L^p spaces

Now that we have reviewed the foundations of measure theory, let us now put it to work to set up the basic theory of one of the fundamental families of function spaces in analysis, namely the L^p *spaces* (also known as *Lebesgue spaces*). These spaces serve as important model examples for the general theory of topological and normed vector spaces, which we will discuss a little bit in this lecture and then in much greater detail in later lectures.

Just as scalar quantities live in the space of real or complex numbers, and vector quantities live in vector spaces, functions $f : X \to \mathbf{C}$ (or other objects closely related to functions, such as measures) live in function spaces. Like other spaces in mathematics (e.g., vector spaces, metric spaces, topological spaces, etc.) a function space V is not just mere sets of objects (in this case, the objects are functions), but they also come with various important structures that allow one to do some useful operations inside these spaces and from one space to another. For example, function spaces tend to have several (though usually not all) of the following types of structures, which are usually related to each other by various compatibility conditions:

- **Vector space structure.** One can often add two functions f, g in a function space V and expect to get another function $f + g$ in that space V; similarly, one can multiply a function f in V by a scalar c and get another function cf in V. Usually, these operations obey the axioms of a vector space, though it is important to caution that the dimension of a function space is typically infinite. (In some cases, the space of scalars is a more complicated ring than the real or complex field, in which case we need the notion of a module rather than a vector space, but we will not use this more general notion in this course.) Virtually all of the function spaces we shall encounter in this course will be vector spaces. Because the field of scalars is real or complex, vector spaces also come with the notion of convexity, which turns out to be crucial in many aspects of analysis. As a consequence (and in marked contrast to algebra or number theory), much of the theory in real analysis does not seem to extend to other fields of scalars (in particular, real analysis fails spectacularly in the finite characteristic setting).

- **Algebra structure.** Sometimes (though not always) we also wish to multiply two functions f, g in V and get another function fg in V; when combined with the vector space structure and assuming some compatibility conditions (e.g., the distributive law), this makes V an algebra. This multiplication operation is often just pointwise multiplication, but there are other important multiplication operations on function spaces too, such as[2] *convolution*.
- **Norm structure.** We often want to distinguish *large* functions in V from *small* ones, especially in analysis, in which small terms in an expression are routinely discarded or deemed to be acceptable errors. One way to do this is to assign a magnitude or norm $\|f\|_V$ to each function that measures its size. Unlike the situation with scalars, where there is basically a single notion of magnitude, functions have a wide variety of useful notions of size, each measuring a different aspect (or combination of aspects) of the function, such as height, width, oscillation, regularity, decay, and so forth. Typically, each such norm gives rise to a separate function space (although sometimes it is useful to consider a single function space with multiple norms on it). We usually require the norm to be compatible with the vector space structure (and algebra structure, if present), for instance by demanding that the *triangle inequality* hold.
- **Metric structure.** We also want to tell whether two functions f, g in a function space V are *near together* or *far apart*. A typical way to do this is to impose a metric $d : V \times V \to \mathbf{R}^+$ on the space V. If both a norm $\|\|_V$ and a vector space structure are available, there is an obvious way to do this: define the distance between two functions f, g in V to be[3] $d(f,g) := \|f - g\|_V$. It is often important

[2]One sometimes sees other algebraic structures than multiplication appear in function spaces, such as *commutators* and *derivations*, but again we will not encounter those in this course. Another common algebraic operation for function spaces is conjugation or adjoint, leading to the notion of a **-algebra*.

[3]This will be the only type of metric on function spaces encountered in this course. But there are some non-linear function spaces of importance in non-linear analysis (e.g., spaces of maps from one manifold to another) which have no vector space structure or norm, but still have a metric.

to know if the vector space is complete[4] with respect to the given metric; this allows one to take limits of Cauchy sequences, and (with a norm and vector space structure) sum absolutely convergent series, as well as use some useful results from point set topology such as the *Baire category theorem*; see Section 1.7. All of these operations are of course vital in analysis.

- **Topological structure.** It is often important to know when a sequence (or, occasionally, *nets*) of functions f_n in V *converges* in some sense to a limit f (which, hopefully, is still in V); there are often many distinct modes of convergence (e.g., pointwise convergence, uniform convergence, etc.) that one wishes to carefully distinguish from each other. Also, in order to apply various powerful topological theorems (or to justify various formal operations involving limits, suprema, etc.), it is important to know when certain subsets of V enjoy key topological properties (most notably compactness and connectedness), and to know which operations on V are continuous. For all of this, one needs a topology on V. If one already has a metric, then one of course has a topology generated by the open balls of that metric. But there are many important topologies on function spaces in analysis that do not arise from metrics. We also often require the topology to be compatible with the other structures on the function space; for instance, we usually require the vector space operations of addition and scalar multiplication to be continuous. In some cases, the topology on V extends to some natural superspace W of more general functions that contain V. In such cases, it is often important to know whether V is closed in W, so that limits of sequences in V stay in V.
- **Functional structures.** Since numbers are easier to understand and deal with than functions, it is not surprising that we often study functions f in a function space V by first applying some functional $\lambda : V \to \mathbf{C}$ to V to identify some key numerical quantity $\lambda(f)$ associated to f. Norms $f \mapsto \|f\|_V$ are of course one important example of a functional, integration $f \mapsto \int_X f\ d\mu$ provides another, and evaluation $f \mapsto f(x)$ at a point x provides a third important class. (Note, though, that while evaluation is the fundamental feature of a function in set theory, it is often a quite minor operation in analysis; indeed, in many function spaces, evaluation is not even defined at all, for instance because the functions in the space are only defined almost

[4]Compactness would be an even better property than completeness to have, but function spaces unfortunately tend be non-compact in various rather nasty ways, although there are useful partial substitutes for compactness that are available; see, e.g., Section 1.6 of *Poincaré's Legacies, Vol. I.*

everywhere!) An inner product $\langle , \rangle$ on V (see below) also provides a large family $f \mapsto \langle f, g \rangle$ of useful functionals. It is of particular interest to study functionals that are compatible with the vector space structure (i.e., are linear) and with the topological structure (i.e., are continuous); this will give rise to the important notion of duality on function spaces.

- **Inner product structure.** One often would like to pair a function f in a function space V with another object g (which is often, though not always, another function in the same function space V) and obtain a number $\langle f, g \rangle$, that typically measures the amount of *interaction* or *correlation* between f and g. Typical examples include inner products arising from integration, such as $\langle f, g \rangle := \int_X f\overline{g}\ d\mu$; integration itself can also be viewed as a pairing, $\langle f, \mu \rangle := \int_X f\ d\mu$. Of course, we usually require such inner products to be compatible with the other structures present on the space (e.g., to be compatible with the vector space structure, we usually require the inner product to be *bilinear* or *sesquilinear*). Inner products, when available, are incredibly useful in understanding the metric and norm geometry of a space, due to such fundamental facts as the *Cauchy-Schwarz inequality* and the *parallelogram law*. They also give rise to the important notion of *orthogonality* between functions.

- **Group actions.** We often expect our function spaces to enjoy various symmetries; we might wish to rotate, reflect, translate, modulate, or dilate our functions and expect to preserve most of the structure of the space when doing so. In modern mathematics, symmetries are usually encoded by *group actions* (or actions of other group-like objects, such as semigroups or groupoids; one also often upgrades groups to more structured objects such as Lie groups). As usual, we typically require the group action to preserve the other structures present on the space, e.g., one often restricts attention to group actions that are linear (to preserve the vector space structure), continuous (to preserve topological structure), unitary (to preserve inner product structure), isometric (to preserve metric structure), and so forth. Besides giving us useful symmetries to spend, the presence of such group actions allows one to apply the powerful techniques of representation theory, Fourier analysis, and ergodic theory. However, as this is a foundational real analysis class, we will not discuss these important topics much here (and in fact will not deal with group actions much at all).

- **Order structure.** In some cases, we want to utilise the notion of a function f being *non-negative*, or *dominating* another function

g. One might also want to take the max or *supremum* of two or more functions in a function space V, or split a function into *positive* and *negative* components. Such order structures interact with the other structures on a space in many useful ways (e.g., via the *Stone-Weierstrass theorem*, Theorem 1.10.18). Much like convexity, order structure is specific to the real line and is another reason why much of real analysis breaks down over other fields. (The complex plane is of course an extension of the real line and so is able to exploit the order structure of that line, usually by treating the real and imaginary components separately.)

There are of course many ways to combine various flavours of these structures together, and there are entire subfields of mathematics that are devoted to studying particularly common and useful categories of such combinations (e.g., topological vector spaces, normed vector spaces, Banach spaces, Banach algebras, von Neumann algebras, C^* algebras, Frechet spaces, Hilbert spaces, group algebras, etc.) The study of these sorts of spaces is known collectively as functional analysis. We will study some (but certainly not all) of these combinations in an abstract and general setting later in this course, but to begin with we will focus on the L^p spaces, which are very good model examples for many of the above general classes of spaces, and also of importance in many applications of analysis (such as probability or PDE).

1.3.1. L^p spaces. In this section, $(X, \mathcal{X}, \mu)$ will be a fixed measure space; notions such as "measurable", "measure", "almost everywhere", etc., will always be with respect to this space, unless otherwise specified. Similarly, unless otherwise specified, all subsets of X mentioned are restricted to be measurable, as are all scalar functions on X.

For the sake of concreteness, we shall select the field of scalars to be the complex numbers $\mathbf{C}$. The theory of real Lebesgue spaces is virtually identical to that of complex Lebesgue spaces, and the former can largely be deduced from the latter as a special case.

We already have the notion of an absolutely integrable function on X, which is a function $f : X \to \mathbf{C}$ such that $\int_X |f|\ d\mu$ is finite. More generally, given any[5] exponent $0 < p < \infty$, we can define a pth-power integrable function to be a function $f : X \to \mathbf{C}$ such that $\int_X |f|^p\ d\mu$ is finite.

Remark 1.3.1. One can also extend these notions to functions that take values in the extended complex plane $\mathbf{C} \cup \{\infty\}$, but one easily observes that pth power integrable functions must be finite almost everywhere, and so

[5] Besides $p = 1$, the case of most interest is the case of square-integrable functions, when $p = 2$. We will also extend this notion later to $p = \infty$, which is also an important special case.

there is essentially no increase in generality afforded by extending the range in this manner.

Following the *Lebesgue philosophy* (that one should ignore whatever is going on on a set of measure zero), let us declare two measurable functions to be equivalent if they agree almost everywhere. This is easily checked to be an equivalence relation, which does not affect the property of being pth-power integrable. Thus, we can define the Lebesgue space $L^p(X, \mathcal{X}, \mu)$ to be the space of pth-power integrable functions, quotiented out by this equivalence relation. Thus, strictly speaking, a typical element of $L^p(X, \mathcal{X}, \mu)$ is not actually a specific function f, but is instead an equivalence class $[f]$, consisting of all functions equivalent to a single function f. However, we shall abuse notation and speak loosely of a function f "belonging" to $L^p(X, \mathcal{X}, \mu)$, where it is understood that f is only defined up to equivalence, or more imprecisely is "defined almost everywhere". For the purposes of integration, this equivalence is quite harmless, but this convention does mean that we can no longer evaluate a function f in $L^p(X, \mathcal{X}, \mu)$ at a single point x if that point x has zero measure. It takes a little bit of getting used to the idea of a function that cannot actually be evaluated at any specific point, but with some practice you will find that it will not cause[6] any significant conceptual difficulty.

Exercise 1.3.1. If $(X, \mathcal{X}, \mu)$ is a measure space and $\overline{\mathcal{X}}$ is the completion of $\mathcal{X}$, show that the spaces $L^p(X, \mathcal{X}, \mu)$ and $L^p(X, \overline{\mathcal{X}}, \mu)$ are isomorphic using the obvious candidate for the isomorphism. Because of this, when dealing with L^p spaces, we will usually not be too concerned with whether the underlying measure space is complete.

Remark 1.3.2. Depending on which of the three structures $X, \mathcal{X}, \mu$ of the measure space one wishes to emphasise, the space $L^p(X, \mathcal{X}, \mu)$ is often abbreviated $L^p(X)$, $L^p(\mathcal{X})$, $L^p(X, \mu)$, or even just L^p. Since for this discussion the measure space $(X, \mathcal{X}, \mu)$ will be fixed, we shall usually use the L^p abbreviation in this section. When the space X is discrete (i.e., $\mathcal{X} = 2^X$) and μ is a counting measure, then $L^p(X, \mathcal{X}, \mu)$ is usually abbreviated $\ell^p(X)$ or just ℓ^p (and the almost everywhere equivalence relation trivialises and can thus be completely ignored).

At present, the Lebesgue spaces L^p are just sets. We now begin to place several of the structures mentioned in the introduction to upgrade these sets to richer spaces.

[6] One could also take a more abstract view, dispensing with the set X altogether and defining the Lebesgue space $L^p(\mathcal{X}, \mu)$ on abstract measure spaces $(\mathcal{X}, \mu)$, but we will not do so here. Another way to think about elements of L^p is that they are functions which are *unreliable* on an unknown set of measure zero, but remain *reliable* almost everywhere.

We begin with vector space structure. Fix $0 < p < \infty$, and let $f, g \in L^p$ be two pth-power integrable functions. From the crude pointwise (or more precisely, pointwise almost everywhere) inequality

$$\begin{aligned} |f(x) + g(x)|^p &\leq (2 \max(|f(x)|, |g(x)|))^p \\ &= 2^p \max(|f(x)|^p, |g(x)|^p) \\ &\leq 2^p(|f(x)|^p + |g(x)|^p), \end{aligned} \tag{1.16}$$

we see that the sum of two pth-power integrable functions is also pth-power integrable. It is also easy to see that any scalar multiple of a pth-power integrable function is also pth-power integrable. These operations respect almost everywhere equivalence, and so L^p becomes a (complex) vector space.

Next, we set up the norm structure. If $f \in L^p$, we define the L^p *norm* $\|f\|_{L^p}$ of f to be the number

$$\|f\|_{L^p} := (\int_X |f|^p \, d\mu)^{1/p}. \tag{1.17}$$

This is a finite non-negative number by definition of L^p; in particular, we have the identity

$$\|f^r\|_{L^p} = \|f\|_{L^{pr}}^r \tag{1.18}$$

for all $0 < p, r < \infty$.

The L^p norm has the following three basic properties:

Lemma 1.3.3. *Let* $0 < p < \infty$ *and* $f, g \in L^p$.

(i) Non-degeneracy. $\|f\|_{L^p} = 0$ *if and only if* $f = 0$.

(ii) Homogeneity. $\|cf\|_{L^p} = |c|\|f\|_{L^p}$ *for all complex numbers* c.

(iii) (Quasi-)triangle inequality. *We have* $\|f + g\|_{L^p} \leq C(\|f\|_{L^p} + \|g\|_{L^p})$ *for some constant* C *depending on* p. *If* $p \geq 1$, *then we can take* $C = 1$ (*this fact is also known as* Minkowski's inequality).

Proof. The claims (i) and (ii) are obvious. (Note how important it is that we equate functions that vanish almost everywhere in order to get (i).) The quasi-triangle inequality follows from a variant of the estimates in (1.16) and is left as an exercise. For the triangle inequality, we have to be more efficient than the crude estimate (1.16). By the non-degeneracy property we may take $\|f\|_{L^p}$ and $\|g\|_{L^p}$ to be non-zero. Using homogeneity, we can normalise $\|f\|_{L^p} + \|g\|_{L^p}$ to equal 1, thus (by homogeneity again) we can write $f = (1 - \theta)F$ and $g = \theta G$ for some $0 < \theta < 1$ and $F, G \in L^p$ with $\|F\|_{L^p} = \|G\|_{L^p} = 1$. Our task is now to show that

$$\int_X |(1 - \theta)F(x) + \theta G(x)|^p \, d\mu \leq 1. \tag{1.19}$$

But observe that for $1 \leq p < \infty$, the function $x \mapsto |x|^p$ is convex on $\mathbf{C}$, and in particular that

$$|(1-\theta)F(x) + \theta G(x)|^p \leq (1-\theta)|F(x)|^p + \theta|G(x)|^p. \tag{1.20}$$

(If one wishes, one can use the complex triangle inequality to first reduce to the case when F, G are non-negative, in which case one only needs convexity on $[0, +\infty)$ rather than all of $\mathbf{C}$.) The claim (1.19) then follows from (1.20) and the normalisations of F, G. □

Exercise 1.3.2. Let $0 < p \leq 1$ and $f, g \in L^p$.

(i) Establish the variant $\|f + g\|_{L^p}^p \leq \|f\|_{L^p}^p + \|g\|_{L^p}^p$ of the triangle inequality.

(ii) If furthermore f and g are non-negative (almost everywhere), establish also the reverse triangle inequality $\|f + g\|_{L^p} \geq \|f\|_{L^p} + \|g\|_{L^p}$.

(iii) Show that the best constant C in the quasi-triangle inequality is $2^{\frac{1}{p}-1}$. In particular, the triangle inequality is false for $p < 1$.

(iv) Now suppose instead that $1 < p < \infty$ or $0 < p < 1$. If $f, g \in L^p$ are such that $\|f + g\|_{L^p} = \|f\|_{L^p} + \|g\|_{L^p}$, show that one of the functions f, g is a non-negative scalar multiple of the other (up to equivalence, of course). What happens when $p = 1$?

A vector space V with a function $\|\| : V \to [0, +\infty)$ obeying the non-degeneracy, homogeneity, and (quasi-)triangle inequality is known as a (quasi-)normed vector space, and the function $f \mapsto \|f\|$ is then known as a (quasi-)norm; thus L^p is a normed vector space for $1 \leq p < \infty$ but only a quasi-normed vector space for $0 < p < 1$. A function $\|\| : V \to [0, +\infty)$ obeying the homogeneity and triangle inequality, but not necessarily the non-degeneracy property, is known as a seminorm; thus for instance the L^p norms for $1 \leq p < \infty$ would have been seminorms if we did not equate functions that agreed almost everywhere. (Conversely, given a seminormed vector space $(V, \|\|)$, one can convert it into a normed vector space by quotienting out the subspace $\{f \in V : \|f\| = 0\}$. We leave the details as an exercise for the reader.)

Exercise 1.3.3. Let $\|\| : V \to [0, +\infty)$ be a function on a vector space which obeys the non-degeneracy and homogeneity properties. Show that $\|\|$ is a norm if and only if the closed unit ball $\{x : \|x\| \leq 1\}$ is convex. Show that the same equivalence also holds for the open unit ball. This fact emphasises the geometric nature of the triangle inequality.

Exercise 1.3.4. If $f \in L^p$ for some $0 < p < \infty$, show that the support $\{x \in X : f(x) \neq 0\}$ of f (which is defined only up to sets of measure zero) is a σ-finite set. (Because of this, we can often reduce from the non-σ-finite case to the σ-finite case in many, though not all, questions concerning L^p spaces.)

We now are able to define L^p norms and spaces in the limit $p = \infty$. We say that a function $f : X \to \mathbf{C}$ is *essentially bounded* if there exists an M such that $|f(x)| \leq M$ for almost every $x \in X$, and define $\|f\|_{L^\infty}$ to be the least M that serves as such a bound. We let L^∞ denote the space of essentially bounded functions, quotiented out by equivalence, and given the norm $\|\cdot\|_{L^\infty}$. It is not hard to see that this is also a normed vector space. Observe that a sequence $f_n \in L^\infty$ converges to a limit $f \in L^\infty$ if and only if f_n converges essentially uniformly to f, i.e., it converges uniformly to f outside of a set of measure zero. (Compare with Egorov's theorem (Theorem 1.1.21), which equates pointwise convergence with uniform convergence outside of a set of arbitrarily small measure.)

Now we explain why we call this norm the L^∞ norm:

Example 1.3.4. Let f be a (generalised) step function, thus $f = A1_E$ for some amplitude $A > 0$ and some set E. Let us assume that E has positive finite measure. Then $\|f\|_{L^p} = A\mu(E)^{1/p}$ for all $0 < p < \infty$, and also $\|f\|_{L^\infty} = A$. Thus in this case, at least, the L^∞ norm is the limit of the L^p norms. This example illustrates also that the L^p norms behave like combinations of the *height* A of a function, and the *width* $\mu(E)$ of such a function, though of course the concepts of height and width are not formally defined for functions that are not step functions.

Exercise 1.3.5.
- If $f \in L^\infty \cap L^{p_0}$ for some $0 < p_0 < \infty$, show that $\|f\|_{L^p} \to \|f\|_{L^\infty}$ as $p \to \infty$. (*Hint*: Use the monotone convergence theorem, Theorem 1.1.21.)
- If $f \notin L^\infty$, show that $\|f\|_{L^p} \to \infty$ as $p \to \infty$.

Once one has a vector space structure and a (quasi-)norm structure, we immediately get a (quasi-)metric structure:

Exercise 1.3.6. Let $(V, \|\|)$ be a normed vector space. Show that the function $d : V \times V \to [0, +\infty)$ defined by $d(f, g) := \|f - g\|$ is a metric on V which is translation invariant (thus $d(f + h, g + h) = d(f, g)$ for all $f, g \in V$) and homogeneous (thus $d(cf, cg) = |c|d(f, g)$ for all $f, g \in V$ and scalars c). Conversely, show that every translation-invariant homogeneous metric on V arises from precisely one norm in this manner. Establish a similar claim relating quasi-norms with quasi-metrics (which are defined as metrics, but with the triangle inequality replaced by a quasi-triangle inequality), or between seminorms and semimetrics (which are defined as metrics, but where distinct points are allowed to have a zero separation; these are also known as *pseudometrics*).

The (quasi-)metric structure in turn generates a topological structure in the usual manner using the (quasi-)metric balls as a base for the topology. In particular, a sequence of functions $f_n \in L^p$ converges to a limit $f \in L^p$ if $\|f_n - f\|_{L^p} \to 0$ as $n \to \infty$. We refer to this type of convergence as a convergence in L^p norm or a strong convergence in L^p (we will discuss other modes of convergence in later lectures). As is usual in (quasi-)metric spaces (or more generally for Hausdorff spaces), the limit, if it exists, is unique. (This is however not the case for topological structures induced by seminorms or semimetrics, though we can solve this problem by quotienting out the degenerate elements as discussed earlier.)

Recall that any series $\sum_{n=1}^\infty a_n$ of scalars is convergent if it is absolutely convergent (i.e., if $\sum_{n=1}^\infty |a_n| < \infty$). This fact turns out to be closely related to the fact that the field of scalars $\mathbf{C}$ is complete. This can be seen from the following result:

Exercise 1.3.7. Let $(V, \|\|)$ be a normed vector space (and hence also a metric space and a topological space). Show that the following are equivalent:

- V is a complete metric space (i.e., every Cauchy sequence converges).
- Every sequence $f_n \in V$ which is absolutely convergent (i.e., $\sum_{n=1}^\infty \|f_n\| < \infty$) is also conditionally convergent (i.e., $\sum_{n=1}^N f_n$ converges to a limit as $N \to \infty$).

Remark 1.3.5. The situation is more complicated for complete quasi-normed vector spaces; not every absolutely convergent series is conditionally convergent. On the other hand, if $\|f_n\|$ decays faster than a sufficiently large negative power of n, one recovers conditional convergence; see [**Ta**].

Remark 1.3.6. Let X be a topological space, and let $BC(X)$ be the space of bounded continuous functions on X; this is a vector space. We can place the uniform norm $\|f\|_u := \sup_{x \in X} |f(x)|$ on this space; this makes $BC(X)$ into a normed vector space. It is not hard to verify that this space is complete, and so every absolutely convergent series in $BC(X)$ is conditionally convergent. This fact is better known as the *Weierstrass M-test.*

A space obeying the properties in Exercise 1.3.5 (i.e., a complete normed vector space) is known as a *Banach space.* We will study Banach spaces in more detail later in this course. For now, we give one of the fundamental examples of Banach spaces.

Proposition 1.3.7. *L^p is a Banach space for every $1 \leq p \leq \infty$.*

Proof. By Exercise 1.3.7, it suffices to show that any series $\sum_{n=1}^\infty f_n$ of functions in L^p which is absolutely convergent is also conditionally convergent. This is easy in the case $p = \infty$ and is left as an exercise. In the case

$1 \leq p < \infty$, we write $M := \sum_{n=1}^\infty \|f_n\|_{L^p}$, which is a finite quantity by hypothesis. By the triangle inequality, we have $\| \sum_{n=1}^N |f_n| \|_{L^p} \leq M$ for all N. By monotone convergence (Theorem 1.1.21), we conclude $\| \sum_{n=1}^\infty |f_n| \|_{L^p} \leq M$. In particular, $\sum_{n=1}^\infty f_n(x)$ is absolutely convergent for almost every x. Write the limit of this series as $F(x)$. By dominated convergence (Theorem 1.1.21), we see that $\sum_{n=1}^N f_n(x)$ converges in L^p norm to F, and we are done. □

An important fact is that functions in L^p can be approximated by simple functions:

Proposition 1.3.8. *If $0 < p < \infty$, then the space of simple functions with finite measure support is a dense subspace of L^p.*

Remark 1.3.9. The concept of a non-trivial dense subspace is one which only comes up in infinite dimensions, and it is hard to visualise directly. Very roughly speaking, the infinite number of degrees of freedom in an infinite dimensional space gives a subspace an infinite number of "opportunities" to come as close as one desires to any given point in that space, which is what allows such spaces to be dense.

Proof. The only non-trivial thing to show is the density. An application of the monotone convergence theorem (Theorem 1.1.21) shows that the space of bounded L^p functions are dense in L^p. Another application of monotone convergence (and Exercise 1.3.4) then shows that the space of bounded L^p functions of finite measure support are dense in the space of bounded L^p functions. Finally, by discretising the range of bounded L^p functions, we see that the space of simple functions with finite measure support is dense in the space of bounded L^p functions with finite support. □

Remark 1.3.10. Since not every function in L^p is a simple function with finite measure support, we thus see that the space of simple functions with finite measure support with the L^p norm is an example of a normed vector space which is not complete.

Exercise 1.3.8. Show that the space of simple functions (not necessarily with finite measure support) is a dense subspace of L^∞. Is the same true if one reinstates the finite measure support restriction?

Exercise 1.3.9. Suppose that μ is σ-finite and $\mathcal{X}$ is separable (i.e., countably generated). Show that L^p is separable (i.e., has a countable dense subset) for all $1 \leq p < \infty$. Give a counterexample that shows that L^∞ need not be separable. (*Hint*: Try using a counting measure.)

Next, we turn to algebra properties of L^p spaces. The key fact here is

Proposition 1.3.11 (Hölder's inequality)**.** *Let $f \in L^p$ and $g \in L^q$ for some $0 < p, q \leq \infty$. Then $fg \in L^r$ and $\|fg\|_{L^r} \leq \|f\|_{L^p}\|g\|_{L^q}$, where the exponent r is defined by the formula $\frac{1}{r} = \frac{1}{p} + \frac{1}{q}$.*

Proof. This will be a variant of the proof of the triangle inequality in Lemma 1.3.3, again relying ultimately on convexity. The claim is easy when $p = \infty$ or $q = \infty$ and is left as an exercise for the reader in this case, so we assume $p, q < \infty$. Raising f and g to the power r using (1.17), we may assume $r = 1$, which makes $1 < p, q < \infty$ dual exponents in the sense that $\frac{1}{p} + \frac{1}{q} = 1$. The claim is obvious if either $\|f\|_{L^p}$ or $\|g\|_{L^q}$ are zero, so we may assume they are non-zero; by homogeneity we may then normalise $\|f\|_{L^p} = \|g\|_{L^q} = 1$. Our task is now to show that

$$\int_X |fg| \; d\mu \leq 1. \tag{1.21}$$

Here, we use the convexity of the exponential function $t \mapsto e^t$ on $[0, +\infty)$, which implies the convexity of the function $t \mapsto |f(x)|^{p(1-t)}|g(x)|^{qt}$ for $t \in [0, 1]$ for any x. In particular we have

$$|f(x)g(x)| \leq \frac{1}{p}|f(x)|^p + \frac{1}{q}|g(x)|^q, \tag{1.22}$$

and the claim (1.21) follows from the normalisations on p, q, f, g. □

Remark 1.3.12. For a different proof of this inequality (based on the *tensor power trick*), see Section 1.9 of *Structure and Randomness.*

Remark 1.3.13. One can also use Hölder's inequality to prove the triangle inequality for L^p, $1 \leq p < \infty$ (i.e., *Minkowski's inequality*). From the complex triangle inequality $|f + g| \leq |f| + |g|$, it suffices to check the case when f, g are non-negative. In this case we have the identity

$$\|f + g\|^p_{L^p} = \|f|f + g|^{p-1}\|_{L^1} + \|g|f + g|^{p-1}\|_{L^1}, \tag{1.23}$$

while Hölder's inequality gives $\|f|f + g|^{p-1}\|_{L^1} \leq \|f\|_{L^p}\|f + g\|^{p-1}_{L^p}$ and $\|g|f+g|^{p-1}\|_{L^1} \leq \|g\|_{L^p}\|f+g\|^{p-1}_{L^p}$. The claim then follows from some algebra (and checking the degenerate cases separately, e.g., when $\|f + g\|_{L^p} = 0$).

Remark 1.3.14. The proofs of Hölder's inequality and Minkowski's inequality both relied on convexity of various functions in **C** or $[0, +\infty)$. One way to emphasise this is to deduce both inequalities from *Jensen's inequality*, which is an inequality that manifestly exploits this convexity. We will not take this approach here, but see for instance [**LiLo2000**] for a discussion.

Example 1.3.15. It is instructive to test Hölder's inequality (and also Exercises 1.3.10–1.3.14 below) in the special case when f, g are generalised step functions, say $f = A1_E$ and $g = B1_F$ with A, B non-zero. The inequality then simplifies to

$$\mu(E \cap F)^{1/r} \leq \mu(E)^{1/p}\mu(F)^{1/q}, \tag{1.24}$$

which can be easily deduced from the hypothesis $\frac{1}{p} + \frac{1}{q} = \frac{1}{r}$ and the trivial inequalities $\mu(E \cap F) \leq \mu(E)$ and $\mu(E \cap F) \leq \mu(F)$. One then easily sees (when p, q are finite) that equality in (1.24) only holds if $\mu(E \cap F) = \mu(E) = \mu(F)$, or in other words if E and F agree almost everywhere. Note the above computations also explain why the condition $\frac{1}{p} + \frac{1}{q} = \frac{1}{r}$ is necessary.

Exercise 1.3.10. Let $0 < p, q < \infty$, and let $f \in L^p, g \in L^q$ be such that Hölder's inequality is obeyed with equality. Show that of the functions f^p, g^q, one of them is a scalar multiple of the other (up to equivalence, of course). What happens if p or q is infinite?

An important corollary of Hölder's inequality is the *Cauchy-Schwarz inequality*

$$|\int_X f(x)\overline{g(x)}\ d\mu| \leq \|f\|_{L^2}\|g\|_{L^2}, \tag{1.25}$$

which can of course be proven by many other means.

Exercise 1.3.11. If $f \in L^p$ for some $0 < p \leq \infty$ and is also supported on a set E of finite measure, show that $f \in L^q$ for all $0 < q \leq p$, with $\|f\|_{L^q} \leq \mu(E)^{\frac{1}{q}-\frac{1}{p}}\|f\|_{L^p}$. When does equality occur?

Exercise 1.3.12. If $f \in L^p$ for some $0 < p < \infty$ and every set of positive measure in X has measure at least m, show that $f \in L^q$ for all $p < q \leq \infty$, with $\|f\|_{L^q} \leq m^{\frac{1}{q}-\frac{1}{p}}\|f\|_{L^p}$. When does equality occur? (This result is especially useful for the ℓ^p spaces, in which μ is a counting measure and m can be taken to be 1.)

Exercise 1.3.13. If $f \in L^{p_0} \cap L^{p_1}$ for some $0 < p_0 < p_1 \leq \infty$, show that $f \in L^p$ for all $p_0 \leq p \leq p_1$ and that $\|f\|_{L^p} \leq \|f\|_{L^{p_0}}^{1-\theta}\|f\|_{L^{p_1}}^{\theta}$, where $0 < \theta < 1$ is such that $\frac{1}{p} = \frac{1-\theta}{p_0} + \frac{\theta}{p_1}$. Another way of saying this is that the function $\frac{1}{p} \mapsto \log\|f\|_{L^p}$ is convex. When does equality occur? This convexity is a prototypical example of *interpolation*, about which we shall say more in Section 1.11.

Exercise 1.3.14. If $f \in L^{p_0}$ for some $0 < p_0 \leq \infty$ and its support $E := \{x \in X : f(x) \neq 0\}$ has finite measure, show that $f \in L^p$ for all $0 < p < p_0$ and that $\|f\|_{L^p}^p \to \mu(E)$ as $p \to 0$. (Because of this, the measure of the support of f is sometimes known as the L^0 *norm* of f, or more precisely the L^0 norm raised to the power 0.)

1.3.2. Linear functionals on L^p. Given an exponent $1 \leq p \leq \infty$, define the dual exponent $1 \leq p' \leq \infty$ by the formula $\frac{1}{p} + \frac{1}{p'} = 1$ (thus $p' = p/(p-1)$ for $1 < p < \infty$, while 1 and ∞ are duals of each other). From Hölder's inequality, we see that for any $g \in L^{p'}$, the functional $\lambda_g : L^p \to \mathbf{C}$ defined by

$$\lambda_g(f) := \int_X f\overline{g}\, d\mu \tag{1.26}$$

is well defined on L^p; the functional is also clearly linear. Furthermore, Hölder's inequality also tells us that this functional is continuous.

A deep and important fact about L^p spaces is that, in most cases, the converse is true: the recipe (1.26) is the *only* way to create continuous linear functionals on L^p.

Theorem 1.3.16 (Dual of L^p)**.** *Let $1 \leq p < \infty$, and assume μ is σ-finite. Let $\lambda : L^p \to \mathbf{C}$ be a continuous linear functional. Then there exists a unique $g \in L^{p'}$ such that $\lambda = \lambda_g$.*

This result should be compared with the Radon-Nikodym theorem (Corollary 1.2.5). Both theorems start with an abstract function $\mu : \mathcal{X} \to \mathbf{R}$ or $\lambda : L^p \to \mathbf{C}$, and create a function out of it. Indeed, we shall see shortly that the two theorems are essentially equivalent to each other. We will develop Theorem 1.3.16 further in Section 1.5, once we introduce the notion of a dual space.

To prove Theorem 1.3.16, we first need a simple and useful lemma:

Lemma 1.3.17 (Continuity is equivalent to boundedness for linear operators)**.** *Let $T : X \to Y$ be a linear transformation from one normed vector space $(X, \|\|_X)$ to another $(Y, \|\|_Y)$. Then the following are equivalent:*

(i) *T is continuous.*

(ii) *T is continuous at 0.*

(iii) *There exists a constant C such that $\|Tx\|_Y \leq C\|x\|_X$ for all $x \in X$.*

Proof. It is clear that (i) implies (ii), and that (iii) implies (ii). Next, from linearity we have $Tx = Tx_0 + T(x - x_0)$ for any $x, x_0 \in X$, which (together with the continuity of addition, which follows from the triangle inequality) shows that continuity of T at 0 implies continuity of T at any x_0, so that (ii) implies (i). The only remaining task is to show that (i) implies (iii). By continuity, the inverse image of the unit ball in Y must be an open neighbourhood of 0 in X, thus there exists some radius $r > 0$ such that $\|Tx\|_Y < 1$ whenever $\|x\|_X < r$. The claim then follows (with $C := 1/r$) by homogeneity. (Alternatively, one can deduce (iii) from (ii) by contradiction. If (iii) failed, then there exists a sequence x_n of non-zero elements of X

such that $\|Tx_n\|_Y/\|x_n\|_X$ goes to infinity. By homogeneity, we can arrange matters so that $\|x_n\|_X$ goes to zero, but $\|Tx_n\|_Y$ stays away from zero, thus contradicting continuity at 0.) □

Proof of Theorem 1.3.16. The uniqueness claim is similar to the uniqueness claim in the Radon-Nikodym theorem (Exercise 1.2.2) and is left as an exercise to the reader; the hard part is establishing existence.

Let us first consider the case when μ is finite. The linear functional $\lambda : L^p \to \mathbf{C}$ induces a functional $\nu : \mathcal{X} \to \mathbf{C}$ on sets E by the formula

$$\nu(E) := \lambda(1_E). \tag{1.27}$$

Since λ is linear, ν is finitely additive (and sends the empty set to zero). Also, if $E_1, E_2, \ldots$ are a sequence of disjoint sets, then $1_{\bigcup_{n=1}^N E_n}$ converges in L^p to $1_{\bigcup_{n=1}^\infty E_n}$ as $n \to \infty$ (by the dominated convergence theorem and the finiteness of μ), and thus (by continuity of λ and finite additivity of ν), ν is countably additive as well. Finally, from (1.27) we also see that $\nu(E) = 0$ whenever $\mu(E) = 0$, thus ν is absolutely continuous with respect to μ. Applying the Radon-Nikodym theorem (Corollary 1.2.5) to both the real and imaginary components of ν, we conclude that $\nu = \mu_g$ for some $g \in L^1$. Thus by (1.27) we have

$$\lambda(1_E) = \lambda_g(1_E) \tag{1.28}$$

for all measurable E. By linearity, this implies that λ and λ_g agree on simple functions. Taking uniform limits (using Exercise 1.3.8) and using continuity (and the finite measure of μ), we conclude that λ and λ_g agree on all bounded functions. Taking monotone limits (working on the positive and negative supports of the real and imaginary parts of g separately), we conclude that λ and λ_g agree on all functions in L^p, and in particular that $\int_X f\overline{g}\, d\mu$ is absolutely convergent for all $f \in L^p$.

To finish the theorem in this case, we need to establish that g lies in $L^{p'}$. By taking real and imaginary parts, we may assume without loss of generality that g is real; by splitting into the regions where g is positive and negative, we may assume that g is non-negative.

We already know that $\lambda_g = \lambda$ is a continuous functional from L^p to $\mathbf{C}$. By Lemma 1.3.17, this implies a bound of the form $|\lambda_g(f)| \leq C\|f\|_{L^p}$ for some $C > 0$.

Suppose first that $p > 1$. Heuristically, we would like to test this inequality with $f := g^{p'-1}$, since we formally have $\lambda_g(f) = \|g\|_{L^{p'}}^{p'}$ and $\|f\|_{L^p} = \|g\|_{L^{p'}}^{p'-1}$. (Not coincidentally, this is also the choice that would make Hölder's inequality an equality; see Exercise 1.3.10.) Cancelling the $\|g\|_{L^{p'}}$ factors would then give the desired finiteness of $\|g\|_{L^{p'}}$.

We cannot quite make that argument work, because it is circular: it assumes $\|g\|_{L^{p'}}$ is finite in order to show that $\|g\|_{L^{p'}}$ is finite! But this can be easily remedied. We test the inequality with $f_N := \min(g, N)^{p'-1}$ for some large N; this lies in L^p. We have $\lambda_g(f_N) \geq \|\min(g, N)\|_{L^{p'}}^{p'}$ and $\|f_N\|_{L^p} = \|\min(g, N)\|_{L^{p'}}^{p'-1}$, and hence $\|\min(g, N)\|_{L^{p'}} \leq C$ for all N. Letting N go to infinity and using monotone convergence (Theorem 1.1.21), we obtain the claim.

In the $p = 1$ case, we instead use $f := 1_{g>N}$ as the test functions, to conclude that g is bounded almost everywhere by N. We leave the details to the reader.

This handles the case when μ is finite. When μ is σ-finite, we can write X as the union of an increasing sequence E_n of sets of finite measure. On each such set, the above arguments let us write $\lambda = \lambda_{g_n}$ for some $g_n \in L^{p'}(E_n)$. The uniqueness arguments tell us that the g_n are all compatible with each other, in particular if $n < m$, then g_n and g_m agree on E_n. Thus all the g_n are in fact restrictions of a single function g to E_n. The previous arguments also tell us that the $L^{p'}$ norm of g_n is bounded by the same constant C uniformly in n, so by monotone convergence (Theorem 1.1.21), g has bounded $L^{p'}$ norm also, and we are done. □

Remark 1.3.18. When $1 < p < \infty$, the hypothesis that μ is σ-finite can be dropped, but not when $p = 1$; see, e.g., [**Fo2000**, Section 6.2] for further discussion. In these lectures, though, we will be content with working in the σ-finite setting. On the other hand, the claim fails when $p = \infty$ (except when X is finite); we will see this in Section 1.5, when we discuss the Hahn-Banach theorem.

Remark 1.3.19. We have seen how the Lebesgue-Radon-Nikodym theorem can be used to establish Theorem 1.3.16. The converse is also true: Theorem 1.3.16 can be used to deduce the Lebesgue-Radon-Nikodym theorem (a fact essentially observed by von Neumann). For simplicity, let us restrict our attention to the unsigned finite case, thus μ and m are unsigned and finite. This implies that the sum $\mu+m$ is also unsigned and finite. We observe that the linear functional $\lambda : f \mapsto \int_X f\ d\mu$ is continuous on $L^1(\mu + m)$, hence by Theorem 1.3.16, there must exist a function $g \in L^\infty(\mu + m)$ such that

$$\int_X f\ d\mu = \int_X f\overline{g}\ d(\mu + m) \tag{1.29}$$

for all $f \in L^1(\mu+m)$. It is easy to see that g must be real and non-negative, and also at most 1 almost everywhere. If E is the set where $m = 1$, we see by setting $f = 1_E$ in (1.29) that E has m-measure zero, and so $\mu \lfloor_E$ is singular. Outside of E, we see from (1.29) and some rearrangement that

$$\int_{X \backslash E} (1-g) f \, d\mu = \int_X fg \, dm \tag{1.30}$$

and one then easily verifies that μ agrees with $m_{\frac{g}{1-g}}$ outside of E'. This gives the desired Lebesgue-Radon-Nikodym decomposition $\mu = m_{\frac{g}{1-g}} + \mu \lfloor_E$.

Remark 1.3.20. The argument used in Remark 1.3.19 also shows that the Radon-Nikodym theorem implies the Lebesgue-Radon-Nikodym theorem.

Remark 1.3.21. One can give an alternate proof of Theorem 1.3.16, which relies on the geometry (and in particular, the uniform convexity) of L^p spaces rather than on the Radon-Nikodym theorem, and can thus be viewed as giving an independent proof of that theorem; see Exercise 1.4.14.

Notes. This lecture first appeared at

`terrytao.wordpress.com/2009/01/09`.

Thanks to Xiaochuan Li for corrections.

Hilbert spaces

In the next few lectures, we will be studying four major classes of function spaces. In decreasing order of generality, these classes are the *topological vector spaces*, the *normed vector spaces*, the *Banach spaces*, and the *Hilbert spaces*. In order to motivate the discussion of the more general classes of spaces, we will first focus on the most special class—that of (real and complex) Hilbert spaces. These spaces can be viewed as generalisations of (real and complex) Euclidean spaces such as $\mathbf{R}^n$ and $\mathbf{C}^n$ to infinite-dimensional settings, and indeed much of one's Euclidean geometry intuition concerning lengths, angles, orthogonality, subspaces, etc., will transfer readily to arbitrary Hilbert spaces. In contrast, this intuition is not always accurate in the more general vector spaces mentioned above. In addition to Euclidean spaces, another fundamental example[7] of Hilbert spaces comes from the Lebesgue spaces $L^2(X, \mathcal{X}, \mu)$ of a measure space $(X, \mathcal{X}, \mu)$.

Hilbert spaces are the natural abstract framework in which to study two important (and closely related) concepts, orthogonality and unitarity, allowing us to generalise familiar concepts and facts from Euclidean geometry such as the Cartesian coordinate system, rotations and reflections, and the Pythagorean theorem to Hilbert spaces. (For instance, the Fourier transform (Section 1.12) is a unitary transformation and can thus be viewed as a kind of generalised rotation.) Furthermore, the *Hodge duality* on Euclidean

[7] There are of course many other Hilbert spaces of importance in complex analysis, harmonic analysis, and PDE, such as *Hardy spaces* $\mathcal{H}^2$, *Sobolev spaces* $H^s = W^{s,2}$, and the space HS of *Hilbert-Schmidt operators*; see for instance Section 1.14 for a discussion of Sobolev spaces. Complex Hilbert spaces also play a fundamental role in the foundations of quantum mechanics, being the natural space to hold all the possible states of a quantum system (possibly after projectivising the Hilbert space), but we will not discuss this subject here.

spaces has a partial analogue for Hilbert spaces, namely the *Riesz representation theorem* for Hilbert spaces, which makes the theory of duality and adjoints for Hilbert spaces especially simple (when compared with the more subtle theory of duality for, say, Banach spaces; see Section 1.5).

These notes are only the most basic introduction to the theory of Hilbert spaces. In particular, the theory of linear transformations between two Hilbert spaces, which is perhaps the most important aspect of the subject, is not covered much at all here.

1.4.1. Inner product spaces. The Euclidean norm

$$|(x_1, \dots, x_n)| := \sqrt{x_1^2 + \dots + x_n^2} \tag{1.31}$$

in real Euclidean space $\mathbf{R}^n$ can be expressed in terms of the *dot product* $\cdot : \mathbf{R}^n \times \mathbf{R}^n \to \mathbf{R}$, defined as

$$(x_1, \dots, x_n) \cdot (y_1, \dots, y_n) := x_1 y_1 + \dots + x_n y_n \tag{1.32}$$

by the well-known formula

$$|x| = (x \cdot x)^{1/2}. \tag{1.33}$$

In particular, we have the positivity property

$$x \cdot x \geq 0 \tag{1.34}$$

with equality if and only if $x = 0$. One reason why it is more advantageous to work with the dot product than the norm is that while the norm function is only sublinear, the dot product is *bilinear*, thus

$$(cx + dy) \cdot z = c(x \cdot z) + d(y \cdot z); \quad z \cdot (cx + dy) = c(z \cdot x) + d(z \cdot y) \tag{1.35}$$

for all vectors x, y and scalars c, d, and also symmetric,

$$x \cdot y = y \cdot x. \tag{1.36}$$

These properties make the inner product easier to manipulate algebraically than the norm.

The above discussion was for the real vector space $\mathbf{R}^n$, but one can develop analogous statements for the complex vector space $\mathbf{C}^n$, in which the norm

$$\|(z_1, \dots, z_n)\| := \sqrt{|z_1|^2 + \dots + |z_n|^2} \tag{1.37}$$

can be represented in terms of the complex inner product $\langle , \rangle : \mathbf{C}^n \times \mathbf{C}^n \to \mathbf{C}$ defined by the formula

$$(z_1, \dots, z_n) \cdot (w_1, \dots, w_n) := z_1 \overline{w_1} + \dots + z_n \overline{w_n} \tag{1.38}$$

by the analogue of (1.33), namely

$$\|x\| = (\langle x, x \rangle)^{1/2}. \tag{1.39}$$

In particular, as before with (1.34), we have the positivity property

$$\langle x, x\rangle \geq 0 \tag{1.40}$$

with equality if and only if $x = 0$. The bilinearity property (1.35) is modified to the *sesquilinearity* property

$$\langle cx + dy, z\rangle = c\langle x, z\rangle + d\langle y, z\rangle, \quad \langle z, cx + dy\rangle = \overline{c}\langle z, x\rangle + \overline{d}\langle z, y\rangle \tag{1.41}$$

while the symmetry property (1.36) needs to be replaced with

$$\langle x, y\rangle = \overline{\langle y, x\rangle} \tag{1.42}$$

in order to be compatible with sesquilinearity.

We can formalise all these properties axiomatically as follows.

Definition 1.4.1 (Inner product space)**.** A *complex inner product space* $(V, \langle,\rangle)$ is a complex vector space V, together with an inner product $\langle,\rangle : V \times V \to \mathbf{C}$ which is sesquilinear (i.e., (1.41) holds for all $x, y \in V$ and $c, d \in \mathbf{C}$) and symmetric in the sesquilinear sense (i.e., (1.42) holds for all $x, y \in V$), and obeys the positivity property (1.40) for all $x \in V$, with equality if and only if $x = 0$. We will usually abbreviate $(V, \langle,\rangle)$ as V.

A real inner product space is defined similarly, but with all references to $\mathbf{C}$ replaced by $\mathbf{R}$ (and all references to complex conjugation dropped).

Example 1.4.2. $\mathbf{R}^n$ with the standard dot product (1.32) is a real inner product space, and $\mathbf{C}^n$ with the complex inner product (1.38) is a complex inner product space.

Example 1.4.3. If $(X, \mathcal{X}, \mu)$ is a measure space, then the complex L^2 space $L^2(X, \mathcal{X}, \mu) = L^2(X, \mathcal{X}, \mu; \mathbf{C})$ with the complex inner product

$$\langle f, g\rangle := \int_X f\overline{g}\, d\mu \tag{1.43}$$

(which is well defined by the Cauchy-Schwarz inequality) is easily verified to be a complex inner product space, and similarly for the real L^2 space (with the complex conjugate signs dropped, of course). Note that the finite dimensional examples $\mathbf{R}^n, \mathbf{C}^n$ can be viewed as the special case of the L^2 examples in which X is $\{1, \ldots, n\}$ with the discrete σ-algebra and counting measure.

Example 1.4.4. Any subspace of a (real or complex) inner product space is again a (real or complex) inner product space, simply by restricting the inner product to the subspace.

Example 1.4.5. Also, any real inner product space V can be *complexified* into the complex inner product space $V_{\mathbf{C}}$, defined as the space of formal combinations $x + iy$ of vectors $x, y \in V$ (with the obvious complex vector space structure), and with inner product

$$\langle a + ib, c + id \rangle := \langle a, c\rangle + i\langle b, c\rangle - i\langle a, d\rangle + \langle b, d\rangle. \tag{1.44}$$

Example 1.4.6. Fix a probability space $(X, \mathcal{X}, \mu)$. The space of square-integrable real-valued random variables of mean zero is an inner product space if one uses covariance as the inner product. (What goes wrong if one drops the mean zero assumption?)

Given a (real or complex) inner product space V, we can define the *norm* $\|x\|$ of any vector $x \in V$ by the formula (1.39), which is well defined thanks to the positivity property; in the case of the L^2 spaces, this norm of course corresponds to the usual L^2 norm. We have the following basic facts:

Lemma 1.4.7. *Let V be a real or complex inner product space.*

(i) Cauchy-Schwarz inequality*. For any $x, y \in V$, we have $|\langle x, y\rangle| \leq \|x\|\|y\|$.*
(ii) *The function $x \mapsto \|x\|$ is a norm on V. (Thus every inner product space is a normed vector space.)*

Proof. We shall just verify the complex case, as the real case is similar (and slightly easier). The positivity property tells us that the quadratic form $\langle ax + by, ax + by\rangle$ is non-negative for all complex numbers a, b. Using sesquilinearity and symmetry, we can expand this form as

$$|a|^2\|x\|^2 + 2\operatorname{Re}(a\bar{b}\langle x, y\rangle) + |b|^2\|y\|^2. \tag{1.45}$$

Optimising in a, b (see also Section 1.10 of *Structure and Randomness*), we obtain the Cauchy-Schwarz inequality. To verify the norm property, the only non-trivial verification is that of the triangle inequality $\|x+y\| \leq \|x\| + \|y\|$. But on expanding $\|x + y\|^2 = \langle x + y, x + y\rangle$, we see that

$$\|x + y\|^2 = \|x\|^2 + 2\operatorname{Re}(\langle x, y\rangle) + \|y\|^2, \tag{1.46}$$

and the claim then follows from the Cauchy-Schwarz inequality. □

Observe from the Cauchy-Schwarz inequality that the inner product $\langle , \rangle : H \times H \to \mathbf{C}$ is continuous.

Exercise 1.4.1. Let $T : V \to W$ be a linear map from one (real or complex) inner product space to another. Show that T preserves the inner product structure (i.e., $\langle Tx, Ty\rangle = \langle x, y\rangle$ for all $x, y \in V$) if and only if T is an isometry (i.e., $\|Tx\| = \|x\|$ for all $x \in V$). (*Hint*: In the real case, express $\langle x, y\rangle$ in terms of $\|x+y\|^2$ and $\|x-y\|^2$. In the complex case, use $x+y, x-y, x+iy, x-iy$ instead of $x+y, x-y$.)

Inspired by the above exercise, we say that two inner product spaces are *isomorphic* if there exists an invertible isometry from one space to the other; such invertible isometries are known as *isomorphisms*.

Exercise 1.4.2. Let V be a real or complex inner product space. If $x_1, \ldots,$ x_n are a finite collection of vectors in V, show that the *Gram matrix* $(\langle x_i, x_j\rangle)_{1\leq i,j\leq n}$ is Hermitian and positive semidefinite, and it is positive definite if and only if the $x_1, \ldots, x_n$ are linearly independent. Conversely, given a Hermitian positive semidefinite matrix $(a_{ij})_{1\leq i,j\leq n}$ with real (resp., complex) entries, show that there exists a real (resp., complex) inner product space V and vectors $x_1, \ldots, x_n$ such that $\langle x_i, x_j\rangle = a_{ij}$ for all $1 \leq i, j \leq n$.

In analogy with the Euclidean case, we say that two vectors x, y in a (real or complex) vector space are *orthogonal* if $\langle x, y\rangle = 0$. (With this convention, we see in particular that 0 is orthogonal to every vector, and is the only vector with this property.)

Exercise 1.4.3 (Pythagorean theorem)**.** Let V be a real or complex inner product space. If $x_1, \ldots, x_n$ are a finite set of pairwise orthogonal vectors, then $\|x_1 + \cdots + x_n\|^2 = \|x_1\|^2 + \cdots + \|x_n\|^2$. In particular, we see that $\|x_1 + x_2\| \geq \|x_1\|$ whenever x_2 is orthogonal to x_1.

A (possibly infinite) collection $(e_\alpha)_{\alpha\in A}$ of vectors in a (real or complex) inner product space is said to be *orthonormal* if they are pairwise orthogonal and all of unit length.

Exercise 1.4.4. Let $(e_\alpha)_{\alpha\in A}$ be an orthonormal system of vectors in a real or complex inner product space. Show that this system is (algebraically) linearly independent (thus any non-trivial finite linear combination of vectors in this system is non-zero). If x lies in the algebraic span of this system (i.e., it is a finite linear combination of vectors in the system), establish the *inversion formula*

$$x = \sum_{\alpha\in A} \langle x, e_\alpha\rangle e_\alpha \tag{1.47}$$

(with only finitely many of the terms non-zero) and the (finite) *Plancherel formula*

$$\|x\|^2 = \sum_{\alpha\in A} |\langle x, e_\alpha\rangle|^2. \tag{1.48}$$

Exercise 1.4.5 (Gram-Schmidt theorem)**.** Let $e_1, \ldots, e_n$ be a finite orthonormal system in a real or complex inner product space, and let v be a vector not in the span of $e_1, \ldots, e_n$. Show that there exists a vector e_{n+1} with $\operatorname{span}(e_1, \ldots, e_n, e_{n+1}) = \operatorname{span}(e_1, \ldots, e_n, v)$ such that $e_1, \ldots, e_{n+1}$ is an orthonormal system. Conclude that an n-dimensional real or complex inner product space is isomorphic to $\mathbf{R}^n$ or $\mathbf{C}^n$, respectively. Thus, any statement about inner product spaces which only involves a finite-dimensional subspace of that space can be verified just by checking it on Euclidean spaces.

Exercise 1.4.6 (Parallelogram law)**.** For any inner product space V, establish the *parallelogram law*

$$\|x+y\|^2 + \|x-y\|^2 = 2\|x\|^2 + 2\|y\|^2. \tag{1.49}$$

Show that this inequality fails for $L^p(X, \mathcal{X}, \mu)$ for $p \neq 2$ as soon as X contains at least two disjoint sets of non-empty finite measure. On the other hand, establish the *Hanner inequalities*

$$\|f+g\|_p^p + \|f-g\|_p^p \geq (\|f\|_p + \|g\|_p)^p + |\|f\|_p - \|g\|_p|^p \tag{1.50}$$

and

$$(\|f+g\|_p + \|f-g\|_p)^p + |\|f+g\|_p - \|f-g\|_p|^p \leq 2^p(\|f\|_p^p + \|g\|_p^p) \tag{1.51}$$

for $1 \leq p \leq 2$, with the inequalities being reversed for $2 \leq p < \infty$. (*Hint*: (1.51) can be deduced from (1.50) by a simple substitution. For (1.50), reduce to the case when f, g are non-negative, and then exploit the inequality

$$\begin{aligned} |x+y|^p + |x-y|^p \geq ((1+r)^{p-1} + (1-r)^{p-1})x^p \\ + ((1+r)^{p-1} - (1-r)^{p-1})r^{1-p}y^p \end{aligned} \tag{1.52}$$

for all non-negative x, y, $0 < r < 1$, and $1 \leq p \leq 2$, with the inequality being reversed for $2 \leq p < \infty$, and with equality being attained when $y < x$ and $r = y/x$.)

1.4.2. Hilbert spaces. Thus far, our discussion of inner product spaces has been largely algebraic in nature; this is because we have not been able to take limits inside these spaces and do some actual analysis. This can be rectified by adding an additional axiom:

Definition 1.4.8 (Hilbert spaces)**.** A (real or complex) *Hilbert space* is a (real or complex) inner product space which is complete (or equivalently, an inner product space which is also a Banach space).

Example 1.4.9. From Proposition 1.3.7, (real or complex) $L^2(X, \mathcal{X}, \mu)$ is a Hilbert space for any measure space $(X, \mathcal{X}, \mu)$. In particular, $\mathbf{R}^n$ and $\mathbf{C}^n$ are Hilbert spaces.

Exercise 1.4.7. Show that a subspace of a Hilbert space H will itself be a Hilbert space if and only if it is closed. (In particular, proper dense subspaces of Hilbert spaces are not Hilbert spaces.)

Example 1.4.10. By Example 1.4.9, the space $l^2(\mathbf{Z})$ of doubly infinite square-summable sequences is a Hilbert space. Inside this space, the space $c_c(\mathbf{Z})$ of sequences of finite support is a proper dense subspace (as can be seen for instance by Proposition 1.3.8, though this can also be seen much more directly), and so cannot be a Hilbert space.

Exercise 1.4.8. Let V be an inner product space. Show that there exists a Hilbert space $\overline{V}$ which contains a dense subspace isomorphic to V; we refer to $\overline{V}$ as a *completion* of V. Furthermore, this space is essentially unique in the sense that if $\overline{V}$, $\overline{V}'$ are two such completions, then there exists an isomorphism from $\overline{V}$ to $\overline{V}'$ which is the identity on V (if one identifies V with the dense subspaces of $\overline{V}$ and $\overline{V'}$. Because of this fact, inner product spaces are sometimes known as *pre-Hilbert spaces*, and can always be identified with dense subspaces of actual Hilbert spaces.

Exercise 1.4.9. Let H, H' be two Hilbert spaces. Define the *direct sum* $H \oplus H'$ of the two spaces to be the vector space $H \times H'$ with inner product $\langle (x, x'), (y, y') \rangle_{H \oplus H'} := \langle x, y \rangle_H + \langle x', y' \rangle_{H'}$. Show that $H \oplus H'$ is also a Hilbert space.

Example 1.4.11. If H is a complex Hilbert space, one can define the *complex conjugate* $\overline{H}$ of that space to be the set of formal conjugates $\{\overline{x} : x \in H\}$ of vectors in H, with complex vector space structure $\overline{x} + \overline{y} := \overline{x + y}$ and $c\overline{x} := \overline{\overline{c}x}$, and inner product $\langle \overline{x}, \overline{y} \rangle_{\overline{H}} := \langle y, x \rangle_H$. One easily checks that $\overline{H}$ is again a complex Hilbert space. Note the map $x \mapsto \overline{x}$ is not a complex linear isometry; instead, it is a complex *antilinear* isometry.

A key application of the completeness axiom is to be able to define the *nearest point* from a vector to a closed convex body.

Proposition 1.4.12 (Existence of minimisers)**.** *Let H be a Hilbert space, let K be a non-empty closed convex subset of H, and let x be a point in H. Then there exists a unique y in K that minimises the distance $\|y - x\|$ to x. Furthermore, for any other z in K, we have* $\operatorname{Re}\langle z - y, y - x \rangle \geq 0$.

Recall that a subset K of a real or complex vector space is *convex* if $(1 - t)v + tw \in K$ whenever $v, w \in K$ and $0 \leq t \leq 1$.

Proof. Observe from the parallelogram law (1.49) that we have the (geometrically obvious) fact that if y and y' are distinct and equidistant from x, then their midpoint $(y + y')/2$ is strictly closer to x than either of y or y'. This (and convexity) ensures that the distance minimiser, if it exists, is unique. Also, if y is the distance minimiser and z is in K, then $(1 - \theta)y + \theta z$ is at least as distant from x as y is for any $0 < \theta < 1$, by convexity. Squaring this and rearranging, we conclude that

$$2 \operatorname{Re}\langle z - y, y - x \rangle + \theta \|z - y\|^2 \geq 0. \tag{1.53}$$

Letting $\theta \to 0$ we obtain the final claim in the proposition.

It remains to show existence. Write $D := \inf_{y \in K} \|x - y\|$. It is clear that D is finite and non-negative. If the infimum is attained, then we would be done. We cannot conclude immediately that this is the case, but we can certainly find a sequence $y_n \in K$ such that $\|x - y_n\| \to D$. On the other hand, the midpoints $\frac{y_n + y_m}{2}$ lie in K by convexity and so $\|x - \frac{y_n+y_m}{2}\| \geq D$. Using the parallelogram law (1.49) we deduce that $\|y_n - y_m\| \to 0$ as $n, m \to \infty$, and so y_n is a Cauchy sequence; by completeness, it converges to a limit y, which lies in K since K is closed. From the triangle inequality we see that $\|x - y_n\| \to \|x - y\|$, and thus $\|x - y\| = D$, and so y is a distance minimiser. □

Exercise 1.4.10. Show by constructing counterexamples that the existence of the distance minimiser y can fail if either the closure or convexity hypothesis on K is dropped, or if H is merely an inner product space rather than a Hilbert space. (*Hint*: For the last case, let H be the inner product space $C([0,1]) \subset L^2([0,1])$, and let K be the subspace of continuous functions supported on $[0,1/2]$.) On the other hand, show that existence (but not uniqueness) can be recovered if K is assumed to be compact rather than convex.

Exercise 1.4.11. Using the Hanner inequalities (Exercise 1.4.6), show that Proposition 1.4.12 also holds for the L^p spaces as long as $1 < p < \infty$. (The specific feature of the L^p spaces that is allowing this is known as *uniform convexity*.) Give counterexamples to show that the proposition can fail for L^1 and for L^∞.

Proposition 1.4.12 has some importance in *calculus of variations*, but we will not pursue those applications here.

Since every subspace is necessarily convex, we have a corollary:

Exercise 1.4.12 (Orthogonal projections)**.** Let V be a closed subspace of a Hilbert space H. Then for every $x \in H$ there exists a unique decomposition $x = x_V + x_{V^\perp}$, where $x_V \in V$ and $x_{V^\perp}$ is orthogonal to every element of V. Furthermore, x_V is the closest element of V to x.

Let $\pi_V : H \to V$ be the map $\pi_V : x \mapsto x_V$, where x_V is given by the above exercise; we refer to π_V as the *orthogonal projection* from H onto V. It is not hard to see that π_V is linear, and from the Pythagorean theorem we see that π_V is a contraction (thus $\|\pi_V x\| \leq \|x\|$ for all $x \in V$). In particular, π_V is continuous.

Exercise 1.4.13 (Orthogonal complement)**.** Given a subspace V of a Hilbert space H, define the *orthogonal complement* $V^\perp$ of V to be the set of all vectors in H that are orthogonal to every element of V. Establish the following claims:

- $V^\perp$ is a closed subspace of H, and that $(V^\perp)^\perp$ is the closure of V.
- $V^\perp$ is the trivial subspace $\{0\}$ if and only if V is dense.
- If V is closed, then H is isomorphic to the direct sum of V and $V^\perp$.
- If V, W are two closed subspaces of H, then $(V+W)^\perp = V^\perp \cap W^\perp$ and $(V \cap W)^\perp = \overline{V^\perp + W^\perp}$.

Every vector v in a Hilbert space gives rise to a continuous linear functional $\lambda_v : H \to \mathbf{C}$, defined by the formula $\lambda_v(w) := \langle w, v\rangle$ (the continuity follows from the Cauchy-Schwarz inequality). The *Riesz representation theorem for Hilbert spaces* gives a converse:

Theorem 1.4.13 (Riesz representation theorem for Hilbert spaces)**.** *Let H be a complex Hilbert space, and let $\lambda : H \to \mathbf{C}$ be a continuous linear functional on H. Then there exists a unique v in H such that $\lambda = \lambda_v$. A similar claim holds for real Hilbert spaces (replacing $\mathbf{C}$ by $\mathbf{R}$ throughout).*

Proof. We just show the claim for complex Hilbert spaces, since the claim for real Hilbert spaces is very similar. First, we show uniqueness: if $\lambda_v = \lambda_{v'}$, then $\lambda_{v-v'} = 0$, and in particular $\langle v - v', v - v'\rangle = 0$, and so $v = v'$.

Now we show existence. We may assume that λ is not identically zero, since the claim is obvious otherwise. Observe that the kernel $V := \{x \in H : \lambda(x) = 0\}$ is then a proper subspace of H, which is closed since λ is continuous. By Exercise 1.4.13, the orthogonal complement $V^\perp$ must contain at least one non-trivial vector w, which we can normalise to have unit magnitude. Since w does not lie in V, $\lambda(w)$ is non-zero. Now observe that for any x in H, $x - \frac{\lambda(x)}{\lambda(w)} w$ lies in the kernel of λ, i.e., it lies in V. Taking inner products with w, we conclude that

$$\langle x, w\rangle - \frac{\lambda(x)}{\lambda(w)} = 0, \tag{1.54}$$

and thus

$$\lambda(x) = \langle x, \overline{\lambda(w)} w\rangle. \tag{1.55}$$

Thus we have $\lambda = \lambda_{\overline{\lambda(w)}w}$, and the claim follows. □

Remark 1.4.14. This result gives an alternate proof of the $p = 2$ case of Theorem 1.3.16, and by modifying Remark 1.26, it can be used to give an alternate proof of the Lebesgue-Radon-Nikodym theorem; this proof is due to von Neumann.

Remark 1.4.15. In the next set of notes, when we define the notion of a dual space, we can reinterpret the Riesz representation theorem as providing a canonical isomorphism $H^* \equiv \overline{H}$.

Exercise 1.4.14. Using Exercise 1.4.11, give an alternate proof of the $1 < p < \infty$ case of Theorem 1.3.16.

One important consequence of the Riesz representation theorem is the existence of adjoints:

Exercise 1.4.15 (Existence of adjoints)**.** Let $T : H \to H'$ be a continuous linear transformation. Show that that there exists a unique continuous linear transformation $T^\dagger : H' \to H$ with the property that $\langle Tx, y\rangle = \langle x, T^\dagger y\rangle$ for all $x \in H$ and $y \in H'$. The transformation $T^\dagger$ is called the (Hilbert space) *adjoint* of T; it is of course compatible with the notion of an adjoint matrix from linear algebra.

Exercise 1.4.16. Let $T : H \to H'$ be a continuous linear transformation.

- Show that $(T^\dagger)^\dagger = T$.
- Show that T is an isometry if and only if $T^\dagger\mathrm{T} = \mathrm{id}_H$.
- Show that T is an isomorphism if and only if $T^\dagger\mathrm{T} = \mathrm{id}_H$ and $TT^\dagger = \mathrm{id}_{H'}$.
- If $S : H' \to H''$ is another continuous linear transformation, show that $(ST)^\dagger = T^\dagger S^\dagger$.

Remark 1.4.16. An isomorphism of complex Hilbert spaces is also known as a *unitary transformation*. (For real Hilbert spaces, the term *orthogonal transformation* is used instead.) Note that unitary and orthogonal $n \times n$ matrices generate unitary and orthogonal transformations on $\mathbf{C}^n$ and $\mathbf{R}^n$, respectively.

Exercise 1.4.17. Show that the projection map $\pi_V : H \to V$ from a Hilbert space to a closed subspace is the adjoint of the inclusion map $\iota_V : V \to H$.

1.4.3. Orthonormal bases. In the section on inner product spaces, we studied finite linear combinations of orthonormal systems. Now that we have completeness, we turn to *infinite* linear combinations.

We begin with countable linear combinations:

Exercise 1.4.18. Suppose that $e_1, e_2, e_3, \ldots$ is a countable orthonormal system in a complex Hilbert space H, and $c_1, c_2, \ldots$ is a sequence of complex numbers. (As usual, similar statements will hold here for real Hilbert spaces and real numbers.)

(i) Show that the series $\sum_{n=1}^\infty c_n e_n$ is conditionally convergent in H if and only if c_n is square-summable.

(ii) If c_n is square-summable, show that $\sum_{n=1}^\infty c_n e_n$ is unconditionally convergent in H, i.e., every permutation of the $c_n e_n$ sums to the same value.

(iii) Show that the map $(c_n)_{n=1}^\infty \mapsto \sum_{n=1}^\infty c_n e_n$ is an isometry from the Hilbert space $\ell^2(\mathbf{N})$ to H. The image V of this isometry is the smallest closed subspace of H that contains $e_1, e_2, \ldots$, and which we shall therefore call the (Hilbert space) *span* of $e_1, e_2, \ldots$.

(iv) Take adjoints of (ii) and conclude that for any $x \in H$, we have $\pi_V(x) = \sum_{n=1}^\infty \langle x, e_n \rangle e_n$ and $\|\pi_V(x)\| = (\sum_{n=1}^\infty |\langle x, e_n \rangle|^2)^{1/2}$. Conclude in particular the *Bessel inequality* $\sum_{n=1}^\infty |\langle x, e_n \rangle|^2 \leq \|x\|^2$.

Remark 1.4.17. Note the contrast here between conditional and unconditional summability (which needs only square-summability of the coefficients c_n) and absolute summability (which requires the stronger condition that the c_n are absolutely summable). In particular there exist non-absolutely summable series that are still unconditionally summable, in contrast to the situation for scalars, in which one has the *Riemann rearrangement theorem*.

Now we can handle arbitrary orthonormal systems $(e_\alpha)_{\alpha \in A}$. If $(c_\alpha)_{\alpha \in A}$ is square-summable, then at most countably many of the c_α are non-zero (by Exercise 1.3.4). Using parts (i), (ii) of Exercise 1.4.18, we can then form the sum $\sum_{\alpha \in A} c_\alpha e_\alpha$ in an unambiguous manner. It is not hard to use Exercise 1.4.18 to then conclude that this gives an isometric embedding of $\ell^2(A)$ into H. The image of this isometry is the smallest closed subspace of H that contains the orthonormal system, which we call the (Hilbert space) *span* of that system. (It is the closure of the algebraic span of the system.)

Exercise 1.4.19. Let $(e_\alpha)_{\alpha \in A}$ be an orthonormal system in H. Show that the following statements are equivalent:

(i) The Hilbert space span of $(e_\alpha)_{\alpha \in A}$ is all of H.

(ii) The algebraic span of $(e_\alpha)_{\alpha \in A}$ (i.e., the finite linear combinations of the e_α) is dense in H.

(iii) One has the *Parseval identity* $\|x\|^2 = \sum_{\alpha \in A} |\langle x, e_\alpha \rangle|^2$ for all $x \in H$.

(iv) One has the *inversion formula* $x = \sum_{\alpha \in A} \langle x, e_\alpha \rangle e_\alpha$ for all $x \in H$ (in particular, the coefficients $\langle x, e_\alpha \rangle$ are square-summable).

(v) The only vector that is orthogonal to all the e_α is the zero vector.

(vi) There is an isomorphism from $\ell^2(A)$ to H that maps δ_α to e_α for all $\alpha \in A$ (where δ_α is the *Kronecker delta* at α).

A system $(e_\alpha)_{\alpha \in A}$ obeying any (and hence all) of the properties in Exercise 1.4.19 is known as an *orthonormal basis* of the Hilbert space H. All Hilbert spaces have such a basis:

Proposition 1.4.18. *Every Hilbert space has at least one orthonormal basis.*

Proof. We use the standard Zorn's lemma argument (see Section 2.4). Every Hilbert space has at least one orthonormal system, namely the empty system. We order the orthonormal systems by inclusion, and observe that the union of any totally ordered set of orthonormal systems is again an orthonormal system. By Zorn's lemma, there must exist a maximal orthonormal system $(e_\alpha)_{\alpha \in A}$. There cannot be any unit vector orthogonal to all the elements of this system, since otherwise one could add that vector to the system and contradict orthogonality. Applying Exercise 1.4.19 in the contrapositive, we obtain an orthonormal basis as claimed. □

Exercise 1.4.20. Show that every vector space V has at least one algebraic basis, i.e., a set of basis vectors such that every vector in V can be expressed uniquely as a finite linear combination of basis vectors. (Such bases are also known as *Hamel bases*.)

Corollary 1.4.19. *Every Hilbert space is isomorphic to $\ell^2(A)$ for some set A.*

Exercise 1.4.21. Let A, B be sets. Show that $\ell^2(A)$ and $\ell^2(B)$ are isomorphic iff A and B have the same cardinality. (*Hint*: The case when A or B is finite is easy, so suppose A and B are both infinite. If $\ell^2(A)$ and $\ell^2(B)$ are isomorphic, show that B can be covered by a family of at most countable sets indexed by A, and vice versa. Then apply the *Schröder-Bernstein theorem* (Section 1.13 of *Volume II*).

We can now classify Hilbert spaces up to isomorphism by a single cardinal, the dimension of that space:

Exercise 1.4.22. Show that all orthonormal bases of a given Hilbert space H have the same cardinality. This cardinality is called the (Hilbert space) *dimension* of the Hilbert space.

Exercise 1.4.23. Show that a Hilbert space is *separable* (i.e., has a countable dense subset) if and only if its dimension is at most countable. Conclude in particular that up to isomorphism, there is exactly one separable infinite-dimensional Hilbert space.

Exercise 1.4.24. Let H, H' be complex Hilbert spaces. Show that there exists another Hilbert space $H \otimes H'$, together with a map $\otimes : H \times H' \to H \otimes H'$ with the following properties:

(i) The map $\otimes$ is bilinear, thus $(cx + dy) \otimes x' = c(x \otimes x') + d(y \otimes x')$ and $x \otimes (cx' + dy') = c(x \otimes x') + d(x \otimes y')$ for all $x, y \in H, x', y' \in H', c, d \in \mathbf{C}$.

(ii) We have $\langle x \otimes x', y \otimes y' \rangle_{H \otimes H'} = \langle x, y \rangle_H \langle x', y' \rangle_{H'}$ for all $x, y \in H, x', y' \in H'$.

(iii) The (algebraic) span of $\{x \otimes x' : x \in H, x' \in H'\}$ is dense in $H \otimes H'$.

Furthermore, show that $H \otimes H'$ and $\otimes$ are unique up to isomorphism in the sense that if $H \tilde{\otimes} H'$ and $\tilde{\otimes} : H \times H' \to H \tilde{\otimes} H'$ are another pair of objects obeying the above properties, then there exists an isomorphism $\Phi : H \otimes H' \to H \tilde{\otimes} H'$ such that $x \tilde{\otimes} x' = \Phi(x \otimes x')$ for all $x \in H, x' \in H'$. (*Hint*: To prove existence, create orthonormal bases for H and H' and take formal tensor products of these bases.) The space $H \otimes H'$ is called the (Hilbert space) *tensor product* of H and H', and $x \otimes x'$ is the tensor product of x and x'.

Exercise 1.4.25. Let $(X, \mathcal{X}, \mu)$ and $(Y, \mathcal{Y}, \nu)$ be measure spaces. Show that $L^2(X \times Y, \mathcal{X} \times \mathcal{Y}, \mu \times \nu)$ is the tensor product of $L^2(X, \mathcal{X}, \mu)$ and $L^2(Y, \mathcal{Y}, \mu)$, if one defines the tensor product $f \otimes g$ of $f \in L^2(X, \mathcal{X}, \mu)$ and $g \in L^2(Y, \mathcal{Y}, \mu)$ as $f \otimes g(x, y) := f(x)g(y)$.

We do not yet have enough theory in other areas to give the really useful applications of Hilbert space theory yet, but let us just illustrate a simple one, namely the development of *Fourier series* on the unit circle $\mathbf{R}/\mathbf{Z}$. We can give this space the usual Lebesgue measure (identifying the unit circle with $[0, 1)$, if one wishes), giving rise to the complex Hilbert space $L^2(\mathbf{R}/\mathbf{Z})$. On this space we can form the *characters* $e_n(x) := e^{2\pi i n x}$ for all integers n; one easily verifies that $(e_n)_{n \in \mathbf{Z}}$ is an orthonormal system. We claim that it is in fact an orthonormal basis. By Exercise 1.4.19, it suffices to show that the algebraic span of the e_n, i.e., the space of trigonometric polynomials, is dense in $L^2(\mathbf{R}/\mathbf{Z})$. But[8] from an explicit computation (e.g., using *Fejér kernels*) one can show that the indicator function of any interval can be approximated to arbitrary accuracy in the L^2 norm by trigonometric polynomials, and is thus in the closure of the trigonometric polynomials. By linearity, the same is then true of an indicator function of a finite union of intervals; since Lebesgue measurable sets in $\mathbf{R}/\mathbf{Z}$ can be approximated to arbitrary accuracy by finite unions of intervals, the same is true for indicators of measurable sets. By linearity, the same is true for simple functions, and by density (Proposition 1.3.8) the same is true for arbitrary L^2 functions, and the claim follows.

The Fourier transform $\hat{f} : \mathbf{Z} \to \mathbf{C}$ of a function $f \in L^2(\mathbf{R}/\mathbf{Z})$ is defined as

$$\hat{f}(n) := \langle f, e_n \rangle = \int_0^1 f(x) e^{-2\pi i n x}\, dx. \tag{1.56}$$

[8]One can also use the Stone-Weierstrass theorem here; see Theorem 1.10.18.

From Exercise 1.4.19, we obtain the *Parseval identity*

$$\sum_{n\in\mathbf{Z}} |\hat f(n)|^2 = \int_{\mathbf{R}/\mathbf{Z}} |f(x)|^2\ dx$$

(in particular, $\hat f \in \ell^2(\mathbf{Z})$) and the *inversion formula*

$$f = \sum_{n\in\mathbf{Z}} \hat f(n) e_n,$$

where the right-hand side is unconditionally convergent. Indeed, the Fourier transform $f \mapsto \hat f$ is a unitary transformation between $L^2(\mathbf{R}/\mathbf{Z})$ and $\ell^2(\mathbf{Z})$. (These facts are collectively referred to as *Plancherel's theorem* for the unit circle.) We will develop Fourier analysis on other spaces than the unit circle in Section 1.12.

Remark 1.4.20. Of course, much of the theory here generalises the corresponding theory in finite-dimensional linear algebra; we will continue this theme much later in the course when we turn to the spectral theorem. However, not every aspect of finite-dimensional linear algebra will carry over so easily. For instance, it turns out to be quite difficult to take the determinant or trace of a linear transformation from a Hilbert space to itself in general (unless the transformation is particularly well behaved, e.g., of trace class). The *Jordan normal form* also does not translate to the infinite-dimensional setting, leading to the notorious *invariant subspace problem* in the subject. It is also worth cautioning that while the theory of orthonormal bases in finite-dimensional Euclidean spaces generalises very nicely to the Hilbert space setting, the more general theory of bases in finite dimensions becomes much more subtle in infinite-dimensional Hilbert spaces, unless the basis is "almost orthonormal" in some sense (e.g., if it forms a *frame*).

Notes. This lecture first appeared at

`terrytao.wordpress.com/2009/01/17`.

Thanks to Américo Tavares, S, and Xiaochuan Liu for corrections.

Uhlrich Groh and Dmitriy raised the interesting open problem of whether any closed subset K of H for which distance minimisers to every point x existed and are unique were necessarily convex, thus providing a converse to Proposition 1.4.12. (Sets with this property are known as *Chebyshev sets*.)