is not necessarily the simplest; we give it in order to show how the inequality is related to Jensen’s inequality. Another proof is outlined in the exercises.

2.3 THEOREM (Hölder’s inequality)

Let \(p\) and \(q\) be dual indices, i.e., \(1/p + 1/q = 1\) with \(1 \leq p \leq \infty.\) Let \(f \in L^p(\Omega)\) and \(g \in L^q(\Omega).\) Then the pointwise product, given by \((fg)(x) = f(x)g(x),\) is in \(L^1(\Omega)\) and

\[
\left| \int_{\Omega} fg \, d\mu \right| \leq \int_{\Omega} |f||g| \, d\mu \leq \|f\|_p \|g\|_q.
\]

(1)
The first inequality in (1) is an equality if and only if

(i) \( f(x)g(x) = c^{i\theta}|f(x)||g(x)| \) for some real constant \( \theta \) and for \( \mu \)-almost every \( x \).

If \( f \neq 0 \) the second inequality in (1) is an equality if and only if there is a constant \( \lambda \in \mathbb{R} \) such that:

(ii) If \( 1 < p < \infty \), \( |g(x)| = \lambda|f(x)|^{p-1} \) for \( \mu \)-almost every \( x \).

(iii) If \( p = 1 \), \( |g(x)| \leq \lambda \) for \( \mu \)-almost every \( x \) and \( |g(x)| = \lambda \) when \( f(x) \neq 0 \).

(iv) If \( p = \infty \), \( |f(x)| \leq \lambda \) for \( \mu \)-almost every \( x \) and \( |f(x)| = \lambda \) when \( g(x) \neq 0 \).

REMARKS. (1) The special case \( p = q = 2 \) is the Schwarz inequality

\[
\left| \int_\Omega fg \right|^2 \leq \int_\Omega |f|^2 \int_\Omega |g|^2.
\]

(2) If \( f_1, \ldots, f_m \) are functions on \( \Omega \) with \( f_i \in L^p(\Omega) \) and \( \sum_{i=1}^m 1/p_i = 1 \) then

\[
\left| \int_\Omega \prod_{i=1}^m f_i \ d\mu \right| \leq \prod_{i=1}^m \| f_i \|_{p_i}.
\]

This generalization is a simple consequence of (1) with \( f := f_1 \) and \( g := \prod_{j=2}^m f_j \). Then use induction on \( \int_\Omega |g|^p \).

PROOF. The left inequality in (1) is a triviality, so we may as well suppose \( f \geq 0 \) and \( g \geq 0 \) (note that condition (i) is what is needed for equality here).

The cases \( p = \infty \) and \( q = \infty \) are trivial so we suppose that \( 1 < p, q < \infty \). Set \( A = \{ x : g(x) > 0 \} \subset \Omega \) and let \( B = \Omega \sim A = \{ x : g(x) = 0 \} \). Since

\[
\int_\Omega f^p \ d\mu = \int_A f^p \ d\mu + \int_B f^p \ d\mu,
\]

since \( \int_\Omega g^p \ d\mu = \int_A g^p \ d\mu \), and since \( \int_\Omega fg \ d\mu = \int_A fg \ d\mu \), we see that it suffices—in order to prove (1)—to assume that \( \Omega = A \). (Why is \( \int fg \ d\mu \) defined?) Introduce a new measure on \( \Omega - A \) by \( \nu(dx) = g(x)^{\alpha} \ d\mu(dx) \). Also, set \( F'(x) = f(x)g(x)^{-\alpha/p} \) (which makes sense since \( g(x) > 0 \) a.e.). Then, with respect to the measure \( \nu \), we have that \( \langle F' \rangle = \int_\Omega fg \ d\mu / \int_\Omega g^\alpha \ d\mu \). On the other hand, with \( J(t) = |t|^p, \int_\Omega J \circ h' \ d\nu = \int_\Omega f^p \ d\mu \). Our conclusion (1) is then an immediate consequence of Jensen’s inequality—as is the condition for equality.
2.4 THEOREM (Minkowski's inequality)

Suppose that $\Omega$ and $\Gamma$ are any two spaces with sigma-finite measures $\mu$ and $\nu$ respectively. Let $f$ be a nonnegative function on $\Omega \times \Gamma$ which is $\mu \times \nu$-measurable. Let $1 \leq p < \infty$. Then

$$
\int_{\Gamma} \left( \int_{\Omega} f(x,y)^p \mu(dx) \right)^{1/p} \nu(dy) 
\geq \left( \int_{\Omega} \left( \int_{\Gamma} f(x,y)^p \nu(dy) \right)^p \mu(dx) \right)^{1/p}
$$

in the sense that the finiteness of the left side implies the finiteness of the right side.

Equality and finiteness in (1) for $1 < p < \infty$ imply the existence of a $\mu$ measurable function $\alpha : \Omega \to \mathbb{R}^+$ and a $\nu$ measurable function $\beta : \Gamma \to \mathbb{R}^+$ such that

$$f(x,y) = \alpha(x)\beta(y) \quad \text{for } \mu \times \nu\text{-almost every } (x,y).$$

A special case of this is the triangle inequality. For $f, g \in L^p(\Omega, d\mu)$ (possibly complex functions)

$$\|f + g\|_p \leq \|f\|_p + \|g\|_p \quad \text{for } 1 \leq p \leq \infty. \quad (2)$$

If $f \neq 0$ and if $1 < p < \infty$, there is equality in (2) if and only if $g = \lambda f$ for some $\lambda \geq 0$.

PROOF. First we note that the two functions

$$\int_{\Omega} f(x,y)^p \mu(dx) \quad \text{and} \quad H(x) := \int_{\Gamma} f(x,y)^p \nu(dy)$$

are measurable functions. This follows from Theorem 1.12 (Fubini's theorem) and the assumption that $f$ is $\mu \times \nu$-measurable. We can assume that $f > 0$ on a set of positive $\mu \times \nu$ measure, for otherwise there is nothing to prove. We can also assume that the right side of (1) is finite; if not we can truncate $f$ so that it is finite and then use a monotone convergence argument to remove the truncation. Sigma-finiteness is again used in this step.
The right side of (1) can be written as follows:

\[ \int_{\Omega} H(x)^p \mu(dx) = \int_{\Omega} \left( \int_{\Gamma} f(x, y) \nu(dy) \right) H(x)^{p-1} \mu(dx) \]

\[ = \int_{\Gamma} \left( \int_{\Omega} f(x, y) H(x)^{p-1} \mu(dx) \right) \nu(dy). \]

The last equation follows by Fubini’s theorem. Using Theorem 2.3 (Hölder’s inequality) on the right side we obtain

\[ \int_{\Omega} H(x)^p \mu(dx) \leq \int_{\Gamma} \left( \int_{\Omega} f(x, y)^p \mu(dx) \right)^{1/p} \]

\[ \times \left( \int_{\Omega} H(x)^{p-1} \mu(dx) \right)^{\frac{p-1}{p}} \nu(dy). \]  \(3\)

Dividing both sides of (3) by

\[ \left( \int_{\Omega} H(x)^p \mu(dx) \right)^{(p-1)/p}, \]

which is neither zero nor infinity (by our assumptions about \(f\)), yields (1).

The equality sign in the use of Hölder’s inequality implies that for \(\nu\)-almost every \(y\) there exists a number \(\lambda(y)\) (i.e., independent of \(x\)) such that

\[ \lambda(y) H(x) = f(x, y) \text{ for } \mu\text{-almost every } x. \]  \(4\)

As mentioned above, \(\Pi\) is \(\mu\)-measurable. To see that \(\lambda\) is \(\nu\)-measurable we note that

\[ \lambda(y) \int_{\Omega} \Pi(x)^p \mu(dx) = \int_{\Omega} f(x, y)^p \mu(dx), \]

and this yields the desired result since the right side is \(\nu\)-measurable (by Fubini’s theorem).

It remains to prove (2). First, by observing that

\[ |f(x) + g(x)| \leq |f(x)| + |g(x)|, \]  \(5\)

the problem is reduced to proving (2) for nonnegative functions. Evidently, (5) implies (2) when \(p = 1\) or \(\infty\), so we can assume \(1 < p < \infty\). We set \(F(x, 1) = |f(x)|, F(x, 2) = |g(x)|\) and let \(\nu\) be the counting measure of the set \(\Gamma = \{1, 2\}\), namely \(\nu(\{1\}) - \nu(\{2\}) - 1\). Then the inequality (2) is seen to be a special case of (1). (Note the use of Fubini’s theorem here.)
Equality in (2) entails the existence of constants \( \lambda_1 \) and \( \lambda_2 \) (independent of \( x \)) such that

\[
|f(x)| = \lambda_1(|f(x)| + |g(x)|) \quad \text{and} \quad |g(x)| = \lambda_2(|f(x)| + |g(x)|).
\] 

(6)

Thus, \( |g(x)| = \lambda f(x)| \) almost everywhere for some constant \( \lambda \). However, equality in (5) means that \( g(x) = \lambda f(x) \) with \( \lambda \) real and nonnegative.

\[\blacksquare\]

- If \( 1 < p < \infty \), then \( L^p(\Omega) \) possesses another geometric structure that has many consequences, among them the characterization of the dual of \( L^p(\Omega) \) (2.14) and, in connection with weak convergence, Mazur’s theorem (2.13). This structure is called \textbf{uniform convexity} and will be described next. The version we give is optimal and is due to [Hanner]; the proof is in [Ball-Carlen-Lieb]. It improves the triangle (or convexity) inequality

\[
\|f + g\|_p \leq \|f\|_p + \|g\|_p.
\]

2.5 \textbf{THEOREM (Hanner’s inequality)}

\[\text{Let} \ f \ \text{and} \ g \ \text{be functions in} \ L^p(\Omega). \ \text{If} \ 1 \leq p \leq 2, \ \text{we have}
\]

\[
\|f + g\|_p^p + \|f - g\|_p^p \geq (\|f\|_p + \|g\|_p)^p + \|f\|_p - \|g\|_p|^p,
\]

(1)

\[
\left(\|f + g\|_p + \|f - g\|_p\right)^p + \|f + g\|_p - \|f - g\|_p|^p \leq 2^p (\|f\|_p^p + \|g\|_p^p).
\]

(2)

If \( 2 \leq p < \infty \), the inequalities are reversed.

\textbf{REMARK.} When \( \|f\|_p = \|g\|_p \), (2) improves the triangle inequality

\[
\|f + g\|_p \leq \|f\|_p + \|g\|_p
\]

because, by convexity of \( t \mapsto |t|^p \), the left side of (2) is not smaller than \( 2\|f + g\|_p^2 \). To be more precise, it is easy to prove (Exercise 4) that the left side of (2) is bounded below for \( 1 \leq p \leq 2 \) and for

\[
\|f - g\|_p \leq \|f - f\|_p + \|f - g\|_p^p
\]

by

\[
2\|f + g\|_p^p + p(\|f\|_p^p + \|g\|_p^p - \|f - g\|_p^p)^2.
\]

The geometric meaning of Theorem 2.5 is explored in Exercise 5.

\textbf{PROOF.} (1) and (2) are identities when \( p = 2 \) ((1) is then called the \textbf{parallelogram identity}) and reduce to the triangle inequality if \( p = 1 \). (2) is derived from (1) by the replacements \( f \to f + g \) and \( g \to f - g \). Thus, we concentrate on proving (1) for \( p \neq 2 \). We can obviously assume that \( R := \|g\|_p/\|f\|_p \leq 1 \) and that \( \|f\|_p = 1 \). For \( 0 \leq r \leq 1 \) define

\[
\alpha(r) = (1 + r)^{p-1} + (1 - r)^{p-1}
\]