5. Vector Space $\mathbb{R}^n$

5.1 Subspaces and Spanning

In Section 2.2 we introduced the set $\mathbb{R}^n$ of all n-tuples (called *vectors*), and began our investigation of the matrix transformations $\mathbb{R}^n \to \mathbb{R}^m$ given by matrix multiplication by an $m \times n$ matrix. Particular attention was paid to the euclidean plane $\mathbb{R}^2$ where certain simple geometric transformations were seen to be matrix transformations. Then in Section 2.6 we introduced linear transformations, showed that they are all matrix transformations, and found the matrices of rotations and reflections in $\mathbb{R}^2$. We returned to this in Section 4.4 where we showed that projections, reflections, and rotations of $\mathbb{R}^2$ and $\mathbb{R}^3$ were all linear, and where we related areas and volumes to determinants.

In this chapter we investigate $\mathbb{R}^n$ in full generality, and introduce some of the most important concepts and methods in linear algebra. The n-tuples in $\mathbb{R}^n$ will continue to be denoted $\mathbf{x}$, $\mathbf{y}$, and so on, and will be written as rows or columns depending on the context.

Subspaces of $\mathbb{R}^n$

Definition 5.1 Subspace of $\mathbb{R}^n$

*A set[1] U of vectors in $\mathbb{R}^n$ is called a **subspace** of $\mathbb{R}^n$ if it satisfies the following properties:*

S1. The zero vector $\mathbf{0} \in U$.

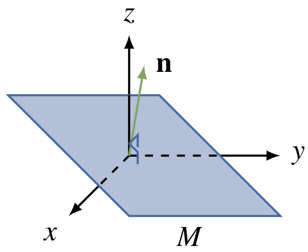
S2. If $\mathbf{x} \in U$ and $\mathbf{y} \in U$, then $\mathbf{x}+\mathbf{y} \in U$.

S3. If $\mathbf{x} \in U$, then $a\mathbf{x} \in U$ for every real number a.

We say that the subset U is **closed under addition** if S2 holds, and that U is **closed under scalar multiplication** if S3 holds.

Clearly $\mathbb{R}^n$ is a subspace of itself, and this chapter is about these subspaces and their properties. The set $U = \{\mathbf{0}\}$, consisting of only the zero vector, is also a subspace because $\mathbf{0}+\mathbf{0}=\mathbf{0}$ and $a\mathbf{0}=\mathbf{0}$ for each a in $\mathbb{R}$; it is called the **zero subspace**. Any subspace of $\mathbb{R}^n$ other than $\{\mathbf{0}\}$ or $\mathbb{R}^n$ is called a **proper** subspace.

[1] We use the language of sets. Informally, a **set** X is a collection of objects, called the **elements** of the set. The fact that x is an element of X is denoted $x \in X$. Two sets X and Y are called equal (written $X = Y$) if they have the same elements. If every element of X is in the set Y, we say that X is a **subset** of Y, and write $X \subseteq Y$. Hence $X \subseteq Y$ and $Y \subseteq X$ both hold if and only if $X = Y$.



We saw in Section 4.2 that every plane M through the origin in $\mathbb{R}^3$ has equation $ax+by+cz=0$ where a, b, and c are not all zero. Here $\mathbf{n} = \begin{bmatrix} a \\ b \\ c \end{bmatrix}$ is a normal for the plane and

$$M = \{\mathbf{v} \text{ in } \mathbb{R}^3 \mid \mathbf{n}\cdot\mathbf{v} = 0\}$$

where $\mathbf{v} = \begin{bmatrix} x \\ y \\ z \end{bmatrix}$ and $\mathbf{n}\cdot\mathbf{v}$ denotes the dot product introduced in Section 2.2 (see the diagram).[2] Then M is a subspace of $\mathbb{R}^3$. Indeed we show that M satisfies S1, S2, and S3 as follows:

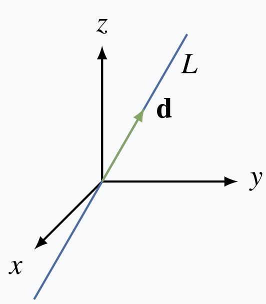
S1. $\mathbf{0} \in M$ *because* $\mathbf{n}\cdot\mathbf{0} = 0$*;*

S2. *If* $\mathbf{v} \in M$ *and* $\mathbf{v}_1 \in M$ *, then* $\mathbf{n}\cdot(\mathbf{v}+\mathbf{v}_1) = \mathbf{n}\cdot\mathbf{v} + \mathbf{n}\cdot\mathbf{v}_1 = 0+0 = 0$ *, so* $\mathbf{v}+\mathbf{v}_1 \in M$*;*

S3. *If* $\mathbf{v} \in M$ *, then* $\mathbf{n}\cdot(a\mathbf{v}) = a(\mathbf{n}\cdot\mathbf{v}) = a(0) = 0$ *, so* $a\mathbf{v} \in M$*.*

This proves the first part of

Example 5.1.1



Planes and lines through the origin in $\mathbb{R}^3$ are all subspaces of $\mathbb{R}^3$.

Solution. We dealt with planes above. If L is a line through the origin with direction vector $\mathbf{d}$, then $L = \{t\mathbf{d} \mid t \in \mathbb{R}\}$ (see the diagram). We leave it as an exercise to verify that L satisfies S1, S2, and S3.

Example 5.1.1 shows that lines through the origin in $\mathbb{R}^2$ are subspaces; in fact, they are the *only* proper subspaces of $\mathbb{R}^2$ (Exercise 5.1.24). Indeed, we shall see in Example 5.2.14 that lines and planes through the origin in $\mathbb{R}^3$ are the only proper subspaces of $\mathbb{R}^3$. Thus the geometry of lines and planes through the origin is captured by the subspace concept. (Note that *every* line or plane is just a translation of one of these.)

Subspaces can also be used to describe important features of an $m \times n$ matrix A. The **null space** of A, denoted $\text{null } A$, and the **image space** of A, denoted $\text{im } A$, are defined by

$$\text{null } A = \{\mathbf{x} \in \mathbb{R}^n \mid A\mathbf{x} = \mathbf{0}\} \quad \text{and} \quad \text{im } A = \{A\mathbf{x} \mid \mathbf{x} \in \mathbb{R}^n\}$$

In the language of Chapter 2, $\text{null } A$ consists of all solutions $\mathbf{x}$ in $\mathbb{R}^n$ of the homogeneous system $A\mathbf{x} = \mathbf{0}$, and $\text{im } A$ is the set of all vectors $\mathbf{y}$ in $\mathbb{R}^m$ such that $A\mathbf{x} = \mathbf{y}$ *has* a solution $\mathbf{x}$. Note that $\mathbf{x}$ is in $\text{null } A$ if it

[2] We are using set notation here. In general $\{q \mid p\}$ means the set of all objects q with property p.

satisfies the *condition* $A\mathbf{x} = \mathbf{0}$, while $\text{im } A$ consists of vectors of the *form* $A\mathbf{x}$ for some $\mathbf{x}$ in $\mathbb{R}^n$. These two ways to describe subsets occur frequently.

Example 5.1.2

If A is an $m \times n$ matrix, then:

1. $\text{null } A$ is a subspace of $\mathbb{R}^n$.
2. $\text{im } A$ is a subspace of $\mathbb{R}^m$.

Solution.

1. The zero vector $\mathbf{0} \in \mathbb{R}^n$ lies in $\text{null } A$ because $A\mathbf{0} = \mathbf{0}$.[3] If $\mathbf{x}$ and $\mathbf{x}_1$ are in $\text{null } A$, then $\mathbf{x} + \mathbf{x}_1$ and $a\mathbf{x}$ are in $\text{null } A$ because they satisfy the required condition:

$$A(\mathbf{x}+\mathbf{x}_1) = A\mathbf{x} + A\mathbf{x}_1 = \mathbf{0} + \mathbf{0} = \mathbf{0} \quad \text{and} \quad A(a\mathbf{x}) = a(A\mathbf{x}) = a\mathbf{0} = \mathbf{0}$$

   Hence $\text{null } A$ satisfies S1, S2, and S3, and so is a subspace of $\mathbb{R}^n$.

2. The zero vector $\mathbf{0} \in \mathbb{R}^m$ lies in $\text{im } A$ because $\mathbf{0} = A\mathbf{0}$. Suppose that $\mathbf{y}$ and $\mathbf{y}_1$ are in $\text{im } A$, say $\mathbf{y} = A\mathbf{x}$ and $\mathbf{y}_1 = A\mathbf{x}_1$ where $\mathbf{x}$ and $\mathbf{x}_1$ are in $\mathbb{R}^n$. Then

$$\mathbf{y} + \mathbf{y}_1 = A\mathbf{x} + A\mathbf{x}_1 = A(\mathbf{x}+\mathbf{x}_1) \quad \text{and} \quad a\mathbf{y} = a(A\mathbf{x}) = A(a\mathbf{x})$$

   show that $\mathbf{y} + \mathbf{y}_1$ and $a\mathbf{y}$ are both in $\text{im } A$ (they have the required form). Hence $\text{im } A$ is a subspace of $\mathbb{R}^m$.

There are other important subspaces associated with a matrix A that clarify basic properties of A. If A is an $n \times n$ matrix and λ is any number, let

$$E_\lambda(A) = \{\mathbf{x} \in \mathbb{R}^n \mid A\mathbf{x} = \lambda\mathbf{x}\}$$

A vector $\mathbf{x}$ is in $E_\lambda(A)$ if and only if $(\lambda I - A)\mathbf{x} = \mathbf{0}$, so Example 5.1.2 gives:

Example 5.1.3

$E_\lambda(A) = \text{null } (\lambda I - A)$ is a subspace of $\mathbb{R}^n$ for each $n \times n$ matrix A and number λ.

$E_\lambda(A)$ is called the **eigenspace** of A corresponding to λ. The reason for the name is that, in the terminology of Section 3.3, λ is an **eigenvalue** of A if $E_\lambda(A) \neq \{\mathbf{0}\}$. In this case the nonzero vectors in $E_\lambda(A)$ are called the **eigenvectors** of A corresponding to λ.

The reader should not get the impression that *every* subset of $\mathbb{R}^n$ is a subspace. For example:

$$U_1 = \left\{ \left[\begin{array}{c} x \\ y \end{array} \right] \middle| x \geq 0 \right\} \text{ satisfies S1 and S2, but not S3;}$$

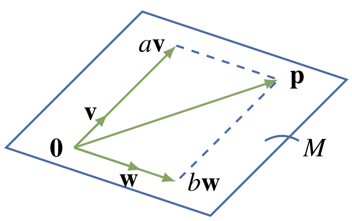
[3] We are using $\mathbf{0}$ to represent the zero vector in both $\mathbb{R}^m$ and $\mathbb{R}^n$. This abuse of notation is common and causes no confusion once everybody knows what is going on.

$$U_2 = \left\{ \begin{bmatrix} x \\ y \end{bmatrix} \middle| x^2 = y^2 \right\} \text{ satisfies S1 and S3, but not S2;}$$

Hence neither U_1 nor U_2 is a subspace of $\mathbb{R}^2$. (However, see Exercise 5.1.20.)

Spanning Sets

Let $\mathbf{v}$ and $\mathbf{w}$ be two nonzero, nonparallel vectors in $\mathbb{R}^3$ with their tails at the origin. The plane M through the origin containing these vectors is described in Section 4.2 by saying that $\mathbf{n} = \mathbf{v} \times \mathbf{w}$ is a *normal* for M, and that M consists of all vectors $\mathbf{p}$ such that $\mathbf{n} \cdot \mathbf{p} = 0$.[4] While this is a very useful way to look at planes, there is another approach that is at least as useful in $\mathbb{R}^3$ and, more importantly, works for all subspaces of $\mathbb{R}^n$ for any $n \geq 1$.



The idea is as follows: Observe that, by the diagram, a vector $\mathbf{p}$ is in M if and only if it has the form

$$\mathbf{p} = a\mathbf{v} + b\mathbf{w}$$

for certain real numbers a and b (we say that $\mathbf{p}$ is a *linear combination* of $\mathbf{v}$ and $\mathbf{w}$). Hence we can describe M as

$$M = \{a\mathbf{x} + b\mathbf{w} \mid a,\ b \in \mathbb{R}\}.^5$$

and we say that $\{\mathbf{v},\ \mathbf{w}\}$ is a *spanning set* for M. It is this notion of a spanning set that provides a way to describe all subspaces of $\mathbb{R}^n$.

As in Section 1.3, given vectors $\mathbf{x}_1,\ \mathbf{x}_2,\ \dots,\ \mathbf{x}_k$ in $\mathbb{R}^n$, a vector of the form

$$t_1\mathbf{x}_1 + t_2\mathbf{x}_2 + \cdots + t_k\mathbf{x}_k \quad \text{where the } t_i \text{ are scalars}$$

is called a **linear combination** of the $\mathbf{x}_i$, and t_i is called the **coefficient** of $\mathbf{x}_i$ in the linear combination.

Definition 5.2 Linear Combinations and Span in $\mathbb{R}^n$

The set of all such linear combinations is called the ***span*** *of the $\mathbf{x}_i$ and is denoted*

$$\operatorname{span}\{\mathbf{x}_1,\ \mathbf{x}_2,\ \dots,\ \mathbf{x}_k\} = \{t_1\mathbf{x}_1 + t_2\mathbf{x}_2 + \cdots + t_k\mathbf{x}_k \mid t_i \text{ in } \mathbb{R}\}$$

If $V = \operatorname{span}\{\mathbf{x}_1,\ \mathbf{x}_2,\ \dots,\ \mathbf{x}_k\}$, we say that V is ***spanned*** *by the vectors $\mathbf{x}_1,\ \mathbf{x}_2,\ \dots,\ \mathbf{x}_k$, and that the vectors $\mathbf{x}_1,\ \mathbf{x}_2,\ \dots,\ \mathbf{x}_k$* ***span*** *the space V.*

Here are two examples:

$$\operatorname{span}\{\mathbf{x}\} = \{t\mathbf{x} \mid t \in \mathbb{R}\}$$

which we write as $\operatorname{span}\{\mathbf{x}\} = \mathbb{R}\mathbf{x}$ for simplicity.

$$\operatorname{span}\{\mathbf{x},\ \mathbf{y}\} = \{r\mathbf{x} + s\mathbf{y} \mid r,\ s \in \mathbb{R}\}$$

[4] The vector $\mathbf{n} = \mathbf{v} \times \mathbf{w}$ is nonzero because $\mathbf{v}$ and $\mathbf{w}$ are not parallel.

[5] In particular, this implies that any vector $\mathbf{p}$ orthogonal to $\mathbf{v} \times \mathbf{w}$ must be a linear combination $\mathbf{p} = a\mathbf{v} + b\mathbf{w}$ of $\mathbf{v}$ and $\mathbf{w}$ for some a and b. Can you prove this directly?

In particular, the above discussion shows that, if $\mathbf{v}$ and $\mathbf{w}$ are two nonzero, nonparallel vectors in $\mathbb{R}^3$, then

$$M = \text{span}\,\{\mathbf{v},\ \mathbf{w}\}$$

is the plane in $\mathbb{R}^3$ containing $\mathbf{v}$ and $\mathbf{w}$. Moreover, if $\mathbf{d}$ is any nonzero vector in $\mathbb{R}^3$ (or $\mathbb{R}^2$), then

$$L = \text{span}\,\{\mathbf{v}\} = \{t\mathbf{d} \mid t \in \mathbb{R}\} = \mathbb{R}\mathbf{d}$$

is the line with direction vector $\mathbf{d}$. Hence lines and planes can both be described in terms of spanning sets.

Example 5.1.4

Let $\mathbf{x} = (2,\ -1,\ 2,\ 1)$ and $\mathbf{y} = (3,\ 4,\ -1,\ 1)$ in $\mathbb{R}^4$. Determine whether $\mathbf{p} = (0,\ -11,\ 8,\ 1)$ or $\mathbf{q} = (2,\ 3,\ 1,\ 2)$ are in $U = \text{span}\,\{\mathbf{x},\ \mathbf{y}\}$.

Solution. The vector $\mathbf{p}$ is in U if and only if $\mathbf{p} = s\mathbf{x} + t\mathbf{y}$ for scalars s and t. Equating components gives equations

$$2s + 3t = 0, \quad -s + 4t = -11, \quad 2s - t = 8, \quad \text{and} \quad s + t = 1$$

This linear system has solution $s = 3$ and $t = -2$, so $\mathbf{p}$ is in U. On the other hand, asking that $\mathbf{q} = s\mathbf{x} + t\mathbf{y}$ leads to equations

$$2s + 3t = 2, \quad -s + 4t = 3, \quad 2s - t = 1, \quad \text{and} \quad s + t = 2$$

and this system has *no* solution. So $\mathbf{q}$ does *not* lie in U.

Theorem 5.1.1: Span Theorem

Let $U = \text{span}\,\{\mathbf{x}_1,\ \mathbf{x}_2,\ \dots,\ \mathbf{x}_k\}$ *in* $\mathbb{R}^n$. *Then:*

1. U *is a subspace of* $\mathbb{R}^n$ *containing each* $\mathbf{x}_i$.
2. *If* W *is a subspace of* $\mathbb{R}^n$ *and each* $\mathbf{x}_i \in W$, *then* $U \subseteq W$.

Proof.

1. The zero vector $\mathbf{0}$ is in U because $\mathbf{0} = 0\mathbf{x}_1 + 0\mathbf{x}_2 + \cdots + 0\mathbf{x}_k$ is a linear combination of the $\mathbf{x}_i$. If $\mathbf{x} = t_1\mathbf{x}_1 + t_2\mathbf{x}_2 + \cdots + t_k\mathbf{x}_k$ and $\mathbf{y} = s_1\mathbf{x}_1 + s_2\mathbf{x}_2 + \cdots + s_k\mathbf{x}_k$ are in U, then $\mathbf{x} + \mathbf{y}$ and $a\mathbf{x}$ are in U because
$$\begin{aligned} \mathbf{x} + \mathbf{y} &= (t_1 + s_1)\mathbf{x}_1 + (t_2 + s_2)\mathbf{x}_2 + \cdots + (t_k + s_k)\mathbf{x}_k, \quad \text{and} \\ a\mathbf{x} &= (at_1)\mathbf{x}_1 + (at_2)\mathbf{x}_2 + \cdots + (at_k)\mathbf{x}_k \end{aligned}$$
Finally each $\mathbf{x}_i$ is in U (for example, $\mathbf{x}_2 = 0\mathbf{x}_1 + 1\mathbf{x}_2 + \cdots + 0\mathbf{x}_k$) so S1, S2, and S3 are satisfied for U, proving (1).

2. Let $\mathbf{x} = t_1\mathbf{x}_1 + t_2\mathbf{x}_2 + \cdots + t_k\mathbf{x}_k$ where the t_i are scalars and each $\mathbf{x}_i \in W$. Then each $t_i\mathbf{x}_i \in W$ because W satisfies S3. But then $\mathbf{x} \in W$ because W satisfies S2 (verify). This proves (2). □

Condition (2) in Theorem 5.1.1 can be expressed by saying that $\text{span}\,\{\mathbf{x}_1,\ \mathbf{x}_2,\ \dots,\ \mathbf{x}_k\}$ is the *smallest* subspace of $\mathbb{R}^n$ that contains each $\mathbf{x}_i$. This is useful for showing that two subspaces U and W are equal, since this amounts to showing that both $U \subseteq W$ and $W \subseteq U$. Here is an example of how it is used.

Example 5.1.5

If $\mathbf{x}$ and $\mathbf{y}$ are in $\mathbb{R}^n$, show that $\operatorname{span}\{\mathbf{x}, \mathbf{y}\} = \operatorname{span}\{\mathbf{x}+\mathbf{y}, \mathbf{x}-\mathbf{y}\}$.

Solution. Since both $\mathbf{x}+\mathbf{y}$ and $\mathbf{x}-\mathbf{y}$ are in $\operatorname{span}\{\mathbf{x}, \mathbf{y}\}$, Theorem 5.1.1 gives

$$\operatorname{span}\{\mathbf{x}+\mathbf{y}, \mathbf{x}-\mathbf{y}\} \subseteq \operatorname{span}\{\mathbf{x}, \mathbf{y}\}$$

But $\mathbf{x} = \frac{1}{2}(\mathbf{x}+\mathbf{y}) + \frac{1}{2}(\mathbf{x}-\mathbf{y})$ and $\mathbf{y} = \frac{1}{2}(\mathbf{x}+\mathbf{y}) - \frac{1}{2}(\mathbf{x}-\mathbf{y})$ are both in $\operatorname{span}\{\mathbf{x}+\mathbf{y}, \mathbf{x}-\mathbf{y}\}$, so

$$\operatorname{span}\{\mathbf{x}, \mathbf{y}\} \subseteq \operatorname{span}\{\mathbf{x}+\mathbf{y}, \mathbf{x}-\mathbf{y}\}$$

again by Theorem 5.1.1. Thus $\operatorname{span}\{\mathbf{x}, \mathbf{y}\} = \operatorname{span}\{\mathbf{x}+\mathbf{y}, \mathbf{x}-\mathbf{y}\}$, as desired.

It turns out that many important subspaces are best described by giving a spanning set. Here are three examples, beginning with an important spanning set for $\mathbb{R}^n$ itself. Column j of the $n \times n$ identity matrix I_n is denoted $\mathbf{e}_j$ and called the jth **coordinate vector** in $\mathbb{R}^n$, and the set $\{\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_n\}$ is called the **standard basis** of $\mathbb{R}^n$. If $\mathbf{x} = \left[\begin{array}{c} x_1 \\ x_2 \\ \vdots \\ x_n \end{array}\right]$ is any vector in $\mathbb{R}^n$, then $\mathbf{x} = x_1\mathbf{e}_1 + x_2\mathbf{e}_2 + \cdots + x_n\mathbf{e}_n$, as the reader can verify. This proves:

Example 5.1.6

$\mathbb{R}^n = \operatorname{span}\{\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_n\}$ where $\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_n$ are the columns of I_n.

If A is an $m \times n$ matrix A, the next two examples show that it is a routine matter to find spanning sets for $\operatorname{null} A$ and $\operatorname{im} A$.

Example 5.1.7

Given an $m \times n$ matrix A, let $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_k$ denote the basic solutions to the system $A\mathbf{x} = \mathbf{0}$ given by the gaussian algorithm. Then

$$\operatorname{null} A = \operatorname{span}\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_k\}$$

Solution. If $\mathbf{x} \in \operatorname{null} A$, then $A\mathbf{x} = \mathbf{0}$ so Theorem 1.3.2 shows that $\mathbf{x}$ is a linear combination of the basic solutions; that is, $\operatorname{null} A \subseteq \operatorname{span}\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_k\}$. On the other hand, if $\mathbf{x}$ is in $\operatorname{span}\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_k\}$, then $\mathbf{x} = t_1\mathbf{x}_1 + t_2\mathbf{x}_2 + \cdots + t_k\mathbf{x}_k$ for scalars t_i, so

$$A\mathbf{x} = t_1A\mathbf{x}_1 + t_2A\mathbf{x}_2 + \cdots + t_kA\mathbf{x}_k = t_1\mathbf{0} + t_2\mathbf{0} + \cdots + t_k\mathbf{0} = \mathbf{0}$$

This shows that $\mathbf{x} \in \operatorname{null} A$, and hence that $\operatorname{span}\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_k\} \subseteq \operatorname{null} A$. Thus we have equality.

Example 5.1.8

Let $\mathbf{c}_1, \mathbf{c}_2, \dots, \mathbf{c}_n$ denote the columns of the $m \times n$ matrix A. Then

$$\text{im } A = \text{span}\{\mathbf{c}_1, \mathbf{c}_2, \dots, \mathbf{c}_n\}$$

Solution. If $\{\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_n\}$ is the standard basis of $\mathbb{R}^n$, observe that

$$\left[\begin{array}{cccc} A\mathbf{e}_1 & A\mathbf{e}_2 & \cdots & A\mathbf{e}_n \end{array}\right] = A\left[\begin{array}{cccc} \mathbf{e}_1 & \mathbf{e}_2 & \cdots & \mathbf{e}_n \end{array}\right] = AI_n = A = \left[\begin{array}{ccc} \mathbf{c}_1 & \mathbf{c}_2 & \cdots \mathbf{c}_n \end{array}\right].$$

Hence $\mathbf{c}_i = A\mathbf{e}_i$ is in $\text{im } A$ for each i, so $\text{span}\{\mathbf{c}_1, \mathbf{c}_2, \dots, \mathbf{c}_n\} \subseteq \text{im } A$.

Conversely, let $\mathbf{y}$ be in $\text{im } A$, say $\mathbf{y} = A\mathbf{x}$ for some $\mathbf{x}$ in $\mathbb{R}^n$. If $\mathbf{x} = \left[\begin{array}{c} x_1 \\ x_2 \\ \vdots \\ x_n \end{array}\right]$, then Definition 2.5 gives

$$\mathbf{y} = A\mathbf{x} = x_1\mathbf{c}_1 + x_2\mathbf{c}_2 + \cdots + x_n\mathbf{c}_n \text{ is in } \text{span}\{\mathbf{c}_1, \mathbf{c}_2, \dots, \mathbf{c}_n\}$$

This shows that $\text{im } A \subseteq \text{span}\{\mathbf{c}_1, \mathbf{c}_2, \dots, \mathbf{c}_n\}$, and the result follows.

Exercises for 5.1

We often write vectors in $\mathbb{R}^n$ as rows.

Exercise 5.1.1 In each case determine whether U is a subspace of $\mathbb{R}^3$. Support your answer.

a. $U = \{(1, s, t) \mid s \text{ and } t \text{ in } \mathbb{R}\}$.

b. $U = \{(0, s, t) \mid s \text{ and } t \text{ in } \mathbb{R}\}$.

c. $U = \{(r, s, t) \mid r, s, \text{ and } t \text{ in } \mathbb{R}, -r + 3s + 2t = 0\}$.

d. $U = \{(r, 3s, r-2) \mid r \text{ and } s \text{ in } \mathbb{R}\}$.

e. $U = \{(r, 0, s) \mid r^2 + s^2 = 0, r \text{ and } s \text{ in } \mathbb{R}\}$.

f. $U = \{(2r, -s^2, t) \mid r, s, \text{ and } t \text{ in } \mathbb{R}\}$.

Exercise 5.1.2 In each case determine if $\mathbf{x}$ lies in $U = \text{span}\{\mathbf{y}, \mathbf{z}\}$. If $\mathbf{x}$ is in U, write it as a linear combination of $\mathbf{y}$ and $\mathbf{z}$; if $\mathbf{x}$ is not in U, show why not.

a. $\mathbf{x} = (2, -1, 0, 1)$, $\mathbf{y} = (1, 0, 0, 1)$, and $\mathbf{z} = (0, 1, 0, 1)$.

b. $\mathbf{x} = (1, 2, 15, 11)$, $\mathbf{y} = (2, -1, 0, 2)$, and $\mathbf{z} = (1, -1, -3, 1)$.

c. $\mathbf{x} = (8, 3, -13, 20)$, $\mathbf{y} = (2, 1, -3, 5)$, and $\mathbf{z} = (-1, 0, 2, -3)$.

d. $\mathbf{x} = (2, 5, 8, 3)$, $\mathbf{y} = (2, -1, 0, 5)$, and $\mathbf{z} = (-1, 2, 2, -3)$.

Exercise 5.1.3 In each case determine if the given vectors span $\mathbb{R}^4$. Support your answer.

a. $\{(1, 1, 1, 1), (0, 1, 1, 1), (0, 0, 1, 1), (0, 0, 0, 1)\}$.

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

Exercise 5.1.4 Is it possible that $\{(1, 2, 0), (2, 0, 3)\}$ can span the subspace $U = \{(r, s, 0) \mid r \text{ and } s \text{ in } \mathbb{R}\}$? Defend your answer.

Exercise 5.1.5 Give a spanning set for the zero subspace $\{\mathbf{0}\}$ of $\mathbb{R}^n$.

Exercise 5.1.6 Is $\mathbb{R}^2$ a subspace of $\mathbb{R}^3$? Defend your answer.

Exercise 5.1.7 If $U = \text{span}\{\mathbf{x}, \mathbf{y}, \mathbf{z}\}$ in $\mathbb{R}^n$, show that $U = \text{span}\{\mathbf{x}+t\mathbf{z}, \mathbf{y}, \mathbf{z}\}$ for every t in $\mathbb{R}$.

Exercise 5.1.8 If $U = \text{span}\{\mathbf{x}, \mathbf{y}, \mathbf{z}\}$ in $\mathbb{R}^n$, show that $U = \text{span}\{\mathbf{x}+\mathbf{y}, \mathbf{y}+\mathbf{z}, \mathbf{z}+\mathbf{x}\}$.

Exercise 5.1.9 If $a \neq 0$ is a scalar, show that $\text{span}\{a\mathbf{x}\} = \text{span}\{\mathbf{x}\}$ for every vector $\mathbf{x}$ in $\mathbb{R}^n$.

Exercise 5.1.10 If $a_1, a_2, \dots, a_k$ are nonzero scalars, show that $\text{span}\{a_1\mathbf{x}_1, a_2\mathbf{x}_2, \dots, a_k\mathbf{x}_k\} = \text{span}\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_k\}$ for any vectors $\mathbf{x}_i$ in $\mathbb{R}^n$.

Exercise 5.1.11 If $\mathbf{x} \neq \mathbf{0}$ in $\mathbb{R}^n$, determine all subspaces of $\text{span}\{\mathbf{x}\}$.

Exercise 5.1.12 Suppose that $U = \text{span}\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_k\}$ where each $\mathbf{x}_i$ is in $\mathbb{R}^n$. If A is an $m \times n$ matrix and $A\mathbf{x}_i = \mathbf{0}$ for each i, show that $A\mathbf{y} = \mathbf{0}$ for every vector $\mathbf{y}$ in U.

Exercise 5.1.13 If A is an $m \times n$ matrix, show that, for each invertible $m \times m$ matrix U, $\text{null}(A) = \text{null}(UA)$.

Exercise 5.1.14 If A is an $m \times n$ matrix, show that, for each invertible $n \times n$ matrix V, $\text{im}(A) = \text{im}(AV)$.

Exercise 5.1.15 Let U be a subspace of $\mathbb{R}^n$, and let $\mathbf{x}$ be a vector in $\mathbb{R}^n$.

a. If $a\mathbf{x}$ is in U where $a \neq 0$ is a number, show that $\mathbf{x}$ is in U.

b. If $\mathbf{y}$ and $\mathbf{x}+\mathbf{y}$ are in U where $\mathbf{y}$ is a vector in $\mathbb{R}^n$, show that $\mathbf{x}$ is in U.

Exercise 5.1.16 In each case either show that the statement is true or give an example showing that it is false.

a. If $U \neq \mathbb{R}^n$ is a subspace of $\mathbb{R}^n$ and $\mathbf{x}+\mathbf{y}$ is in U, then $\mathbf{x}$ and $\mathbf{y}$ are both in U.

b. If U is a subspace of $\mathbb{R}^n$ and $r\mathbf{x}$ is in U for all r in $\mathbb{R}$, then $\mathbf{x}$ is in U.

c. If U is a subspace of $\mathbb{R}^n$ and $\mathbf{x}$ is in U, then $-\mathbf{x}$ is also in U.

d. If $\mathbf{x}$ is in U and $U = \text{span}\{\mathbf{y}, \mathbf{z}\}$, then $U = \text{span}\{\mathbf{x}, \mathbf{y}, \mathbf{z}\}$.

e. The empty set of vectors in $\mathbb{R}^n$ is a subspace of $\mathbb{R}^n$.

f. $\begin{bmatrix} 0 \\ 1 \end{bmatrix}$ is in $\text{span}\left\{\begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 2 \\ 0 \end{bmatrix}\right\}$.

Exercise 5.1.17

a. If A and B are $m \times n$ matrices, show that $U = \{\mathbf{x} \text{ in } \mathbb{R}^n \mid A\mathbf{x} = B\mathbf{x}\}$ is a subspace of $\mathbb{R}^n$.

b. What if A is $m \times n$, B is $k \times n$, and $m \neq k$?

Exercise 5.1.18 Suppose that $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_k$ are vectors in $\mathbb{R}^n$. If $\mathbf{y} = a_1\mathbf{x}_1 + a_2\mathbf{x}_2 + \cdots + a_k\mathbf{x}_k$ where $a_1 \neq 0$, show that $\text{span}\{\mathbf{x}_1 \mathbf{x}_2, \dots, \mathbf{x}_k\} = \text{span}\{\mathbf{y}_1, \mathbf{x}_2, \dots, \mathbf{x}_k\}$.

Exercise 5.1.19 If $U \neq \{\mathbf{0}\}$ is a subspace of $\mathbb{R}$, show that $U = \mathbb{R}$.

Exercise 5.1.20 Let U be a nonempty subset of $\mathbb{R}^n$. Show that U is a subspace if and only if S2 and S3 hold.

Exercise 5.1.21 If S and T are nonempty sets of vectors in $\mathbb{R}^n$, and if $S \subseteq T$, show that $\text{span}\{S\} \subseteq \text{span}\{T\}$.

Exercise 5.1.22 Let U and W be subspaces of $\mathbb{R}^n$. Define their **intersection** $U \cap W$ and their **sum** $U + W$ as follows:

$U \cap W = \{\mathbf{x} \in \mathbb{R}^n \mid \mathbf{x} \text{ belongs to both } U \text{ and } W\}$.

$U + W = \{\mathbf{x} \in \mathbb{R}^n \mid \mathbf{x} \text{ is a sum of a vector in } U$ and a vector in $W\}$.

a. Show that $U \cap W$ is a subspace of $\mathbb{R}^n$.

b. Show that $U + W$ is a subspace of $\mathbb{R}^n$.

Exercise 5.1.23 Let P denote an invertible $n \times n$ matrix. If λ is a number, show that

$$E_\lambda(PAP^{-1}) = \{P\mathbf{x} \mid \mathbf{x} \text{ is in } E_\lambda(A)\}$$

for each $n \times n$ matrix A.

Exercise 5.1.24 Show that every proper subspace U of $\mathbb{R}^2$ is a line through the origin. [*Hint*: If $\mathbf{d}$ is a nonzero vector in U, let $L = \mathbb{R}\mathbf{d} = \{r\mathbf{d} \mid r \text{ in } \mathbb{R}\}$ denote the line with direction vector $\mathbf{d}$. If $\mathbf{u}$ is in U but not in L, argue geometrically that every vector $\mathbf{v}$ in $\mathbb{R}^2$ is a linear combination of $\mathbf{u}$ and $\mathbf{d}$.]

5.2 Independence and Dimension

Some spanning sets are better than others. If $U = \text{span}\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_k\}$ is a subspace of $\mathbb{R}^n$, then every vector in U can be written as a linear combination of the $\mathbf{x}_i$ in at least one way. Our interest here is in spanning sets where each vector in U has a *exactly one* representation as a linear combination of these vectors.

Linear Independence

Given $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_k$ in $\mathbb{R}^n$, suppose that two linear combinations are equal:

$$r_1\mathbf{x}_1 + r_2\mathbf{x}_2 + \cdots + r_k\mathbf{x}_k = s_1\mathbf{x}_1 + s_2\mathbf{x}_2 + \cdots + s_k\mathbf{x}_k$$

We are looking for a condition on the set $\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_k\}$ of vectors that guarantees that this representation is *unique*; that is, $r_i = s_i$ for each i. Taking all terms to the left side gives

$$(r_1 - s_1)\mathbf{x}_1 + (r_2 - s_2)\mathbf{x}_2 + \cdots + (r_k - s_k)\mathbf{x}_k = \mathbf{0}$$

so the required condition is that this equation forces all the coefficients $r_i - s_i$ to be zero.

Definition 5.3 Linear Independence in $\mathbb{R}^n$

*With this in mind, we call a set $\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_k\}$ of vectors **linearly independent** (or simply **independent**) if it satisfies the following condition:*

$$\text{If } t_1\mathbf{x}_1 + t_2\mathbf{x}_2 + \cdots + t_k\mathbf{x}_k = \mathbf{0} \text{ then } t_1 = t_2 = \cdots = t_k = 0$$

We record the result of the above discussion for reference.

Theorem 5.2.1

*If $\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_k\}$ is an independent set of vectors in $\mathbb{R}^n$, then every vector in $\text{span}\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_k\}$ has a **unique** representation as a linear combination of the $\mathbf{x}_i$.*

It is useful to state the definition of independence in different language. Let us say that a linear combination **vanishes** if it equals the zero vector, and call a linear combination **trivial** if every coefficient is zero. Then the definition of independence can be compactly stated as follows:

> A set of vectors is independent if and only if the only linear combination that vanishes is the trivial one.

Hence we have a procedure for checking that a set of vectors is independent:

Independence Test

To verify that a set $\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_k\}$ of vectors in $\mathbb{R}^n$ is independent, proceed as follows:

1. *Set a linear combination equal to zero:* $t_1\mathbf{x}_1 + t_2\mathbf{x}_2 + \cdots + t_k\mathbf{x}_k = \mathbf{0}$.
2. *Show that $t_i = 0$ for each i (that is, the linear combination is trivial).*

Of course, if some nontrivial linear combination vanishes, the vectors are not independent.

Example 5.2.1

Determine whether $\{(1, 0, -2, 5), (2, 1, 0, -1), (1, 1, 2, 1)\}$ is independent in $\mathbb{R}^4$.

Solution. Suppose a linear combination vanishes:

$$r(1, 0, -2, 5) + s(2, 1, 0, -1) + t(1, 1, 2, 1) = (0, 0, 0, 0)$$

Equating corresponding entries gives a system of four equations:

$$r + 2s + t = 0,\ s + t = 0,\ -2r + 2t = 0, \text{ and } 5r - s + t = 0$$

The only solution is the trivial one $r = s = t = 0$ (verify), so these vectors are independent by the independence test.

Example 5.2.2

Show that the standard basis $\{\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_n\}$ of $\mathbb{R}^n$ is independent.

Solution. The components of $t_1\mathbf{e}_1 + t_2\mathbf{e}_2 + \cdots + t_n\mathbf{e}_n$ are $t_1, t_2, \dots, t_n$ (see the discussion preceding Example 5.1.6) So the linear combination vanishes if and only if each $t_i = 0$. Hence the independence test applies.

Example 5.2.3

If $\{\mathbf{x}, \mathbf{y}\}$ is independent, show that $\{2\mathbf{x} + 3\mathbf{y}, \mathbf{x} - 5\mathbf{y}\}$ is also independent.

Solution. If $s(2\mathbf{x} + 3\mathbf{y}) + t(\mathbf{x} - 5\mathbf{y}) = \mathbf{0}$, collect terms to get $(2s + t)\mathbf{x} + (3s - 5t)\mathbf{y} = \mathbf{0}$. Since $\{\mathbf{x}, \mathbf{y}\}$ is independent this combination must be trivial; that is, $2s + t = 0$ and $3s - 5t = 0$. These equations have only the trivial solution $s = t = 0$, as required.

Example 5.2.4

Show that the zero vector in $\mathbb{R}^n$ does not belong to any independent set.

Solution. No set $\{\mathbf{0},\ \mathbf{x}_1,\ \mathbf{x}_2,\ \dots,\ \mathbf{x}_k\}$ of vectors is independent because we have a vanishing, nontrivial linear combination $1\cdot\mathbf{0}+0\mathbf{x}_1+0\mathbf{x}_2+\cdots+0\mathbf{x}_k=\mathbf{0}$.

Example 5.2.5

Given $\mathbf{x}$ in $\mathbb{R}^n$, show that $\{\mathbf{x}\}$ is independent if and only if $\mathbf{x}\neq\mathbf{0}$.

Solution. A vanishing linear combination from $\{\mathbf{x}\}$ takes the form $t\mathbf{x}=\mathbf{0}$, t in $\mathbb{R}$. This implies that $t=0$ because $\mathbf{x}\neq\mathbf{0}$.

The next example will be needed later.

Example 5.2.6

Show that the nonzero rows of a row-echelon matrix R are independent.

Solution. We illustrate the case with 3 leading 1s; the general case is analogous. Suppose R has the form $R=\left[\begin{array}{cccccc} 0 & 1 & * & * & * & * \\ 0 & 0 & 0 & 1 & * & * \\ 0 & 0 & 0 & 0 & 1 & * \\ 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right]$ where $*$ indicates a nonspecified number. Let R_1, R_2, and R_3 denote the nonzero rows of R. If $t_1R_1+t_2R_2+t_3R_3=0$ we show that $t_1=0$, then $t_2=0$, and finally $t_3=0$. The condition $t_1R_1+t_2R_2+t_3R_3=0$ becomes

$$(0,\ t_1,\ *,\ *,\ *,\ *)+(0,\ 0,\ 0,\ t_2,\ *,\ *)+(0,\ 0,\ 0,\ 0,\ t_3,\ *)=(0,\ 0,\ 0,\ 0,\ 0,\ 0)$$

Equating second entries show that $t_1=0$, so the condition becomes $t_2R_2+t_3R_3=0$. Now the same argument shows that $t_2=0$. Finally, this gives $t_3R_3=0$ and we obtain $t_3=0$.

A set of vectors in $\mathbb{R}^n$ is called **linearly dependent** (or simply **dependent**) if it is *not* linearly independent, equivalently if some nontrivial linear combination vanishes.

Example 5.2.7

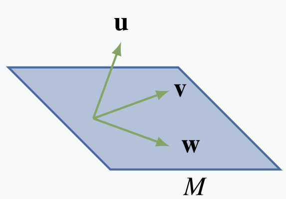
If $\mathbf{v}$ and $\mathbf{w}$ are nonzero vectors in $\mathbb{R}^3$, show that $\{\mathbf{v},\ \mathbf{w}\}$ is dependent if and only if $\mathbf{v}$ and $\mathbf{w}$ are parallel.

Solution. If $\mathbf{v}$ and $\mathbf{w}$ are parallel, then one is a scalar multiple of the other (Theorem 4.1.4), say $\mathbf{v}=a\mathbf{w}$ for some scalar a. Then the nontrivial linear combination $\mathbf{v}-a\mathbf{w}=\mathbf{0}$ vanishes, so $\{\mathbf{v},\ \mathbf{w}\}$ is dependent.

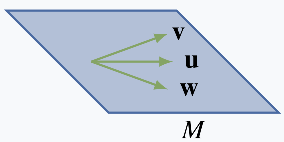
Conversely, if $\{\mathbf{v},\ \mathbf{w}\}$ is dependent, let $s\mathbf{v}+t\mathbf{w}=\mathbf{0}$ be nontrivial, say $s\neq 0$. Then $\mathbf{v}=-\frac{t}{s}\mathbf{w}$ so $\mathbf{v}$ and $\mathbf{w}$ are parallel (by Theorem 4.1.4). A similar argument works if $t\neq 0$.

With this we can give a geometric description of what it means for a set $\{\mathbf{u},\ \mathbf{v},\ \mathbf{w}\}$ in $\mathbb{R}^3$ to be independent. Note that this requirement means that $\{\mathbf{v},\ \mathbf{w}\}$ is also independent ($a\mathbf{v} + b\mathbf{w} = \mathbf{0}$ means that $0\mathbf{u} + a\mathbf{v} + b\mathbf{w} = \mathbf{0}$), so $M = \text{span}\,\{\mathbf{v},\ \mathbf{w}\}$ is the plane containing $\mathbf{v}$, $\mathbf{w}$, and $\mathbf{0}$ (see the discussion preceding Example 5.1.4). So we assume that $\{\mathbf{v},\ \mathbf{w}\}$ is independent in the following example.

Example 5.2.8



$\{\mathbf{u},\ \mathbf{v},\ \mathbf{w}\}$ independent



$\{\mathbf{u},\ \mathbf{v},\ \mathbf{w}\}$ not independent

Let $\mathbf{u}$, $\mathbf{v}$, and $\mathbf{w}$ be nonzero vectors in $\mathbb{R}^3$ where $\{\mathbf{v},\ \mathbf{w}\}$ independent. Show that $\{\mathbf{u},\ \mathbf{v},\ \mathbf{w}\}$ is independent if and only if $\mathbf{u}$ is not in the plane $M = \text{span}\,\{\mathbf{v},\ \mathbf{w}\}$. This is illustrated in the diagrams.

Solution. If $\{\mathbf{u},\ \mathbf{v},\ \mathbf{w}\}$ is independent, suppose $\mathbf{u}$ is in the plane $M = \text{span}\,\{\mathbf{v},\ \mathbf{w}\}$, say $\mathbf{u} = a\mathbf{v} + b\mathbf{w}$, where a and b are in $\mathbb{R}$. Then $1\mathbf{u} - a\mathbf{v} - b\mathbf{w} = \mathbf{0}$, contradicting the independence of $\{\mathbf{u},\ \mathbf{v},\ \mathbf{w}\}$.
On the other hand, suppose that $\mathbf{u}$ is not in M; we must show that $\{\mathbf{u},\ \mathbf{v},\ \mathbf{w}\}$ is independent. If $r\mathbf{u} + s\mathbf{v} + t\mathbf{w} = \mathbf{0}$ where r, s, and t are in $\mathbb{R}^3$, then $r = 0$ since otherwise $\mathbf{u} = -\frac{s}{r}\mathbf{v} + \frac{-t}{r}\mathbf{w}$ is in M. But then $s\mathbf{v} + t\mathbf{w} = \mathbf{0}$, so $s = t = 0$ by our assumption. This shows that $\{\mathbf{u},\ \mathbf{v},\ \mathbf{w}\}$ is independent, as required.

By the inverse theorem, the following conditions are equivalent for an $n \times n$ matrix A:

1. *A is invertible.*
2. *If $A\mathbf{x} = \mathbf{0}$ where $\mathbf{x}$ is in $\mathbb{R}^n$, then $\mathbf{x} = \mathbf{0}$.*
3. *$A\mathbf{x} = \mathbf{b}$ has a solution $\mathbf{x}$ for every vector $\mathbf{b}$ in $\mathbb{R}^n$.*

While condition 1 makes no sense if A is not square, conditions 2 and 3 are meaningful for any matrix A and, in fact, are related to independence and spanning. Indeed, if $\mathbf{c}_1,\ \mathbf{c}_2,\ \dots,\ \mathbf{c}_n$ are the columns of A, and if we write $\mathbf{x} = \left[\begin{array}{c} x_1 \\ x_2 \\ \vdots \\ x_n \end{array}\right]$, then

$$A\mathbf{x} = x_1\mathbf{c}_1 + x_2\mathbf{c}_2 + \cdots + x_n\mathbf{c}_n$$

by Definition 2.5. Hence the definitions of independence and spanning show, respectively, that condition 2 is equivalent to the independence of $\{\mathbf{c}_1,\ \mathbf{c}_2,\ \dots,\ \mathbf{c}_n\}$ and condition 3 is equivalent to the requirement that $\text{span}\,\{\mathbf{c}_1,\ \mathbf{c}_2,\ \dots,\ \mathbf{c}_n\} = \mathbb{R}^m$. This discussion is summarized in the following theorem:

Theorem 5.2.2

If A is an $m \times n$ matrix, let $\{\mathbf{c}_1,\ \mathbf{c}_2,\ \dots,\ \mathbf{c}_n\}$ denote the columns of A.

1. *$\{\mathbf{c}_1,\ \mathbf{c}_2,\ \dots,\ \mathbf{c}_n\}$ is independent in $\mathbb{R}^m$ if and only if $A\mathbf{x} = \mathbf{0}$, $\mathbf{x}$ in $\mathbb{R}^n$, implies $\mathbf{x} = \mathbf{0}$.*
2. *$\mathbb{R}^m = \text{span}\,\{\mathbf{c}_1,\ \mathbf{c}_2,\ \dots,\ \mathbf{c}_n\}$ if and only if $A\mathbf{x} = \mathbf{b}$ has a solution $\mathbf{x}$ for every vector $\mathbf{b}$ in $\mathbb{R}^m$.*

For a *square* matrix A, Theorem 5.2.2 characterizes the invertibility of A in terms of the spanning and independence of its columns (see the discussion preceding Theorem 5.2.2). It is important to be able to discuss these notions for *rows*. If $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_k$ are $1 \times n$ rows, we define $\text{span}\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_k\}$ to be the set of all linear combinations of the $\mathbf{x}_i$ (as matrices), and we say that $\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_k\}$ is linearly independent if the only vanishing linear combination is the trivial one (that is, if $\{\mathbf{x}_1^T, \mathbf{x}_2^T, \dots, \mathbf{x}_k^T\}$ is independent in $\mathbb{R}^n$, as the reader can verify).[6]

Theorem 5.2.3

The following are equivalent for an $n \times n$ matrix A:

1. *A is invertible.*
2. *The columns of A are linearly independent.*
3. *The columns of A span $\mathbb{R}^n$.*
4. *The rows of A are linearly independent.*
5. *The rows of A span the set of all $1 \times n$ rows.*

Proof. Let $\mathbf{c}_1, \mathbf{c}_2, \dots, \mathbf{c}_n$ denote the columns of A.

(1) $\Leftrightarrow$ (2). By Theorem 2.4.5, A is invertible if and only if $A\mathbf{x} = \mathbf{0}$ implies $\mathbf{x} = \mathbf{0}$; this holds if and only if $\{\mathbf{c}_1, \mathbf{c}_2, \dots, \mathbf{c}_n\}$ is independent by Theorem 5.2.2.

(1) $\Leftrightarrow$ (3). Again by Theorem 2.4.5, A is invertible if and only if $A\mathbf{x} = \mathbf{b}$ has a solution for every column B in $\mathbb{R}^n$; this holds if and only if $\text{span}\{\mathbf{c}_1, \mathbf{c}_2, \dots, \mathbf{c}_n\} = \mathbb{R}^n$ by Theorem 5.2.2.

(1) $\Leftrightarrow$ (4). The matrix A is invertible if and only if A^T is invertible (by Corollary 2.4.1 to Theorem 2.4.4); this in turn holds if and only if A^T has independent columns (by (1) $\Leftrightarrow$ (2)); finally, this last statement holds if and only if A has independent rows (because the rows of A are the transposes of the columns of A^T).

(1) $\Leftrightarrow$ (5). The proof is similar to (1) $\Leftrightarrow$ (4). □

Example 5.2.9

Show that $S = \{(2, -2, 5), (-3, 1, 1), (2, 7, -4)\}$ is independent in $\mathbb{R}^3$.

Solution. Consider the matrix $A = \left[\begin{array}{rrr} 2 & -2 & 5 \\ -3 & 1 & 1 \\ 2 & 7 & -4 \end{array}\right]$ with the vectors in S as its rows. A routine computation shows that $\det A = -117 \neq 0$, so A is invertible. Hence S is independent by Theorem 5.2.3. Note that Theorem 5.2.3 also shows that $\mathbb{R}^3 = \text{span } S$.

[6] It is best to view columns and rows as just two different *notations* for ordered n-tuples. This discussion will become redundant in Chapter 6 where we define the general notion of a vector space.

Dimension

It is common geometrical language to say that $\mathbb{R}^3$ is 3-dimensional, that planes are 2-dimensional and that lines are 1-dimensional. The next theorem is a basic tool for clarifying this idea of "dimension". Its importance is difficult to exaggerate.

Theorem 5.2.4: Fundamental Theorem

Let U be a subspace of $\mathbb{R}^n$. If U is spanned by m vectors, and if U contains k linearly independent vectors, then $k \leq m$.

This proof is given in Theorem 6.3.2 in much greater generality.

Definition 5.4 Basis of $\mathbb{R}^n$

*If U is a subspace of $\mathbb{R}^n$, a set $\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_m\}$ of vectors in U is called a **basis** of U if it satisfies the following two conditions:*

1. *$\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_m\}$ is linearly independent.*
2. *$U = \text{span}\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_m\}$.*

The most remarkable result about bases[7] is:

Theorem 5.2.5: Invariance Theorem

If $\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_m\}$ and $\{\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_k\}$ are bases of a subspace U of $\mathbb{R}^n$, then $m = k$.

Proof. We have $k \leq m$ by the fundamental theorem because $\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_m\}$ spans U, and $\{\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_k\}$ is independent. Similarly, by interchanging $\mathbf{x}$'s and $\mathbf{y}$'s we get $m \leq k$. Hence $m = k$. $\square$

The invariance theorem guarantees that there is no ambiguity in the following definition:

Definition 5.5 Dimension of a Subspace of $\mathbb{R}^n$

*If U is a subspace of $\mathbb{R}^n$ and $\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_m\}$ is any basis of U, the number, m, of vectors in the basis is called the **dimension** of U, denoted*

$$\dim U = m$$

The importance of the invariance theorem is that the dimension of U can be determined by counting the number of vectors in *any* basis.[8]

[7] The plural of "basis" is "bases".
[8] We will show in Theorem 5.2.6 that every subspace of $\mathbb{R}^n$ does indeed *have* a basis.

Let $\{\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_n\}$ denote the standard basis of $\mathbb{R}^n$, that is the set of columns of the identity matrix. Then $\mathbb{R}^n = \text{span}\,\{\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_n\}$ by Example 5.1.6, and $\{\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_n\}$ is independent by Example 5.2.2. Hence it is indeed a basis of $\mathbb{R}^n$ in the present terminology, and we have

Example 5.2.10

$\dim(\mathbb{R}^n) = n$ and $\{\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_n\}$ is a basis.

This agrees with our geometric sense that $\mathbb{R}^2$ is two-dimensional and $\mathbb{R}^3$ is three-dimensional. It also says that $\mathbb{R}^1 = \mathbb{R}$ is one-dimensional, and $\{1\}$ is a basis. Returning to subspaces of $\mathbb{R}^n$, we define

$$\dim\{\mathbf{0}\} = 0$$

This amounts to saying $\{\mathbf{0}\}$ has a basis containing *no* vectors. This makes sense because $\mathbf{0}$ cannot belong to *any* independent set (Example 5.2.4).

Example 5.2.11

Let $U = \left\{ \left[\begin{array}{c} r \\ s \\ r \end{array} \right] \mid r, s \text{ in } \mathbb{R} \right\}$. Show that U is a subspace of $\mathbb{R}^3$, find a basis, and calculate $\dim U$.

Solution. Clearly, $\left[\begin{array}{c} r \\ s \\ r \end{array} \right] = r\mathbf{u} + s\mathbf{v}$ where $\mathbf{u} = \left[\begin{array}{c} 1 \\ 0 \\ 1 \end{array} \right]$ and $\mathbf{v} = \left[\begin{array}{c} 0 \\ 1 \\ 0 \end{array} \right]$. It follows that $U = \text{span}\,\{\mathbf{u}, \mathbf{v}\}$, and hence that U is a subspace of $\mathbb{R}^3$. Moreover, if $r\mathbf{u} + s\mathbf{v} = \mathbf{0}$, then $\left[\begin{array}{c} r \\ s \\ r \end{array} \right] = \left[\begin{array}{c} 0 \\ 0 \\ 0 \end{array} \right]$ so $r = s = 0$. Hence $\{\mathbf{u}, \mathbf{v}\}$ is independent, and so a **basis** of U. This means $\dim U = 2$.

Example 5.2.12

Let $B = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n\}$ be a basis of $\mathbb{R}^n$. If A is an invertible $n \times n$ matrix, then $D = \{A\mathbf{x}_1, A\mathbf{x}_2, \dots, A\mathbf{x}_n\}$ is also a basis of $\mathbb{R}^n$.

Solution. Let $\mathbf{x}$ be a vector in $\mathbb{R}^n$. Then $A^{-1}\mathbf{x}$ is in $\mathbb{R}^n$ so, since B is a basis, we have $A^{-1}\mathbf{x} = t_1\mathbf{x}_1 + t_2\mathbf{x}_2 + \cdots + t_n\mathbf{x}_n$ for t_i in $\mathbb{R}$. Left multiplication by A gives $\mathbf{x} = t_1(A\mathbf{x}_1) + t_2(A\mathbf{x}_2) + \cdots + t_n(A\mathbf{x}_n)$, and it follows that D spans $\mathbb{R}^n$. To show independence, let $s_1(A\mathbf{x}_1) + s_2(A\mathbf{x}_2) + \cdots + s_n(A\mathbf{x}_n) = \mathbf{0}$, where the s_i are in $\mathbb{R}$. Then $A(s_1\mathbf{x}_1 + s_2\mathbf{x}_2 + \cdots + s_n\mathbf{x}_n) = \mathbf{0}$ so left multiplication by A^{-1} gives $s_1\mathbf{x}_1 + s_2\mathbf{x}_2 + \cdots + s_n\mathbf{x}_n = \mathbf{0}$. Now the independence of B shows that each $s_i = 0$, and so proves the independence of D. Hence D is a basis of $\mathbb{R}^n$.

While we have found bases in many subspaces of $\mathbb{R}^n$, we have not yet shown that *every* subspace *has* a basis. This is part of the next theorem, the proof of which is deferred to Section 6.4 (Theorem 6.4.1) where it will be proved in more generality.

Theorem 5.2.6

Let $U \neq \{\mathbf{0}\}$ be a subspace of $\mathbb{R}^n$. Then:

1. *U has a basis and $\dim U \leq n$.*
2. *Any independent set in U can be enlarged (by adding vectors from the standard basis) to a basis of U.*
3. *Any spanning set for U can be cut down (by deleting vectors) to a basis of U.*

Example 5.2.13

Find a basis of $\mathbb{R}^4$ containing $S = \{\mathbf{u},\ \mathbf{v}\}$ where $\mathbf{u} = (0,\ 1,\ 2,\ 3)$ and $\mathbf{v} = (2,\ -1,\ 0,\ 1)$.

Solution. By Theorem 5.2.6 we can find such a basis by adding vectors from the standard basis of $\mathbb{R}^4$ to S. If we try $\mathbf{e}_1 = (1,\ 0,\ 0,\ 0)$, we find easily that $\{\mathbf{e}_1,\ \mathbf{u},\ \mathbf{v}\}$ is independent. Now add another vector from the standard basis, say $\mathbf{e}_2$.
Again we find that $B = \{\mathbf{e}_1,\ \mathbf{e}_2,\ \mathbf{u},\ \mathbf{v}\}$ is independent. Since B has $4 = \dim \mathbb{R}^4$ vectors, then B must span $\mathbb{R}^4$ by Theorem 5.2.7 below (or simply verify it directly). Hence B is a basis of $\mathbb{R}^4$.

Theorem 5.2.6 has a number of useful consequences. Here is the first.

Theorem 5.2.7

Let U be a subspace of $\mathbb{R}^n$ where $\dim U = m$ and let $B = \{\mathbf{x}_1,\ \mathbf{x}_2,\ \dots,\ \mathbf{x}_m\}$ be a set of m vectors in U. Then B is independent if and only if B spans U.

Proof. Suppose B is independent. If B does not span U then, by Theorem 5.2.6, B can be enlarged to a basis of U containing more than m vectors. This contradicts the invariance theorem because $\dim U = m$, so B spans U. Conversely, if B spans U but is not independent, then B can be cut down to a basis of U containing fewer than m vectors, again a contradiction. So B is independent, as required. □

As we saw in Example 5.2.13, Theorem 5.2.7 is a “labour-saving” result. It asserts that, given a subspace U of dimension m and a set B of exactly m vectors in U, to prove that B is a basis of U it suffices to show either that B spans U or that B is independent. It is not necessary to verify both properties.

Theorem 5.2.8

Let $U \subseteq W$ be subspaces of $\mathbb{R}^n$. Then:

1. *$\dim U \leq \dim W$.*
2. *If $\dim U = \dim W$, then $U = W$.*

Proof. Write $\dim W = k$, and let B be a basis of U.

1. If $\dim U > k$, then B is an independent set in W containing more than k vectors, contradicting the fundamental theorem. So $\dim U \leq k = \dim W$.

2. If $\dim U = k$, then B is an independent set in W containing $k = \dim W$ vectors, so B spans W by Theorem 5.2.7. Hence $W = \text{span } B = U$, proving (2). □

It follows from Theorem 5.2.8 that if U is a subspace of $\mathbb{R}^n$, then $\dim U$ is one of the integers $0, 1, 2, \dots, n$, and that:

$$\begin{array}{lll} \dim U = 0 & \text{if and only if} & U = \{\mathbf{0}\}, \\ \dim U = n & \text{if and only if} & U = \mathbb{R}^n \end{array}$$

The other subspaces of $\mathbb{R}^n$ are called **proper**. The following example uses Theorem 5.2.8 to show that the proper subspaces of $\mathbb{R}^2$ are the lines through the origin, while the proper subspaces of $\mathbb{R}^3$ are the lines and planes through the origin.

Example 5.2.14

1. If U is a subspace of $\mathbb{R}^2$ or $\mathbb{R}^3$, then $\dim U = 1$ if and only if U is a line through the origin.

2. If U is a subspace of $\mathbb{R}^3$, then $\dim U = 2$ if and only if U is a plane through the origin.

Proof.

1. Since $\dim U = 1$, let $\{\mathbf{u}\}$ be a basis of U. Then $U = \text{span}\{\mathbf{u}\} = \{t\mathbf{u} \mid t \text{ in } \mathbb{R}\}$, so U is the line through the origin with direction vector $\mathbf{u}$. Conversely each line L with direction vector $\mathbf{d} \neq \mathbf{0}$ has the form $L = \{t\mathbf{d} \mid t \text{ in } \mathbb{R}\}$. Hence $\{\mathbf{d}\}$ is a basis of U, so U has dimension 1.

2. If $U \subseteq \mathbb{R}^3$ has dimension 2, let $\{\mathbf{v}, \mathbf{w}\}$ be a basis of U. Then $\mathbf{v}$ and $\mathbf{w}$ are not parallel (by Example 5.2.7) so $\mathbf{n} = \mathbf{v} \times \mathbf{w} \neq \mathbf{0}$. Let $P = \{\mathbf{x} \text{ in } \mathbb{R}^3 \mid \mathbf{n} \cdot \mathbf{x} = 0\}$ denote the plane through the origin with normal $\mathbf{n}$. Then P is a subspace of $\mathbb{R}^3$ (Example 5.1.1) and both $\mathbf{v}$ and $\mathbf{w}$ lie in P (they are orthogonal to $\mathbf{n}$), so $U = \text{span}\{\mathbf{v}, \mathbf{w}\} \subseteq P$ by Theorem 5.1.1. Hence

$$U \subseteq P \subseteq \mathbb{R}^3$$

Since $\dim U = 2$ and $\dim(\mathbb{R}^3) = 3$, it follows from Theorem 5.2.8 that $\dim P = 2$ or 3, whence $P = U$ or $\mathbb{R}^3$. But $P \neq \mathbb{R}^3$ (for example, $\mathbf{n}$ is not in P) and so $U = P$ is a plane through the origin.

Conversely, if U is a plane through the origin, then $\dim U = 0$, 1, 2, or 3 by Theorem 5.2.8. But $\dim U \neq 0$ or 3 because $U \neq \{\mathbf{0}\}$ and $U \neq \mathbb{R}^3$, and $\dim U \neq 1$ by (1). So $\dim U = 2$. □

Note that this proof shows that if $\mathbf{v}$ and $\mathbf{w}$ are nonzero, nonparallel vectors in $\mathbb{R}^3$, then $\text{span}\{\mathbf{v}, \mathbf{w}\}$ is the plane with normal $\mathbf{n} = \mathbf{v} \times \mathbf{w}$. We gave a geometrical verification of this fact in Section 5.1.

Exercises for 5.2

In Exercises 5.2.1-5.2.6 we write vectors $\mathbb{R}^n$ as rows.

Exercise 5.2.1 Which of the following subsets are independent? Support your answer.

a. $\{(1,\ -1,\ 0),\ (3,\ 2,\ -1),\ (3,\ 5,\ -2)\}$ in $\mathbb{R}^3$

b. $\{(1,\ 1,\ 1),\ (1,\ -1,\ 1),\ (0,\ 0,\ 1)\}$ in $\mathbb{R}^3$

c. $\{(1,\ -1,\ 1,\ -1),\ (2,\ 0,\ 1,\ 0),\ (0,\ -2,\ 1,\ -2)\}$ in $\mathbb{R}^4$

d. $\{(1,\ 1,\ 0,\ 0),\ (1,\ 0,\ 1,\ 0),\ (0,\ 0,\ 1,\ 1),\ (0,\ 1,\ 0,\ 1)\}$ in $\mathbb{R}^4$

Exercise 5.2.2 Let $\{\mathbf{x},\ \mathbf{y},\ \mathbf{z},\ \mathbf{w}\}$ be an independent set in $\mathbb{R}^n$. Which of the following sets is independent? Support your answer.

a. $\{\mathbf{x}-\mathbf{y},\ \mathbf{y}-\mathbf{z},\ \mathbf{z}-\mathbf{x}\}$

b. $\{\mathbf{x}+\mathbf{y},\ \mathbf{y}+\mathbf{z},\ \mathbf{z}+\mathbf{x}\}$

c. $\{\mathbf{x}-\mathbf{y},\ \mathbf{y}-\mathbf{z},\ \mathbf{z}-\mathbf{w},\ \mathbf{w}-\mathbf{x}\}$

d. $\{\mathbf{x}+\mathbf{y},\ \mathbf{y}+\mathbf{z},\ \mathbf{z}+\mathbf{w},\ \mathbf{w}+\mathbf{x}\}$

Exercise 5.2.3 Find a basis and calculate the dimension of the following subspaces of $\mathbb{R}^4$.

a. $\operatorname{span}\{(1,\ -1,\ 2,\ 0),\ (2,\ 3,\ 0,\ 3),\ (1,\ 9,\ -6,\ 6)\}$

b. $\operatorname{span}\{(2,\ 1,\ 0,\ -1),\ (-1,\ 1,\ 1,\ 1),\ (2,\ 7,\ 4,\ 1)\}$

c. $\operatorname{span}\{(-1,\ 2,\ 1,\ 0),\ (2,\ 0,\ 3,\ -1),\ (4,\ 4,\ 11,\ -3),\ (3,\ -2,\ 2,\ -1)\}$

d. $\operatorname{span}\{(-2,\ 0,\ 3,\ 1),\ (1,\ 2,\ -1,\ 0),\ (-2,\ 8,\ 5,\ 3),\ (-1,\ 2,\ 2,\ 1)\}$

Exercise 5.2.4 Find a basis and calculate the dimension of the following subspaces of $\mathbb{R}^4$.

a. $U = \left\{ \left[\begin{array}{c} a \\ a+b \\ a-b \\ b \end{array} \right] \middle|\ a \text{ and } b \text{ in } \mathbb{R} \right\}$

b. $U = \left\{ \left[\begin{array}{c} a+b \\ a-b \\ b \\ a \end{array} \right] \middle|\ a \text{ and } b \text{ in } \mathbb{R} \right\}$

c. $U = \left\{ \left[\begin{array}{c} a \\ b \\ c+a \\ c \end{array} \right] \middle|\ a,\ b, \text{ and } c \text{ in } \mathbb{R} \right\}$

d. $U = \left\{ \left[\begin{array}{c} a-b \\ b+c \\ a \\ b+c \end{array} \right] \middle|\ a,\ b, \text{ and } c \text{ in } \mathbb{R} \right\}$

e. $U = \left\{ \left[\begin{array}{c} a \\ b \\ c \\ d \end{array} \right] \middle|\ a+b-c+d = 0 \text{ in } \mathbb{R} \right\}$

f. $U = \left\{ \left[\begin{array}{c} a \\ b \\ c \\ d \end{array} \right] \middle|\ a+b = c+d \text{ in } \mathbb{R} \right\}$

Exercise 5.2.5 Suppose that $\{\mathbf{x},\ \mathbf{y},\ \mathbf{z},\ \mathbf{w}\}$ is a basis of $\mathbb{R}^4$. Show that:

a. $\{\mathbf{x}+a\mathbf{w},\ \mathbf{y},\ \mathbf{z},\ \mathbf{w}\}$ is also a basis of $\mathbb{R}^4$ for any choice of the scalar a.

b. $\{\mathbf{x}+\mathbf{w},\ \mathbf{y}+\mathbf{w},\ \mathbf{z}+\mathbf{w},\ \mathbf{w}\}$ is also a basis of $\mathbb{R}^4$.

c. $\{\mathbf{x},\ \mathbf{x}+\mathbf{y},\ \mathbf{x}+\mathbf{y}+\mathbf{z},\ \mathbf{x}+\mathbf{y}+\mathbf{z}+\mathbf{w}\}$ is also a basis of $\mathbb{R}^4$.

Exercise 5.2.6 Use Theorem 5.2.3 to determine if the following sets of vectors are a basis of the indicated space.

a. $\{(3,\ -1),\ (2,\ 2)\}$ in $\mathbb{R}^2$

b. $\{(1,\ 1,\ -1),\ (1,\ -1,\ 1),\ (0,\ 0,\ 1)\}$ in $\mathbb{R}^3$

c. $\{(-1,\ 1,\ -1),\ (1,\ -1,\ 2),\ (0,\ 0,\ 1)\}$ in $\mathbb{R}^3$

d. $\{(5,\ 2,\ -1),\ (1,\ 0,\ 1),\ (3,\ -1,\ 0)\}$ in $\mathbb{R}^3$

e. $\{(2,\ 1,\ -1,\ 3),\ (1,\ 1,\ 0,\ 2),\ (0,\ 1,\ 0,\ -3),\ (-1,\ 2,\ 3,\ 1)\}$ in $\mathbb{R}^4$

f. $\{(1,\ 0,\ -2,\ 5),\ (4,\ 4,\ -3,\ 2),\ (0,\ 1,\ 0,\ -3),\ (1,\ 3,\ 3,\ -10)\}$ in $\mathbb{R}^4$

Exercise 5.2.7 In each case show that the statement is true or give an example showing that it is false.

a. If $\{\mathbf{x},\ \mathbf{y}\}$ is independent, then $\{\mathbf{x},\ \mathbf{y},\ \mathbf{x}+\mathbf{y}\}$ is independent.

b. If $\{\mathbf{x},\ \mathbf{y},\ \mathbf{z}\}$ is independent, then $\{\mathbf{y},\ \mathbf{z}\}$ is independent.

c. If $\{\mathbf{y},\ \mathbf{z}\}$ is dependent, then $\{\mathbf{x},\ \mathbf{y},\ \mathbf{z}\}$ is dependent for any $\mathbf{x}$.

d. If all of $\mathbf{x}_1,\ \mathbf{x}_2,\ \dots,\ \mathbf{x}_k$ are nonzero, then $\{\mathbf{x}_1,\ \mathbf{x}_2,\ \dots,\ \mathbf{x}_k\}$ is independent.

e. If one of $\mathbf{x}_1,\ \mathbf{x}_2,\ \dots,\ \mathbf{x}_k$ is zero, then $\{\mathbf{x}_1,\ \mathbf{x}_2,\ \dots,\ \mathbf{x}_k\}$ is dependent.

f. If $a\mathbf{x}+b\mathbf{y}+c\mathbf{z}=\mathbf{0}$, then $\{\mathbf{x},\ \mathbf{y},\ \mathbf{z}\}$ is independent.

g. If $\{\mathbf{x},\ \mathbf{y},\ \mathbf{z}\}$ is independent, then $a\mathbf{x}+b\mathbf{y}+c\mathbf{z}=\mathbf{0}$ for some a, b, and c in $\mathbb{R}$.

h. If $\{\mathbf{x}_1,\ \mathbf{x}_2,\ \dots,\ \mathbf{x}_k\}$ is dependent, then $t_1\mathbf{x}_1+t_2\mathbf{x}_2+\cdots+t_k\mathbf{x}_k=\mathbf{0}$ for some numbers t_i in $\mathbb{R}$ not all zero.

i. If $\{\mathbf{x}_1,\ \mathbf{x}_2,\ \dots,\ \mathbf{x}_k\}$ is independent, then $t_1\mathbf{x}_1+t_2\mathbf{x}_2+\cdots+t_k\mathbf{x}_k=\mathbf{0}$ for some t_i in $\mathbb{R}$.

j. Every non-empty subset of a linearly independent set is again linearly independent.

k. Every set containing a spanning set is again a spanning set.

Exercise 5.2.8 If A is an $n\times n$ matrix, show that $\det A=0$ if and only if some column of A is a linear combination of the other columns.

Exercise 5.2.9 Let $\{\mathbf{x},\ \mathbf{y},\ \mathbf{z}\}$ be a linearly independent set in $\mathbb{R}^4$. Show that $\{\mathbf{x},\ \mathbf{y},\ \mathbf{z},\ \mathbf{e}_k\}$ is a basis of $\mathbb{R}^4$ for some $\mathbf{e}_k$ in the standard basis $\{\mathbf{e}_1,\ \mathbf{e}_2,\ \mathbf{e}_3,\ \mathbf{e}_4\}$.

Exercise 5.2.10 If $\{\mathbf{x}_1,\ \mathbf{x}_2,\ \mathbf{x}_3,\ \mathbf{x}_4,\ \mathbf{x}_5,\ \mathbf{x}_6\}$ is an independent set of vectors, show that the subset $\{\mathbf{x}_2,\ \mathbf{x}_3,\ \mathbf{x}_5\}$ is also independent.

Exercise 5.2.11 Let A be any $m\times n$ matrix, and let $\mathbf{b}_1,\ \mathbf{b}_2,\ \mathbf{b}_3,\ \dots,\ \mathbf{b}_k$ be columns in $\mathbb{R}^m$ such that the system $A\mathbf{x}=\mathbf{b}_i$ has a solution $\mathbf{x}_i$ for each i. If $\{\mathbf{b}_1,\ \mathbf{b}_2,\ \mathbf{b}_3,\ \dots,\ \mathbf{b}_k\}$ is independent in $\mathbb{R}^m$, show that $\{\mathbf{x}_1,\ \mathbf{x}_2,\ \mathbf{x}_3,\ \dots,\ \mathbf{x}_k\}$ is independent in $\mathbb{R}^n$.

Exercise 5.2.12 If $\{\mathbf{x}_1,\ \mathbf{x}_2,\ \mathbf{x}_3,\ \dots,\ \mathbf{x}_k\}$ is independent, show $\{\mathbf{x}_1,\ \mathbf{x}_1+\mathbf{x}_2,\ \mathbf{x}_1+\mathbf{x}_2+\mathbf{x}_3,\ \dots,\ \mathbf{x}_1+\mathbf{x}_2+\cdots+\mathbf{x}_k\}$ is also independent.

Exercise 5.2.13 If $\{\mathbf{y},\ \mathbf{x}_1,\ \mathbf{x}_2,\ \mathbf{x}_3,\ \dots,\ \mathbf{x}_k\}$ is independent, show that $\{\mathbf{y}+\mathbf{x}_1,\ \mathbf{y}+\mathbf{x}_2,\ \mathbf{y}+\mathbf{x}_3,\ \dots,\ \mathbf{y}+\mathbf{x}_k\}$ is also independent.

Exercise 5.2.14 If $\{\mathbf{x}_1,\ \mathbf{x}_2,\ \dots,\ \mathbf{x}_k\}$ is independent in $\mathbb{R}^n$, and if $\mathbf{y}$ is not in $\operatorname{span}\{\mathbf{x}_1,\ \mathbf{x}_2,\ \dots,\ \mathbf{x}_k\}$, show that $\{\mathbf{x}_1,\ \mathbf{x}_2,\ \dots,\ \mathbf{x}_k,\ \mathbf{y}\}$ is independent.

Exercise 5.2.15 If A and B are matrices and the columns of AB are independent, show that the columns of B are independent.

Exercise 5.2.16 Suppose that $\{\mathbf{x},\ \mathbf{y}\}$ is a basis of $\mathbb{R}^2$, and let $A=\left[\begin{array}{cc} a & b \\ c & d \end{array}\right]$.

a. If A is invertible, show that $\{a\mathbf{x}+b\mathbf{y},\ c\mathbf{x}+d\mathbf{y}\}$ is a basis of $\mathbb{R}^2$.

b. If $\{a\mathbf{x}+b\mathbf{y},\ c\mathbf{x}+d\mathbf{y}\}$ is a basis of $\mathbb{R}^2$, show that A is invertible.

Exercise 5.2.17 Let A denote an $m\times n$ matrix.

a. Show that $\operatorname{null} A=\operatorname{null}(UA)$ for every invertible $m\times m$ matrix U.

b. Show that $\dim(\operatorname{null} A)=\dim(\operatorname{null}(AV))$ for every invertible $n\times n$ matrix V. [*Hint*: If $\{\mathbf{x}_1,\ \mathbf{x}_2,\ \dots,\ \mathbf{x}_k\}$ is a basis of $\operatorname{null} A$, show that $\{V^{-1}\mathbf{x}_1,\ V^{-1}\mathbf{x}_2,\ \dots,\ V^{-1}\mathbf{x}_k\}$ is a basis of $\operatorname{null}(AV)$.]

Exercise 5.2.18 Let A denote an $m\times n$ matrix.

a. Show that $\operatorname{im} A=\operatorname{im}(AV)$ for every invertible $n\times n$ matrix V.

b. Show that $\dim(\operatorname{im} A)=\dim(\operatorname{im}(UA))$ for every invertible $m\times m$ matrix U. [*Hint*: If $\{\mathbf{y}_1,\ \mathbf{y}_2,\ \dots,\ \mathbf{y}_k\}$ is a basis of $\operatorname{im}(UA)$, show that $\{U^{-1}\mathbf{y}_1,\ U^{-1}\mathbf{y}_2,\ \dots,\ U^{-1}\mathbf{y}_k\}$ is a basis of $\operatorname{im} A$.]

Exercise 5.2.19 Let U and W denote subspaces of $\mathbb{R}^n$, and assume that $U\subseteq W$. If $\dim U=n-1$, show that either $W=U$ or $W=\mathbb{R}^n$.

Exercise 5.2.20 Let U and W denote subspaces of $\mathbb{R}^n$, and assume that $U\subseteq W$. If $\dim W=1$, show that either $U=\{\mathbf{0}\}$ or $U=W$.

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 $\mathbb{R}$.

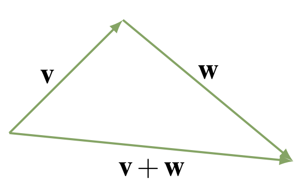
There is an important consequence of the Cauchy inequality. Given $\mathbf{x}$ and $\mathbf{y}$ in $\mathbb{R}^n$, use Example 5.3.2 and the fact that $\mathbf{x}\cdot\mathbf{y} \leq \|\mathbf{x}\|\|\mathbf{y}\|$ to compute

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

Taking positive square roots gives:

Corollary 5.3.1: Triangle Inequality

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



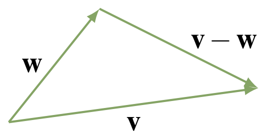
The reason for the name comes from the observation that in $\mathbb{R}^3$ the inequality asserts that the sum of the lengths of two sides of a triangle is not less than the length of the third side. This is illustrated in the diagram.

[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.

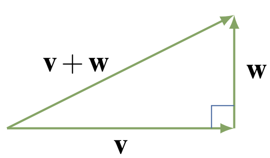
If $\mathbf{x} \neq \mathbf{0}$, it follows from item (6) of Theorem 5.3.1 that $\frac{1}{\|\mathbf{x}\|}\mathbf{x}$ is a unit vector, that is it has length 1.

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

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 $\mathbf{x} = \left[\begin{array}{c}\sqrt{x}\\ \sqrt{y}\end{array}\right]$ *and* $\mathbf{y} = \left[\begin{array}{c}\sqrt{y}\\ \sqrt{x}\end{array}\right]$.]

Exercise 5.3.11 Use the Cauchy inequality to prove that:

a. $r_1 + r_2 + \dots + r_n \leq n(r_1^2 + r_2^2 + \dots + r_n^2)$ 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 $\mathbf{x} = \left[\begin{array}{c}1\\1\end{array}\right]$, $\mathbf{y} = \left[\begin{array}{c}1\\0\end{array}\right]$ and $\mathbf{z} = \left[\begin{array}{c}-2\\3\end{array}\right]$, 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

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

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

Let $\mathbf{c}_{j_1}, \mathbf{c}_{j_2}, \dots, \mathbf{c}_{j_r}$ denote the columns of R containing leading 1s. Then $\{\mathbf{c}_{j_1}, \mathbf{c}_{j_2}, \dots, \mathbf{c}_{j_r}\}$ is independent because the leading 1s are in different rows (and have zeros below and to the left of them). Let U denote the subspace of all columns in $\mathbb{R}^m$ in which the last $m-r$ entries are zero. Then $\dim U = r$ (it is just $\mathbb{R}^r$ with extra zeros). Hence the independent set $\{\mathbf{c}_{j_1}, \mathbf{c}_{j_2}, \dots, \mathbf{c}_{j_r}\}$ is a basis of U by Theorem 5.2.7. Since each $\mathbf{c}_{j_i}$ is in $\operatorname{col} R$, it follows that $\operatorname{col} R = U$, proving (2). □

With Lemma 5.4.2 we can fill a gap in the definition of the rank of a matrix given in Chapter 1. Let A be any matrix and suppose A is carried to some row-echelon matrix R by row operations. Note that R is not unique. In Section 1.2 we defined the **rank** of A, denoted $\operatorname{rank} A$, to be the number of leading 1s in R, that is the number of nonzero rows of R. The fact that this number does not depend on the choice of R was not proved in Section 1.2. However part 1 of Lemma 5.4.2 shows that

$$\operatorname{rank} A = \dim(\operatorname{row} A)$$

and hence that $\operatorname{rank} A$ is independent of R.

Lemma 5.4.2 can be used to find bases of subspaces of $\mathbb{R}^n$ (written as rows). Here is an example.

Example 5.4.1

Find a basis of $U = \operatorname{span}\{(1, 1, 2, 3), (2, 4, 1, 0), (1, 5, -4, -9)\}$.

Solution. U is the row space of $\begin{bmatrix} 1 & 1 & 2 & 3 \\ 2 & 4 & 1 & 0 \\ 1 & 5 & -4 & -9 \end{bmatrix}$. This matrix has row-echelon form $\begin{bmatrix} 1 & 1 & 2 & 3 \\ 0 & 1 & -\frac{3}{2} & -3 \\ 0 & 0 & 0 & 0 \end{bmatrix}$, so $\{(1, 1, 2, 3), (0, 1, -\frac{3}{2}, -3)\}$ is basis of U by Lemma 5.4.2.

Note that $\{(1, 1, 2, 3), (0, 2, -3, -6)\}$ is another basis that avoids fractions.

Lemmas 5.4.1 and 5.4.2 are enough to prove the following fundamental theorem.

Theorem 5.4.1: Rank Theorem

Let A denote any $m \times n$ matrix of rank r. Then

$$\dim(\operatorname{col} A) = \dim(\operatorname{row} A) = r$$

Moreover, if A is carried to a row-echelon matrix R by row operations, then

1. *The r nonzero rows of R are a basis of $\operatorname{row} A$.*
2. *If the leading 1s lie in columns $j_1, j_2, \dots, j_r$ of R, then columns $j_1, j_2, \dots, j_r$ of A are a basis of $\operatorname{col} A$.*

Proof. We have $\operatorname{row} A = \operatorname{row} R$ by Lemma 5.4.1, so (1) follows from Lemma 5.4.2. Moreover, $R = UA$ for some invertible matrix U by Theorem 2.5.1. Now write $A = \begin{bmatrix} \mathbf{c}_1 & \mathbf{c}_2 & \dots & \mathbf{c}_n \end{bmatrix}$ where $\mathbf{c}_1, \mathbf{c}_2, \dots, \mathbf{c}_n$ are the columns of A. Then

$$R = UA = U\begin{bmatrix} \mathbf{c}_1 & \mathbf{c}_2 & \cdots & \mathbf{c}_n \end{bmatrix} = \begin{bmatrix} U\mathbf{c}_1 & U\mathbf{c}_2 & \cdots & U\mathbf{c}_n \end{bmatrix}$$

Thus, in the notation of (2), the set $B = \{U\mathbf{c}_{j_1}, U\mathbf{c}_{j_2}, \dots, U\mathbf{c}_{j_r}\}$ is a basis of col R by Lemma 5.4.2. So, to prove (2) and the fact that $\dim(\text{col } A) = r$, it is enough to show that $D = \{\mathbf{c}_{j_1}, \mathbf{c}_{j_2}, \dots, \mathbf{c}_{j_r}\}$ is a basis of col A. First, D is linearly independent because U is invertible (verify), so we show that, for each j, column $\mathbf{c}_j$ is a linear combination of the $\mathbf{c}_{j_i}$. But $U\mathbf{c}_j$ is column j of R, and so is a linear combination of the $U\mathbf{c}_{j_i}$, say $U\mathbf{c}_j = a_1U\mathbf{c}_{j_1} + a_2U\mathbf{c}_{j_2} + \cdots + a_rU\mathbf{c}_{j_r}$ where each a_i is a real number.

Since U is invertible, it follows that $\mathbf{c}_j = a_1\mathbf{c}_{j_1} + a_2\mathbf{c}_{j_2} + \cdots + a_r\mathbf{c}_{j_r}$ and the proof is complete. □

Example 5.4.2

Compute the rank of $A = \left[\begin{array}{rrrr} 1 & 2 & 2 & -1 \\ 3 & 6 & 5 & 0 \\ 1 & 2 & 1 & 2 \end{array}\right]$ and find bases for row A and col A.

Solution. The reduction of A to row-echelon form is as follows:

$$\left[\begin{array}{rrrr} 1 & 2 & 2 & -1 \\ 3 & 6 & 5 & 0 \\ 1 & 2 & 1 & 2 \end{array}\right] \rightarrow \left[\begin{array}{rrrr} 1 & 2 & 2 & -1 \\ 0 & 0 & -1 & 3 \\ 0 & 0 & -1 & 3 \end{array}\right] \rightarrow \left[\begin{array}{rrrr} 1 & 2 & 2 & -1 \\ 0 & 0 & -1 & 3 \\ 0 & 0 & 0 & 0 \end{array}\right]$$

Hence rank $A = 2$, and $\left\{\left[\begin{array}{rrrr} 1 & 2 & 2 & -1 \end{array}\right], \left[\begin{array}{rrrr} 0 & 0 & 1 & -3 \end{array}\right]\right\}$ is a basis of row A by Lemma 5.4.2. Since the leading 1s are in columns 1 and 3 of the row-echelon matrix, Theorem 5.4.1 shows that columns 1 and 3 of A are a basis $\left\{\left[\begin{array}{r} 1 \\ 3 \\ 1 \end{array}\right], \left[\begin{array}{r} 2 \\ 5 \\ 1 \end{array}\right]\right\}$ of col A.

Theorem 5.4.1 has several important consequences. The first, Corollary 5.4.1 below, follows because the rows of A are independent (respectively span row A) if and only if their transposes are independent (respectively span col A).

Corollary 5.4.1

If A is any matrix, then rank $A = \text{rank}\,(A^T)$.

If A is an $m \times n$ matrix, we have col $A \subseteq \mathbb{R}^m$ and row $A \subseteq \mathbb{R}^n$. Hence Theorem 5.2.8 shows that $\dim(\text{col } A) \leq \dim(\mathbb{R}^m) = m$ and $\dim(\text{row } A) \leq \dim(\mathbb{R}^n) = n$. Thus Theorem 5.4.1 gives:

Corollary 5.4.2

If A is an $m \times n$ matrix, then rank $A \leq m$ *and* rank $A \leq n$.

Corollary 5.4.3

rank $A = \text{rank}\,(UA) = \text{rank}\,(AV)$ *whenever U and V are invertible.*

Proof. Lemma 5.4.1 gives $\text{rank}\, A = \text{rank}\,(UA)$. Using this and Corollary 5.4.1 we get

$$\text{rank}\,(AV) = \text{rank}\,(AV)^T = \text{rank}\,(V^T A^T) = \text{rank}\,(A^T) = \text{rank}\, A$$

□

The next corollary requires a preliminary lemma.

Lemma 5.4.3

Let A, U, and V be matrices of sizes $m \times n$, $p \times m$, and $n \times q$ respectively.

1. $\text{col}\,(AV) \subseteq \text{col}\, A$, *with equality if* $VV' = I_n$ *for some* V'.
2. $\text{row}\,(UA) \subseteq \text{row}\, A$, *with equality if* $U'U = I_m$ *for some* U'.

Proof. For (1), write $V = \left[\mathbf{v}_1,\ \mathbf{v}_2,\ \dots,\ \mathbf{v}_q\right]$ where $\mathbf{v}_j$ is column j of V. Then we have $AV = \left[A\mathbf{v}_1,\ A\mathbf{v}_2,\ \dots,\ A\mathbf{v}_q\right]$, and each $A\mathbf{v}_j$ is in $\text{col}\, A$ by Definition 2.4. It follows that $\text{col}\,(AV) \subseteq \text{col}\, A$. If $VV' = I_n$, we obtain $\text{col}\, A = \text{col}\,[(AV)V'] \subseteq \text{col}\,(AV)$ in the same way. This proves (1).

As to (2), we have $\text{col}\,\left[(UA)^T\right] = \text{col}\,(A^T U^T) \subseteq \text{col}\,(A^T)$ by (1), from which $\text{row}\,(UA) \subseteq \text{row}\, A$. If $U'U = I_m$, this is equality as in the proof of (1). □

Corollary 5.4.4

If A is $m \times n$ and B is $n \times m$, then $\text{rank}\, AB \leq \text{rank}\, A$ *and* $\text{rank}\, AB \leq \text{rank}\, B$.

Proof. By Lemma 5.4.3, $\text{col}\,(AB) \subseteq \text{col}\, A$ and $\text{row}\,(BA) \subseteq \text{row}\, A$, so Theorem 5.4.1 applies. □

In Section 5.1 we discussed two other subspaces associated with an $m \times n$ matrix A: the null space $\text{null}\,(A)$ and the image space $\text{im}\,(A)$

$$\text{null}\,(A) = \{\mathbf{x} \text{ in } \mathbb{R}^n \mid A\mathbf{x} = \mathbf{0}\} \text{ and } \text{im}\,(A) = \{A\mathbf{x} \mid \mathbf{x} \text{ in } \mathbb{R}^n\}$$

Using rank, there are simple ways to find bases of these spaces. If A has rank r, we have $\text{im}\,(A) = \text{col}\,(A)$ by Example 5.1.8, so $\dim[\text{im}\,(A)] = \dim[\text{col}\,(A)] = r$. Hence Theorem 5.4.1 provides a method of finding a basis of $\text{im}\,(A)$. This is recorded as part (2) of the following theorem.

Theorem 5.4.2

Let A denote an $m \times n$ matrix of rank r. Then

1. *The $n - r$ basic solutions to the system $A\mathbf{x} = \mathbf{0}$ provided by the gaussian algorithm are a basis of* $\text{null}\,(A)$, *so* $\dim[\text{null}\,(A)] = n - r$.
2. *Theorem 5.4.1 provides a basis of* $\text{im}\,(A) = \text{col}\,(A)$, *and* $\dim[\text{im}\,(A)] = r$.

Proof. It remains to prove (1). We already know (Theorem 2.2.1) that $\text{null}\,(A)$ is spanned by the $n - r$ basic solutions of $A\mathbf{x} = \mathbf{0}$. Hence using Theorem 5.2.7, it suffices to show that $\dim[\text{null}\,(A)] = n - r$. So let $\{\mathbf{x}_1,\ \dots,\ \mathbf{x}_k\}$ be a basis of $\text{null}\,(A)$, and extend it to a basis $\{\mathbf{x}_1,\ \dots,\ \mathbf{x}_k,\ \mathbf{x}_{k+1},\ \dots,\ \mathbf{x}_n\}$ of $\mathbb{R}^n$ (by

Theorem 5.2.6). It is enough to show that $\{A\mathbf{x}_{k+1}, \dots, A\mathbf{x}_n\}$ is a basis of $\text{im}(A)$; then $n-k=r$ by the above and so $k=n-r$ as required.

Spanning. Choose $A\mathbf{x}$ in $\text{im}(A)$, $\mathbf{x}$ in $\mathbb{R}^n$, and write $\mathbf{x}=a_1\mathbf{x}_1+\cdots+a_k\mathbf{x}_k+a_{k+1}\mathbf{x}_{k+1}+\cdots+a_n\mathbf{x}_n$ where the a_i are in $\mathbb{R}$. Then $A\mathbf{x}=a_{k+1}A\mathbf{x}_{k+1}+\cdots+a_nA\mathbf{x}_n$ because $\{\mathbf{x}_1, \dots, \mathbf{x}_k\} \subseteq \text{null}(A)$.

Independence. Let $t_{k+1}A\mathbf{x}_{k+1}+\cdots+t_nA\mathbf{x}_n=\mathbf{0}$, t_i in $\mathbb{R}$. Then $t_{k+1}\mathbf{x}_{k+1}+\cdots+t_n\mathbf{x}_n$ is in $\text{null}\,A$, so $t_{k+1}\mathbf{x}_{k+1}+\cdots+t_n\mathbf{x}_n=t_1\mathbf{x}_1+\cdots+t_k\mathbf{x}_k$ for some $t_1, \dots, t_k$ in $\mathbb{R}$. But then the independence of the $\mathbf{x}_i$ shows that $t_i=0$ for every i. □

Example 5.4.3

If $A=\left[\begin{array}{rrrr} 1 & -2 & 1 & 1 \\ -1 & 2 & 0 & 1 \\ 2 & -4 & 1 & 0 \end{array}\right]$, find bases of $\text{null}(A)$ and $\text{im}(A)$, and so find their dimensions.

Solution. If $\mathbf{x}$ is in $\text{null}(A)$, then $A\mathbf{x}=\mathbf{0}$, so $\mathbf{x}$ is given by solving the system $A\mathbf{x}=\mathbf{0}$. The reduction of the augmented matrix to reduced form is

$$\left[\begin{array}{rrrr|r} 1 & -2 & 1 & 1 & 0 \\ -1 & 2 & 0 & 1 & 0 \\ 2 & -4 & 1 & 0 & 0 \end{array}\right] \to \left[\begin{array}{rrrr|r} 1 & -2 & 0 & -1 & 0 \\ 0 & 0 & 1 & 2 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right]$$

Hence $r=\text{rank}(A)=2$. Here, $\text{im}(A)=\text{col}(A)$ has basis $\left\{\left[\begin{array}{r} 1 \\ -1 \\ 2 \end{array}\right], \left[\begin{array}{r} 1 \\ 0 \\ 1 \end{array}\right]\right\}$ by Theorem 5.4.1 because the leading 1s are in columns 1 and 3. In particular, $\dim[\text{im}(A)]=2=r$ as in Theorem 5.4.2.

Turning to $\text{null}(A)$, we use gaussian elimination. The leading variables are x_1 and x_3, so the nonleading variables become parameters: $x_2=s$ and $x_4=t$. It follows from the reduced matrix that $x_1=2s+t$ and $x_3=-2t$, so the general solution is

$$\mathbf{x}=\left[\begin{array}{c} x_1 \\ x_2 \\ x_3 \\ x_4 \end{array}\right]=\left[\begin{array}{c} 2s+t \\ s \\ -2t \\ t \end{array}\right]=s\mathbf{x}_1+t\mathbf{x}_2 \text{ where } \mathbf{x}_1=\left[\begin{array}{c} 2 \\ 1 \\ 0 \\ 0 \end{array}\right], \text{ and } \mathbf{x}_2=\left[\begin{array}{r} 1 \\ 0 \\ -2 \\ 1 \end{array}\right].$$

Hence $\text{null}(A)$. But $\mathbf{x}_1$ and $\mathbf{x}_2$ *are* solutions (basic), so

$$\text{null}(A)=\text{span}\{\mathbf{x}_1, \mathbf{x}_2\}$$

However Theorem 5.4.2 asserts that $\{\mathbf{x}_1, \mathbf{x}_2\}$ is a *basis* of $\text{null}(A)$. (In fact it is easy to verify directly that $\{\mathbf{x}_1, \mathbf{x}_2\}$ is independent in this case.) In particular, $\dim[\text{null}(A)]=2=n-r$, as Theorem 5.4.2 asserts.

Let A be an $m\times n$ matrix. Corollary 5.4.2 of Theorem 5.4.1 asserts that $\text{rank}\,A\le m$ and $\text{rank}\,A\le n$, and it is natural to ask when these extreme cases arise. If $\mathbf{c}_1, \mathbf{c}_2, \dots, \mathbf{c}_n$ are the columns of A, Theorem 5.2.2 shows that $\{\mathbf{c}_1, \mathbf{c}_2, \dots, \mathbf{c}_n\}$ spans $\mathbb{R}^m$ if and only if the system $A\mathbf{x}=\mathbf{b}$ is consistent for every $\mathbf{b}$ in $\mathbb{R}^m$, and that $\{\mathbf{c}_1, \mathbf{c}_2, \dots, \mathbf{c}_n\}$ is independent if and only if $A\mathbf{x} = \mathbf{0}$, $\mathbf{x}$ in $\mathbb{R}^n$, implies $\mathbf{x} = \mathbf{0}$. The next two useful theorems improve on both these results, and relate them to when the rank of A is n or m.

Theorem 5.4.3

The following are equivalent for an $m \times n$ matrix A:

1. rank $A = n$.
2. *The rows of A span $\mathbb{R}^n$.*
3. *The columns of A are linearly independent in $\mathbb{R}^m$.*
4. *The $n \times n$ matrix $A^T A$ is invertible.*
5. *$CA = I_n$ for some $n \times m$ matrix C.*
6. *If $A\mathbf{x} = \mathbf{0}$, $\mathbf{x}$ in $\mathbb{R}^n$, then $\mathbf{x} = \mathbf{0}$.*

Proof. (1) $\Rightarrow$ (2). We have $\text{row } A \subseteq \mathbb{R}^n$, and $\dim(\text{row } A) = n$ by (1), so $\text{row } A = \mathbb{R}^n$ by Theorem 5.2.8. This is (2).

(2) $\Rightarrow$ (3). By (2), $\text{row } A = \mathbb{R}^n$, so $\text{rank } A = n$. This means $\dim(\text{col } A) = n$. Since the n columns of A span $\text{col } A$, they are independent by Theorem 5.2.7.

(3) $\Rightarrow$ (4). If $(A^T A)\mathbf{x} = \mathbf{0}$, $\mathbf{x}$ in $\mathbb{R}^n$, we show that $\mathbf{x} = \mathbf{0}$ (Theorem 2.4.5). We have

$$\|A\mathbf{x}\|^2 = (A\mathbf{x})^T A\mathbf{x} = \mathbf{x}^T A^T A\mathbf{x} = \mathbf{x}^T \mathbf{0} = \mathbf{0}$$

Hence $A\mathbf{x} = \mathbf{0}$, so $\mathbf{x} = \mathbf{0}$ by (3) and Theorem 5.2.2.

(4) $\Rightarrow$ (5). Given (4), take $C = (A^T A)^{-1} A^T$.

(5) $\Rightarrow$ (6). If $A\mathbf{x} = \mathbf{0}$, then left multiplication by C (from (5)) gives $\mathbf{x} = \mathbf{0}$.

(6) $\Rightarrow$ (1). Given (6), the columns of A are independent by Theorem 5.2.2. Hence $\dim(\text{col } A) = n$, and (1) follows. □

Theorem 5.4.4

The following are equivalent for an $m \times n$ matrix A:

1. rank $A = m$.
2. *The columns of A span $\mathbb{R}^m$.*
3. *The rows of A are linearly independent in $\mathbb{R}^n$.*
4. *The $m \times m$ matrix AA^T is invertible.*
5. *$AC = I_m$ for some $n \times m$ matrix C.*
6. *The system $A\mathbf{x} = \mathbf{b}$ is consistent for every $\mathbf{b}$ in $\mathbb{R}^m$.*

Proof. (1) $\Rightarrow$ (2). By (1), $\dim\left(\text{col } A = m\right.$, so $\text{col } A = \mathbb{R}^m$ by Theorem 5.2.8.

(2) $\Rightarrow$ (3). By (2), $\text{col } A = \mathbb{R}^m$, so $\text{rank } A = m$. This means $\dim\left(\text{row } A\right) = m$. Since the m rows of A span $\text{row } A$, they are independent by Theorem 5.2.7.

(3) $\Rightarrow$ (4). We have $\text{rank } A = m$ by (3), so the $n \times m$ matrix A^T has rank m. Hence applying Theorem 5.4.3 to A^T in place of A shows that $(A^T)^T A^T$ is invertible, proving (4).

(4) $\Rightarrow$ (5). Given (4), take $C = A^T(AA_T)^{-1}$ in (5).

(5) $\Rightarrow$ (6). Comparing columns in $AC = I_m$ gives $A\mathbf{c}_j = \mathbf{e}_j$ for each j, where $\mathbf{c}_j$ and $\mathbf{e}_j$ denote column j of C and I_m respectively. Given $\mathbf{b}$ in $\mathbb{R}^m$, write $\mathbf{b} = \sum_{j=1}^{m} r_j\mathbf{e}_j$, r_j in $\mathbb{R}$. Then $A\mathbf{x} = \mathbf{b}$ holds with $\mathbf{x} = \sum_{j=1}^{m} r_j\mathbf{c}_j$ as the reader can verify.

(6) $\Rightarrow$ (1). Given (6), the columns of A span $\mathbb{R}^m$ by Theorem 5.2.2. Thus $\text{col } A = \mathbb{R}^m$ and (1) follows. □

Example 5.4.4

Show that $\begin{bmatrix} 3 & x+y+z \\ x+y+z & x^2+y^2+z^2 \end{bmatrix}$ is invertible if x, y, and z are not all equal.

Solution. The given matrix has the form A^TA where $A = \begin{bmatrix} 1 & x \\ 1 & y \\ 1 & z \end{bmatrix}$ has independent columns because x, y, and z are not all equal (verify). Hence Theorem 5.4.3 applies.

Theorem 5.4.3 and Theorem 5.4.4 relate several important properties of an $m \times n$ matrix A to the invertibility of the square, symmetric matrices A^TA and AA^T. In fact, even if the columns of A are not independent or do not span $\mathbb{R}^m$, the matrices A^TA and AA^T are both symmetric and, as such, have real eigenvalues as we shall see. We return to this in Chapter 7.

Exercises for 5.4

Exercise 5.4.1 In each case find bases for the row and column spaces of A and determine the rank of A.

a. $\begin{bmatrix} 2 & -4 & 6 & 8 \\ 2 & -1 & 3 & 2 \\ 4 & -5 & 9 & 10 \\ 0 & -1 & 1 & 2 \end{bmatrix}$

b. $\begin{bmatrix} 2 & -1 & 1 \\ -2 & 1 & 1 \\ 4 & -2 & 3 \\ -6 & 3 & 0 \end{bmatrix}$

c. $\begin{bmatrix} 1 & -1 & 5 & -2 & 2 \\ 2 & -2 & -2 & 5 & 1 \\ 0 & 0 & -12 & 9 & -3 \\ -1 & 1 & 7 & -7 & 1 \end{bmatrix}$

d. $\begin{bmatrix} 1 & 2 & -1 & 3 \\ -3 & -6 & 3 & -2 \end{bmatrix}$

Exercise 5.4.2 In each case find a basis of the subspace U.

a. $U = \text{span}\{(1, -1, 0, 3), (2, 1, 5, 1), (4, -2, 5, 7)\}$

b. $U = \text{span}\{(1, -1, 2, 5, 1), (3, 1, 4, 2, 7), (1, 1, 0, 0, 0), (5, 1, 6, 7, 8)\}$

c. $U = \text{span}\left\{\begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \\ 1 \end{bmatrix}\right\}$

d.

$$U = \text{span}\left\{ \begin{bmatrix} 1 \\ 5 \\ -6 \end{bmatrix}, \begin{bmatrix} 2 \\ 6 \\ -8 \end{bmatrix}, \begin{bmatrix} 3 \\ 7 \\ -10 \end{bmatrix}, \begin{bmatrix} 4 \\ 8 \\ 12 \end{bmatrix} \right\}$$

Exercise 5.4.3

a. Can a 3×4 matrix have independent columns? Independent rows? Explain.

b. If A is 4×3 and rank $A = 2$, can A have independent columns? Independent rows? Explain.

c. If A is an $m \times n$ matrix and rank $A = m$, show that $m \leq n$.

d. Can a nonsquare matrix have its rows independent and its columns independent? Explain.

e. Can the null space of a 3×6 matrix have dimension 2? Explain.

f. Suppose that A is 5×4 and null $(A) = \mathbb{R}\mathbf{x}$ for some column $\mathbf{x} \neq \mathbf{0}$. Can dim $(\text{im } A) = 2$?

Exercise 5.4.4 If A is $m \times n$ show that

$$\text{col}\,(A) = \{A\mathbf{x} \mid \mathbf{x} \text{ in } \mathbb{R}^n\}$$

Exercise 5.4.5 If A is $m \times n$ and B is $n \times m$, show that $AB = 0$ if and only if col $B \subseteq$ null A.

Exercise 5.4.6 Show that the rank does not change when an elementary row or column operation is performed on a matrix.

Exercise 5.4.7 In each case find a basis of the null space of A. Then compute rank A and verify (1) of Theorem 5.4.2.

a. $A = \begin{bmatrix} 3 & 1 & 1 \\ 2 & 0 & 1 \\ 4 & 2 & 1 \\ 1 & -1 & 1 \end{bmatrix}$

b. $A = \begin{bmatrix} 3 & 5 & 5 & 2 & 0 \\ 1 & 0 & 2 & 2 & 1 \\ 1 & 1 & 1 & -2 & -2 \\ -2 & 0 & -4 & -4 & -2 \end{bmatrix}$

Exercise 5.4.8 Let $A = \mathbf{c}\mathbf{r}$ where $\mathbf{c} \neq \mathbf{0}$ is a column in $\mathbb{R}^m$ and $\mathbf{r} \neq \mathbf{0}$ is a row in $\mathbb{R}^n$.

a. Show that col $A = \text{span}\,\{\mathbf{c}\}$ and row $A = \text{span}\,\{\mathbf{r}\}$.

b. Find dim (null A).

c. Show that null A = null $\mathbf{r}$.

Exercise 5.4.9 Let A be $m \times n$ with columns $\mathbf{c}_1, \mathbf{c}_2, \ldots, \mathbf{c}_n$.

a. If $\{\mathbf{c}_1, \ldots, \mathbf{c}_n\}$ is independent, show null $A = \{\mathbf{0}\}$.

b. If null $A = \{\mathbf{0}\}$, show that $\{\mathbf{c}_1, \ldots, \mathbf{c}_n\}$ is independent.

Exercise 5.4.10 Let A be an $n \times n$ matrix.

a. Show that $A^2 = 0$ if and only if col $A \subseteq$ null A.

b. Conclude that if $A^2 = 0$, then rank $A \leq \frac{n}{2}$.

c. Find a matrix A for which col A = null A.

Exercise 5.4.11 Let B be $m \times n$ and let AB be $k \times n$. If rank B = rank (AB), show that null B = null (AB). [*Hint*: Theorem 5.4.1.]

Exercise 5.4.12 Give a careful argument why rank (A^T) = rank A.

Exercise 5.4.13 Let A be an $m \times n$ matrix with columns $\mathbf{c}_1, \mathbf{c}_2, \ldots, \mathbf{c}_n$. If rank $A = n$, show that $\{A^T\mathbf{c}_1, A^T\mathbf{c}_2, \ldots, A^T\mathbf{c}_n\}$ is a basis of $\mathbb{R}^n$.

Exercise 5.4.14 If A is $m \times n$ and $\mathbf{b}$ is $m \times 1$, show that $\mathbf{b}$ lies in the column space of A if and only if rank $[A\ \mathbf{b}]$ = rank A.

Exercise 5.4.15

a. Show that $A\mathbf{x} = \mathbf{b}$ has a solution if and only if rank A = rank $[A\ \mathbf{b}]$. [*Hint*: Exercises 5.4.12 and 5.4.14.]

b. If $A\mathbf{x} = \mathbf{b}$ has no solution, show that rank $[A\ \mathbf{b}] = 1 +$ rank A.

Exercise 5.4.16 Let X be a $k \times m$ matrix. If I is the $m \times m$ identity matrix, show that $I + X^TX$ is invertible.

[*Hint*: $I + X^TX = A^TA$ where $A = \begin{bmatrix} I \\ X \end{bmatrix}$ in block form.]

Exercise 5.4.17 If A is $m \times n$ of rank r, show that A can be factored as $A = PQ$ where P is $m \times r$ with r independent columns, and Q is $r \times n$ with r independent rows. [*Hint*: Let $UAV = \left[\begin{array}{cc} I_r & 0 \\ 0 & 0 \end{array} \right]$ by Theorem 2.5.3, and write $U^{-1} = \left[\begin{array}{cc} U_1 & U_2 \\ U_3 & U_4 \end{array} \right]$ and $V^{-1} = \left[\begin{array}{cc} V_1 & V_2 \\ V_3 & V_4 \end{array} \right]$ in block form, where U_1 and V_1 are $r \times r$.]

Exercise 5.4.18

a. Show that if A and B have independent columns, so does AB.

b. Show that if A and B have independent rows, so does AB.

Exercise 5.4.19 A matrix obtained from A by deleting rows and columns is called a **submatrix** of A. If A has an invertible $k \times k$ submatrix, show that rank $A \geq k$. [*Hint*: Show that row and column operations carry $A \to \left[\begin{array}{cc} I_k & P \\ 0 & Q \end{array} \right]$ in block form.] *Remark*: It can be shown that rank A is the largest integer r such that A has an invertible $r \times r$ submatrix.

5.5 Similarity and Diagonalization

In Section 3.3 we studied diagonalization of a square matrix A, and found important applications (for example to linear dynamical systems). We can now utilize the concepts of subspace, basis, and dimension to clarify the diagonalization process, reveal some new results, and prove some theorems which could not be demonstrated in Section 3.3.

Before proceeding, we introduce a notion that simplifies the discussion of diagonalization, and is used throughout the book.

Similar Matrices

Definition 5.11 Similar Matrices

If A and B are $n \times n$ matrices, we say that A and B are **similar***, and write $A \sim B$, if $B = P^{-1}AP$ for some invertible matrix P.*

Note that $A \sim B$ if and only if $B = QAQ^{-1}$ where Q is invertible (write $P^{-1} = Q$). The language of similarity is used throughout linear algebra. For example, a matrix A is diagonalizable if and only if it is similar to a diagonal matrix.

If $A \sim B$, then necessarily $B \sim A$. To see why, suppose that $B = P^{-1}AP$. Then $A = PBP^{-1} = Q^{-1}BQ$ where $Q = P^{-1}$ is invertible. This proves the second of the following properties of similarity (the others are left as an exercise):

$$\begin{array}{l} 1.\ A \sim A \text{ for all square matrices } A. \\ 2.\ \text{If } A \sim B, \text{ then } B \sim A. \\ 3.\ \text{If } A \sim B \text{ and } B \sim A, \text{ then } A \sim C. \end{array} \tag{5.2}$$

These properties are often expressed by saying that the similarity relation $\sim$ is an **equivalence relation** on the set of $n \times n$ matrices. Here is an example showing how these properties are used.

Example 5.5.1

If A is similar to B and either A or B is diagonalizable, show that the other is also diagonalizable.

Solution. We have $A \sim B$. Suppose that A is diagonalizable, say $A \sim D$ where D is diagonal. Since $B \sim A$ by (2) of (5.2), we have $B \sim A$ and $A \sim D$. Hence $B \sim D$ by (3) of (5.2), so B is diagonalizable too. An analogous argument works if we assume instead that B is diagonalizable.

Similarity is compatible with inverses, transposes, and powers:

$$\text{If } A \sim B \text{ then} \quad A^{-1} \sim B^{-1}, \quad A^T \sim B^T, \quad \text{and} \quad A^k \sim B^k \text{ for all integers } k \geq 1.$$

The proofs are routine matrix computations using Theorem 3.3.1. Thus, for example, if A is diagonalizable, so also are A^T, A^{-1} (if it exists), and A^k (for each $k \geq 1$). Indeed, if $A \sim D$ where D is a diagonal matrix, we obtain $A^T \sim D^T$, $A^{-1} \sim D^{-1}$, and $A^k \sim D^k$, and each of the matrices D^T, D^{-1}, and D^k is diagonal.

We pause to introduce a simple matrix function that will be referred to later.

Definition 5.12 Trace of a Matrix

*The **trace** $\text{tr}\, A$ of an $n \times n$ matrix A is defined to be the sum of the main diagonal elements of A.*

In other words:

$$\text{If } A = \left[a_{ij}\right], \text{ then } \text{tr}\, A = a_{11} + a_{22} + \cdots + a_{nn}.$$

It is evident that $\text{tr}\,(A+B) = \text{tr}\, A + \text{tr}\, B$ and that $\text{tr}\,(cA) = c\, \text{tr}\, A$ holds for all $n \times n$ matrices A and B and all scalars c. The following fact is more surprising.

Lemma 5.5.1

Let A and B be $n \times n$ matrices. Then $\text{tr}\,(AB) = \text{tr}\,(BA)$.

Proof. Write $A = \left[a_{ij}\right]$ and $B = \left[b_{ij}\right]$. For each i, the $(i,\ i)$-entry d_i of the matrix AB is given as follows: $d_i = a_{i1}b_{1i} + a_{i2}b_{2i} + \cdots + a_{in}b_{ni} = \sum_j a_{ij}b_{ji}$. Hence

$$\text{tr}\,(AB) = d_1 + d_2 + \cdots + d_n = \sum_i d_i = \sum_i \left(\sum_j a_{ij}b_{ji} \right)$$

Similarly we have $\text{tr}\,(BA) = \sum_i (\sum_j b_{ij}a_{ji})$. Since these two double sums are the same, Lemma 5.5.1 is proved. □

As the name indicates, similar matrices share many properties, some of which are collected in the next theorem for reference.

Theorem 5.5.1

If A and B are similar $n \times n$ matrices, then A and B have the same determinant, rank, trace, characteristic polynomial, and eigenvalues.

Proof. Let $B = P^{-1}AP$ for some invertible matrix P. Then we have

$$\det B = \det\left(P^{-1}\right) \det A \det P = \det A \text{ because } \det\left(P^{-1}\right) = 1/\det P$$

Similarly, $\operatorname{rank} B = \operatorname{rank}\left(P^{-1}AP\right) = \operatorname{rank} A$ by Corollary 5.4.3. Next Lemma 5.5.1 gives

$$\operatorname{tr}\left(P^{-1}AP\right) = \operatorname{tr}\left[P^{-1}(AP)\right] = \operatorname{tr}\left[(AP)P^{-1}\right] = \operatorname{tr} A$$

As to the characteristic polynomial,

$$\begin{aligned} c_B(x) = \det(xI - B) &= \det\left\{x(P^{-1}IP) - P^{-1}AP\right\} \\ &= \det\left\{P^{-1}(xI - A)P\right\} \\ &= \det(xI - A) \\ &= c_A(x) \end{aligned}$$

Finally, this shows that A and B have the same eigenvalues because the eigenvalues of a matrix are the roots of its characteristic polynomial. □

Example 5.5.2

Sharing the five properties in Theorem 5.5.1 does not guarantee that two matrices are similar. The matrices $A = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$ and $I = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ have the same determinant, rank, trace, characteristic polynomial, and eigenvalues, but they are not similar because $P^{-1}IP = I$ for any invertible matrix P.

Diagonalization Revisited

Recall that a square matrix A is **diagonalizable** if there exists an invertible matrix P such that $P^{-1}AP = D$ is a diagonal matrix, that is if A is similar to a diagonal matrix D. Unfortunately, not all matrices are diagonalizable, for example $\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$ (see Example 3.3.10). Determining whether A is diagonalizable is closely related to the eigenvalues and eigenvectors of A. Recall that a number λ is called an **eigenvalue** of A if $A\mathbf{x} = \lambda\mathbf{x}$ for some nonzero column $\mathbf{x}$ in $\mathbb{R}^n$, and any such nonzero vector $\mathbf{x}$ is called an **eigenvector** of A corresponding to λ (or simply a λ-eigenvector of A). The eigenvalues and eigenvectors of A are closely related to the **characteristic polynomial** $c_A(x)$ of A, defined by

$$c_A(x) = \det(xI - A)$$

If A is $n \times n$ this is a polynomial of degree n, and its relationship to the eigenvalues is given in the following theorem (a repeat of Theorem 3.3.2).

Theorem 5.5.2

Let A be an $n \times n$ matrix.

1. *The eigenvalues λ of A are the roots of the characteristic polynomial $c_A(x)$ of A.*
2. *The λ-eigenvectors $\mathbf{x}$ are the nonzero solutions to the homogeneous system*

$$(\lambda I - A)\mathbf{x} = \mathbf{0}$$

of linear equations with $\lambda I - A$ as coefficient matrix.

Example 5.5.3

Show that the eigenvalues of a triangular matrix are the main diagonal entries.

Solution. Assume that A is triangular. Then the matrix $xI - A$ is also triangular and has diagonal entries $(x - a_{11}), (x - a_{22}), \dots, (x - a_{nn})$ where $A = \left[a_{ij}\right]$. Hence Theorem 3.1.4 gives

$$c_A(x) = (x - a_{11})(x - a_{22}) \cdots (x - a_{nn})$$

and the result follows because the eigenvalues are the roots of $c_A(x)$.

Theorem 3.3.4 asserts (in part) that an $n \times n$ matrix A is diagonalizable if and only if it has n eigenvectors $\mathbf{x}_1, \dots, \mathbf{x}_n$ such that the matrix $P = \left[\begin{array}{ccc} \mathbf{x}_1 & \cdots & \mathbf{x}_n \end{array}\right]$ with the $\mathbf{x}_i$ as columns is invertible. This is equivalent to requiring that $\{\mathbf{x}_1, \dots, \mathbf{x}_n\}$ is a basis of $\mathbb{R}^n$ consisting of eigenvectors of A. Hence we can restate Theorem 3.3.4 as follows:

Theorem 5.5.3

Let A be an $n \times n$ matrix.

1. *A is diagonalizable if and only if $\mathbb{R}^n$ has a basis $\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n\}$ consisting of eigenvectors of A.*
2. *When this is the case, the matrix $P = \left[\begin{array}{cccc} \mathbf{x}_1 & \mathbf{x}_2 & \cdots & \mathbf{x}_n \end{array}\right]$ is invertible and $P^{-1}AP = \operatorname{diag}(\lambda_1, \lambda_2, \dots, \lambda_n)$ where, for each i, λ_i is the eigenvalue of A corresponding to $\mathbf{x}_i$.*

The next result is a basic tool for determining when a matrix is diagonalizable. It reveals an important connection between eigenvalues and linear independence: Eigenvectors corresponding to distinct eigenvalues are necessarily linearly independent.

Theorem 5.5.4

Let $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_k$ be eigenvectors corresponding to distinct eigenvalues $\lambda_1, \lambda_2, \dots, \lambda_k$ of an $n \times n$ matrix A. Then $\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_k\}$ is a linearly independent set.

Proof. We use induction on k. If $k = 1$, then $\{\mathbf{x}_1\}$ is independent because $\mathbf{x}_1 \neq \mathbf{0}$. In general, suppose the theorem is true for some $k \geq 1$. Given eigenvectors $\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_{k+1}\}$, suppose a linear combination vanishes:

$$t_1\mathbf{x}_1 + t_2\mathbf{x}_2 + \cdots + t_{k+1}\mathbf{x}_{k+1} = \mathbf{0} \tag{5.3}$$

We must show that each $t_i = 0$. Left multiply (5.3) by A and use the fact that $A\mathbf{x}_i = \lambda_i\mathbf{x}_i$ to get

$$t_1\lambda_1\mathbf{x}_1 + t_2\lambda_2\mathbf{x}_2 + \cdots + t_{k+1}\lambda_{k+1}\mathbf{x}_{k+1} = \mathbf{0} \tag{5.4}$$

If we multiply (5.3) by λ_1 and subtract the result from (5.4), the first terms cancel and we obtain

$$t_2(\lambda_2 - \lambda_1)\mathbf{x}_2 + t_3(\lambda_3 - \lambda_1)\mathbf{x}_3 + \cdots + t_{k+1}(\lambda_{k+1} - \lambda_1)\mathbf{x}_{k+1} = \mathbf{0}$$

Since $\mathbf{x}_2, \mathbf{x}_3, \dots, \mathbf{x}_{k+1}$ correspond to distinct eigenvalues $\lambda_2, \lambda_3, \dots, \lambda_{k+1}$, the set $\{\mathbf{x}_2, \mathbf{x}_3, \dots, \mathbf{x}_{k+1}\}$ is independent by the induction hypothesis. Hence,

$$t_2(\lambda_2 - \lambda_1) = 0, \quad t_3(\lambda_3 - \lambda_1) = 0, \quad \dots, \quad t_{k+1}(\lambda_{k+1} - \lambda_1) = 0$$

and so $t_2 = t_3 = \cdots = t_{k+1} = 0$ because the λ_i are distinct. Hence (5.3) becomes $t_1\mathbf{x}_1 = \mathbf{0}$, which implies that $t_1 = 0$ because $\mathbf{x}_1 \neq \mathbf{0}$. This is what we wanted. □

Theorem 5.5.4 will be applied several times; we begin by using it to give a useful condition for when a matrix is diagonalizable.

Theorem 5.5.5

If A is an $n \times n$ matrix with n distinct eigenvalues, then A is diagonalizable.

Proof. Choose one eigenvector for each of the n distinct eigenvalues. Then these eigenvectors are independent by Theorem 5.5.4, and so are a basis of $\mathbb{R}^n$ by Theorem 5.2.7. Now use Theorem 5.5.3. □

Example 5.5.4

Show that $A = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 2 & 3 \\ -1 & 1 & 0 \end{bmatrix}$ is diagonalizable.

Solution. A routine computation shows that $c_A(x) = (x-1)(x-3)(x+1)$ and so has distinct eigenvalues 1, 3, and -1. Hence Theorem 5.5.5 applies.

However, a matrix can have multiple eigenvalues as we saw in Section 3.3. To deal with this situation, we prove an important lemma which formalizes a technique that is basic to diagonalization, and which will be used three times below.

Lemma 5.5.2

Let $\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_k\}$ be a linearly independent set of eigenvectors of an $n \times n$ matrix A, extend it to a basis $\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_k, \dots, \mathbf{x}_n\}$ of $\mathbb{R}^n$, and let

$$P = \left[\begin{array}{cccc} \mathbf{x}_1 & \mathbf{x}_2 & \cdots & \mathbf{x}_n \end{array} \right]$$

be the (invertible) $n \times n$ matrix with the $\mathbf{x}_i$ as its columns. If $\lambda_1, \lambda_2, \dots, \lambda_k$ are the (not necessarily distinct) eigenvalues of A corresponding to $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_k$ respectively, then $P^{-1}AP$ has block form

$$P^{-1}AP = \left[\begin{array}{cc} \operatorname{diag}(\lambda_1, \lambda_2, \dots, \lambda_k) & B \\ 0 & A_1 \end{array} \right]$$

where B has size $k \times (n-k)$ and A_1 has size $(n-k) \times (n-k)$.

Proof. If $\{\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_n\}$ is the standard basis of $\mathbb{R}^n$, then

$$\begin{aligned} \left[\begin{array}{cccc} \mathbf{e}_1 & \mathbf{e}_2 & \dots & \mathbf{e}_n \end{array} \right] = I_n = P^{-1}P = P^{-1} \left[\begin{array}{cccc} \mathbf{x}_1 & \mathbf{x}_2 & \cdots & \mathbf{x}_n \end{array} \right] \\ = \left[\begin{array}{cccc} P^{-1}\mathbf{x}_1 & P^{-1}\mathbf{x}_2 & \cdots & P^{-1}\mathbf{x}_n \end{array} \right] \end{aligned}$$

Comparing columns, we have $P^{-1}\mathbf{x}_i = \mathbf{e}_i$ for each $1 \leq i \leq n$. On the other hand, observe that

$$P^{-1}AP = P^{-1}A \left[\begin{array}{cccc} \mathbf{x}_1 & \mathbf{x}_2 & \cdots & \mathbf{x}_n \end{array} \right] = \left[\begin{array}{cccc} (P^{-1}A)\mathbf{x}_1 & (P^{-1}A)\mathbf{x}_2 & \cdots & (P^{-1}A)\mathbf{x}_n \end{array} \right]$$

Hence, if $1 \leq i \leq k$, column i of $P^{-1}AP$ is

$$(P^{-1}A)\mathbf{x}_i = P^{-1}(\lambda_i \mathbf{x}_i) = \lambda_i(P^{-1}\mathbf{x}_i) = \lambda_i \mathbf{e}_i$$

This describes the first k columns of $P^{-1}AP$, and Lemma 5.5.2 follows. □

Note that Lemma 5.5.2 (with $k = n$) shows that an $n \times n$ matrix A is diagonalizable if $\mathbb{R}^n$ has a basis of eigenvectors of A, as in (1) of Theorem 5.5.3.

Definition 5.13 Eigenspace of a Matrix

If λ is an eigenvalue of an $n \times n$ matrix A, define the **eigenspace** *of A corresponding to λ by*

$$E_\lambda(A) = \{\mathbf{x} \text{ in } \mathbb{R}^n \mid A\mathbf{x} = \lambda \mathbf{x}\}$$

This is a subspace of $\mathbb{R}^n$ and the eigenvectors corresponding to λ are just the nonzero vectors in $E_\lambda(A)$. In fact $E_\lambda(A)$ is the null space of the matrix $(\lambda I - A)$:

$$E_\lambda(A) = \{\mathbf{x} \mid (\lambda I - A)\mathbf{x} = \mathbf{0}\} = \operatorname{null}(\lambda I - A)$$

Hence, by Theorem 5.4.2, the basic solutions of the homogeneous system $(\lambda I - A)\mathbf{x} = \mathbf{0}$ given by the gaussian algorithm form a basis for $E_\lambda(A)$. In particular

$$\dim E_\lambda(A) \text{ is the number of basic solutions } \mathbf{x} \text{ of } (\lambda I - A)\mathbf{x} = \mathbf{0} \tag{5.5}$$

Now recall (Definition 3.7) that the **multiplicity**[11] of an eigenvalue λ of A is the number of times λ occurs as a root of the characteristic polynomial $c_A(x)$ of A. In other words, the multiplicity of λ is the largest integer $m \geq 1$ such that

$$c_A(x) = (x-\lambda)^m g(x)$$

for some polynomial $g(x)$. Because of (5.5), the assertion (without proof) in Theorem 3.3.5 can be stated as follows: A square matrix is diagonalizable if and only if the multiplicity of each eigenvalue λ equals $\dim\left[E_\lambda(A)\right]$. We are going to prove this, and the proof requires the following result which is valid for *any* square matrix, diagonalizable or not.

Lemma 5.5.3

Let λ be an eigenvalue of multiplicity m of a square matrix A. Then $\dim\left[E_\lambda(A)\right] \leq m$.

Proof. Write $\dim\left[E_\lambda(A)\right] = d$. It suffices to show that $c_A(x) = (x-\lambda)^d g(x)$ for some polynomial $g(x)$, because m is the highest power of $(x-\lambda)$ that divides $c_A(x)$. To this end, let $\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_d\}$ be a basis of $E_\lambda(A)$. Then Lemma 5.5.2 shows that an invertible $n \times n$ matrix P exists such that

$$P^{-1}AP = \left[\begin{array}{cc} \lambda I_d & B \\ 0 & A_1 \end{array}\right]$$

in block form, where I_d denotes the $d \times d$ identity matrix. Now write $A' = P^{-1}AP$ and observe that $c_{A'}(x) = c_A(x)$ by Theorem 5.5.1. But Theorem 3.1.5 gives

$$\begin{aligned} c_A(x) = c_{A'}(x) = \det(xI_n - A') &= \det \left[\begin{array}{cc} (x-\lambda)I_d & -B \\ 0 & xI_{n-d} - A_1 \end{array}\right] \\ &= \det\left[(x-\lambda)I_d\right] \det\left[(xI_{n-d} - A_1)\right] \\ &= (x-\lambda)^d g(x) \end{aligned}$$

where $g(x) = cA_1(x)$. This is what we wanted. □

It is impossible to ignore the question when equality holds in Lemma 5.5.3 for each eigenvalue λ. It turns out that this characterizes the diagonalizable $n \times n$ matrices A for which $c_A(x)$ **factors completely** over $\mathbb{R}$. By this we mean that $c_A(x) = (x-\lambda_1)(x-\lambda_2)\cdots(x-\lambda_n)$, where the λ_i are *real* numbers (not necessarily distinct); in other words, every eigenvalue of A is real. This need not happen (consider $A = \left[\begin{array}{rr} 0 & -1 \\ 1 & 0 \end{array}\right]$), and we investigate the general case below.

Theorem 5.5.6

The following are equivalent for a square matrix A for which $c_A(x)$ factors completely.

1. *A is diagonalizable.*
2. $\dim\left[E_\lambda(A)\right]$ *equals the multiplicity of λ for every eigenvalue λ of the matrix A.*

[11] This is often called the *algebraic* multiplicity of λ.

Proof. Let A be $n \times n$ and let $\lambda_1, \lambda_2, \dots, \lambda_k$ be the distinct eigenvalues of A. For each i, let m_i denote the multiplicity of λ_i and write $d_i = \dim\left[E_{\lambda_i}(A)\right]$. Then

$$c_A(x) = (x-\lambda_1)^{m_1}(x-\lambda_2)^{m_2}\dots(x-\lambda_k)^{m_k}$$

so $m_1 + \cdots + m_k = n$ because $c_A(x)$ has degree n. Moreover, $d_i \leq m_i$ for each i by Lemma 5.5.3.

(1) $\Rightarrow$ (2). By (1), $\mathbb{R}^n$ has a basis of n eigenvectors of A, so let t_i of them lie in $E_{\lambda_i}(A)$ for each i. Since the subspace spanned by these t_i eigenvectors has dimension t_i, we have $t_i \leq d_i$ for each i by Theorem 5.2.4. Hence

$$n = t_1 + \cdots + t_k \leq d_1 + \cdots + d_k \leq m_1 + \cdots + m_k = n$$

It follows that $d_1 + \cdots + d_k = m_1 + \cdots + m_k$ so, since $d_i \leq m_i$ for each i, we must have $d_i = m_i$. This is (2).

(2) $\Rightarrow$ (1). Let B_i denote a basis of $E_{\lambda_i}(A)$ for each i, and let $B = B_1 \cup \cdots \cup B_k$. Since each B_i contains m_i vectors by (2), and since the B_i are pairwise disjoint (the λ_i are distinct), it follows that B contains n vectors. So it suffices to show that B is linearly independent (then B is a basis of $\mathbb{R}^n$). Suppose a linear combination of the vectors in B vanishes, and let $\mathbf{y}_i$ denote the sum of all terms that come from B_i. Then $\mathbf{y}_i$ lies in $E_{\lambda_i}(A)$, so the nonzero $\mathbf{y}_i$ are independent by Theorem 5.5.4 (as the λ_i are distinct). Since the sum of the $\mathbf{y}_i$ is zero, it follows that $\mathbf{y}_i = \mathbf{0}$ for each i. Hence all coefficients of terms in $\mathbf{y}_i$ are zero (because B_i is independent). Since this holds for each i, it shows that B is independent. □

Example 5.5.5

If $A = \left[\begin{array}{rrr} 5 & 8 & 16 \\ 4 & 1 & 8 \\ -4 & -4 & -11 \end{array}\right]$ and $B = \left[\begin{array}{rrr} 2 & 1 & 1 \\ 2 & 1 & -2 \\ -1 & 0 & -2 \end{array}\right]$ show that A is diagonalizable but B is not.

Solution. We have $c_A(x) = (x+3)^2(x-1)$ so the eigenvalues are $\lambda_1 = -3$ and $\lambda_2 = 1$. The corresponding eigenspaces are $E_{\lambda_1}(A) = \text{span}\,\{\mathbf{x}_1, \mathbf{x}_2\}$ and $E_{\lambda_2}(A) = \text{span}\,\{\mathbf{x}_3\}$ where

$$\mathbf{x}_1 = \left[\begin{array}{r} -1 \\ 1 \\ 0 \end{array}\right], \mathbf{x}_2 = \left[\begin{array}{r} -2 \\ 0 \\ 1 \end{array}\right], \mathbf{x}_3 = \left[\begin{array}{r} 2 \\ 1 \\ -1 \end{array}\right]$$

as the reader can verify. Since $\{\mathbf{x}_1, \mathbf{x}_2\}$ is independent, we have $\dim(E_{\lambda_1}(A)) = 2$ which is the multiplicity of λ_1. Similarly, $\dim(E_{\lambda_2}(A)) = 1$ equals the multiplicity of λ_2. Hence A is diagonalizable by Theorem 5.5.6, and a diagonalizing matrix is $P = \left[\begin{array}{ccc} \mathbf{x}_1 & \mathbf{x}_2 & \mathbf{x}_3 \end{array}\right]$.
Turning to B, $c_B(x) = (x+1)^2(x-3)$ so the eigenvalues are $\lambda_1 = -1$ and $\lambda_2 = 3$. The corresponding eigenspaces are $E_{\lambda_1}(B) = \text{span}\,\{\mathbf{y}_1\}$ and $E_{\lambda_2}(B) = \text{span}\,\{\mathbf{y}_2\}$ where

$$\mathbf{y}_1 = \left[\begin{array}{r} -1 \\ 2 \\ 1 \end{array}\right], \mathbf{y}_2 = \left[\begin{array}{r} 5 \\ 6 \\ -1 \end{array}\right]$$

Here $\dim(E_{\lambda_1}(B)) = 1$ is *smaller* than the multiplicity of λ_1, so the matrix B is *not* diagonalizable, again by Theorem 5.5.6. The fact that $\dim(E_{\lambda_1}(B)) = 1$ means that there is no possibility of finding *three* linearly independent eigenvectors.

Complex Eigenvalues

All the matrices we have considered have had real eigenvalues. But this need not be the case: The matrix $A = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$ has characteristic polynomial $c_A(x) = x^2 + 1$ which has no real roots. Nonetheless, this matrix is diagonalizable; the only difference is that we must use a larger set of scalars, the complex numbers. The basic properties of these numbers are outlined in Appendix A.

Indeed, nearly everything we have done for real matrices can be done for complex matrices. The methods are the same; the only difference is that the arithmetic is carried out with complex numbers rather than real ones. For example, the gaussian algorithm works in exactly the same way to solve systems of linear equations with complex coefficients, matrix multiplication is defined the same way, and the matrix inversion algorithm works in the same way.

But the complex numbers are better than the real numbers in one respect: While there are polynomials like $x^2 + 1$ with real coefficients that have no real root, this problem does not arise with the complex numbers: *Every* nonconstant polynomial with complex coefficients has a complex root, and hence factors completely as a product of linear factors. This fact is known as the fundamental theorem of algebra.[12]

Example 5.5.6

Diagonalize the matrix $A = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$.

Solution. The characteristic polynomial of A is

$$c_A(x) = \det(xI - A) = x^2 + 1 = (x - i)(x + i)$$

where $i^2 = -1$. Hence the eigenvalues are $\lambda_1 = i$ and $\lambda_2 = -i$, with corresponding eigenvectors $\mathbf{x}_1 = \begin{bmatrix} 1 \\ -i \end{bmatrix}$ and $\mathbf{x}_2 = \begin{bmatrix} 1 \\ i \end{bmatrix}$. Hence A is diagonalizable by the complex version of Theorem 5.5.5, and the complex version of Theorem 5.5.3 shows that $P = \begin{bmatrix} \mathbf{x}_1 & \mathbf{x}_2 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ -i & i \end{bmatrix}$ is invertible and $P^{-1}AP = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} = \begin{bmatrix} i & 0 \\ 0 & -i \end{bmatrix}$. Of course, this can be checked directly.

We shall return to complex linear algebra in Section 8.7.

[12]This was a famous open problem in 1799 when Gauss solved it at the age of 22 in his Ph.D. dissertation.

Symmetric Matrices[13]

On the other hand, many of the applications of linear algebra involve a real matrix A and, while A will have complex eigenvalues by the fundamental theorem of algebra, it is always of interest to know when the eigenvalues are, in fact, real. While this can happen in a variety of ways, it turns out to hold whenever A is symmetric. This important theorem will be used extensively later. Surprisingly, the theory of *complex* eigenvalues can be used to prove this useful result about *real* eigenvalues.

Let $\overline{z}$ denote the conjugate of a complex number z. If A is a complex matrix, the **conjugate matrix** $\overline{A}$ is defined to be the matrix obtained from A by conjugating every entry. Thus, if $A = \left[z_{ij}\right]$, then $\overline{A} = \left[\overline{z}_{ij}\right]$. For example,

$$\text{If } A = \left[\begin{array}{cc} -i+2 & 5 \\ i & 3+4i \end{array}\right] \text{ then } \overline{A} = \left[\begin{array}{cc} i+2 & 5 \\ -i & 3-4i \end{array}\right]$$

Recall that $\overline{z+w} = \overline{z} + \overline{w}$ and $\overline{zw} = \overline{z}\,\overline{w}$ hold for all complex numbers z and w. It follows that if A and B are two complex matrices, then

$$\overline{A+B} = \overline{A} + \overline{B}, \quad \overline{AB} = \overline{A}\,\overline{B} \quad \text{and } \overline{\lambda A} = \overline{\lambda}\,\overline{A}$$

hold for all complex scalars λ. These facts are used in the proof of the following theorem.

Theorem 5.5.7

Let A be a symmetric real matrix. If λ is any complex eigenvalue of A, then λ is real.[14]

Proof. Observe that $\overline{A} = A$ because A is real. If λ is an eigenvalue of A, we show that λ is real by showing that $\overline{\lambda} = \lambda$. Let $\mathbf{x}$ be a (possibly complex) eigenvector corresponding to λ, so that $\mathbf{x} \neq \mathbf{0}$ and $A\mathbf{x} = \lambda\mathbf{x}$. Define $c = \mathbf{x}^T\overline{\mathbf{x}}$.

If we write $\mathbf{x} = \left[\begin{array}{c} z_1 \\ z_2 \\ \vdots \\ z_n \end{array}\right]$ where the z_i are complex numbers, we have

$$c = \mathbf{x}^T\overline{\mathbf{x}} = z_1\overline{z_1} + z_2\overline{z_2} + \cdots + z_n\overline{z_n} = |\overline{z_1}|^2 + |\overline{z_2}|^2 + \cdots + |\overline{z_n}|^2$$

Thus c is a real number, and $c > 0$ because at least one of the $z_i \neq 0$ (as $\mathbf{x} \neq \mathbf{0}$). We show that $\overline{\lambda} = \lambda$ by verifying that $\lambda c = \overline{\lambda} c$. We have

$$\lambda c = \lambda(\mathbf{x}^T\overline{\mathbf{x}}) = (\lambda\mathbf{x})^T\overline{\mathbf{x}} = (A\mathbf{x})^T\overline{\mathbf{x}} = \mathbf{x}^T A^T\overline{\mathbf{x}}$$

At this point we use the hypothesis that A is symmetric and real. This means $A^T = A = \overline{A}$ so we continue the calculation:

[13]This discussion uses complex conjugation and absolute value. These topics are discussed in Appendix A.

[14]This theorem was first proved in 1829 by the great French mathematician Augustin Louis Cauchy (1789–1857).

$$\begin{aligned}\lambda c = \mathbf{x}^T A^T \overline{\mathbf{x}} = \mathbf{x}^T (\overline{A}\,\overline{\mathbf{x}}) = \mathbf{x}^T (\overline{A\mathbf{x}}) &= \mathbf{x}^T (\overline{\lambda \mathbf{x}}) \\ &= \mathbf{x}^T (\overline{\lambda}\,\overline{\mathbf{x}}) \\ &= \overline{\lambda}\mathbf{x}^T \overline{\mathbf{x}} \\ &= \overline{\lambda} c\end{aligned}$$

as required. □

The technique in the proof of Theorem 5.5.7 will be used again when we return to complex linear algebra in Section 8.7.

Example 5.5.7

Verify Theorem 5.5.7 for every real, symmetric 2×2 matrix A.

Solution. If $A = \left[\begin{array}{cc} a & b \\ b & c \end{array}\right]$ we have $c_A(x) = x^2 - (a+c)x + (ac - b^2)$, so the eigenvalues are given by $\lambda = \frac{1}{2}[(a+c) \pm \sqrt{(a+c)^2 - 4(ac-b^2)}]$. But here

$$(a+c)^2 - 4(ac-b^2) = (a-c)^2 + 4b^2 \geq 0$$

for any choice of a, b, and c. Hence, the eigenvalues are real numbers.

Exercises for 5.5

Exercise 5.5.1 By computing the trace, determinant, and rank, show that A and B are *not* similar in each case.

a. $A = \left[\begin{array}{rr} 1 & 2 \\ 2 & 1 \end{array}\right]$, $B = \left[\begin{array}{rr} 1 & 1 \\ -1 & 1 \end{array}\right]$

b. $A = \left[\begin{array}{rr} 3 & 1 \\ 2 & -1 \end{array}\right]$, $B = \left[\begin{array}{rr} 1 & 1 \\ 2 & 1 \end{array}\right]$

c. $A = \left[\begin{array}{rr} 2 & 1 \\ 1 & -1 \end{array}\right]$, $B = \left[\begin{array}{rr} 3 & 0 \\ 1 & -1 \end{array}\right]$

d. $A = \left[\begin{array}{rr} 3 & 1 \\ -1 & 2 \end{array}\right]$, $B = \left[\begin{array}{rr} 2 & -1 \\ 3 & 2 \end{array}\right]$

e. $A = \left[\begin{array}{rrr} 2 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{array}\right]$, $B = \left[\begin{array}{rrr} 1 & -2 & 1 \\ -2 & 4 & -2 \\ -3 & 6 & -3 \end{array}\right]$

f. $A = \left[\begin{array}{rrr} 1 & 2 & -3 \\ 1 & -1 & 2 \\ 0 & 3 & -5 \end{array}\right]$, $B = \left[\begin{array}{rrr} -2 & 1 & 3 \\ 6 & -3 & -9 \\ 0 & 0 & 0 \end{array}\right]$

Exercise 5.5.2 Show that $\left[\begin{array}{rrrr} 1 & 2 & -1 & 0 \\ 2 & 0 & 1 & 1 \\ 1 & 1 & 0 & -1 \\ 4 & 3 & 0 & 0 \end{array}\right]$ and $\left[\begin{array}{rrrr} 1 & -1 & 3 & 0 \\ -1 & 0 & 1 & 1 \\ 0 & -1 & 4 & 1 \\ 5 & -1 & -1 & -4 \end{array}\right]$ are *not* similar.

Exercise 5.5.3 If $A \sim B$, show that:

a. $A^T \sim B^T$

b. $A^{-1} \sim B^{-1}$

c. $rA \sim rB$ for r in $\mathbb{R}$

d. $A^n \sim B^n$ for $n \geq 1$

Exercise 5.5.4 In each case, decide whether the matrix A is diagonalizable. If so, find P such that $P^{-1}AP$ is diagonal.

a. $\left[\begin{array}{lll} 1 & 0 & 0 \\ 1 & 2 & 1 \\ 0 & 0 & 1 \end{array}\right]$

b. $\left[\begin{array}{rrr} 3 & 0 & 6 \\ 0 & -3 & 0 \\ 5 & 0 & 2 \end{array}\right]$

c. $\left[\begin{array}{rrr} 3 & 1 & 6 \\ 2 & 1 & 0 \\ -1 & 0 & -3 \end{array}\right]$

d. $\left[\begin{array}{lll} 4 & 0 & 0 \\ 0 & 2 & 2 \\ 2 & 3 & 1 \end{array}\right]$

Exercise 5.5.5 If A is invertible, show that AB is similar to BA for all B.

Exercise 5.5.6 Show that the only matrix similar to a scalar matrix $A = rI$, r in $\mathbb{R}$, is A itself.

Exercise 5.5.7 Let λ be an eigenvalue of A with corresponding eigenvector $\mathbf{x}$. If $B = P^{-1}AP$ is similar to A, show that $P^{-1}\mathbf{x}$ is an eigenvector of B corresponding to λ.

Exercise 5.5.8 If $A \sim B$ and A has any of the following properties, show that B has the same property.

a. Idempotent, that is $A^2 = A$.

b. Nilpotent, that is $A^k = 0$ for some $k \geq 1$.

c. Invertible.

Exercise 5.5.9 Let A denote an $n \times n$ upper triangular matrix.

a. If all the main diagonal entries of A are distinct, show that A is diagonalizable.

b. If all the main diagonal entries of A are equal, show that A is diagonalizable only if it is *already* diagonal.

c. Show that $\left[\begin{array}{lll} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 2 \end{array}\right]$ is diagonalizable but that $\left[\begin{array}{lll} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 2 \end{array}\right]$ is not diagonalizable.

Exercise 5.5.10 Let A be a diagonalizable $n \times n$ matrix with eigenvalues $\lambda_1, \lambda_2, \dots, \lambda_n$ (including multiplicities). Show that:

a. $\det A = \lambda_1 \lambda_2 \cdots \lambda_n$

b. $\operatorname{tr} A = \lambda_1 + \lambda_2 + \cdots + \lambda_n$

Exercise 5.5.11 Given a polynomial $p(x) = r_0 + r_1 x + \cdots + r_n x^n$ and a square matrix A, the matrix $p(A) = r_0 I + r_1 A + \cdots + r_n A^n$ is called the **evaluation** of $p(x)$ at A. Let $B = P^{-1}AP$. Show that $p(B) = P^{-1}p(A)P$ for all polynomials $p(x)$.

Exercise 5.5.12 Let P be an invertible $n \times n$ matrix. If A is any $n \times n$ matrix, write $T_P(A) = P^{-1}AP$. Verify that:

a. $T_P(I) = I$

b. $T_P(AB) = T_P(A)T_P(B)$

c. $T_P(A+B) = T_P(A) + T_P(B)$

d. $T_P(rA) = rT_P(A)$

e. $T_P(A^k) = [T_P(A)]^k$ for $k \geq 1$

f. If A is invertible, $T_P(A^{-1}) = [T_P(A)]^{-1}$.

g. If Q is invertible, $T_Q[T_P(A)] = T_{PQ}(A)$.

Exercise 5.5.13

a. Show that two diagonalizable matrices are similar if and only if they have the same eigenvalues with the same multiplicities.

b. If A is diagonalizable, show that $A \sim A^T$.

c. Show that $A \sim A^T$ if $A = \left[\begin{array}{ll} 1 & 1 \\ 0 & 1 \end{array}\right]$

Exercise 5.5.14 If A is 2×2 and diagonalizable, show that $C(A) = \{X \mid XA = AX\}$ has dimension 2 or 4. [*Hint*: If $P^{-1}AP = D$, show that X is in $C(A)$ if and only if $P^{-1}XP$ is in $C(D)$.]

Exercise 5.5.15 If A is diagonalizable and $p(x)$ is a polynomial such that $p(\lambda) = 0$ for all eigenvalues λ of A, show that $p(A) = 0$ (see Example 3.3.9). In particular, show $c_A(A) = 0$. [*Remark*: $c_A(A) = 0$ for *all* square matrices A—this is the Cayley-Hamilton theorem, see Theorem 11.1.2.]

Exercise 5.5.16 Let A be $n \times n$ with n distinct real eigenvalues. If $AC = CA$, show that C is diagonalizable.

Exercise 5.5.17 Let $A = \begin{bmatrix} 0 & a & b \\ a & 0 & c \\ b & c & 0 \end{bmatrix}$ and $B = \begin{bmatrix} c & a & b \\ a & b & c \\ b & c & a \end{bmatrix}$.

a. Show that $x^3 - (a^2+b^2+c^2)x - 2abc$ has real roots by considering A.

b. Show that $a^2+b^2+c^2 \geq ab+ac+bc$ by considering B.

Exercise 5.5.18 Assume the 2×2 matrix A is similar to an upper triangular matrix. If $\text{tr } A = 0 = \text{tr } A^2$, show that $A^2 = 0$.

Exercise 5.5.19 Show that A is similar to A^T for all 2×2 matrices A. [*Hint*: Let $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$. If $c = 0$ treat the cases $b = 0$ and $b \neq 0$ separately. If $c \neq 0$, reduce to the case $c = 1$ using Exercise 5.5.12(d).]

Exercise 5.5.20 Refer to Section 3.4 on linear recurrences. Assume that the sequence $x_0, x_1, x_2, \dots$ satisfies

$$x_{n+k} = r_0 x_n + r_1 x_{n+1} + \cdots + r_{k-1} x_{n+k-1}$$

for all $n \geq 0$. Define

$$A = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ r_0 & r_1 & r_2 & \cdots & r_{k-1} \end{bmatrix}, \; V_n = \begin{bmatrix} x_n \\ x_{n+1} \\ \vdots \\ x_{n+k-1} \end{bmatrix}.$$

Then show that:

a. $V_n = A^n V_0$ for all n.

b. $c_A(x) = x^k - r_{k-1}x^{k-1} - \cdots - r_1 x - r_0$

c. If λ is an eigenvalue of A, the eigenspace E_λ has dimension 1, and $\mathbf{x} = (1, \lambda, \lambda^2, \dots, \lambda^{k-1})^T$ is an eigenvector. [*Hint*: Use $c_A(\lambda) = 0$ to show that $E_\lambda = \mathbb{R}\mathbf{x}$.]

d. A is diagonalizable if and only if the eigenvalues of A are distinct. [*Hint*: See part (c) and Theorem 5.5.4.]

e. If $\lambda_1, \lambda_2, \dots, \lambda_k$ are distinct real eigenvalues, there exist constants $t_1, t_2, \dots, t_k$ such that $x_n = t_1\lambda_1^n + \cdots + t_k\lambda_k^n$ holds for all n. [*Hint*: If D is diagonal with $\lambda_1, \lambda_2, \dots, \lambda_k$ as the main diagonal entries, show that $A^n = PD^nP^{-1}$ has entries that are linear combinations of $\lambda_1^n, \lambda_2^n, \dots, \lambda_k^n$.]

Exercise 5.5.21 Suppose A is 2×2 and $A^2 = 0$. If $\text{tr } A \neq 0$ show that $A = 0$.

5.6 Best Approximation and Least Squares

Often an exact solution to a problem in applied mathematics is difficult to obtain. However, it is usually just as useful to find arbitrarily close approximations to a solution. In particular, finding “linear approximations” is a potent technique in applied mathematics. One basic case is the situation where a system of linear equations has no solution, and it is desirable to find a “best approximation” to a solution to the system. In this section best approximations are defined and a method for finding them is described. The result is then applied to “least squares” approximation of data.

Suppose A is an $m \times n$ matrix and $\mathbf{b}$ is a column in $\mathbb{R}^m$, and consider the system

$$A\mathbf{x} = \mathbf{b}$$

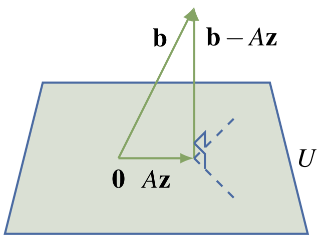
of m linear equations in n variables. This need not have a solution. However, given any column $\mathbf{z} \in \mathbb{R}^n$, the distance $\|\mathbf{b} - A\mathbf{z}\|$ is a measure of how far $A\mathbf{z}$ is from $\mathbf{b}$. Hence it is natural to ask whether there is a column $\mathbf{z}$ in $\mathbb{R}^n$ that is as close as possible to a solution in the sense that

$$\|\mathbf{b} - A\mathbf{z}\|$$

is the minimum value of $\|\mathbf{b} - A\mathbf{x}\|$ as $\mathbf{x}$ ranges over all columns in $\mathbb{R}^n$.

The answer is "yes", and to describe it define

$$U = \{A\mathbf{x} \mid \mathbf{x} \text{ lies in } \mathbb{R}^n\}$$



This is a subspace of $\mathbb{R}^n$ (verify) and we want a vector $A\mathbf{z}$ in U as close as possible to $\mathbf{b}$. That there is such a vector is clear geometrically if $n = 3$ by the diagram. In general such a vector $A\mathbf{z}$ exists by a general result called the *projection theorem* that will be proved in Chapter 8 (Theorem 8.1.3). Moreover, the projection theorem gives a simple way to compute $\mathbf{z}$ because it also shows that the vector $\mathbf{b} - A\mathbf{z}$ is *orthogonal* to every vector $A\mathbf{x}$ in U. Thus, for all $\mathbf{x}$ in $\mathbb{R}^n$,

$$\begin{aligned} 0 = (A\mathbf{x}) \cdot (\mathbf{b} - A\mathbf{z}) = (A\mathbf{x})^T (\mathbf{b} - A\mathbf{z}) &= \mathbf{x}^T A^T (\mathbf{b} - A\mathbf{z}) \\ &= \mathbf{x} \cdot [A^T (\mathbf{b} - A\mathbf{z})] \end{aligned}$$

In other words, the vector $A^T(\mathbf{b} - A\mathbf{z})$ in $\mathbb{R}^n$ is orthogonal to *every* vector in $\mathbb{R}^n$ and so must be zero (being orthogonal to itself). Hence $\mathbf{z}$ satisfies

$$(A^T A)\mathbf{z} = A^T \mathbf{b}$$

Definition 5.14 Normal Equations

This is a system of linear equations called the ***normal equations*** *for* $\mathbf{z}$.

Note that this system can have more than one solution (see Exercise 5.6.5). However, the $n \times n$ matrix $A^T A$ is invertible if (and only if) the columns of A are linearly independent (Theorem 5.4.3); so, in this case, $\mathbf{z}$ is uniquely determined and is given explicitly by $\mathbf{z} = (A^T A)^{-1} A^T \mathbf{b}$. However, the most efficient way to find $\mathbf{z}$ is to apply gaussian elimination to the normal equations.

This discussion is summarized in the following theorem.

Theorem 5.6.1: Best Approximation Theorem

Let A be an $m \times n$ matrix, let $\mathbf{b}$ be any column in $\mathbb{R}^m$, and consider the system

$$A\mathbf{x} = \mathbf{b}$$

of m equations in n variables.

1. *Any solution $\mathbf{z}$ to the normal equations*

$$(A^T A)\mathbf{z} = A^T \mathbf{b}$$

   is a best approximation to a solution to $A\mathbf{x} = \mathbf{b}$ in the sense that $\|\mathbf{b} - A\mathbf{z}\|$ is the minimum value of $\|\mathbf{b} - A\mathbf{x}\|$ as $\mathbf{x}$ ranges over all columns in $\mathbb{R}^n$.

2. *If the columns of A are linearly independent, then $A^T A$ is invertible and $\mathbf{z}$ is given uniquely by $\mathbf{z} = (A^T A)^{-1} A^T \mathbf{b}$.*

We note in passing that if A is $n \times n$ and invertible, then

$$\mathbf{z} = (A^TA)^{-1}A^T\mathbf{b} = A^{-1}\mathbf{b}$$

is the solution to the system of equations, and $\|\mathbf{b} - A\mathbf{z}\| = 0$. Hence if A has independent columns, then $(A^TA)^{-1}A^T$ is playing the role of the inverse of the nonsquare matrix A. The matrix $A^T(AA^T)^{-1}$ plays a similar role when the rows of A are linearly independent. These are both special cases of the **generalized inverse** of a matrix A (see Exercise 5.6.14). However, we shall not pursue this topic here.

Example 5.6.1

The system of linear equations

$$\begin{aligned} 3x - y &= 4 \\ x + 2y &= 0 \\ 2x + y &= 1 \end{aligned}$$

has no solution. Find the vector $\mathbf{z} = \left[\begin{array}{c} x_0 \\ y_0 \end{array} \right]$ that best approximates a solution.

Solution. In this case,

$$A = \left[\begin{array}{rr} 3 & -1 \\ 1 & 2 \\ 2 & 1 \end{array} \right], \text{ so } A^TA = \left[\begin{array}{rrr} 3 & 1 & 2 \\ -1 & 2 & 1 \end{array} \right] \left[\begin{array}{rr} 3 & -1 \\ 1 & 2 \\ 2 & 1 \end{array} \right] = \left[\begin{array}{rr} 14 & 1 \\ 1 & 6 \end{array} \right]$$

is invertible. The normal equations $(A^TA)\mathbf{z} = A^T\mathbf{b}$ are

$$\left[\begin{array}{rr} 14 & 1 \\ 1 & 6 \end{array} \right] \mathbf{z} = \left[\begin{array}{r} 14 \\ -3 \end{array} \right], \text{ so } \mathbf{z} = \frac{1}{83} \left[\begin{array}{r} 87 \\ -56 \end{array} \right]$$

Thus $x_0 = \frac{87}{83}$ and $y_0 = \frac{-56}{83}$. With these values of x and y, the left sides of the equations are, approximately,

$$\begin{aligned} 3x_0 - y_0 &= \tfrac{317}{83} = 3.82 \\ x_0 + 2y_0 &= \tfrac{-25}{83} = -0.30 \\ 2x_0 + y_0 &= \tfrac{118}{83} = 1.42 \end{aligned}$$

This is as close as possible to a solution.

Example 5.6.2

The average number g of goals per game scored by a hockey player seems to be related linearly to two factors: the number x_1 of years of experience and the number x_2 of goals in the preceding 10 games. The data on the following page were collected on four players. Find the linear function

$g = a_0 + a_1x_1 + a_2x_2$ that best fits these data.

g	x_1	x_2
0.8	5	3
0.8	3	4
0.6	1	5
0.4	2	1

Solution. If the relationship is given by $g = r_0 + r_1x_1 + r_2x_2$, then the data can be described as follows:

$$\begin{bmatrix} 1 & 5 & 3 \\ 1 & 3 & 4 \\ 1 & 1 & 5 \\ 1 & 2 & 1 \end{bmatrix} \begin{bmatrix} r_0 \\ r_1 \\ r_2 \end{bmatrix} = \begin{bmatrix} 0.8 \\ 0.8 \\ 0.6 \\ 0.4 \end{bmatrix}$$

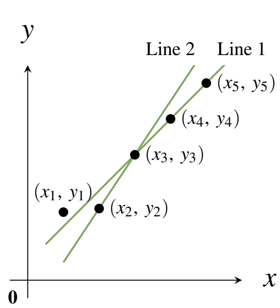
Using the notation in Theorem 5.6.1, we get

$$\begin{aligned} \mathbf{z} &= (A^TA)^{-1}A^T\mathbf{b} \\ &= \frac{1}{42}\begin{bmatrix} 119 & -17 & -19 \\ -17 & 5 & 1 \\ -19 & 1 & 5 \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 & 1 \\ 5 & 3 & 1 & 2 \\ 3 & 4 & 5 & 1 \end{bmatrix} \begin{bmatrix} 0.8 \\ 0.8 \\ 0.6 \\ 0.4 \end{bmatrix} = \begin{bmatrix} 0.14 \\ 0.09 \\ 0.08 \end{bmatrix} \end{aligned}$$

Hence the best-fitting function is $g = 0.14 + 0.09x_1 + 0.08x_2$. The amount of computation would have been reduced if the normal equations had been constructed and then solved by gaussian elimination.

Least Squares Approximation

In many scientific investigations, data are collected that relate two variables. For example, if x is the number of dollars spent on advertising by a manufacturer and y is the value of sales in the region in question, the manufacturer could generate data by spending $x_1, x_2, \dots, x_n$ dollars at different times and measuring the corresponding sales values $y_1, y_2, \dots, y_n$.

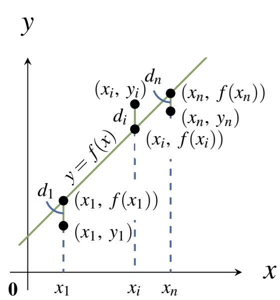


Suppose it is known that a linear relationship exists between the variables x and y—in other words, that $y = a + bx$ for some constants a and b. If the data are plotted, the points $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ may appear to lie on a straight line and estimating a and b requires finding the "best-fitting" line through these data points. For example, if five data points occur as shown in the diagram, line 1 is clearly a better fit than line 2. In general, the problem is to find the values of the constants a and b such that the line $y = a + bx$ best approximates the data in question. Note that an exact fit would be obtained if a and b were such that $y_i = a + bx_i$ were true for each data point (x_i, y_i). But this is too much to expect. Experimental errors in measurement are bound to occur, so the choice of a and b should be made in such a way that the errors between the observed values y_i and the corresponding fitted values $a+bx_i$ are in some sense minimized. Least squares approximation is a way to do this.

The first thing we must do is explain exactly what we mean by the *best fit* of a line $y = a+bx$ to an observed set of data points $(x_1,\ y_1),\ (x_2,\ y_2),\ \dots,\ (x_n,\ y_n)$. For convenience, write the linear function r_0+r_1x as

$$f(x) = r_0 + r_1 x$$

so that the fitted points (on the line) have coordinates $(x_1,\ f(x_1)),\ \dots,\ (x_n,\ f(x_n))$.



The second diagram is a sketch of what the line $y = f(x)$ might look like. For each i the observed data point $(x_i,\ y_i)$ and the fitted point $(x_i,\ f(x_i))$ need not be the same, and the distance d_i between them measures how far the line misses the observed point. For this reason d_i is often called the **error** at x_i, and a natural measure of how close the line $y = f(x)$ is to the observed data points is the sum $d_1+d_2+\cdots+d_n$ of all these errors. However, it turns out to be better to use the sum of squares

$$S = d_1^2 + d_2^2 + \cdots + d_n^2$$

as the measure of error, and the line $y = f(x)$ is to be chosen so as to make this sum as small as possible. This line is said to be the **least squares approximating line** for the data points $(x_1,\ y_1),\ (x_2,\ y_2),\ \dots,\ (x_n,\ y_n)$.

The square of the error d_i is given by $d_i^2 = [y_i - f(x_i)]^2$ for each i, so the quantity S to be minimized is the sum:

$$S = [y_1 - f(x_1)]^2 + [y_2 - f(x_2)]^2 + \cdots + [y_n - f(x_n)]^2$$

Note that all the numbers x_i and y_i are *given* here; what is required is that the *function* f be chosen in such a way as to minimize S. Because $f(x) = r_0 + r_1 x$, this amounts to choosing r_0 and r_1 to minimize S. This problem can be solved using Theorem 5.6.1. The following notation is convenient.

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \quad \mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \quad \text{and} \quad f(\mathbf{x}) = \begin{bmatrix} f(x_1) \\ f(x_2) \\ \vdots \\ f(x_n) \end{bmatrix} = \begin{bmatrix} r_0 + r_1 x_1 \\ r_0 + r_1 x_2 \\ \vdots \\ r_0 + r_1 x_n \end{bmatrix}$$

Then the problem takes the following form: Choose r_0 and r_1 such that

$$S = [y_1 - f(x_1)]^2 + [y_2 - f(x_2)]^2 + \cdots + [y_n - f(x_n)]^2 = \|\mathbf{y} - f(\mathbf{x})\|^2$$

is as small as possible. Now write

$$M = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix} \quad \text{and} \quad \mathbf{r} = \begin{bmatrix} r_0 \\ r_1 \end{bmatrix}$$

Then $M\mathbf{r} = f(\mathbf{x})$, so we are looking for a column $\mathbf{r} = \begin{bmatrix} r_0 \\ r_1 \end{bmatrix}$ such that $\|\mathbf{y} - M\mathbf{r}\|^2$ is as small as possible. In other words, we are looking for a best approximation $\mathbf{z}$ to the system $M\mathbf{r} = \mathbf{y}$. Hence Theorem 5.6.1 applies directly, and we have

Theorem 5.6.2

Suppose that n data points $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ *are given, where at least two of* $x_1, x_2, \dots, x_n$ *are distinct. Put*

$$\mathbf{y} = \left[\begin{array}{c} y_1 \\ y_2 \\ \vdots \\ y_n \end{array} \right] \quad M = \left[\begin{array}{cc} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{array} \right]$$

Then the least squares approximating line for these data points has equation

$$y = z_0 + z_1 x$$

where $\mathbf{z} = \left[\begin{array}{c} z_0 \\ z_1 \end{array} \right]$ *is found by gaussian elimination from the normal equations*

$$(M^T M)\mathbf{z} = M^T \mathbf{y}$$

The condition that at least two of $x_1, x_2, \dots, x_n$ *are distinct ensures that* $M^T M$ *is an invertible matrix, so* $\mathbf{z}$ *is unique:*

$$\mathbf{z} = (M^T M)^{-1} M^T \mathbf{y}$$

Example 5.6.3

Let data points $(x_1, y_1), (x_2, y_2), \dots, (x_5, y_5)$ be given as in the accompanying table. Find the least squares approximating line for these data.

x	y
1	1
3	2
4	3
6	4
7	5

Solution. In this case we have

$$\begin{aligned} M^T M &= \left[\begin{array}{cccc} 1 & 1 & \cdots & 1 \\ x_1 & x_2 & \cdots & x_5 \end{array} \right] \left[\begin{array}{cc} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_5 \end{array} \right] \\ &= \left[\begin{array}{cc} 5 & x_1 + \cdots + x_5 \\ x_1 + \cdots + x_5 & x_1^2 + \cdots + x_5^2 \end{array} \right] = \left[\begin{array}{cc} 5 & 21 \\ 21 & 111 \end{array} \right] \end{aligned}$$

and $$M^T\mathbf{y} = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ x_1 & x_2 & \cdots & x_5 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_5 \end{bmatrix}$$
$$= \begin{bmatrix} y_1 + y_2 + \cdots + y_5 \\ x_1y_1 + x_2y_2 + \cdots + x_5y_5 \end{bmatrix} = \begin{bmatrix} 15 \\ 78 \end{bmatrix}$$

so the normal equations $(M^TM)\mathbf{z} = M^T\mathbf{y}$ for $\mathbf{z} = \begin{bmatrix} z_0 \\ z_1 \end{bmatrix}$ become

$$\begin{bmatrix} 5 & 21 \\ 21 & 111 \end{bmatrix} = \begin{bmatrix} z_0 \\ z_1 \end{bmatrix} = \begin{bmatrix} 15 \\ 78 \end{bmatrix}$$

The solution (using gaussian elimination) is $\mathbf{z} = \begin{bmatrix} z_0 \\ z_1 \end{bmatrix} = \begin{bmatrix} 0.24 \\ 0.66 \end{bmatrix}$ to two decimal places, so the least squares approximating line for these data is $y = 0.24 + 0.66x$. Note that M^TM is indeed invertible here (the determinant is 114), and the exact solution is

$$\mathbf{z} = (M^TM)^{-1}M^T\mathbf{y} = \tfrac{1}{114}\begin{bmatrix} 111 & -21 \\ -21 & 5 \end{bmatrix}\begin{bmatrix} 15 \\ 78 \end{bmatrix} = \tfrac{1}{114}\begin{bmatrix} 27 \\ 75 \end{bmatrix} = \tfrac{1}{38}\begin{bmatrix} 9 \\ 25 \end{bmatrix}$$

Least Squares Approximating Polynomials

Suppose now that, rather than a straight line, we want to find a polynomial

$$y = f(x) = r_0 + r_1x + r_2x^2 + \cdots + r_mx^m$$

of degree m that best approximates the data pairs $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$. As before, write

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \quad \mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \quad \text{and} \quad f(\mathbf{x}) = \begin{bmatrix} f(x_1) \\ f(x_2) \\ \vdots \\ f(x_n) \end{bmatrix}$$

For each x_i we have two values of the variable y, the observed value y_i, and the computed value $f(x_i)$. The problem is to choose $f(x)$—that is, choose $r_0, r_1, \ldots, r_m$ —such that the $f(x_i)$ are as close as possible to the y_i. Again we define “as close as possible” by the least squares condition: We choose the r_i such that

$$\|\mathbf{y} - f(\mathbf{x})\|^2 = [y_1 - f(x_1)]^2 + [y_2 - f(x_2)]^2 + \cdots + [y_n - f(x_n)]^2$$

is as small as possible.

Definition 5.15 Least Squares Approximation

*A polynomial $f(x)$ satisfying this condition is called a **least squares approximating polynomial** of degree m for the given data pairs.*

If we write

$$M = \begin{bmatrix} 1 & x_1 & x_1^2 & \cdots & x_1^m \\ 1 & x_2 & x_2^2 & \cdots & x_2^m \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_n & x_n^2 & \cdots & x_n^m \end{bmatrix} \quad \text{and} \quad \mathbf{r} = \begin{bmatrix} r_0 \\ r_1 \\ \vdots \\ r_m \end{bmatrix}$$

we see that $f(\mathbf{x}) = M\mathbf{r}$. Hence we want to find $\mathbf{r}$ such that $\|\mathbf{y} - M\mathbf{r}\|^2$ is as small as possible; that is, we want a best approximation $\mathbf{z}$ to the system $M\mathbf{r} = \mathbf{y}$. Theorem 5.6.1 gives the first part of Theorem 5.6.3.

Theorem 5.6.3

Let n data pairs $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ be given, and write

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \quad M = \begin{bmatrix} 1 & x_1 & x_1^2 & \cdots & x_1^m \\ 1 & x_2 & x_2^2 & \cdots & x_2^m \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_n & x_n^2 & \cdots & x_n^m \end{bmatrix} \quad \mathbf{z} = \begin{bmatrix} z_0 \\ z_1 \\ \vdots \\ z_m \end{bmatrix}$$

1. *If $\mathbf{z}$ is any solution to the normal equations*

$$(M^T M)\mathbf{z} = M^T \mathbf{y}$$

   then the polynomial

$$z_0 + z_1 x + z_2 x^2 + \cdots + z_m x^m$$

   is a least squares approximating polynomial of degree m for the given data pairs.

2. *If at least $m+1$ of the numbers $x_1, x_2, \dots, x_n$ are distinct (so $n \geq m+1$), the matrix $M^T M$ is invertible and $\mathbf{z}$ is uniquely determined by*

$$\mathbf{z} = (M^T M)^{-1} M^T \mathbf{y}$$

Proof. It remains to prove (2), and for that we show that the columns of M are linearly independent (Theorem 5.4.3). Suppose a linear combination of the columns vanishes:

$$r_0 \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} + r_1 \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} + \cdots + r_m \begin{bmatrix} x_1^m \\ x_2^m \\ \vdots \\ x_n^m \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

If we write $q(x) = r_0 + r_1x + \cdots + r_mx^m$, equating coefficients shows that

$$q(x_1) = q(x_2) = \cdots = q(x_n) = 0$$

Hence $q(x)$ is a polynomial of degree m with at least $m+1$ distinct roots, so $q(x)$ must be the zero polynomial (see Appendix D or Theorem 6.5.4). Thus $r_0 = r_1 = \cdots = r_m = 0$ as required. □

Example 5.6.4

Find the least squares approximating quadratic $y = z_0 + z_1x + z_2x^2$ for the following data points.

$$(-3,\ 3),\ (-1,\ 1),\ (0,\ 1),\ (1,\ 2),\ (3,\ 4)$$

Solution. This is an instance of Theorem 5.6.3 with $m = 2$. Here

$$\mathbf{y} = \begin{bmatrix} 3 \\ 1 \\ 1 \\ 2 \\ 4 \end{bmatrix} \quad M = \begin{bmatrix} 1 & -3 & 9 \\ 1 & -1 & 1 \\ 1 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & 3 & 9 \end{bmatrix}$$

Hence,

$$M^TM = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ -3 & -1 & 0 & 1 & 3 \\ 9 & 1 & 0 & 1 & 9 \end{bmatrix} \begin{bmatrix} 1 & -3 & 9 \\ 1 & -1 & 1 \\ 1 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & 3 & 9 \end{bmatrix} = \begin{bmatrix} 5 & 0 & 20 \\ 0 & 20 & 0 \\ 20 & 0 & 164 \end{bmatrix}$$

$$M^T\mathbf{y} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ -3 & -1 & 0 & 1 & 3 \\ 9 & 1 & 0 & 1 & 9 \end{bmatrix} \begin{bmatrix} 3 \\ 1 \\ 1 \\ 2 \\ 4 \end{bmatrix} = \begin{bmatrix} 11 \\ 4 \\ 66 \end{bmatrix}$$

The normal equations for $\mathbf{z}$ are

$$\begin{bmatrix} 5 & 0 & 20 \\ 0 & 20 & 0 \\ 20 & 0 & 164 \end{bmatrix} \mathbf{z} = \begin{bmatrix} 11 \\ 4 \\ 66 \end{bmatrix} \quad \text{whence } \mathbf{z} = \begin{bmatrix} 1.15 \\ 0.20 \\ 0.26 \end{bmatrix}$$

This means that the least squares approximating quadratic for these data is
$y = 1.15 + 0.20x + 0.26x^2$.

Other Functions

There is an extension of Theorem 5.6.3 that should be mentioned. Given data pairs (x_1, y_1), (x_2, y_2), $\dots$, (x_n, y_n), that theorem shows how to find a polynomial

$$f(x) = r_0 + r_1 x + \cdots + r_m x^m$$

such that $\|\mathbf{y} - f(\mathbf{x})\|^2$ is as small as possible, where $\mathbf{x}$ and $f(\mathbf{x})$ are as before. Choosing the appropriate polynomial $f(x)$ amounts to choosing the coefficients $r_0, r_1, \dots, r_m$, and Theorem 5.6.3 gives a formula for the optimal choices. Here $f(x)$ is a linear combination of the functions $1, x, x^2, \dots, x^m$ where the r_i are the coefficients, and this suggests applying the method to other functions. If $f_0(x), f_1(x), \dots, f_m(x)$ are given functions, write

$$f(x) = r_0 f_0(x) + r_1 f_1(x) + \cdots + r_m f_m(x)$$

where the r_i are real numbers. Then the more general question is whether $r_0, r_1, \dots, r_m$ can be found such that $\|\mathbf{y} - f(\mathbf{x})\|^2$ is as small as possible where

$$f(\mathbf{x}) = \left[\begin{array}{c} f(x_1) \\ f(x_2) \\ \vdots \\ f(x_m) \end{array} \right]$$

Such a function $f(\mathbf{x})$ is called a **least squares best approximation** for these data pairs of the form $r_0 f_0(x) + r_1 f_1(x) + \cdots + r_m f_m(x)$, r_i in $\mathbb{R}$. The proof of Theorem 5.6.3 goes through to prove

Theorem 5.6.4

Let n data pairs (x_1, y_1), (x_2, y_2), $\dots$, (x_n, y_n) be given, and suppose that $m+1$ functions $f_0(x), f_1(x), \dots, f_m(x)$ are specified. Write

$$\mathbf{y} = \left[\begin{array}{c} y_1 \\ y_2 \\ \vdots \\ y_n \end{array} \right] \quad M = \left[\begin{array}{cccc} f_0(x_1) & f_1(x_1) & \cdots & f_m(x_1) \\ f_0(x_2) & f_1(x_2) & \cdots & f_m(x_2) \\ \vdots & \vdots & & \vdots \\ f_0(x_n) & f_1(x_n) & \cdots & f_m(x_n) \end{array} \right] \quad \mathbf{z} = \left[\begin{array}{c} z_1 \\ z_2 \\ \vdots \\ z_m \end{array} \right]$$

1. *If $\mathbf{z}$ is any solution to the normal equations*

$$(M^T M)\mathbf{z} = M^T \mathbf{y}$$

then the function

$$z_0 f_0(x) + z_1 f_1(x) + \cdots + z_m f_m(x)$$

is the best approximation for these data among all functions of the form $r_0 f_0(x) + r_1 f_1(x) + \cdots + r_m f_m(x)$ where the r_i are in $\mathbb{R}$.

2. *If $M^T M$ is invertible (that is, if $\operatorname{rank}(M) = m+1$), then $\mathbf{z}$ is uniquely determined; in fact, $\mathbf{z} = (M^T M)^{-1}(M^T \mathbf{y})$.*

Clearly Theorem 5.6.4 contains Theorem 5.6.3 as a special case, but there is no simple test in general for whether M^TM is invertible. Conditions for this to hold depend on the choice of the functions $f_0(x), f_1(x), \dots, f_m(x)$.

Example 5.6.5

Given the data pairs $(-1, 0)$, $(0, 1)$, and $(1, 4)$, find the least squares approximating function of the form $r_0x + r_1 2^x$.

Solution. The functions are $f_0(x) = x$ and $f_1(x) = 2^x$, so the matrix M is

$$M = \begin{bmatrix} f_0(x_1) & f_1(x_1) \\ f_0(x_2) & f_1(x_2) \\ f_0(x_3) & f_1(x_3) \end{bmatrix} = \begin{bmatrix} -1 & 2^{-1} \\ 0 & 2^0 \\ 1 & 2^1 \end{bmatrix} = \frac{1}{2}\begin{bmatrix} -2 & 1 \\ 0 & 2 \\ 2 & 4 \end{bmatrix}$$

In this case $M^TM = \frac{1}{4}\begin{bmatrix} 8 & 6 \\ 6 & 21 \end{bmatrix}$ is invertible, so the normal equations

$$\frac{1}{4}\begin{bmatrix} 8 & 6 \\ 6 & 21 \end{bmatrix}\mathbf{z} = \begin{bmatrix} 4 \\ 9 \end{bmatrix}$$

have a unique solution $\mathbf{z} = \frac{1}{11}\begin{bmatrix} 10 \\ 16 \end{bmatrix}$. Hence the best-fitting function of the form $r_0x + r_1 2^x$ is $\overline{f}(x) = \frac{10}{11}x + \frac{16}{11}2^x$. Note that $\overline{f}(\mathbf{x}) = \begin{bmatrix} \overline{f}(-1) \\ \overline{f}(0) \\ \overline{f}(1) \end{bmatrix} = \begin{bmatrix} \frac{-2}{11} \\ \frac{16}{11} \\ \frac{42}{11} \end{bmatrix}$, compared with $\mathbf{y} = \begin{bmatrix} 0 \\ 1 \\ 4 \end{bmatrix}$

Exercises for 5.6

Exercise 5.6.1 Find the best approximation to a solution of each of the following systems of equations.

a.
$$\begin{aligned} x + y - z &= 5 \\ 2x - y + 6z &= 1 \\ 3x + 2y - z &= 6 \\ -x + 4y + z &= 0 \end{aligned}$$

b.
$$\begin{aligned} 3x + y + z &= 6 \\ 2x + 3y - z &= 1 \\ 2x - y + z &= 0 \\ 3x - 3y + 3z &= 8 \end{aligned}$$

Exercise 5.6.2 Find the least squares approximating line $y = z_0 + z_1x$ for each of the following sets of data points.

a. $(1, 1), (3, 2), (4, 3), (6, 4)$

b. $(2, 4), (4, 3), (7, 2), (8, 1)$

c. $(-1, -1), (0, 1), (1, 2), (2, 4), (3, 6)$

d. $(-2, 3), (-1, 1), (0, 0), (1, -2), (2, -4)$

Exercise 5.6.3 Find the least squares approximating quadratic $y = z_0 + z_1x + z_2x^2$ for each of the following sets of data points.

a. $(0, 1), (2, 2), (3, 3), (4, 5)$

b. $(-2, 1), (0, 0), (3, 2), (4, 3)$

Exercise 5.6.4 Find a least squares approximating function of the form $r_0x + r_1x^2 + r_2 2^x$ for each of the following sets of data pairs.

a. $(-1, 1)$, $(0, 3)$, $(1, 1)$, $(2, 0)$

b. $(0, 1)$, $(1, 1)$, $(2, 5)$, $(3, 10)$

Exercise 5.6.5 Find the least squares approximating function of the form $r_0 + r_1x^2 + r_2 \sin \frac{\pi x}{2}$ for each of the following sets of data pairs.

a. $(0, 3)$, $(1, 0)$, $(1, -1)$, $(-1, 2)$

b. $(-1, \frac{1}{2})$, $(0, 1)$, $(2, 5)$, $(3, 9)$

Exercise 5.6.6 If M is a square invertible matrix, show that $\mathbf{z} = M^{-1}\mathbf{y}$ (in the notation of Theorem 5.6.3).

Exercise 5.6.7 Newton's laws of motion imply that an object dropped from rest at a height of 100 metres will be at a height $s = 100 - \frac{1}{2}gt^2$ metres t seconds later, where g is a constant called the acceleration due to gravity. The values of s and t given in the table are observed. Write $x = t^2$, find the least squares approximating line $s = a + bx$ for these data, and use b to estimate g.

Then find the least squares approximating quadratic $s = a_0 + a_1t + a_2t^2$ and use the value of a_2 to estimate g.

t	1	2	3
s	95	80	56

Exercise 5.6.8 A naturalist measured the heights y_i (in metres) of several spruce trees with trunk diameters x_i (in centimetres). The data are as given in the table. Find the least squares approximating line for these data and use it to estimate the height of a spruce tree with a trunk of diameter 10 cm.

x_i	5	7	8	12	13	16
y_i	2	3.3	4	7.3	7.9	10.1

Exercise 5.6.9 The yield y of wheat in bushels per acre appears to be a linear function of the number of days x_1 of sunshine, the number of inches x_2 of rain, and the number of pounds x_3 of fertilizer applied per acre. Find the best fit to the data in the table by an equation of the form $y = r_0 + r_1x_1 + r_2x_2 + r_3x_3$. [*Hint*: If a calculator for inverting A^TA is not available, the inverse is given in the answer.]

y	x_1	x_2	x_3
28	50	18	10
30	40	20	16
21	35	14	10
23	40	12	12
23	30	16	14

Exercise 5.6.10

a. Use $m = 0$ in Theorem 5.6.3 to show that the best-fitting horizontal line $y = a_0$ through the data points $(x_1, y_1), \dots, (x_n, y_n)$ is

$$y = \frac{1}{n}(y_1 + y_2 + \cdots + y_n)$$

the average of the y coordinates.

b. Deduce the conclusion in (a) without using Theorem 5.6.3.

Exercise 5.6.11 Assume $n = m+1$ in Theorem 5.6.3 (so M is square). If the x_i are distinct, use Theorem 3.2.6 to show that M is invertible. Deduce that $\mathbf{z} = M^{-1}\mathbf{y}$ and that the least squares polynomial is the interpolating polynomial (Theorem 3.2.6) and actually passes through all the data points.

Exercise 5.6.12 Let A be any $m \times n$ matrix and write $K = \{\mathbf{x} \mid A^TA\mathbf{x} = \mathbf{0}\}$. Let $\mathbf{b}$ be an m-column. Show that, if $\mathbf{z}$ is an n-column such that $\|\mathbf{b} - A\mathbf{z}\|$ is minimal, then *all* such vectors have the form $\mathbf{z} + \mathbf{x}$ for some $\mathbf{x} \in K$. [*Hint*: $\|\mathbf{b} - A\mathbf{y}\|$ is minimal if and only if $A^TA\mathbf{y} = A^T\mathbf{b}$.]

Exercise 5.6.13 Given the situation in Theorem 5.6.4, write

$$f(x) = r_0p_0(x) + r_1p_1(x) + \cdots + r_mp_m(x)$$

Suppose that $f(x)$ has at most k roots for any choice of the coefficients $r_0, r_1, \dots, r_m$, not all zero.

a. Show that M^TM is invertible if at least $k+1$ of the x_i are distinct.

b. If at least two of the x_i are distinct, show that there is always a best approximation of the form $r_0 + r_1e^x$.

c. If at least three of the x_i are distinct, show that there is always a best approximation of the form $r_0 + r_1x + r_2e^x$. [Calculus is needed.]

Exercise 5.6.14 If A is an $m \times n$ matrix, it can be proved that there exists a unique $n \times m$ matrix $A^\#$ satisfying the following four conditions: $AA^\#A = A$; $A^\#AA^\# = A^\#$; $AA^\#$ and $A^\#A$ are symmetric. The matrix $A^\#$ is called the **generalized inverse** of A, or the **Moore-Penrose** inverse.

a. If A is square and invertible, show that $A^\# = A^{-1}$.

b. If rank $A = m$, show that $A^\# = A^T(AA^T)^{-1}$.

c. If rank $A = n$, show that $A^\# = (A^TA)^{-1}A^T$.

5.7 An Application to Correlation and Variance

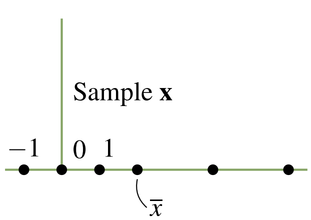
Suppose the heights $h_1, h_2, \dots, h_n$ of n men are measured. Such a data set is called a **sample** of the heights of all the men in the population under study, and various questions are often asked about such a sample: What is the average height in the sample? How much variation is there in the sample heights, and how can it be measured? What can be inferred from the sample about the heights of all men in the population? How do these heights compare to heights of men in neighbouring countries? Does the prevalence of smoking affect the height of a man?

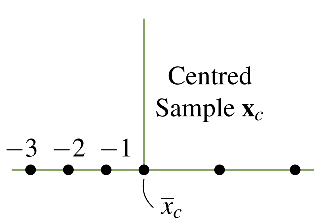
The analysis of samples, and of inferences that can be drawn from them, is a subject called *mathematical statistics*, and an extensive body of information has been developed to answer many such questions. In this section we will describe a few ways that linear algebra can be used.

It is convenient to represent a sample $\{x_1, x_2, \dots, x_n\}$ as a **sample vector**[15] $\mathbf{x} = \left[\begin{array}{cccc} x_1 & x_2 & \cdots & x_n \end{array}\right]$ in $\mathbb{R}^n$. This being done, the dot product in $\mathbb{R}^n$ provides a convenient tool to study the sample and describe some of the statistical concepts related to it. The most widely known statistic for describing a data set is the **sample mean** $\bar{x}$ defined by[16]

$$\bar{x} = \frac{1}{n}(x_1 + x_2 + \cdots + x_n) = \frac{1}{n}\sum_{i=1}^{n} x_i$$

The mean $\bar{x}$ is "typical" of the sample values x_i, but may not itself be one of them. The number $x_i - \bar{x}$ is called the **deviation** of x_i from the mean $\bar{x}$. The deviation is positive if $x_i > \bar{x}$ and it is negative if $x_i < \bar{x}$. Moreover, the sum of these deviations is zero:





$$\sum_{i=1}^{n}(x_i - \bar{x}) = \left(\sum_{i=1}^{n} x_i\right) - n\bar{x} = n\bar{x} - n\bar{x} = 0 \tag{5.6}$$

This is described by saying that the sample mean $\bar{x}$ is *central* to the sample values x_i.

If the mean $\bar{x}$ is subtracted from each data value x_i, the resulting data $x_i - \bar{x}$ are said to be **centred**. The corresponding data vector is

$$\mathbf{x}_c = \left[\begin{array}{cccc} x_1 - \bar{x} & x_2 - \bar{x} & \cdots & x_n - \bar{x} \end{array}\right]$$

and (5.6) shows that the mean $\bar{x}_c = 0$. For example, we have plotted the sample $\mathbf{x} = \left[\begin{array}{ccccc} -1 & 0 & 1 & 4 & 6 \end{array}\right]$ in the first diagram. The mean is $\bar{x} = 2$,

[15] We write vectors in $\mathbb{R}^n$ as row matrices, for convenience.

[16] The mean is often called the "average" of the sample values x_i, but statisticians use the term "mean".

and the centred sample $\mathbf{x}_c = \begin{bmatrix} -3 & -2 & -1 & 2 & 4 \end{bmatrix}$ is also plotted. Thus, the effect of centring is to shift the data by an amount $\overline{x}$ (to the left if $\overline{x}$ is positive) so that the mean moves to 0.

Another question that arises about samples is how much variability there is in the sample

$$\mathbf{x} = \begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix}$$

that is, how widely are the data "spread out" around the sample mean $\overline{x}$. A natural measure of variability would be the sum of the deviations of the x_i about the mean, but this sum is zero by (5.6); these deviations cancel out. To avoid this cancellation, statisticians use the *squares* $(x_i - \overline{x})^2$ of the deviations as a measure of variability. More precisely, they compute a statistic called the **sample variance** s_x^2 defined[17] as follows:

$$s_x^2 = \frac{1}{n-1}[(x_1-\overline{x})^2 + (x_2-\overline{x})^2 + \cdots + (x_n-\overline{x})^2] = \frac{1}{n-1}\sum_{i=1}^{n}(x_i-\overline{x})^2$$

The sample variance will be large if there are many x_i at a large distance from the mean $\overline{x}$, and it will be small if all the x_i are tightly clustered about the mean. The variance is clearly nonnegative (hence the notation s_x^2), and the square root s_x of the variance is called the **sample standard deviation**.

The sample mean and variance can be conveniently described using the dot product. Let

$$\mathbf{1} = \begin{bmatrix} 1 & 1 & \cdots & 1 \end{bmatrix}$$

denote the row with every entry equal to 1. If $\mathbf{x} = \begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix}$, then $\mathbf{x}\cdot\mathbf{1} = x_1 + x_2 + \cdots + x_n$, so the sample mean is given by the formula

$$\overline{x} = \tfrac{1}{n}(\mathbf{x}\cdot\mathbf{1})$$

Moreover, remembering that $\overline{x}$ is a scalar, we have $\overline{x}\mathbf{1} = \begin{bmatrix} \overline{x} & \overline{x} & \cdots & \overline{x} \end{bmatrix}$, so the centred sample vector $\mathbf{x}_c$ is given by

$$\mathbf{x}_c = \mathbf{x} - \overline{x}\mathbf{1} = \begin{bmatrix} x_1-\overline{x} & x_2-\overline{x} & \cdots & x_n-\overline{x} \end{bmatrix}$$

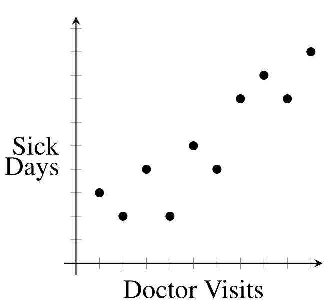
Thus we obtain a formula for the sample variance:

$$s_x^2 = \tfrac{1}{n-1}\|\mathbf{x}_c\|^2 = \tfrac{1}{n-1}\|\mathbf{x} - \overline{x}\mathbf{1}\|^2$$

Linear algebra is also useful for comparing two different samples. To illustrate how, consider two examples.

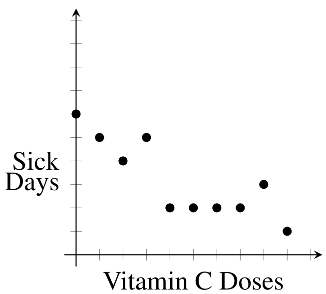
The following table represents the number of sick days at work per year and the yearly number of visits to a physician for 10 individuals.

Individual	1	2	3	4	5	6	7	8	9	10
Doctor visits	2	6	8	1	5	10	3	9	7	4
Sick days	2	4	8	3	5	9	4	7	7	2



The data are plotted in the **scatter diagram** where it is evident that, roughly speaking, the more visits to the doctor the more sick days. This is an example of a *positive correlation* between sick days and doctor visits.

[17]Since there are n sample values, it seems more natural to divide by n here, rather than by $n-1$. The reason for using $n-1$ is that then the sample variance s^2x provides a better estimate of the variance of the entire population from which the sample was drawn.



Now consider the following table representing the daily doses of vitamin C and the number of sick days.

Individual	1	2	3	4	5	6	7	8	9	10
Vitamin C	1	5	7	0	4	9	2	8	6	3
Sick days	5	2	2	6	2	1	4	3	2	5

The scatter diagram is plotted as shown and it appears that the more vitamin C taken, the fewer sick days. In this case there is a *negative correlation* between daily vitamin C and sick days.

In both these situations, we have **paired samples**, that is observations of two variables are made for ten individuals: doctor visits and sick days in the first case; daily vitamin C and sick days in the second case. The scatter diagrams point to a relationship between these variables, and there is a way to use the sample to compute a number, called the correlation coefficient, that measures the degree to which the variables are associated.

To motivate the definition of the correlation coefficient, suppose two paired samples $\mathbf{x} = \left[\begin{array}{cccc} x_1 & x_2 & \cdots & x_n \end{array} \right]$, and $\mathbf{y} = \left[\begin{array}{cccc} y_1 & y_2 & \cdots & y_n \end{array} \right]$ are given and consider the centred samples

$$\mathbf{x}_c = \left[\begin{array}{cccc} x_1 - \overline{x} & x_2 - \overline{x} & \cdots & x_n - \overline{x} \end{array} \right] \text{ and } \mathbf{y}_c = \left[\begin{array}{cccc} y_1 - \overline{y} & y_2 - \overline{y} & \cdots & y_n - \overline{y} \end{array} \right]$$

If x_k is large among the x_i's, then the deviation $x_k - \overline{x}$ will be positive; and $x_k - \overline{x}$ will be negative if x_k is small among the x_i's. The situation is similar for $\mathbf{y}$, and the following table displays the sign of the quantity $(x_i - \overline{x})(y_k - \overline{y})$ in all four cases:

Sign of $(x_i - \overline{x})(y_k - \overline{y})$:

	x_i large	x_i small
y_i large	positive	negative
y_i small	negative	positive

Intuitively, if $\mathbf{x}$ and $\mathbf{y}$ are positively correlated, then two things happen:

1. *Large values of the x_i tend to be associated with large values of the y_i, and*
2. *Small values of the x_i tend to be associated with small values of the y_i.*

It follows from the table that, if $\mathbf{x}$ and $\mathbf{y}$ are positively correlated, then the dot product

$$\mathbf{x}_c \cdot \mathbf{y}_c = \sum_{i=1}^{n} (x_i - \overline{x})(y_i - \overline{y})$$

is positive. Similarly $\mathbf{x}_c \cdot \mathbf{y}_c$ is negative if $\mathbf{x}$ and $\mathbf{y}$ are negatively correlated. With this in mind, the **sample correlation coefficient**[18] r is defined by

$$r = r(\mathbf{x},\ \mathbf{y}) = \frac{\mathbf{x}_c \cdot \mathbf{y}_c}{\|\mathbf{x}_c\|\ \|\mathbf{y}_c\|}$$

[18]The idea of using a single number to measure the degree of relationship between different variables was pioneered by Francis Galton (1822–1911). He was studying the degree to which characteristics of an offspring relate to those of its parents. The idea was refined by Karl Pearson (1857–1936) and r is often referred to as the Pearson correlation coefficient.

Bearing the situation in $\mathbb{R}^3$ in mind, r is the cosine of the "angle" between the vectors $\mathbf{x}_c$ and $\mathbf{y}_c$, and so we would expect it to lie between -1 and 1. Moreover, we would expect r to be near 1 (or -1) if these vectors were pointing in the same (opposite) direction, that is the "angle" is near zero (or π).

This is confirmed by Theorem 5.7.1 below, and it is also borne out in the examples above. If we compute the correlation between sick days and visits to the physician (in the first scatter diagram above) the result is $r = 0.90$ as expected. On the other hand, the correlation between daily vitamin C doses and sick days (second scatter diagram) is $r = -0.84$.

However, a word of caution is in order here. We *cannot* conclude from the second example that taking more vitamin C will reduce the number of sick days at work. The (negative) correlation may arise because of some third factor that is related to both variables. For example, case it may be that less healthy people are inclined to take more vitamin C. Correlation does *not* imply causation. Similarly, the correlation between sick days and visits to the doctor does not mean that having many sick days *causes* more visits to the doctor. A correlation between two variables may point to the existence of other underlying factors, but it does not necessarily mean that there is a causality relationship between the variables.

Our discussion of the dot product in $\mathbb{R}^n$ provides the basic properties of the correlation coefficient:

Theorem 5.7.1

Let $\mathbf{x} = \left[\begin{array}{cccc} x_1 & x_2 & \cdots & x_n \end{array}\right]$ *and* $\mathbf{y} = \left[\begin{array}{cccc} y_1 & y_2 & \cdots & y_n \end{array}\right]$ *be (nonzero) paired samples, and let* $r = r(\mathbf{x},\ \mathbf{y})$ *denote the correlation coefficient. Then:*

1. $-1 \leq r \leq 1$.
2. $r = 1$ *if and only if there exist* a *and* $b > 0$ *such that* $y_i = a + bx_i$ *for each* i.
3. $r = -1$ *if and only if there exist* a *and* $b < 0$ *such that* $y_i = a + bx_i$ *for each* i.

Proof. The Cauchy inequality (Theorem 5.3.2) proves (1), and also shows that $r = \pm 1$ if and only if one of $\mathbf{x}_c$ and $\mathbf{y}_c$ is a scalar multiple of the other. This in turn holds if and only if $\mathbf{y}_c = b\mathbf{x}_c$ for some $b \neq 0$, and it is easy to verify that $r = 1$ when $b > 0$ and $r = -1$ when $b < 0$.

Finally, $\mathbf{y}_c = b\mathbf{x}_c$ means $y_i - \overline{y} = b(x_i - \overline{x})$ for each i; that is, $y_i = a + bx_i$ where $a = \overline{y} - b\overline{x}$. Conversely, if $y_i = a + bx_i$, then $\overline{y} = a + b\overline{x}$ (verify), so $y_1 - \overline{y} = (a + bx_i) - (a + b\overline{x}) = b(x_1 - \overline{x})$ for each i. In other words, $\mathbf{y}_c = b\mathbf{x}_c$. This completes the proof. □

Properties (2) and (3) in Theorem 5.7.1 show that $r(\mathbf{x},\ \mathbf{y}) = 1$ means that there is a linear relation with *positive* slope between the paired data (so large x values are paired with large y values). Similarly, $r(\mathbf{x},\ \mathbf{y}) = -1$ means that there is a linear relation with *negative* slope between the paired data (so small x values are paired with small y values). This is borne out in the two scatter diagrams above.

We conclude by using the dot product to derive some useful formulas for computing variances and correlation coefficients. Given samples $\mathbf{x} = \left[\begin{array}{cccc} x_1 & x_2 & \cdots & x_n \end{array}\right]$ and $\mathbf{y} = \left[\begin{array}{cccc} y_1 & y_2 & \cdots & y_n \end{array}\right]$, the key observation is the following formula:

$$\mathbf{x}_c \cdot \mathbf{y}_c = \mathbf{x} \cdot \mathbf{y} - n\overline{x}\,\overline{y} \tag{5.7}$$

Indeed, remembering that $\overline{x}$ and $\overline{y}$ are scalars:

$$\begin{aligned}
\mathbf{x}_c \cdot \mathbf{y}_c &= (\mathbf{x} - \overline{x}\mathbf{1}) \cdot (\mathbf{y} - \overline{y}\mathbf{1}) \\
&= \mathbf{x} \cdot \mathbf{y} - \mathbf{x} \cdot (\overline{y}\mathbf{1}) - (\overline{x}\mathbf{1}) \cdot \mathbf{y} + (\overline{x}\mathbf{1})(\overline{y}\mathbf{1}) \\
&= \mathbf{x} \cdot \mathbf{y} - \overline{y}(\mathbf{x} \cdot \mathbf{1}) - \overline{x}(\mathbf{1} \cdot \mathbf{y}) + \overline{x}\overline{y}(\mathbf{1} \cdot \mathbf{1}) \\
&= \mathbf{x} \cdot \mathbf{y} - \overline{y}(n\overline{x}) - \overline{x}(n\overline{y}) + \overline{x}\,\overline{y}(n) \\
&= \mathbf{x} \cdot \mathbf{y} - n\overline{x}\,\overline{y}
\end{aligned}$$

Taking $\mathbf{y} = \mathbf{x}$ in (5.7) gives a formula for the variance $s_x^2 = \frac{1}{n-1}\|\mathbf{x}_c\|^2$ of $\mathbf{x}$.

Variance Formula

If x is a sample vector, then $s_x^2 = \frac{1}{n-1}\left(\|\mathbf{x}_c\|^2 - n\overline{x}^2\right)$.

We also get a convenient formula for the correlation coefficient, $r = r(\mathbf{x},\ \mathbf{y}) = \frac{\mathbf{x}_c \cdot \mathbf{y}_c}{\|\mathbf{x}_c\|\ \|\mathbf{y}_c\|}$. Moreover, (5.7) and the fact that $s_x^2 = \frac{1}{n-1}\|\mathbf{x}_c\|^2$ give:

Correlation Formula

If $\mathbf{x}$ and $\mathbf{y}$ are sample vectors, then

$$r = r(\mathbf{x},\ \mathbf{y}) = \frac{\mathbf{x} \cdot \mathbf{y} - n\overline{x}\,\overline{y}}{(n-1)s_x s_y}$$

Finally, we give a method that simplifies the computations of variances and correlations.

Data Scaling

Let $\mathbf{x} = \left[\begin{array}{cccc} x_1 & x_2 & \cdots & x_n \end{array}\right]$ *and* $\mathbf{y} = \left[\begin{array}{cccc} y_1 & y_2 & \cdots & y_n \end{array}\right]$ *be sample vectors. Given constants a, b, c, and d, consider new samples* $\mathbf{z} = \left[\begin{array}{cccc} z_1 & z_2 & \cdots & z_n \end{array}\right]$ *and* $\mathbf{w} = \left[\begin{array}{cccc} w_1 & w_2 & \cdots & w_n \end{array}\right]$ *where* $z_i = a + bx_i$*, for each i and* $w_i = c + dy_i$ *for each i. Then:*

a. $\overline{z} = a + b\overline{x}$

b. $s_z^2 = b^2 s_x^2$, so $s_z = |b|s_x$

c. *If b and d have the same sign, then* $r(\mathbf{x},\ \mathbf{y}) = r(\mathbf{z},\ \mathbf{w})$.

The verification is left as an exercise. For example, if $\mathbf{x} = \left[\begin{array}{cccccc} 101 & 98 & 103 & 99 & 100 & 97 \end{array}\right]$, subtracting 100 yields $\mathbf{z} = \left[\begin{array}{cccccc} 1 & -2 & 3 & -1 & 0 & -3 \end{array}\right]$. A routine calculation shows that $\overline{z} = -\frac{1}{3}$ and $s_z^2 = \frac{14}{3}$, so $\overline{x} = 100 - \frac{1}{3} = 99.67$, and $s_z^2 = \frac{14}{3} = 4.67$.

Exercises for 5.7

Exercise 5.7.1 The following table gives IQ scores for 10 fathers and their eldest sons. Calculate the means, the variances, and the correlation coefficient r. (The data scaling formula is useful.)

	1	2	3	4	5	6	7	8	9	10
Father's IQ	140	131	120	115	110	106	100	95	91	86
Son's IQ	130	138	110	99	109	120	105	99	100	94

Exercise 5.7.2 The following table gives the number of years of education and the annual income (in thousands) of 10 individuals. Find the means, the variances, and the correlation coefficient. (Again the data scaling formula is useful.)

Individual	1	2	3	4	5	6	7	8	9	10
Years of education	12	16	13	18	19	12	18	19	12	14
Yearly income (1000's)	31	48	35	28	55	40	39	60	32	35

Exercise 5.7.3 If $\mathbf{x}$ is a sample vector, and $\mathbf{x}_c$ is the centred sample, show that $\overline{x}_c = 0$ and the standard deviation of $\mathbf{x}_c$ is s_x.

Exercise 5.7.4 Prove the data scaling formulas found on page 326: (a), (b), and (c).

Supplementary Exercises for Chapter 5

Exercise 5.1 In each case either show that the statement is true or give an example showing that it is false. Throughout, $\mathbf{x}$, $\mathbf{y}$, $\mathbf{z}$, $\mathbf{x}_1$, $\mathbf{x}_2$, $\dots$, $\mathbf{x}_n$ denote vectors in $\mathbb{R}^n$.

a. If U is a subspace of $\mathbb{R}^n$ and $\mathbf{x}+\mathbf{y}$ is in U, then $\mathbf{x}$ and $\mathbf{y}$ are both in U.

b. If U is a subspace of $\mathbb{R}^n$ and $r\mathbf{x}$ is in U, then $\mathbf{x}$ is in U.

c. If U is a nonempty set and $s\mathbf{x}+t\mathbf{y}$ is in U for any s and t whenever $\mathbf{x}$ and $\mathbf{y}$ are in U, then U is a subspace.

d. If U is a subspace of $\mathbb{R}^n$ and $\mathbf{x}$ is in U, then $-\mathbf{x}$ is in U.

e. If $\{\mathbf{x}, \mathbf{y}\}$ is independent, then $\{\mathbf{x}, \mathbf{y}, \mathbf{x}+\mathbf{y}\}$ is independent.

f. If $\{\mathbf{x}, \mathbf{y}, \mathbf{z}\}$ is independent, then $\{\mathbf{x}, \mathbf{y}\}$ is independent.

g. If $\{\mathbf{x}, \mathbf{y}\}$ is not independent, then $\{\mathbf{x}, \mathbf{y}, \mathbf{z}\}$ is not independent.

h. If all of $\mathbf{x}_1$, $\mathbf{x}_2$, $\dots$, $\mathbf{x}_n$ are nonzero, then $\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n\}$ is independent.

i. If one of $\mathbf{x}_1$, $\mathbf{x}_2$, $\dots$, $\mathbf{x}_n$ is zero, then $\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n\}$ is not independent.

j. If $a\mathbf{x}+b\mathbf{y}+c\mathbf{z}=\mathbf{0}$ where a, b, and c are in $\mathbb{R}$, then $\{\mathbf{x}, \mathbf{y}, \mathbf{z}\}$ is independent.

k. If $\{\mathbf{x}, \mathbf{y}, \mathbf{z}\}$ is independent, then $a\mathbf{x}+b\mathbf{y}+c\mathbf{z}=\mathbf{0}$ for some a, b, and c in $\mathbb{R}$.

l. If $\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n\}$ is not independent, then $t_1\mathbf{x}_1+t_2\mathbf{x}_2+\cdots+t_n\mathbf{x}_n=\mathbf{0}$ for t_i in $\mathbb{R}$ not all zero.

m. If $\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n\}$ is independent, then $t_1\mathbf{x}_1+t_2\mathbf{x}_2+\cdots+t_n\mathbf{x}_n=\mathbf{0}$ for some t_i in $\mathbb{R}$.

n. Every set of four non-zero vectors in $\mathbb{R}^4$ is a basis.

o. No basis of $\mathbb{R}^3$ can contain a vector with a component $\mathbf{0}$.

p. $\mathbb{R}^3$ has a basis of the form $\{\mathbf{x},\ \mathbf{x}+\mathbf{y},\ \mathbf{y}\}$ where $\mathbf{x}$ and $\mathbf{y}$ are vectors.

q. Every basis of $\mathbb{R}^5$ contains one column of I_5.

r. Every nonempty subset of a basis of $\mathbb{R}^3$ is again a basis of $\mathbb{R}^3$.

s. If $\{\mathbf{x}_1,\ \mathbf{x}_2,\ \mathbf{x}_3,\ \mathbf{x}_4\}$ and $\{\mathbf{y}_1,\ \mathbf{y}_2,\ \mathbf{y}_3,\ \mathbf{y}_4\}$ are bases of $\mathbb{R}^4$, then $\{\mathbf{x}_1+\mathbf{y}_1,\ \mathbf{x}_2+\mathbf{y}_2,\ \mathbf{x}_3+\mathbf{y}_3,\ \mathbf{x}_4+\mathbf{y}_4\}$ is also a basis of $\mathbb{R}^4$.