5.3 Orthogonality

Length and orthogonality are basic concepts in geometry and, in $\mathbb{R}^2$ and $\mathbb{R}^3$, they both can be defined using the dot product. In this section we extend the dot product to vectors in $\mathbb{R}^n$, and so endow $\mathbb{R}^n$ with euclidean geometry. We then introduce the idea of an orthogonal basis—one of the most useful concepts in linear algebra, and begin exploring some of its applications.

Dot Product, Length, and Distance

If $\mathbf{x} = (x_1,\ x_2,\ \dots,\ x_n)$ and $\mathbf{y} = (y_1,\ y_2,\ \dots,\ y_n)$ are two n-tuples in $\mathbb{R}^n$, recall that their **dot product** was defined in Section 2.2 as follows:

$$\mathbf{x} \cdot \mathbf{y} = x_1y_1 + x_2y_2 + \cdots + x_ny_n$$

Observe that if $\mathbf{x}$ and $\mathbf{y}$ are written as columns then $\mathbf{x} \cdot \mathbf{y} = \mathbf{x}^T\mathbf{y}$ is a matrix product (and $\mathbf{x} \cdot \mathbf{y} = \mathbf{x}\mathbf{y}^T$ if they are written as rows). Here $\mathbf{x} \cdot \mathbf{y}$ is a 1×1 matrix, which we take to be a number.

Definition 5.6 Length in $\mathbb{R}^n$

As in $\mathbb{R}^3$, the ***length*** $\|\mathbf{x}\|$ *of the vector is defined by*

$$\|\mathbf{x}\| = \sqrt{\mathbf{x} \cdot \mathbf{x}} = \sqrt{x_1^2 + x_2^2 + \cdots + x_n^2}$$

Where $\sqrt{(\quad)}$ indicates the positive square root.

A vector $\mathbf{x}$ of length 1 is called a **unit vector**. If $\mathbf{x} \neq \mathbf{0}$, then $\|\mathbf{x}\| \neq 0$ and it follows easily that $\frac{1}{\|\mathbf{x}\|}\mathbf{x}$ is a unit vector (see Theorem 5.3.6 below), a fact that we shall use later.

Example 5.3.1

If $\mathbf{x} = (1,\ -1,\ -3,\ 1)$ and $\mathbf{y} = (2,\ 1,\ 1,\ 0)$ in $\mathbb{R}^4$, then $\mathbf{x} \cdot \mathbf{y} = 2 - 1 - 3 + 0 = -2$ and $\|\mathbf{x}\| = \sqrt{1+1+9+1} = \sqrt{12} = 2\sqrt{3}$. Hence $\frac{1}{2\sqrt{3}}\mathbf{x}$ is a unit vector; similarly $\frac{1}{\sqrt{6}}\mathbf{y}$ is a unit vector.

These definitions agree with those in $\mathbb{R}^2$ and $\mathbb{R}^3$, and many properties carry over to $\mathbb{R}^n$:

Theorem 5.3.1

Let $\mathbf{x}$, $\mathbf{y}$, and $\mathbf{z}$ denote vectors in $\mathbb{R}^n$. Then:

1. $\mathbf{x} \cdot \mathbf{y} = \mathbf{y} \cdot \mathbf{x}$.
2. $\mathbf{x} \cdot (\mathbf{y} + \mathbf{z}) = \mathbf{x} \cdot \mathbf{y} + \mathbf{x} \cdot \mathbf{z}$.
3. $(a\mathbf{x}) \cdot \mathbf{y} = a(\mathbf{x} \cdot \mathbf{y}) = \mathbf{x} \cdot (a\mathbf{y})$ *for all scalars a.*

4. $\|\mathbf{x}\|^2 = \mathbf{x} \cdot \mathbf{x}$.

5. $\|\mathbf{x}\| \geq 0$, *and* $\|\mathbf{x}\| = 0$ *if and only if* $\mathbf{x} = \mathbf{0}$.

6. $\|a\mathbf{x}\| = |a|\|\mathbf{x}\|$ *for all scalars* a.

Proof. (1), (2), and (3) follow from matrix arithmetic because $\mathbf{x} \cdot \mathbf{y} = \mathbf{x}^T\mathbf{y}$; (4) is clear from the definition; and (6) is a routine verification since $|a| = \sqrt{a^2}$. If $\mathbf{x} = (x_1, x_2, \dots, x_n)$, then $\|\mathbf{x}\| = \sqrt{x_1^2 + x_2^2 + \cdots + x_n^2}$ so $\|\mathbf{x}\| = 0$ if and only if $x_1^2 + x_2^2 + \cdots + x_n^2 = 0$. Since each x_i is a real number this happens if and only if $x_i = 0$ for each i; that is, if and only if $\mathbf{x} = \mathbf{0}$. This proves (5). □

Because of Theorem 5.3.1, computations with dot products in $\mathbb{R}^n$ are similar to those in $\mathbb{R}^3$. In particular, the dot product

$$(\mathbf{x}_1 + \mathbf{x}_2 + \cdots + \mathbf{x}_m) \cdot (\mathbf{y}_1 + \mathbf{y}_2 + \cdots + \mathbf{y}_k)$$

equals the sum of mk terms, $\mathbf{x}_i \cdot \mathbf{y}_j$, one for each choice of i and j. For example:

$$\begin{aligned}(3\mathbf{x} - 4\mathbf{y}) \cdot (7\mathbf{x} + 2\mathbf{y}) &= 21(\mathbf{x} \cdot \mathbf{x}) + 6(\mathbf{x} \cdot \mathbf{y}) - 28(\mathbf{y} \cdot \mathbf{x}) - 8(\mathbf{y} \cdot \mathbf{y}) \\ &= 21\|\mathbf{x}\|^2 - 22(\mathbf{x} \cdot \mathbf{y}) - 8\|\mathbf{y}\|^2\end{aligned}$$

holds for all vectors $\mathbf{x}$ and $\mathbf{y}$.

Example 5.3.2

Show that $\|\mathbf{x} + \mathbf{y}\|^2 = \|\mathbf{x}\|^2 + 2(\mathbf{x} \cdot \mathbf{y}) + \|\mathbf{y}\|^2$ for any $\mathbf{x}$ and $\mathbf{y}$ in $\mathbb{R}^n$.

Solution. Using Theorem 5.3.1 several times:

$$\begin{aligned}\|\mathbf{x} + \mathbf{y}\|^2 &= (\mathbf{x} + \mathbf{y}) \cdot (\mathbf{x} + \mathbf{y}) = \mathbf{x} \cdot \mathbf{x} + \mathbf{x} \cdot \mathbf{y} + \mathbf{y} \cdot \mathbf{x} + \mathbf{y} \cdot \mathbf{y} \\ &= \|\mathbf{x}\|^2 + 2(\mathbf{x} \cdot \mathbf{y}) + \|\mathbf{y}\|^2\end{aligned}$$

Example 5.3.3

Suppose that $\mathbb{R}^n = \text{span}\{\mathbf{f}_1, \mathbf{f}_2, \dots, \mathbf{f}_k\}$ for some vectors $\mathbf{f}_i$. If $\mathbf{x} \cdot \mathbf{f}_i = 0$ for each i where $\mathbf{x}$ is in $\mathbb{R}^n$, show that $\mathbf{x} = \mathbf{0}$.

Solution. We show $\mathbf{x} = \mathbf{0}$ by showing that $\|\mathbf{x}\| = 0$ and using (5) of Theorem 5.3.1. Since the $\mathbf{f}_i$ span $\mathbb{R}^n$, write $\mathbf{x} = t_1\mathbf{f}_1 + t_2\mathbf{f}_2 + \cdots + t_k\mathbf{f}_k$ where the t_i are in $\mathbb{R}$. Then

$$\begin{aligned}\|\mathbf{x}\|^2 &= \mathbf{x} \cdot \mathbf{x} = \mathbf{x} \cdot (t_1\mathbf{f}_1 + t_2\mathbf{f}_2 + \cdots + t_k\mathbf{f}_k) \\ &= t_1(\mathbf{x} \cdot \mathbf{f}_1) + t_2(\mathbf{x} \cdot \mathbf{f}_2) + \cdots + t_k(\mathbf{x} \cdot \mathbf{f}_k) \\ &= t_1(0) + t_2(0) + \cdots + t_k(0) \\ &= 0\end{aligned}$$

We saw in Section 4.2 that if $\mathbf{u}$ and $\mathbf{v}$ are nonzero vectors in $\mathbb{R}^3$, then $\frac{\mathbf{u}\cdot\mathbf{v}}{\|\mathbf{u}\|\|\mathbf{v}\|} = \cos\theta$ where θ is the angle between $\mathbf{u}$ and $\mathbf{v}$. Since $|\cos\theta| \leq 1$ for any angle θ, this shows that $|\mathbf{u}\cdot\mathbf{v}| \leq \|\mathbf{u}\|\|\mathbf{v}\|$. In this form the result holds in $\mathbb{R}^n$.

Theorem 5.3.2: Cauchy Inequality[9]

If $\mathbf{x}$ and $\mathbf{y}$ are vectors in $\mathbb{R}^n$, then

$$|\mathbf{x}\cdot\mathbf{y}| \leq \|\mathbf{x}\|\|\mathbf{y}\|$$

Moreover $|\mathbf{x}\cdot\mathbf{y}| = \|\mathbf{x}\|\|\mathbf{y}\|$ if and only if one of $\mathbf{x}$ and $\mathbf{y}$ is a multiple of the other.

Proof. The inequality holds if $\mathbf{x} = \mathbf{0}$ or $\mathbf{y} = \mathbf{0}$ (in fact it is equality). Otherwise, write $\|\mathbf{x}\| = a > 0$ and $\|\mathbf{y}\| = b > 0$ for convenience. A computation like that preceding Example 5.3.2 gives

$$\|b\mathbf{x} - a\mathbf{y}\|^2 = 2ab(ab - \mathbf{x}\cdot\mathbf{y}) \text{ and } \|b\mathbf{x} + a\mathbf{y}\|^2 = 2ab(ab + \mathbf{x}\cdot\mathbf{y}) \tag{5.1}$$

It follows that $ab - \mathbf{x}\cdot\mathbf{y} \geq 0$ and $ab + \mathbf{x}\cdot\mathbf{y} \geq 0$, and hence that $-ab \leq \mathbf{x}\cdot\mathbf{y} \leq ab$. Hence $|\mathbf{x}\cdot\mathbf{y}| \leq ab = \|\mathbf{x}\|\|\mathbf{y}\|$, proving the Cauchy inequality.

If equality holds, then $|\mathbf{x}\cdot\mathbf{y}| = ab$, so $\mathbf{x}\cdot\mathbf{y} = ab$ or $\mathbf{x}\cdot\mathbf{y} = -ab$. Hence Equation 5.1 shows that $b\mathbf{x} - a\mathbf{y} = 0$ or $b\mathbf{x} + a\mathbf{y} = 0$, so one of $\mathbf{x}$ and $\mathbf{y}$ is a multiple of the other (even if $a = 0$ or $b = 0$). □

The Cauchy inequality is equivalent to $(\mathbf{x}\cdot\mathbf{y})^2 \leq \|\mathbf{x}\|^2\|\mathbf{y}\|^2$. In $\mathbb{R}^5$ this becomes

$$(x_1y_1 + x_2y_2 + x_3y_3 + x_4y_4 + x_5y_5)^2 \leq (x_1^2 + x_2^2 + x_3^2 + x_4^2 + x_5^2)(y_1^2 + y_2^2 + y_3^2 + y_4^2 + y_5^2)$$

for all x_i and y_i in $\mathbb{R}$.

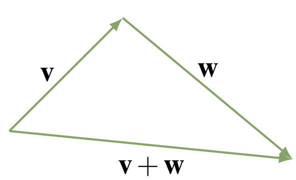
There is an important consequence of the Cauchy inequality. Given $\mathbf{x}$ and $\mathbf{y}$ in $\mathbb{R}^n$, use Example 5.3.2 and the fact that $\mathbf{x}\cdot\mathbf{y} \leq \|\mathbf{x}\|\|\mathbf{y}\|$ to compute

$$\|\mathbf{x}+\mathbf{y}\|^2 = \|\mathbf{x}\|^2 + 2(\mathbf{x}\cdot\mathbf{y}) + \|\mathbf{y}\|^2 \leq \|\mathbf{x}\|^2 + 2\|\mathbf{x}\|\|\mathbf{y}\| + \|\mathbf{y}\|^2 = (\|\mathbf{x}+\mathbf{y}\|)^2$$

Taking positive square roots gives:

Corollary 5.3.1: Triangle Inequality

If $\mathbf{x}$ and $\mathbf{y}$ are vectors in $\mathbb{R}^n$, then $\|\mathbf{x}+\mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|$.



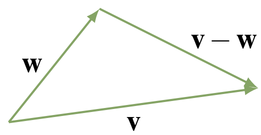
The reason for the name comes from the observation that in $\mathbb{R}^3$ the inequality asserts that the sum of the lengths of two sides of a triangle is not less than the length of the third side. This is illustrated in the diagram.

[9]Augustin Louis Cauchy (1789–1857) was born in Paris and became a professor at the École Polytechnique at the age of 26. He was one of the great mathematicians, producing more than 700 papers, and is best remembered for his work in analysis in which he established new standards of rigour and founded the theory of functions of a complex variable. He was a devout Catholic with a long-term interest in charitable work, and he was a royalist, following King Charles X into exile in Prague after he was deposed in 1830. Theorem 5.3.2 first appeared in his 1812 memoir on determinants.

Definition 5.7 Distance in $\mathbb{R}^n$

If $\mathbf{x}$ and $\mathbf{y}$ are two vectors in $\mathbb{R}^n$, we define the ***distance*** *$d(\mathbf{x},\ \mathbf{y})$ between $\mathbf{x}$ and $\mathbf{y}$ by*

$$d(\mathbf{x},\ \mathbf{y}) = \|\mathbf{x} - \mathbf{y}\|$$



The motivation again comes from $\mathbb{R}^3$ as is clear in the diagram. This distance function has all the intuitive properties of distance in $\mathbb{R}^3$, including another version of the triangle inequality.

Theorem 5.3.3

If $\mathbf{x}$, $\mathbf{y}$, and $\mathbf{z}$ are three vectors in $\mathbb{R}^n$ we have:

1. $d(\mathbf{x},\ \mathbf{y}) \geq 0$ *for all $\mathbf{x}$ and $\mathbf{y}$.*
2. $d(\mathbf{x},\ \mathbf{y}) = 0$ *if and only if $\mathbf{x} = \mathbf{y}$.*
3. $d(\mathbf{x},\ \mathbf{y}) = d(\mathbf{y},\ \mathbf{x})$ *for all $\mathbf{x}$ and $\mathbf{y}$.*
4. $d(\mathbf{x},\ \mathbf{z}) \leq d(\mathbf{x},\ \mathbf{y}) + d(\mathbf{y},\ \mathbf{z})$ *for all $\mathbf{x}$, $\mathbf{y}$, and $\mathbf{z}$.* *Triangle inequality.*

Proof. (1) and (2) restate part (5) of Theorem 5.3.1 because $d(\mathbf{x},\ \mathbf{y}) = \|\mathbf{x} - \mathbf{y}\|$, and (3) follows because $\|\mathbf{u}\| = \|-\mathbf{u}\|$ for every vector $\mathbf{u}$ in $\mathbb{R}^n$. To prove (4) use the Corollary to Theorem 5.3.2:

$$\begin{aligned} d(\mathbf{x},\ \mathbf{z}) = \|\mathbf{x} - \mathbf{z}\| &= \|(\mathbf{x} - \mathbf{y}) + (\mathbf{y} - \mathbf{z})\| \\ &\leq \|(\mathbf{x} - \mathbf{y})\| + \|(\mathbf{y} - \mathbf{z})\| = d(\mathbf{x},\ \mathbf{y}) + d(\mathbf{y},\ \mathbf{z}) \end{aligned}$$

□

Orthogonal Sets and the Expansion Theorem

Definition 5.8 Orthogonal and Orthonormal Sets

We say that two vectors $\mathbf{x}$ and $\mathbf{y}$ in $\mathbb{R}^n$ are ***orthogonal*** *if $\mathbf{x} \cdot \mathbf{y} = 0$, extending the terminology in $\mathbb{R}^3$ (See Theorem 4.2.3). More generally, a set $\{\mathbf{x}_1,\ \mathbf{x}_2,\ \dots,\ \mathbf{x}_k\}$ of vectors in $\mathbb{R}^n$ is called an* ***orthogonal set*** *if*

$$\mathbf{x}_i \cdot \mathbf{x}_j = 0 \text{ for all } i \neq j \quad \text{and} \quad \mathbf{x}_i \neq \mathbf{0} \text{ for all } i$$[10]

Note that $\{\mathbf{x}\}$ is an orthogonal set if $\mathbf{x} \neq \mathbf{0}$. A set $\{\mathbf{x}_1,\ \mathbf{x}_2,\ \dots,\ \mathbf{x}_k\}$ of vectors in $\mathbb{R}^n$ is called ***orthonormal*** *if it is orthogonal and, in addition, each $\mathbf{x}_i$ is a unit vector:*

$$\|\mathbf{x}_i\| = 1 \text{ for each } i.$$

[10]The reason for insisting that orthogonal sets consist of *nonzero* vectors is that we will be primarily concerned with orthogonal bases.

Example 5.3.4

The standard basis $\{\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n\}$ is an orthonormal set in $\mathbb{R}^n$.

The routine verification is left to the reader, as is the proof of:

Example 5.3.5

If $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k\}$ is orthogonal, so also is $\{a_1\mathbf{x}_1, a_2\mathbf{x}_2, \ldots, a_k\mathbf{x}_k\}$ for any nonzero scalars a_i.

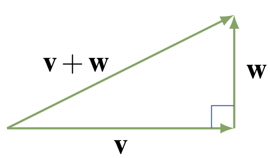
If $\mathbf{x} \neq \mathbf{0}$, it follows from item (6) of Theorem 5.3.1 that $\frac{1}{\|\mathbf{x}\|}\mathbf{x}$ is a unit vector, that is it has length 1.

Definition 5.9 Normalizing an Orthogonal Set

Hence if $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k\}$ *is an orthogonal set, then* $\{\frac{1}{\|\mathbf{x}_1\|}\mathbf{x}_1, \frac{1}{\|\mathbf{x}_2\|}\mathbf{x}_2, \cdots, \frac{1}{\|\mathbf{x}_k\|}\mathbf{x}_k\}$ *is an orthonormal set, and we say that it is the result of* ***normalizing*** *the orthogonal set* $\{\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_k\}$.

Example 5.3.6

If $\mathbf{f}_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \\ -1 \end{bmatrix}$, $\mathbf{f}_2 = \begin{bmatrix} 1 \\ 0 \\ 1 \\ 2 \end{bmatrix}$, $\mathbf{f}_3 = \begin{bmatrix} -1 \\ 0 \\ 1 \\ 0 \end{bmatrix}$, and $\mathbf{f}_4 = \begin{bmatrix} -1 \\ 3 \\ -1 \\ 1 \end{bmatrix}$ then $\{\mathbf{f}_1, \mathbf{f}_2, \mathbf{f}_3, \mathbf{f}_4\}$ is an orthogonal set in $\mathbb{R}^4$ as is easily verified. After normalizing, the corresponding orthonormal set is $\{\frac{1}{2}\mathbf{f}_1, \frac{1}{\sqrt{6}}\mathbf{f}_2, \frac{1}{\sqrt{2}}\mathbf{f}_3, \frac{1}{2\sqrt{3}}\mathbf{f}_4\}$



The most important result about orthogonality is Pythagoras' theorem. Given orthogonal vectors $\mathbf{v}$ and $\mathbf{w}$ in $\mathbb{R}^3$, it asserts that

$$\|\mathbf{v}+\mathbf{w}\|^2 = \|\mathbf{v}\|^2 + \|\mathbf{w}\|^2$$

as in the diagram. In this form the result holds for any orthogonal set in $\mathbb{R}^n$.

Theorem 5.3.4: Pythagoras' Theorem

If $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k\}$ *is an orthogonal set in* $\mathbb{R}^n$, *then*

$$\|\mathbf{x}_1 + \mathbf{x}_2 + \cdots + \mathbf{x}_k\|^2 = \|\mathbf{x}_1\|^2 + \|\mathbf{x}_2\|^2 + \cdots + \|\mathbf{x}_k\|^2.$$

Proof. The fact that $\mathbf{x}_i \cdot \mathbf{x}_j = 0$ whenever $i \neq j$ gives

$$\begin{aligned}\|\mathbf{x}_1+\mathbf{x}_2+\cdots+\mathbf{x}_k\|^2 &= (\mathbf{x}_1+\mathbf{x}_2+\cdots+\mathbf{x}_k)\cdot(\mathbf{x}_1+\mathbf{x}_2+\cdots+\mathbf{x}_k)\\ &= (\mathbf{x}_1\cdot\mathbf{x}_1+\mathbf{x}_2\cdot\mathbf{x}_2+\cdots+\mathbf{x}_k\cdot\mathbf{x}_k)+\sum_{i\neq j}\mathbf{x}_i\cdot\mathbf{x}_j\\ &= \|\mathbf{x}_1\|^2+\|\mathbf{x}_2\|^2+\cdots+\|\mathbf{x}_k\|^2+0\end{aligned}$$

This is what we wanted. □

If $\mathbf{v}$ and $\mathbf{w}$ are orthogonal, nonzero vectors in $\mathbb{R}^3$, then they are certainly not parallel, and so are linearly independent Example 5.2.7. The next theorem gives a far-reaching extension of this observation.

Theorem 5.3.5

Every orthogonal set in $\mathbb{R}^n$ is linearly independent.

Proof. Let $\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_k\}$ be an orthogonal set in $\mathbb{R}^n$ and suppose a linear combination vanishes, say: $t_1\mathbf{x}_1+t_2\mathbf{x}_2+\cdots+t_k\mathbf{x}_k=\mathbf{0}$. Then

$$\begin{aligned}0=\mathbf{x}_1\cdot\mathbf{0}=\mathbf{x}_1\cdot(t_1\mathbf{x}_1+t_2\mathbf{x}_2+\cdots+t_k\mathbf{x}_k)\\ &= t_1(\mathbf{x}_1\cdot\mathbf{x}_1)+t_2(\mathbf{x}_1\cdot\mathbf{x}_2)+\cdots+t_k(\mathbf{x}_1\cdot\mathbf{x}_k)\\ &= t_1\|\mathbf{x}_1\|^2+t_2(0)+\cdots+t_k(0)\\ &= t_1\|\mathbf{x}_1\|^2\end{aligned}$$

Since $\|\mathbf{x}_1\|^2\neq 0$, this implies that $t_1=0$. Similarly $t_i=0$ for each i. □

Theorem 5.3.5 suggests considering orthogonal bases for $\mathbb{R}^n$, that is orthogonal sets that span $\mathbb{R}^n$. These turn out to be the best bases in the sense that, when expanding a vector as a linear combination of the basis vectors, there are explicit formulas for the coefficients.

Theorem 5.3.6: Expansion Theorem

Let $\{\mathbf{f}_1, \mathbf{f}_2, \dots, \mathbf{f}_m\}$ be an orthogonal basis of a subspace U of $\mathbb{R}^n$. If $\mathbf{x}$ is any vector in U, we have

$$\mathbf{x}=\left(\frac{\mathbf{x}\cdot\mathbf{f}_1}{\|\mathbf{f}_1\|^2}\right)\mathbf{f}_1+\left(\frac{\mathbf{x}\cdot\mathbf{f}_2}{\|\mathbf{f}_2\|^2}\right)\mathbf{f}_1+\cdots+\left(\frac{\mathbf{x}\cdot\mathbf{f}_m}{\|\mathbf{f}_m\|^2}\right)\mathbf{f}_m$$

Proof. Since $\{\mathbf{f}_1, \mathbf{f}_2, \dots, \mathbf{f}_m\}$ spans U, we have $\mathbf{x}=t_1\mathbf{f}_1+t_2\mathbf{f}_2+\cdots+t_m\mathbf{f}_m$ where the t_i are scalars. To find t_1 we take the dot product of both sides with $\mathbf{f}_1$:

$$\begin{aligned}\mathbf{x}\cdot\mathbf{f}_1 &= (t_1\mathbf{f}_1+t_2\mathbf{f}_2+\cdots+t_m\mathbf{f}_m)\cdot\mathbf{f}_1\\ &= t_1(\mathbf{f}_1\cdot\mathbf{f}_1)+t_2(\mathbf{f}_2\cdot\mathbf{f}_1)+\cdots+t_m(\mathbf{f}_m\cdot\mathbf{f}_1)\\ &= t_1\|\mathbf{f}_1\|^2+t_2(0)+\cdots+t_m(0)\\ &= t_1\|\mathbf{f}_1\|^2\end{aligned}$$

Since $\mathbf{f}_1 \neq \mathbf{0}$, this gives $t_1 = \frac{\mathbf{x}\cdot\mathbf{f}_1}{\|\mathbf{f}_1\|^2}$. Similarly, $t_i = \frac{\mathbf{x}\cdot\mathbf{f}_i}{\|\mathbf{f}_i\|^2}$ for each i. □

The expansion in Theorem 5.3.6 of $\mathbf{x}$ as a linear combination of the orthogonal basis $\{\mathbf{f}_1,\ \mathbf{f}_2,\ \dots,\ \mathbf{f}_m\}$ is called the **Fourier expansion** of $\mathbf{x}$, and the coefficients $t_1 = \frac{\mathbf{x}\cdot\mathbf{f}_i}{\|\mathbf{f}_i\|^2}$ are called the **Fourier coefficients**. Note that if $\{\mathbf{f}_1,\ \mathbf{f}_2,\ \dots,\ \mathbf{f}_m\}$ is actually orthonormal, then $t_i = \mathbf{x}\cdot\mathbf{f}_i$ for each i. We will have a great deal more to say about this in Section 10.5.

Example 5.3.7

Expand $\mathbf{x} = (a,\ b,\ c,\ d)$ as a linear combination of the orthogonal basis $\{\mathbf{f}_1,\ \mathbf{f}_2,\ \mathbf{f}_3,\ \mathbf{f}_4\}$ of $\mathbb{R}^4$ given in Example 5.3.6.

Solution. We have $\mathbf{f}_1 = (1,\ 1,\ 1,\ -1)$, $\mathbf{f}_2 = (1,\ 0,\ 1,\ 2)$, $\mathbf{f}_3 = (-1,\ 0,\ 1,\ 0)$, and $\mathbf{f}_4 = (-1,\ 3,\ -1,\ 1)$ so the Fourier coefficients are

$$t_1 = \frac{\mathbf{x}\cdot\mathbf{f}_1}{\|\mathbf{f}_1\|^2} = \frac{1}{4}(a+b+c+d) \qquad t_3 = \frac{\mathbf{x}\cdot\mathbf{f}_3}{\|\mathbf{f}_3\|^2} = \frac{1}{2}(-a+c)$$

$$t_2 = \frac{\mathbf{x}\cdot\mathbf{f}_2}{\|\mathbf{f}_2\|^2} = \frac{1}{6}(a+c+2d) \qquad t_4 = \frac{\mathbf{x}\cdot\mathbf{f}_4}{\|\mathbf{f}_4\|^2} = \frac{1}{12}(-a+3b-c+d)$$

The reader can verify that indeed $\mathbf{x} = t_1\mathbf{f}_1 + t_2\mathbf{f}_2 + t_3\mathbf{f}_3 + t_4\mathbf{f}_4$.

A natural question arises here: Does every subspace U of $\mathbb{R}^n$ *have* an orthogonal basis? The answer is "yes"; in fact, there is a systematic procedure, called the Gram-Schmidt algorithm, for turning any basis of U into an orthogonal one. This leads to a definition of the projection onto a subspace U that generalizes the projection along a vector used in $\mathbb{R}^2$ and $\mathbb{R}^3$. All this is discussed in Section 8.1.

Exercises for 5.3

We often write vectors in $\mathbb{R}^n$ as row n-tuples.

Exercise 5.3.1 Obtain orthonormal bases of $\mathbb{R}^3$ by normalizing the following.

a. $\{(1,\ -1,\ 2),\ (0,\ 2,\ 1),\ (5,\ 1,\ -2)\}$

b. $\{(1,\ 1,\ 1),\ (4,\ 1,\ -5),\ (2,\ -3,\ 1)\}$

Exercise 5.3.2 In each case, show that the set of vectors is orthogonal in $\mathbb{R}^4$.

a. $\{(1,\ -1,\ 2,\ 5),\ (4,\ 1,\ 1,\ -1),\ (-7,\ 28,\ 5,\ 5)\}$

b. $\{(2,\ -1,\ 4,\ 5),\ (0,\ -1,\ 1,\ -1),\ (0,\ 3,\ 2,\ -1)\}$

Exercise 5.3.3 In each case, show that B is an orthogonal basis of $\mathbb{R}^3$ and use Theorem 5.3.6 to expand $\mathbf{x} = (a,\ b,\ c)$ as a linear combination of the basis vectors.

a. $B = \{(1,\ -1,\ 3),\ (-2,\ 1,\ 1),\ (4,\ 7,\ 1)\}$

b. $B = \{(1,\ 0,\ -1),\ (1,\ 4,\ 1),\ (2,\ -1,\ 2)\}$

c. $B = \{(1,\ 2,\ 3),\ (-1,\ -1,\ 1),\ (5,\ -4,\ 1)\}$

d. $B = \{(1,\ 1,\ 1),\ (1,\ -1,\ 0),\ (1,\ 1,\ -2)\}$

Exercise 5.3.4 In each case, write $\mathbf{x}$ as a linear combination of the orthogonal basis of the subspace U.

a. $\mathbf{x} = (13,\ -20,\ 15)$; $U = \text{span}\,\{(1,\ -2,\ 3),\ (-1,\ 1,\ 1)\}$

b. $\mathbf{x} = (14,\ 1,\ -8,\ 5)$;
$U = \text{span}\{(2,\ -1,\ 0,\ 3),\ (2,\ 1,\ -2,\ -1)\}$

Exercise 5.3.5 In each case, find all $(a,\ b,\ c,\ d)$ in $\mathbb{R}^4$ such that the given set is orthogonal.

a. $\{(1,\ 2,\ 1,\ 0),\ (1,\ -1,\ 1,\ 3),\ (2,\ -1,\ 0,\ -1),\ (a,\ b,\ c,\ d)\}$

b. $\{(1,\ 0,\ -1,\ 1),\ (2,\ 1,\ 1,\ -1),\ (1,\ -3,\ 1,\ 0),\ (a,\ b,\ c,\ d)\}$

Exercise 5.3.6 If $\|\mathbf{x}\| = 3$, $\|\mathbf{y}\| = 1$, and $\mathbf{x}\cdot\mathbf{y} = -2$, compute:

a. $\|3\mathbf{x} - 5\mathbf{y}\|$

b. $\|2\mathbf{x} + 7\mathbf{y}\|$

c. $(3\mathbf{x} - \mathbf{y})\cdot(2\mathbf{y} - \mathbf{x})$

d. $(\mathbf{x} - 2\mathbf{y})\cdot(3\mathbf{x} + 5\mathbf{y})$

Exercise 5.3.7 In each case either show that the statement is true or give an example showing that it is false.

a. Every independent set in $\mathbb{R}^n$ is orthogonal.

b. If $\{\mathbf{x},\ \mathbf{y}\}$ is an orthogonal set in $\mathbb{R}^n$, then $\{\mathbf{x},\ \mathbf{x}+\mathbf{y}\}$ is also orthogonal.

c. If $\{\mathbf{x},\ \mathbf{y}\}$ and $\{\mathbf{z},\ \mathbf{w}\}$ are both orthogonal in $\mathbb{R}^n$, then $\{\mathbf{x},\ \mathbf{y},\ \mathbf{z},\ \mathbf{w}\}$ is also orthogonal.

d. If $\{\mathbf{x}_1,\ \mathbf{x}_2\}$ and $\{\mathbf{y}_1,\ \mathbf{y}_2,\ \mathbf{y}_3\}$ are both orthogonal and $\mathbf{x}_i\cdot\mathbf{y}_j = 0$ for all i and j, then $\{\mathbf{x}_1,\ \mathbf{x}_2,\ \mathbf{y}_1,\ \mathbf{y}_2,\ \mathbf{y}_3\}$ is orthogonal.

e. If $\{\mathbf{x}_1,\ \mathbf{x}_2,\ \dots,\ \mathbf{x}_n\}$ is orthogonal in $\mathbb{R}^n$, then $\mathbb{R}^n = \text{span}\{\mathbf{x}_1,\ \mathbf{x}_2,\ \dots,\ \mathbf{x}_n\}$.

f. If $\mathbf{x} \neq \mathbf{0}$ in $\mathbb{R}^n$, then $\{\mathbf{x}\}$ is an orthogonal set.

Exercise 5.3.8 Let $\mathbf{v}$ denote a nonzero vector in $\mathbb{R}^n$.

a. Show that $P = \{\mathbf{x} \text{ in } \mathbb{R}^n \mid \mathbf{x}\cdot\mathbf{v} = 0\}$ is a subspace of $\mathbb{R}^n$.

b. Show that $\mathbb{R}\mathbf{v} = \{t\mathbf{v} \mid t \text{ in } \mathbb{R}\}$ is a subspace of $\mathbb{R}^n$.

c. Describe P and $\mathbb{R}\mathbf{v}$ geometrically when $n = 3$.

Exercise 5.3.9 If A is an $m \times n$ matrix with orthonormal columns, show that $A^TA = I_n$. [*Hint*: If $\mathbf{c}_1,\ \mathbf{c}_2,\ \dots,\ \mathbf{c}_n$ are the columns of A, show that column j of A^TA has entries $\mathbf{c}_1\cdot\mathbf{c}_j,\ \mathbf{c}_2\cdot\mathbf{c}_j,\ \dots,\ \mathbf{c}_n\cdot\mathbf{c}_j$].

Exercise 5.3.10 Use the Cauchy inequality to show that $\sqrt{xy} \leq \frac{1}{2}(x+y)$ for all $x \geq 0$ and $y \geq 0$. Here $\sqrt{xy}$ and $\frac{1}{2}(x+y)$ are called, respectively, the *geometric mean* and *arithmetic mean* of x and y.

[*Hint*: Use $\mathbf{x} = \begin{bmatrix} \sqrt{x} \\ \sqrt{y} \end{bmatrix}$ *and* $\mathbf{y} = \begin{bmatrix} \sqrt{y} \\ \sqrt{x} \end{bmatrix}$.]

Exercise 5.3.11 Use the Cauchy inequality to prove that:

a. $r_1 + r_2 + \cdots + r_n \leq n(r_1^2 + r_2^2 + \cdots + r_n^2)$ for all r_i in $\mathbb{R}$ and all $n \geq 1$.

b. $r_1r_2 + r_1r_3 + r_2r_3 \leq r_1^2 + r_2^2 + r_3^2$ for all $r_1,\ r_2,$ and r_3 in $\mathbb{R}$. [*Hint*: See part (a).]

Exercise 5.3.12

a. Show that $\mathbf{x}$ and $\mathbf{y}$ are orthogonal in $\mathbb{R}^n$ if and only if $\|\mathbf{x}+\mathbf{y}\| = \|\mathbf{x}-\mathbf{y}\|$.

b. Show that $\mathbf{x}+\mathbf{y}$ and $\mathbf{x}-\mathbf{y}$ are orthogonal in $\mathbb{R}^n$ if and only if $\|\mathbf{x}\| = \|\mathbf{y}\|$.

Exercise 5.3.13

a. Show that $\|\mathbf{x}+\mathbf{y}\|^2 = \|\mathbf{x}\|^2 + \|\mathbf{y}\|^2$ if and only if $\mathbf{x}$ is orthogonal to $\mathbf{y}$.

b. If $\mathbf{x} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$, $\mathbf{y} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ and $\mathbf{z} = \begin{bmatrix} -2 \\ 3 \end{bmatrix}$, show that $\|\mathbf{x}+\mathbf{y}+\mathbf{z}\|^2 = \|\mathbf{x}\|^2 + \|\mathbf{y}\|^2 + \|\mathbf{z}\|^2$ but $\mathbf{x}\cdot\mathbf{y} \neq 0$, $\mathbf{x}\cdot\mathbf{z} \neq 0$, and $\mathbf{y}\cdot\mathbf{z} \neq 0$.

Exercise 5.3.14

a. Show that $\mathbf{x}\cdot\mathbf{y} = \frac{1}{4}[\|\mathbf{x}+\mathbf{y}\|^2 - \|\mathbf{x}-\mathbf{y}\|^2]$ for all $\mathbf{x}$, $\mathbf{y}$ in $\mathbb{R}^n$.

b. Show that $\|\mathbf{x}\|^2 + \|\mathbf{y}\|^2 = \frac{1}{2}\left[\|\mathbf{x}+\mathbf{y}\|^2 + \|\mathbf{x}-\mathbf{y}\|^2\right]$ for all $\mathbf{x}$, $\mathbf{y}$ in $\mathbb{R}^n$.

Exercise 5.3.15 If A is $n \times n$, show that every eigenvalue of A^TA is nonnegative. [*Hint*: Compute $\|A\mathbf{x}\|^2$ where $\mathbf{x}$ is an eigenvector.]

Exercise 5.3.16 If $\mathbb{R}^n = \text{span}\{\mathbf{x}_1,\ \dots,\ \mathbf{x}_m\}$ and $\mathbf{x}\cdot\mathbf{x}_i = 0$ for all i, show that $\mathbf{x} = 0$. [*Hint*: Show $\|\mathbf{x}\| = 0$.]

Exercise 5.3.17 If $\mathbb{R}^n = \text{span}\{\mathbf{x}_1,\ \dots,\ \mathbf{x}_m\}$ and $\mathbf{x}\cdot\mathbf{x}_i = \mathbf{y}\cdot\mathbf{x}_i$ for all i, show that $\mathbf{x} = \mathbf{y}$. [*Hint*: Exercise 5.3.16]

Exercise 5.3.18 Let $\{\mathbf{e}_1,\ \dots,\ \mathbf{e}_n\}$ be an orthogonal basis of $\mathbb{R}^n$. Given $\mathbf{x}$ and $\mathbf{y}$ in $\mathbb{R}^n$, show that

$$\mathbf{x}\cdot\mathbf{y} = \frac{(\mathbf{x}\cdot\mathbf{e}_1)(\mathbf{y}\cdot\mathbf{e}_1)}{\|\mathbf{e}_1\|^2} + \cdots + \frac{(\mathbf{x}\cdot\mathbf{e}_n)(\mathbf{y}\cdot\mathbf{e}_n)}{\|\mathbf{e}_n\|^2}$$

5.4 Rank of a Matrix

In this section we use the concept of dimension to clarify the definition of the rank of a matrix given in Section 1.2, and to study its properties. This requires that we deal with rows and columns in the same way. While it has been our custom to write the n-tuples in $\mathbb{R}^n$ as columns, in this section we will frequently write them as rows. Subspaces, independence, spanning, and dimension are defined for rows using matrix operations, just as for columns. If A is an $m \times n$ matrix, we define:

Definition 5.10 Column and Row Space of a Matrix

The ***column space****,* col A*, of* A *is the subspace of* $\mathbb{R}^m$ *spanned by the columns of* A*.*
The ***row space****,* row A*, of* A *is the subspace of* $\mathbb{R}^n$ *spanned by the rows of* A*.*

Much of what we do in this section involves these subspaces. We begin with:

Lemma 5.4.1

Let A *and* B *denote* $m \times n$ *matrices.*

1. *If* $A \to B$ *by elementary row operations, then* $\text{row } A = \text{row } B$*.*
2. *If* $A \to B$ *by elementary column operations, then* $\text{col } A = \text{col } B$*.*

Proof. We prove (1); the proof of (2) is analogous. It is enough to do it in the case when $A \to B$ by a single row operation. Let $R_1, R_2, \dots, R_m$ denote the rows of A. The row operation $A \to B$ either interchanges two rows, multiplies a row by a nonzero constant, or adds a multiple of a row to a different row. We leave the first two cases to the reader. In the last case, suppose that a times row p is added to row q where $p < q$. Then the rows of B are $R_1, \dots, R_p, \dots, R_q + aR_p, \dots, R_m$, and Theorem 5.1.1 shows that

$$\text{span}\{R_1, \dots, R_p, \dots, R_q, \dots, R_m\} = \text{span}\{R_1, \dots, R_p, \dots, R_q + aR_p, \dots, R_m\}$$

That is, $\text{row } A = \text{row } B$. □

If A is any matrix, we can carry $A \to R$ by elementary row operations where R is a row-echelon matrix. Hence $\text{row } A = \text{row } R$ by Lemma 5.4.1; so the first part of the following result is of interest.

Lemma 5.4.2

If R *is a row-echelon matrix, then*

1. *The nonzero rows of* R *are a basis of* $\text{row } R$*.*
2. *The columns of* R *containing leading ones are a basis of* $\text{col } R$*.*

Proof. The rows of R are independent by Example 5.2.6, and they span $\text{row } R$ by definition. This proves (1).