6.4 The Gram-Schmidt Procedure

Now that we understand the usefulness of orthonormal bases, how do we go about finding them? For example, does $\mathbb{P}_m$, with inner product given by integration on $[0, 1]$ have an orthonormal basis? As we will see, the next result will lead to answers to these questions. The algorithm used in the next proof is called the *Gram-Schmidt procedure*. It gives a method for turning a linearly independent list into an orthonormal list with the same span as the original list.

Theorem 11 (Gram-Schmidt). *If $\{\mathbf{x}_1, \dots, \mathbf{x}_p\}$ is a linearly independent list of vectors in W, then there exists an orthogonal list $\{\mathbf{v}_1, \dots, \mathbf{v}_p\}$ of vectors in W such that*

$$\text{Span}\{\mathbf{x}_1, \dots, \mathbf{x}_j\} = \text{Span}\{\mathbf{v}_1, \dots, \mathbf{v}_j\} \tag{1}$$

for $j = 1, \dots, p$. More specifically,

$$\begin{aligned}
\mathbf{v}_1 &= \mathbf{x}_1 \\
\mathbf{v}_2 &= \mathbf{x}_2 - \frac{\langle \mathbf{x}_2, \mathbf{v}_1 \rangle}{\langle \mathbf{v}_1, \mathbf{v}_1 \rangle}\mathbf{v}_1 \\
\mathbf{v}_3 &= \mathbf{x}_3 - \frac{\langle \mathbf{x}_3, \mathbf{v}_1 \rangle}{\langle \mathbf{v}_1, \mathbf{v}_1 \rangle}\mathbf{v}_1 - \frac{\langle \mathbf{x}_3, \mathbf{v}_2 \rangle}{\langle \mathbf{v}_2, \mathbf{v}_2 \rangle}\mathbf{v}_2 \\
&\vdots \\
\mathbf{v}_p &= \mathbf{x}_p - \frac{\langle \mathbf{x}_p, \mathbf{v}_1 \rangle}{\langle \mathbf{v}_1, \mathbf{v}_1 \rangle}\mathbf{v}_1 - \frac{\langle \mathbf{x}_p, \mathbf{v}_2 \rangle}{\langle \mathbf{v}_2, \mathbf{v}_2 \rangle}\mathbf{v}_2 - \cdots - \frac{\langle \mathbf{x}_p, \mathbf{v}_{p-1} \rangle}{\langle \mathbf{v}_{p-1}, \mathbf{v}_{p-1} \rangle}\mathbf{v}_{p-1}
\end{aligned}$$

Then $\{\mathbf{v}_1, \dots, \mathbf{v}_p\}$ is an orthogonal basis for V. Normalizing each $\mathbf{v}_j$ results in an orthonormal basis.

Proof. For $1 \leq k \leq p$, let $W_k = \text{Span}\{\mathbf{x}_1, \dots, \mathbf{x}_k\}$. Set $\mathbf{v}_1 = \mathbf{x}_1$, so that $\text{Span}\{\mathbf{v}_1\} = \text{Span}\{\mathbf{x}_1\}$.

Suppose for some $k < p$, we have constructed $\mathbf{v}_1, \dots, \mathbf{v}_k$ so that $\{\mathbf{v}_1, \dots, \mathbf{v}_k\}$ is an orthogonal basis for W_k. Define

$$\mathbf{v}_{k+1} = \mathbf{x}_{k+1} - P_{W_k}\mathbf{x}_{k+1}.$$

By the Orthogonal Decomposition Theorem, $\mathbf{v}_{k+1}$ is orthogonal to W_k. Note that $P_{W_k}\mathbf{x}_{k+1}$ is in W_k and hence also in W_{k+1}. Since $\mathbf{x}_{k+1}$ is in W_{k+1}, so is $\mathbf{v}_{k+1}$, because W_{k+1} is a subspace and is closed under subtraction. Furthermore, $\mathbf{v}_{k+1} \neq \mathbf{0}$ because $\mathbf{x}_{k+1}$ is not in $W_k = \text{Span}\{\mathbf{x}_1, \dots, \mathbf{x}_k\}$. Hence $\{\mathbf{v}_1, \dots, \mathbf{v}_{k+1}\}$ is an orthogonal set of nonzero vectors in the $(k+1)$-dimensional space W_{k+1}. By the Basis Theorem in §4.5, this set is an orthogonal basis for W_{k+1}. Hence $W_{k+1} = \text{Span}\{\mathbf{v}_1, \dots, \mathbf{v}_{k+1}\}$.

When $k + 1 = p$, the procedure stops. We may form an orthonormal basis from an orthogonal basis by simply normalizing each vector in the orthogonal basis after finishing Gram-Schmidt. □

Corollary. *Every finite-dimensional inner-product space has an orthonormal basis.*

Proof. Choose a basis of V. Apply the Gram-Schmidt procedure to it, producing an orthonormal list. This orthonormal list is linearly independent and its span equals V. Thus it is an orthonormal basis of V. □

Corollary. *Every orthonormal list of vectors in V can be extended to an orthonormal basis of V.*

Proof. Suppose $\{\mathbf{e}_1, \dots, \mathbf{e}_m\}$ is an orthonormal list of vectors in V. Then $\{\mathbf{e}_1, \dots, \mathbf{e}_m\}$ is linearly independent, and hence it can be extended to a basis $\{\mathbf{e}_1, \dots, \mathbf{e}_m, \mathbf{v}_1, \dots, \mathbf{v}_n\}$ of V. Now apply the Gram-Schmidt procedure to $\{\mathbf{e}_1, \dots, \mathbf{e}_m, \mathbf{v}_1, \dots, \mathbf{v}_n\}$, producing an orthonormal list

$$\{\mathbf{e}_1, \dots, \mathbf{e}_m, \mathbf{f}_1, \dots, \mathbf{f}_n\}; \tag{2}$$

here the Gram-Schmidt procedure leaves the first m vectors unchanged because they are already orthonormal. Clearly (2) is an orthonormal basis of V because it is linearly independent and its span equals V. Hence we have our extension of $\{\mathbf{e}_1, \dots, \mathbf{e}_m\}$ to an orthonormal basis of V. □

Theorem 12 (The QR Factorization). *If A is an $m \times n$ matrix with linearly independent columns, then A can be factored as $A = QR$, where Q is an $m \times n$ matrix whose columns form an orthonormal basis for* Col A *and R is an $n \times n$ upper-triangular invertible matrix with positive entries on its diagonal.*

Example 1. *Use Gram-Schmidt procedure to produce an orthonormal basis for* $W = \text{Span}\left\{ \begin{bmatrix} 3 \\ -4 \\ 5 \end{bmatrix}, \begin{bmatrix} -3 \\ 14 \\ -7 \end{bmatrix} \right\}$.

Example 2. *As an illustration of this procedure, consider the problem of finding a polynomial $\mathbf{u}$ with real coefficients and degree at most 5 that on the interval $[-\pi, \pi]$ approximates $\sin x$ as well as possible, in the sense that*

$$\int_{-\pi}^{\pi} |\sin x - \mathbf{u}(x)|^2 dx$$

is as small as possible. To solve this problem, let $C[-\pi, \pi]$ denote the real vector space of continuous real-valued functions on $[-\pi, \pi]$ with inner product

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

Let $\mathbf{v} \in C[-\pi, \pi]$ be the function defined by $\mathbf{v}(x) = \sin x$. Let U denote the subspace of $C[-\pi, \pi]$ consisting of the polynomials with real coefficients and degree at most 5. Our problem can now be reformulated as follows: find $\mathbf{u} \in U$ such that $\|\mathbf{v} - \mathbf{u}\|$ is as small as possible.

Solution. To compute the solution to our approximation problem, first apply the Gram-Schmidt procedure, using the inner product given by (3) to the basis $\{1, x, x^2, x^3, x^4, x^5\}$ of U, producing an orthonormal basis $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3, \mathbf{e}_4, \mathbf{e}_5, \mathbf{e}_6\}$ of U. Then, again using the inner product given by (3), compute $P_U\mathbf{v}$ using

$$P_U\mathbf{v} = \langle \mathbf{v}, \mathbf{e}_1 \rangle \mathbf{e}_1 + \cdots + \langle \mathbf{v}, \mathbf{e}_m \rangle \mathbf{e}_m$$

with $m = 6$. Doing this computation shows that $P_U\mathbf{v}$ is the function

$$0.987862x - 0.155271x^3 + 0.00564312x^5, \tag{4}$$

where the π's that appear in the exact answer have been replaced with a good decimal approximation.

By The Best Approximation Theorem, the polynomial above should be about as good an approximation to $\sin x$ on $[-\pi, \pi]$ as is possible using polynomials of degree at most 5. To see how good this approximation is, we may compare the graphs of both $\sin x$ and our approximation (4) over the interval $[-\pi, \pi]$.

Our approximation (4) is so accurate that the two graphs are almost identical - our eyes may see only one graph!

Another well-known approximation to $\sin x$ by a polynomial of degree 5 is given by the Taylor polynomial

$$x - \frac{x^3}{3!} + \frac{x^5}{5!}. \tag{5}$$

To see how good this approximation is, we may compare the graphs of both $\sin x$ and the Taylor polynomial (5) over the interval $[-\pi, \pi]$.

The Taylor polynomial is an excellent approximation to $\sin x$ for x near 0. But for $|x| > 2$, the Taylor polynomial is not so accurate, especially compared to (4). For example, taking $x = 3$, our approximation (4) estimates $\sin 3$ with an error of about 0.001, but the Taylor series (5) estimates $\sin 3$ with an error of about 0.4. Thus at $x = 3$, the error in the Taylor series is hundreds of times larger than the error given by (4). Linear algebra has helped us discover an approximation to $\sin x$ that improves upon what we learned in calculus! □