Gram-Schmidt for functions: Legendre polynomials

S. G. Johnson, MIT course 18.06, Spring 2009 (supplement to textbook section 8.5)

March 16, 2009

Most of 18.06 is about column vectors in $\mathbb{R}^m$ or $\mathbb{R}^n$ and $m \times n$ matrices. But we observed early on that there are many other perfectly good "vector spaces" for which we can add, subtract, and multiply vectors by scalars. Some of the most important examples of these are vector spaces of functions. In this note, we will consider the vector space of functions $f(x)$ defined for $x \in [-1,1]$.

For column vectors, we also defined dot products $\vec{x} \cdot \vec{y} = \vec{x}^T \vec{y}$ and from this defined orthogonal vectors, and from that orthonormal vectors and orthonormal bases. An orthonormal basis of vectors is one of the nicest (most useful, most easily applied and analyzed) ways of describing a vector space or subspace. To get an orthonormal basis, we derived the Gram-Schmidt process. Here, we will do *exactly* the same things, but for *functions*.

Functions already form a vector space (we can add/subtract them and multiply them by scalars). The only new thing we need to talk about their orthogonality is some kind of dot product. (A vector space that also has a dot product is called a *Hilbert space*.[1]) Dot products of column vectors are defined by multiplying the components and adding them up. To define the "dot product" $f(x) \cdot g(x)$ of two functions, we will do the same thing: we will multiply the "components" (the values at each x) and "add them up" via an *integral*. The simplest and most intuitive definition of a dot product for functions on $[-1,1]$ (but not the only possibility!) is therefore:

$$f(x) \cdot g(x) = \int_{-1}^{1} f(x) g(x)\, dx.$$

For example, consider the function $f(x) = 1$. The squared "length" of this function is $\|1\|^2 = \int_{-1}^{1} 1^2 dx = 2$. A "unit vector" $q_1(x)$ parallel to this function is therefore:

$$q_1(x) = \frac{1}{\|1\|} = \boxed{\sqrt{1/2}}.$$

Take another function $f(x) = x$. The "unit vector" parallel to this function is:

$$q_2(x) = \frac{x}{\|x\|} = \frac{x}{\sqrt{\int_{-1}^{1} x^2 dx}} = \frac{x}{\sqrt{2/3}} = \boxed{x\sqrt{3/2}}.$$

Now, what is the dot product $q_1(x) \cdot q_2(x)$? This is proportional to $\int_{-1}^{1} 1x\, dx = 0$: the two functions are orthogonal and length one, so they are *orthonormal!* What is their span? If we take all linear combinations $c_1 q_1(x) + c_2 q_2(x)$ for any c_1 and c_2, we get *all polynomials of degree 0 or 1.* So, q_1 and q_2 are an orthonormal basis for polynomials of degree ≤ 1.

What if we want an orthonormal basis for functions of degree ≤ 2? Well, clearly we need to add another function x^2 to get degree-2 polynomials. But this is not orthogonal: $x^2 \cdot 1 = \int_{-1}^{1} x^2 1\, dx = 2/3 \neq 1$. How

---

[1] Technically, this is a lie. To qualify as a Hilbert space, a vector space must not only have an inner product but also a property called "completeness" (all Cauchy sequences must have limits). This is mainly a detail for functional analysts, and is only there to exclude pathological cases—any vector space you are likely to come up with or encounter will almost certainly be complete. We won't worry about it in 18.06!

can we get an orthogonal basis function $q_3(x)$? The same way we would if these were column vectors: by Gram-Schmidt, we just need to *subtract the components of* q_1 *and* q_2 *from* x^2*:*

$$
\begin{aligned}
q_3(x) &= \frac{x^2 - q_1(x)[q_1(x)] \cdot x^2] - q_2(x)[q_2(x) \cdot x^2]}{\|\cdots\|} = \frac{x^2 - \sqrt{1/2}[2/3\sqrt{1/2}] - x\sqrt{3/2}[0]}{\|\cdots\|} \\
&= \frac{x^2 - 1/3}{\sqrt{\int_{-1}^{1}(x^2-1/3)^2 dx}} = \frac{x^2 - 1/3}{\sqrt{\int_{-1}^{1}(x^4 - 2x^2/3 + 1/9)dx}} = \frac{x^2-1/3}{\sqrt{8/45}} = \boxed{(3x^2-1)\sqrt{5/8}}.
\end{aligned}
$$

If we continue this process, what we are doing is taking the functions $1, x, x^2, x^3, x^4$, and so on, and applying Gram-Schmidt to them: the functions $q_1, q_2, \ldots, q_n$ will form an orthonormal basis for all polynomials of degree $\leq n-1$.

There is another name for these functions: they are called the *Legendre polynomials*, and play an important role in the understanding of functions, polynomials, integration, differential equations, and many other areas. (Technically, the Legendre polynomials are only proportional to the $q_i's$ defined here, since by convention the Legendre polynomials are normalized to a length other than 1, but that's just a minor detail.) The next few are:

$$q_4(x) = (5x^3 - 3x)\frac{\sqrt{7/2}}{2}$$

$$q_5(x) = (35x^4 - 30x^2 + 3)\frac{\sqrt{9/2}}{8}$$

$$q_6(x) = (63x^5 - 70x^3 + 15x)\frac{\sqrt{11/2}}{8}$$

and so on. If we look at *all* Legendre polynomials (up to $n = \infty$), we can describe *any* function that can be written in terms of *any* polynomial on $[-1,1]$, including *infinite series* (e.g. functions with convergent Taylor series).

Just about anything that is possible with orthonormal bases of column vectors we can also do with this orthonormal basis of functions. For example, if we want to write some function $b(x)$in this basis, i.e. write

$$b(x) = c_1 q_1(x) + c_2 q_2(x) + c_3 q_3(x) + \cdots$$

for some coefficients c_i, to get the coefficients c_i we just take the dot product $q_i(x) \cdot b(x)$: just like $\vec{q}_i \vec{q}_i^T \vec{b}$ is the projection of a column vector $\vec{b}$ onto $\vec{q}_i$, the projection of $b(x)$ onto $q_i(x)$ is $q_i(x)[q_i(x) \cdot b(x)] = c_i q_i(x)$.

As another example, suppose we have some function $b(x)$ and we want to find the "closest" function $\hat{b}(x)$ of degree 3 or less, in the sense that we want $\|\hat{b}(x) - b(x)\|^2$ to be minimized. This is *exactly* a least-squares problem where we want the closest function to $b(x)$ in the span of q_1, q_2, q_3, and q_4. And we already know that the solution to this problem is the orthogonal projection of $b(x)$ onto the subspace. If these were vectors, we would compute $\hat{b} = QQ^T b$ where Q is the matrix with columns $q_{1\ldots4}$. But that is exactly the same thing as writing $\hat{b} = q_1(q_1^T b) + q_2(q_2^T b) + q_3(q_3^T b) + q_4(q_4^T b)$, and we can do *exactly this* for functions, just with the dot product $q_i^T b$ defined as the integral.

The Legendre polynomials have a number of other beautiful properties that we won't derive here. For example, the degree-n Legendre polynomial has exactly n roots in the interval $[-1,1]$ (and there is a beautiful proof of this from orthogonality). Google will turn up many, many web pages on them. And you can form other sets of orthogonal polynomials, each with their own beautiful and important properties, by defining the dot product integral slightly differently—you will do one such case for homework. Other examples include Hermite polynomials, Chebyshev polynomials, Laguerre polynomials, Jacobi polynomials, and so on.