Orthonormal Bases in Hilbert Spaces

MA247 Mathematical Excursions

David McCormick (0504432)

February 12, 2007

1 Hilbert Spaces

There are many different function spaces used in analysis; one of the most familiar is the vector space of continuous functions $f: [a, b] \to \mathbb{R}$, commonly denoted $\mathcal{C}([a, b])$. In order to do analysis on this space, or indeed any vector space, we need a concept analogous to that of the absolute value function on $\mathbb{R}$. Virtually all of the deep results of real analysis depend on relatively few simple properties of the absolute value function, chief among which is the triangle inequality. If we can define a sense of "size" on a vector space, and it satisfies properties similar to those needed from the absolute value function, then it is called a *norm*:

Definition 1.1. *A* norm *on a vector space V over $\mathbb{R}$ (or $\mathbb{C}$) is a function $\|\cdot\|: V \times V \to \mathbb{R}$ satisfying*

1. $\|x\| \geq 0$ *for all* $x \in V$*;*
2. $\|x\| = 0$ *if and only if* $x = 0_V$*;*
3. $\|\lambda x\| = |\lambda|\|x\|$ *for all* $\lambda \in \mathbb{R}$ *and* $x \in V$*; and*
4. $\|x + y\| \leq \|x\| + \|y\|$ *for all* $x, y \in V$*.*

Note that by taking $y = -x$ in property 4 and using property 3, we obtain property 1, so we could (strictly speaking) omit property 1. A vector space together with a norm on that vector space is unsurprisingly called a *normed vector space*. The definitions of convergence and continuity carry over to any normed vector space without alteration from the real case, except that we replace $|\cdot|$ by $\|\cdot\|$.

One of the simplest examples of a norm is the Euclidean norm on $\mathbb{R}^n$: given a vector $\mathbf{x} = (x_1, \ldots, x_n) \in \mathbb{R}^n$, we define

$$\|\mathbf{x}\|_2 := \sqrt{\sum_{i=1}^{n} x_i^2}.$$

We can define many norms on $\mathcal{C}([a, b])$; one of the most commonly used is the *supremum norm*, which we will call the L^∞ norm (for reasons which will become apparent later):

Definition 1.2. *On the space $\mathcal{C}([a, b]) := \{f : [a, b] \to \mathbb{R} \mid f$ is continuous$\}$, we define the L^∞ norm as*

$$\|f\|_{L^\infty} := \sup \{|f(x)| \mid x \in [a, b]\}.$$

Using standard results on uniformly convergent sequences of functions, we obtain the following very useful theorem:

Lemma 1.3. *$\mathcal{C}([a, b])$ with the L^∞ norm is complete; that is, every Cauchy sequence in $\mathcal{C}([a, b])$ converges to some function in the L^∞-sense.*

We call a normed vector space which is complete a *Banach space*; both $\mathbb{R}^n$ and $\mathcal{C}([a, b])$ are Banach spaces.

In a sense, a norm is a generalisation of the Euclidean norm in $\mathbb{R}^n$; however, the Euclidean norm is derived from the standard inner product on $\mathbb{R}^n$: given $\mathbf{x} = (x_1, \dots, x_n)$ and $\mathbf{y} = (y_1, \dots, y_n)$ in $\mathbb{R}^n$, we define

$$\langle \mathbf{x}, \mathbf{y} \rangle := \sum_{i=1}^{n} x_i y_i$$

and then $\|\mathbf{x}\|_2 = \sqrt{\langle \mathbf{x}, \mathbf{x} \rangle}$. What happens, then, if we try to generalise the idea of an inner product to any vector space? If we can, then we would then have enough structure to talk about orthogonal vectors in any vector space, which is a very useful concept. In a similar vein to a norm, we distill down the most important properties of the standard inner product on $\mathbb{R}^n$ and require them as properties of any inner product:

Definition 1.4 (taken from Rynne and Youngson [**8**, defn. 3.1])**.** *An* inner product *on a vector space V over $\mathbb{R}$ is a function $\langle \cdot, \cdot \rangle \colon V \times V \to \mathbb{R}$ satisfying*

1. $\langle x, x \rangle \geq 0$ *for all* $x \in V$*;*
2. $\langle x, x \rangle = 0$ *if and only if* $x = 0_V$*;*
3. $\langle x, y \rangle = \langle y, x \rangle$ *for all* $x, y \in V$*;*
4. $\langle \lambda x + \mu y, z \rangle = \lambda \langle x, z \rangle + \mu \langle y, z \rangle$ *for all* $\lambda, \mu \in \mathbb{R}$ *and* $x, y, z \in V$*.*

A vector space together with an inner product on that space is known as an *inner product space.* Given any inner product space, we can define $\|x\| := \sqrt{\langle x, y \rangle}$, and this is indeed a norm on our vector space. To see this, note that properties 1 and 2 for a norm follow directly from properties 1 and 2 respectively of the inner product; for property 3 of a norm, we have

$$\|\lambda x\| = \sqrt{\langle \lambda x, \lambda x \rangle} = \sqrt{\lambda^2} \sqrt{\langle x, x \rangle} = |\lambda| \|x\|.$$

Property 4 of a norm, the triangle inequality, is the only one which gives us trouble. To prove it, we require the Cauchy–Schwarz inequality (sometimes known as the Cauchy–Bunyakovskii–Schwarz inequality):

Theorem 1.5 (Cauchy–Schwarz inequality)**.** *Let V be an inner product space. Then for any $x, y \in V$, $|\langle x, y \rangle| \leq \|x\| \cdot \|y\|$.*

Proof. The following elegant argument is taken from Körner [**5**, lemma 4.2]. If $x = 0$ or $y = 0$, then $\langle x, y \rangle = 0$ and we are done. If not, then

$$\begin{aligned} 0 &\leq \langle \lambda x + y, \lambda x + y \rangle \\ &= \lambda^2 \|x\|^2 + 2\lambda \langle x, y \rangle + \|y\|^2 \\ &= \left(\lambda \|x\| + \frac{\langle x, y \rangle}{\|x\|} \right)^2 + \|y\|^2 - \frac{\langle x, y \rangle^2}{\|x\|^2}. \end{aligned}$$

Setting $\lambda = -\dfrac{\langle x, y \rangle}{\|x\|^2}$ gives

$$0 \leq \|y\|^2 - \frac{\langle x, y \rangle^2}{\|x\|^2},$$

and the result follows by rearranging and taking square roots. □

We can now prove the triangle inequality for an inner product space:

$$\begin{aligned} \|x + y\|^2 &= \langle x + y, x + y \rangle \\ &= \|x\|^2 + 2\langle x, y \rangle + \|y\|^2 \\ &\leq \|x\|^2 + 2\|x\|\|y\| + \|y\|^2 && \text{by theorem 1.5} \\ &= (\|x\| + \|y\|)^2, \end{aligned}$$

and hence $\|x+y\| \leq \|x\|+\|y\|$. So any inner product space is also a normed space; we say that the inner product *induces* a norm on the vector space. If an inner product space is complete with respect to its induced norm, then we call it a *Hilbert space.* Clearly every Hilbert space is a Banach space.

Unfortunately the converse does not hold: in general, given a norm on a vector space, it is impossible to define an inner product for which all the required properties hold. This is because norms derived from inner products have one key property in addition to those in definition 1.1:

Lemma 1.6 (Parallelogram law, taken from Robinson [**6**, lemma 3.11])**.** *Let V be an inner product space with induced norm $\|\cdot\|$. Then*

$$\|x+y\|^2+\|x-y\|^2=2\left(\|x\|^2+\|y\|^2\right)$$

for all $x, y \in V$.

Proof. This is a simple matter of expanding out the expressions:

$$\begin{aligned}\|x+y\|^2+\|x-y\|^2 &= \langle x+y, x+y\rangle+\langle x-y, x-y\rangle \\ &= \langle x, x\rangle+\langle x, y\rangle+\langle y, x\rangle+\langle y, y\rangle+\langle x, x\rangle-\langle x, y\rangle-\langle y, x\rangle+\langle y, y\rangle \\ &= 2\left(\|x\|^2+\|y\|^2\right).\end{aligned}$$

□

From this we can immediately see that $\mathcal{C}([a, b])$ with the L^∞ norm is *not* an inner product space. Consider the functions $f, g\colon [a, b] \rightarrow \mathbb{R}$ given by

$$f(x)=\frac{x-a}{b-a} \qquad g(x)=\left(\frac{x-a}{b-a}\right)^2.$$

Then $\|f\|_{L^\infty}=1$, $\|g\|_{L^\infty}=1$ and so $\|f+g\|_{L^\infty}=2$ and $\|f-g\|_{L^\infty}=\frac{1}{4}$, hence

$$\|f+g\|_{L^\infty}^2+\|f-g\|_{L^\infty}^2=4+\frac{1}{16} \neq 4=2\left(\|f\|_{L^\infty}^2+\|g\|_{L^\infty}^2\right).$$

Hence the parallelogram law does not hold and so the L^∞ norm is not derived from an inner product.

On the other hand, given a norm which *is* derived from an inner product, the reverse of the parallelogram law allows us to reconstruct the inner product:

Lemma 1.7 (Polarisation identity)**.** *Let V be an inner product space over $\mathbb{R}$ with induced norm $\|\cdot\|$. Then*

$$\langle x, y\rangle=\frac{1}{4}\left(\|x+y\|^2-\|x-y\|^2\right).$$

Proof. Again this is a simple matter of expanding out the expressions, starting from the right-hand side:

$$\begin{aligned}\|x+y\|^2-\|x-y\|^2 &= \langle x+y, x+y\rangle-\langle x-y, x-y\rangle \\ &= \langle x, x\rangle+2\langle x, y\rangle+\langle y, y\rangle-\langle x, x\rangle+2\langle x, y\rangle-\langle y, y\rangle \\ &= 4\langle x, y\rangle.\end{aligned}$$

□

So far we have considered vector spaces over $\mathbb{R}$; we now consider vector spaces over $\mathbb{C}$, and genuine differences present themselves immediately. We encounter the problem even in moving from $\mathbb{R}^n$ to $\mathbb{C}^n$: if we try to define the norm of a vector $\mathbf{x}=(x_1, \ldots, x_n) \in \mathbb{C}^n$ as $\sum_{i=1}^n x_i^2$, then we run into a problem: in general x_i is complex, and hence x_i^2 is not necessarily real. If we want the length of a vector in $\mathbb{C}^n$ under the Euclidean norm to be real, then we can use $|x_i|^2=x_i \overline{x_i}$, where $\overline{x}$ denotes the complex conjugate. So in $\mathbb{C}^n$ the Euclidean norm is defined as

$$\|\mathbf{x}\|_2:=\sqrt{\sum_{i=1}^n|x_i|^2},$$

and the inner product on $\mathbb{C}^n$ as

$$\langle\mathbf{x}, \mathbf{y}\rangle:=\sum_{i=1}^n x_i \overline{y_i}.$$

Definition 1.8. *An* inner product *on a vector space V over $\mathbb{C}$ is a function $\langle\cdot,\cdot\rangle\colon V\times V\to\mathbb{C}$ satisfying*

1. $\langle x,x\rangle\geq 0$ *for all* $x\in V$*;*
2. $\langle x,x\rangle = 0$ *if and only if* $x=0_V$*;*
3. $\langle x,y\rangle=\overline{\langle y,x\rangle}$ *for all* $x,y\in V$*;*
4. $\langle\lambda x+\mu y,z\rangle=\lambda\langle x,z\rangle+\mu\langle y,z\rangle$ *for all* $\lambda,\mu\in\mathbb{R}$ *and* $x,y,z\in V$*. (Taken from [**8**].)*

Note that property 3 implies that $\langle x,x\rangle=\overline{\langle x,x\rangle}$ and hence that $\langle x,x\rangle$ is real, and so the requirement that $\langle x,x\rangle\geq 0$ still makes sense. This also implies that we no longer have linearity in the second argument (which we were careful not to state for the real case): we now have conjugate linearity, so

$$\langle x,\lambda y+\mu z\rangle=\overline{\lambda}\langle x,y\rangle+\overline{\mu}\langle x,z\rangle.$$

We check that the Cauchy–Schwarz inequality still holds:

Proof of theorem 1.5, complex case. If $x=0$ or $y=0$, then $\langle x,y\rangle=0$ and we are done. If not, then

$$\begin{aligned}0&\leq\langle\lambda x+y,\lambda x+y\rangle\\&=\lambda\overline{\lambda}\|x\|^2+\lambda\langle x,y\rangle+\overline{\lambda}\langle y,x\rangle+\|y\|^2\end{aligned}$$

Setting $\lambda=-\frac{\overline{\langle x,y\rangle}}{\|x\|^2}$ gives

$$\begin{aligned}0&\leq\frac{-\overline{\langle x,y\rangle}}{\|x\|^2}\cdot\frac{-\langle x,y\rangle}{\|x\|^2}\cdot\|x\|^2-\frac{\overline{\langle x,y\rangle}\langle x,y\rangle}{\|x\|^2}-\frac{\langle x,y\rangle\overline{\langle x,y\rangle}}{\|x\|^2}+\|y\|^2\\\implies 0&\leq\|y\|^2-\frac{|\langle x,y\rangle|^2}{\|x\|^2},\end{aligned}$$

and the result follows by rearranging and taking square roots. □

The parallelogram law still holds, its proof being word-for-word identical, but the polarisation identity not only requires a change of proof, but a change of statement:

Lemma 1.9 (Polarisation identity, complex case)**.** *Let V be an inner product space over $\mathbb{C}$ with induced norm $\|\cdot\|$. Then*

$$\langle x,y\rangle=\frac{1}{4}\left(\|x+y\|^2-\|x-y\|^2+i\|x+iy\|^2-i\|x-iy\|^2\right).$$

Proof. Again this is a simple matter of expanding out the expressions, starting from the right-hand side:

$$\begin{aligned}&\|x+y\|^2-\|x-y\|^2+i\|x+iy\|^2-i\|x-iy\|^2\\&\quad=\langle x+y,x+y\rangle-\langle x-y,x-y\rangle+i\langle x+iy,x+iy\rangle-i\langle x-iy,x-iy\rangle\\&\quad=\langle x,x\rangle+\langle x,y\rangle+\langle y,x\rangle+\langle y,y\rangle-\langle x,x\rangle+\langle x,y\rangle+\langle y,x\rangle-\langle y,y\rangle\\&\qquad+i\left[\langle x,x\rangle-i\langle x,y\rangle+i\langle y,x\rangle+\langle y,y\rangle\right]-i\left[\langle x,x\rangle+i\langle x,y\rangle-i\langle y,x\rangle+\langle y,y\rangle\right]\\&\quad=2\langle x,y\rangle+2\langle y,x\rangle+i(-i)\langle x,y\rangle+i^2\langle y,x\rangle-i^2\langle x,y\rangle-i(-i)\langle y,x\rangle\\&\quad=2\langle x,y\rangle+2\langle y,x\rangle+\langle x,y\rangle-\langle y,x\rangle+\langle x,y\rangle-\langle y,x\rangle\\&\quad=4\langle x,y\rangle.\end{aligned}$$

□

We will consider both real and complex Hilbert spaces; however, to avoid stating things twice, in general we will just state the complex case if the real case can clearly be derived by ignoring any conjugation operations.

Up until now, we have derived the basic theory of Hilbert spaces without actually having any examples to lay our hands on, except for $\mathbb{R}^n$ and $\mathbb{C}^n$ with which we are already familiar. As with Banach spaces, the true power of Hilbert space theory (and indeed, one of the main reasons for its existence) is the application of ideas from finite-dimensional spaces such as $\mathbb{C}^n$ to *infinite-dimensional* spaces, i.e. spaces which cannot be spanned by any finite set of basis vectors.

The Hilbert space known as ℓ^2 arises naturally if we simply consider extending $\mathbb{C}^n$ to have an infinite number of basis vectors, that is to say, it is a space of *sequences* of real or complex numbers. We define an inner product of two sequences $(x_n)_{n=1}^{\infty}$ and $(y_n)_{n=1}^{\infty}$ as

$$\langle x_n, y_n \rangle = \sum_{i=1}^{\infty} x_i \overline{y_i}$$

just as in $\mathbb{C}^n$. This gives the norm of a sequence as

$$\|x_n\|_{\ell^2} = \left(\sum_{i=1}^{\infty} |x_i|^2 \right)^{1/2}$$

But these are *infinite series*, and we have no guarantee in general that these converge. Consider again the sum $\sum_{i=1}^{\infty} x_i \overline{y_i}$; we now seek conditions under which this sum converges absolutely. The sum to n terms gives $\sum_{i=1}^{n} |x_i \overline{y_i}| = \sum_{i=1}^{n} |x_i||y_i|$; using the Cauchy–Schwarz inequality for $\mathbb{R}^n$ we obtain

$$\sum_{i=1}^{n} |x_i \overline{y_i}| = \sum_{i=1}^{n} |x_i||y_i| \leq \left(\sum_{i=1}^{n} |x_i|^2 \right)^{1/2} \left(\sum_{i=1}^{n} |y_i|^2 \right)^{1/2} \leq \left(\sum_{i=1}^{\infty} |x_i|^2 \right)^{1/2} \left(\sum_{i=1}^{\infty} |y_i|^2 \right)^{1/2}.$$

So if the two sums on the right converge, then the inner product of the two sequences will converge absolutely. But the sums on the right are simply the norms of (x_n) and (y_n); so if the norms of the sequences are finite then the inner product is well-defined. Hence we make the following definition:

Definition 1.10. *The sequence space $\ell^2(\mathbb{C})$ is defined as*

$$\ell^2(\mathbb{C}) := \left\{ (x_n)_{n=1}^{\infty} \,\middle|\, x_n \in \mathbb{C}, \sum_{n=1}^{\infty} |x_n|^2 < \infty \right\}.$$

The space of those sequences in $\ell^2(\mathbb{C})$ which are real-valued is denoted $\ell^2(\mathbb{R})$.

With the inner product defined above, $\ell^2(\mathbb{C})$ is indeed an inner product space:

Proposition 1.11. $\langle x_n, y_n \rangle = \sum_{i=1}^{\infty} x_i \overline{y_i}$ *is an inner product on $\ell^2(\mathbb{C})$.*

Proof. We first verify property 3:

$$\langle x_n, y_n \rangle = \sum_{i=1}^{\infty} x_i \overline{y_i} = \sum_{i=1}^{\infty} \overline{y_i} x_i = \overline{\langle y_n, x_n \rangle}.$$

Hence $\langle x_n, x_n \rangle = \sum_{i=1}^{\infty} x_i \overline{x_i} = \sum_{i=1}^{\infty} |x_i|^2$ is real and non-negative, satisfying property 1. Clearly the sequence $0_{\ell^2} := (0, 0, 0, \dots)$ has $\langle 0_{\ell^2}, 0_{\ell^2} \rangle = 0$, and if any other sequence (x_n) had $\langle x_n, x_n \rangle = 0$, then since $|x_i|^2$ is always nonnegative we must have $x_n = 0$ for all n, proving property 2. Finally we check linearity as in property 4:

$$\begin{aligned} \langle \lambda x + \mu y, z \rangle &= \sum_{i=1}^{\infty} (\lambda x_i + \mu y_i) \overline{z_i} \\ &= \lambda \sum_{i=1}^{\infty} x_i \overline{z_i} + \mu \sum_{i=1}^{\infty} y_i \overline{z_i} \\ &= \lambda \langle x, z \rangle + \mu \langle y, z \rangle. \end{aligned}$$

□

$\ell^2(\mathbb{C})$ is indeed complete with respect to the inner product, and is hence a Hilbert space. The proof is omitted here; it is similar to the proof that $\mathbb{R}^n$ is complete, and can be found in Young [**12**, thm. 3.2].

2 L^2 Space and Lebesgue Integration

We come now to the most important Hilbert space of them all: L^2. Basically, our aim is to turn $\mathcal{C}([a,b])$ into a Hilbert space. As the L^∞-norm is useless for defining an inner product, we look elsewhere.

For sequences, we considered the space ℓ^2 with inner product given by $\langle x_n, y_n\rangle = \sum_{i=1}^{\infty} x_i \overline{y_i}$. This is like evaluating the function $f(n) = x_n$ at each of the points $n \in \mathbb{N}$, and summing them up. If we turn $\mathbb{N}$ into $\mathbb{R}$, we can (heuristically at least) turn the sum into an integral, and so for two functions $f\colon [a,b] \to \mathbb{R}$ we can define the inner product

$$\langle f, g\rangle = \int_a^b f(x)g(x)\,\mathrm{d}x.$$

In the case of functions $f\colon [a,b] \to \mathbb{C}$, we define

$$\langle f, g\rangle = \int_a^b f(x)\overline{g(x)}\,\mathrm{d}x.$$

Here of course the integral is complex; recall that the integral of a function $f\colon [a,b] \to \mathbb{C}$ is simply given by

$$\int_a^b f := \int_a^b \operatorname{Re} f + i \int_a^b \operatorname{Im} f.$$

In either the real or the complex case, the induced norm is

$$\|f\|_{L^2} = \sqrt{\langle f, f\rangle} = \left(\int_a^b |f|^2\,\mathrm{d}x\right)^{1/2}.$$

Assuming for the moment that the above inner product *is* an inner product, we can use the Cauchy–Schwarz inequality to get

$$\int_a^b f(x)\overline{g(x)}\,\mathrm{d}x \le \left(\int_a^b |f|^2\,\mathrm{d}x\right)^{1/2} \left(\int_a^b |g|^2\,\mathrm{d}x\right)^{1/2}.$$

So if $\int_a^b |f|^2$ is finite, then the L^2-norm of the function is well-defined, and hence so is the inner product.

Now, however, we run into trouble; for which functions is this integral defined? This question, and others like it, vexed the mathematicians of the late nineteenth century until Lebesgue came up with a definition of the integral in 1902 that circumvented just about all the problems mathematicians had been having with integration up to this point. In order to make any progress, we need to make use of this *Lebesgue integral*; however, defining it rigorously takes an entire course, so we will describe it heuristically and state some key properties of the Lebesgue integral without proof.

The basic idea of Lebesgue integration is so beautifully simple, that it is somewhat surprising that nobody thought of it earlier. In elementary integration, whether by regulated functions or the Riemann integral, one partitions the x-axis into a number of small intervals, approximates the function over those intervals as a constant, and computes the area of the rectangles thus created; and by taking finer and finer partitions one arrives at a limit. One example, based on the mid-ordinate rule, is shown in figure 1.

Lebesgue's stroke of genius was not to partition the x-axis, but to partition the y-axis. Take an interval in the codomain of the function; find its preimage, find the length of the preimage, and multiply that length by some point in the interval. In essence then, we slice the area horizontally rather than vertically; an approximation is shown in figure 2.

What is actually done is that the integral of a *simple function* is defined first; a simple function has finite range, and so we can take the preimage of each point individually; the integral can thus be computed exactly without having to choose an arbitrary point in the interval. To extend this definition to any function, we simply take a supremum of the integrals of simple functions which are less than or equal to our function, much as in defining the regulated integral you take the limit of the integrals of a sequence of step functions. We denote the Lebesgue integral of a function $f\colon A \to \mathbb{R}$ by $\int_A f\,\mathrm{d}\mu$, the μ signifying Lebesgue measure.

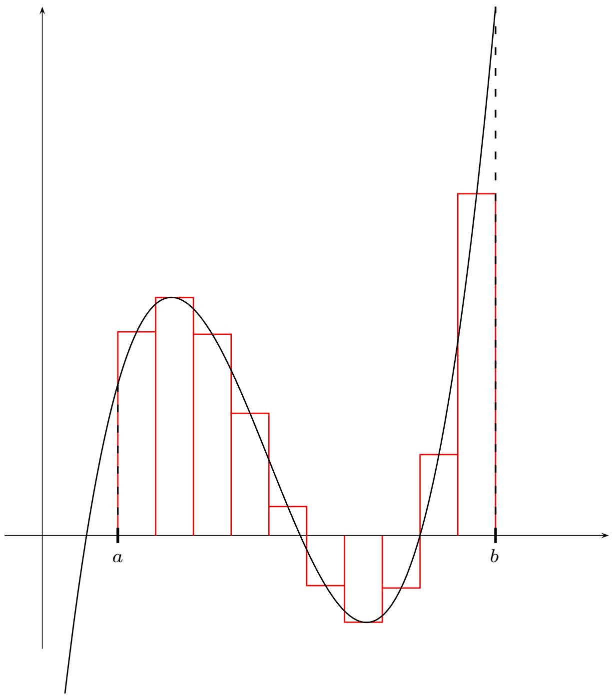

Figure 1: An approximation to the Riemann integral.

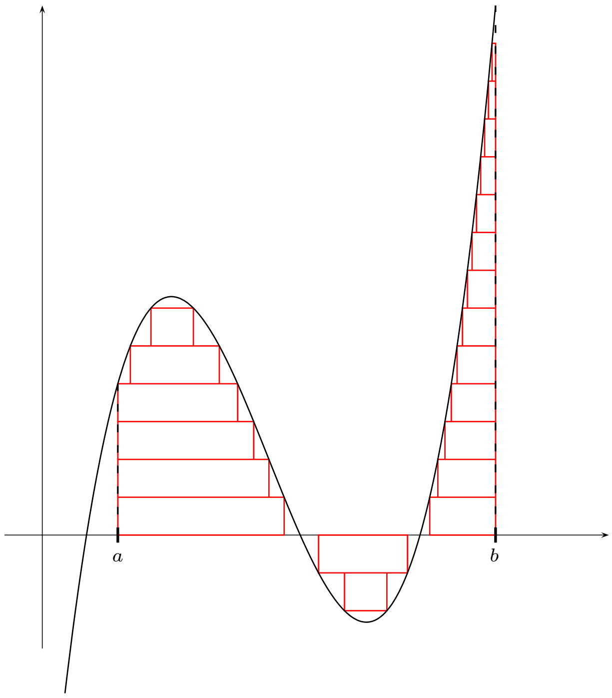

Figure 2: An approximation to the Lebesgue integral.

So the problem of finding the integral of a function is reduced to finding the length of preimages under the function. This is less trivial than it sounds; in principle, the preimages can be extremely intricate sets. Lebesgue came up with the notion of *Lebesgue measure* as a way of assigning a measure or 'length' to more-or-less arbitrary subsets of the real line. Lebesgue laid down three fundamental rules [**2**] in defining the measure of a set:

1. The measure should be translation invariant; that is to say, if the set is shifted along the axis the measure should not change.

2. The measure of an interval $[a, b]$, or indeed (a, b), should be equal to $b - a$.

3. The measure of a *countable* union of pairwise-disjoint sets should be equal to the sum of the measures of the individual sets (this is known as being *countably additive*).

These three rules in fact give a unique measure for a subset of $\mathbb{R}$ [**2**].

Since intervals are somehow nice sets for which we know (or rather have predetermined) their measure, we use these to define the Lebesgue measure of a bounded set X. We cover the set with (at most) countably many open intervals, which are pairwise disjoint, and whose union contains the set X. (We allow some of the intervals to be empty.) This is known as a *countable open cover* [**2**].

The *outer measure* of X is defined as the infimum over all possible countable open covers of X; i.e. the least length of open intervals required to completely cover X. To define the *inner measure*, we take the complement of X in some interval $[a, b]$ containing X, and subtract the outer measure of X^C from the outer measure of $[a, b]$ (which is $b - a$). This is essentially taking the supremum over all countable sets of open intervals contained in X.

If the outer measure and the inner measure of a set X are equal, then we define their common value to be the (Lebesgue) measure $\mu(X)$ of X. The measure thus defined is indeed translation-invariant and countably additive; we call a set measurable if we can assign it a measure as prescribed above.

So far the only sets we have considered are intervals, and we already know their measure. In general, computing the measure of a set is complicated; however, one class of sets plays an important role in the theory not only of integration, but also of Hilbert spaces; that is the class of sets of *measure zero.* If a set has measure zero, this means that, for any $\varepsilon > 0$, we can find a countable collection of open intervals whose length is ε which completely contains the set.

A simple example is the empty set, or even any finite set. In fact, any countable subset of $\mathbb{R}$ must have measure zero; to see this, we enumerate the points with the natural numbers, and place an interval of length $\frac{\varepsilon}{2}$ around the first point, an interval of $\frac{\varepsilon}{4}$ around the second point, an interval of $\frac{\varepsilon}{8}$ around the third point, and so on; this gives a countable collection of open intervals of total length ε. However, it should be emphasised that not all sets of measure zero are countable; for example, the Cantor middle-thirds set is uncountable and has measure zero.

The point of all this is that a set of measure zero will not contribute to an integral; for this reason we call a set of measure zero a *null set.* So if we have two functions f and g defined on some set X, then their integrals will be the same even if they take different values at the points of some subset $N \subseteq X$ which has measure zero. As Capiński and Kopp [**4**] put it, the two functions are equal "for all practical purposes". Formally, we say that f has property P *almost everywhere* or *a.e.* for short, if property P holds for all points of its domain except possibly a null set.

One of the fundamental improvements of the Lebesgue integral, as compared with the Riemann integral, is that we can now integrate over any measurable set: the function need not be defined on an interval. However, we do require that the function be measurable; that is, for any interval in the codomain, the preimage of that interval is a measurable subset of $\mathbb{R}$ (this is the definition given in Capiński and Kopp [**4**, defn. 3.1]).

So far we have not come across any sets which we cannot measure; however, if (and only if) you accept the axiom of choice, then such sets can be constructed. However, virtually all sets you can imagine are indeed measurable, so the restrictions of the Lebesgue integral are hardly restrictions at all.

Given any measurable function $f \colon X \to \mathbb{R}$, we form $f_+ = \max\{f, 0\}$ and $f_- = \max\{-f, 0\}$, and then define $\int_X f \, \mathrm{d}\mu = \int_X f_+ \, \mathrm{d}\mu - \int_X f_- \, \mathrm{d}\mu$ if both the integrals on the right are finite; if so, we call f *integrable.*

We now state without proof some intuitive and useful results from the theory of Lebesgue integration; most of these should be familiar from elementary integration. The following propositions are taken from Rudin [**7**, ch. 11] and Capiński and Kopp [**4**, ch. 4]:

Proposition 2.1 (adapted from Capiński and Kopp [**4**, thm. 4.7] and Rudin [**7**, rem. 11.23])**.** *Suppose f and g are integrable functions. If A is measurable, and $f \leq g$ on A, then $\int_A f \, \mathrm{d}\mu \leq \int_A g \, \mathrm{d}\mu$.*

The following theorem says that null sets are precisely those which are negligible for the purposes of integration:

Theorem 2.2 (taken from Capiński and Kopp [**4**, thm. 4.8])**.** *Suppose $f \colon A \to \mathbb{R}$ is a non-negative measurable function. Then $f = 0$ almost everywhere, if and only if $\int_A f \, \mathrm{d}\mu = 0$.*

This finally gives us an example of a Lebesgue-integrable function which is *not* Riemann integrable: the characteristic function of the rationals:

$$f \colon \mathbb{R} \to \mathbb{R} \qquad f(x) = \begin{cases} 1 & \text{if } x \in \mathbb{Q} \\ 0 & \text{if } x \in \mathbb{R} \setminus \mathbb{Q} \end{cases}$$

Since this is zero except for a null set (since $\mathbb{Q}$ is null in $\mathbb{R}$), the function f is 0 almost everywhere; hence its integral is 0.

Theorem 2.3 (Linearity of the integral, taken from Capiński and Kopp [**4**, thm. 4.19 and prop. 4.20])**.** *For any integrable functions $f, g \colon X \to \mathbb{R}$, and any real numbers α, β, $\alpha f + \beta g$ is also integrable and*

$$\int_X (\alpha f + \beta g) \, \mathrm{d}\mu = \alpha \int_X f \, \mathrm{d}\mu + \beta \int_X g \, \mathrm{d}\mu.$$

It should be emphasised that the Lebesgue integral does not suddenly change the integrals of familiar functions: any function which is regulated is Riemann integrable, and any function which is Riemann integrable is Lebesgue integrable, and the integrals always give the same value if more than one is defined. Therefore the familiar techniques of integration still work, and as such, we retain the notation $\int f(x) \, \mathrm{d}x$ rather than resorting to $\int f \, \mathrm{d}\mu$.

We can now return to the world of Hilbert spaces and finally define L^2 as a function space:

Definition 2.4. *For any measurable set X, the function space $L^2(X, \mathbb{C})$ is defined as*

$$L^2(X, \mathbb{C}) := \left\{ f \colon X \to \mathbb{C} \,\middle|\, \int_X |f(x)|^2 \, \mathrm{d}x < \infty \right\}.$$

The space of real-valued functions in $L^2(X, \mathbb{C})$ is denoted $L^2(X, \mathbb{R})$.

When there is no danger of confusion, we sometimes write $L^2(X)$ for $L^2(X, \mathbb{C})$ or $L^2(X, \mathbb{R})$, as dictated by the context.

It is not immediately obvious that the L^2 space is a vector space. Clearly though, if $|f|^2$ is integrable, then so is $|af|^2$, so it remains to check that it is closed under addition: note that

$$|f + g|^2 = |f|^2 + 2|f||g| + |g|^2 \leq 2|f|^2 + 2|g|^2,$$

and so if $|f|^2$ and $|g|^2$ are integrable it is clear that $|f + g|^2$ is too.

There is a problem with this definition, however: $\int_X |f|^2 \, \mathrm{d}x = 0$ implies that $f = 0$ almost everywhere, but *not* that $f \equiv 0$; we thus need to identify equivalence classes of functions which differ on at most a null set. Once we have done so, the equivalence class containing the zero function is indeed the zero vector in L^2. However, this is cumbersome, and we speak of functions in L^2 when there is no danger of confusion.

With the inner product defined above and functions boxed into equivalence classes, L^2 is indeed an inner product space:

Proposition 2.5. $\langle f, g \rangle = \int_X f \overline{g} \, \mathrm{d}x$ *is an inner product on $L^2(X, \mathbb{C})$.*

Proof. We first verify property 3:

$$\langle f, g\rangle = \int_X f\overline{g}\,\mathrm{d}x = \int_X \overline{g}f\,\mathrm{d}x = \overline{\langle g, f\rangle}.$$

Hence $\langle f, f\rangle = \int_X |f|^2\,\mathrm{d}x$ is real and non-negative, proving property 1. Theorem 2.2 immediatiely yields property 2, so it remains to prove property 4:

$$\begin{aligned}\langle \alpha f + \beta g, h\rangle &= \int_X (\alpha f + \beta g)\overline{h}\,\mathrm{d}x \\ &= \alpha\int_X f\overline{h}\,\mathrm{d}x + \beta\int_X g\overline{h}\,\mathrm{d}x && \text{by theorem 2.3} \\ &= \alpha\langle f, g\rangle + \beta\langle f, h\rangle.\end{aligned}$$

□

So L^2 is clearly an inner product space. Is it a Hilbert space, i.e. is it complete? The answer is yes:

Theorem 2.6. *Let (f_n) be a Cauchy sequence of functions in $L^2(X, \mathbb{C})$. Then (f_n) converges (in the L^2 sense) to some function f in $L^2(X, \mathbb{C})$, and hence $L^2(X, \mathbb{C})$ is complete.*

The proof of this fact is quite involved and involves a number of deep convergence theorems derived from the Lebesgue integral, so we will not reproduce it here; a proof is given in Capiński and Kopp [**4**, thm. 5.24]. However, this is one situation in which the use of the Lebesgue integral is vital: the space of all square-integrable functions would *not* be complete if we used the Riemann integral.

Indeed, this is the reason we do not use the space of continuous functions: it is not complete with respect to the L^2-norm. The L^2 space is in fact the completion with respect to the continuous functions, which basically means that it consists of the continuous functions together with all possible functions obtained by taking limits of continuous functions. To see that this can indeed give us discontinuous functions, consider the following example due to Robinson [**6**, ex. 4.29]:

$$f_k\colon [0,1] \to \mathbb{R} \qquad f_k(x) = \begin{cases} 0 & \text{if } 0 \le x \le \frac{1}{2} - \frac{1}{k} \\ k\left[x - \left(\frac{1}{2} - \frac{1}{k}\right)\right] & \text{if } \frac{1}{2} - \frac{1}{k} \le x \le \frac{1}{2} \\ 1 & \text{if } \frac{1}{2} \le x \le 1 \end{cases}$$

The limit function $f\colon [0,1] \to \mathbb{R}$ is given by

$$f(x) = \begin{cases} 0 & \text{if } 0 \le x < \frac{1}{2} \\ 1 & \text{if } \frac{1}{2} < x \le 1 \end{cases}$$

but this is not continuous; it is not even defined at $\frac{1}{2}$, but whatever value we choose for $f\left(\frac{1}{2}\right)$ we get

$$\begin{aligned}\|f_k - f\|_{L^2}^2 &= \int_0^1 |f_k(x) - f(x)|^2\,\mathrm{d}x \\ &= \int_{\frac{1}{2}-\frac{1}{k}}^{\frac{1}{2}} \left|k\left(x - \tfrac{1}{2} + \tfrac{1}{k}\right)\right|^2\,\mathrm{d}x \\ &= \frac{k^2}{3}\left[x - \tfrac{1}{2} + \tfrac{1}{k}\right]_{x=\frac{1}{2}-\frac{1}{k}}^{\frac{1}{2}} = \frac{1}{3k} \to 0 \text{ as } k \to \infty.\end{aligned}$$

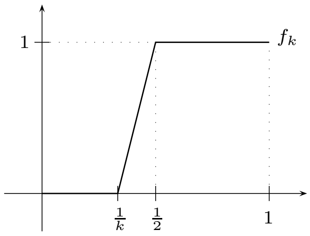

Figure 3: A sequence of continuous functions with a discontinuous limit under the L^2 norm.

3 Orthogonality and Orthonormal Bases

One of the main reasons that inner product spaces, and in particular Hilbert spaces, are so useful is that we can talk about angles between vectors. In $\mathbb{R}^n$ we define the angle θ between two vectors $\mathbf{x}$ and $\mathbf{y}$ to be

$$\theta = \arccos\left(\frac{\langle \mathbf{x}, \mathbf{y}\rangle}{\|\mathbf{x}\|\|\mathbf{y}\|}\right).$$

In general, though, the inner product of two vectors x and y can be complex, and so the angle would be also; this is not particularly helpful to our intuition. However, one special case we can talk about is when the inner product is zero, i.e. the situation when two vectors are *orthogonal*. The definition of orthogonality is exactly that in $\mathbb{R}^n$ or $\mathbb{C}^n$:

Definition 3.1. *Two vectors v, w in an inner product space are called* orthogonal *if $\langle v, w\rangle = 0$.*

One of the most useful concepts in a vector space is that of a basis; that is to say, a set of (linearly independent) vectors from which you can construct every vector in the vector space using linear combinations. There are many bases of $\mathbb{R}^n$, however, the one which is the most useful is the "standard" basis, which consists of the following:

$$\begin{aligned}
\mathbf{e}_1 &= (1, 0, \ldots, 0, 0)\\
\mathbf{e}_2 &= (0, 1, \ldots, 0, 0)\\
&\vdots\\
\mathbf{e}_n &= (0, 0, \ldots, 0, 1)
\end{aligned}$$

The key property that this possesses is that it is *orthonormal.* A set of vectors is orthonormal if each vector has length 1 and is orthogonal to every other vector in the set:

Definition 3.2. *A set of vectors $\{e_k \mid k \in I\}$ in an inner product space (where I is some index set) is* orthonormal *if*

$$\langle e_i, e_j\rangle = \delta_{ij},$$

where δ_{ij}, the Kronecker delta, is 1 if $i = j$ and 0 if $i \neq j$.

$\mathbb{R}^n$ and $\mathbb{C}^n$ are examples of finite-dimensional Hilbert spaces; that is, any basis will have a finite number of basis vectors (in fact, n of them); expressing a vector in $\mathbb{R}^n$ or $\mathbb{C}^n$ as a vector in terms of an orthonormal basis is easy. In infinite-dimensional Hilbert spaces, however, such an expression must involve an infinite sum, and hence issues of convergence. To say that $x = \sum_{i=1}^{\infty} x_i e_i$ for some vectors $\{e_i\}$, we mean that

$$\left\| x - \sum_{i=1}^{n} x_i e_i \right\| \to 0 \text{ as } n \to \infty.$$

Using this we can finally define a basis of a Hilbert space:

Definition 3.3. *A set $\{e_i \mid i \in \mathbb{N}\}$ is a* basis *for a Hilbert space H if every x can be expressed uniquely in the form*

$$x = \sum_{i=1}^{\infty} x_i e_i$$

for some x_i in the field of scalars. If in addition $\{e_i \mid i \in \mathbb{N}\}$ is an orthonormal set, then we refer to it as an orthonormal basis*, or a* complete orthonormal sequence*.*

Note that we are only considering a countable set of basis vectors; if we try and have uncountably many then we run into trouble trying to define $\sum x_i e_i$.

Our aim is now to ascertain what orthonormal bases look like, especially in the space L^2. Answering this is somewhat tricky, so we will first tackle the ℓ^2 space. At this juncture, it should be pointed out that $\ell^2(\mathbb{C})$ is in fact just a special case of L^2, specifically that of $L^2(\mathbb{N}, \mathbb{C})$. (Identifying it this way means that the measure we are integrating with respect to is no longer Lebesgue measure; the "length" of a preimage is now the number of points which map to that point, and the integral becomes a (countably infinite) summation. However, the result that L^2 is complete does not depend on what measure we use.)

Our results so far have suggested that $\ell^2(\mathbb{C})$ is simply an infinite-dimensional analogue of $\mathbb{C}^n$, and the next result seems to confirm our suspicions:

Theorem 3.4. *The set* $\{e_i \mid i \in \mathbb{N}\}$ *where* $e_i = (0, 0, \dots, 0, 1, 0, \dots)$ *(where the* 1 *lies in the* i^{th} *position) forms an orthonormal basis for* $\ell^2(\mathbb{C})$*.*

That the e_i are orthonormal is clear; the meat of the proof is in showing that they form a basis, which seems obvious, but proving it is somewhat tricky. For one, we now allow infinite linear combinations as opposed to only finite linear combinations.

To answer this, we will need some machinery for talking about orthonormal bases. (For much of the rest of this section we follow the exposition of Robinson [**6**, §5.1].) For instance, we would like some way to determine the coefficients x_i. Given a basis $\{e_i \mid 1 \leq i \leq n\}$ in $\mathbb{C}^n$, we can write

$$\mathbf{x} = \sum_{i=1}^{n} x_i e_i = \sum_{i=1}^{n} \langle x, e_i \rangle e_i,$$

that is, the vector is equal to the sum of the projections onto each of the basis vectors. Suppose we try the same in an infinite-dimensional Hilbert space with an orthonormal basis $\{e_i \mid i \in N\}$; when can we say

$$x = \sum_{i=1}^{\infty} \langle x, e_i \rangle e_i?$$

We now must consider issues of convergence; when does the right-hand side of this expression actually make sense, i.e. when does it converge? This question is answered in the next sequence of propositions, beginning with Bessel's inequality:

Proposition 3.5 (Bessel's inequality)**.** *If* $\{e_i \mid i \in \mathbb{N}\}$ *is an orthonormal set in an inner product space* V*, then for any* $x \in V$ *we have*

$$\sum_{i=1}^{\infty} |\langle x, e_i \rangle|^2 \leq \|x\|^2.$$

Proof. The proposition and its proof are taken from Robinson [**6**, cor. 5.11]. Let $x_n := \sum_{i=1}^{n} \langle x, e_i \rangle e_i$ be the n^{th} partial sum. Pythagoras' Theorem states that the square of the length of a vector in $\mathbb{R}^n$ is equal to the sum of the squares of the projections onto each vector of an orthonormal basis, i.e.

$$\|x_n\|^2 = \sum_{i=1}^{n} |\langle x, e_i \rangle|^2.$$

Now consider

$$\begin{aligned}
\|x - x_n\|^2 &= \langle x - x_n, x - x_n \rangle = \|x\|^2 - \langle x_n, x \rangle - \langle x, x_n \rangle + \|x_n\|^2 \\
&= \|x\|^2 - \langle \sum_{i=1}^{n} \langle x, e_i \rangle e_i, x \rangle - \langle x, \sum_{i=1}^{n} \langle x, e_i \rangle e_i \rangle \\
&= \|x\|^2 - \sum_{i=1}^{n} \langle x, e_i \rangle \langle e_i, x \rangle - \sum_{i=1}^{n} \overline{\langle x, e_i \rangle} \langle x, e_i \rangle + \|x_n\|^2 \\
&= \|x\|^2 - \|x_n\|^2.
\end{aligned}$$

Hence

$$\sum_{i=1}^{n} |\langle x, e_i \rangle|^2 = \|x_n\|^2 = \|x\|^2 - \|x - x_n\|^2 \leq \|x\|^2. \qquad \square$$

We use this to say when a sum $\sum_{i=1}^{\infty} \alpha_i e_i$ converges whenever $\{e_i \mid i \in \mathbb{N}\}$ is an orthonormal set:

Lemma 3.6. *Let* H *be a Hilbert space and* $\{e_i \mid i \in \mathbb{N}\}$ *an orthonormal set in* H*. Then the series* $\sum_{i=1}^{\infty} \alpha_i e_i$ *converges if and only if* $\sum_{i=1}^{\infty} |\alpha_i|^2 < \infty$*.*

Proof. This proof is adapted from Young [**12**, thm. 4.11] and Robinson [**6**, lemma 5.12].

($\Longrightarrow$) Suppose $\sum_{i=1}^{\infty} \alpha_i e_i = x$. For $k \leq n$ in $\mathbb{N}$, we have $\langle \sum_{i=1}^{n} \alpha_i e_i, e_k \rangle = \sum_{i=1}^{n} \alpha_i \langle e_i, e_k \rangle = \alpha_k$. On letting $n \to \infty$ and using the continuity of the inner product we obtain $\langle x, e_k \rangle = \lim_{n\to\infty} \alpha_k = \alpha_k$. Hence, by Bessel's inequality,

$$\sum_{k=1}^{\infty} |\alpha_k|^2 = \sum_{k=1}^{\infty} |\langle x, e_k \rangle|^2 \leq \|x\|^2 < \infty.$$

($\Longleftarrow$) Set $x_n = \sum_{i=1}^{n} \alpha_i e_i$. By Pythagoras' Theorem, for $n > m$ in $\mathbb{N}$,

$$\|x_n - x_m\|^2 = \left\| \sum_{i=m+1}^{n} \alpha_i e_i \right\|^2 = \sum_{i=m+1}^{n} |\alpha_i|^2.$$

Now as $\sum_{i=1}^{\infty} |\alpha_i|^2 < \infty$, the sequence of partial sums is Cauchy and so we can make the right-hand side arbitrarily small. Hence (x_n) is a Cauchy sequence in H and as H is a Hilbert space it converges to some point of H. $\square$

This lemma has the following immediate corollary:

Corollary 3.7. *Let H be a Hilbert space and $\{e_i \mid i \in \mathbb{N}\}$ an orthonormal set in H. Then for any $x \in H$, $\sum_{i=1}^{\infty} \langle x, e_i \rangle e_i$ converges.*

This leads us to the following proposition:

Proposition 3.8. *Let $\{e_i \mid i \in \mathbb{N}\}$ be an orthonormal set in a Hilbert space H. Then the following are equivalent:*

1. *$\{e_i \mid i \in \mathbb{N}\}$ is an orthonormal basis for H;*
2. *$x = \sum_{i=1}^{\infty} \langle x, e_i \rangle e_i$ for all $x \in H$;*
3. *$\|x\|^2 = \sum_{i=1}^{\infty} |\langle x, e_i \rangle|^2$ for all $x \in H$; and*
4. *$\langle x, e_i \rangle = 0$ for all i implies that $x = 0$.*

Proof. This proposition and its proof are taken from Robinson [**6**, prop. 5.14].

- 1 $\iff$ 2: If $\{e_i \mid i \in \mathbb{N}\}$ is an orthonormal basis for H then we can write

$$x = \sum_{i=1}^{\infty} \alpha_i e_i, \quad \text{i.e.} \quad x = \lim_{n\to\infty} \sum_{i=1}^{n} \alpha_i e_i.$$

  As in lemma 3.6, if $k \leq n$ we have $\langle \sum_{i=1}^{n} \alpha_i e_i, e_k \rangle = \alpha_k$, and so by continuity of the inner product we obtain $\langle x, e_k \rangle = \alpha_k$, and hence 2 holds. The same argument shows that if we assume 2, then this expansion is unique and so $\{e_i \mid i \in \mathbb{N}\}$ is a basis.

- 2 $\implies$ 3: We have

$$\|x\|^2 = \left\| \sum_{i=1}^{\infty} \langle x, e_i \rangle e_i \right\|^2 = \sum_{i=1}^{\infty} |\langle x, e_i \rangle|^2$$

  by Pythagoras' theorem and continuity of the inner product.

- 3 $\implies$ 4: If $\langle x, e_i \rangle = 0$ for all i then $\|x\|^2 = \sum_{i=1}^{\infty} |\langle x, e_i \rangle|^2 = 0$ which implies that $x = 0$.

- 4 $\implies$ 2: Take $x \in H$ and let $y = x - \sum_{i=1}^{\infty} \langle x, e_i \rangle e_i$. For each natural number k we have

$$\langle y, e_k \rangle = \langle x, e_k \rangle - \lim_{n\to\infty} \langle \sum_{i=1}^{n} \langle x, e_i \rangle e_i, e_k \rangle = 0$$

  since eventually $n \geq k$. It follows from 4 that $y = 0$, and hence that $x = \sum_{i=1}^{\infty} \langle x, e_i \rangle e_i$. $\square$

The last of the equivalent properties encapsulates the basic thought that if we have a vector which is orthogonal to all the e_i, then either $\{e_i \mid i \in \mathbb{N}\}$ is not complete, i.e. it does not form a basis, or it must be zero. This gives us a simple proof of theorem 3.4:

Proof of Theorem 3.4. Let $x = (x_n)_{n=1}^{\infty}$ be a sequence in ℓ^2. If $\langle x, e_i \rangle = 0$ for all i, then $x_i = 0$ for all i, and so clearly $x = 0$. Hence by proposition 3.8, $\{e_i \mid i \in \mathbb{N}\}$ form a basis of ℓ^2. □

4 The Fourier Basis of $L^2([-\pi, \pi])$

We come now to the fundamental question posed by this essay: *what do orthonormal bases look like in the L^2 space?* We first consider the basis of $L^2([-\pi, \pi])$, although we could of course consider any closed interval since the functions obtained will then simply be scaled and transposed; we initially only consider real-valued functions and denote the space by $L^2_{\mathbb{R}}([-\pi, \pi])$

Our first answer is best served by historical considerations of the origin of L^2, which came into being as a tool for talking about *Fourier series*. Fourier series arise mainly in the study of partial differential equations, and seek to represent a function as an infinite series of trigonometric functions, as follows:

$$f(x) = \frac{a_0}{2} + \sum_{n=1}^{\infty} (a_n \cos nx + b_n \sin nx)$$

This is an infinite series of continuous functions, so each of the functions

$$1, \cos x, \sin x, \cos 2x, \sin 2x, \ldots$$

belong to $L^2_{\mathbb{R}}([-\pi, \pi])$. What happens if we compute the inner product of these functions? We do so now:

$$\begin{aligned}
\langle 1, \cos nx \rangle &= \int_{-\pi}^{\pi} 1 \cdot \cos nx \, \mathrm{d}x = \left[\frac{1}{n} \sin nx \right]_{-\pi}^{\pi} = 0 \\
\langle 1, \sin nx \rangle &= \int_{-\pi}^{\pi} 1 \cdot \sin nx \, \mathrm{d}x = \left[-\frac{1}{n} \cos nx \right]_{-\pi}^{\pi} = 0 \\
\langle \cos nx, \cos mx \rangle &= \int_{-\pi}^{\pi} \cos nx \cos mx \, \mathrm{d}x \\
&= \int_{-\pi}^{\pi} \frac{1}{2} \left[\cos(n-m)x + \cos(n+m)x \right] \mathrm{d}x \\
&= \frac{1}{2} \left[\frac{1}{n-m} \sin(n-m)x + \frac{1}{n+m} \sin(n+m)x \right]_{-\pi}^{\pi} = 0 \quad \text{if } n \neq m \\
\langle \sin nx, \sin mx \rangle &= \int_{-\pi}^{\pi} \sin nx \sin mx \, \mathrm{d}x \\
&= \int_{-\pi}^{\pi} \frac{1}{2} \left[\cos(n-m)x - \cos(n+m)x \right] \mathrm{d}x \\
&= \frac{1}{2} \left[\frac{1}{n-m} \sin(n-m)x - \frac{1}{n+m} \sin(n+m)x \right]_{-\pi}^{\pi} = 0 \quad \text{if } n \neq m \\
\langle \sin nx, \cos mx \rangle &= \int_{-\pi}^{\pi} \sin nx \cos mx \, \mathrm{d}x \\
&= \int_{-\pi}^{\pi} \frac{1}{2} \left[\sin(n-m)x + \sin(n+m)x \right] \mathrm{d}x \\
&= -\frac{1}{2} \left[\frac{1}{n-m} \cos(n-m)x + \frac{1}{n+m} \cos(n+m)x \right]_{-\pi}^{\pi} = 0 \quad \text{if } n \neq m \\
\langle \sin nx, \cos nx \rangle &= \int_{-\pi}^{\pi} \sin nx \cos nx \, \mathrm{d}x = \int_{-\pi}^{\pi} \frac{1}{2} \left[\sin 2nx \right] \mathrm{d}x = -\frac{1}{2} \left[\frac{1}{2n} \cos 2nx \right]_{-\pi}^{\pi} = 0
\end{aligned}$$

So the functions 1, $\cos x$, $\sin x$, $\cos 2x$, $\sin 2x$, ... are orthogonal with respect to the L^2 inner product! They are not, however, orthonormal, since:

$$\begin{aligned}
\|1\|_{L^2}^2 &= \langle 1, 1\rangle = \int_{-\pi}^{\pi} \mathrm{d}x = 2\pi \\
\|\cos nx\|_{L^2}^2 &= \langle \cos nx, \cos nx\rangle = \int_{-\pi}^{\pi} \cos^2 nx \,\mathrm{d}x \\
&= \int_{-\pi}^{\pi} \frac{1}{2}\left[1 + \cos 2nx\right] \,\mathrm{d}x = \frac{1}{2}\left[x + \frac{1}{2n}\sin 2nx\right]_{-\pi}^{\pi} = \pi \\
\|\sin nx\|_{L^2}^2 &= \langle \sin nx, \sin nx\rangle = \int_{-\pi}^{\pi} \sin^2 nx \,\mathrm{d}x \\
&= \int_{-\pi}^{\pi} \frac{1}{2}\left[1 - \cos 2nx\right] \,\mathrm{d}x = \frac{1}{2}\left[x - \frac{1}{2n}\sin 2nx\right]_{-\pi}^{\pi} = \pi
\end{aligned}$$

Normalising them appropriately, we get the orthonormal sequence of functions:

$$\frac{1}{\sqrt{2\pi}}, \frac{1}{\sqrt{\pi}}\cos x, \frac{1}{\sqrt{\pi}}\sin x, \frac{1}{\sqrt{\pi}}\cos 2x, \frac{1}{\sqrt{\pi}}\sin 2x, \ldots$$

If this sequence were in fact an orthonormal basis, then we would be able to say that every function in $L^2_{\mathbb{R}}([-\pi, \pi])$ has an expression in terms of that basis, i.e. a Fourier series, which converges to the function in question. Analysts of the nineteenth century searched long and hard for such a condition, and this is their holy grail:

Theorem 4.1. *The sequence*

$$\frac{1}{\sqrt{2\pi}}, \frac{1}{\sqrt{\pi}}\cos x, \frac{1}{\sqrt{\pi}}\sin x, \frac{1}{\sqrt{\pi}}\cos 2x, \frac{1}{\sqrt{\pi}}\sin 2x, \ldots$$

of functions in $L^2_{\mathbb{R}}([-\pi, \pi])$ form an orthonormal basis of that space.

We use condition 4 of proposition 3.8: we suppose that there is a function orthogonal to every member of this basis, and show that it must be identically zero.

Lemma 4.2. *If f is continuous on $[-\pi, \pi]$, and f is orthogonal to each of the functions*

$$\frac{1}{\sqrt{2\pi}}, \frac{1}{\sqrt{\pi}}\cos x, \frac{1}{\sqrt{\pi}}\sin x, \frac{1}{\sqrt{\pi}}\cos 2x, \frac{1}{\sqrt{\pi}}\sin 2x, \ldots$$

then $f(x) = 0$ for all $x \in [-\pi, \pi]$.

Proof. The proposition and its proof are taken from Bressoud [**2**, lemma 8.15] and Saxe [**9**, thm. 4.6] respectively.

Suppose that $f \neq 0$. Then as f is continuous, we know that there exists an x_0 at which $|f|$ achieves its maximum; assume without loss of generality that $f(x_0) > 0$. Let $\delta > 0$ be small enough to ensure that $f(x) > \frac{f(x_0)}{2}$ for all $x \in (x_0 - \delta, x_0 + \delta)$.

Consider the function

$$t(x) = 1 + \cos(x_0 - x) - \cos\delta.$$

This function is a finite linear combination of elements of the Fourier basis; hence f is orthogonal to t, and furthermore it it orthogonal to t^n for every $n \in \mathbb{N}$. This will lead us to a contradiction. Consider

$$\begin{aligned}
0 = \langle f, t^n\rangle &= \int_{-\pi}^{\pi} f(x)t^n(x)\,\mathrm{d}x \\
&= \int_{-\pi}^{x_0-\delta} f(x)t^n(x)\,\mathrm{d}x + \int_{x_0-\delta}^{x_0+\delta} f(x)t^n(x)\,\mathrm{d}x + \int_{x_0+\delta}^{\pi} f(x)t^n(x)\,\mathrm{d}x
\end{aligned}$$

Notice that $t(x) > 1$ for all $x \in (x_0 - \delta, x_0 + \delta)$, and $|t(x)| \leq 1$ for all x outside that range. So as $f(x) \leq f(x_0)$ for all $x \in [-\pi, \pi]$, the first and third integrals are bounded in absolute value by $2\pi f(x_0)$ for all $n \in \mathbb{N}$.

However, consider the middle integral: if $[a, b] \subset (x_0 - \delta, x_0 + \delta)$ then

$$\int_a^b f(x)t^n(x)\,\mathrm{d}x \leq \int_{x_0-\delta}^{x_0+\delta} f(x)t^n(x)\,\mathrm{d}x.$$

Since t is continuous on $[a, b]$, it achieves a minimum value, m say, somewhere in that interval. Since $t(x) > 1$ for all $x \in (x_0 - \delta, x_0 + \delta)$, we have $m > 1$. Then

$$\int_a^b f(x)t^n(x)\,\mathrm{d}x \geq \frac{f(x_0)}{2} \cdot m^n \cdot (b - a),$$

which grows without bound as $n \to \infty$. This contradicts the assumption that $\langle f, t^n \rangle = 0$. The hypothesis that led to this contradiction was to assume that $f \neq 0$; hence we must have $f = 0$ for all $x \in [-\pi, \pi]$. □

Lemma 4.3. *If f is integrable on $[-\pi, \pi]$, and f is orthogonal to each of the functions*

$$\frac{1}{\sqrt{2\pi}}, \frac{1}{\sqrt{\pi}}\cos x, \frac{1}{\sqrt{\pi}}\sin x, \frac{1}{\sqrt{\pi}}\cos 2x, \frac{1}{\sqrt{\pi}}\sin 2x, \ldots$$

then $f(x) = 0$ almost everywhere in $[-\pi, \pi]$.

Proof. The proposition and its proof are taken from Bressoud [**2**, lemma 8.16] and Saxe [**9**, thm. 4.6] respectively.

Define

$$F(x) = \int_{-\pi}^{x} f(t)\,\mathrm{d}t.$$

This is a continuous function of x. Integration by parts yields

$$\int_{-\pi}^{\pi} F(x)\cos nx\,\mathrm{d}x = \left[\frac{1}{n}F(x)\sin nx\right]_{-\pi}^{\pi} - \frac{1}{n}\int_{-\pi}^{\pi} f(x)\sin nx\,\mathrm{d}x.$$

Now f is orthogonal to every member of the Fourier basis, so $\int_{-\pi}^{\pi} f(x)\sin nx\,\mathrm{d}x = 0$, and the first term vanishes as well. Hence

$$\int_{-\pi}^{\pi} F(x)\cos nx\,\mathrm{d}x = 0.$$

Similarly,

$$\int_{-\pi}^{\pi} F(x)\sin nx\,\mathrm{d}x = \left[-\frac{1}{n}F(x)\cos nx\right]_{-\pi}^{\pi} + \frac{1}{n}\int_{-\pi}^{\pi} f(x)\cos nx\,\mathrm{d}x$$

and again we have $\int_{-\pi}^{\pi} f(x)\cos nx\,\mathrm{d}x = 0$ and hence

$$\int_{-\pi}^{\pi} F(x)\sin nx\,\mathrm{d}x = 0.$$

So F is orthogonal to all the members of the Fourier basis, except possibly the member $\frac{1}{\sqrt{2\pi}}$. Letting $C_0 = \langle F, \frac{1}{\sqrt{2\pi}} \rangle$, we see that $F - C_0$ is orthogonal to every member of the Fourier basis. As this is continuous, we can invoke lemma 4.2 to see that $F - C_0$ is identically zero. From this it follows that $f = F'$ is 0 almost everywhere. □

Proof of Theorem 4.1. By lemma 4.3, if $f \in L^2_{\mathbb{R}}([-\pi, \pi])$ and f is orthogonal the functions 1, $\cos nx$, $\sin nx$ for all $n \geq 1$, then $f(x) = 0$ for all $x \in [-\pi, \pi]$. Hence by proposition 3.8, the functions

$$\frac{1}{\sqrt{2\pi}}, \frac{1}{\sqrt{\pi}}\cos x, \frac{1}{\sqrt{\pi}}\sin x, \frac{1}{\sqrt{\pi}}\cos 2x, \frac{1}{\sqrt{\pi}}\sin 2x, \ldots$$

form an orthonormal basis of $L^2_{\mathbb{R}}([-\pi, \pi])$. □

This is a very useful result; every function $f \in L^2_{\mathbb{R}}([-\pi, \pi])$ has a unique Fourier series representation which converges to it in the L^2 sense. Note that convergence in the L^2 norm, or convergence "in mean" as it is sometimes known, does not imply uniform convergence or pointwise convergence; it is, however, a deep theorem of Lennart Carleson and Richard Hunt that convergence in the L^2 norm implies pointwise convergence almost everywhere, as seen in Bressoud [**2**, thm. 8.9].

We can, in fact, go one better. The reader will note that we restricted our attention to real-valued functions at the start of the section; however, we can move to the complex case using de Moivre's formula:

$$e^{iz} = \cos z + i \sin z.$$

Given a function in $L^2_{\mathbb{C}}([-\pi, \pi])$, we can express its real and imaginary parts in terms of sines and cosines, by theorem 4.1. Then, using the formulae

$$\cos z = \frac{e^{iz} + e^{-iz}}{2} \quad \text{and} \quad \sin z = \frac{e^{iz} - e^{-iz}}{2i}$$

we can express each of the real and imaginary parts in terms of the functions e^{ikx}, with k ranging over $\mathbb{Z}$, and combining these gives an expression for the function. Noting that

$$\langle e^{inx}, e^{imx} \rangle = \int_{-\pi}^{\pi} e^{i(n-m)x} \, \mathrm{d}x = \begin{cases} 2\pi & \text{if } n = m \\ 0 & \text{if } n \neq m \end{cases}$$

we arrive at the following corollary:

Corollary 4.4. *The set*

$$\left\{ \frac{1}{\sqrt{2\pi}} e^{inx} \mid n \in \mathbb{Z} \right\}$$

of functions in $L^2_{\mathbb{C}}([-\pi, \pi])$ form an orthonormal basis of that space.

The completeness of the Fourier basis leads to solutions of many partial differential equations, not least the wave equation.

5 Other Bases of L^2

While the Fourier basis of L^2 is important, and quite possibly the most useful, there are others. We can use the *Gram–Schmidt algorithm* to construct new bases from more-or-less arbitrary collections of vectors. This is an inductive process which, given a vector space V and any basis $\{u_i \mid i \in I\}$ for some index set I, we set $e_1 = \frac{u_1}{\|u_1\|}$ and then inductively

$$v_{n+1} = u_{n+1} - \sum_{k=1}^{n} \langle u_{n+1}, e_k \rangle e_k, \qquad e_{n+1} = \frac{v_{n+1}}{\|v_{n+1}\|}.$$

Then the set $\{e_i \mid i \in I\}$ is an *orthonormal* basis of V [**8**].

If we apply the Gram–Schmidt process to the functions

$$1, x, x^2, x^3, \ldots$$

over the domain $[-1, 1]$, we obtain the orthogonal basis consisting of the *Legendre polynomials*:

$$P_n(x) = \frac{1}{2^n n!} \frac{\mathrm{d}^n}{\mathrm{d}x^n} (x^2 - 1)^n$$

for $n \in \mathbb{N}$, together with the constant function $P_0(x) = 1$ [**9**]. The first few are shown in table 1 on the following page.

Order	$P_n(x)$
0	1
1	x
2	$\frac{1}{2}(3x^2 - 1)$
3	$\frac{1}{2}(5x^3 - 3x)$
4	$\frac{1}{8}(35x^4 - 30x^2 + 3)$
5	$\frac{1}{8}(63x^5 - 70x^3 + 15x)$
6	$\frac{1}{16}(231x^6 - 315x^4 + 105x^2 - 5)$

Table 1: The first few Legendre polynomials. [**10**]

Normalising these we get

$$\left\{ \sqrt{\frac{2n+1}{2}} P_n(x) \mid n \in \mathbb{N} \right\}$$

which forms an orthonormal basis of $L^2([-1, 1])$.

Just as the Fourier basis relates to solutions of partial differential equations, so does the Legendre basis: the unnormalised Legendre polynomials are solutions to the differential equation

$$\frac{d}{dx}\left[(1-x^2)\frac{d}{dx}P(x)\right] + n(n+1)P(x) = 0.$$

Each of the Legendre polynomials $P_n(x)$ correspond to each $n \in \mathbb{N} \cup \{0\}$ in the above equation. They appear in solving Laplace's equation of the potential [**11**].

The final basis of L^2 that we discuss is somewhat different to those we have discussed thus far; this is the so-called wavelet basis or *Haar basis* of $L^2([0, 1])$, named for Alfréd Haar (1885–1933, Hungary). This example is fundamentally different from the other two examples in that the functions in the basis are *not* continuous, and they are not connected with differential equations. They instead appear in the study of "wavelets". The Haar basis is defined in Saxe [**9**, §4.1] as $H_{0,0}(x) = 1$ and

$$H_{n,k}(x) = \begin{cases} -2^{n/2} & \text{if } \dfrac{k-1}{2^n} \le x < \dfrac{k-\frac{1}{2}}{2^n} \\ 2^{n/2} & \text{if } \dfrac{k-\frac{1}{2}}{2^n} \le x < \dfrac{k}{2^n} \\ 0 & \text{otherwise} \end{cases}$$

for $n \ge 1$ and $1 \le k \le 2^n$. The study of wavelets has exploded in recent years and has led to many advances into the study of data compression and led to, among other things, the creation of the JPEG format [**11**] for image compression.

In each of these cases, translating concrete problems in partial differential equations (otherwise known as "physics") into abstract theoretical tools has yielded powerful results that can be translated back to the concrete setting to new and possibly surprising ends. Fourier analysis started in solving the heat equation, but it came of age when, having translated the problem into the language of Hilbert spaces, it spawned new forms of analysis leading to, among other things, the mathematical theory of quantum mechanics.

Simply put, physics as we know it today would not have come into being without Fourier analysis and Hilbert space theory. However, it is also fair to say that mathematics as we know it today would not have come into being without Fourier's insights into the series which bear his name; for without the questions raised by Fourier, mathematical analysis as we know it today would not exist. It is, thus, perhaps the greatest triumph of mathematics, which Fourier spawned, that we have succeeeded in taming the infinite in order that we might understand, as Newton put it, "the system of the world".

References

1. David M. Bressoud. *A Radical Approach to Real Analysis.* The Mathematical Association of America, Inc., second edition, 2007.

2. David M. Bressoud. *A Radical Approach to Lebesgue's Theory of Integration.* To appear, 2007. Preprint retrieved from `http://www.macalester.edu/~bressoud/books/book7-jan07.pdf` [accessed 27th January 2007].

3. Frederick W. Byron, Jr. and Robert W. Fuller. *Mathematics of Classical and Quantum Physics: Volume One.* Addison-Wesley Publishing Company, Inc., 1969.

4. Marek Capiński and Ekkehard Kopp. *Measure, Integral and Probability.* Springer-Verlag London Ltd., second edition, 2004.

5. T. W. Körner. *A Companion to Analysis: A Second First and First Second Course in Analysis.* American Mathematical Society, 2004.

6. James Robinson. Applied Analysis lecture notes. University of Warwick, 2006. Available from `http://www.maths.warwick.ac.uk/~jcr/notes.pdf` [accessed 7th February 2007].

7. Walter Rudin. *Principles of Mathematical Analysis.* McGraw-Hill International Editions, third edition, 1976.

8. Bryan P. Rynne and Martin A. Youngson. *Linear Functional Analysis.* Springer-Verlag London Ltd., 2000.

9. Karen Saxe. *Beginning Functional Analysis.* Springer-Verlag New York, Inc., 2002.

10. Eric W. Weisstein. "Legendre Polynomial." From MathWorld–A Wolfram Web Resource, 2007. URL `\url{http://mathworld.wolfram.com/LegendrePolynomial.html}`. [accessed 11th February 2007].

11. Wikipedia. Wikipedia, the free encyclopedia, 2007. URL `\url{http://en.wikipedia.org/}`.

12. Nicholas Young. *An Introduction to Hilbert Space.* Cambridge University Press, 1988.