1 Inner product spaces

> We come now to the approximation process most commonly employed and most highly developed: least squares. An abstract vantage point from which it is convenient to survey the common features of various least square approximations is provided by the theory of inner product spaces. If the subjects of algebra, geometry, and analysis can be said to have a "center of gravity," it surely lies in this theory.
>
> *Philip J. Davis*

Up until this point, we have been treating vector spaces and linear transformations as purely algebraic objects, which, emphatically, they are. Reflect upon how much we have accomplished through algebra alone! For instance, you now know how to diagonalize a linear operator using its eigenvectors and eigenvalues; as a consequence, you can solve [almost] any system of linear ODE with constant coefficients. Although we have not, by far, exhausted the list of purely algebraic topics, it is time to introduce geometry. As the opening quote suggests, modern linear algebra is a union of algebra, geometry, and analysis. I hope that after reading this section you will catch a glimpse of beautiful synergy between these seemingly very different branches of mathematics.

1.1 Dot product as a special case of a an inner product

Chances are, you first saw vectors drawn as arrows. That is how vectors are typically introduced in Physics and Calculus; later they acquire components and become n-tuples of numbers. In Multivariate Calculus the n-dimensional Euclidean space $\mathbb{R}^n$ is introduced essentially as a collection of arrows issuing from the origin. Each arrow has length—the usual expression given by the Pythagorian theorem. Furthermore, if two vectors are drawn, there is a well-defined (acute) angle between them. As one learns in Calculus, the easiest way to compute both the lengths and the angles in $\mathbb{R}^n$ is through the use of the *dot product*. The latter is commonly postulated by the formula

$$\langle x, y\rangle = x_1\, y_1 + x_2\, y_2 + \ldots + x_n\, y_n, \tag{1}$$

which we will write in an equivalent form:

$$\langle x, y\rangle = x^T\, y.$$

Recall that the superscript "T" is for 'transpose', which is to say, "interchange rows with columns"; it is the single quote in `Matlab`:

```
s = x'*y; % dot product of two vectors
```

With the dot product defined by (1), it is easy to see that the length of a vector should be:

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

Meanwhile, the angle between x and y can be computed from the relation:

$$\cos\theta = \frac{\langle x, y \rangle}{|x|\,|y|}.$$

The latter formula is not obvious; it is usually justified using the Law of Cosines from geometry.

If you are familiar with the dot product, have you ever wondered why it is defined by Equation (1)? Well, it does not have to be. The dot product on $\mathbb{R}^n$ is just a special case of an *inner product*. In general, an inner product $\langle \cdot, \cdot \rangle$ is a mapping which takes a pair of vectors into a number subject to the following axioms:

(IP1) Inner product of any vector with itself must be a nonnegative number: $\langle x, x \rangle \geq 0$. Moreover, $\langle x, x \rangle = 0$ implies $x = 0$.

(IP2) Inner product is symmetric: $\langle x, y \rangle = \langle y, x \rangle$.

(IP3) Inner product is linear in the first argument: $\langle \sum_n c_n\, x_n, y \rangle = \sum_n c_n \langle x_n, y \rangle$; as a consequence of symmetry, inner product is linear in the second argument as well.

Any bilinear mapping that takes a pair of vectors into a scalar is called a bilinear *form*. One can summarize the conditions imposed on an inner product in one sentence as follows: Inner product is a positive definite, symmetric, bilinear form.

Whenever we work with vectors in $\mathbb{R}^n$ we will usually use the dot product (1) as the default inner product. However, on occasion, we will use other inner products. For instance, the following is an inner product on $\mathbb{R}^3$ which is clearly different from the dot product:

$$\langle x, y \rangle = x_1\, y_1 + 2\, x_2\, y_2 + 3\, x_3\, y_3. \tag{2}$$

We can call it a *weighted dot product*. As an easy exercise, verify that Equation (2) satisfies all of the necessary axioms.

As another example, let V be the space of functions of one variable defined on the interval $[-1, 1]$. Define the inner product on V by:

$$\langle f, g \rangle = \int_{-1}^{1} f(x)\, g(x)\, dx \tag{3}$$

We will call the inner product (3) the L^2-inner product. It is the continuous analogue of the dot product (1) and is therefore very common.

Suppose now that a vector space V has an inner product $\langle \cdot, \cdot \rangle$ and is therefore an *inner product space*. Using the inner product, we define the length of a vector, which we will usually call 'norm', by

$$|x| = \sqrt{\langle x, x \rangle} \tag{4}$$

Notice that Equation (4) is indistinguishable from that in $\mathbb{R}^n$; it is well defined because $\langle \cdot, \cdot \rangle$ is nonnegative definite.

Similarly, we define the (acute) angle between two vectors by

$$\cos\theta = \frac{\langle x, y \rangle}{|x|\,|y|}. \tag{5}$$

Again, this is the same formula as in Multivariate Calculus. If $\langle x, y \rangle = 0$ then, clearly, $\cos\theta = 0$ and the vectors are perpendicular. We will commonly use the term 'orthogonal' instead of 'perpendicular'.

1.2 Basis expansions

Let V be an n-dimensional vector space equipped with a basis $\mathcal{B} = \{v_1, \ldots, v_n\}$. Consider the problem of finding the components of $w \in V$ with respect the the basis $\mathcal{B}$. That is, we want to find coefficients in the linear combination:

$$w = c_1\, v_1 + \ldots + c_n\, v_n. \tag{6}$$

Generally, this can be a very difficult problem. However, let us equip V with an inner product $\langle \cdot, \cdot \rangle$—any inner product. Taking the inner products of both sides of Equation (6) with the basis vectors, we get n relations of the form

$$\langle w, v_m \rangle = c_1\, \langle v_1, v_m \rangle + \ldots + c_n\, \langle v_n, v_m \rangle. \tag{7}$$

These are the usual linear equations with scalar coefficients and scalar unknowns which we can write in matrix form as:

$$G\, c = b.$$

Here the n-by-n matrix G has entries $\langle v_n, v_m \rangle$ and is called Gram's matrix; the right-hand side vector b has known components $\langle w, v_m \rangle$ and c is the vector of coefficients that are to be determined. The standard linear system $G\,c = b$ can be solved through Gaussian elimination; symbolically, $c = G^{-1}b$.

Gram's matrices $G = \langle v_n, v_m \rangle$ have a number of interesting properties. They are clearly *symmetric*: $G^T = G$. More importantly, they are nonsingular as long as the vectors $\{v_n\}$ are linearly independent, which must be true if $\{v_n\}$ form a basis. Some of the exercises at the end ask you to confirm these and other properties of symmetric matrices computationally.

Having dealt with the general problem of basis expansion in inner product spaces, let us consider a particularly important special case where the basis vectors are mutually orthogonal: $\langle v_n, v_m \rangle = 0$ if $n \neq m$. It is easy to see that the Gram matrix in this case is diagonal with diagonal entries $\langle v_n, v_n \rangle$. Consequently, the explicit solution of the expansion problem in this case is:

$$w = \frac{\langle w, v_1 \rangle}{\langle v_1, v_1 \rangle}\, v_1 + \ldots + \frac{\langle w, v_n \rangle}{\langle v_n, v_n \rangle}\, v_n \tag{8}$$

Equation (9) is called "orthogonal basis expansion" or "Fourier expansion". It further simplifies to

$$w = \langle w, v_1 \rangle\, v_1 + \ldots + \langle w, v_n \rangle\, v_n$$

if the basis vectors have unit length. Henceforth we will call bases of orthogonal vectors simply 'orthogonal bases' and the corresponding basis expansions 'orthogonal expansions'. Orthogonal bases where vectors have unit length are usually called 'orthonormal.'

As a simple example, consider the standard basis of $\mathbb{R}^3$: $\{i, j, k\}$. It is an orthonormal basis with respect to the standard dot product. Therefore, any vector $w \in \mathbb{R}^3$ can be written as:

$$w = \langle w, i \rangle\, i + \langle w, j \rangle\, j + \langle w, k \rangle\, k.$$

As another example, consider quadratics on $[-1, 1]$ with L^2-inner product (3). The Legendre polynomials $\{P_0 = 1, P_1 = x, P_2 = \frac{3}{2}\,x^2 - \frac{1}{2}\}$ are easily seen to be orthogonal (but not orthonormal). Therefore, any quadratic q can be written as the following combination:

$$q = \frac{\langle q, P_0 \rangle}{\langle P_0, P_0 \rangle}\, P_0 + \frac{\langle q, P_1 \rangle}{\langle P_1, P_1 \rangle}\, P_1 + \frac{\langle q, P_2 \rangle}{\langle P_2, P_2 \rangle}\, P_2. \tag{9}$$

For instance, let $q = x^2 + x + 1$. Then

$$\begin{aligned}
\langle q, P_0 \rangle &= \int_{-1}^{1} (x^2 + x + 1)\, 1\, dx = \frac{8}{3} \\
\langle q, P_1 \rangle &= \int_{-1}^{1} (x^2 + x + 1)\, x\, dx = \frac{2}{3} \\
\langle q, P_2 \rangle &= \int_{-1}^{1} (x^2 + x + 1) \left(\frac{3}{2} x^2 - \frac{1}{2} \right) dx = \frac{4}{15}
\end{aligned}$$

Furthermore,

$$\begin{aligned}
\langle P_0, P_0 \rangle &= \int_{-1}^{1} 1^2\, dx = 2 \\
\langle P_1, P_1 \rangle &= \int_{-1}^{1} x^2\, dx = \frac{2}{3} \\
\langle P_2, P_2 \rangle &= \int_{-1}^{1} \left(\frac{3}{2} x^2 - \frac{1}{2} \right)^2 dx = \frac{2}{5}
\end{aligned}$$

Finally,

$$\begin{aligned}
\frac{\langle q, P_0 \rangle}{\langle P_0, P_0 \rangle} P_0 + \frac{\langle q, P_1 \rangle}{\langle P_1, P_1 \rangle} P_1 + \frac{\langle q, P_2 \rangle}{\langle P_2, P_2 \rangle} P_2 &= \frac{(8/3)}{2}\, 1 + \frac{(2/3)}{(2/3)}\, x + \frac{(4/15)}{2/5} \left(\frac{3}{2} x^2 - \frac{1}{2} \right) \\
&= \frac{4}{3} + x + x^2 - \frac{1}{3} = q,
\end{aligned}$$

as required.

1.3 Fourier approximation

We have shown that any quadratic can be written as an orthogonal expansion of the first three Legendre polynomials:

$$q = \frac{\langle q, P_0 \rangle}{\langle P_0, P_0 \rangle} P_0 + \frac{\langle q, P_1 \rangle}{\langle P_1, P_1 \rangle} P_1 + \frac{\langle q, P_2 \rangle}{\langle P_2, P_2 \rangle} P_2. \tag{10}$$

Suppose that q is not a quadratic but is some function on $[-1, 1]$. Although Equation (10) clearly cannot hold in that case, its right-hand side still makes sense and can be regarded as a quadratic approximation to q. For example, let $q = \cos(x)$. Then, inspired by Equation (10), we can approximate the cosine on $[-1, 1]$ as follows:

$$\cos(x) \approx \sin(1) + \frac{5}{2}\,(-4\,\sin(1) + 6\,\cos(1))\left(\frac{3}{2}\,x^2 - \frac{1}{2}\right)$$

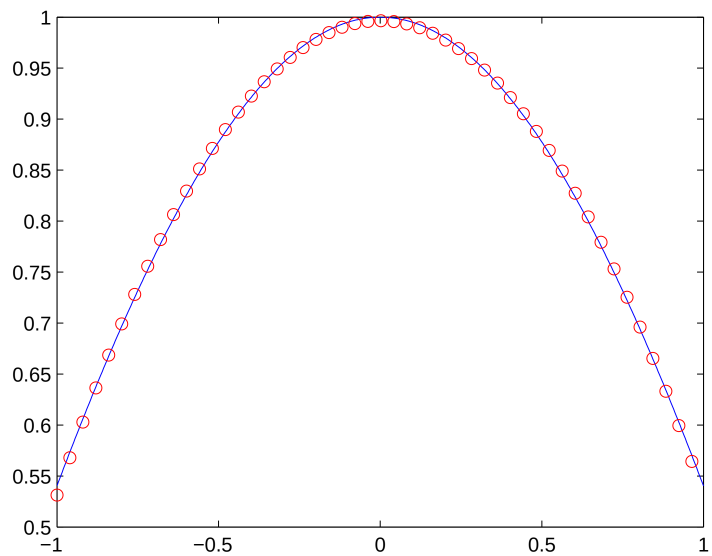

Figure 1: Approximation of $\cos(x)$ with a linear combination of the first three Legendre polynomials

As Figure 1 clearly suggests, the approximation is excellent throughout the interval. Compare that with Figure 2 below which shows the Taylor approximation.

The Taylor approximation is slightly better near the origin—the center of expansion. However, the accuracy of Taylor approximation quickly degrades away from the origin. In contrast, the accuracy of the Fourier approximation is consistent throughout the interval.

As another example, consider the space of functions on $[0, \pi]$ with "integrable squares":

$$V = \{f(x) \mid \int_0^\pi f(x)^2\,dx < \infty\}$$

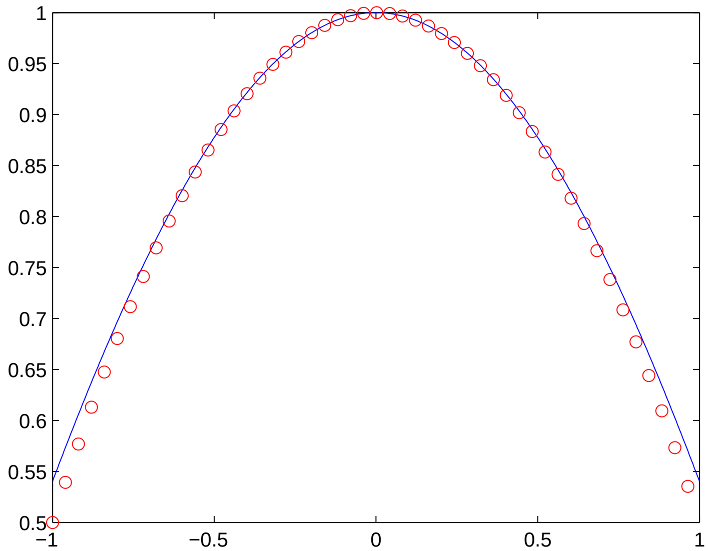

Figure 2: Approximation of $\cos(x)$ with the Taylor quadratic

The integrability of squares allows for the L^2 inner product on V:

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

You can easily check (exercise) that the functions

$$\phi_n(x) = \sqrt{\frac{2}{\pi}}\, \sin(n\, x), \quad n = 1, 2, 3, \ldots$$

belong to V, are orthogonal under the L^2-inner product, and have unit L^2-norm. Consequently, any function in V can be approximated with a sum of sines of integer frequencies:

$$f(x) \approx \sum_{n=1}^{N} \langle f, \phi_n\rangle\, \phi_n(x).$$

The greater N is, the better is the approximation. For example, let $f(x) = x$, which is clearly not a combination of sines. Then, using integration by parts, one can show that

$$\langle f, \phi_n\rangle = -\sqrt{\frac{2}{\pi}}\, \frac{\pi}{n}\, \cos(\pi\, n)$$

Consequently, after simplification, we can write:

$$x \approx \sum_{n=1}^{N} \left(-\frac{2\,\cos(\pi\,n)}{n} \right)\,\sin(n\,x) \tag{11}$$

Figure 3 shows the Fourier approximation (11) with $N = 50$. The Fourier sum struggles to converge near $x = \pi$ where all sines vanish but the function has value π; apart from that, however, the match is very decent.

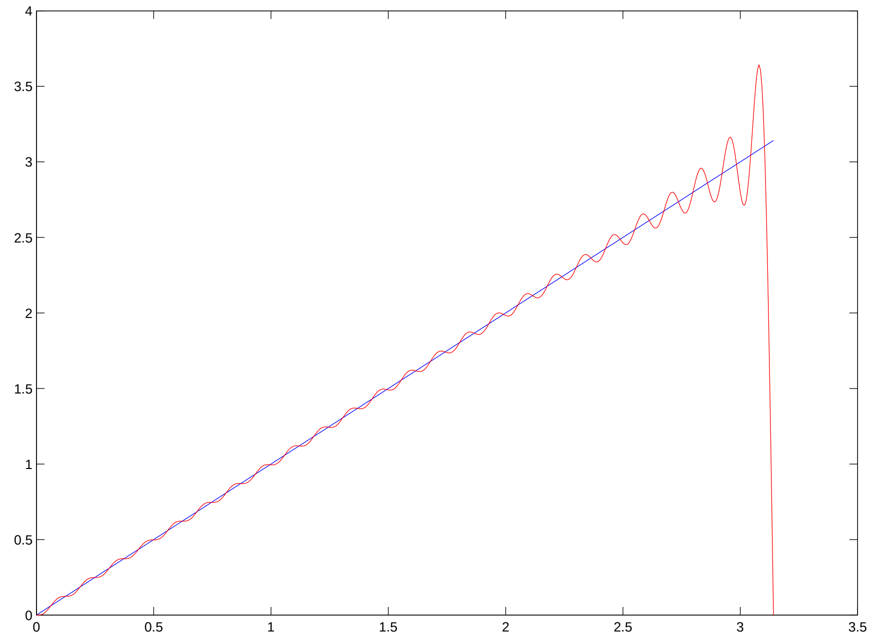

Figure 3: Approximation of x with sines of integer frequencies ($N = 50$); the oscillations near $x = \pi$ are due to the fact that all sines vanish there.

Here is the plot of the approximation (11) with $N = 50$ on the interval $[0, 10\,\pi]$:

If you had Circuits, you will recognize it as the plot of a "ramp waveform"—a special voltage that is often used to investigate AC circuits. You now know that it can be synthesized from simple harmonic voltages.

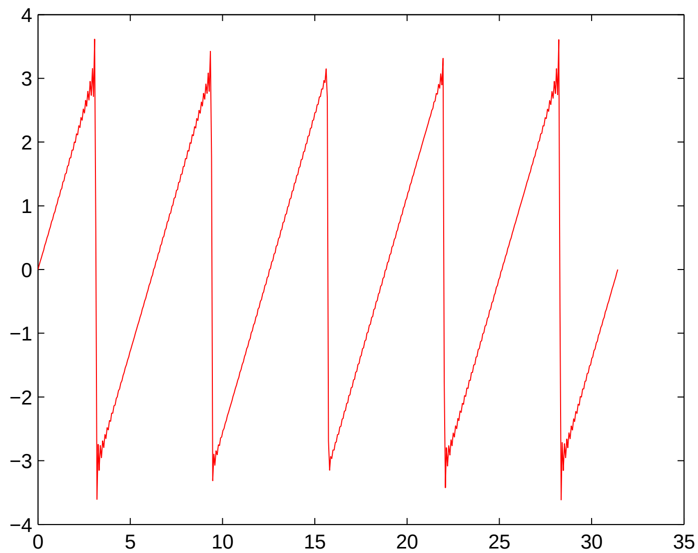

Figure 4: Ramp waveform

Exercises

1. Define the inner product on $\mathbb{R}^2$ by the formula:

$$\langle x, y \rangle = x_1\, y_1 + 4\, x_2\, y_2.$$

   Here we use the convention of denoting the n-th component of a vector with a subscript. Find a two-by-two matrix A such that:

$$\langle x, y \rangle = x^T\, A\, y.$$

2. Let

$$A = \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix}$$

   For $x, y \in \mathbb{R}^2$ suppose we define

$$[x, y] = x^T\, A\, y.$$

   Is $[x, y]$ an inner product on $\mathbb{R}^2$? If you answered "No", explain what breaks down. If you answered "Yes" either provide a formal argument, if you can, or generate convincing evidence in `Matlab`.

3. As you know from this handout, the first three Legendre polynomials are

$$P_0 = 1, \quad P_1 = x, \quad P_2 = \frac{3}{2}\,x^2 - \frac{1}{2}.$$

In general, Legendre polynomials are defined by the orthogonality relationship and the normalization condition:

$$\langle P_n, P_m \rangle = \int_{-1}^{1} P_n(x)\, P_m(x)\, dx = \begin{cases} 0, & \text{if } n \neq m, \\ \frac{2}{2\,n+1}, & \text{if } n = m. \end{cases}$$

Find P_3 and P_4.

4. Let $f = \cos(4\,x)$. Use orthogonal expansion to approximate f on $[-1, 1]$ with a linear combination of Legendre polynomials $P_0, \ldots, P_4$. Illustrate your results with a plot similar to Figure 1.

5. Let $f = x$. Approximate f on $[0, 2\pi]$ with the trigonometric sum of the form:

$$a_0 + \sum_{n=1}^{N} (a_n\, \cos(n\,x) + b_n\, \sin(n\,x)).$$

Which is to say, find the Fourier coefficients a_n and b_n. Illustrate your results with plots similar to Figures 3 and 4 (with $N = 50$). Compare to the approximation with sines alone.

6. Repeat the previous exercise with

$$f = \begin{cases} +1, & 0 \leq x \leq \pi, \\ -1, & \pi < x \leq 2\,\pi. \end{cases}$$

Where is the Fourier approximation bad in this case and what is the [obvious] cause?

7. Consider quadratics on the interval $[0, 1]$ with the L^2 inner product:

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

Find an orthogonal basis of this space. Validate your result with an orthogonal expansion of $q = x^2 + x + 1$.