Orthogonal Functions
Class Notes by Bob Parker

1. Orthogonality and the Inner Product

You are surely familiar with the ordinary **dot product** between two vectors in ordinary space: if $\mathbf{x}, \mathbf{y} \in \mathbb{R}^3$ then

$$\mathbf{x} \cdot \mathbf{y} = x_1 y_1 + x_2 y_2 + x_3 y_3 = x_j y_j \tag{1.1}$$

where x_j and y_j are components in a Cartesian reference system; the second form is of course the one using Einstein summation. As we can easily show the dot product also can be written:

$$\mathbf{x} \cdot \mathbf{y} = |\mathbf{x}|\ |\mathbf{y}| \cos\theta \tag{1.2}$$

where $|\mathbf{x}| = (x_j x_j)^{1/2}$, the Euclidean length of $\mathbf{x}$, similarly for $|\mathbf{y}|$, and θ is the angle between the two vectors. This is very simple. Obviously, these ideas generalize effortlessly to vectors $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ for integer $n > 3$. Then the concept of the angle between to vectors is harder to grasp intuitively, but we can simply use (1.2) as its *definition.*

When two vectors in space are at right angles, we see that their dot product is zero. So in general we will define the condition of **orthogonality** as the situation when $\mathbf{x} \cdot \mathbf{y} = 0$, and both $|\mathbf{x}|, |\mathbf{y}| > 0$.

Suppose in the n-dimensional space $\mathbb{R}^n$ we have a Cartesian coordinate system defined by the directions $\hat{\mathbf{x}}_k$ and an origin somewhere, O. My notation here is that the hat $\hat{}$ means a vector of unit length, so $|\hat{\mathbf{x}}_k| = 1$. A Cartesian axis system is one where all the axes are orthogonal, and so obviously

$$\hat{\mathbf{x}}_j \cdot \hat{\mathbf{x}}_k = \begin{cases} 1, & j = k \\ 0, & j \neq k\,. \end{cases} \tag{1.3}$$

This is more compactly written with the Kronecker delta:

$$\hat{\mathbf{x}}_j \cdot \hat{\mathbf{x}}_k = \delta_{jk}\,. \tag{1.4}$$

The vector $\mathbf{x}$ can be written in terms of this system:

$$\mathbf{x} = (x_1, x_2, x_3, \cdots x_n)\,. \tag{1.5}$$

Another way of writing this is as an expansion of the original vector in terms of the Cartesian axes, an expansion in an **orthogonal basis**:

$$\mathbf{x} = x_1\hat{\mathbf{x}}_1 + x_2\hat{\mathbf{x}}_2 + x_3\hat{\mathbf{x}}_3 + \cdots + x_n\hat{\mathbf{x}}_n = x_j\hat{\mathbf{x}}_j\,. \tag{1.6}$$

If we wanted to find the k-component of $\mathbf{x}$ we could take $\mathbf{x}$ and form the dot product with the unit $\hat{\mathbf{x}}_k$:

$$\mathbf{x} \cdot \hat{\mathbf{x}}_k = x_1\hat{\mathbf{x}}_1 \cdot \hat{\mathbf{x}}_k + x_2\hat{\mathbf{x}}_2 \cdot \hat{\mathbf{x}}_k + \cdots + x_k\hat{\mathbf{x}}_k \cdot \hat{\mathbf{x}}_k + \cdots + x_n\hat{\mathbf{x}}_n \cdot \hat{\mathbf{x}}_k \tag{1.7}$$

$$= 0 + 0 + 0 + \cdots + x_k + \cdots + 0 = x_k\,. \tag{1.8}$$

All the dot products vanish except the one with the $\hat{\mathbf{x}}_k$ (and I assume that $k \neq 1$, 2 or n). Again this is very simple, so simple it seems quite unnecessary to go through it in such detail.

Let us now move onto something a bit more advanced. In place of the space $\mathbb{R}^n$ we will introduce the space $L_2(a, b)$, pronounced "ell two". This is a **linear vector space**, a set of real-valued *functions* with the property that

$$\| f \| = \left(\int_a^b f(x)^2 \, dx \right)^{1/2} < \infty . \tag{1.9}$$

(Strictly speaking, L_2 is the completion of this space, but we will not go into that here.) As you should already know the notation $\| \cdot \|$ is read as the **norm**. The norm of f, which by (1.9) is always finite for members of L_2, is a measure of the size of the function f, clearly by analogy with the Euclidean length of a vector.

We are going to treat the functions in L_2 as if they were vectors. We already have an idea of the length of the vector through the norm. We will define an analog for the dot product called the **inner product**:

$$(f, g) = \int_a^b f(x)\, g(x) \, dx \tag{1.10}$$

defined for any two functions f and g in $L_2(a, b)$. We can see that this definition has the same properties as the dot product; for example, it commutes: $(f, g) = (g, f)$, and so on. It has the same relationship to the norm as the dot product does to the length:

$$\| f \| = (f, f)^{1/2} \quad \text{and} \quad |\mathbf{x}| = (\mathbf{x} \cdot \mathbf{x})^{1/2} . \tag{1.11}$$

And now we can say that two elements of L_2 are **orthogonal** if

$$(f, g) = 0, \ \text{and} \ \| f \|, \ \| g \| > 0 . \tag{1.12}$$

Let us prove something now: **Schwarz's inequality.** This states that

$$|(f, g)| \le \| f \| \ \| g \| . \tag{1.13}$$

For ordinary vectors the analogy is that $|\mathbf{x} \cdot \mathbf{y}| \le |\mathbf{x}| \ |\mathbf{y}|$, which is obvious from (1.2). The result (1.13) is clearly true when $g = 0$, meaning the function $g(x)$ vanishes everywhere on the interval (a, b). So we only need examine the case $\| g \| \neq 0$. Consider the function given by $h = f - \gamma\, g$ where γ is a real number. By the (1.9) and the left half of (1.11) applied to h:

$$0 \le \| h \|^2 = \| f - \gamma\, g \|^2 = (f - \gamma\, g, f - \gamma\, g) \tag{1.14}$$

$$0 \le (f, f) - 2\gamma\, (f, g) + \gamma^2\, (g, g) . \tag{1.15}$$

We get (1.15) by treating the inner product algebraically just like a dot product, or ordinary multiplication, which is allowed if one looks at the definition (1.10). Now in (1.15) simply choose the value of $\gamma = (f, g)/(g, g)$ which exists because the denominator does not vanish by assumption; then

$$0 \le (f,f) - 2(f,g)^2/(g,g) + (f,g)^2/(g,g) = (f,f) - (f,g)^2/(g,g)\,. \tag{1.16}$$

And tidying up this is

$$(f,f)\,(g,g) \ge (f,g)^2 \tag{1.17}$$

and hence (1.13). QED □

We should mention there are many variants of L_2; this is the family of **Hilbert spaces**. Because Hilbert spaces must have an inner product (and therefore a norm) the idea of orthogonality and expansions in an orthogonal basis works in all of them. A simple variant is weighted L_2, in which the norm is

$$\| f \| = \left(\int_a^b f(x)^2 \, w(x) \, dx \right)^{1/2} \tag{1.18}$$

where $w(x) > 0$. Another important Hilbert space is one based on complex valued functions, where the inner product is

$$(f,\, g) = \int_a^b f(x)\, g(x)^* \, dx \tag{1.19}$$

where the asterisk is complex conjugate; we discuss this further in Section 5. See Section 7 for another Hilbert space, this one defined for functions on a sphere.

Exercises

1. How many parameters must one specify to change from one Cartesian coordinate system to another in three-dimensional space? How many in $\mathbb{R}^n$?
   Hint: You must move the origin and rotate the axes.

2. Use Schwarz's inequality for L_2 to prove

   $$\| f + g \| \le \| f \| + \| g \|$$

   which is known as the **triangle inequality,** a property obvious in $\mathbb{R}^2$ and needed for every norm.

3. Which of the following functions belongs to the space $L_2(-1, +1)$? Explain carefully how you reach your conclusion.

   $$f_1(x) = \cos x;\;\; f_2(x) = \frac{\sin \pi x}{\pi x}\,;\;\; f_3(x) = \frac{1}{|x|^{1/2}}\,;\;\; f_4(x) = \frac{\cos x}{|x|^{1/4}}\,.$$

   Which of these functions belongs in $L_2(-\infty, \infty)$?

2. Infinite Sets of Orthogonal Functions

We are going to look at the analog of (1.6) in which we take a function in L_2 and expand it in a (usually infinite) sequence of orthogonal functions, a kind of basis for L_2. The idea is to write

$$f = \sum_{n=1}^{\infty} c_n g_n \tag{2.1}$$

where c_n are real numbers (the "components" of f) and g_n comprise an infinite sequence of mutually orthogonal functions, corresponding to the Cartesian basis in (1.6). In ordinary space or in n-dimensional space there are infinitely many different Cartesian axis systems. There are infinitely many ways of performing the expansion (2.1) as well.

Several questions arise that do not have analogs in finite-dimensional spaces. (As I have hinted $L_2(a, b)$ is an example of an **infinite-dimensional linear vector space,** one whose members cannot be built up by linear combinations of a finite number of basis functions.) For example, (2.1) is an infinite sum; when does it converge and does a converged sum always give the right answer? Not every infinite collection of orthogonal functions is rich enough to expand every function in L_2; this is the question of **completeness.** How do we generate suitable sets of orthogonal functions? How do we get the coefficients c_n when we know the function f and the basis functions g_n?

This last question is easy to answer. We can use the same process as I used in equations (1.7-8). Take the inner production of both sides of (2.1) with the function g_k. Then

$$(f, g_k) = (\sum_{n=1}^{\infty} c_n g_n, g_k) \tag{2.2}$$

$$= \sum_{n=1}^{\infty} c_n (g_n, g_k) \tag{2.3}$$

$$= c_k \| g_k \|^2 . \tag{2.4}$$

Every one of the inner products in the sum in (2.3) vanishes, except the one with $n = k$. Rearranging (2.4) we have the result that

$$c_k = (f, g_k)/\| g_k \|^2 . \tag{2.5}$$

We could make (2.5) even simpler if we had $\| g_n \| = 1$, and this is sometimes the case, but convention does not demand it.

Before discussing the more complicated issues of generation and completeness, let us look at an example or two. The first example is that of the ordinary **Fourier Series**, discussed at length in Chapter 12 of Riley et al. We choose a concrete interval (a, b) to work on. It is convenient to pick $a = -\pi$ and $b = \pi$. Now the functions we need are the cosines and sines that are periodic on $(-\pi, \pi)$ plus the constant function $g_0(x) = 1$. For this case it is more tidy to let the sum in (2.3) begin at $n = 0$

rather than 1; there is some arbitrariness in labeling here, but please be tolerant. I will say for $n = 0, 1, 2, \cdots$

$$g_{2n}(x) = \cos nx, \quad \text{and} \quad g_{2n+1}(x) = \sin(n+1)x\,. \tag{2.6}$$

We need to show that these functions are indeed mutually orthogonal. The cosines are all even functions, which means $g_{2n}(x) = g_{2n}(-x)$ and the sines are all odd functions, with $g_{2n+1}(x) = -g_{2n+1}(-x)$; see Figure 1. So

$$(g_{2n}, g_{2n+1}) = \int_{-\pi}^{\pi} g_{2n}(x)\, g_{2n+1}(x)\, dx \tag{2.7}$$

$$= \int_{0}^{\pi} g_{2n}(x)\, g_{2n+1}(x)\, dx + \int_{0}^{\pi} g_{2n}(-x)\, g_{2n+1}(-x)\, dx \tag{2.8}$$

$$= \int_{0}^{\pi} g_{2n}(x)\, [g_{2n+1}(x) + g_{2n+1}(-x)]\, dx = 0\,. \tag{2.9}$$

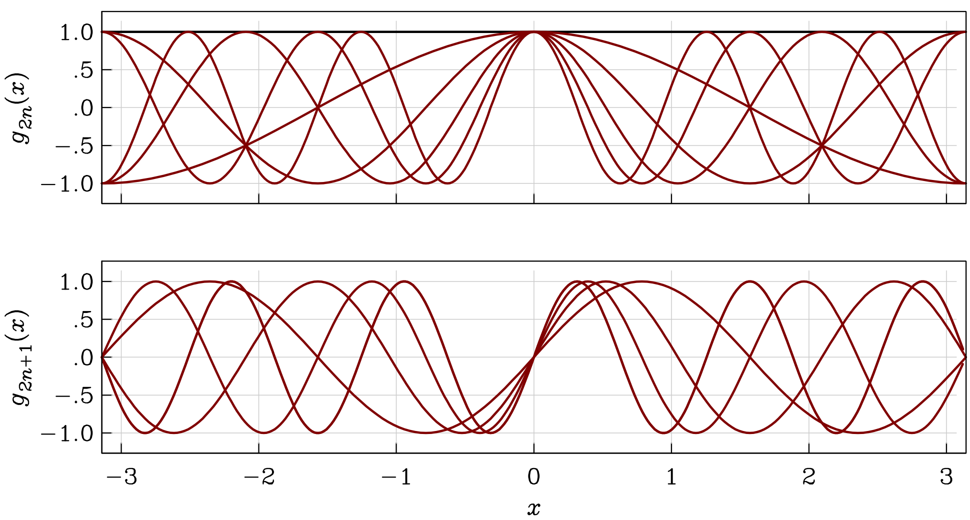
That takes care of sines multiplied by cosines. Even though we are using a *real* vector space for L_2 (and the complex version is in fact more elegant; see the Exercises) it is worth knowing that complex variables almost always simplify calculations with sines and cosines. So we use them here. First, we have the elementary and nearly obvious result that for integer $n \neq 0$:

$$\int_{-\pi}^{\pi} \mathrm{e}^{inx}\, dx = \left.\frac{\mathrm{e}^{inx}}{in}\right|_{-\pi}^{\pi} = \frac{\mathrm{e}^{in\pi} - \mathrm{e}^{-in\pi}}{in} = \frac{2\sin n\pi}{n} = 0\,. \tag{2.10}$$

Write $\cos nx = \mathrm{Re}\, \mathrm{e}^{inx}$. Then when $m \neq n$ we have

$$(g_{2n}, g_{2m}) = \mathrm{Re} \int_{-\pi}^{\pi} \cos mx \mathrm{e}^{inx}\, dx = ½\mathrm{Re} \int_{-\pi}^{\pi} (\mathrm{e}^{imx} + \mathrm{e}^{-imx})\, \mathrm{e}^{inx}\, dx \tag{2.11}$$

Figure 1: The first 10 members of the Fourier basis (2.6).



$$= \tfrac{1}{2}\mathrm{Re}\left[\int_{-\pi}^{\pi} e^{i(m+n)x}\,dx + \int_{-\pi}^{\pi} e^{-i(m-n)x}\,dx\right] = 0\,. \tag{2.12}$$

A similar calculation works for the sines. Finally, we need

$$\| g_0 \|^2 = \int_{-\pi}^{\pi} 1\,dx = 2\pi \tag{2.13}$$

$$\| g_{2n} \|^2 = \int_{-\pi}^{\pi} \cos^2 nx\,dx = \pi,\ \ n \geq 1 \tag{2.14}$$

$$\| g_{2n+1} \|^2 = \int_{-\pi}^{\pi} \sin^2 (n+1)x\,dx = \pi,\ n \geq 0\,. \tag{2.15}$$

This rather messy collection of results allows us to find the ordinary Fourier expansion of a function on $(-\pi, \pi)$. We just use (2.5) and the definition of the inner product (1.10). The expansion (2.1) is then the **Fourier series** for f:

$$S(f) = \sum_{n=0}^{\infty} \frac{(f, g_n)}{\| g_n \|^2}\, g_n \tag{2.16}$$

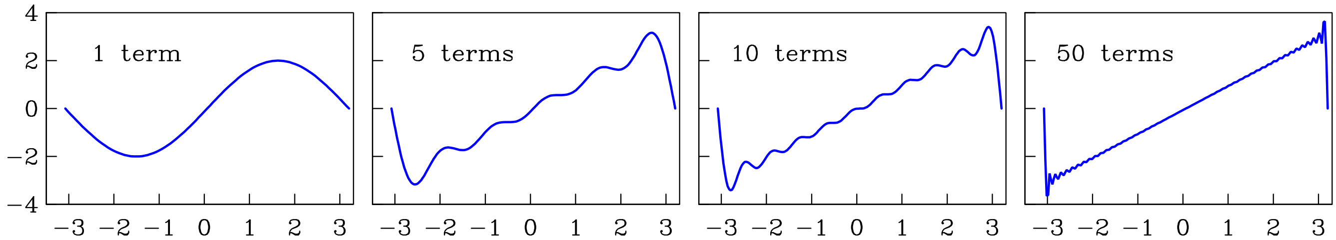
Let us calculate the expansion coefficients for a simple function, and study how the Fourier series converges experimentally. Almost the simplest function I can think of is $f(x) = x$, obviously in $L_2(-\pi, \pi)$. Then, because this is an odd function in x, all the cosine terms vanish. You will see in Exercise 1 that a complex basis is much simpler than the real one we are using here, but I will just use the complex exponential to help us evaluate the integral again; from (2.5) and (2.16):

$$c_{2n+1} = \frac{1}{\pi}\,\mathrm{Im}\int_{-\pi}^{\pi} x\, e^{i(n+1)x}\,dx \tag{2.17}$$

$$= \frac{1}{\pi}\,\mathrm{Im}\left[\left.\frac{x e^{i(n+1)x}}{i(n+1)}\right|_{-\pi}^{\pi} - \int_{-\pi}^{\pi} 1 \cdot \frac{e^{i(n+1)x}}{i(n+1)}\,dx\right] \tag{2.18}$$

$$= \mathrm{Im}\,\frac{1}{\pi}\left[\frac{-2(-1)^{n+1}\pi}{i(n+1)} - 0\right] = \frac{2(-1)^n}{n+1} \tag{2.19}$$

Figure 2: Partial sum of the series in (2.20).



where I integrated by parts to get (2.18) and used $e^{i\pi} = -1$ in (2.19). So rearranging slightly we conclude that

$$S(f) = 2\left[\sin x - \frac{\sin 2x}{2} + \frac{\sin 3x}{3} - \frac{\sin 4x}{4} + \cdots\right]. \tag{2.20}$$

In Figure 2 we see partial sums of this series. Notice how the series converges (trivially) at $x = \pm\pi$, but not to the correct answer. Also observe the persistent overshoot near the ends, something known as **Gibbs' phenomenon.** See Section 6 (but not Riley et al.!) for the theory: the overshoot never goes away, no matter how many terms are taken. It can be proved that $S(f)$ is convergent **pointwise** to $f(x)$ in this case, except at the ends of the interval. It is not **uniformly convergent** to $f(x)$. In uniform convergence, one measures the maximum deviation of the partial sum from the true function. There is a sense in which the series does converge to f, which we describe briefly in Section 3.

Another example of a complete family in L_2 is the set of sine functions on a different interval, $(0, \pi)$: let

$$g_n(x) = \sin nx, \quad n = 1, 2, 3, \cdots \tag{2.21}$$

which are shown in Figure 3. Even if we shift the origin back to the center of the interval, this is a different set of functions from the sines and cosines in the first example; for example, the constant function $f(x) = 1$ is not in the set. With the Fourier basis in (2.6) exactly one term is needed in the series expansion, whereas with (2.21) an infinite number of terms is required. Without going into details of the derivation we find that here if $f(x) = 1$ for x on $(0, \pi)$ then

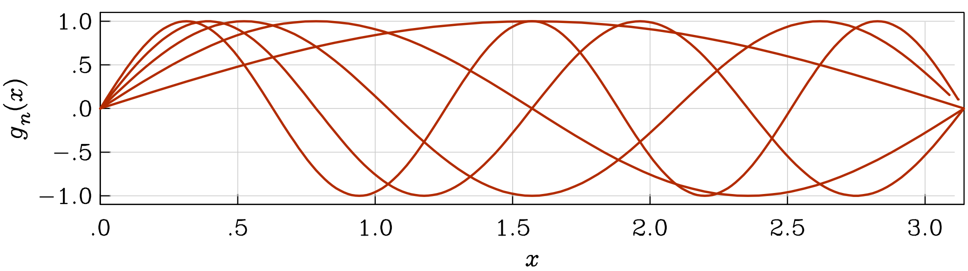
$$c_n = \begin{cases} 4/\pi n, & n \text{ odd} \\ 0, & n \text{ even} \end{cases} \tag{2.22}$$

and thus

$$S(f) = \frac{4}{\pi}\left[\sin x + \frac{\sin 3x}{3} + \frac{\sin 5x}{5} + \frac{\sin 7x}{7} + \cdots\right]. \tag{2.23}$$

So far it looks as if the only functions available for orthogonal sets are sines and cosines, but that is not the case, although they do have special properties. For example, if one insists on extending the interval outside the original (a, b) the function $f(x)$ is reproduced periodically by the periodic trigonometric functions.

Figure 3: The sine Fourier basis (2.21) with $n = 1, 2, \cdots 5$.



As a final example in this section, I introduce a set of orthogonal polynomials, functions that are composed of sums of the powers x^n. You will be familiar with the ordinary Taylor series for elementary functions, for example:

$$\ln(1+x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} + \cdots \tag{2.24}$$

valid when $-1 < x \leq 1$. The powers x, x^2, etc are not orthogonal on $(-1, 1)$, or any other interval for that matter, so we cannot use (2.16) to form a power series like (2.24). But powers can be built into an orthogonal set and do form a basis for $L_2(a, b)$ for finite intervals. There are the **Legendre polynomials**: for x in $(-1, 1)$ we define

$$P_n(x) = \frac{1}{2^n n!} \frac{d^n}{dx^n}\left[(x^2-1)^n\right], \quad n = 0, 1, 2, \cdots \tag{2.25}$$

Here are the first few:

$$P_0(x) = 1; \quad P_1(x) = x; \quad P_2(x) = ½(3x^2 - 1); \quad P_3(x) = ½(5x^3 - 3x)\,. \tag{2.26}$$

I show a few more in Figure 4. The polynomial $P_n(x)$ is always of degree n, is composed of even powers when n is even, and odd powers when n is odd, and most important, these polynomials are orthogonal functions: $(P_m, P_n) = 0$ when $m \neq n$. How do we obtain these functions, and others like them, since there are infinitely many such sets? I will discuss this question in the Section 4. But first a very brief section on convergence.

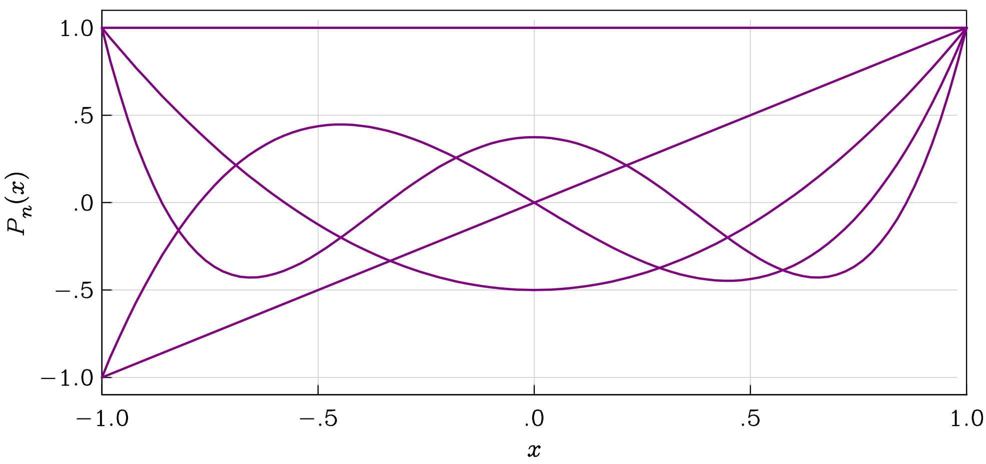
Exercises

1. It is easy to extend the definition of the inner product to functions $f : \mathbb{R} \to \mathbb{C}$, that is, complex-valued functions with a real argument. Here is how it is done:

$$(f, g) = \int_a^b f(x) g(x)^* \, dx$$

where * means take the complex conjugate. This ensures that (f, f) is positive and real, so that (1.11) still works. Show that with this inner product the

Figure 4: The first five Legendre polynomials (2.25).



functions

$$e_n(x) = e^{2\pi inx}, \ \ n = 0, \pm 1, \pm 2, \cdots$$

are orthogonal on (0, 1). Show that $\| e_n \| = 1$. Show the Fourier series

$$S(f) = \sum_{-\infty}^{\infty} c_n e^{2\pi inx}$$

is exactly equivalent for real functions to the one obtained from (2.6), after a suitable scaling and change of origin in x.

2. Show that the set of functions

$$g_n(x) = \cos nx, \quad n = 0, 1, 2, \cdots$$

are orthogonal on the interval $(0, \pi)$, and find $\| g_n \|$. Assuming this set to be complete (it is), find the coefficients c_n in a Fourier expansion for the function $f(x) = x$. Plot partial the partial sums of this series as shown in Figure 2. Is there evidence of Gibbs' phenomenon here?

3. In the Taylor series (2.24), set $x = e^{i\theta}$ and compare the imaginary part with the result (2.20). Is the real part of this function so defined in $L_2(-\pi, \pi)$? If so, is the real part of the right side a valid Fourier series? Plot the partial sums of the real part and compare them with the function.

4. Prove **Parseval's Theorem** for an expansion in an orthogonal basis:

$$\| f \|^2 = \sum_{n=1}^{\infty} c_n^2 \, \| g_n \|^2$$

where f and c_n are as in (2.1).

In Question 3 the Fourier series for the real and imaginary parts have the same coefficients. Use this fact and Parseval's Theorem to prove:

$$\sum_{n=1}^{\infty} \frac{1}{n^2} = \frac{\pi^2}{6}$$

and

$$\int_{-\pi}^{\pi} \ln^2 2(1+\cos\theta) \, d\theta = \frac{2\pi^3}{3} .$$

3. Convergence

In elementary calculus courses there is usually a discussion of convergence of series, giving tests and conditions to cover the case of a sequence of numbers; here we have a sequence of *functions*. New issues arise in this case that have no counterpart in the simpler theory: for example, there are several different ways in which a sequence of functions can converge.

The question of the precise conditions under which a Fourier series converges and in what manner was an important topic of analysis in the 19th century. Before discussing this I state, but not do prove (of course), the general result for complete orthogonal series. If the set of functions g_n form an orthogonal basis (that means the set is complete) then it can be shown that:

$$\lim_{N\to\infty} || f - S_N(f) || = 0 \tag{3.1}$$

where

$$S_N(f) = \sum_{n=1}^{N} c_n g_n \tag{3.2}$$

and where c_n defined in the usual way, by (2.6). Thus the distance in the 2-norm between the partial sums and the original function always shrinks to zero; thus we have guaranteed convergence in the sense of the norm of L_2. Doesn't this mean the two functions are equal? Not quite. Because the norm (1.9) depends on an integral, and it does not "see" differences if they occur at isolated points, like the end points of the interval in Figure 2. That series never gets the values right at the end points, yet measured by the norm, the two functions are the same. In fact, examples can be cooked up in which $S(f)$ and f differ at an *infinite* number of points, yet satisfy (3.2). Equation (3.1) is the definition of **convergence with respect to a norm**, in this case the 2-norm defined in (1.9).

This gets us into the idea of different kinds of convergence. The most natural-seeming is *pointwise* convergence, where we examine the discrepancy at every point. It is difficult (maybe impossible) to provide general results about pointwise convergence for general orthogonal functions, but for Fourier series a lot has been done. We have to ask that f be a lot smoother than just belonging to L_2 which is a "big" space of functions. You might think having $f(x)$ continuous would be enough, but that obviously doesn't work: look at Figure 2 again and consider that $f(x) = x$, surely a continuous function! But we know the sine and cosines are really defined in a periodic way, and perhaps we should consider function defined on a circle, so that the value (and derivatives if they exist) at $x = 0$ match those at $x = 2\pi$. Then the function $f(x) = x$ isn't continuous because it jumps from one value to another across the end points. It turns out that even functions continuous on a circle do not always have pointwise convergent Fourier series! See Körner's book for examples. We need stronger conditions. Riley et al. (p 421) give the **Dirichlet conditions.** Here is a slightly different statement of the conditions copied from Keener:

> Suppose $f(x)$ is in $PC^1[0, 2\pi]$, that is, piecewise C^1, which means f has continuous first derivatives on the interval, except possibly at a finite number of points at which there is a jump in $f(x)$, where left and right derivatives must exist. Then the Fourier series of f converges to ½$(f(x^+) + f(x^-))$ for every point in the open interval $(0, 2\pi)$. At $x = 0$ and 2π, the series converges to ½$(f(0^+) + f(2\pi^-))$.

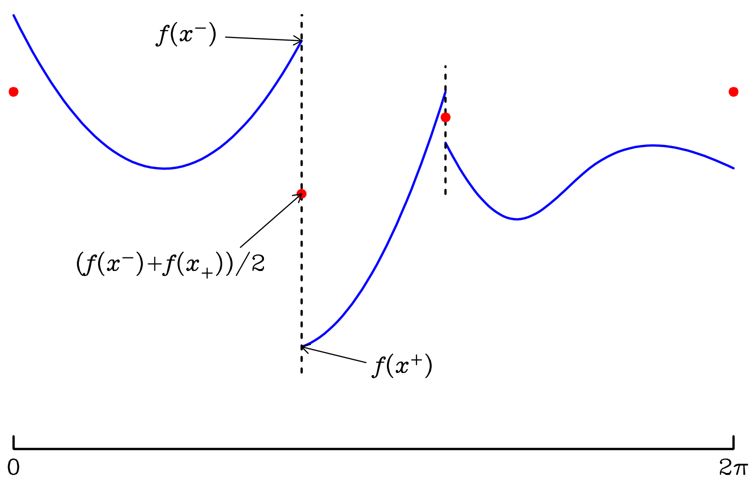
Notice this statement answers two questions: first, the series under these conditions always converges pointwise; second, it tells us what the series converges to: the correct value most of the time, giving possibly incorrect answers only at the discontinuities and the ends of the interval. This kind of convergence is called **convergence in the mean.**

Another form of convergence is **uniform convergence.** This used to be the gold standard of convergence. For continuous functions, you can measure the maximum difference between the series and the function written:

$$\lim_{N \to \infty} \| f - S_N(f) \|_\infty = 0 \,. \tag{3.3}$$

The maximum discrepancy is another kind of norm, called the **uniform norm** (also called the **sup norm**, for supremum, which is the mathematically way of defining a maximum for discontinuous functions). Pointwise convergence does not imply uniform convergence! The Gibbs' phenomenon, which we discuss later, is the most famous example of this. But uniform convergence does include pointwise.

Figure 4a: Convergence in the mean.



4. Where Do Orthogonal Functions Come From?

The first thing to be looked at is the **Gram-Schmidt** process. Suppose we have an infinite set of functions $u_1, u_2, u_3, \cdots$ in L_2 and they are not orthogonal. We can create an orthogonal set from them in a clever way. It is helpful to create the new set to be of unit norm: $\| g_n \| = 1$. We start with the first one, and say $g_1 = u_1 / \| u_1 \|$. Next we take u_2 and remove any part parallel to g_1:

$$g = u_2 - g_1(g_1, u_2)\,. \tag{4.1}$$

The function $g_1(g_1, u_2)$ is called the **orthogonal projection** of u_2 onto g_1. Observe that g has zero inner product with g_1:

$$(g_1, g) = (g_1, u_2) - (g_1, g_1)(g_1, u_2) = 0\,. \tag{4.2}$$

Now it could be that $\| g \|$ vanishes, and if it does we don't get a new function. But if $\| g \| > 0$ we say this is the next orthogonal element: $g_2 = g / \| g \|$. We repeat the process, starting with u_3 and removing the components parallel to u_1 and u_2:

$$g = u_3 - g_1(g_1, u_3) - g_2(g_2, u_3) \tag{4.3}$$

and if this g doesn't vanish we add to the set, building up an orthogonal series of functions, $g_1, g_2, \cdots$.

Thus from any infinite sequence of functions we can get an orthogonal set. That set might be infinite, or it might not; it might be a basis for L_2 or not. These are hard questions. For example, if we start with the functions $u_n(x) = x^n$ for $n = 0, 1, 2 \cdots$, the powers of x on $(-1, 1)$, the Gram-Schmidt process gives us an infinite set of orthogonal polynomials and, yes, *they are the Legendre polynomials* (scaled to get a unit norm). They are a basis. Suppose now that you decide to omit u_0, the constant. Performing Gram-Schimdt on the reduced set gives another series of orthogonal polynomials, not the Legendre polynomials. Are these complete, even though we dropped a function from the first list? They are! But if we dropped the constant function P_0 from the list of Legendre polynomials, that orthogonal set *would not be complete.*

So we now know how to get an orthogonal set of functions out of any sequence. But that is not the way the orthogonal functions arise most often. Their most common appearance is in connection with the eigenvalue problem of special linear operators, called **self-adjoint** operators or when the functions are complex, **Hermitian operators.** You will be seeing this theory properly in Chapter 17 of Riley et al. The basic idea can be appreciated again by analogy to finite-dimensional spaces, where there are linear mappings of vectors into each other, the matrices. The self-adjoint property of operators corresponds with matrix symmetry. Let us write

$$g = Lf \tag{4.4}$$

where f and g are in a smoother subspace of L_2. Self-adjointness of L is by definition that

$$(f, Lg) = (Lf, g) \tag{4.5}$$

for all f and g for which L may act. When a matrix is symmetric ($A = A^T$) and eigenvalues of A are distinct, this implies two things: real eigenvalues and *orthogonal* *eigenvectors.* It means the same thing for self-adjoint operators, and the proof is quite easy. So any self-adjoint operator naturally generates an infinite orthogonal sequence of functions. If it is a basis (that is, the set is complete, which is not always true), this a good way to solve differential equations, because L is frequently a differential operator we are interested in. Eigenfunction expansions are used extensively in scattering problems and of course in normal mode seismology. They essentially convert the differential equation into an algebraic one.

To give a specific example, consider the operator $L = d^2/dx^2$ for functions on $(0, \pi)$; to define L properly we must also specify boundary conditions for L at the endpoints, and so I will take $f(0) = f(\pi) = 0$. For eigenvalue problems the boundary conditions need to homogeneous. We can now verify L is self-adjoint according to (4.5):

$$(f, Lg) - (Lf, g) = \int_0^\pi \left[f(x) \frac{d^2 g}{dx^2} - g(x) \frac{d^2 f}{dx^2} \right] dx \tag{4.6}$$

$$= \left[f(x) \frac{dg}{dx} - g(x) \frac{df}{dg} \right]_0^\pi - \int_0^\pi \left[\frac{df}{dx} \frac{dg}{dx} - \frac{dg}{dx} \frac{df}{dx} \right] dx = 0 \,. \tag{4.7}$$

I integrated by parts to get (4.7), then applied the boundary conditions to make the first bracketed term vanish. The eigenvalue problem for L is:

$$Lu = \lambda u, \text{ with } u(0) = u(\pi) = 0. \tag{4.8}$$

You will easily verify that because $L \sin nx = -n^2 \sin nx$ the eigenvalues and eigenfunctions for (4.8) are

$$\lambda_n = -n^2, \quad u_n(x) = \sin nx, \quad n = 1, 2, 3, \cdots \tag{4.9}$$

So we have recovered the Fourier sine orthogonal family of (2.21)! As you may wish to verify, the other kinds of Fourier basis arise from different homogeneous boundary conditions. So Fourier series are really eigenfunction expansions.

The eigenvalue problem for Legendre polynomials is given by

$$(1 - x^2) \frac{d^2 P_n}{dx^2} - 2x \frac{dP_n}{dx} = \lambda_n P_n(x) \tag{4.10}$$

for x on the interval $(-1, +1)$. The "boundary values" are that the solution be regular at $x = \pm 1$, because as you can see there are singular points in the operator at the end points, and singular solutions do exist, but these are not the ones required. The eigenvalues are $\lambda_n \in \{0, -2, -6, -12, \cdots -l(l+1), \cdots\}$.

One of the odd things at first sight is that the domain of an operator like d^2/dx^2 (that is, the functions that it can act on, twice-differentiable functions), is so much smaller than the space L_2 which merely have to be square integrable, yet the eigenfunctions of L provide a basis for *all of* L_2. This is very useful, because it means nonsmooth functions that may arise in a differential equation, for example as a forcing term, can still be expanded in the basis.

The question of how one proves completeness is quite difficult. For special cases, like Fourier series and Legendre polynomials, authors often appeal to the completeness of the powers of x and use something called **Weierstrass' theorem.** For self-adjoint operators, a sufficient condition is that the operator be **compact**, a notion connected to the property of producing a convergent sequence from a merely bounded one. But differential operators are not compact, so to use this general property one must find a related compact operator that shares the same eigenfunctions; this isn't so hard as it sounds because any two commuting operators have the same eigenfunctions. See Keener's book for a good treatment.

Exercises

1. Consider the infinite set of functions on the interval $(0, \pi)$ given by

$$u_n(x) = \cos^n x, \quad n = 0, 1, 2, \cdots$$

Use the Gram-Schmidt process to generate the first four orthogonal functions $g_n(x)$ in an infinite sequence under the inner product of $L_2(0, \pi)$. Use MATLAB to graph these four. From your result make a guess for the general form of g_n, and then prove it.

Hint: $\int_0^\pi \cos^{2n} x \, dx = \sqrt{\pi}\,\Gamma(n + ½) / \Gamma(n+1)$, with integer $n \ge 0$.

Books

Dym, H. and H. P. McKean, *Fourier Series and Integrals,* Academic Press, New York, 1972.
A mathematician's version of the theory; modern compact notation; terse but elegant. Chapter 1 is on Fourier series, giving dozens of applications and proofs of all the standard results.

Körner, T. W., *Fourier Analysis,* Cambridge University Press, Cambridge, 1988.
Another mathematician's book, full of great applications, rigorous proofs, but haphazardly organized.

Keener, J. P., *Principles of Applied Mathematics: Transformation and Approximation,* Addison-Wesley, Redwood City, 1988.
The general book you should turn to after Riley et al. Unfortunately, it's currently out of print.

Riley, K. F., Hobson, M. P., and Bence, S. J., *Mathematical Methods for Physics and Engineering,* Cambridge Univ. Press, 2002.
Covers a lot of ground; useful introduction. Not my favorite, but chosen for the class. Notation is old-fashioned or nonstandard, generally low level of rigor.

5. The Complex Fourier Basis

It is time to expand our treatment to complex-valued functions of a real argument, typically things like $e^{i\theta}$ which we a already familiar with. This Section was offered as an Exercise earlier, but it is important enough to write out in full here. Complex $L_2(a, b)$ is defined as the set of complex functions on (a, b) such that

$$\| f \| = \left(\int_a^b | f(x) |^2 \, dx \right)^{1/2} < \infty \,. \tag{5.1}$$

Notice this is gives exactly the same answer as real L_2 if f happens to be real. The inner product is slightly different:

$$(f, g) = \int_a^b f(x) \, g(x)^* \, dx \tag{5.2}$$

where $(\cdot)^*$ means complex conjugate. This definition requires some care, because it is no longer true that the inner product commutes; instead we see

$$(f, g) = (g, f)^* \,. \tag{5.3}$$

The greatest use for us will be the complex Fourier basis for $L_2(0, 1)$. We will show that the following functions are orthogonal:

$$e_n(x) = e^{2\pi i n x}, \quad n = 0, \pm 1, \pm 2, \cdots \tag{5.4}$$

We calculate the required inner products from their definition:

$$(e_m, e_n) = \int_0^1 e_m(x) \, e_n(x)^* \, dx = \int_0^1 e^{2\pi i m x} \, (e^{2\pi i n x})^* \, dx \tag{5.5}$$

$$= \int_0^1 e^{2\pi i m x} \, e^{-2\pi i n x} \, dx \tag{5.6}$$

$$= \int_0^1 e^{2\pi i (m-n) x} \, dx \,. \tag{5.7}$$

When $m = n$ the integrand is obviously 1, so

$$(e_n, e_n) = \| e_n \|^2 = 1 \,. \tag{5.8}$$

When $m \neq n$ we have

$$(e_m, e_n) = \left[\frac{e^{2\pi i (m-n) x}}{2\pi i (m-n)} \right]_0^{2\pi} = \frac{1}{2\pi i (m-n)} \left(1 - e^{2\pi i (m-n)} \right) \tag{5.9}$$

$$= \frac{1}{2\pi i (m-n)} \, (1 - 1) = 0 \tag{5.10}$$

because $e^{2\pi i N} = 1$ for any integer N. Hence the e_n are orthogonal and of unit norm.

They are also a complete set in $L_2(0, 1)$, so we can expand arbitrary functions in that space with them. In fact they are exactly equivalent to the sine and cosine basis we discussed earlier on the interval $(-\pi, \pi)$, after a bit of shifting and scaling.

To show the equivalence we write out the sum in complex form, then rearrange a little:

$$S(f) = \sum_{-\infty}^{\infty} c_n e^{2\pi i n x} = \left(\sum_{-\infty}^{-1} + \sum_{0}^{\infty} \right) c_n e^{2\pi i n x} \tag{5.11}$$

$$= c_0 + \sum_{n=1}^{\infty} \left[c_{-n} e^{-2\pi i n x} + c_n e^{2\pi i n x} \right] \tag{5.12}$$

$$= c_0 + \sum_{n=1}^{\infty} [c_{-n}(\cos 2\pi nx - i \sin 2\pi nx) + c_n(\cos 2\pi nx + i \sin 2\pi nx)] \tag{5.13}$$

$$= c_0 + \sum_{n=1}^{\infty} [(c_{-n} + c_n)\cos 2\pi nx + i(-c_{-n} + c_n)\sin 2\pi nx]\,. \tag{5.14}$$

If we demand $c_n + c_{-n}$ to be real and $c_n - c_{-n}$ to be imaginary, this sum is an expansion almost (2.6) except for the interval. But you can easily see that by making $y = 2\pi(x - ½)$, the new interval becomes $(-\pi, \pi)$ and the basis functions are then exactly the same. The complex form is much easier to work with in general, is strongly recommended.

Another important result that we should mention is **Parseval's Theorem**; this was an exercise in Section 2 . One way to look at it is as a conservation of energy as we shall see. The result is straightforward to show if we don't worry too much about rigor: in complex L_2

$$\| f \|^2 = (f, f) = (\sum_n c_n g_n, \sum_n c_n g_n) \tag{5.15}$$

$$= \sum_m \sum_n (c_m g_m, c_n g_n) = \sum_m \sum_n c_m (c_n)^* (g_m, g_n)$$

$$= \sum_n |c_n|^2 (g_n, g_n) = \sum_n |c_n|^2 \| g_n \|^2\,. \tag{5.16}$$

This Parseval's Theorem. It is exactly the same in real L_2, absent the magnitude on c_n. With the complex Fourier basis e_n, which is an orthonormal set, this result is even simpler. The physical interpretation, as I hinted earlier, is that the energy in the system, as given by the squared norm, is the same as the energy summed over the orthogonal modes.

Parseval's Theorem is a neat way of summing some difficult looking infinite series. For example, recall the Fourier sine series for the constant function $f(x) = 1$ on the interval $(0, \pi)$, equation (2.23).

$$g_n(x) = \sin nx, \quad n = 1, 2, 3, \cdots \tag{5.17}$$

Then $\| g_n \|^2 = ½\pi$, and with $f(x) = 1$, we found

$$c_n = 4/n\pi, \quad n \text{ odd}, \quad = 0, \quad \text{otherwise}\,. \tag{5.18}$$

Therefore, according to Parseval

$$\| f \|^2 = \sum_{n \text{ odd}} \left(\frac{4}{n\pi} \right)^2 \| g_n \|^2 \tag{5.19}$$

$$= \frac{16}{\pi^2} \frac{\pi}{2} \sum_{n \text{ odd}} \frac{1}{n^2} = \frac{8}{\pi} \left[1 + \frac{1}{9} + \frac{1}{25} + \cdots \right]. \tag{5.20}$$

But the integral for the norm of f is really easy:

$$\| f \|^2 = \int_0^\pi 1 \, dx = \pi \, . \tag{5.21}$$

From which we conclude

$$\frac{\pi^2}{8} = 1 + \frac{1}{9} + \frac{1}{25} + \cdots \tag{5.22}$$

6. Gibbs' Phenomenon

We keep meeting this issue and it is unsatisfactory for you not to have a treatment, particularly since Riley et al. duck the problem too. So I will give you a brief tour. The question revolves around any Fourier series expansion of a function with a (finite number of) simple jump discontinuities. We saw them in Figure 2, page 6 because the function $f(x) = x$ is not continuous on a circle, and so there is an effective jump between $x = -\pi$ and $x = \pi$. Here I examine the simplest case. We work on the interval $(0, +1)$ and use the complex basis $e_n(x) = \mathrm{e}^{2\pi i n x}$ with $n = 0, \pm 1, \pm 2, \cdots$ which is complete, and we have $\| e_n \| = 1$.

First we write out the expansion of a function f, taking the terms out to $|n| = N$:

$$S_N(f) = \sum_{n=-N}^{N} c_n e_n(x) = \sum_{n=-N}^{N} (f, e_n)\, e_n(x) \tag{6.1}$$

$$= \sum_{n=-N}^{N} \left(\int_0^1 f(y)\, \mathrm{e}^{-2\pi i n y}\, dy \right) \mathrm{e}^{2\pi i n x} \tag{6.2}$$

$$= \int_0^1 dy \left(f(y) \sum_{n=-N}^{N} \mathrm{e}^{2\pi i n (y-x)} \right) = \int_0^1 dy\, f(y)\, D_N(y-x)\,. \tag{6.3}$$

The integral is an example of a **convolution,** an operation we will see more of with Fourier transforms. We can evaluate the function D_N:

$$D_N(u) = \sum_{n=-N}^{N} \mathrm{e}^{2\pi i n u} = \mathrm{e}^{-2\pi i N} \sum_{n=0}^{2N} \mathrm{e}^{2\pi i \nu}\,. \tag{6.4}$$

The sum now is just $2N+1$ terms of a geometric series (like $1 + x + x^2 + \cdots x^{2N}$) whose sum is $(x^{2N+1} - 1)/(x-1)$. Hence:

$$D_N(u) = \mathrm{e}^{-2\pi i N u} \frac{\mathrm{e}^{2\pi i (2N+1) u} - 1}{\mathrm{e}^{2\pi i u} - 1} \tag{6.5}$$

$$= \mathrm{e}^{-2\pi i N u} \frac{\mathrm{e}^{\pi i (2N+1) u} (\mathrm{e}^{\pi i (2N+1) u} - \mathrm{e}^{-\pi i (2N+1) u})}{\mathrm{e}^{\pi i u} (\mathrm{e}^{\pi i u} - \mathrm{e}^{-\pi i u})} \tag{6.6}$$

$$= \mathrm{e}^{-2\pi i N + \pi i (2N+1) u - \pi i u} \frac{(\mathrm{e}^{\pi i (2N+1) u} - \mathrm{e}^{-\pi i (2N+1) u})/2i}{(\mathrm{e}^{\pi i u} - \mathrm{e}^{-\pi i u})/2i} \tag{6.7}$$

$$= \frac{\sin \pi (2N+1) u}{\sin \pi u}\,. \tag{6.8}$$

In the present context this function is called the **Dirichlet kernel.** We are interested in $D_N(x)$ for large values of N, when we have taken a lot of terms in (6.1). A graph of the Dirichlet kernel is shown in Figure 5, with $N = 25$.

Returning to (6.3), let us now choose a particular function for f. We will treat the Heavisde function, in this case, a step at $x = ½$: let

$$f(x) = H(x - ½) = \begin{cases} 0, & x < ½ \\ 1, & x \geq ½ . \end{cases} \tag{6.9}$$

Let us call the sum for this function $Q_N(x)$; then

$$Q_N(x) = \int_0^1 H(y - ½)\, D_N(y)\, dy = \int_{½}^1 D_N(y - x)\, dy . \tag{6.10}$$

We are interested in the behavior near the jump at $x = ½$, so we write $x = \xi + ½$ and

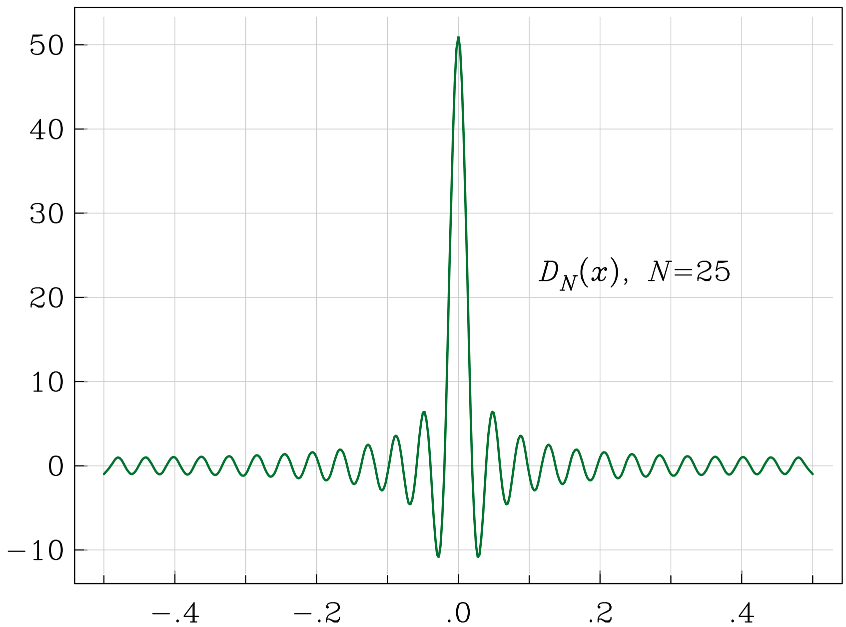
$$Q_N(\xi + ½) = \int_{½}^1 D_N(y - \xi - ½)\, dy = \int_{½}^1 \frac{\sin \pi(2N+1)(y - \xi - ½)}{\sin \pi(y - \xi - ½)}\, dy . \tag{6.11}$$

We make the change of variable: $t = \pi(2N+1)(y - \xi - ½)$; then

$$Q_N(\xi + ½) = \int_{-(2N+1)\pi\xi}^{(2N+1)\pi(½ - \xi)} \frac{\sin t}{t} \cdot \frac{t/(2N+1)}{\sin t/(2N+1)}\, \frac{dt}{\pi} . \tag{6.12}$$

Up to this point we have made no approximations whatever! Now we let N become very large and ξ be small so that $N\xi \sim 1$. Then the second factor in the integrand tends to one since $t/(2N+1)$ is of the order of ξ which is small, and we get the approximation:

Figure 5: The Dirichlet kernel



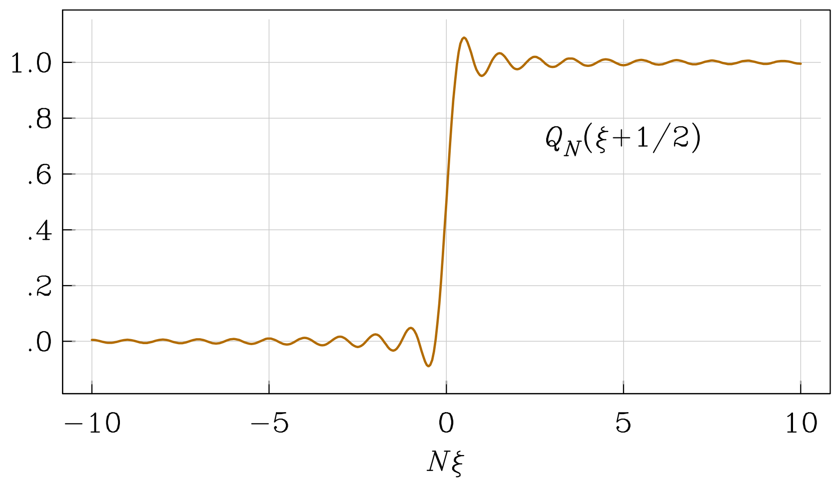
$$\tilde{Q}_N(\xi + ½) = \frac{1}{\pi} \int_{-2\pi N\xi}^{\infty} \frac{\sin t}{t}\, dt = \frac{1}{\pi} \left[\int_0^{\tau} + \int_0^{\infty} \right] \frac{\sin t}{t}\, dt \tag{6.13}$$

$$= ½ + \frac{1}{\pi} \int_0^{\tau} \frac{\sin t}{t}\, dt \tag{6.14}$$

where I have written $\tau = 2\pi N\xi$, and I have used the fact that $\sin t/t$ is even in t. I have also used the fact that the integral on $(0, \infty)$ is $½\pi$ something we will show later on. The integral of $\sin t/t$ is called $si(t)$ and is a commonly used function – in MATLAB it is called sinint. I plot (6.14) in Figure 6 below. Notice how the result depends on the number of terms N only on the way the x axis is scaled. As N becomes large the picture is compressed in x, but its fundamental shape never changes: there is always an overshoot and an undershoot of the same magnitude, which turns out to be about 8.94 percent. This means that the maximum difference between the original step function and the sum of the series does not tend to zero with increasing N; in other words the Fourier series $S(f)$ is not uniformly convergent for this function.

The same analysis applies at any jump discontinuity, not just the unit step, and not just at $x = ½$.

Figure 6: Gibbs' overshoot function.



7. Geophysical Examples of Orthogonal Functions

Any time one encounters a periodic phenomenon, a Fourier series is appropriate. Tides might seem to be an example, but really they are not because the different tidal frequencies are not multiples of some fundamental frequency. There are plenty of Fourier series in signal processing: for example in electromagnetic sounding in the marine environment, Steve Constable uses a *square wave* signal in the source electric current, which is analyzed in terms of its separate components as sine waves; see equation (2.23).

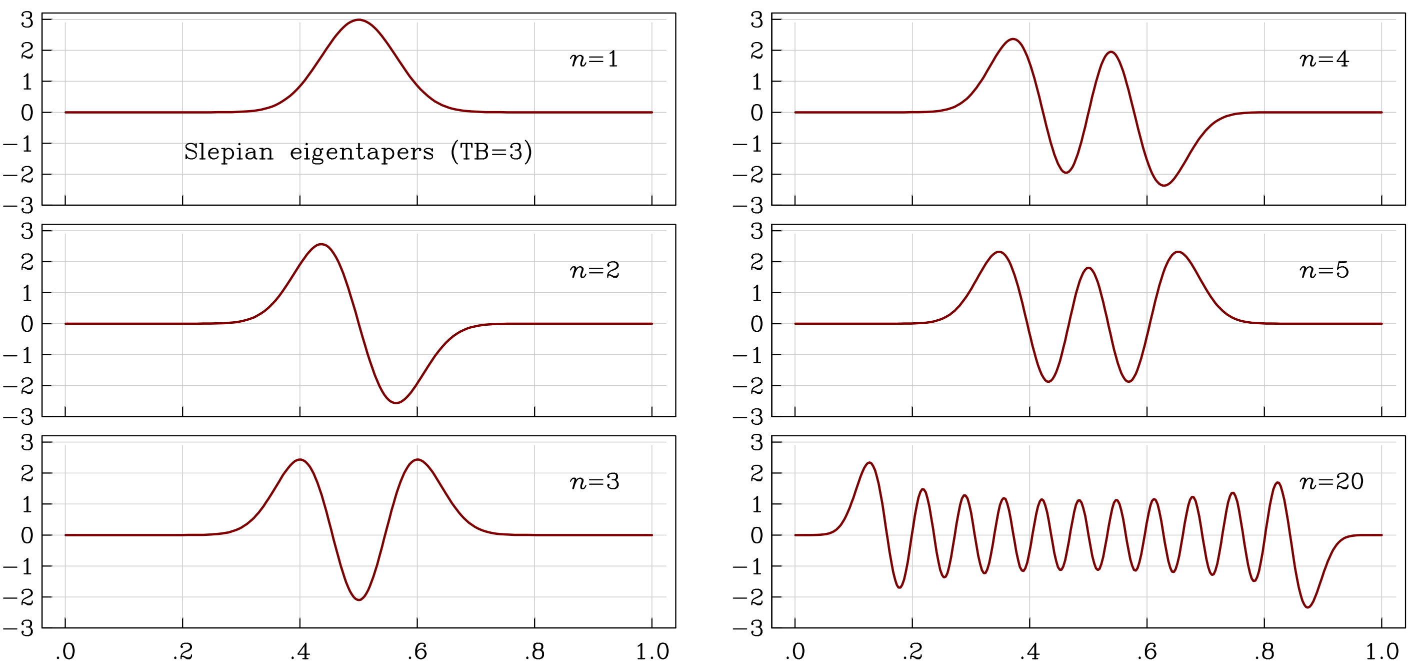
Another signal processing example is a set of orthogonal functions used in the estimation of power spectra. The power spectrum (or, PSD for power spectral density) is the means of representing the frequency content of a random signal. The modern way of computing the PSD involves first multiplying the original time series by a succession of orthogonal functions, called **prolate spheroidal wavefunctions**, eigenfunctions of a certain self adjoint operator. These will be covered in the Geophysical Data Analysis class. They are illustrated below.

The most common set of orthogonal functions in geophysics are the **spherical harmonics.** We can easily define an inner product and norm using integrals over regions other than the real line, for example, over the surface of the sphere. This gives rise to the space $L_2(S^2)$, where S^2 is just a symbol for the surface of the unit sphere. If we write the inner product and norm for complex-valued functions as

$$(f, g) = \int_{S^2} f(\hat{\mathbf{r}})\, g(\hat{\mathbf{r}})^*\, d^2\hat{\mathbf{r}}, \quad \| f \|^2 = \int_{S^2} | f(\hat{\mathbf{r}}) |^2\, d^2\hat{\mathbf{r}} \tag{7.1}$$

then we can build orthogonal functions on the sphere. The most common of these are the spherical harmonics, eigenfunctions of the self adjoint operator ∇_1^2; they are the equivalent of the Fourier series on a sphere. Gravitational and magnetic fields around the Earth are always expanded in spherical harmonics. In seismology, the

Figure 7: Prolate spheroidal, or Slepian, functions.



free oscillations of the Earth as a whole are computed and classified in terms of their expansions as spherical harmonics. Almost every quantity considered globally is decomposed into its spherical harmonic series, including topography, heatflow, surface temperature, etc:

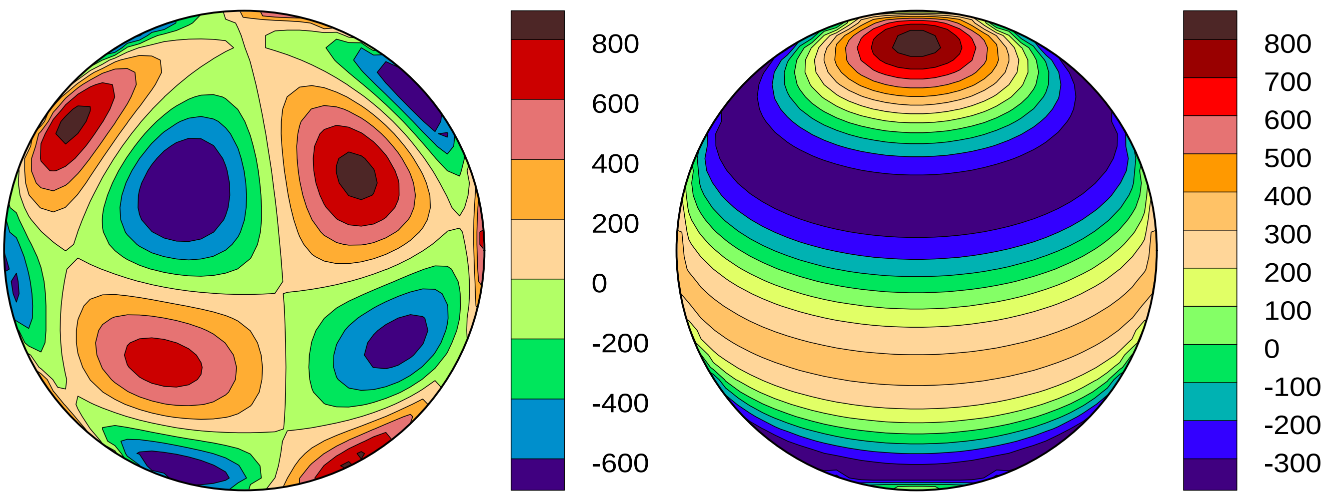
$$f(\hat{\mathbf{r}}) = \sum_{l=0}^{\infty} \sum_{m=-l}^{l} c_{lm} \, Y_l^m(\hat{\mathbf{r}}) \tag{7.2}$$

$$c_{lm} = (f, Y_l^m) \,. \tag{7.3}$$

Notice there are two counters, l and m, instead of the familiar single index in the expansion for reasons we will not go into here.

Below we illustrate two typical spherical harmonic basis functions. As illustrated on the right, when the index $m = 0$, we obtain axisymmetric functions; these functions are can be expressed as the Legendre polynomials defined in (2.25): $Y_l^0 = P_l(\cos\theta)$ where θ is the colatitude.

Figure 8: Two of the spherical harmonic basis functions.



l=5, m=3

l=4, m=0

8. The Reproducing Kernel

The picture we saw in the discussion of Gibbs' phenomenon (Section 6) has a mild generalization worth mentioning here. Consider any complete set of orthogonal functions, arranged in a specific order in the usual way: $g_1, g_2, \ldots g_n \cdots$. Also we will take these to be of unit norm: $\| g_n \| = 1$. Now consider forming the following function from the first N of them:

$$K_N(x, y) = \sum_{n=1}^{N} g_n(x)\, g_n(y)\,. \tag{8.1}$$

This is known as the **reproducing kernel** because of the following property. Take the inner product with some other arbitrary element in the Hilbert space, say f which varies in the argument x. Then

$$(K_N, f) = (\sum_{n=1}^{N} g_n(x)\, g_n(y),\, f(x)) \tag{8.2}$$

$$= \sum_{n=1}^{N} (g_n, f)\, g_n(y) \tag{8.3}$$

$$= \sum_{n=1}^{N} c_n\, g_n(y) = S_N(f)\,. \tag{8.4}$$

In other words, the inner product (8.2) is the version of the function f obtained by truncating the orthogonal expansion (2.1) to N terms.

What this means is that for any fixed position y the function $K_N(x, y)$ looks, as N becomes large, like a narrower and narrower spike concentrated at y, and it is therefore an approximation to a delta function. We have seen how this looks for the Fourier basis in Section 6, the Dirichlet kernel. Exactly the same behavior will be found for *any* complete expansion. For example, for a Hilbert space based on a sphere, the spherical harmonics provide a reproducing kernel, which can be used for deriving the *Spherical Harmonic Addition Theorem.*

When the g_n are orthogonal functions arising as the eigenfunctions of a self-adjoint differential operator, then K_N can be used to derive **Green's function** for the operator since we can write:

$$L_x G_N(x, y) = K_N(x, y) \tag{8.5}$$

and upon expanding G_N in the eigenbasis we easily see that

$$G_N(x, y) = \sum_{n=1}^{N} \frac{g_n(x)\, g_n(y)}{\lambda_n} \tag{8.6}$$

and Green's function results as $N \to \infty$.