8. Orthogonality

In Section 5.3 we introduced the dot product in \( R^n \) and extended the basic geometric notions of length and distance. A set \( \{f_1, f_2, \ldots, f_m\} \) of nonzero vectors in \( R^n \) was called an **orthogonal set** if \( f_i \cdot f_j = 0 \) for all \( i \neq j \), and it was proved that every orthogonal set is independent. In particular, it was observed that the expansion of a vector as a linear combination of orthogonal basis vectors is easy to obtain because formulas exist for the coefficients. Hence the orthogonal bases are the “nice” bases, and much of this chapter is devoted to extending results about bases to orthogonal bases. This leads to some very powerful methods and theorems. Our first task is to show that every subspace of \( R^n \) has an orthogonal basis.

8.1 Orthogonal Complements and Projections

If \( \{v_1, \ldots, v_m\} \) is linearly independent in a general vector space, and if \( v_{m+1} \) is not in span \( \{v_1, \ldots, v_m\} \), then \( \{v_1, \ldots, v_m, v_{m+1}\} \) is independent (Lemma 6.4.1). Here is the analog for orthogonal sets in \( R^n \).

**Lemma 8.1.1: Orthogonal Lemma**

Let \( \{f_1, f_2, \ldots, f_m\} \) be an orthogonal set in \( R^n \). Given \( x \) in \( R^n \), write

\[
f_{m+1} = x - \frac{x \cdot f_1}{\|f_1\|^2}f_1 - \frac{x \cdot f_2}{\|f_2\|^2}f_2 - \cdots - \frac{x \cdot f_m}{\|f_m\|^2}f_m
\]

Then:

1. \( f_{m+1} \cdot f_k = 0 \) for \( k = 1, 2, \ldots, m \).

2. If \( x \) is not in span \( \{f_1, \ldots, f_m\} \), then \( f_{m+1} \neq 0 \) and \( \{f_1, \ldots, f_m, f_{m+1}\} \) is an orthogonal set.

**Proof.** For convenience, write \( t_i = (x \cdot f_i)/\|f_i\|^2 \) for each \( i \). Given \( 1 \leq k \leq m \):

\[
f_{m+1} \cdot f_k = (x - t_1f_1 - \cdots - t_kf_k - \cdots - t_mf_m) \cdot f_k
\]

\[
= x \cdot f_k - t_1(f_1 \cdot f_k) - \cdots - t_k(f_k \cdot f_k) - \cdots - t_m(f_m \cdot f_k)
\]

\[
= x \cdot f_k - t_k\|f_k\|^2
\]

\[
= 0
\]

This proves (1), and (2) follows because \( f_{m+1} \neq 0 \) if \( x \) is not in span \( \{f_1, \ldots, f_m\} \).

The orthogonal lemma has three important consequences for \( R^n \). The first is an extension for orthogonal sets of the fundamental fact that any independent set is part of a basis (Theorem 6.4.1).
Theorem 8.1.1
Let $U$ be a subspace of $\mathbb{R}^n$.

1. Every orthogonal subset $\{f_1, \ldots, f_m\}$ in $U$ is a subset of an orthogonal basis of $U$.

2. $U$ has an orthogonal basis.

Proof.

1. If $\text{span}\{f_1, \ldots, f_m\} = U$, it is already a basis. Otherwise, there exists $x$ in $U$ outside $\text{span}\{f_1, \ldots, f_m\}$. If $f_{m+1}$ is as given in the orthogonal lemma, then $f_{m+1}$ is in $U$ and $\{f_1, \ldots, f_m, f_{m+1}\}$ is orthogonal. If $\text{span}\{f_1, \ldots, f_m, f_{m+1}\} = U$, we are done. Otherwise, the process continues to create larger and larger orthogonal subsets of $U$. They are all independent by Theorem 5.3.5, so we have a basis when we reach a subset containing $\dim U$ vectors.

2. If $U = \{0\}$, the empty basis is orthogonal. Otherwise, if $f \neq 0$ is in $U$, then $\{f\}$ is orthogonal, so (2) follows from (1). \qed

We can improve upon (2) of Theorem 8.1.1. In fact, the second consequence of the orthogonal lemma is a procedure by which any basis $\{x_1, \ldots, x_m\}$ of a subspace $U$ of $\mathbb{R}^n$ can be systematically modified to yield an orthogonal basis $\{f_1, \ldots, f_m\}$ of $U$. The $f_i$ are constructed one at a time from the $x_i$.

To start the process, take $f_1 = x_1$. Then $x_2$ is not in $\text{span}\{f_1\}$ because $\{x_1, x_2\}$ is independent, so take

$$f_2 = x_2 - \frac{x_2 \cdot f_1}{\|f_1\|^2} f_1$$

Thus $\{f_1, f_2\}$ is orthogonal by Lemma 8.1.1. Moreover, $\text{span}\{f_1, f_2\} = \text{span}\{x_1, x_2\}$ (verify), so $x_3$ is not in $\text{span}\{f_1, f_2\}$. Hence $\{f_1, f_2, f_3\}$ is orthogonal where

$$f_3 = x_3 - \frac{x_3 \cdot f_1}{\|f_1\|^2} f_1 - \frac{x_3 \cdot f_2}{\|f_2\|^2} f_2$$

Again, $\text{span}\{f_1, f_2, f_3\} = \text{span}\{x_1, x_2, x_3\}$, so $x_4$ is not in $\text{span}\{f_1, f_2, f_3\}$ and the process continues. At the $m$th iteration we construct an orthogonal set $\{f_1, \ldots, f_m\}$ such that

$$\text{span}\{f_1, f_2, \ldots, f_m\} = \text{span}\{x_1, x_2, \ldots, x_m\} = U$$

Hence $\{f_1, f_2, \ldots, f_m\}$ is the desired orthogonal basis of $U$. The procedure can be summarized as follows.
Theorem 8.1.2: Gram-Schmidt Orthogonalization Algorithm

If \( \{x_1, x_2, \ldots, x_m\} \) is any basis of a subspace \( U \) of \( \mathbb{R}^n \), construct \( f_1, f_2, \ldots, f_m \) in \( U \) successively as follows:

\[
\begin{align*}
    f_1 &= x_1 \\
    f_2 &= x_2 - \frac{x_2}{\|x_1\|^2} f_1 \\
    f_3 &= x_3 - \frac{x_3}{\|x_1\|^2} f_1 - \frac{x_3}{\|x_2\|^2} f_2 \\
    &\quad \vdots \\
    f_k &= x_k - \frac{x_k}{\|x_1\|^2} f_1 - \frac{x_k}{\|x_2\|^2} f_2 - \cdots - \frac{x_k}{\|x_{k-1}\|^2} f_{k-1}
\end{align*}
\]

for each \( k = 2, 3, \ldots, m \). Then

1. \( \{f_1, f_2, \ldots, f_m\} \) is an orthogonal basis of \( U \).
2. \( \text{span} \{f_1, f_2, \ldots, f_k\} = \text{span} \{x_1, x_2, \ldots, x_k\} \) for each \( k = 1, 2, \ldots, m \).

The process (for \( k = 3 \)) is depicted in the diagrams. Of course, the algorithm converts any basis of \( \mathbb{R}^n \) itself into an orthogonal basis.

Example 8.1.1

Find an orthogonal basis of the row space of \( A = \begin{bmatrix} 1 & 1 & -1 & -1 \\ 3 & 2 & 0 & 1 \\ 1 & 0 & 1 & 0 \end{bmatrix} \).

Solution. Let \( x_1, x_2, x_3 \) denote the rows of \( A \) and observe that \( \{x_1, x_2, x_3\} \) is linearly independent. Take \( f_1 = x_1 \). The algorithm gives

\[
\begin{align*}
    f_2 &= x_2 - \frac{x_2}{\|x_1\|^2} f_1 = (3, 2, 0, 1) - \frac{4}{4} (1, 1, -1, -1) = (2, 1, 1, 2) \\
    f_3 &= x_3 - \frac{x_3}{\|x_1\|^2} f_1 - \frac{x_3}{\|x_2\|^2} f_2 = x_3 - \frac{0}{4} f_1 - \frac{3}{10} f_2 = \frac{1}{10} (4, -3, 7, -6)
\end{align*}
\]

Hence \( \{(1, 1, -1, -1), (2, 1, 1, 2), \frac{1}{10} (4, -3, 7, -6)\} \) is the orthogonal basis provided by the algorithm. In hand calculations it may be convenient to eliminate fractions (see the Remark below), so \( \{(1, 1, -1, -1), (2, 1, 1, 2), (4, -3, 7, -6)\} \) is also an orthogonal basis for row \( A \).

---

1 Erhardt Schmidt (1876–1959) was a German mathematician who studied under the great David Hilbert and later developed the theory of Hilbert spaces. He first described the present algorithm in 1907. Jörgen Pederson Gram (1850–1916) was a Danish actuary.
Remark

Observe that the vector \( \frac{x}{\|f_i\|^2} f_i \) is unchanged if a nonzero scalar multiple of \( f_i \) is used in place of \( f_i \). Hence, if a newly constructed \( f_i \) is multiplied by a nonzero scalar at some stage of the Gram-Schmidt algorithm, the subsequent \( f_s \) will be unchanged. This is useful in actual calculations.

Projections

Suppose a point \( x \) and a plane \( U \) through the origin in \( \mathbb{R}^3 \) are given, and we want to find the point \( p \) in the plane that is closest to \( x \). Our geometric intuition assures us that such a point \( p \) exists. In fact (see the diagram), \( p \) must be chosen in such a way that \( x - p \) is perpendicular to the plane. Now we make two observations: first, the plane \( U \) is a subspace of \( \mathbb{R}^3 \) (because \( U \) contains the origin); and second, that the condition that \( x - p \) is perpendicular to the plane \( U \) means that \( x - p \) is orthogonal to every vector in \( U \). In these terms the whole discussion makes sense in \( \mathbb{R}^n \). Furthermore, the orthogonal lemma provides exactly what is needed to find \( p \) in this more general setting.

**Definition 8.1 Orthogonal Complement of a Subspace of \( \mathbb{R}^n \)**

If \( U \) is a subspace of \( \mathbb{R}^n \), define the **orthogonal complement** \( U^\perp \) of \( U \) (pronounced “\( U \)-perp”) by

\[
U^\perp = \{ x \in \mathbb{R}^n \mid x \cdot y = 0 \text{ for all } y \text{ in } U \}
\]

The following lemma collects some useful properties of the orthogonal complement; the proof of (1) and (2) is left as Exercise 8.1.6.

**Lemma 8.1.2**

Let \( U \) be a subspace of \( \mathbb{R}^n \).

1. \( U^\perp \) is a subspace of \( \mathbb{R}^n \).

2. \( \{0\}^\perp = \mathbb{R}^n \) and \( (\mathbb{R}^n)^\perp = \{0\} \).

3. If \( U = \text{span} \{x_1, x_2, \ldots, x_k\} \), then \( U^\perp = \{ x \in \mathbb{R}^n \mid x \cdot x_i = 0 \text{ for } i = 1, 2, \ldots, k \} \).

**Proof.**

3. Let \( U = \text{span} \{x_1, x_2, \ldots, x_k\} \); we must show that \( U^\perp = \{ x \mid x \cdot x_i = 0 \text{ for each } i \} \). If \( x \) is in \( U^\perp \) then \( x \cdot x_i = 0 \) for all \( i \) because each \( x_i \) is in \( U \). Conversely, suppose that \( x \cdot x_i = 0 \) for all \( i \); we must show that \( x \) is in \( U^\perp \), that is, \( x \cdot y = 0 \) for each \( y \) in \( U \). Write \( y = r_1 x_1 + r_2 x_2 + \cdots + r_k x_k \), where each \( r_i \) is in \( \mathbb{R} \). Then, using Theorem 5.3.1,

\[
x \cdot y = r_1 (x \cdot x_1) + r_2 (x \cdot x_2) + \cdots + r_k (x \cdot x_k) = r_1 0 + r_2 0 + \cdots + r_k 0 = 0
\]

as required. \( \square \)
Example 8.1.2

Find $U^\perp$ if $U = \text{span} \{(1, -1, 2, 0), (1, 0, -2, 3)\}$ in $\mathbb{R}^4$.

**Solution.** By Lemma 8.1.2, $x = (x, y, z, w)$ is in $U^\perp$ if and only if it is orthogonal to both $(1, -1, 2, 0)$ and $(1, 0, -2, 3)$; that is,

$$
x - y + 2z = 0
$$
$$
x - 2z + 3w = 0
$$

Gaussian elimination gives $U^\perp = \text{span} \{(2, 4, 1, 0), (3, 3, 0, -1)\}$.

Now consider vectors $x$ and $d \neq 0$ in $\mathbb{R}^3$. The projection $p = \text{proj}_d x$ of $x$ on $d$ was defined in Section 4.2 as in the diagram.

The following formula for $p$ was derived in Theorem 4.2.4

$$
p = \text{proj}_d x = \left( \frac{x^d}{||d||^2} \right) d
$$

where it is shown that $x - p$ is orthogonal to $d$. Now observe that the line $U = \mathbb{R}d = \{rd \mid r \in \mathbb{R}\}$ is a subspace of $\mathbb{R}^3$, that $\{d\}$ is an orthogonal basis of $U$, and that $p \in U$ and $x - p \in U^\perp$ (by Theorem 4.2.4).

In this form, this makes sense for any vector $x$ in $\mathbb{R}^n$ and any subspace $U$ of $\mathbb{R}^n$, so we generalize it as follows. If $\{f_1, f_2, \ldots, f_m\}$ is an orthogonal basis of $U$, we define the projection $p$ of $x$ on $U$ by the formula

$$
p = \left( \frac{x^f_1}{||f_1||^2} \right) f_1 + \left( \frac{x^f_2}{||f_2||^2} \right) f_2 + \cdots + \left( \frac{x^f_m}{||f_m||^2} \right) f_m
$$

Then $p \in U$ and (by the orthogonal lemma) $x - p \in U^\perp$, so it looks like we have a generalization of Theorem 4.2.4.

However there is a potential problem: the formula (8.1) for $p$ must be shown to be independent of the choice of the orthogonal basis $\{f_1, f_2, \ldots, f_m\}$. To verify this, suppose that $\{f_1', f_2', \ldots, f_m'\}$ is another orthogonal basis of $U$, and write

$$
p' = \left( \frac{x^{f_1'}}{||f_1'||^2} \right) f_1' + \left( \frac{x^{f_2'}}{||f_2'||^2} \right) f_2' + \cdots + \left( \frac{x^{f_m'}}{||f_m'||^2} \right) f_m'
$$

As before, $p' \in U$ and $x - p' \in U^\perp$, and we must show that $p' = p$. To see this, write the vector $p - p'$ as follows:

$$p - p' = (x - p') - (x - p)$$

This vector is in $U$ (because $p$ and $p'$ are in $U$) and it is in $U^\perp$ (because $x - p'$ and $x - p$ are in $U^\perp$), and so it must be zero (it is orthogonal to itself!). This means $p' = p$ as desired.

Hence, the vector $p$ in equation (8.1) depends only on $x$ and the subspace $U$, and not on the choice of orthogonal basis $\{f_1, \ldots, f_m\}$ of $U$ used to compute it. Thus, we are entitled to make the following definition:
Orthogonality

Definition 8.2 Projection onto a Subspace of \( \mathbb{R}^n \)

Let \( U \) be a subspace of \( \mathbb{R}^n \) with orthogonal basis \( \{ f_1, f_2, \ldots, f_m \} \). If \( x \) is in \( \mathbb{R}^n \), the vector

\[
\text{proj}_U x = \frac{x \cdot f_1}{\|f_1\|^2} f_1 + \frac{x \cdot f_2}{\|f_2\|^2} f_2 + \cdots + \frac{x \cdot f_m}{\|f_m\|^2} f_m
\]

is called the orthogonal projection of \( x \) on \( U \). For the zero subspace \( U = \{0\} \), we define

\[
\text{proj}_{\{0\}} x = 0
\]

The preceding discussion proves (1) of the following theorem.

Theorem 8.1.3: Projection Theorem

If \( U \) is a subspace of \( \mathbb{R}^n \) and \( x \) is in \( \mathbb{R}^n \), write \( p = \text{proj}_U x \). Then:

1. \( p \) is in \( U \) and \( x - p \) is in \( U^\perp \).
2. \( p \) is the vector in \( U \) closest to \( x \) in the sense that

\[
\|x - p\| < \|x - y\| \quad \text{for all } y \in U, y \neq p
\]

Proof.

1. This is proved in the preceding discussion (it is clear if \( U = \{0\} \)).

2. Write \( x - y = (x - p) + (p - y) \). Then \( p - y \) is in \( U \) and so is orthogonal to \( x - p \) by (1). Hence, the Pythagorean theorem gives

\[
\|x - y\|^2 = \|x - p\|^2 + \|p - y\|^2 > \|x - p\|^2
\]

because \( p - y \neq 0 \). This gives (2). \( \square \)

Example 8.1.3

Let \( U = \text{span} \{x_1, x_2\} \) in \( \mathbb{R}^4 \) where \( x_1 = (1, 1, 0, 1) \) and \( x_2 = (0, 1, 1, 2) \). If \( x = (3, -1, 0, 2) \), find the vector in \( U \) closest to \( x \) and express \( x \) as the sum of a vector in \( U \) and a vector orthogonal to \( U \).

Solution. \( \{x_1, x_2\} \) is independent but not orthogonal. The Gram-Schmidt process gives an orthogonal basis \( \{f_1, f_2\} \) of \( U \) where \( f_1 = x_1 = (1, 1, 0, 1) \) and

\[
f_2 = x_2 - \frac{x_2 \cdot f_1}{\|f_1\|^2} f_1 = x_2 - \frac{3}{3} f_1 = (-1, 0, 1, 1)
\]

Hence, we can compute the projection using \( \{f_1, f_2\} \):

\[
p = \text{proj}_U x = \frac{x \cdot f_1}{\|f_1\|^2} f_1 + \frac{x \cdot f_2}{\|f_2\|^2} f_2 = \frac{4}{3} f_1 + \frac{1}{3} f_2 = \frac{1}{3} \begin{bmatrix} 5 & 4 & -1 & 3 \end{bmatrix}
\]
Thus, \( \mathbf{p} \) is the vector in \( U \) closest to \( \mathbf{x} \), and \( \mathbf{x} - \mathbf{p} = \frac{1}{3}(4, -7, 1, 3) \) is orthogonal to every vector in \( U \). (This can be verified by checking that it is orthogonal to the generators \( \mathbf{x}_1 \) and \( \mathbf{x}_2 \) of \( U \).) The required decomposition of \( \mathbf{x} \) is thus

\[
\mathbf{x} = \mathbf{p} + (\mathbf{x} - \mathbf{p}) = \frac{1}{3}(5, 4, -1, 3) + \frac{1}{3}(4, -7, 1, 3)
\]

Example 8.1.4

Find the point in the plane with equation \( 2x + y - z = 0 \) that is closest to the point \( (2, -1, -3) \).

Solution. We write \( \mathbb{R}^3 \) as rows. The plane is the subspace \( U \) whose points \( (x, y, z) \) satisfy \( z = 2x + y \). Hence

\[
U = \{(s, t, 2s + t) \mid s, t \text{ in } \mathbb{R}\} = \text{span}\{(0, 1, 1), (1, 0, 2)\}
\]

The Gram-Schmidt process produces an orthogonal basis \( \{\mathbf{f}_1, \mathbf{f}_2\} \) of \( U \) where \( \mathbf{f}_1 = (0, 1, 1) \) and \( \mathbf{f}_2 = (1, -1, 1) \). Hence, the vector in \( U \) closest to \( \mathbf{x} = (2, -1, -3) \) is

\[
\text{proj}_U \mathbf{x} = \frac{\mathbf{x} \cdot \mathbf{f}_1}{\|\mathbf{f}_1\|^2} \mathbf{f}_1 + \frac{\mathbf{x} \cdot \mathbf{f}_2}{\|\mathbf{f}_2\|^2} \mathbf{f}_2 = \frac{2}{3} \mathbf{f}_1 + \frac{1}{3} \mathbf{f}_2 = (0, -2, -2)
\]

Thus, the point in \( U \) closest to \( (2, -1, -3) \) is \( (0, -2, -2) \).

The next theorem shows that projection on a subspace of \( \mathbb{R}^n \) is actually a linear operator \( \mathbb{R}^n \to \mathbb{R}^n \).

Theorem 8.1.4

Let \( U \) be a fixed subspace of \( \mathbb{R}^n \). If we define \( T : \mathbb{R}^n \to \mathbb{R}^n \) by

\[
T(\mathbf{x}) = \text{proj}_U \mathbf{x} \quad \text{for all } \mathbf{x} \text{ in } \mathbb{R}^n
\]

1. \( T \) is a linear operator.
2. \( \text{im } T = U \) and \( \ker T = U^\perp \).
3. \( \dim U + \dim U^\perp = n \).

Proof. If \( U = \{\mathbf{0}\} \), then \( U^\perp = \mathbb{R}^n \), and so \( T(\mathbf{x}) = \text{proj}_{\{\mathbf{0}\}} \mathbf{x} = \mathbf{0} \) for all \( \mathbf{x} \). Thus \( T = 0 \) is the zero (linear) operator, so (1), (2), and (3) hold. Hence assume that \( U \neq \{\mathbf{0}\} \).

1. If \( \{\mathbf{f}_1, \mathbf{f}_2, \ldots, \mathbf{f}_m\} \) is an orthonormal basis of \( U \), then

\[
T(\mathbf{x}) = (\mathbf{x} \cdot \mathbf{f}_1)\mathbf{f}_1 + (\mathbf{x} \cdot \mathbf{f}_2)\mathbf{f}_2 + \cdots + (\mathbf{x} \cdot \mathbf{f}_m)\mathbf{f}_m \quad \text{for all } \mathbf{x} \text{ in } \mathbb{R}^n
\]

by the definition of the projection. Thus \( T \) is linear because

\[
(\mathbf{x} + \mathbf{y}) \cdot \mathbf{f}_i = \mathbf{x} \cdot \mathbf{f}_i + \mathbf{y} \cdot \mathbf{f}_i \quad \text{and} \quad (r\mathbf{x}) \cdot \mathbf{f}_i = r(\mathbf{x} \cdot \mathbf{f}_i) \quad \text{for each } i
\]
2. We have \( \text{im } T \subseteq U \) by (8.2) because each \( f_i \) is in \( U \). But if \( \mathbf{x} \) is in \( U \), then \( \mathbf{x} = T(\mathbf{x}) \) by (8.2) and the expansion theorem applied to the space \( U \). This shows that \( U \subseteq \text{im } T \), so \( \text{im } T = U \).

Now suppose that \( \mathbf{x} \) is in \( U^\perp \). Then \( \mathbf{x} \cdot f_i = 0 \) for each \( i \) (again because each \( f_i \) is in \( U \)) so \( \mathbf{x} \) is in \( \ker T \) by (8.2). Hence \( U^\perp \subseteq \ker T \). On the other hand, Theorem 8.1.3 shows that \( \mathbf{x} - T(\mathbf{x}) \) is in \( U^\perp \) for all \( \mathbf{x} \) in \( \mathbb{R}^n \), and it follows that \( \ker T \subseteq U^\perp \). Hence \( \ker T = U^\perp \), proving (2).

3. This follows from (1), (2), and the dimension theorem (Theorem 7.2.4).

\( \square \)

### Exercises for 8.1

**Exercise 8.1.1** In each case, use the Gram-Schmidt algorithm to convert the given basis \( B \) of \( V \) into an orthogonal basis.

- a. \( V = \mathbb{R}^2, B = \{(1, -1), (2, 1)\} \)
- b. \( V = \mathbb{R}^2, B = \{(2, 1), (1, 2)\} \)
- c. \( V = \mathbb{R}^3, B = \{(1, -1, 1), (1, 0, 1), (1, 1, 2)\} \)
- d. \( V = \mathbb{R}^3, B = \{(0, 1, 1), (1, 1, 1), (1, -2, 2)\} \)

**Exercise 8.1.2** In each case, write \( \mathbf{x} \) as the sum of a vector in \( U \) and a vector in \( U^\perp \).

- a. \( \mathbf{x} = (1, 5, 7), U = \text{span}\{(1, -2, 3), (-1, 1, 1)\} \)
- b. \( \mathbf{x} = (2, 1, 6), U = \text{span}\{(3, -1, 2), (2, 0, -3)\} \)
- c. \( \mathbf{x} = (3, 1, 5, 9), U = \text{span}\{(1, 0, 1, 1), (0, 1, -1, 1), (-2, 0, 1, 1)\} \)
- d. \( \mathbf{x} = (2, 0, 1, 6), U = \text{span}\{(1, 1, 1, 1), (1, 1, -1, -1), (1, -1, 1, -1)\} \)
- e. \( \mathbf{x} = (a, b, c, d), U = \text{span}\{(1, 0, 0, 0), (0, 1, 0, 0), (0, 0, 1, 0)\} \)
- f. \( \mathbf{x} = (a, b, c, d), U = \text{span}\{(1, -1, 2, 0), (-1, 1, 1, 1)\} \)

**Exercise 8.1.3** Let \( \mathbf{x} = (1, -2, 1, 6) \) in \( \mathbb{R}^4 \), and let \( U = \text{span}\{(2, 1, 3, -4), (1, 2, 0, 1)\} \).

- a. Compute \( \text{proj}_U \mathbf{x} \).
- b. Show that \( \{(1, 0, 2, -3), (4, 7, 1, 2)\} \) is another orthogonal basis of \( U \).
- c. Use the basis in part (b) to compute \( \text{proj}_U \mathbf{x} \).

**Exercise 8.1.4** In each case, use the Gram-Schmidt algorithm to find an orthogonal basis of the subspace \( U \), and find the vector in \( U \) closest to \( \mathbf{x} \).

- a. \( U = \text{span}\{(1, 1, 1), (0, 1, 1)\}, \mathbf{x} = (-1, 2, 1) \)
- b. \( U = \text{span}\{(1, -1, 0), (-1, 0, 1)\}, \mathbf{x} = (2, 1, 0) \)
- c. \( U = \text{span}\{(1, 0, 1, 0), (1, 1, 1, 0), (1, 1, 0, 0)\}, \mathbf{x} = (2, 0, -1, 3) \)
- d. \( U = \text{span}\{(1, -1, 0, 1), (1, 1, 0, 0), (1, 1, 0, 1)\}, \mathbf{x} = (2, 0, 3, 1) \)

**Exercise 8.1.5** Let \( U = \text{span}\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k\} \), \( \mathbf{v}_i \) in \( \mathbb{R}^n \), and let \( A \) be the \( k \times n \) matrix with the \( \mathbf{v}_i \) as rows.

- a. Show that \( U^\perp = \{\mathbf{x} \in \mathbb{R}^n \mid A\mathbf{x}^T = \mathbf{0}\} \).
- b. Use part (a) to find \( U^\perp \) if \( U = \text{span}\{(1, -1, 2, 1), (1, 0, -1, 1)\} \).

**Exercise 8.1.6**

- a. Prove part 1 of Lemma 8.1.2.
- b. Prove part 2 of Lemma 8.1.2.
Exercise 8.1.7 Let \( U \) be a subspace of \( \mathbb{R}^n \). If \( x \) in \( \mathbb{R}^n \)
can be written in any way at all as \( x = p + q \) with \( p \) in \( U \)
and \( q \) in \( U^\perp \), show that necessarily \( p = \text{proj}_U x \).

Exercise 8.1.8 Let \( U \) be a subspace of \( \mathbb{R}^n \) and let \( x \) be
a vector in \( \mathbb{R}^n \). Using Exercise 8.1.7, or otherwise, show
that \( x \) is in \( U \) if and only if \( x = \text{proj}_U x \).

Exercise 8.1.10 If \( U \) is a subspace of \( \mathbb{R}^n \), show that
\( \text{proj}_U x = x \) for all \( x \) in \( U \).

Exercise 8.1.11 If \( U \) is a subspace of \( \mathbb{R}^n \), show that
\( x = \text{proj}_U x + \text{proj}_{U^\perp} x \) for all \( x \) in \( \mathbb{R}^n \).

Exercise 8.1.12 If \( \{f_1, \ldots, f_m\} \) is an orthogonal basis of
\( \mathbb{R}^n \) and \( U = \text{span} \{f_1, \ldots, f_m\} \), show that
\( U^\perp = \text{span} \{f_{m+1}, \ldots, f_n\} \).

Exercise 8.1.13 If \( U \) is a subspace of \( \mathbb{R}^n \), show that
\( U^\perp \perp = U \). [Hint: Show that \( U \subseteq U^\perp \perp \), then use The-
orem 8.1.4 (3) twice.]

Exercise 8.1.14 If \( U \) is a subspace of \( \mathbb{R}^n \), show how to
find an \( n \times n \) matrix \( A \) such that \( U = \{x \mid Ax = 0\} \). [Hint:
Exercise 8.1.13.]

Exercise 8.1.15 Write \( \mathbb{R}^n \) as rows. If \( A \) is an \( n \times n \) matrix,
write its null space as \( \text{null} A = \{x \in \mathbb{R}^n \mid Ax^T = 0\} \).
Show that:

a. \( \text{null} A = (\text{row} A)^\perp; \quad \text{b.} \quad \text{null} A^T = (\text{col} A)^\perp. \)

Exercise 8.1.16 If \( U \) and \( W \) are subspaces, show that
\( (U + W)^\perp = U^\perp \cap W^\perp \). [See Exercise 5.1.22.]

Exercise 8.1.17 Think of \( \mathbb{R}^n \) as consisting of rows.

a. Let \( E \) be an \( n \times n \) matrix, and let
\( U = \{xE \mid x \in \mathbb{R}^n\} \). Show that the following are equivalent.

i. \( E^2 = E = E^T \) (\( E \) is a projection matrix).

ii. \( (x - xE) \cdot (yE) = 0 \) for all \( x \) and \( y \) in \( \mathbb{R}^n \).

b. If \( E \) is a projection matrix, show that \( I - E \) is also a projection
matrix.

c. If \( EF = 0 = FE \) and \( E \) and \( F \) are projection matrices, show that \( E + F \) is also a projection matrix.

d. If \( A \) is \( m \times n \) and \( AA^T \) is invertible, show that
\( E = A^T (A A^T)^{-1} A \) is a projection matrix.

Exercise 8.1.18 Let \( A \) be an \( n \times n \) matrix of rank \( r \). Show
that there is an invertible \( n \times n \) matrix \( U \) such that \( UA \) is a
row-echelon matrix with the property that the first \( r \) rows
are orthogonal. [Hint: Let \( R \) be the row-echelon form
of \( A \), and use the Gram-Schmidt process on the nonzero
rows of \( R \) from the bottom up. Use Lemma 2.4.1.]

Exercise 8.1.19 Let \( A \) be an \( (n - 1) \times n \) matrix with rows
\( x_1, x_2, \ldots, x_{n-1} \) and let \( A_i \) denote the
\( (n - 1) \times (n - 1) \) matrix obtained from \( A \) by deleting column \( i \). Define the vector \( y \) in \( \mathbb{R}^n \) by
\[ y = \left[ \begin{array}{c} \det A_1 - \det A_2 \\ \det A_3 \\ \vdots \\ \det (-1)^{n+1} \det A_n \end{array} \right] \]
Show that:

a. \( x_i \cdot y = 0 \) for all \( i = 1, 2, \ldots, n - 1 \). [Hint: Write
\( B_i = \begin{bmatrix} x_i \\ A \end{bmatrix} \) and show that \( \det B_i = 0 \).]

b. \( y \neq 0 \) if and only if \( \{x_1, x_2, \ldots, x_{n-1}\} \) is linearly
independent. [Hint: If some \( \det A_i \neq 0 \), the rows
of \( A_i \) are linearly independent. Conversely, if the
\( x_i \) are independent, consider \( A = UR \) where \( R \) is in
reduced row-echelon form.]

c. If \( \{x_1, x_2, \ldots, x_{n-1}\} \) is linearly independent, use
Theorem 8.1.3(3) to show that all solutions to the
system of \( n - 1 \) homogeneous equations
\[ Ax^T = 0 \]
are given by \( ty, t \) a parameter.
8.2 Orthogonal Diagonalization

Recall (Theorem 5.5.3) that an $n \times n$ matrix $A$ is diagonalizable if and only if it has $n$ linearly independent eigenvectors. Moreover, the matrix $P$ with these eigenvectors as columns is a diagonalizing matrix for $A$, that is

$$P^{-1}AP$$

diagonal.

As we have seen, the really nice bases of $\mathbb{R}^n$ are the orthogonal ones, so a natural question is: which $n \times n$ matrices have an orthogonal basis of eigenvectors? These turn out to be precisely the symmetric matrices, and this is the main result of this section.

Before proceeding, recall that an orthogonal set of vectors is called orthonormal if $\|v\| = 1$ for each vector $v$ in the set, and that any orthogonal set $\{v_1, v_2, \ldots, v_k\}$ can be “normalized”, that is converted into an orthonormal set $\{\frac{1}{\|v_1\|}v_1, \frac{1}{\|v_2\|}v_2, \ldots, \frac{1}{\|v_k\|}v_k\}$. In particular, if a matrix $A$ has $n$ orthogonal eigenvectors, they can (by normalizing) be taken to be orthonormal. The corresponding diagonalizing matrix $P$ has orthonormal columns, and such matrices are very easy to invert.

Theorem 8.2.1

The following conditions are equivalent for an $n \times n$ matrix $P$.

1. $P$ is invertible and $P^{-1} = P^T$.
2. The rows of $P$ are orthonormal.
3. The columns of $P$ are orthonormal.

Proof. First recall that condition (1) is equivalent to $PP^T = I$ by Corollary 2.4.1 of Theorem 2.4.5. Let $x_1, x_2, \ldots, x_n$ denote the rows of $P$. Then $x_j^T$ is the $j$th column of $P^T$, so the $(i, j)$-entry of $PP^T$ is $x_i \cdot x_j$. Thus $PP^T = I$ means that $x_i \cdot x_j = 0$ if $i \neq j$ and $x_i \cdot x_j = 1$ if $i = j$. Hence condition (1) is equivalent to (2). The proof of the equivalence of (1) and (3) is similar. \qed

Definition 8.3 Orthogonal Matrices

An $n \times n$ matrix $P$ is called an orthogonormal matrix$^2$ if it satisfies one (and hence all) of the conditions in Theorem 8.2.1.

Example 8.2.1

The rotation matrix

$$\begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix}$$

is orthogonal for any angle $\theta$.

These orthogonal matrices have the virtue that they are easy to invert—simply take the transpose. But they have many other important properties as well. If $T: \mathbb{R}^n \rightarrow \mathbb{R}^n$ is a linear operator, we will prove

$^2$In view of (2) and (3) of Theorem 8.2.1, orthonormal matrix might be a better name. But orthogonal matrix is standard.