5.6 The Gram–Schmidt Orthogonalization Process

In this section we learn a process for constructing an orthonormal basis for an n-dimensional inner product space V. The method involves using projections to transform an ordinary basis $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n\}$ into an orthonormal basis $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n\}$.

We will construct the $\mathbf{u}_i$'s so that

$$\text{Span}(\mathbf{u}_1, \ldots, \mathbf{u}_k) = \text{Span}(\mathbf{x}_1, \ldots, \mathbf{x}_k)$$

for $k = 1, \ldots, n$. To begin the process, let

$$\mathbf{u}_1 = \left(\frac{1}{\|\mathbf{x}_1\|}\right)\mathbf{x}_1 \tag{1}$$

$\text{Span}(\mathbf{u}_1) = \text{Span}(\mathbf{x}_1)$, since $\mathbf{u}_1$ is a unit vector in the direction of $\mathbf{x}_1$. Let $\mathbf{p}_1$ denote the projection of $\mathbf{x}_2$ onto $\text{Span}(\mathbf{x}_1) = \text{Span}(\mathbf{u}_1)$; that is,

$$\mathbf{p}_1 = \langle \mathbf{x}_2, \mathbf{u}_1 \rangle \, \mathbf{u}_1$$

By Theorem 5.5.7,

$$(\mathbf{x}_2 - \mathbf{p}_1) \perp \mathbf{u}_1$$

Note that $\mathbf{x}_2 - \mathbf{p}_1 \neq \mathbf{0}$, since

$$\mathbf{x}_2 - \mathbf{p}_1 = \frac{-\langle \mathbf{x}_2, \mathbf{u}_1 \rangle}{\|\mathbf{x}_1\|}\mathbf{x}_1 + \mathbf{x}_2 \tag{2}$$

and $\mathbf{x}_1$ and $\mathbf{x}_2$ are linearly independent. If we set

$$\mathbf{u}_2 = \frac{1}{\|\mathbf{x}_2 - \mathbf{p}_1\|}(\mathbf{x}_2 - \mathbf{p}_1) \tag{3}$$

then $\mathbf{u}_2$ is a unit vector orthogonal to $\mathbf{u}_1$. It follows from (1), (2), and (3) that $\text{Span}(\mathbf{u}_1, \mathbf{u}_2) \subset \text{Span}(\mathbf{x}_1, \mathbf{x}_2)$. Since $\mathbf{u}_1$ and $\mathbf{u}_2$ are linearly independent, it also follows that $\{\mathbf{u}_1, \mathbf{u}_2\}$ is an orthonormal basis for $\text{Span}(\mathbf{x}_1, \mathbf{x}_2)$, and hence

$$\text{Span}(\mathbf{x}_1, \mathbf{x}_2) = \text{Span}(\mathbf{u}_1, \mathbf{u}_2)$$

To construct $\mathbf{u}_3$, continue in the same manner: Let $\mathbf{p}_2$ be the projection of $\mathbf{x}_3$ onto $\text{Span}(\mathbf{x}_1, \mathbf{x}_2) = \text{Span}(\mathbf{u}_1, \mathbf{u}_2)$; that is,

$$\mathbf{p}_2 = \langle \mathbf{x}_3, \mathbf{u}_1 \rangle \, \mathbf{u}_1 + \langle \mathbf{x}_3, \mathbf{u}_2 \rangle \, \mathbf{u}_2$$

and set

$$\mathbf{u}_3 = \frac{1}{\|\mathbf{x}_3 - \mathbf{p}_2\|}(\mathbf{x}_3 - \mathbf{p}_2)$$

and so on (see Figure 5.6.1).

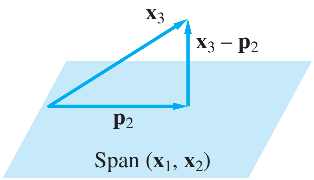

Figure 5.6.1.

Theorem 5.6.1 The Gram–Schmidt Process

Let $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n\}$ *be a basis for the inner product space* V. *Let*

$$\mathbf{u}_1 = \left(\frac{1}{\|\mathbf{x}_1\|}\right)\mathbf{x}_1$$

and define $\mathbf{u}_2, \ldots, \mathbf{u}_n$ *recursively by*

$$\mathbf{u}_{k+1} = \frac{1}{\|\mathbf{x}_{k+1} - \mathbf{p}_k\|}(\mathbf{x}_{k+1} - \mathbf{p}_k) \quad \textit{for } k = 1, \ldots, n-1$$

where

$$\mathbf{p}_k = \langle \mathbf{x}_{k+1}, \mathbf{u}_1 \rangle \mathbf{u}_1 + \langle \mathbf{x}_{k+1}, \mathbf{u}_2 \rangle \mathbf{u}_2 + \cdots + \langle \mathbf{x}_{k+1}, \mathbf{u}_k \rangle \mathbf{u}_k$$

is the projection of $\mathbf{x}_{k+1}$ *onto* $\text{Span}(\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_k)$. *Then the set*

$$\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n\}$$

is an orthonormal basis for V.

Proof We will argue inductively. Clearly, $\text{Span}(\mathbf{u}_1) = \text{Span}(\mathbf{x}_1)$. Suppose that $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_k$ have been constructed so that $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_k\}$ is an orthonormal set and

$$\text{Span}(\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_k) = \text{Span}(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k)$$

Since $\mathbf{p}_k$ is a linear combination of $\mathbf{u}_1, \ldots, \mathbf{u}_k$, it follows that $\mathbf{p}_k \in \text{Span}(\mathbf{x}_1, \ldots, \mathbf{x}_k)$ and $\mathbf{x}_{k+1} - \mathbf{p}_k \in \text{Span}(\mathbf{x}_1, \ldots, \mathbf{x}_{k+1})$.

$$\mathbf{x}_{k+1} - \mathbf{p}_k = \mathbf{x}_{k+1} - \sum_{i=1}^{k} c_i \mathbf{x}_i$$

Since $\mathbf{x}_1, \ldots, \mathbf{x}_{k+1}$ are linearly independent, it follows that $\mathbf{x}_{k+1} - \mathbf{p}_k$ is nonzero and, by Theorem 5.5.7, it is orthogonal to each $\mathbf{u}_i$, $1 \leq i \leq k$. Thus, $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_{k+1}\}$ is an orthonormal set of vectors in $\text{Span}(\mathbf{x}_1, \ldots, \mathbf{x}_{k+1})$. Since $\mathbf{u}_1, \ldots, \mathbf{u}_{k+1}$ are linearly independent, they form a basis for $\text{Span}(\mathbf{x}_1, \ldots, \mathbf{x}_{k+1})$ and, consequently,

$$\text{Span}(\mathbf{u}_1, \ldots, \mathbf{u}_{k+1}) = \text{Span}(\mathbf{x}_1, \ldots, \mathbf{x}_{k+1})$$

It follows by mathematical induction that $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n\}$ is an orthonormal basis for V. ■

EXAMPLE 1 Find an orthonormal basis for P_3 if the inner product on P_3 is defined by

$$\langle p, q \rangle = \sum_{i=1}^{3} p(x_i)q(x_i)$$

where $x_1 = -1$, $x_2 = 0$, and $x_3 = 1$.

Solution

Starting with the basis $\{1, x, x^2\}$, we can use the Gram–Schmidt process to generate an orthonormal basis.

$$\|1\|^2 = \langle 1, 1 \rangle = 3$$

so

$$\mathbf{u}_1 = \left(\frac{1}{\|1\|}\right)1 = \frac{1}{\sqrt{3}}$$

Set

$$p_1 = \left\langle x, \frac{1}{\sqrt{3}} \right\rangle \frac{1}{\sqrt{3}} = \left(-1 \cdot \frac{1}{\sqrt{3}} + 0 \cdot \frac{1}{\sqrt{3}} + 1 \cdot \frac{1}{\sqrt{3}}\right)\frac{1}{\sqrt{3}} = 0$$

Therefore,

$$x - p_1 = x \qquad \text{and} \qquad \|x - p_1\|^2 = \langle x, x \rangle = 2$$

Hence,

$$\mathbf{u}_2 = \frac{1}{\sqrt{2}}x$$

Finally,

$$p_2 = \left\langle x^2, \frac{1}{\sqrt{3}} \right\rangle \frac{1}{\sqrt{3}} + \left\langle x^2, \frac{1}{\sqrt{2}}x \right\rangle \frac{1}{\sqrt{2}}x = \frac{2}{3}$$

$$\|x^2 - p_2\|^2 = \left\langle x^2 - \frac{2}{3}, x^2 - \frac{2}{3} \right\rangle = \frac{2}{3}$$

and hence

$$\mathbf{u}_3 = \frac{\sqrt{6}}{2}\left(x^2 - \frac{2}{3}\right)$$

■

Orthogonal polynomials will be studied in more detail in Section 5.7.

EXAMPLE 2 Let

$$A = \begin{Bmatrix} 1 & -1 & 4 \\ 1 & 4 & -2 \\ 1 & 4 & 2 \\ 1 & -1 & 0 \end{Bmatrix}$$

Find an orthonormal basis for the column space of A.

Solution

The column vectors of A are linearly independent and hence form a basis for a three-dimensional subspace of $\mathbb{R}^4$. The Gram–Schmidt process can be used to construct an orthonormal basis as follows: Set

$$
\begin{aligned}
r_{11} &= \|\mathbf{a}_1\| = 2 \\
\mathbf{q}_1 &= \frac{1}{r_{11}}\mathbf{a}_1 = \left(\frac{1}{2}, \frac{1}{2}, \frac{1}{2}, \frac{1}{2}\right)^T \\
r_{12} &= \langle \mathbf{a}_2, \mathbf{q}_1 \rangle = \mathbf{q}_1^T\mathbf{a}_2 = 3 \\
\mathbf{p}_1 &= r_{12}\mathbf{q}_1 = 3\mathbf{q}_1 \\
\mathbf{a}_2 - \mathbf{p}_1 &= \left(-\frac{5}{2}, \frac{5}{2}, \frac{5}{2}, -\frac{5}{2}\right)^T \\
r_{22} &= \|\mathbf{a}_2 - \mathbf{p}_1\| = 5 \\
\mathbf{q}_2 &= \frac{1}{r_{22}}(\mathbf{a}_2 - \mathbf{p}_1) = \left(-\frac{1}{2}, \frac{1}{2}, \frac{1}{2}, -\frac{1}{2}\right)^T \\
r_{13} &= \langle \mathbf{a}_3, \mathbf{q}_1 \rangle = \mathbf{q}_1^T\mathbf{a}_3 = 2, \qquad r_{23} = \langle \mathbf{a}_3, \mathbf{q}_2 \rangle = \mathbf{q}_2^T\mathbf{a}_3 = -2 \\
\mathbf{p}_2 &= r_{13}\mathbf{q}_1 + r_{23}\mathbf{q}_2 = (2, 0, 0, 2)^T \\
\mathbf{a}_3 - \mathbf{p}_2 &= (2, -2, 2, -2)^T \\
r_{33} &= \|\mathbf{a}_3 - \mathbf{p}_2\| = 4 \\
\mathbf{q}_3 &= \frac{1}{r_{33}}(\mathbf{a}_3 - \mathbf{p}_2) = \left(\frac{1}{2}, -\frac{1}{2}, \frac{1}{2}, -\frac{1}{2}\right)^T
\end{aligned}
$$

The vectors $\mathbf{q}_1, \mathbf{q}_2, \mathbf{q}_3$ form an orthonormal basis for $R(A)$. ■

We can obtain a useful factorization of the matrix A if we keep track of all the inner products and norms computed in the Gram–Schmidt process. For the matrix in Example 2, if the r_{ij}'s are used to form a matrix

$$
R = \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ 0 & r_{22} & r_{23} \\ 0 & 0 & r_{33} \end{bmatrix} = \begin{bmatrix} 2 & 3 & 2 \\ 0 & 5 & -2 \\ 0 & 0 & 4 \end{bmatrix}
$$

and we set

$$
Q = (\mathbf{q}_1, \mathbf{q}_2, \mathbf{q}_3) = \begin{bmatrix} \frac{1}{2} & -\frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} & -\frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} & -\frac{1}{2} \end{bmatrix}
$$

then it is easily verified that $QR = A$. This result is proved in the following theorem.

Theorem 5.6.2 Gram–Schmidt QR Factorization

If A is an $m \times n$ matrix of rank n, then A can be factored into a product QR, where Q is an $m \times n$ matrix with orthonormal column vectors and R is an upper triangular $n \times n$ matrix whose diagonal entries are all positive. [Note: R must be nonsingular since $\det(R) > 0$.]

Proof Let $\mathbf{p}_1, \ldots, \mathbf{p}_{n-1}$ be the projection vectors defined in Theorem 5.6.1, and let $\{\mathbf{q}_1, \mathbf{q}_2, \ldots, \mathbf{q}_n\}$ be the orthonormal basis of $R(A)$ derived from the Gram–Schmidt process. Define

$$\begin{aligned} r_{11} &= \|\mathbf{a}_1\| \\ r_{kk} &= \|\mathbf{a}_k - \mathbf{p}_{k-1}\| \qquad \text{for} \qquad k = 2, \ldots, n \end{aligned}$$

and

$$r_{ik} = \mathbf{q}_i^T \mathbf{a}_k \qquad \text{for} \quad i = 1, \ldots, k-1 \qquad \text{and} \qquad k = 2, \ldots, n$$

By the Gram–Schmidt process,

$$\begin{aligned} r_{11}\mathbf{q}_1 &= \mathbf{a}_1 \\ r_{kk}\mathbf{q}_k &= \mathbf{a}_k - r_{1k}\mathbf{q}_1 - r_{2k}\mathbf{q}_2 - \cdots - r_{k-1,k}\mathbf{q}_{k-1} \quad \text{for} \quad k = 2, \ldots, n \end{aligned} \tag{4}$$

System (4) may be rewritten in the form

$$\begin{aligned} \mathbf{a}_1 &= r_{11}\mathbf{q}_1 \\ \mathbf{a}_2 &= r_{12}\mathbf{q}_1 + r_{22}\mathbf{q}_2 \\ &\vdots \\ \mathbf{a}_n &= r_{1n}\mathbf{q}_1 + \cdots + r_{nn}\mathbf{q}_n \end{aligned}$$

If we set

$$Q = (\mathbf{q}_1, \mathbf{q}_2, \ldots, \mathbf{q}_n)$$

and define R to be the upper triangular matrix

$$R = \begin{pmatrix} r_{11} & r_{12} & \cdots & r_{1n} \\ 0 & r_{22} & \cdots & r_{2n} \\ \vdots & & & \\ 0 & 0 & \cdots & r_{nn} \end{pmatrix}$$

then the jth column of the product QR will be

$$Q\mathbf{r}_j = r_{1j}\mathbf{q}_1 + r_{2j}\mathbf{q}_2 + \cdots + r_{jj}\mathbf{q}_j = \mathbf{a}_j$$

for $j = 1, \ldots, n$. Therefore,

$$QR = (\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_n) = A$$ ■

EXAMPLE 3 Compute the Gram–Schmidt QR factorization of the matrix

$$A = \begin{pmatrix} 1 & -2 & -1 \\ 2 & 0 & 1 \\ 2 & -4 & 2 \\ 4 & 0 & 0 \end{pmatrix}$$

Solution

Step 1. Set

$$r_{11} = \|\mathbf{a}_1\| = 5$$

$$\mathbf{q}_1 = \frac{1}{r_{11}}\mathbf{a}_1 = \left(\frac{1}{5}, \frac{2}{5}, \frac{2}{5}, \frac{4}{5}\right)^T$$

Step 2. Set

$$\begin{aligned} r_{12} &= \mathbf{q}_1^T\mathbf{a}_2 = -2 \\ \mathbf{p}_1 &= r_{12}\mathbf{q}_1 = -2\mathbf{q}_1 \\ \mathbf{a}_2 - \mathbf{p}_1 &= \left(-\frac{8}{5}, \frac{4}{5}, -\frac{16}{5}, \frac{8}{5}\right)^T \\ r_{22} &= \|\mathbf{a}_2 - \mathbf{p}_1\| = 4 \\ \mathbf{q}_2 &= \frac{1}{r_{22}}(\mathbf{a}_2 - \mathbf{p}_1) = \left(-\frac{2}{5}, \frac{1}{5}, -\frac{4}{5}, \frac{2}{5}\right)^T \end{aligned}$$

Step 3. Set

$$\begin{aligned} r_{13} &= \mathbf{q}_1^T\mathbf{a}_3 = 1, \qquad r_{23} = \mathbf{q}_2^T\mathbf{a}_3 = -1 \\ \mathbf{p}_2 &= r_{13}\mathbf{q}_1 + r_{23}\mathbf{q}_2 = \mathbf{q}_1 - \mathbf{q}_2 = \left(\frac{3}{5}, \frac{1}{5}, \frac{6}{5}, \frac{2}{5}\right)^T \\ \mathbf{a}_3 - \mathbf{p}_2 &= \left(-\frac{8}{5}, \frac{4}{5}, \frac{4}{5}, -\frac{2}{5}\right)^T \\ r_{33} &= \|\mathbf{a}_3 - \mathbf{p}_2\| = 2 \\ \mathbf{q}_3 &= \frac{1}{r_{33}}(\mathbf{a}_3 - \mathbf{p}_2) = \left(-\frac{4}{5}, \frac{2}{5}, \frac{2}{5}, -\frac{1}{5}\right)^T \end{aligned}$$

At each step, we have determined a column of Q and a column of R. The factorization is given by

$$A = QR = \begin{bmatrix} \frac{1}{5} & -\frac{2}{5} & -\frac{4}{5} \\ \frac{2}{5} & \frac{1}{5} & \frac{2}{5} \\ \frac{2}{5} & -\frac{4}{5} & \frac{2}{5} \\ \frac{4}{5} & \frac{2}{5} & -\frac{1}{5} \end{bmatrix} \begin{bmatrix} 5 & -2 & 1 \\ 0 & 4 & -1 \\ 0 & 0 & 2 \end{bmatrix}$$ ■

We saw in Section 5.5 that if the columns of an $m \times n$ matrix A form an orthonormal set, then the least squares solution of $A\mathbf{x} = \mathbf{b}$ is simply $\hat{\mathbf{x}} = A^T\mathbf{b}$. If A has rank n, but its column vectors do not form an orthonormal set in $\mathbb{R}^m$, then the QR factorization can be used to solve the least squares problem.

Theorem 5.6.3 *If A is an $m \times n$ matrix of rank n, then the least squares solution of $A\mathbf{x} = \mathbf{b}$ is given by $\hat{\mathbf{x}} = R^{-1}Q^T\mathbf{b}$, where Q and R are the matrices obtained from the factorization given in Theorem 5.6.2. The solution $\hat{\mathbf{x}}$ may be obtained by using back substitution to solve $R\mathbf{x} = Q^T\mathbf{b}$.*