ORTHOGONALITY

If **u** and **v** are vectors in $\mathbb{R}^n$, then we regard **u** and **v** as $n \times 1$ matrices. The transpose $\mathbf{u}^T$ is a $1 \times n$ matrix, and the matrix product $\mathbf{u}^T\mathbf{v}$ is a 1×1 matrix, which we write as a single real number (a scalar) without brackets. The number $\mathbf{u}^T\mathbf{v}$ is called the **inner product** of **u** and **v**, and often it is written as $\mathbf{u}\cdot\mathbf{v}$. This inner product is also referred to as a **dot product**. If

$$\mathbf{u} = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix} \quad \text{and} \quad \mathbf{v} = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix}$$

then the inner product of **u** and **v** is

$$\begin{bmatrix} u_1 & u_2 & \cdots & u_n \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix} = u_1v_1 + u_2v_2 + \cdots + u_nv_n$$

EXAMPLE 1 Compute $\mathbf{u}\cdot\mathbf{v}$ and $\mathbf{v}\cdot\mathbf{u}$ for $\mathbf{u} = \begin{bmatrix} 2 \\ -5 \\ -1 \end{bmatrix}$ and $\mathbf{v} = \begin{bmatrix} 3 \\ 2 \\ -3 \end{bmatrix}$.

SOLUTION

$$\mathbf{u}\cdot\mathbf{v} = \mathbf{u}^T\mathbf{v} = \begin{bmatrix} 2 & -5 & -1 \end{bmatrix} \begin{bmatrix} 3 \\ 2 \\ -3 \end{bmatrix} = (2)(3) + (-5)(2) + (-1)(-3) = -1$$

$$\mathbf{v}\cdot\mathbf{u} = \mathbf{v}^T\mathbf{u} = \begin{bmatrix} 3 & 2 & -3 \end{bmatrix} \begin{bmatrix} 2 \\ -5 \\ -1 \end{bmatrix} = (3)(2) + (2)(-5) + (-3)(-1) = -1$$

It is clear from the calculations in Example 1 why $\mathbf{u}\cdot\mathbf{v} = \mathbf{v}\cdot\mathbf{u}$. This commutativity of the inner product holds in general. The following properties of the inner product are easily deduced from properties of the transpose operation.

THEOREM 1

Let $\mathbf{u}$, $\mathbf{v}$, and $\mathbf{w}$ be vectors in $\mathbb{R}^n$, and let c be a scalar. Then

a. $\mathbf{u} \cdot \mathbf{v} = \mathbf{v} \cdot \mathbf{u}$

b. $(\mathbf{u} + \mathbf{v}) \cdot \mathbf{w} = \mathbf{u} \cdot \mathbf{w} + \mathbf{v} \cdot \mathbf{w}$

c. $(c\mathbf{u}) \cdot \mathbf{v} = c(\mathbf{u} \cdot \mathbf{v}) = \mathbf{u} \cdot (c\mathbf{v})$

d. $\mathbf{u} \cdot \mathbf{u} \geq 0$, and $\mathbf{u} \cdot \mathbf{u} = 0$ if and only if $\mathbf{u} = \mathbf{0}$

Properties (b) and (c) can be combined several times to produce the following useful rule:

$$(c_1\mathbf{u}_1 + \cdots + c_p\mathbf{u}_p) \cdot \mathbf{w} = c_1(\mathbf{u}_1 \cdot \mathbf{w}) + \cdots + c_p(\mathbf{u}_p \cdot \mathbf{w})$$

DEFINITION

If $\mathbf{v}$ is in $\mathbb{R}^n$, with entries $v_1, \ldots, v_n$, then the square root of $\mathbf{v} \cdot \mathbf{v}$ is defined because $\mathbf{v} \cdot \mathbf{v}$ is nonnegative.

The **length** (or **norm**) of $\mathbf{v}$ is the nonnegative scalar $\|\mathbf{v}\|$ defined by

$$\|\mathbf{v}\| = \sqrt{\mathbf{v} \cdot \mathbf{v}} = \sqrt{v_1^2 + v_2^2 + \cdots + v_n^2}, \quad \text{and} \quad \|\mathbf{v}\|^2 = \mathbf{v} \cdot \mathbf{v}$$

For any scalar c, the length of $c\mathbf{v}$ is $|c|$ times the length of $\mathbf{v}$. That is,

$$\|c\mathbf{v}\| = |c|\|\mathbf{v}\|$$

(To see this, compute $\|c\mathbf{v}\|^2 = (c\mathbf{v}) \cdot (c\mathbf{v}) = c^2\mathbf{v} \cdot \mathbf{v} = c^2\|\mathbf{v}\|^2$ and take square roots.)

A vector whose length is 1 is called a **unit vector**. If we *divide* a nonzero vector $\mathbf{v}$ by its length—that is, multiply by $1/\|\mathbf{v}\|$—we obtain a unit vector $\mathbf{u}$ because the length of $\mathbf{u}$ is $(1/\|\mathbf{v}\|)\|\mathbf{v}\|$. The process of creating $\mathbf{u}$ from $\mathbf{v}$ is sometimes called **normalizing** $\mathbf{v}$, and we say that $\mathbf{u}$ is *in the same direction* as $\mathbf{v}$.

EXAMPLE 2 Let $\mathbf{v} = (1, -2, 2, 0)$. Find a unit vector $\mathbf{u}$ in the same direction as $\mathbf{v}$.

SOLUTION First, compute the length of $\mathbf{v}$:

$$\|\mathbf{v}\|^2 = \mathbf{v}\cdot\mathbf{v} = (1)^2 + (-2)^2 + (2)^2 + (0)^2 = 9$$
$$\|\mathbf{v}\| = \sqrt{9} = 3$$

Then, multiply $\mathbf{v}$ by $1/\|\mathbf{v}\|$ to obtain

$$\mathbf{u} = \frac{1}{\|\mathbf{v}\|}\mathbf{v} = \frac{1}{3}\mathbf{v} = \frac{1}{3}\begin{bmatrix} 1 \\ -2 \\ 2 \\ 0 \end{bmatrix} = \begin{bmatrix} 1/3 \\ -2/3 \\ 2/3 \\ 0 \end{bmatrix}$$

To check that $\|\mathbf{u}\| = 1$, it suffices to show that $\|\mathbf{u}\|^2 = 1$.

$$\|\mathbf{u}\|^2 = \mathbf{u}\cdot\mathbf{u} = \left(\tfrac{1}{3}\right)^2 + \left(-\tfrac{2}{3}\right)^2 + \left(\tfrac{2}{3}\right)^2 + (0)^2$$
$$= \tfrac{1}{9} + \tfrac{4}{9} + \tfrac{4}{9} + 0 = 1$$

EXAMPLE 3 Let W be the subspace of $\mathbb{R}^2$ spanned by $\mathbf{x} = (\frac{2}{3}, 1)$. Find a unit vector $\mathbf{z}$ that is a basis for W.

SOLUTION Any nonzero vector in W is a basis for W. To simplify the calculation, "scale" $\mathbf{x}$ to eliminate fractions. That is, multiply $\mathbf{x}$ by 3 to get

$$\mathbf{y} = \begin{bmatrix} 2 \\ 3 \end{bmatrix}$$

Now compute $\|\mathbf{y}\|^2 = 2^2 + 3^2 = 13$, $\|\mathbf{y}\| = \sqrt{13}$, and normalize $\mathbf{y}$ to get

$$\mathbf{z} = \frac{1}{\sqrt{13}}\begin{bmatrix} 2 \\ 3 \end{bmatrix} = \begin{bmatrix} 2/\sqrt{13} \\ 3/\sqrt{13} \end{bmatrix}$$

Another unit vector is $(-2/\sqrt{13}, -3/\sqrt{13})$.

DEFINITION

For $\mathbf{u}$ and $\mathbf{v}$ in $\mathbb{R}^n$, the **distance between u and v**, written as $\text{dist}(\mathbf{u}, \mathbf{v})$, is the length of the vector $\mathbf{u} - \mathbf{v}$. That is,

$$\text{dist}(\mathbf{u}, \mathbf{v}) = \|\mathbf{u} - \mathbf{v}\|$$

In $\mathbb{R}^2$ and $\mathbb{R}^3$, this definition of distance coincides with the usual formulas for the Euclidean distance between two points, as the next two examples show.

EXAMPLE 4 Compute the distance between the vectors $\mathbf{u} = (7, 1)$ and $\mathbf{v} = (3, 2)$.

SOLUTION Calculate

$$\mathbf{u} - \mathbf{v} = \begin{bmatrix} 7 \\ 1 \end{bmatrix} - \begin{bmatrix} 3 \\ 2 \end{bmatrix} = \begin{bmatrix} 4 \\ -1 \end{bmatrix}$$

$$\|\mathbf{u} - \mathbf{v}\| = \sqrt{4^2 + (-1)^2} = \sqrt{17}$$

EXAMPLE 5 If $\mathbf{u} = (u_1, u_2, u_3)$ and $\mathbf{v} = (v_1, v_2, v_3)$, then

$$\begin{aligned} \text{dist}(\mathbf{u}, \mathbf{v}) = \|\mathbf{u} - \mathbf{v}\| &= \sqrt{(\mathbf{u} - \mathbf{v}) \cdot (\mathbf{u} - \mathbf{v})} \\ &= \sqrt{(u_1 - v_1)^2 + (u_2 - v_2)^2 + (u_3 - v_3)^2} \end{aligned}$$

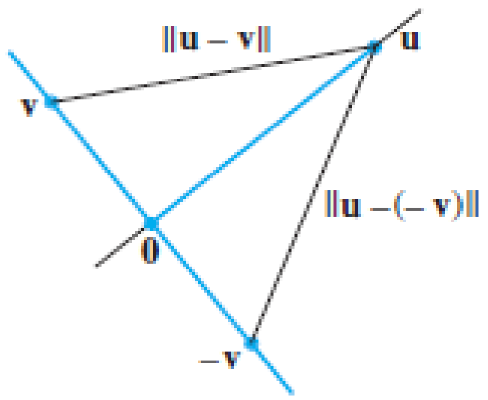
DEFINITION

Two vectors $\mathbf{u}$ and $\mathbf{v}$ in $\mathbb{R}^n$ are **orthogonal** (to each other) if $\mathbf{u} \cdot \mathbf{v} = 0$.

The rest of this chapter depends on the fact that the concept of perpendicular lines in ordinary Euclidean geometry has an analogue in $\mathbb{R}^n$.

Consider $\mathbb{R}^2$ or $\mathbb{R}^3$ and two lines through the origin determined by vectors **u** and **v**. The two lines shown in Figure are geometrically perpendicular if and only if the distance from **u** to **v** is the same as the distance from **u** to $-$**v**. This is the same as requiring the squares of the distances to be the same. Now

$$\begin{aligned}[\operatorname{dist}(\mathbf{u},-\mathbf{v})]^2 &= \|\mathbf{u}-(-\mathbf{v})\|^2 = \|\mathbf{u}+\mathbf{v}\|^2 \\ &= (\mathbf{u}+\mathbf{v})\cdot(\mathbf{u}+\mathbf{v}) \\ &= \mathbf{u}\cdot(\mathbf{u}+\mathbf{v})+\mathbf{v}\cdot(\mathbf{u}+\mathbf{v}) \\ &= \mathbf{u}\cdot\mathbf{u}+\mathbf{u}\cdot\mathbf{v}+\mathbf{v}\cdot\mathbf{u}+\mathbf{v}\cdot\mathbf{v} \\ &= \|\mathbf{u}\|^2+\|\mathbf{v}\|^2+2\mathbf{u}\cdot\mathbf{v} \qquad (1)\end{aligned}$$



The same calculations with **v** and $-$**v** interchanged show that

$$\begin{aligned}[\operatorname{dist}(\mathbf{u},\mathbf{v})]^2 &= \|\mathbf{u}\|^2+\|-\mathbf{v}\|^2+2\mathbf{u}\cdot(-\mathbf{v}) \\ &= \|\mathbf{u}\|^2+\|\mathbf{v}\|^2-2\mathbf{u}\cdot\mathbf{v}\end{aligned}$$

The two squared distances are equal if and only if $2\mathbf{u}\cdot\mathbf{v} = -2\mathbf{u}\cdot\mathbf{v}$, which happens if and only if $\mathbf{u}\cdot\mathbf{v} = 0$.

This calculation shows that when vectors **u** and **v** are identified with geometric points, the corresponding lines through the points and the origin are perpendicular if and only if $\mathbf{u}\cdot\mathbf{v} = 0$. The following definition generalizes to $\mathbb{R}^n$ this notion of perpendicularity (or *orthogonality*, as it is commonly called in linear algebra).

DEFINITION

Two vectors $\mathbf{u}$ and $\mathbf{v}$ in $\mathbb{R}^n$ are **orthogonal** (to each other) if $\mathbf{u} \cdot \mathbf{v} = 0$.

Observe that the zero vector is orthogonal to every vector in $\mathbb{R}^n$ because $\mathbf{0}^T \mathbf{v} = 0$ for all $\mathbf{v}$.

The next theorem provides a useful fact about orthogonal vectors. The proof follows immediately from the calculation in (1) above and the definition of orthogonality.

THEOREM 2

The Pythagorean Theorem

Two vectors $\mathbf{u}$ and $\mathbf{v}$ are orthogonal if and only if $\|\mathbf{u} + \mathbf{v}\|^2 = \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2$.

If a vector $\mathbf{z}$ is orthogonal to every vector in a subspace W of $\mathbb{R}^n$, then $\mathbf{z}$ is said to be **orthogonal to** W. The set of all vectors $\mathbf{z}$ that are orthogonal to W is called the **orthogonal complement** of W and is denoted by $W^\perp$ (and read as "W perpendicular" or simply "W perp").

1. A vector $\mathbf{x}$ is in $W^\perp$ if and only if $\mathbf{x}$ is orthogonal to every vector in a set that spans W.
2. $W^\perp$ is a subspace of $\mathbb{R}^n$.

THEOREM 3

Let A be an $m \times n$ matrix. The orthogonal complement of the row space of A is the null space of A, and the orthogonal complement of the column space of A is the null space of A^T:

$$(\text{Row } A)^{\perp} = \text{Nul } A \quad \text{and} \quad (\text{Col } A)^{\perp} = \text{Nul } A^T$$

PROOF The row–column rule for computing $A\mathbf{x}$ shows that if $\mathbf{x}$ is in Nul A, then $\mathbf{x}$ is orthogonal to each row of A (with the rows treated as vectors in $\mathbb{R}^n$). Since the rows of A span the row space, $\mathbf{x}$ is orthogonal to Row A. Conversely, if $\mathbf{x}$ is orthogonal to Row A, then $\mathbf{x}$ is certainly orthogonal to each row of A, and hence $A\mathbf{x} = \mathbf{0}$. This proves the first statement of the theorem. Since this statement is true for any matrix, it is true for A^T. That is, the orthogonal complement of the row space of A^T is the null space of A^T. This proves the second statement, because Row A^T = Col A. ■

If $\mathbf{u}$ and $\mathbf{v}$ are nonzero vectors in either $\mathbb{R}^2$ or $\mathbb{R}^3$, then there is a nice connection between their inner product and the angle ϑ between the two line segments from the origin to the points identified with $\mathbf{u}$ and $\mathbf{v}$. The formula is

$$\mathbf{u}\cdot\mathbf{v} = \|\mathbf{u}\|\,\|\mathbf{v}\|\cos\vartheta \tag{2}$$

ORTHOGONAL SETS

A set of vectors $\{\mathbf{u}_1, \dots, \mathbf{u}_p\}$ in $\mathbb{R}^n$ is said to be an **orthogonal set** if each pair of distinct vectors from the set is orthogonal, that is, if $\mathbf{u}_i \cdot \mathbf{u}_j = 0$ whenever $i \neq j$.

EXAMPLE 1 Show that $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$ is an orthogonal set, where

$$\mathbf{u}_1 = \begin{bmatrix} 3 \\ 1 \\ 1 \end{bmatrix}, \quad \mathbf{u}_2 = \begin{bmatrix} -1 \\ 2 \\ 1 \end{bmatrix}, \quad \mathbf{u}_3 = \begin{bmatrix} -1/2 \\ -2 \\ 7/2 \end{bmatrix}$$

SOLUTION Consider the three possible pairs of distinct vectors, namely, $\{\mathbf{u}_1, \mathbf{u}_2\}$, $\{\mathbf{u}_1, \mathbf{u}_3\}$, and $\{\mathbf{u}_2, \mathbf{u}_3\}$.

$$\begin{aligned}
\mathbf{u}_1 \cdot \mathbf{u}_2 &= 3(-1) + 1(2) + 1(1) = 0 \\
\mathbf{u}_1 \cdot \mathbf{u}_3 &= 3\left(-\tfrac{1}{2}\right) + 1(-2) + 1\left(\tfrac{7}{2}\right) = 0 \\
\mathbf{u}_2 \cdot \mathbf{u}_3 &= -1\left(-\tfrac{1}{2}\right) + 2(-2) + 1\left(\tfrac{7}{2}\right) = 0
\end{aligned}$$

Each pair of distinct vectors is orthogonal, and so $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$ is an orthogonal set.

THEOREM 4

If $S = \{\mathbf{u}_1, \dots, \mathbf{u}_p\}$ is an orthogonal set of nonzero vectors in $\mathbb{R}^n$, then S is linearly independent and hence is a basis for the subspace spanned by S.

PROOF If $\mathbf{0} = c_1\mathbf{u}_1 + \cdots + c_p\mathbf{u}_p$ for some scalars $c_1, \dots, c_p$, then

$$\begin{aligned}
0 = \mathbf{0} \cdot \mathbf{u}_1 &= (c_1\mathbf{u}_1 + c_2\mathbf{u}_2 + \cdots + c_p\mathbf{u}_p) \cdot \mathbf{u}_1 \\
&= (c_1\mathbf{u}_1) \cdot \mathbf{u}_1 + (c_2\mathbf{u}_2) \cdot \mathbf{u}_1 + \cdots + (c_p\mathbf{u}_p) \cdot \mathbf{u}_1 \\
&= c_1(\mathbf{u}_1 \cdot \mathbf{u}_1) + c_2(\mathbf{u}_2 \cdot \mathbf{u}_1) + \cdots + c_p(\mathbf{u}_p \cdot \mathbf{u}_1) \\
&= c_1(\mathbf{u}_1 \cdot \mathbf{u}_1)
\end{aligned}$$

because $\mathbf{u}_1$ is orthogonal to $\mathbf{u}_2, \dots, \mathbf{u}_p$. Since $\mathbf{u}_1$ is nonzero, $\mathbf{u}_1 \cdot \mathbf{u}_1$ is not zero and so $c_1 = 0$. Similarly, $c_2, \dots, c_p$ must be zero. Thus S is linearly independent. ■

DEFINITION

An **orthogonal basis** for a subspace W of $\mathbb{R}^n$ is a basis for W that is also an orthogonal set.

The next theorem suggests why an orthogonal basis is much nicer than other bases. The weights in a linear combination can be computed easily.

THEOREM 5

Let $\{\mathbf{u}_1, \dots, \mathbf{u}_p\}$ be an orthogonal basis for a subspace W of $\mathbb{R}^n$. For each $\mathbf{y}$ in W, the weights in the linear combination

$$\mathbf{y} = c_1\mathbf{u}_1 + \cdots + c_p\mathbf{u}_p$$

are given by

$$c_j = \frac{\mathbf{y}\cdot\mathbf{u}_j}{\mathbf{u}_j\cdot\mathbf{u}_j} \qquad (j = 1, \dots, p)$$

PROOF As in the preceding proof, the orthogonality of $\{\mathbf{u}_1, \dots, \mathbf{u}_p\}$ shows that

$$\mathbf{y}\cdot\mathbf{u}_1 = (c_1\mathbf{u}_1 + c_2\mathbf{u}_2 + \cdots + c_p\mathbf{u}_p)\cdot\mathbf{u}_1 = c_1(\mathbf{u}_1\cdot\mathbf{u}_1)$$

Since $\mathbf{u}_1\cdot\mathbf{u}_1$ is not zero, the equation above can be solved for c_1. To find c_j for $j = 2, \dots, p$, compute $\mathbf{y}\cdot\mathbf{u}_j$ and solve for c_j. ■

EXAMPLE 2 The set $S = \{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$ in Example 1 is an orthogonal basis for $\mathbb{R}^3$. Express the vector $\mathbf{y} = \begin{bmatrix} 6 \\ 1 \\ -8 \end{bmatrix}$ as a linear combination of the vectors in S.

$$\mathbf{u}_1 = \begin{bmatrix} 3 \\ 1 \\ 1 \end{bmatrix}, \quad \mathbf{u}_2 = \begin{bmatrix} -1 \\ 2 \\ 1 \end{bmatrix}, \quad \mathbf{u}_3 = \begin{bmatrix} -1/2 \\ -2 \\ 7/2 \end{bmatrix}$$

SOLUTION Compute

$$\mathbf{y}\cdot\mathbf{u}_1 = 11, \qquad \mathbf{y}\cdot\mathbf{u}_2 = -12, \qquad \mathbf{y}\cdot\mathbf{u}_3 = -33$$
$$\mathbf{u}_1\cdot\mathbf{u}_1 = 11, \qquad \mathbf{u}_2\cdot\mathbf{u}_2 = 6, \qquad \mathbf{u}_3\cdot\mathbf{u}_3 = 33/2$$

By Theorem 5,

$$\begin{aligned}\mathbf{y} &= \frac{\mathbf{y}\cdot\mathbf{u}_1}{\mathbf{u}_1\cdot\mathbf{u}_1}\mathbf{u}_1 + \frac{\mathbf{y}\cdot\mathbf{u}_2}{\mathbf{u}_2\cdot\mathbf{u}_2}\mathbf{u}_2 + \frac{\mathbf{y}\cdot\mathbf{u}_3}{\mathbf{u}_3\cdot\mathbf{u}_3}\mathbf{u}_3 \\ &= \frac{11}{11}\mathbf{u}_1 + \frac{-12}{6}\mathbf{u}_2 + \frac{-33}{33/2}\mathbf{u}_3 \\ &= \mathbf{u}_1 - 2\mathbf{u}_2 - 2\mathbf{u}_3\end{aligned}$$

An Orthogonal Projection

Given a nonzero vector $\mathbf{u}$ in $\mathbb{R}^n$, consider the problem of decomposing a vector $\mathbf{y}$ in $\mathbb{R}^n$ into the sum of two vectors, one a multiple of $\mathbf{u}$ and the other orthogonal to $\mathbf{u}$. We wish to write

$$\mathbf{y} = \hat{\mathbf{y}} + \mathbf{z} \tag{1}$$

where $\hat{\mathbf{y}} = \alpha\mathbf{u}$ for some scalar α and $\mathbf{z}$ is some vector orthogonal to $\mathbf{u}$.
Given any scalar α, let $\mathbf{z} = \mathbf{y} - \alpha\mathbf{u}$, so that (1) is satisfied. Then $\mathbf{y} - \hat{\mathbf{y}}$ is orthogonal to $\mathbf{u}$ if and only if

$$0 = (\mathbf{y} - \alpha\mathbf{u})\cdot\mathbf{u} = \mathbf{y}\cdot\mathbf{u} - (\alpha\mathbf{u})\cdot\mathbf{u} = \mathbf{y}\cdot\mathbf{u} - \alpha(\mathbf{u}\cdot\mathbf{u})$$

That is, (1) is satisfied with $\mathbf{z}$ orthogonal to $\mathbf{u}$ if and only if $\alpha = \dfrac{\mathbf{y}\cdot\mathbf{u}}{\mathbf{u}\cdot\mathbf{u}}$ and $\hat{\mathbf{y}} = \dfrac{\mathbf{y}\cdot\mathbf{u}}{\mathbf{u}\cdot\mathbf{u}}\mathbf{u}$.
The vector $\hat{\mathbf{y}}$ is called the **orthogonal projection of y onto u**, and the vector $\mathbf{z}$ is called the **component of y orthogonal to u**.

If c is any nonzero scalar and if $\mathbf{u}$ is replaced by $c\mathbf{u}$ in the definition of $\hat{\mathbf{y}}$, then the orthogonal projection of $\mathbf{y}$ onto $c\mathbf{u}$ is exactly the same as the orthogonal projection of $\mathbf{y}$ onto $\mathbf{u}$. Hence this projection is determined by the *subspace* L spanned by $\mathbf{u}$ (the line through $\mathbf{u}$ and $\mathbf{0}$). Sometimes $\hat{\mathbf{y}}$ is denoted by $\operatorname{proj}_L \mathbf{y}$ and is called the **orthogonal projection of y onto** L. That is,

$$\boxed{\hat{\mathbf{y}} = \operatorname{proj}_L \mathbf{y} = \frac{\mathbf{y}\cdot\mathbf{u}}{\mathbf{u}\cdot\mathbf{u}}\mathbf{u}} \qquad (2)$$

EXAMPLE 3 Let $\mathbf{y} = \begin{bmatrix} 7 \\ 6 \end{bmatrix}$ and $\mathbf{u} = \begin{bmatrix} 4 \\ 2 \end{bmatrix}$. Find the orthogonal projection of $\mathbf{y}$ onto $\mathbf{u}$. Then write $\mathbf{y}$ as the sum of two orthogonal vectors, one in Span $\{\mathbf{u}\}$ and one orthogonal to $\mathbf{u}$.

SOLUTION Compute

$$\mathbf{y}\cdot\mathbf{u} = \begin{bmatrix} 7 \\ 6 \end{bmatrix}\cdot\begin{bmatrix} 4 \\ 2 \end{bmatrix} = 40$$

$$\mathbf{u}\cdot\mathbf{u} = \begin{bmatrix} 4 \\ 2 \end{bmatrix}\cdot\begin{bmatrix} 4 \\ 2 \end{bmatrix} = 20$$

The orthogonal projection of $\mathbf{y}$ onto $\mathbf{u}$ is

$$\hat{\mathbf{y}} = \frac{\mathbf{y}\cdot\mathbf{u}}{\mathbf{u}\cdot\mathbf{u}}\mathbf{u} = \frac{40}{20}\mathbf{u} = 2\begin{bmatrix} 4 \\ 2 \end{bmatrix} = \begin{bmatrix} 8 \\ 4 \end{bmatrix}$$

and the component of $\mathbf{y}$ orthogonal to $\mathbf{u}$ is

$$\mathbf{y} - \hat{\mathbf{y}} = \begin{bmatrix} 7 \\ 6 \end{bmatrix} - \begin{bmatrix} 8 \\ 4 \end{bmatrix} = \begin{bmatrix} -1 \\ 2 \end{bmatrix}$$

The sum of these two vectors is **y**. That is,

$$\underset{\mathbf{y}}{\begin{bmatrix} 7 \\ 6 \end{bmatrix}} = \underset{\hat{\mathbf{y}}}{\begin{bmatrix} 8 \\ 4 \end{bmatrix}} + \underset{(\mathbf{y}-\hat{\mathbf{y}})}{\begin{bmatrix} -1 \\ 2 \end{bmatrix}}$$

Note: If the calculations above are correct, then $\{\hat{\mathbf{y}}, \mathbf{y}-\hat{\mathbf{y}}\}$ will be an orthogonal set. As a check, compute

$$\hat{\mathbf{y}}\cdot(\mathbf{y}-\hat{\mathbf{y}}) = \begin{bmatrix} 8 \\ 4 \end{bmatrix}\cdot\begin{bmatrix} -1 \\ 2 \end{bmatrix} = -8 + 8 = 0$$

A set $\{\mathbf{u}_1, \dots, \mathbf{u}_p\}$ is an **orthonormal set** if it is an orthogonal set of unit vectors. If W is the subspace spanned by such a set, then $\{\mathbf{u}_1, \dots, \mathbf{u}_p\}$ is an **orthonormal basis** for W, since the set is automatically linearly independent, by Theorem 4.

The simplest example of an orthonormal set is the standard basis $\{\mathbf{e}_1, \dots, \mathbf{e}_n\}$ for $\mathbb{R}^n$. Any nonempty subset of $\{\mathbf{e}_1, \dots, \mathbf{e}_n\}$ is orthonormal, too. Here is a more complicated example.

EXAMPLE 5 Show that $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ is an orthonormal basis of $\mathbb{R}^3$, where

$$\mathbf{v}_1 = \begin{bmatrix} 3/\sqrt{11} \\ 1/\sqrt{11} \\ 1/\sqrt{11} \end{bmatrix}, \quad \mathbf{v}_2 = \begin{bmatrix} -1/\sqrt{6} \\ 2/\sqrt{6} \\ 1/\sqrt{6} \end{bmatrix}, \quad \mathbf{v}_3 = \begin{bmatrix} -1/\sqrt{66} \\ -4/\sqrt{66} \\ 7/\sqrt{66} \end{bmatrix}$$

SOLUTION Compute

$$\mathbf{v}_1\cdot\mathbf{v}_2 = -3/\sqrt{66} + 2/\sqrt{66} + 1/\sqrt{66} = 0$$
$$\mathbf{v}_1\cdot\mathbf{v}_3 = -3/\sqrt{726} - 4/\sqrt{726} + 7/\sqrt{726} = 0$$
$$\mathbf{v}_2\cdot\mathbf{v}_3 = 1/\sqrt{396} - 8/\sqrt{396} + 7/\sqrt{396} = 0$$

Thus $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ is an orthogonal set. Also,

$$\mathbf{v}_1 \cdot \mathbf{v}_1 = 9/11 + 1/11 + 1/11 = 1$$
$$\mathbf{v}_2 \cdot \mathbf{v}_2 = 1/6 + 4/6 + 1/6 = 1$$
$$\mathbf{v}_3 \cdot \mathbf{v}_3 = 1/66 + 16/66 + 49/66 = 1$$

which shows that $\mathbf{v}_1$, $\mathbf{v}_2$, and $\mathbf{v}_3$ are unit vectors. Thus $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ is an orthonormal set. Since the set is linearly independent, its three vectors form a basis for $\mathbb{R}^3$.

THEOREM 6

An $m \times n$ matrix U has orthonormal columns if and only if $U^T U = I$.

PROOF To simplify notation, we suppose that U has only three columns, each a vector in $\mathbb{R}^m$. The proof of the general case is essentially the same. Let $U = [\,\mathbf{u}_1 \;\; \mathbf{u}_2 \;\; \mathbf{u}_3\,]$ and compute

$$U^T U = \begin{bmatrix} \mathbf{u}_1^T \\ \mathbf{u}_2^T \\ \mathbf{u}_3^T \end{bmatrix} [\,\mathbf{u}_1 \;\; \mathbf{u}_2 \;\; \mathbf{u}_3\,] = \begin{bmatrix} \mathbf{u}_1^T\mathbf{u}_1 & \mathbf{u}_1^T\mathbf{u}_2 & \mathbf{u}_1^T\mathbf{u}_3 \\ \mathbf{u}_2^T\mathbf{u}_1 & \mathbf{u}_2^T\mathbf{u}_2 & \mathbf{u}_2^T\mathbf{u}_3 \\ \mathbf{u}_3^T\mathbf{u}_1 & \mathbf{u}_3^T\mathbf{u}_2 & \mathbf{u}_3^T\mathbf{u}_3 \end{bmatrix} \tag{4}$$

The entries in the matrix at the right are inner products, using transpose notation. The columns of U are orthogonal if and only if

$$\mathbf{u}_1^T\mathbf{u}_2 = \mathbf{u}_2^T\mathbf{u}_1 = 0, \quad \mathbf{u}_1^T\mathbf{u}_3 = \mathbf{u}_3^T\mathbf{u}_1 = 0, \quad \mathbf{u}_2^T\mathbf{u}_3 = \mathbf{u}_3^T\mathbf{u}_2 = 0 \tag{5}$$

The columns of U all have unit length if and only if

$$\mathbf{u}_1^T\mathbf{u}_1 = 1, \quad \mathbf{u}_2^T\mathbf{u}_2 = 1, \quad \mathbf{u}_3^T\mathbf{u}_3 = 1 \tag{6}$$

The theorem follows immediately from (4)–(6). ■

THEOREM 7

Let U be an $m \times n$ matrix with orthonormal columns, and let $\mathbf{x}$ and $\mathbf{y}$ be in $\mathbb{R}^n$. Then

a. $\|U\mathbf{x}\| = \|\mathbf{x}\|$

b. $(U\mathbf{x})\cdot(U\mathbf{y}) = \mathbf{x}\cdot\mathbf{y}$

c. $(U\mathbf{x})\cdot(U\mathbf{y}) = 0$ if and only if $\mathbf{x}\cdot\mathbf{y} = 0$

Properties (a) and (c) say that the linear mapping $\mathbf{x} \mapsto U\mathbf{x}$ preserves lengths and orthogonality.

EXAMPLE 6 Let $U = \begin{bmatrix} 1/\sqrt{2} & 2/3 \\ 1/\sqrt{2} & -2/3 \\ 0 & 1/3 \end{bmatrix}$ and $\mathbf{x} = \begin{bmatrix} \sqrt{2} \\ 3 \end{bmatrix}$. Notice that U has orthonormal columns and

$$U^T U = \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} & 0 \\ 2/3 & -2/3 & 1/3 \end{bmatrix} \begin{bmatrix} 1/\sqrt{2} & 2/3 \\ 1/\sqrt{2} & -2/3 \\ 0 & 1/3 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

Verify that $\|U\mathbf{x}\| = \|\mathbf{x}\|$.

SOLUTION

$$U\mathbf{x} = \begin{bmatrix} 1/\sqrt{2} & 2/3 \\ 1/\sqrt{2} & -2/3 \\ 0 & 1/3 \end{bmatrix} \begin{bmatrix} \sqrt{2} \\ 3 \end{bmatrix} = \begin{bmatrix} 3 \\ -1 \\ 1 \end{bmatrix}$$

$$\|U\mathbf{x}\| = \sqrt{9+1+1} = \sqrt{11}$$

$$\|\mathbf{x}\| = \sqrt{2+9} = \sqrt{11}$$

Theorems 6 and 7 are particularly useful when applied to *square* matrices. An **orthogonal matrix** is a square invertible matrix U such that $U^{-1} = U^T$. By Theorem 6, such a matrix has orthonormal columns. It is easy to see that any *square* matrix with orthonormal columns is an orthogonal matrix.

EXAMPLE 7 The matrix

$$U = \begin{bmatrix} 3/\sqrt{11} & -1/\sqrt{6} & -1/\sqrt{66} \\ 1/\sqrt{11} & 2/\sqrt{6} & -4/\sqrt{66} \\ 1/\sqrt{11} & 1/\sqrt{6} & 7/\sqrt{66} \end{bmatrix}$$

is an orthogonal matrix because it is square and because its columns are orthonormal, by Example 5.

ORTHOGONAL PROJECTIONS

EXAMPLE 1 Let $\{\mathbf{u}_1, \dots, \mathbf{u}_5\}$ be an orthogonal basis for $\mathbb{R}^5$ and let

$$\mathbf{y} = c_1\mathbf{u}_1 + \cdots + c_5\mathbf{u}_5$$

Consider the subspace $W = \text{Span}\{\mathbf{u}_1, \mathbf{u}_2\}$, and write $\mathbf{y}$ as the sum of a vector $\mathbf{z}_1$ in W and a vector $\mathbf{z}_2$ in $W^\perp$.

SOLUTION Write

$$\mathbf{y} = \underbrace{c_1\mathbf{u}_1 + c_2\mathbf{u}_2}_{\mathbf{z}_1} + \underbrace{c_3\mathbf{u}_3 + c_4\mathbf{u}_4 + c_5\mathbf{u}_5}_{\mathbf{z}_2}$$

where $\quad \mathbf{z}_1 = c_1\mathbf{u}_1 + c_2\mathbf{u}_2 \quad$ is in $\text{Span}\{\mathbf{u}_1, \mathbf{u}_2\}$

and $\quad \mathbf{z}_2 = c_3\mathbf{u}_3 + c_4\mathbf{u}_4 + c_5\mathbf{u}_5 \quad$ is in $\text{Span}\{\mathbf{u}_3, \mathbf{u}_4, \mathbf{u}_5\}$.

To show that $\mathbf{z}_2$ is in $W^\perp$, it suffices to show that $\mathbf{z}_2$ is orthogonal to the vectors in the basis $\{\mathbf{u}_1, \mathbf{u}_2\}$ for W. Using properties of the inner product, compute

$$\begin{aligned} \mathbf{z}_2 \cdot \mathbf{u}_1 &= (c_3\mathbf{u}_3 + c_4\mathbf{u}_4 + c_5\mathbf{u}_5) \cdot \mathbf{u}_1 \\ &= c_3\mathbf{u}_3 \cdot \mathbf{u}_1 + c_4\mathbf{u}_4 \cdot \mathbf{u}_1 + c_5\mathbf{u}_5 \cdot \mathbf{u}_1 \\ &= 0 \end{aligned}$$

because $\mathbf{u}_1$ is orthogonal to $\mathbf{u}_3$, $\mathbf{u}_4$, and $\mathbf{u}_5$. A similar calculation shows that $\mathbf{z}_2 \cdot \mathbf{u}_2 = 0$. Thus $\mathbf{z}_2$ is in $W^\perp$.

THEOREM 8

The Orthogonal Decomposition Theorem

Let W be a subspace of $\mathbb{R}^n$. Then each $\mathbf{y}$ in $\mathbb{R}^n$ can be written uniquely in the form

$$\mathbf{y} = \hat{\mathbf{y}} + \mathbf{z} \tag{1}$$

where $\hat{\mathbf{y}}$ is in W and $\mathbf{z}$ is in $W^{\perp}$. In fact, if $\{\mathbf{u}_1, \ldots, \mathbf{u}_p\}$ is any orthogonal basis of W, then

$$\hat{\mathbf{y}} = \frac{\mathbf{y}\cdot\mathbf{u}_1}{\mathbf{u}_1\cdot\mathbf{u}_1}\mathbf{u}_1 + \cdots + \frac{\mathbf{y}\cdot\mathbf{u}_p}{\mathbf{u}_p\cdot\mathbf{u}_p}\mathbf{u}_p \tag{2}$$

and $\mathbf{z} = \mathbf{y} - \hat{\mathbf{y}}$.

PROOF Let $\{\mathbf{u}_1, \ldots, \mathbf{u}_p\}$ be any orthogonal basis for W, and define $\hat{\mathbf{y}}$ by (2). Then $\hat{\mathbf{y}}$ is in W because $\hat{\mathbf{y}}$ is a linear combination of the basis $\mathbf{u}_1, \ldots, \mathbf{u}_p$. Let $\mathbf{z} = \mathbf{y} - \hat{\mathbf{y}}$. Since $\mathbf{u}_1$ is orthogonal to $\mathbf{u}_2, \ldots, \mathbf{u}_p$, it follows from (2) that

$$\begin{aligned}\mathbf{z}\cdot\mathbf{u}_1 = (\mathbf{y} - \hat{\mathbf{y}})\cdot\mathbf{u}_1 &= \mathbf{y}\cdot\mathbf{u}_1 - \left(\frac{\mathbf{y}\cdot\mathbf{u}_1}{\mathbf{u}_1\cdot\mathbf{u}_1}\right)\mathbf{u}_1\cdot\mathbf{u}_1 - 0 - \cdots - 0 \\ &= \mathbf{y}\cdot\mathbf{u}_1 - \mathbf{y}\cdot\mathbf{u}_1 = 0\end{aligned}$$

Thus $\mathbf{z}$ is orthogonal to $\mathbf{u}_1$. Similarly, $\mathbf{z}$ is orthogonal to each $\mathbf{u}_j$ in the basis for W. Hence $\mathbf{z}$ is orthogonal to every vector in W. That is, $\mathbf{z}$ is in $W^{\perp}$.

To show that the decomposition in (1) is unique, suppose $\mathbf{y}$ can also be written as $\mathbf{y} = \hat{\mathbf{y}}_1 + \mathbf{z}_1$, with $\hat{\mathbf{y}}_1$ in W and $\mathbf{z}_1$ in $W^{\perp}$. Then $\hat{\mathbf{y}} + \mathbf{z} = \hat{\mathbf{y}}_1 + \mathbf{z}_1$ (since both sides equal $\mathbf{y}$), and so

$$\hat{\mathbf{y}} - \hat{\mathbf{y}}_1 = \mathbf{z}_1 - \mathbf{z}$$

This equality shows that the vector $\mathbf{v} = \hat{\mathbf{y}} - \hat{\mathbf{y}}_1$ is in W and in $W^{\perp}$ (because $\mathbf{z}_1$ and $\mathbf{z}$ are both in $W^{\perp}$, and $W^{\perp}$ is a subspace). Hence $\mathbf{v}\cdot\mathbf{v} = 0$, which shows that $\mathbf{v} = \mathbf{0}$. This proves that $\hat{\mathbf{y}} = \hat{\mathbf{y}}_1$ and also $\mathbf{z}_1 = \mathbf{z}$. ■

EXAMPLE 2 Let $\mathbf{u}_1 = \begin{bmatrix} 2 \\ 5 \\ -1 \end{bmatrix}$, $\mathbf{u}_2 = \begin{bmatrix} -2 \\ 1 \\ 1 \end{bmatrix}$, and $\mathbf{y} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$. Observe that $\{\mathbf{u}_1, \mathbf{u}_2\}$ is an orthogonal basis for $W = \text{Span}\{\mathbf{u}_1, \mathbf{u}_2\}$. Write $\mathbf{y}$ as the sum of a vector in W and a vector orthogonal to W.

SOLUTION The orthogonal projection of $\mathbf{y}$ onto W is

$$\begin{aligned}\hat{\mathbf{y}} &= \frac{\mathbf{y} \cdot \mathbf{u}_1}{\mathbf{u}_1 \cdot \mathbf{u}_1}\mathbf{u}_1 + \frac{\mathbf{y} \cdot \mathbf{u}_2}{\mathbf{u}_2 \cdot \mathbf{u}_2}\mathbf{u}_2 \\ &= \frac{9}{30}\begin{bmatrix} 2 \\ 5 \\ -1 \end{bmatrix} + \frac{3}{6}\begin{bmatrix} -2 \\ 1 \\ 1 \end{bmatrix} = \frac{9}{30}\begin{bmatrix} 2 \\ 5 \\ -1 \end{bmatrix} + \frac{15}{30}\begin{bmatrix} -2 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} -2/5 \\ 2 \\ 1/5 \end{bmatrix}\end{aligned}$$

Also

$$\mathbf{y} - \hat{\mathbf{y}} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} - \begin{bmatrix} -2/5 \\ 2 \\ 1/5 \end{bmatrix} = \begin{bmatrix} 7/5 \\ 0 \\ 14/5 \end{bmatrix}$$

Theorem 8 ensures that $\mathbf{y} - \hat{\mathbf{y}}$ is in $W^{\perp}$. To check the calculations, however, it is a good idea to verify that $\mathbf{y} - \hat{\mathbf{y}}$ is orthogonal to both $\mathbf{u}_1$ and $\mathbf{u}_2$ and hence to all of W. The desired decomposition of $\mathbf{y}$ is

$$\mathbf{y} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} = \begin{bmatrix} -2/5 \\ 2 \\ 1/5 \end{bmatrix} + \begin{bmatrix} 7/5 \\ 0 \\ 14/5 \end{bmatrix}$$

If $\mathbf{y}$ is in $W = \text{Span}\{\mathbf{u}_1, \ldots, \mathbf{u}_p\}$, then $\text{proj}_W \mathbf{y} = \mathbf{y}$.

This fact also follows from the next theorem.

THEOREM 9

The Best Approximation Theorem

Let W be a subspace of $\mathbb{R}^n$, let $\mathbf{y}$ be any vector in $\mathbb{R}^n$, and let $\hat{\mathbf{y}}$ be the orthogonal projection of $\mathbf{y}$ onto W. Then $\hat{\mathbf{y}}$ is the closest point in W to $\mathbf{y}$, in the sense that

$$\|\mathbf{y} - \hat{\mathbf{y}}\| < \|\mathbf{y} - \mathbf{v}\| \tag{3}$$

for all $\mathbf{v}$ in W distinct from $\hat{\mathbf{y}}$.

PROOF Take $\mathbf{v}$ in W distinct from $\hat{\mathbf{y}}$. Then $\hat{\mathbf{y}} - \mathbf{v}$ is in W. By the Orthogonal Decomposition Theorem, $\mathbf{y} - \hat{\mathbf{y}}$ is orthogonal to W. In particular, $\mathbf{y} - \hat{\mathbf{y}}$ is orthogonal to $\hat{\mathbf{y}} - \mathbf{v}$ (which is in W). Since

$$\mathbf{y} - \mathbf{v} = (\mathbf{y} - \hat{\mathbf{y}}) + (\hat{\mathbf{y}} - \mathbf{v})$$

the Pythagorean Theorem gives

$$\|\mathbf{y} - \mathbf{v}\|^2 = \|\mathbf{y} - \hat{\mathbf{y}}\|^2 + \|\hat{\mathbf{y}} - \mathbf{v}\|^2$$

Now $\|\hat{\mathbf{y}} - \mathbf{v}\|^2 > 0$ because $\hat{\mathbf{y}} - \mathbf{v} \neq \mathbf{0}$, and so inequality (3) follows immediately. ■

EXAMPLE 3 If $\mathbf{u}_1 = \begin{bmatrix} 2 \\ 5 \\ -1 \end{bmatrix}$, $\mathbf{u}_2 = \begin{bmatrix} -2 \\ 1 \\ 1 \end{bmatrix}$, $\mathbf{y} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$, and $W = \text{Span}\{\mathbf{u}_1, \mathbf{u}_2\}$, as in Example 2, then the closest point in W to $\mathbf{y}$ is

$$\hat{\mathbf{y}} = \frac{\mathbf{y}\cdot\mathbf{u}_1}{\mathbf{u}_1\cdot\mathbf{u}_1}\mathbf{u}_1 + \frac{\mathbf{y}\cdot\mathbf{u}_2}{\mathbf{u}_2\cdot\mathbf{u}_2}\mathbf{u}_2 = \begin{bmatrix} -2/5 \\ 2 \\ 1/5 \end{bmatrix}$$

EXAMPLE 4 The distance from a point $\mathbf{y}$ in $\mathbb{R}^n$ to a subspace W is defined as the distance from $\mathbf{y}$ to the nearest point in W. Find the distance from $\mathbf{y}$ to $W = \text{Span}\{\mathbf{u}_1, \mathbf{u}_2\}$, where

$$\mathbf{y} = \begin{bmatrix} -1 \\ -5 \\ 10 \end{bmatrix}, \quad \mathbf{u}_1 = \begin{bmatrix} 5 \\ -2 \\ 1 \end{bmatrix}, \quad \mathbf{u}_2 = \begin{bmatrix} 1 \\ 2 \\ -1 \end{bmatrix}$$

SOLUTION By the Best Approximation Theorem, the distance from $\mathbf{y}$ to W is $\|\mathbf{y} - \hat{\mathbf{y}}\|$, where $\hat{\mathbf{y}} = \text{proj}_W\,\mathbf{y}$. Since $\{\mathbf{u}_1, \mathbf{u}_2\}$ is an orthogonal basis for W,

$$\hat{\mathbf{y}} = \frac{15}{30}\mathbf{u}_1 + \frac{-21}{6}\mathbf{u}_2 = \frac{1}{2}\begin{bmatrix} 5 \\ -2 \\ 1 \end{bmatrix} - \frac{7}{2}\begin{bmatrix} 1 \\ 2 \\ -1 \end{bmatrix} = \begin{bmatrix} -1 \\ -8 \\ 4 \end{bmatrix}$$

$$\mathbf{y} - \hat{\mathbf{y}} = \begin{bmatrix} -1 \\ -5 \\ 10 \end{bmatrix} - \begin{bmatrix} -1 \\ -8 \\ 4 \end{bmatrix} = \begin{bmatrix} 0 \\ 3 \\ 6 \end{bmatrix}$$

$$\|\mathbf{y} - \hat{\mathbf{y}}\|^2 = 3^2 + 6^2 = 45$$

The distance from $\mathbf{y}$ to W is $\sqrt{45} = 3\sqrt{5}$. ■

THEOREM 10

If $\{\mathbf{u}_1, \dots, \mathbf{u}_p\}$ is an orthonormal basis for a subspace W of $\mathbb{R}^n$, then

$$\text{proj}_W\, \mathbf{y} = (\mathbf{y}\cdot\mathbf{u}_1)\mathbf{u}_1 + (\mathbf{y}\cdot\mathbf{u}_2)\mathbf{u}_2 + \cdots + (\mathbf{y}\cdot\mathbf{u}_p)\mathbf{u}_p \tag{4}$$

If $U = [\,\mathbf{u}_1 \quad \mathbf{u}_2 \quad \cdots \quad \mathbf{u}_p\,]$, then

$$\text{proj}_W\, \mathbf{y} = UU^T\mathbf{y} \quad \text{for all } \mathbf{y} \text{ in } \mathbb{R}^n \tag{5}$$

PROOF Formula (4) follows immediately from (2) in Theorem 8. Also, (4) shows that $\text{proj}_W\, \mathbf{y}$ is a linear combination of the columns of U using the weights $\mathbf{y}\cdot\mathbf{u}_1$, $\mathbf{y}\cdot\mathbf{u}_2, \dots, \mathbf{y}\cdot\mathbf{u}_p$. The weights can be written as $\mathbf{u}_1^T\mathbf{y}, \mathbf{u}_2^T\mathbf{y}, \dots, \mathbf{u}_p^T\mathbf{y}$, showing that they are the entries in $U^T\mathbf{y}$ and justifying (5). ■

EXAMPLE 1 Let $W = \text{Span}\,\{\mathbf{x}_1, \mathbf{x}_2\}$, where $\mathbf{x}_1 = \begin{bmatrix} 3 \\ 6 \\ 0 \end{bmatrix}$ and $\mathbf{x}_2 = \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix}$. Construct an orthogonal basis $\{\mathbf{v}_1, \mathbf{v}_2\}$ for W.

SOLUTION The component of $\mathbf{x}_2$ orthogonal to $\mathbf{x}_1$ is $\mathbf{x}_2 - \mathbf{p}$, which is in W because it is formed from $\mathbf{x}_2$ and a multiple of $\mathbf{x}_1$. Let $\mathbf{v}_1 = \mathbf{x}_1$ and

$$\mathbf{v}_2 = \mathbf{x}_2 - \mathbf{p} = \mathbf{x}_2 - \frac{\mathbf{x}_2\cdot\mathbf{x}_1}{\mathbf{x}_1\cdot\mathbf{x}_1}\mathbf{x}_1 = \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix} - \frac{15}{45}\begin{bmatrix} 3 \\ 6 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 2 \end{bmatrix}$$

Then $\{\mathbf{v}_1, \mathbf{v}_2\}$ is an orthogonal set of nonzero vectors in W. Since $\dim W = 2$, the set $\{\mathbf{v}_1, \mathbf{v}_2\}$ is a basis for W. ■

The next example fully illustrates the Gram–Schmidt process.

EXAMPLE 2 Let $\mathbf{x}_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$, $\mathbf{x}_2 = \begin{bmatrix} 0 \\ 1 \\ 1 \\ 1 \end{bmatrix}$, and $\mathbf{x}_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \end{bmatrix}$. Then $\{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3\}$ is clearly linearly independent and thus is a basis for a subspace W of $\mathbb{R}^4$. Construct an orthogonal basis for W.

SOLUTION

Step 1. Let $\mathbf{v}_1 = \mathbf{x}_1$ and $W_1 = \text{Span}\,\{\mathbf{x}_1\} = \text{Span}\,\{\mathbf{v}_1\}$.

Step 2. Let $\mathbf{v}_2$ be the vector produced by subtracting from $\mathbf{x}_2$ its projection onto the subspace W_1. That is, let

$$
\begin{aligned}
\mathbf{v}_2 &= \mathbf{x}_2 - \operatorname{proj}_{W_1} \mathbf{x}_2 \\
&= \mathbf{x}_2 - \frac{\mathbf{x}_2 \cdot \mathbf{v}_1}{\mathbf{v}_1 \cdot \mathbf{v}_1}\mathbf{v}_1 \qquad \text{Since } \mathbf{v}_1 = \mathbf{x}_1 \\
&= \begin{bmatrix} 0 \\ 1 \\ 1 \\ 1 \end{bmatrix} - \frac{3}{4}\begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} -3/4 \\ 1/4 \\ 1/4 \\ 1/4 \end{bmatrix}
\end{aligned}
$$

As in Example 1, $\mathbf{v}_2$ is the component of $\mathbf{x}_2$ orthogonal to $\mathbf{x}_1$, and $\{\mathbf{v}_1, \mathbf{v}_2\}$ is an orthogonal basis for the subspace W_2 spanned by $\mathbf{x}_1$ and $\mathbf{x}_2$.

Step 2′ (optional). If appropriate, scale $\mathbf{v}_2$ to simplify later computations. Since $\mathbf{v}_2$ has fractional entries, it is convenient to scale it by a factor of 4 and replace $\{\mathbf{v}_1, \mathbf{v}_2\}$ by the orthogonal basis

$$
\mathbf{v}_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \quad \mathbf{v}_2' = \begin{bmatrix} -3 \\ 1 \\ 1 \\ 1 \end{bmatrix}
$$

Step 3. Let $\mathbf{v}_3$ be the vector produced by subtracting from $\mathbf{x}_3$ its projection onto the subspace W_2. Use the orthogonal basis $\{\mathbf{v}_1, \mathbf{v}_2'\}$ to compute this projection onto W_2:

$$
\operatorname{proj}_{W_2} \mathbf{x}_3 = \underbrace{\frac{\mathbf{x}_3 \cdot \mathbf{v}_1}{\mathbf{v}_1 \cdot \mathbf{v}_1}\mathbf{v}_1}_{\text{Projection of } \mathbf{x}_3 \text{ onto } \mathbf{v}_1} + \underbrace{\frac{\mathbf{x}_3 \cdot \mathbf{v}_2'}{\mathbf{v}_2' \cdot \mathbf{v}_2'}\mathbf{v}_2'}_{\text{Projection of } \mathbf{x}_3 \text{ onto } \mathbf{v}_2'} = \frac{2}{4}\begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} + \frac{2}{12}\begin{bmatrix} -3 \\ 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 2/3 \\ 2/3 \\ 2/3 \end{bmatrix}
$$

Then $\mathbf{v}_3$ is the component of $\mathbf{x}_3$ orthogonal to W_2, namely,

$$
\mathbf{v}_3 = \mathbf{x}_3 - \operatorname{proj}_{W_2} \mathbf{x}_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \end{bmatrix} - \begin{bmatrix} 0 \\ 2/3 \\ 2/3 \\ 2/3 \end{bmatrix} = \begin{bmatrix} 0 \\ -2/3 \\ 1/3 \\ 1/3 \end{bmatrix}
$$

Observe that $\mathbf{v}_3$ is in W, because $\mathbf{x}_3$ and $\operatorname{proj}_{W_2}\mathbf{x}_3$ are both in W. Thus $\{\mathbf{v}_1, \mathbf{v}_2', \mathbf{v}_3\}$ is an orthogonal set of nonzero vectors and hence a linearly independent set in W. Note that W is three-dimensional since it was defined by a basis of three vectors. Hence, $\{\mathbf{v}_1, \mathbf{v}_2', \mathbf{v}_3\}$ is an orthogonal basis for W. ■

EXAMPLE 3 Example 1 constructed the orthogonal basis

$$\mathbf{v}_1 = \begin{bmatrix} 3 \\ 6 \\ 0 \end{bmatrix}, \quad \mathbf{v}_2 = \begin{bmatrix} 0 \\ 0 \\ 2 \end{bmatrix}$$

An orthonormal basis is

$$\mathbf{u}_1 = \frac{1}{\|\mathbf{v}_1\|}\mathbf{v}_1 = \frac{1}{\sqrt{45}}\begin{bmatrix} 3 \\ 6 \\ 0 \end{bmatrix} = \begin{bmatrix} 1/\sqrt{5} \\ 2/\sqrt{5} \\ 0 \end{bmatrix}$$

$$\mathbf{u}_2 = \frac{1}{\|\mathbf{v}_2\|}\mathbf{v}_2 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

■

THEOREM 11

The Gram–Schmidt Process

Given a basis $\{\mathbf{x}_1, \dots, \mathbf{x}_p\}$ for a nonzero subspace W of $\mathbb{R}^n$, define

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

Then $\{\mathbf{v}_1, \dots, \mathbf{v}_p\}$ is an orthogonal basis for W. In addition

$$\text{Span}\,\{\mathbf{v}_1, \dots, \mathbf{v}_k\} = \text{Span}\,\{\mathbf{x}_1, \dots, \mathbf{x}_k\} \qquad \text{for } 1 \le k \le p \tag{1}$$

PROOF For $1 \leq k \leq p$, let $W_k = \text{Span}\{\mathbf{x}_1, \ldots, \mathbf{x}_k\}$. 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 $\mathbf{v}_1, \ldots, \mathbf{v}_k$ so that $\{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ is an orthogonal basis for W_k. Define

$$\mathbf{v}_{k+1} = \mathbf{x}_{k+1} - \text{proj}_{W_k} \mathbf{x}_{k+1} \tag{2}$$

By the Orthogonal Decomposition Theorem, $\mathbf{v}_{k+1}$ is orthogonal to W_k. Note that $\text{proj}_{W_k} \mathbf{x}_{k+1}$ 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 $W_k = \text{Span}\{\mathbf{x}_1, \ldots, \mathbf{x}_k\}$. Hence $\{\mathbf{v}_1, \ldots, \mathbf{v}_{k+1}\}$ is an orthogonal set of nonzero vectors in the $(k+1)$-dimensional space W_{k+1}. This set is an orthogonal basis for W_{k+1}. Hence $W_{k+1} = \text{Span}\{\mathbf{v}_1, \ldots, \mathbf{v}_{k+1}\}$. When $k + 1 = p$, the process stops. ■

An orthonormal basis is constructed easily from an orthogonal basis $\{\mathbf{v}_1, \ldots, \mathbf{v}_p\}$: simply normalize (i.e., "scale") all the $\mathbf{v}_k$. When working problems by hand, this is easier than normalizing each $\mathbf{v}_k$ as soon as it is found (because it avoids unnecessary writing of square roots).

EXAMPLE Let $W = \text{Span}\{\mathbf{x}_1, \mathbf{x}_2\}$, where $\mathbf{x}_1 = \begin{bmatrix} 3 \\ 6 \\ 0 \end{bmatrix}$ and $\mathbf{x}_2 = \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix}$. Construct an orthogonal basis $\{\mathbf{v}_1, \mathbf{v}_2\}$ for W.

SOLUTION The subspace W is shown in Figure , along with $\mathbf{x}_1, \mathbf{x}_2$, and the projection $\mathbf{p}$ of $\mathbf{x}_2$ onto $\mathbf{x}_1$. The component of $\mathbf{x}_2$ orthogonal to $\mathbf{x}_1$ is $\mathbf{x}_2 - \mathbf{p}$, which is in W because it is formed from $\mathbf{x}_2$ and a multiple of $\mathbf{x}_1$. Let $\mathbf{v}_1 = \mathbf{x}_1$ and

$$\mathbf{v}_2 = \mathbf{x}_2 - \mathbf{p} = \mathbf{x}_2 - \frac{\mathbf{x}_2 \cdot \mathbf{x}_1}{\mathbf{x}_1 \cdot \mathbf{x}_1}\mathbf{x}_1 = \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix} - \frac{15}{45}\begin{bmatrix} 3 \\ 6 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 2 \end{bmatrix}$$

Then $\{\mathbf{v}_1, \mathbf{v}_2\}$ is an orthogonal set of nonzero vectors in W. Since $\dim W = 2$, the set $\{\mathbf{v}_1, \mathbf{v}_2\}$ is a basis for W.