FOURIER SERIES – AN APPLICATION OF ORTHONORMAL BASES

The point of these notes is to discuss how the concept of orthogonality gets used in signal processing. One should think of there are being two motivating problems:

Motivating Question 1 There are probably twenty or thirty radio stations transmitting in the Ann Arbor area. When I turn my radio to a particular station, how does it pick out that signal from all the others?

Motivating Question 2 I want to store a sound file but I have a limited amount of space to do so. What are the most important aspects of the sound to hold onto?

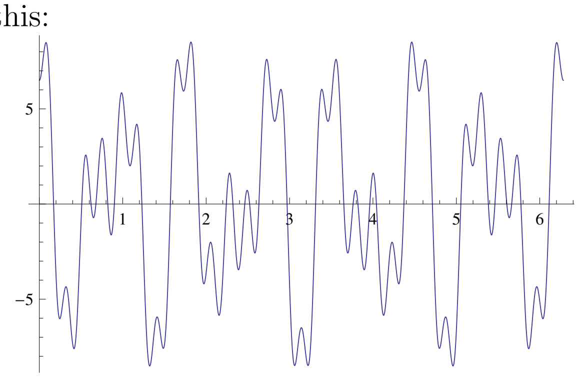
We'll start by discussing Motivating Question 1, and move to discussing Motivating Question 2 later.

1. Breaking a sum of cosines up into pieces

We'll start with the following toy model for Question 1. Suppose that one station is broadcasting the signal $a_7 \cos(7\theta)$, another is broadcasting $a_{11} \cos(11\theta)$ and a third is broadcasting $a_{29} \cos(29\theta)$. All our radio can see is the sum

$$h(\theta) = a_7 \cos(7\theta) + a_{11} \cos(11\theta) + a_{29} \cos(29\theta).$$

It looks something like this:



How can we extract the coefficients a_7, a_{11} and a_{29} from the signal h?

To answer this question, we introduce the following notation: For functions f and g on $[0, 2\pi]$, we set

$$\langle f, g \rangle = \frac{1}{\pi} \int_{\theta=0}^{2\pi} f(\theta) g(\theta) d\theta.$$

A key observation: If j and k are two positive integers, then

$$\langle \cos(j\theta), \cos(k\theta) \rangle = \begin{cases} 1 & \text{if } j = k \\ 0 & \text{if } j \neq k \end{cases}. \qquad (1)$$

Let's check this.

$$\begin{aligned} \langle \cos(j\theta), \cos(k\theta) \rangle &= \tfrac{1}{\pi} \int_0^{2\pi} \cos(j\theta) \cos(k\theta) d\theta \\ &= \tfrac{1}{\pi} \int_0^{2\pi} \tfrac{\cos((j-k)\theta) + \cos((j+k)\theta)}{2} d\theta \\ &= \tfrac{1}{2\pi} \int_0^{2\pi} \cos((j-k)\theta) d\theta + \tfrac{1}{2\pi} \int_0^{2\pi} \cos((j+k)\theta) d\theta \end{aligned}$$

The second integral is always 0, as it is the average value of $\cos((j+k)\theta)$ over a full cycle. If $j \neq k$, the first integral is 0 for the same reason. If $j = k$, then the term inside the first integral is $\cos(0) = 1$, so the integral is 1.

So we have

$$\begin{aligned}\langle h(\theta), \cos 7\theta\rangle &= a_7\langle\cos(7\theta), \cos(7\theta)\rangle + a_{11}\langle\cos(11\theta), \cos(7\theta)\rangle + a_{29}\langle\cos(29\theta), \cos(7\theta)\rangle \\ &= a_7 \cdot 1 + a_{11} \cdot 0 + a_{29} \cdot 0 \\ &= a_7\end{aligned}$$

Similarly

$$\langle h(\theta), \cos(11\theta)\rangle = a_{11} \quad \text{and} \quad \langle h(\theta), \cos(29\theta)\rangle = a_{29}$$

In general:

(2) If $\quad h(\theta) = a_1 \cos\theta + a_2\cos(2\theta) + a_3\cos(3\theta) + \cdots + a_n\cos(n\theta) \quad$ then $\quad a_k = \langle h(\theta), \cos(k\theta)\rangle.$

So we can recover the coefficients of the various cosines from the signal h by taking this integral.

Comments on the realism of the model: If we are talking about an AM radio signal, this isn't too crazy. The signal put out by an AM station at frequency k is $a(t)\cos(kt)$, where $a(t)$ is a function which varies on a much slower time scale than the radio wave. (AM radio uses frequencies of 500 to $1000kHz$. The middle of a piano is $440Hz$ – one thousand times slower. On the time scale with which the radio signal oscillates, the sound wave is essentially constant.) We talked about the case where the coefficient of $\cos(kt)$ is actually a constant; the case where it is varying extremely slowly is not a big difference. One way in which our discussion was not accurate was that we suggested averaging over one period, from 0 to 2π. In practice, it is better to integrate over many periods and average. (The other inaccuracy, when talking about radios specifically, is that radio's don't compute integrals in the sense of having a chip which does numeric integration; they use a resonant electric circuit which has the effect of computing the integral. In other applications, we actually do compute the integral.)

FM Radio is more complicated.

2. Why are we talking about this in this class?

The quantity $\langle f, g\rangle$ acts a lot like dot product:

$$\begin{aligned}\langle f, g+h\rangle &= \langle f, g\rangle + \langle f, h\rangle \\ \langle f+g, h\rangle &= \langle f, h\rangle + \langle g, h\rangle \\ \langle f, kg\rangle &= k\langle f, g\rangle = \langle f, kg\rangle \\ \langle f, g\rangle &= \langle g, f\rangle \\ \langle f, f\rangle &\geq 0\end{aligned}$$

The difference is that $\langle\ ,\ \rangle$ works with functions rather than vectors. In a more sophisticated course, you would learn language to make this not a big deal. (Our text covers this language in Chapters 4 and 5.5.) For now, let's just go with it.

If we think of $\langle\ ,\ \rangle$ as like dot product, then equation (1) says that the functions $\cos(k\theta)$ are orthonormal. And (2) says that, if h is in the span of the functions $\cos(\theta)$, $\cos(2\theta)$, ..., then we can find the coefficients of the various cosines by using the same method we do for orthonormal functions.

3. Getting a basis for the space of functions

The cosines do not span all functions on $[0, 2\pi]$. For example, any linear combination of cosines has average 0, so we can't get functions with nonzero average this way.

We can do better by using, cosines, sines and constants. Consider the list of functions $\cos(\theta)$, $\cos(2\theta)$, $\cos(3\theta)$, …, $\sin(\theta)$, $\sin(2\theta)$, $\sin(3\theta)$, …, $1/\sqrt{2}$. We can check that these are, again, orthonormal.

Saying what a basis means in this setting is complicated. But, roughly, these functions are a basis. In other words, if $h(\theta)$ is any "reasonable" function, and we set

$$\begin{aligned} a_k &= \langle h(\theta), \cos(k\theta)\rangle \\ b_k &= \langle h(\theta), \sin(k\theta)\rangle \\ c &= \langle h(\theta), 1/\sqrt{2}\rangle \end{aligned}$$

then

$$h(\theta) = \sum_{k=1}^{\infty} a_k \cos(k\theta) + \sum_{k=1}^{\infty} b_k \sin(k\theta) + c/\sqrt{2}.$$

For the record, we write this out without the $\langle\ ,\rangle$ notation, and clean up the constant term a little. Let

$$\begin{aligned} a_k &= \tfrac{1}{\pi}\int_0^{2\pi} h(\theta)\cos(k\theta)d\theta \\ b_k &= \tfrac{1}{\pi}\int_0^{2\pi} h(\theta)\sin(k\theta)d\theta \\ d &= \tfrac{1}{2\pi}\int_0^{2\pi} h(\theta)d\theta \end{aligned}$$

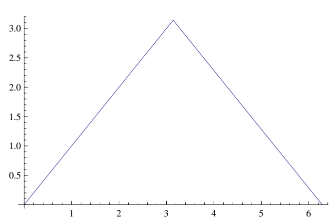
Then

$$h(\theta) = \sum_{k=1}^{\infty} a_k \cos(k\theta) + \sum_{k=1}^{\infty} b_k \sin(k\theta) + d.$$

Here's a simple example. Let

$$h(\theta) = \begin{cases} \theta & 0 \le \theta \le \pi \\ 2\pi - \theta & \pi \le \theta \le 2\pi \end{cases}.$$

Here is a picture of h.



We can compute

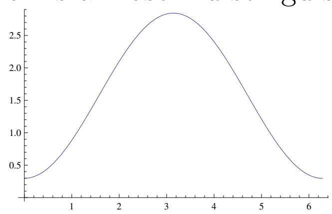
$$\begin{aligned} a_k &= \tfrac{1}{\pi}\int_0^{2\pi} h(\theta)\cos(k\theta)d\theta = \\ &= \tfrac{1}{\pi}\int_0^{\pi} \theta\cos(k\theta)d\theta + \tfrac{1}{\pi}\int_{\pi}^{2\pi}(2\pi-\theta)\cos(k\theta)d\theta \\ &= \begin{cases} -\frac{4}{\pi k^2} & k \text{ odd} \\ 0 & k \text{ even} \end{cases} \end{aligned}$$

(The last integrals are computed by integrating by parts; we've left out the details.)

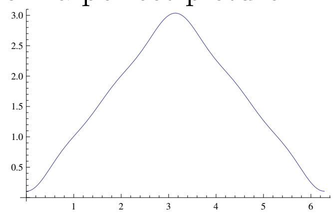
Also, $b_k = 0$ and $d = \pi/2$. We deduce that we should have

$$h(\theta) = -\frac{4}{\pi}\sum_{m=0}^{\infty} \frac{\cos((2m+1)\theta)}{(2m+1)^2} + \frac{\pi}{2}.$$

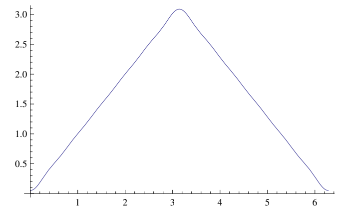
The table below shows the effects of adding in more and more times of the sum. Just adding 3 cosine terms is already a very good approximation; 6 terms gives us something which is almost indistinguishable from a perfect picture.



$\frac{\pi}{2} - \frac{4}{\pi}\cos\theta$



$\frac{\pi}{2} - \frac{4}{\pi}\cos\theta - \frac{4}{9\pi}\cos(3\theta) - \frac{4}{25\pi}\cos(5\theta)$



$\frac{\pi}{2} - \frac{4}{\pi}\sum_{m=0}^{5} \frac{\cos(2m+1)\theta}{(2m+1)^2}$

4. Storing a signal efficiently

Now, let's talk about Question 2. Suppose we have a signal $h(\theta)$, and we want to store only the most important aspects of it. The pictures we just saw suggest an idea: Compute the Fourier coefficients of h, and just store the first few.

We don't have to use sines and cosines for this idea. We could use any collection of orthogonal functions. For example, MP3 files use a collection of orthogonal functions known as "wavelets" to get pretty good sound with very little storage.

By contrast, a WAV file actually stores the values of the function $h(\theta)$, sampled $44,100$ times a second. Notice that a WAV file is usually about 8 to 10 times larger than the corresponding MP3 file. Storing coefficients in an orthonormal basis allows pretty good reproduction with much less data.

The same principles can be used for storing visual images. However, there are many other complications here, and none of the standard image formats is doing anything as simple as simply storing Fourier coefficients or simply storing function values.

5. Orthogonal polynomials

Suppose that we are given a function $f(x)$ and we want to find a low degree polynomial which is pretty close to $f(x)$ on the interval $(-1,1)$. To be concrete, let's try to find a quadratic, $a + bx + cx^2$ which is as close as possible to e^x on $(-1,1)$.

We redefine $\langle\ ,\ \rangle$ to be

$$\langle f(x), g(x)\rangle = \frac{1}{2}\int_{-1}^{1} f(x)g(x)dx.$$

In order to proceed, we want an orthonormal basis for the vector space of quadratic polynomials. There is an obvious basis for the set of quadratic polynomials: Namely, 1, x and x^2. This basis is NOT orthonormal: Notice that, for example, $\langle 1, x^2\rangle = (1/2)\int_{-1}^{1} x^2 dx = 1/3$, not 0.

But we know how to convert a non-orthonormal basis into an orthonormal one: The Gram-Schmit algorithm. Let's do it:

The function 1 already has $\langle 1, 1, \rangle = 1$.

We already have $\langle x, 1\rangle = 0$, so x is orthogonal to 1. However, $\langle x, x\rangle = (1/2)\int_{-1}^{1} x^2 dx = 1/3$, not 1. So the "vector" x doesn't have length 1. So we rescale it to get $\sqrt{3}x$.

We now project x^2. I leave out the details of computing the integrals.

$$x^2-\langle x^2,1\rangle 1-\langle x^2,x\rangle x = x^2-\left(\frac{1}{2}\int_{-1}^{1} x^2 dx\right)1-\left(\frac{1}{2}\int_{-1}^{1} x\cdot x^2 dx\right)x = x^2-1/3-0\cdot x = x^2-1/3.$$

Finally, we must rescale. We have

$$\langle x^2-1/3, x^2-1/3\rangle = \frac{1}{2}\int_{-1}^{1}(x^2-1/3)^2 dx = \frac{4}{45}.$$

So we rescale to get a vector of length 1, namely, $\sqrt{45/4}(x^2-1/3)$.

Our orthonormal basis is

$$1, \quad \sqrt{3}x, \quad \sqrt{\frac{45}{4}}(x^2-1/3).$$

Let's try this with the example $f(x)=e^x$. We have

$$\begin{aligned} \langle e^x, 1\rangle &= \tfrac{1}{2}\int_{-1}^{1} e^x dx &&= 1.1752 \\ \langle e^x, \sqrt{3}x\rangle &= \tfrac{1}{2}\int_{-1}^{1} e^x \sqrt{3}x dx &&= 0.6372 \\ \langle e^x, \sqrt{\tfrac{45}{4}}(x^2-1/3)\rangle &= \tfrac{1}{2}\int_{-1}^{1} e^x \sqrt{\tfrac{45}{4}}(x^2-1/3)dx &&= 0.1600 \end{aligned}$$

So our approximation is

$$1.1752 + 0.6372\left(\sqrt{3}x\right) + 0.1600\left(\sqrt{\frac{45}{4}}(x^2-1/3)\right).$$

In the left hand figure, we have plotted e^x in green and the above polynomial in red. As you can see, they are almost on top of one another. In blue, I have plotted the Taylor series $1+x+x^2/2$ – pretty good, but not nearly as close. On the right, I have plotted the error |true value − approximation| using orthogonal polynomials (red) and using Taylor series (blue). You can see that orthogonal polynomials produce half the error overall, although Taylor series are better near the center of the range.