Construction of orthogonal systems: the Gram–Schmidt Process, QR factorization, Fourier series

A. Eremenko

August 27, 2024

Suppose that we have a sequence of linearly independent vectors

$$\mathbf{v}_1, \mathbf{v}_2, \ldots,$$

finite or infinite. We will show how to construct an orthogonal sequence

$$\mathbf{u}_1, \mathbf{u}_1, \ldots,$$

with the property that for every k,

$$\mathrm{Span}(\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k) = \mathrm{Span}(\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_k) \quad \text{for every} \quad k \geq 1.$$

In particular, starting from any basis, we can construct an orthogonal basis in this way.

1. We can set $\mathbf{u}_1 = \mathbf{v}_1$.

2. Look for $\mathbf{u}_2$ of the form $\mathbf{u}_2 = \mathbf{v}_2 - c\mathbf{u}_1$ which is clearly in the $\mathrm{Span}(\mathbf{v}_1, \mathbf{v}_2)$. We have to choose c so that $\mathbf{u}_2 \perp \mathbf{u}_1$ that is

$$0 = (\mathbf{v}_2 - c\mathbf{u}_1, \mathbf{u}_1) = (\mathbf{v}_2, \mathbf{u}_1) - c(\mathbf{u}_1, \mathbf{u}_1),$$

so

$$c = \frac{\mathbf{v}_2, \mathbf{u}_1)}{(\mathbf{u}_1, \mathbf{u}_1)}.$$

3. Once $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_k$ have been already constructed, we look for $\mathbf{u}_{k+1}$ in the form

$$u_{k+1} = \mathbf{v}_{k+1} - c_1\mathbf{u}_1 - c_2\mathbf{u}_2 - \ldots - c_k\mathbf{u}_k,$$

where the constants c_j are determined from the condition that $(\mathbf{u}_{k+1}, \mathbf{u}_j) = 0,\ 1 \leq j \leq k$. Multiplying on $\mathbf{u}_j$ we obtain

$$0 = (\mathbf{u}_{k+1}, \mathbf{u}_j) = c_j(\mathbf{u}_j, \mathbf{u}_j),$$

since $\mathbf{u}_1, u_2, \ldots, u_j$ are orthogonal to each other. This gives the formula

$$c_j = \frac{(\mathbf{v}_{k+1}, \mathbf{u}_j)}{(\mathbf{u}_j, \mathbf{u}_j)}, \quad 1 \leq j \leq k.$$

This is called the Gram–Schmidt orthogonalization.

Sometimes we want an orthonormal system, rather then just orthogonal. Once you already constructed an orthogonal system, to make is orthonormal, one simply divide each vector by its length. But notice that the length of each $\mathbf{u}_j$ has already been computed at each step, so it makes sense to normalize on each step, rather than in the end. Then we obtain the following modified process:

1a. Take $\mathbf{w}_1 = \mathbf{v}_1/\|\mathbf{v}_1\|$.

2a. Look for $\mathbf{u}_2 = \mathbf{v}_2 - c\mathbf{w}_1$, which is orthogonal to w_1. This requires

$$c = (\mathbf{v}_2, \mathbf{w}_1).$$

Normalize $\mathbf{u}_2$ by replacing it with $\mathbf{w}_2 = \mathbf{u}_2/\|\mathbf{u}_2\|$.

3a. Suppose that $\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_k$ have already been constructed. They are orthonormal. Look for

$$\mathbf{u}_{k+1} = \mathbf{y}_{k+1} - c_1\mathbf{w}_1 - c_2\mathbf{w}_2 - \ldots - c_k\mathbf{w}_k.$$

and make it orthogonal to $\mathbf{w}_1, \ldots, \mathbf{w}_k$ by choosing

$$c_j = (\mathbf{y}_{k+1}, \mathbf{w}_j)), \quad 1 \leq j \leq k.$$

Normalize: $\mathbf{w}_k = \mathbf{u}_{k+1}/\|\mathbf{u}_{k+1}\|$.

For examples, see the book.

All this can be phrased in terms of certain matrix factorization. Let us arrange the given linearly independent sequence $\mathbf{a}_1, \mathbf{a}_2, \ldots,$ into columns of a matrix A. The outcome (an orthonormal system) applied to this sequence $\mathbf{a}_j$ makes the columns of a matrix $Q = [\mathbf{q}_1, \mathbf{q}_2, \ldots]$. The relation of them is

$$A = QR, \tag{1}$$

where R is an upper-triangular matrix.

Indeed, we have $A = [\mathbf{a}_1, \ldots,] = (a_{i,j})$, so that $a_{i,j}$ is the i-th entry of the column $\mathbf{a}_j$, and each $\mathbf{a}_j$ is a linear combination

$$\mathbf{a}_j = \sum_{i=1}^{j} c_{i,j}\mathbf{q}_i. \tag{2}$$

which means

$$a_{k,j} = \sum_{i=1}^{j} q_{k,i}c_{i,j}$$

This means that $A = QR$, where $R = (c_{i,j})$ an upper triangular matrix. The entries $c_{i,j}$ can be obtained in the usual way, by multiplying expansions (2) on $\mathbf{q}_i$ and using that $\mathbf{q}_j$ are orthonormal:

$$c_{i,j} = (\mathbf{a}_j, \mathbf{q}_i).$$

Thus every (rectangular) matrix with linearly independent columns can be factored as in (1) where Q is a square matrix with orthonormal columns, and R is upper triangular, of the same size as A. This is called the QR-factorization. Square matrices with orthonormal columns are exactly orthogonal matrices (check this).

Projections and orthogonal bases. If we now an orthonormal basis is a subspace U of a vector space V, then the formulas for the projection substantially simplify. Arranging this basis $\mathbf{u}_1, \ldots, \mathbf{u}_n$ into the matrix $A = [u_1, \ldots, u_n]$ we obtain $A^T A = I$, and the projection of $\mathbf{x} \in V$ onto U becomes:

$$A(A^T A)^{-1}A^T\mathbf{x} = AA^T\mathbf{x} = (\mathbf{x}, \mathbf{u}_1)\mathbf{u}_1 + (\mathbf{x}, \mathbf{u}_2) + \ldots + (\mathbf{x}, u_n)\mathbf{u}_n. \tag{3}$$

Orthogonal systems in function spaces.

Trigonometric system. Consider the system of functions in the space $PC[0, 1]$ with standard inner product

$$\sin t, \sin 2t, \ldots, \sin nt, \ldots \tag{4}$$

This is an orthogonal system. Let's denote $\phi_n = \sin nt$, then we have

$$(\phi_n, \phi_m) \quad = \quad \int_0^\pi \sin nt \sin mt\, dt$$

$$= \frac{1}{2}\int_0^{\pi} (\cos(m-n)t - \cos(m+n)t)\ dt$$

$$= \frac{1}{2}\int_0^{\pi} \cos(m-n)tdt = \begin{cases} 0, & m \neq n \\ \pi/2 & m = n. \end{cases}$$

By dividing each vector on its length we obtain an orthonormal system

$$\sqrt{\frac{2}{\pi}}\sin nt, \quad n = 1, 2, 3, \ldots.$$

Now suppose that we have an arbitrary function $f \in PC[0, \pi]$. Consider its projection on the span of the first n functions $\phi_1, \ldots, \phi_n$. This projection is given by

$$\phi = c_1\phi_1 + c_2\phi_2 + \ldots + c_n\phi_n, \tag{5}$$

where

$$c_k = \frac{(f, \phi_k)}{(\phi_k, \phi_k)} = \frac{2}{\pi}\int_0^{\pi} f(t)\sin kt\, dt, \tag{6}$$

which is similar to (3). This is the linear combination of $\phi_k(t) = \sin kt$ with $k \leq n$ which is the best approximation of f among all such combinations. The best approximation here has the following precise sense: among all functions ϕ in the span of $\{\phi_k\}_{k=1}^n$ it minimizes the error

$$\int_0^{\pi} \|f(t) - \phi(t)\|^2 dt.$$

It is plausible that when n increases this approximation becomes better and the error tends to zero, that is we have in some precise sense

$$f(t) = \sum_{k=1}^{\infty} c_k \sin kt.$$

This is indeed so, and the series in the right hand side is called the Fourier series of f, and the formula for the coefficients (6) is called the Fourier formula.

Orthogonal polynomials. Let us consider the sequence of linearly independent polynomials

$$1, t, t^2, t^3, \ldots$$

We consider them as functions in the space $PC[-1,1]$ with the standard dot product. Let us apply the orthogonalization process (without normalization on each step) to this sequence. So we set

$$P_1(t) = 1, \quad \|P_0\| = \sqrt{2}.$$

$$P_2(t) = t - c, \quad c = \frac{\int_{-1}^{1} t \cdot 1\, dt}{\|P_1\|^2} = 0,$$

so

$$P_2(t) = t, \quad \|P_2\| = \left(\int_{-1}^{1} t^2\, dt\right)^{1/2} = \sqrt{2/3}.$$

$$P_3(t) = t^2 - c_1 P_1 - c_2 P_2,$$

$$c_1 = \frac{\int_{-1}^{1} t^2\, dt}{\|P_1\|^2} = \frac{1}{3},$$

$$c_2 = \frac{\int_{-1}^{1} t^3\, dt}{\|P_2\|^2} = 0,$$

So

$$P_3(t) = t^2 - 1/3, \quad \|P_3\| = \left(\int_{-1}^{1} (t^2 - 1/3)^2 dt\right)^{1/2} = \sqrt{8/45},$$

and so on.