Any typographical or other corrections about these notes are welcome.

1 Review of inner products

An **inner product space** is a vector space V together with a choice of inner product. Recall that an inner product must be bilinear, symmetric, and positive definite. Since it is positive definite, the quantity $\langle \vec{u}, \vec{u} \rangle$ is never negative, and is never 0 unless $\vec{v} = \vec{0}$. Therefore its square root is well-defined; we define the **norm** of a vector $\vec{u} \in V$ to be

$$\|\vec{u}\| = \sqrt{\langle \vec{u}, \vec{u} \rangle}.$$

Observe that the norm of a vector is a nonnegative number, and the only vector with norm 0 is the zero vector $\vec{0}$ itself.

In an inner product space, we call two vectors $\vec{u}, \vec{v}$ **orthogonal** if $\langle \vec{u}, \vec{v} \rangle = 0$. We will also write $\vec{u} \perp \vec{v}$ as a shorthand to mean that $\vec{u}, \vec{v}$ are orthogonal.

Because an inner product must be bilinear and symmetry, we also obtain the following expression for the squared norm of a sum of two vectors, which is analogous the to law of cosines in plane geometry.

$$\begin{aligned}
\|\vec{u}+\vec{v}\|^2 &= \langle \vec{u}+\vec{v}, \vec{u}+\vec{v} \rangle \\
&= \langle \vec{u}+\vec{v}, \vec{u} \rangle + \langle \vec{u}+\vec{v}, \vec{v} \rangle \\
&= \langle \vec{u}, \vec{u} \rangle + \langle \vec{v}, \vec{u} \rangle + \langle \vec{u}, \vec{v} \rangle + \langle \vec{v}, \vec{v} \rangle \\
&= \|\vec{u}\|^2 + \|\vec{v}\|^2 + 2\langle \vec{u}, \vec{v} \rangle .
\end{aligned}$$

In particular, this gives the following version of the Pythagorean theorem for inner product spaces.

Pythagorean theorem for inner products

If $\vec{u}, \vec{v}$ are orthogonal vectors in an inner product space, then

$$\|\vec{u}+\vec{v}\|^2 = \|\vec{u}\|^2 + \|\vec{v}\|^2.$$

Proof. This follows from the equation $\|\vec{u}+\vec{v}\|^2 = \|\vec{u}\|^2 + \|\vec{v}\|^2 + 2\langle \vec{u}, \vec{v} \rangle$ (found above), together with the fact that $\langle \vec{u}, \vec{v} \rangle = 0$ (by defintion of orthogonality). □

2 The approximation problem and its solution via orthogonality

These notes are concerned with a process for solving the following problem, which is analogous (indeed, a generalization of) the least-squares problem that we discussed a couple weeks ago.

The approximation problem in inner product spaces

Given a list of vectors $\{\vec{v}_1, \cdots, \vec{v}_n\}$ in an inner product space and an additional vector $\vec{b}$, what linear combination of $\{\vec{v}_1, \cdots, \vec{v}_n\}$ is closest to $\vec{b}$? In other words, which values $c_1, \cdots, c_n$ minimize the quantity

$$\left\| \sum_{i=1}^{n} c_i \vec{v}_i - \vec{b} \right\|?$$

As in the case of the least-squares problem, the solution of this approximation problem is intimately tied with the idea of orthogonality. The problem can be reduced to solve a linear system of equations, once we have the following observation: the best possible linear combination is the one that makes the difference to $\vec{b}$ orthogonal to all of the vectors at our disposal. This is made precise in the following theorem.

Main theorem on approximation

Suppose that we are given $S = \{\vec{v}_1, \cdots, \vec{v}_n\}$ and $\vec{b}$ in an inner product space. If a linear combination $\vec{v}$ of S satisfies the orthogonality conditions

$$\vec{v}_i \perp (\vec{v} - \vec{b})$$

for all i, then $\vec{v}$ minimizes $\|\vec{v} - \vec{b}\|$ (among all possible linear combinations of S).

The proof of this theorem follows the same method as the analogous theorem from the notes on least-squares. Indeed, it is the same proof; we simply replace any reference to a dot product with a use of whatever inner product is in use.

Proof. Assume that $\vec{v}$ is a linear combination of S satisfying the given orthogonality conditions. Let $\vec{w}$ be *any other* linear combination of S (i.e. $\vec{w}$ is a linear combination of $\{\vec{v}_1, \cdots, \vec{v}_n\}$, and $\vec{w} \neq \vec{v}$). Then $\vec{w} - \vec{v}$ is a linear combination of $\{\vec{v}_1, \cdots, \vec{v}_n\}$. Since each $\vec{v}_i$ is orthogonal to $\vec{v} - \vec{b}$, and the inner product is bilinear, it follows that $\left\langle (\vec{w} - \vec{v}), \vec{v} - \vec{b} \right\rangle = 0$, i.e.

$$(\vec{w} - \vec{v}) \perp (\vec{v} - \vec{b}).$$

By the Pythagorean theorem for inner product spaces, it follows that

$$\begin{aligned} \|\vec{w} - \vec{b}\|^2 &= \|(\vec{w} - \vec{v}) + (\vec{v} - \vec{b})\|^2 \\ &= \|\vec{w} - \vec{v}\|^2 + \|\vec{v} - \vec{b}\|^2. \end{aligned}$$

Now, since $\vec{w} \neq \vec{v}$, it follows (from positive definiteness of the inner product) that $\|\vec{w} - \vec{v}\|^2 > 0$. Therefore

$$\|\vec{w} - \vec{b}\|^2 > \|\vec{v} - \vec{b}\|^2.$$

Therefore, $\vec{w}$ is further from $\vec{b}$ than $\vec{v}$ is, so indeed $\vec{v}$ is the closest possible linear combination of $\{\vec{v}_1, \cdots, \vec{v}_n\}$ to $\vec{b}$. □

What makes this theorem so powerful is that this list of orthogonality conditions amounts to a *linear system of equations.* Indeed, if we write

$$\vec{v} = c_1\vec{v}_1 + c_2\vec{v}_2 + \cdots + c_n\vec{v}_n,$$

then the condition $\vec{v}_i \perp (\vec{v} - \vec{b})$ can be written

$$\begin{aligned} \left\langle \vec{v}_i, c_1\vec{v}_1 + c_2\vec{v}_2 + \cdots + c_n\vec{v}_n - \vec{b} \right\rangle &= 0 \\ \langle \vec{v}_i, c_1\vec{v}_1 + c_2\vec{v}_2 + \cdots + c_n\vec{v}_n \rangle &= \left\langle \vec{v}_i, \vec{b} \right\rangle \\ c_1 \langle \vec{v}_i, \vec{v}_1 \rangle + c_2 \langle \vec{v}_i, \vec{v}_2 \rangle + \cdots + \cdots + c_n \langle \vec{v}_i, \vec{v}_n \rangle &= \left\langle \vec{v}_i, \vec{b} \right\rangle, \end{aligned}$$

which constitutes a linear equation in the variables $c_1, c_2, \cdots c_n$. One way to write the resulting linear system of equations, analogous to the matrix form of the normal equation (from our discussion of least-squares) is the following.

$$\begin{pmatrix} \langle \vec{v}_1, \vec{v}_1 \rangle & \langle \vec{v}_1, \vec{v}_2 \rangle & \cdots & \langle \vec{v}_1, \vec{v}_n \rangle \\ \langle \vec{v}_2, \vec{v}_1 \rangle & \langle \vec{v}_2, \vec{v}_2 \rangle & \cdots & \langle \vec{v}_2, \vec{v}_n \rangle \\ \vdots & \vdots & & \vdots \\ \langle \vec{v}_n, \vec{v}_1 \rangle & \langle \vec{v}_n, \vec{v}_2 \rangle & \cdots & \langle \vec{v}_n, \vec{v}_n \rangle \end{pmatrix} \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{pmatrix} = \begin{pmatrix} \left\langle \vec{v}_1, \vec{b} \right\rangle \\ \left\langle \vec{v}_2, \vec{b} \right\rangle \\ \vdots \\ \left\langle \vec{v}_n, \vec{b} \right\rangle \end{pmatrix}$$

3 Examples: approximating functions with polynomials

One fairly common situation in practice is curve-fitting: you may have some data from a measuring instrument or laboratory that can be graphed as a function of one variable, and perhaps that data is noisy in some way. it is often convenient to approximate the resulting curve using a low-degree polynomial. The machinery of the previous section gives a way to find a polynomial of "best fit." The examples in this section show one way that this could be done.

These examples take place in the inner product space $\mathcal{C}[0,1]$ (this is out notation from class; the book writes $\mathcal{C}^{(0)}[0,1]$)), with the inner product

$$\langle f, g \rangle = \int_0^1 f(x)g(x)\ dx.$$

Here the approximation problem can be understood as follows: we are looking to approximate a given function as a linear combination of some other previously chosen functions, where we measure the accuracy of an approximation by the integral of the squared difference between the functions.

Example 1

Find the linear polynomial $p(x) = c_1 + c_2 x$ that best approximates $f(x) = e^x$ on the interval $[0,1]$. More precisely, $p(x)$ should minimize the value of the integral

$$\int_0^1 (p(x) - f(x))^2\ dx = \int_0^1 (c_1 + c_2 x - e^x)^2\ dx.$$

Solution to Example 1

We can map this problem onto the notation of the previous sections as follows. The target we wish to approximate is e^x; this is the vector $\vec{b}$. The vectors we wish to use to approximate it are 1 and x (regarded as "vectors" in the vector space $\mathcal{C}[0,1]$), so these are $\vec{v}_1, \vec{v}_2$. Thus we obtain the following pair of orthogonality conditions from the main theorem:

$$\begin{aligned} 1 &\perp c_1 + c_2 x - e^x \\ x &\perp c_1 + c_2 x - e^x \end{aligned}$$

which translate into the following pair of linear equations:

$$\begin{aligned} c_1 \langle 1, 1 \rangle + c_2 \langle 1, x \rangle &= \langle 1, e^x \rangle \\ c_2 \langle x, 1 \rangle + c_2 \langle x, x \rangle &= \langle x, e^x \rangle . \end{aligned}$$

We can evaluate the inner products involved as follows (note: on homework for this class, feel free to use a tool such as Wolfram Alpha from a web browser, or Mathematica if you have it installed, to evaluate such integrals):

$$\begin{aligned} \langle 1, 1 \rangle = \int_0^1 dx &= 1 \\ \langle 1, x \rangle = \int_0^1 x \, dx &= \frac{1}{2} \\ \langle x, 1 \rangle = \int_0^1 x \, dx &= \frac{1}{2} \\ \langle x, x \rangle = \int_0^1 x^2 \, dx &= \frac{1}{3} \\ \langle 1, e^x \rangle = \int_0^1 e^x \, dx &= -e + 1 \\ \langle x, e^x \rangle - \int_0^1 x e^x &= 1 \end{aligned}$$

Therefore we must solve the linear system

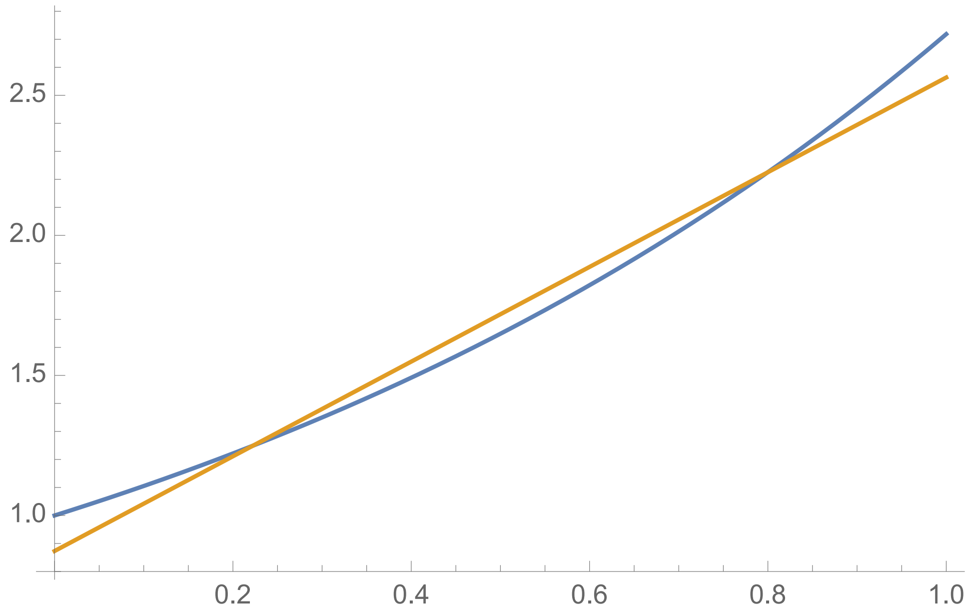
$$\begin{aligned} c_1 + \frac{1}{2} c_2 &= -1 + e \\ \frac{1}{2} c_1 + \frac{1}{3} c_2 &= 1. \end{aligned}$$

This can be done using matrix inversion or row reduction, the result is

$$c_1 = -10 + 4e \approx 0.873, \quad c_2 = 18 - 6e \approx 1.690$$

Thus the best-fit line is $g(x) = (-10 + 4e) + (18 - 6e)x$.

The figure below shows a plot of $y = e^x$ and the best-fit line $y = p(x)$ found above.



Example 2

Find the *quadratic* polynomial $p(x) = c_1 + c_2x + c_3x^2$ the minimizes

$$\int_0^1 (p(x) - e^x)^2 \, dx.$$

Solution to Example 2

We process similarly to before. Now we are taking a linear combination of three functions: $1, x, x^2$. To keep the solution brief, we go straight to the matrix equation described at the end of the previous section. In this case, this equation is

$$\begin{pmatrix} \langle 1,1\rangle & \langle 1,x\rangle & \langle 1,x^2\rangle \\ \langle x,1\rangle & \langle x,x\rangle & \langle x,x^2\rangle \\ \langle x^2,1\rangle & \langle x^2,x\rangle & \langle x^2,x^2\rangle \end{pmatrix} \begin{pmatrix} c_1 \\ c_2 \\ c_2 \end{pmatrix} = \begin{pmatrix} \langle 1,e^x\rangle \\ \langle x,e^x\rangle \\ \langle x^2,e^x\rangle \end{pmatrix}.$$

Using Mathematica to compute all of these integrals, this matrix equation becomes the following.

$$\begin{pmatrix} 1 & \frac{1}{2} & \frac{1}{3} \\ \frac{1}{2} & \frac{1}{3} & \frac{1}{4} \\ \frac{1}{3} & \frac{1}{4} & \frac{1}{5} \end{pmatrix} \begin{pmatrix} c_1 \\ c_2 \\ c_3 \end{pmatrix} = \begin{pmatrix} -1+e \\ 1 \\ -2+e \end{pmatrix}$$

And solving this system (e.g. using Mathematica to avoid painful calculations) gives

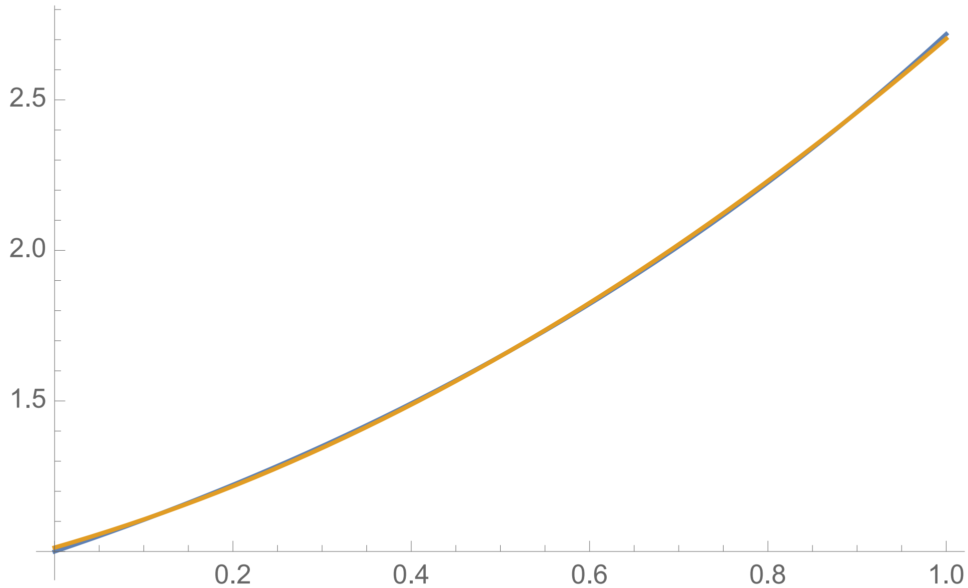
$$c_1 = -105 + 39e, \quad c_2 = 98 - 216e, \quad c_3 = -570 + 210e$$

Or, approximately,

$$c_1 = 1.101, \; c_2 = 0.851, \; c_3 = 0.839.$$

Thus the best-fit quadratic function is roughly $1.01 + 0.851x + 0.839x^2$.

This quadratic function is an excellent fit on the interval [0,1], as the following plot shows.



It is interesting to note that there is another well-known quadratic approximation of e^x, namely its degree-2 Taylor approximation $1+x+\frac{1}{2}x^2$. The approximation above is a better fit in the "least squared error" sense on [0,1], but the quality of its fit is specific to the interval [0,1]. Note that the two approximations have similar coefficients, but they are slightly different.

4 Fourier approximation

We consider in this section a particularly useful case of the approximation problem we've described, where we attempt to approximate functions on the interval $[-\pi, \pi]$[1] by functions from the following list.

$$1, \sin x, \cos x, \sin(2x), \cos(2x), \sin(3x), \cos(3x), \cdots, \sin(nx), \cos(nx), \cdots$$

This list of functions has several wonderful properties. Chief among them is that they form an orthogonal set. That is:

Proposition

Any two of the functions listed above are orthogonal to each other.

We omit the proof of this proposition, because the proof itself it not crucial to the intended applications (if you wish to work the proof out for yourself, I will note that the main ingredient is the use of trigonometric identities for products). The wonderful consequence of orthogonality is that the approximation problem becomes vastly simpler, as described in the following.

[1]There is nothing special about this interval, except that it serves as a period for the functions $\sin x, \cos x$, which simplifies notation. If we were doing Fourier approximation on a different interval, we would typically perform a horizontal scaling on all the approximation functions.

Simplified version of the main theorem, for the orthogonal case

Suppose that $S = \{\vec{v}_1, \cdots, \vec{v}_n\}$ and $\vec{b}$ are given, in some inner product space. Suppose further that any two elements of S are orthogonal to each other. That is, if $i \neq j$ then

$$\langle \vec{v}_i, \vec{v}_j \rangle = 0.$$

Then the linear combination of S that best approximates $\vec{b}$ is given by the formula

$$\vec{v} = \sum_{i=1}^{n} \frac{\left\langle \vec{v}_i, \vec{b} \right\rangle}{\langle \vec{v}_i, \vec{v}_i \rangle} \vec{v}_i.$$

Note that, using the notion of *projection*, we could also write this formula as

$$\vec{v} = \sum_{i=1}^{n} \mathrm{proj}_{\vec{v}_i} \vec{b}.$$

Proof. By the main theorem from earlier in this document, the best linear combination

$$\vec{v} = c_1 \vec{v}_1 + c_2 \vec{v}_2 + \cdots + c_n \vec{v}_n$$

should be chosen such that for each i,

$$\vec{v}_i \perp (\vec{v} - \vec{b}).$$

This orthogonality condition can be rewritten, using inner products, as the equation

$$\langle \vec{v}_i, \vec{v}_1 \rangle c_1 + \langle \vec{v}_i, \vec{v}_2 \rangle c_2 + \cdots + \langle \vec{v}_i, \vec{v}_n \rangle c_n = \left\langle \vec{v}_i, \vec{b} \right\rangle.$$

However, all but one of the inner products on the left side of this equation are equal to 0, due to the orthogonality of the set $\{\vec{v}_1, \cdots, \vec{v}_n\}$. The only one that doesn't vanish is $\langle \vec{v}_i, \vec{v}_i \rangle$. So this equation simplifies to

$$\langle \vec{v}_i, \vec{v}_i \rangle c_i = \left\langle \vec{v}_i, \vec{b} \right\rangle.$$

Thus for each choice of i, the orthogonality condition with $\vec{v}_i$ uniquely identifies the coefficient of $\vec{v}_i$ to be

$$c_i = \frac{\left\langle \vec{v}_i, \vec{b} \right\rangle}{\langle \vec{v}_i, \vec{v}_i \rangle}.$$

Plugging this coefficients into $\vec{v} = \sum_{i=1}^{n} c_i \vec{v}_i$ gives the desired equation. □

We may now combine the proposition (on the orthogonality of the functions in question) with the simplified version of the main theorem to carry out Fourier approximation.

Example: the square wave

Consider the following piecewise-continuous function.

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

Find the linear combination $g(x)$ of

$$\{1, \sin x, \cos x, \sin(2x), \cos(2x), \sin(3x), \cos(3x)\}$$

that minimizes $\int_{-\pi}^{\pi} (g(x) - f(x))^2 \ dx$.

Full disclosure about something I am being slightly imprecise about: since $f(x)$ is not a continuous function, it is not actually a member of the inner product space $\mathcal{C}[-\pi, \pi]$. So I should really be speaking of the space of piecewise-continuous functions, which introduces some technicalities that are not crucial for the applications we are interested in. Since this distinction does not affect any of the methods involved, I will brush this issue under the rug and not worry about it.

Solution: The function we are attempting to minimize can be written $\|g(x) - f(x)\|^2$. So this is an example of the type of approximation problem we have been studying. Due to the orthogonality of the functions to be combined, we may apply the simplified version of the main theorem. The function we seek is therefore given by

$$\begin{aligned} g(x) \quad = \quad & \frac{\langle 1, f\rangle}{\langle 1, 1\rangle} \cdot 1 + \frac{\langle \sin x, f\rangle}{\langle \sin x, \sin x\rangle} \cdot \sin x + \frac{\langle \cos x, f\rangle}{\langle \cos x, \cos x\rangle} \cdot \cos x \\ & + \frac{\langle \sin(2x), f\rangle}{\langle \sin(2x), \sin(2x)\rangle} \cdot \sin(2x) + \frac{\langle \cos(2x), f\rangle}{\langle \cos(2x), \cos(2x)\rangle} \cdot \cos(2x) \\ & + \frac{\langle \sin(3x), f\rangle}{\langle \sin(3x), \sin(3x)\rangle} \cdot \sin(3x) + \frac{\langle \cos(3x), f\rangle}{\langle \cos(3x), \cos(3x)\rangle} \cdot \cos(3x) \end{aligned}$$

Each of these inner products can be computed as an integral. To compute inner products with $f(x)$ requires a bit of thought; the key fact is that integrals involving piecewise functions can be broken into pieces. So when taking inner products with $f(x)$, we can use the formula

$$\langle g, f\rangle = \int_{-\pi}^{\pi} g(x)f(x) \ dx = \int_{-\pi}^{0} g(x)(-1) \ dx + \int_{0}^{\pi} g(x)(1) \ dx = -\int_{-\pi}^{0} g(x) \ dx + \int_{0}^{\pi} g(x) \ dx.$$

So all of the inner products involving f can be found as follows.

$$\begin{aligned}
\langle 1, f\rangle &= -\int_{-\pi}^{0} dx + \int_{0}^{\pi} dx = -\pi + \pi \\
&= 0 \\
\langle \sin x, f\rangle &= -\int_{-\pi}^{0} \sin x \; dx + \int_{0}^{\pi} \sin x \; dx = 2 + 2 \\
&= 4 \\
\langle \cos x, f\rangle &= -\int_{-\pi}^{0} \cos x \; dx + \int_{0}^{\pi} \cos x \; dx = 0 + 0 \\
&= 0 \\
\langle \sin(2x), f\rangle &= -\int_{-\pi}^{0} \sin(2x) \; dx + \int_{0}^{\pi} \sin(2x) \; dx = 0 + 0 \\
&= 0 \\
\langle \cos(2x), f\rangle &= -\int_{-\pi}^{0} \cos(2x) \; dx + \int_{0}^{\pi} \cos(2x) \; dx = 0 + 0 \\
&= 0 \\
\langle \sin(3x), f\rangle &= -\int_{-\pi}^{0} \sin(3x) \; dx + \int_{0}^{\pi} \sin(3x) \; dx = \frac{2}{3} + \frac{2}{3} \\
&= \frac{4}{3} \\
\langle \cos(3x), f\rangle &= -\int_{-\pi}^{0} \cos(3x) \; dx + \int_{0}^{\pi} \cos(3x) \; dx = 0 + 0 \\
&= 0
\end{aligned}$$

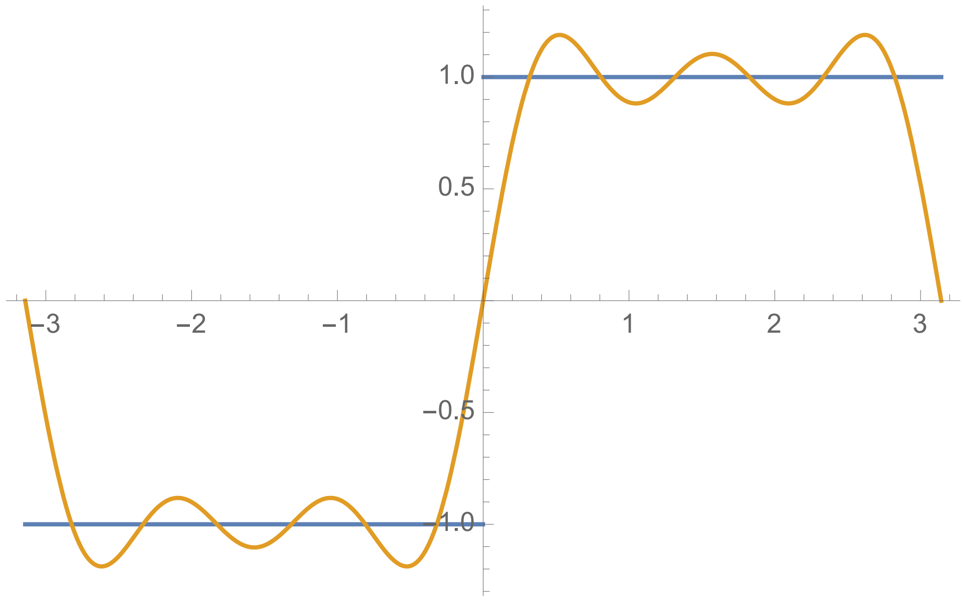
So in fact only two coefficients will be nonzero: that of $\sin x$ and that of $\sin(3x)$. We need two more inner products to write the result. Both can be computed using a trigonometric identity (the half-angle formula), or simply found with a computer.

$$\begin{aligned}
\langle \sin(x), \sin(x)\rangle = \int_{-\pi}^{\pi} \sin^2 x \; dx &= \pi \\
\langle \sin(3x), \sin(3x)\rangle = \int_{-\pi}^{\pi} \sin^2(3x) \; dx &= \pi
\end{aligned}$$

Hence the approximation we seek is

$$g(x) = \frac{4}{\pi} \sin(x) + \frac{4}{3\pi} \sin(3x).$$

The plot below shows the graph of the original "square wave" together with this Fourier approximation.



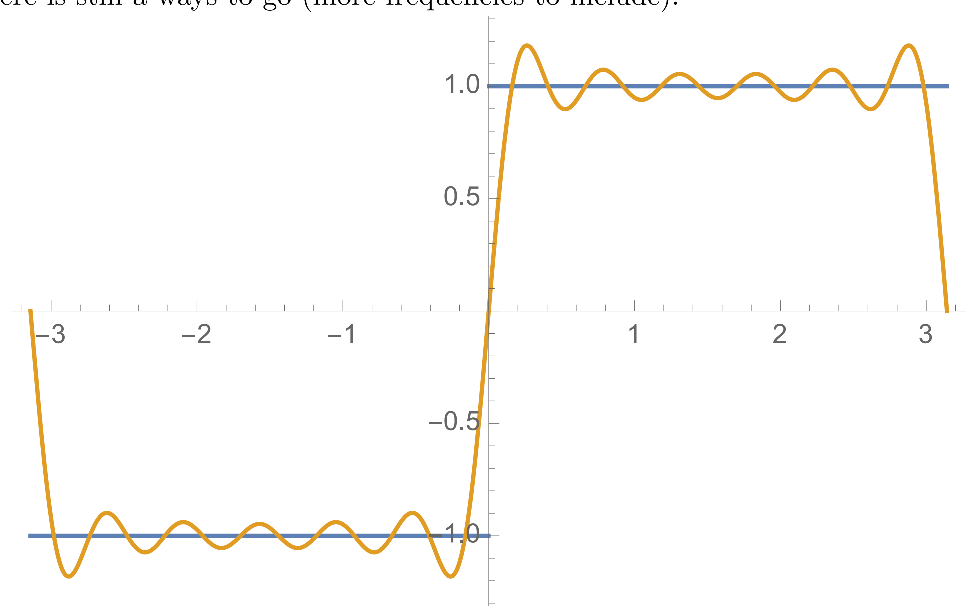
Note. In fact, it is not too difficult to derive the following general formulas for the inner product of the square wave with arbitrary "frequencies" of sine and cosine.

$$\begin{aligned} \langle \cos(nx), f(x) \rangle &= 0 \text{ for all } n \\ \langle \sin(nx), f(x) \rangle &= \begin{cases} \frac{4}{n\pi} & \text{if } n \text{ is odd} \\ 0 & \text{if } n \text{ is even} \end{cases} \end{aligned}$$

Hence we can work out what the approximation will be for any number of included frequencies. For example, if we include all functions from our list up to $\sin(13x), \cos(13x)$, then the approximation for the square wave will be

$$f(x) = \frac{4}{\pi}\sin(x) + \frac{4}{3\pi}\sin(3x) + \frac{4}{5\pi}\sin(5x) + \frac{4}{7\pi}\sin(7x) + \frac{4}{9\pi}\sin(9x) + \frac{4}{11\pi}\sin(11x)$$

The lot of this function is shown below. Note that it is a somewhat better approximation, but there is still a ways to go (more frequencies to include).



5 Additional remarks on Fourier series

What we have been discussing so far are *Fourier approximations,* which approximate a given function on $[-\pi, \pi]$ by a *finite* linear combination of sines and cosines. In fact, what makes the list of functions

$$1, \sin x, \cos x, \sin(2x), \cos(2x), \sin(3x), \cos(3x), \cdots, \sin(nx), \cos(nx), \cdots$$

truly miraculous is what happens when you pass to the limit, and consider the *infinite* sum given by the following formula:

$$\frac{\langle 1, f\rangle}{\langle 1, 1\rangle} \cdot 1 + \sum_{n=1}^{\infty} \frac{\langle \sin(nx), f\rangle}{\langle \sin(nx), \sin(nx)\rangle} \cdot \sin(nx) + \sum_{n=1}^{\infty} \frac{\langle \cos(nx), f\rangle}{\langle \cos(nx), \cos(nx)\rangle} \cdot \cos(nx)$$

For example, the Fourier series for the square wave can be written

$$\sum_{n=0}^{\infty} \frac{4}{(2n+1)\pi} \sin((2n+1)x).$$

So the square wave can be decomposed into an (infinite) sum that includes odd frequencies only.

This infinite sum is called the *Fourier series* of $f(x)$. An amazing fact is that, if $f(x)$ is a continuous function on $[-\pi, \pi]$, this series converges exactly to $f(x)$ (with one small caveat: at the endpoints $\pm\pi$, the sum converges to the average of $f(-\pi)$ and $f(\pi)$). If $f(x)$ is only piecewise continuous, this sum still converges, but the result may differ from x at the discontinuities.

The way this is usually summarized is: the list of functions above (1, sines, and cosines) form an *orthogonal basis*, since (in a suitable sense, if we allow infinite sums) they *span* the space of functions on this interval, in addition to being linearly independent.

There are other orthonormal bases for such spaces of functions that have many useful applications. One that has been particularly useful recently, such as in the most recent version of the JPEG image compression algorithm, comes from the theory of *wavalets.* If you want to learn more, you should take Amherst's course on Wavelets and Fourier Analysis (Math 320).

A final comment: although sines and cosines are familiar from calculus and fairly tangible, in most more advanced courses it turns out to be somewhat more convenient to introduce complex numbers and instead use the functions e^{inx} (where n ranges over all integers, positive and negative). One benefit of this setup is that it is no longer necessary to distinguish between the two types of functions (sine and cosine). So this is the point of view that you will often see adopted in more advanced texts.