CHAPTER 1

BASICS OF LINEAR ALGEBRA

Undoubtedly, one of the subjects in mathematics that has become more indispensable than ever is linear algebra. Several application problems involve at some stage solving linear systems, the computation of eigenvalues and eigenvectors, linear transformations, bases of vector subspaces, and matrix factorizations, to mention a few. One very important characteristic of linear algebra is that as a first course, it requires only very basic prerequisites so that it can be taught very early at undergraduate level; at the same time, mastering vector spaces, linear transformations and their natural extensions to function spaces is essential for researchers in any area of applied mathematics. Linear algebra has innumerable applications, including differential equations, least-square solutions and optimization, demography, electrical engineering, fractal geometry, communication networks, compression, search engines, social sciences, etc. In the next sections we briefly review the concepts of linear algebra that we will need later on.

1.1 NOTATION AND TERMINOLOGY

We start this section by defining an \( m \times n \) matrix as a rectangular array of elements arranged in \( m \) rows and \( n \) columns, and we say the matrix is of order \( m \times n \). We usually denote the elements of a matrix \( A \) of order \( m \times n \) as \( a_{ij} \), where \( i = 1, \ldots, m \), \( j = 1, \ldots, n \), and we
write the matrix $A$ as

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}.$$ 

Although the elements of a matrix can be real or complex numbers, here we will mostly consider the entries of a given matrix to be real unless otherwise stated. In most cases, we will take care to state the complex version of some definitions and results.

**Matrix addition and multiplication.** Given two arbitrary matrices $A_{m \times n}$ and $B_{m \times n}$, we define the matrix

$$C = A + B$$

by adding the entries of $A$ and $B$ componentwise. That is, we have

$$c_{ij} = a_{ij} + b_{ij},$$

for $i = 1, \ldots, m$, $j = 1, \ldots, n$.

This means that the addition of matrices is well defined only for matrices of the same order.

Now consider two arbitrary matrices $A_{m \times p}$ and $B_{p \times n}$. Then we define the product matrix $C_{m \times n} = A \cdot B$ as

$$c_{ij} = \sum_{k=1}^{p} a_{ik} b_{kj},$$

where $i = 1, \ldots, m$, $j = 1, \ldots, n$. This means that to obtain the entry $(i, j)$ of the product matrix, we multiply the $i$-th row of $A$ with the $j$-th column of $B$ entry-wise and add their products.

Observe that for the product to be well-defined, the number of columns of $A$ has to agree with the number of rows of $B$.

**Example 1.1**

Let $A = \begin{bmatrix} 4 & -5 \\ 3 & 2 \\ 1 & 6 \end{bmatrix}$, $B = \begin{bmatrix} -1 & 2 \\ 0 & 3 \end{bmatrix}$. Then, for instance, to obtain the entry $c_{32}$ of the product matrix $C = AB$, we multiply entrywise the third row of $A$ with the second column of $B$: $(1)(2) + (6)(3) = 20$. Thus, we get

$$C = A \cdot B = \begin{bmatrix} 4 & -5 \\ 3 & 2 \\ 1 & 6 \end{bmatrix} \begin{bmatrix} -1 & 2 \\ 0 & 3 \end{bmatrix} = \begin{bmatrix} -4 & -7 \\ -3 & 12 \\ -1 & 20 \end{bmatrix}.$$
• MATLAB command: \( A + B, \ A \ast B \).

Given a matrix \( B_{p \times n} \), we can denote its columns with \( b_1, \ldots, b_n \), where each \( b_i \) is a \( p \)-dimensional vector. We will write accordingly,

\[
B = [b_1 \ \cdots \ b_n].
\]

In such a case, we can write a product of matrices as

\[
AB = A[b_1 \ \cdots \ b_n] = [Ab_1 \ \cdots \ Ab_n],
\]

so that \( Ab_1, \ldots, Ab_n \) are the columns of the matrix \( AB \).

Similarly, if we denote with \( a_1, \ldots, a_m \) the rows of a matrix \( A_{m \times p} \), then

\[
AB = \begin{bmatrix}
a_1 \\
\vdots \\
a_m
\end{bmatrix}B = \begin{bmatrix}
a_1B \\
\vdots \\
a_mB
\end{bmatrix}.
\]

For an arbitrary matrix \( A_{m \times n} \), we denote with \( A^T \) its transpose matrix, that is, the matrix of order \( n \times m \), where the rows of \( A \) have been exchanged for columns and vice versa. For instance,

\[
\text{If } A = \begin{bmatrix} 6 & -4 & 7 \\ -2 & 3 & 8 \end{bmatrix}, \text{ then } A^T = \begin{bmatrix} 6 & -2 \\ -4 & 3 \\ 7 & 8 \end{bmatrix}.
\]

**Remark 1.1** If \( A_{m \times n} \) is a complex matrix, its adjoint matrix is denoted as \( A^* \), where \( A^* = A^T \), that is, the conjugate transpose. For instance,

\[
\text{If } A = \begin{bmatrix} 3 & -4 + i \\ -2i & 5 + 2i \end{bmatrix}, \text{ then } A^* = \begin{bmatrix} 3 & 2i \\ -4 - i & 5 - 2i \end{bmatrix}.
\]

• MATLAB command: \( A' \)

The sum and product of matrices satisfy the following properties (see Exercise 1.1):

\[
(A + B)^T = A^T + B^T, \quad (AB)^T = B^T A^T.
\]

**Definition 1.2** The trace of a square matrix \( A \) of order \( n \) is defined as the sum of its diagonal elements, that is,

\[
\text{tr}(A) = \sum_{i=1}^{n} a_{ii}.
\]
- **MATLAB command:** `trace(A)`

A particular and very special case is that of matrices of order $n \times 1$. Such a matrix is usually called an $n$-dimensional vector. That is, here we consider vectors as column-vectors, and we will use the notation

$$
\mathbf{x} = \begin{bmatrix}
x_1 \\
\vdots \\
x_n
\end{bmatrix} = [x_1 \cdots x_n]^T
$$

for a typical vector. This substitutes the usual notation $\mathbf{x} = (x_1, \ldots, x_n)$, which we reserve to denote a point, and at the same time this notation will allow us to perform matrix-vector operations in agreement with their dimensions. This also closely follows the notation used in MATLAB.

The first two vectors below are column vectors, whereas the third is a row vector.

$$
\begin{bmatrix}
9 \\
4 \\
3
\end{bmatrix}, \quad [4 - 3 5]^T, \quad [1 8 5].
$$

### 1.2 Vector and Matrix Norms

It is always important and useful to have a notion for the "size" or magnitude of a vector or a matrix, just as we understand the magnitude of a real number by using its absolute value. In fact, a norm can be understood as the generalization of the absolute value function to a higher dimensional case. This is especially useful in numerical analysis for estimating the magnitude of the error when approximating the solution to a given problem.

**Definition 1.3** A vector norm, denoted by $\| \cdot \|$, is a real function that satisfies the following properties for arbitrary $n$-dimensional vectors $\mathbf{x}$ and $\mathbf{y}$ and for arbitrary real or complex $\alpha$:

(i) $\|\mathbf{x}\| \geq 0$,

(ii) $\|\mathbf{x}\| = 0$ if and only if $\mathbf{x} = \mathbf{0}$,

(iii) $\|\alpha \mathbf{x}\| = |\alpha| \|\mathbf{x}\|$,

(iv) $\|\mathbf{x} + \mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|$.

There are several norms that can be defined for vectors. Here we mention the most commonly used.

Let $\mathbf{x} = [x_1 \cdots x_n]^T, \quad x_i \in \mathbb{R}$. Then, we define
\[ \|x\|_1 = \sum_{i=1}^{n} |x_i| \quad \text{Sum norm} \]
\[ \|x\|_2 = \left( \sum_{i=1}^{n} x_i^2 \right)^{1/2} \quad \text{Euclidean norm} \tag{1.4} \]
\[ \|x\|_\infty = \max_{i=1,\ldots,n} |x_i| \quad \text{Maximum norm} \]

- MATLAB commands: \texttt{norm}(x, 1), \texttt{norm}(x), \texttt{norm}(x,\infty)

\section*{Example 1.2}

Let \( x = [3 \quad -2 \quad 4 \quad \sqrt{7}]^T \). Then,
\[ \|x\|_1 = |3| + |-2| + |4| + |\sqrt{7}| \approx 11.6458. \]
\[ \|x\|_2 = \sqrt{9 + 4 + 16 + 7} = 6. \]
\[ \|x\|_\infty = \max \{ |3|, |-2|, |4|, |\sqrt{7}| \} = 4. \]

\section*{Remark 1.4}

In general, for an arbitrary vector \( x \in \mathbb{R}^n \), the following inequalities (see Exercise 1.4) relate the three norms above:
\[ \|x\|_\infty \leq \|x\|_2 \leq \|x\|_1. \tag{1.5} \]

\section*{Example 1.3}

The unit ball in \( \mathbb{R}^n \) is the set \( \{ x \in \mathbb{R}^n : \|x\| \leq 1 \} \). The geometrical shape of the unit ball varies according to what norm is used. For the case \( n = 2 \), the unit balls for the three norms in (1.4) are shown in Figure 1.1.

\section*{Note:}
For simplicity of notation, \( \| \cdot \| \) will always denote the Euclidean norm \( \| \cdot \|_2 \) for vectors, unless otherwise stated.

We now introduce the notion of norms for a matrix, in some sense as a generalization of vector norms.

\section*{Definition 1.5}
A matrix norm is a real function that for arbitrary matrices \( A \) and \( B \) and arbitrary real or complex \( \alpha \), satisfies
\begin{enumerate}
  \item \( \|A\| \geq 0 \),
  \item \( \|A\| = 0 \) if and only if \( A = 0 \),
  \item \( \|\alpha A\| = |\alpha| \|A\| \),
\end{enumerate}
Let $A$ be an $m \times n$ matrix. Here are some of the most common matrix norms

1. Maximum column sum norm

\[
\|A\|_1 = \max_{j=1,...,n} \sum_{i=1}^{m} |a_{ij}|
\]

2. Frobenius norm

\[
\|A\|_F = \left( \sum_{i=1}^{m} \sum_{j=1}^{n} |a_{ij}|^2 \right)^{1/2}
\]

3. Maximum row sum norm

\[
\|A\|_{\infty} = \max_{i=1,...,m} \sum_{j=1}^{n} |a_{ij}|
\]

- MATLAB command: `norm(A,1), norm(A,'fro'), norm(A,inf)`

**Remark 1.6** The Frobenius norm is also known as the Euclidean norm for matrices, and in some textbooks it is denoted as $\| \cdot \|_2$. We will reserve this notation for a different matrix norm.

**Example 1.4**

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

\[
\|A\|_1 = \max \{ 7, 6, 5 \} = 7.
\]

\[
\|A\|_F = (1 + 9 + 4 + 4 + 0 + 4 + 16 + 9 + 1)^{1/2} \approx 6.9282.
\]

\[
\|A\|_{\infty} = \max \{ 6, 4, 8 \} = 8.
\]
**Remark 1.7** There are two other ways of defining the Frobenius norm, which are useful for certain computations, as we illustrate later on. Let \( A_{m \times n} \) be an arbitrary matrix, and denote its columns with \( a_1, \ldots, a_n \). Then,

\[
\text{(a) } \| A \|_F^2 = \| a_1 \|_2^2 + \cdots + \| a_n \|_2^2. \tag{1.7}
\]

\[
\text{(b) } \| A \|_F^2 = \text{tr}(A^T A). \tag{1.8}
\]

It seems natural to think that vector norms can be directly generalized to obtain matrix norms (after all, \( m \times n \) matrices can be thought of as vectors of dimension \( m \cdot n \)). However, not all vector norms directly become matrix norms (see Exercise 1.11).

A matrix norm can be **induced** or constructed directly from a vector norm by defining for \( 1 \leq p \leq \infty \)

\[
\| A \|_p = \max_{\| x \|_p = 1} \| A x \|_p, \tag{1.9}
\]

where the norms on the right-hand side represent the vector norm. This is the way to correctly extend or use a vector norm into obtaining a matrix norm.

Given a vector \( x \) of norm \( \| x \| \), when it is multiplied by \( A \), we get the new vector \( A x \) of norm \( \| A x \| \). Thus, we can interpret the matrix norm (1.9) as a natural way to measure how much the vector \( x \) can be stretched or shrunk by \( A \).

The definition of the \( p \)-norm of a matrix in (1.9) is not easy to implement or compute in general. Luckily enough, there are alternative ways to compute such a norm for some particular values of \( p \).

**EXAMPLE 1.5**

It can be proved (see Exercise 1.12) that for \( p = 1 \) and \( p = \infty \), the \( p \)-norms in (1.9) can be directly computed through the corresponding definitions in (1.6).

**EXAMPLE 1.6**

For \( p = 2 \), the \( p \)-norm in (1.9) can be computed as the square root of the largest eigenvalue of the symmetric matrix \( A^T A \). That is,

\[
\| A \|_2 = \max \{ \sqrt{\lambda} : \lambda \text{ is an eigenvalue of } A^T A \}. \tag{1.10}
\]

For instance, for the matrix of Example 1.4, we get \( \| A \|_2 \approx 5.9198 \).

**Remark 1.8** It is possible to show that an equivalent way to define the matrix norm (1.9) is

\[
\| A \|_p = \sup_{x \neq 0} \frac{\| A x \|_p}{\| x \|_p}. \tag{1.11}
\]
Remark 1.9 In general, inequalities for matrix norms similar to the ones in (1.5) are not true. However, it is still true that

\[ ||A||_2 \leq ||A||_F, \]

(1.12)

where \( || \cdot ||_2 \) is given by (1.9).

Finally, we mention another important property that relates matrix and vector norms. We say that the norm of an \( m \times n \) matrix \( A \) is consistent with the norm of an \( n \)-dimensional vector \( x \) if

\[ ||Ax|| \leq ||A||_F ||x||, \]

(1.13)

for any \( x \in \mathbb{R}^n \). Then, it is straightforward to see that the \( p \)-norms defined by (1.9) or (1.11) are all consistent.

1.3 DOT PRODUCT AND ORTHOGONALITY

Quite often we will need to use the so-called dot product or inner product between two vectors \( x = [x_1 \cdots x_n]^T \) and \( y = [y_1 \cdots y_n]^T \), defined as

\[ x \cdot y = \langle x, y \rangle = x_1y_1 + \cdots + x_ny_n. \]

(1.14)

As vectors can also be considered to be \( n \times 1 \) matrices, it is common to represent the inner product as matrix multiplication. Thus, we will consider all these notations to be equivalent:

\[ x \cdot y = \langle x, y \rangle = x^T y. \]

(1.15)

Observe that with the definition above,

\[ ||x||_2^2 = x^T x. \]

For arbitrary vectors \( x, y, z \in \mathbb{R}^n \) and arbitrary real scalar \( c \), the dot product satisfies the following properties:

1. \( \langle x, x \rangle \geq 0 \),
2. \( \langle x, x \rangle = 0 \) if and only if \( x = 0 \),
3. \( \langle cx, y \rangle = c \langle x, y \rangle \),
4. \( \langle x + y, z \rangle = \langle x, z \rangle + \langle y, z \rangle \),
5. \( \langle x, y \rangle = \langle y, x \rangle \).

Remark 1.10 In the complex case, we have

\[ (3a) \langle x, cy \rangle = \overline{c} \langle x, y \rangle \]

(5a) \[ \langle x, y \rangle = \overline{\langle y, x \rangle}. \]

(1.16)
Following (1.15), we also see that for arbitrary $A_{m \times n}$, $x \in \mathbb{R}^n$, $y \in \mathbb{R}^m$,

$$< Ax, y > = (A_y)^T y = x^T A^T y = < x, A^T y >.$$  \hfill (1.17)

**Remark 1.11** The dot product introduced here is a particular case of the general inner product function studied in the context of inner product vector spaces, where the elements are not restricted to real $n$-dimensional vectors.

There is a special kind of vectors that is very useful in several instances in linear algebra and matrix computations. These are the so-called **orthonormal vectors**, which are orthogonal (perpendicular) to each other and they are unit, that is, $x$ and $y$ are orthonormal if

$$x^T y = 0, \quad \text{and} \quad \|x\| = \|y\| = 1.$$  

For example, the following set of vectors is orthonormal:

$$v_1 = \left[\frac{2}{\sqrt{5}} \quad 0 \quad -\frac{1}{\sqrt{5}}\right]^T, \quad v_2 = \left[\frac{1}{\sqrt{5}} \quad 0 \quad \frac{2}{\sqrt{5}}\right]^T, \quad v_3 = [0 \quad -1 \quad 0]^T.$$  

In fact, we can readily verify that

$$v_1^T v_2 = v_1^T v_3 = v_2^T v_3 = 0, \quad \text{and} \quad \|v_1\| = \|v_2\| = \|v_3\| = 1.$$  

### 1.4 SPECIAL MATRICES

We will be using matrices at almost every place in this book, matrices of different kinds and properties. Here we list some of the most common types of matrices we will encounter.

#### 1.4.1 Diagonal and triangular matrices

A square matrix $A$ of order $n$ is called **diagonal** if $a_{ij} = 0$ for all $i \neq j$.

**Example 1.7**

$$A = \begin{bmatrix} 4 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -9 \end{bmatrix}.$$  

- **MATLAB** command: `diag`

A square matrix $A$ of order $n$ is called **upper** (resp. **lower**) **triangular** if $a_{ij} = 0$ for all $i > j$. (resp. for all $i < j$).
**Example 1.8**

\[ A = \begin{bmatrix} 4 & 2 & 5 \\ 0 & 1 & 3 \\ 0 & 0 & 9 \end{bmatrix}, \quad B = \begin{bmatrix} 4 & 0 & 0 \\ 5 & 1 & 0 \\ 7 & 1 & 9 \end{bmatrix}. \]

The matrix \( A \) is upper triangular, and \( B \) is lower triangular.

- MATLAB commands: `triu`, `tril`

**Remark 1.12** If the matrix \( A \) is rectangular, say order \( m \times n \), we say it is upper (resp. lower) trapezoidal if \( a_{ij} = 0 \) for all \( i > j \) (resp. for all \( i < j \)).

A more general case of triangular matrices is that of block triangular matrices. For example, the following \( 5 \times 5 \) matrix is block upper triangular

\[
A = \begin{bmatrix}
-8 & 1 & 0 & 9 & 3 \\
4 & -2 & 8 & 5 & 5 \\
0 & 0 & 5 & 7 & 1 \\
0 & 0 & -9 & -8 & 4 \\
0 & 0 & 0 & 0 & -6
\end{bmatrix} = \begin{bmatrix} A_{11} & A_{12} & A_{13} \\ & A_{22} & A_{23} \\ & & A_{33} \end{bmatrix},
\]

where for instance, the zero in the (2,1) entry of the matrix on the right represents the corresponding \( 2 \times 2 \) zero block of the matrix on the left. In a similar way, a block diagonal matrix can be defined. We will encounter this type of matrices in Section 1.10 when we compute eigenvalues of a matrix and when compute special vector subspaces later in Chapter 7.

### 1.4.2 Hessenberg matrices

A matrix \( A_{n \times n} \) is called an upper Hessenberg matrix if \( a_{ij} = 0 \) for all \( i > j + 1 \). They take the form:

\[
A = \begin{bmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1n-1} & a_{1n} \\ a_{21} & a_{22} & a_{23} & \cdots & a_{2n-1} & a_{2n} \\ 0 & a_{32} & a_{33} & \cdots & a_{3n-1} & a_{3n} \\ 0 & 0 & a_{43} & \cdots & a_{4n-1} & a_{4n} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & a_{n(n-1)} & a_{nn} \end{bmatrix}.
\]

(1.18)

In other words, it is an upper triangular matrix with an additional subdiagonal below the main diagonal.

A matrix \( A \) is called lower Hessenberg if \( A^T \) is upper Hessenberg.

- MATLAB command: `hess(A)`
1.4.3 Nonsingular and inverse matrices

In the set of real numbers \( \mathbb{R} \), the number \( 1 \) is the multiplicative identity, meaning that \( a \cdot 1 = 1 \cdot a = a \), for any real number \( a \). This is generalized to what is called the identity matrix of order \( n \), denoted by \( I \). It consists of a diagonal of ones, and every other entry is zero (see (1.21) below), with the property that

\[
AI = IA = A,
\]

for any matrix \( A \).

- **MATLAB command:** `eye(n)`.

A square matrix \( A \) of order \( n \) is called **nonsingular** if its determinant is not zero: \( \det(A) \neq 0 \). Otherwise it is called **singular**.

**EXAMPLE 1.9**

\[
\begin{bmatrix}
4 & 2 & 9 \\
1 & 1 & 0 \\
7 & 1 & 3 \\
\end{bmatrix}
\]

\[= -48. \] Hence, the matrix is nonsingular.

- **MATLAB command:** `det(A)`.

**Remark 1.13** Two very useful properties of determinants are the following:

\[
\det(AB) = \det A \det B, \tag{1.19}
\]

\[
\det A^{-1} = 1/\det A.
\]

In one dimension, every real number \( a \neq 0 \) has a unique multiplicative inverse \( 1/a \), and it is obvious that \( a \left( \frac{1}{a} \right) = \left( \frac{1}{a} \right) a = 1 \). This has a natural generalization to matrices. A nonsingular matrix \( A_{n \times n} \) has a unique \( n \times n \) **inverse matrix**, denoted by \( A^{-1} \) with the property that

\[
AA^{-1} = A^{-1}A = I. \tag{1.20}
\]

For this reason, nonsingular matrices are also called **invertible**.

- **MATLAB command:** `inv(A)`.
EXAMPLE 1.10

Let $A = \begin{bmatrix} 4 & 0 & 9 \\ 1 & -1 & 0 \\ -2 & 1 & 0 \end{bmatrix}$. Since $\det(A) = -9$, the matrix is nonsingular and has an inverse: $A^{-1} = \begin{bmatrix} 0 & -1 & -1 \\ 0 & -2 & -1 \\ 1/9 & 4/9 & 4/9 \end{bmatrix}$, and we can verify that in fact we have

$$AA^{-1} = A^{-1}A = I = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}. \quad (1.21)$$

Computing the inverse of a matrix is not an easy task. There is more than one analytical or exact way to do this, e.g., using the adjoint method, we have

$$A^{-1} = \text{adjoint}(A)/\det(A).$$

However, this method is rarely used in practice, and we do not discuss it here. A second and more efficient method is through Gauss elimination (see Section 3.1). In general, the computation of the inverse of a matrix has to be done numerically, and great care has to be taken due to potentially large accumulation of errors for some matrices. Thus, in practice it is customary to avoid computing the inverse of a matrix explicitly, and some other options must be used. We will discuss these issues later on.

Remark 1.14 For the inverse of a product of nonsingular matrices, we have

$$(A_1A_2\cdots A_k)^{-1} = A_k^{-1}A_{k-1}^{-1}\cdots A_1^{-1}. \quad (1.22)$$

1.4.4 Symmetric and positive definite matrices

A square matrix $A$ of order $n$ is called symmetric if $A = A^T$, that is, the matrix equals its transpose. If $A = (a_{ij})$, then we can also say $A$ is symmetric if $a_{ij} = a_{ji}$, for $i, j = 1, \ldots, n$.

EXAMPLE 1.11

The following matrices are symmetric:

$$A = \begin{bmatrix} 1 & 4 \\ 4 & 9 \end{bmatrix}, \quad A = \begin{bmatrix} 1 & -2 & 7 \\ -2 & 3 & 5 \\ 7 & 5 & 6 \end{bmatrix}.$$
EXAMPLE 1.12

For an arbitrary matrix $A$, the matrices $A + A^T$ and $AA^T$ are symmetric.

Remark 1.15 If $A_{n \times n}$ is complex, we say $A$ is **Hermitian** if $A = A^*$. For instance, the following matrix is Hermitian:

$$A = \begin{bmatrix} 1 & 4 - 2i \\ 4 + 2i & -9 \end{bmatrix}.$$ 

In the set of real numbers, we can define a number $a$ to be positive if $a x^2 > 0$, for all numbers $x \neq 0$. The generalization of this definition for matrices leads to the so-called **positive definite** matrices. A square matrix $A$ is positive definite if

$$x^T A x > 0,$$

for all vectors $x \neq 0$. If instead we have $x^T A x \geq 0$, then the matrix is said to be positive semidefinite.

In several applications, especially optimization, we will encounter matrices that are both, symmetric and positive definite. That is why several textbooks simply assume positive definite matrices to be symmetric; we call such matrices symmetric positive definite (spd). See also Remark 1.16 below.

- **MATLAB command:** chol(A)  (Choleski)

EXAMPLE 1.13

The following matrices are spd:

$$A = \begin{bmatrix} 4 & 10 \\ 10 & 26 \end{bmatrix}, \quad B = \begin{bmatrix} 100 & 15 & 0.01 \\ 15 & 2.3 & 0.01 \\ 0.01 & 0.01 & 1 \end{bmatrix}.$$ 

It is obvious that $A$ is symmetric. To see why it is positive definite, we can verify that $x^T A x > 0$ for any vector $x \neq 0$:

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \begin{bmatrix} 4 & 10 \\ 10 & 26 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = (2x_1 + 5x_2)^2 + x_2^2 > 0.$$

EXAMPLE 1.14

The matrix $A^T A$ is symmetric positive semidefinite, for any matrix $A_{m \times n}$. 
EXAMPLE 1.15

The matrix $A = \begin{bmatrix} a & b \\ -b & a \end{bmatrix}$ with $a > 0$ is positive definite.

**Remark 1.16** It is important to state that a positive definite matrix does not need to be symmetric, as in the case of the matrix $A = \begin{bmatrix} 1 & 0 \\ 2 & 2 \end{bmatrix}$.

**Definition 1.17** Let $A$ be a square matrix of order $n$, and let $m < n$. A principal submatrix of $A$ is a matrix obtained by deleting any $n - m$ rows and corresponding columns. A leading principal submatrix is obtained by deleting the last $n - m$ rows and columns.

**Notation.** We can use and index set $I \subseteq \{1, \ldots, n\}$ to represent which rows and columns of $A$ are used to form the principal submatrix. For instance, if

$$A = \begin{bmatrix} 1 & 4 & 7 \\ 2 & 5 & 8 \\ 3 & 6 & 9 \end{bmatrix}, \text{ then } A(\{1, 3\}) = \begin{bmatrix} 1 & 7 \\ 3 & 9 \end{bmatrix}, \quad A(\{1, 2\}) = \begin{bmatrix} 1 & 4 \\ 2 & 5 \end{bmatrix}.$$

The last one is a leading principal submatrix of $A$.

**Remark 1.18** A positive definite matrix has all its principal submatrices positive definite. Thus, in particular, all the diagonal elements of a positive definite matrix are positive.

### 1.4.5 Matrix exponential

A central tool in the theory and applications of differential equations, dynamical systems, and control theory, is the matrix exponential of a square matrix $A$ of order $n$. This special type of matrix is defined as

$$e^A = \sum_{k=0}^{\infty} \frac{A^k}{k!} = I + \frac{A}{1!} + \frac{A^2}{2!} + \cdots + \frac{A^k}{k!} + \cdots. \quad (1.24)$$

This is an extension of the real exponential function

$$e^x = 1 + \frac{x}{1!} + \frac{x^2}{2!} + \cdots + \frac{x^k}{k!} + \cdots \quad (x \in \mathbb{R})$$

to $n$ dimensions. The series (1.24) always converges for any square matrix $A$. 

EXAMPLE 1.16

Let $A = \begin{bmatrix} 3 & 0 \\ 0 & 5 \end{bmatrix}$. Since $A^0 = I$ and

$$A^2 = \begin{bmatrix} 3^2 & 0 \\ 0 & 5^2 \end{bmatrix}, \quad A^3 = \begin{bmatrix} 3^3 & 0 \\ 0 & 5^3 \end{bmatrix}, \ldots, A^k = \begin{bmatrix} 3^k & 0 \\ 0 & 5^k \end{bmatrix},$$

we have

$$e^A = \begin{bmatrix} \sum_{k=0}^{\infty} \frac{3^k}{k!} & 0 \\ 0 & \sum_{k=0}^{\infty} \frac{5^k}{k!} \end{bmatrix} = \begin{bmatrix} e^3 & 0 \\ 0 & e^5 \end{bmatrix}.$$ 

EXAMPLE 1.17

Let $A = \begin{bmatrix} 0 & 1 & -2 \\ 0 & 0 & 4 \\ 0 & 0 & 0 \end{bmatrix}$. Then, $A^2 = \begin{bmatrix} 0 & 0 & 4 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$, $A^3 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$, and $A^k$ is the zero matrix, for any $k \geq 3$. Then $e^A = I + A + \frac{1}{2} A^2$, and therefore

$$e^A = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 4 \\ 0 & 0 & 1 \end{bmatrix}.$$ 

In Chapter 7, we will consider several other examples of a matrix exponential.

- MATLAB command: `expm(A)`

1.4.6 Permutation matrices

In matrix computations, we often need to perform permutations of rows or columns of a matrix; that is, we exchange the entries of two given rows or columns, respectively. This can be simply achieved through matrix multiplication by a special type of square matrix.

A permutation matrix $P$ can be defined as the identity matrix with some of its rows reordered as needed.

EXAMPLE 1.18

The following are permutation matrices:

$$P_1 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}, \quad P_2 = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}, \quad P_3 = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}.$$
Observe that for the matrices above, we have

\[ P_1^2 = I, \quad P_2^4 = I, \quad P_3^3 = I. \]

Thus, to recover the identity matrix, we need to multiply the permutation matrix \( P \) by itself as many times as the number of rows that have been reordered in the identity (or, as the number of zeros on the diagonal of \( P \)).

In general, due to the way they are defined, permutation matrices satisfy the important property

\[ P^T P = I = P P^T. \]  \hspace{1cm} (1.25)

Therefore, they are invertible, and \( P^{-1} = P^T \).

Multiplying an arbitrary matrix \( A \) by a permutation matrix \( P \) from the left will result in row permutations in \( A \), and multiplying from the right will result in column permutations.

**EXAMPLE 1.19**

Let \( A = \begin{bmatrix} 1 & 5 & 8 \\ 2 & 0 & 5 \\ 3 & 3 & 6 \end{bmatrix} \), and \( P = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} \). Then, since the permutation matrix \( P \) is the identity after permuting its rows or columns 2 and 3, we can permute the corresponding rows (or columns) of \( A \) by multiplying:

\[ P A = \begin{bmatrix} 1 & 5 & 8 \\ 3 & 3 & 6 \\ 2 & 0 & 5 \end{bmatrix}, \quad \text{or} \quad A P = \begin{bmatrix} 1 & 8 & 5 \\ 2 & 5 & 0 \\ 3 & 6 & 3 \end{bmatrix}. \]

It is clear that if \( P \) is a permutation matrix, then \( P^T \) also is (but in general, \( P^T \neq P \), as we see from the matrices \( P_2, P_3 \) in Example 1.18). Then, we can simultaneously interchange rows and columns of a given matrix \( A \) through the operation \( P^T A P \). In the example below, \( AP \) is the matrix \( A \) with the first and third columns interchanged. Then, the product \( P^T A P \) interchanges the first and third row of \( AP \).

**EXAMPLE 1.20**

Let \( A = \begin{bmatrix} 0 & 7 & 9 & 0 \\ 3 & 0 & 0 & 0 \\ 6 & 4 & 0 & 2 \\ 1 & 0 & 5 & 0 \end{bmatrix} \) and \( P = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \). Then,

\[ P^T A P = \begin{bmatrix} 0 & 4 & 6 & 2 \\ 0 & 0 & 3 & 0 \\ 9 & 7 & 0 & 0 \\ 5 & 0 & 1 & 0 \end{bmatrix}. \]
• MATLAB command: colperm($A$), flipud($A$), fliplr($A$)

1.4.7 Orthogonal matrices

And last, but definitely not least, we consider the type of matrix that is essential and the favorite one in matrix computations, numerical analysis, and applied mathematics in general, due to its very important and remarkable properties.

An $m \times n$ matrix $Q$ is called orthogonal if

$$Q^TQ = I. \quad (1.26)$$

This definition implies (see Exercise 1.38) that the columns of an orthogonal matrix $Q_{m \times n}$ are orthonormal, in the $\| \cdot \|_2$ norm.

In particular, if the matrix $Q$ is square, then $Q^TQ = QQ^T = I$. In this case, the following statements are equivalent:

(a) $Q$ is orthogonal.
(b) $Q^{-1} = Q^T$.
(c) The columns of $Q$ are orthonormal.
(d) The rows of $Q$ are orthonormal.

■ EXAMPLE 1.21

Permutation matrices are obviously orthogonal. See the property (1.25).

■ EXAMPLE 1.22

The following matrices (the first for any real $\alpha$) are also orthogonal.

$$Q = \begin{bmatrix}
\cos \alpha & -\sin \alpha \\
\sin \alpha & \cos \alpha
\end{bmatrix}, \quad Q = \begin{bmatrix}
\frac{1}{\sqrt{3}} & -\frac{1}{\sqrt{2}} \\
\frac{1}{\sqrt{3}} & 0 \\
\frac{1}{\sqrt{3}} & \frac{1}{\sqrt{2}}
\end{bmatrix}.$$

Remark 1.19 In the complex case, we say a matrix $U_{m \times n}$ is unitary if

$$U^*U = I,$$
where as before $U^*$ denotes the adjoint of $U$.

The definition (1.26) and the statements (a)–(d) above, immediately set orthogonal matrices apart from most other matrices, and it would be enough to consider them very important in theory and applications. But there is more.

A very important and useful fact about orthogonal matrices is that they preserve the norm of a vector or a matrix. More precisely, we have the following theorem.

**Theorem 1.20** Let $Q$ be an orthogonal matrix. For any vector $x \in \mathbb{R}^n$ and an arbitrary matrix $A$ we have

$$\|Qx\|_2 = \|x\|_2, \quad \text{and} \quad \|QA\|_2 = \|A\|_2.$$  \hspace{1cm} (1.27)

**Proof.** First, observe that $\|Qx\|_2^2 = x^TQ^TQx = x^Tx = \|x\|_2^2$. Using this fact, we now have

$$\|QA\|_2 = \max_{\|x\|_2 = 1} \|QAx\|_2 = \max_{\|x\|_2 = 1} \|Ax\|_2 = \|A\|_2.$$  \hspace{1cm}

The next important property of an orthogonal matrix is especially related to what is known as sensitivity in the solution of several problems. First we need the following definition.

**Definition 1.21** The **condition number** of a square matrix $A$ is defined as

$$\text{cond}(A) = \|A\|_2 \|A^{-1}\|.$$  \hspace{1cm} (1.28)

This number represents how well or ill-posed a matrix can be, in the sense that how much we can rely on the computations performed with such a matrix, such as solving systems of equations, due to potential accumulation of round-off errors. Computations involving matrices with large condition numbers are potentially very inaccurate; roughly speaking, with $\text{cond}(A) \approx 10^k$, one can expect to lose about $k$ digits of precision in numerical computations.

For an arbitrary matrix $A$, the condition number has the property

$$\text{cond}(A) \geq 1,$$

and for any orthogonal matrix $Q$, we have

$$\text{cond}(Q) = 1,$$

the minimum possible. See Exercise 1.45.

It is important to remark that due to these properties, errors are not magnified when performing matrix computations with orthogonal matrices. For instance, when reducing a matrix $A$ to a triangular form, operations with orthogonal matrices such as successive
products of the type $QA$ are performed safely because the condition number of $QA$ is the same as that of $A$. In general, with products of the type $MA$, there is the actual risk that the condition number of $MA$ is much larger than that of $A$, making the computations numerically unstable and unreliable. This does not happen with orthogonal matrices. See also Exercises 1.42 and 1.43.

- **MATLAB command:** \texttt{cond(A)}

**Remark 1.22** It is important to stress again that orthogonal matrices are the ideal tools to use in numerical computations. They are stable, errors are not magnified, and there is no need to compute inverses. In addition, their condition number is 1, the best possible.

Now we introduce one of the most important orthogonal matrices that are extensively used in matrix computations and that we will use explicitly in Chapter 3. This matrix has the additional and convenient property of symmetry.

**Definition 1.23** (Householder matrix). Let $u$ be a unit vector in $\mathbb{R}^n$. Then the matrix

$$H = I - 2uu^T$$

(1.29)

is called Householder matrix.

**Theorem 1.24** The matrix $H$ in (1.29) is orthogonal and symmetric.

**Proof.** Since $(uu^T)^T = uu^T$, we have $H^T = H$ and hence $H$ is symmetric. In addition, since $u^Tu = 1$, we get

$$H^TH = (I - 2uu^T)(I - 2uu^T) = (I - 2uu^T)(I - 2uu^T)$$

$$= I - 4uu^T + 4uu^Tu^Tu = I,$$

and therefore $H$ is orthogonal.

$\Box$

A Householder matrix (1.29) is also known as Householder reflection because given an arbitrary vector $x \in \mathbb{R}^n$, the vector $Hx$ is a reflection of $x$ with respect to the hyperplane $u^\perp$, which is the set of all vectors perpendicular to $u$. In other words, the vectors $x$ and $Hx$ have exactly the same orthogonal projection onto that hyperplane. See e.g. Figure 1.2.

- **EXAMPLE 1.23**

  Let $u = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix}^T$. Then, the associated Householder matrix is

  $$H = I - 2uu^T = \begin{bmatrix} 0 & -1 \\ -1 & 0 \end{bmatrix}.$$
Now let \( x = [0 \quad 2]^T \); then \( Hx = [-2 \quad 0]^T \), which as we can see from Figure 1.2 is a reflection of \( x \) with respect to \( u^\perp \).

**Remark 1.25** For an arbitrary vector \( u \neq 0 \), a Householder matrix is defined as

\[
H = I - \frac{2uu^T}{u^Tu}.
\]  

(1.30)

**Orthogonal extensions.** One important application of Householder matrices is extending a given vector \( x \in \mathbb{R}^n \) to a set of \( n \) orthonormal vectors. The idea is to get an orthogonal matrix \( H \), with its first column being the vector \( x \), normalized, if necessary. This tool will be crucial, e.g., when proving several theorems in Chapter 3.

Let \( x \in \mathbb{R}^n \) be a vector with \( \|x\|^2 = x^Tx = 1 \). Define the vector

\[
u = x - e_1.
\]

First observe that \( u^Tx = x^Tx - e_1^Tx = 1 - e_1^Tx \) and that \( u^Tu = u^Tx - x^Te_1 - e_1^Tx + e_1^Te_1 = 2(1 - e_1^Tx) \), so that \( u^Tu = 2u^Tx \). Then,

\[
Hx = \left( I - \frac{2uu^T}{u^Tu} \right)x = x - \frac{2u(u^Tx)}{u^Tu} = x - \frac{2(u^Tx)u}{u^Tu} = x - u = e_1.
\]

The columns of this Householder matrix \( H \) are of course orthonormal, but according to the equality \( Hx = e_1 \), we have that \( x = H^Te_1 = He_1 \). That is, \( x \) is the first column of \( H \).

**Example 1.24**

Let \( x = \frac{1}{6} \begin{bmatrix} 4 \\ 0 \\ -4 \\ 2 \end{bmatrix} \). Take \( u = x - e_1 = \frac{1}{6} \begin{bmatrix} -2 \\ 0 \\ -4 \\ 2 \end{bmatrix} \).
Then, the matrix

\[
H = I - \frac{2uu^T}{u^Tu} = \frac{1}{6} \begin{bmatrix}
4 & 0 & -4 & 2 \\
0 & 6 & 0 & 0 \\
-4 & 0 & -2 & 4 \\
2 & 0 & 4 & 4
\end{bmatrix}
\]

is orthogonal, with \( x \) as its first column.

We will see some more applications of Householder matrices in Chapter 3.

1.5 VECTOR SPACES

Although vector spaces can be very general, we restrict our attention to vector spaces in finite dimensions, and unless otherwise stated, we will always consider the scalars to be real numbers.

We know that if we add, subtract, multiply, or divide (except division by zero) two real numbers, we get a real number again. We know that there is the number zero and the number one, with the properties that \( x + 0 = 0 + x = x \) and \( x \cdot 1 = 1 \cdot x = x \). We also know that the real numbers have the commutative and distributive properties; namely: \( x + y = y + x \), \( x \cdot y = y \cdot x \) and \( x + (y + z) = (x + y) + z \), and so on. Because of all these properties, working with real numbers is simple and flexible, and we usually take all these properties for granted.

As linear algebra deals not only with real numbers but also, and especially, with vectors and matrices, we would like to provide vectors and matrices with properties similar to those of real numbers, so that we can handle them in a simple manner. It happens that this is possible not only for vectors and matrices but also for several other objects. Thus, in general, we have the following definition.

**Definition 1.26** A set \( V \) of elements (called vectors) having two operations, addition and scalar multiplication, is a vector space if the following properties are satisfied:

1. \( u + v \in V \), for all \( u, v \in V \).
2. \( cu \in V \), for any scalar \( c \) and \( u \in V \).
3. \( u + v = v + u \), for all \( u, v \in V \).
4. \( u + (v + w) = (u + v) + w \), for all \( u, v, w \in V \).
5. There exists a zero vector such that \( u + 0 = u \), for all \( u \in V \).
6. \( 1u = u \), for all \( u \in V \).
7. For every \( u \) in \( V \), there exists a negative vector \(-u\), such that \( u + (-u) = 0 \).

8. \( c(u + v) = cu + cv \), for every scalar \( c \) and all \( u, v \in V \).

9. \( (c_1 + c_2)u = c_1u + c_2u \), for every scalars \( c_1, c_2 \) and all \( u \in V \).

10. \( c_1(c_2u) = (c_1c_2)u \), for every scalars \( c_1, c_2 \) and all \( u \in V \).

**EXAMPLE 1.25**

The real numbers with the usual addition and multiplication obviously form a vector space.

**EXAMPLE 1.26**

The set \( \mathbb{R}^n \) with the usual operations of addition and scalar multiplication forms a vector space. This is somehow the "canonical" example of a vector space. In fact, when we think about vectors, we tend to think about objects of the form \([x_1 \ldots x_n]^T\), with the usual geometric representation as arrows. But vectors are elements of more general sets.

**EXAMPLE 1.27**

The set of matrices with the usual addition and scalar multiplication (a real number times a matrix), forms a vector space.

**EXAMPLE 1.28**

The set of functions with the real numbers as their domain, and with the usual addition and scalar multiplication, forms a vector space.

Given a vector space \( V \), some of its subsets may be vector spaces on their own, with the same operations of addition and scalar multiplication. More formally, we can characterize them through the following definition.

**Definition 1.27** Let \( V \) be a vector space and \( U \) a nonempty subset of \( V \). Then, we say \( U \) is a vector subspace of \( V \) if

(i) \( u + v \in U \), for all \( u, v \in U \).

(ii) \( cu \in U \), for all scalars \( c \) and all \( u \in U \).
This definition simply says that operations performed with elements of \( U \) will generate vectors that will stay within \( U \).

\[ \text{Example 1.29} \]

Let \( V = \mathbb{R}^2 \). Then \( V \) is a vector space (see Example 1.26), and let \( U = \{(x, y) \in \mathbb{R}^2 : y = x\} \). In other words, \( U \) is the line \( y = x \). Clearly, \( U \) is a subset of \( V \). We need to verify the conditions in Definition 1.27.

(i) Let \( u, v \) be arbitrary elements in \( U \). Then, \( u = (u_1, u_1) \) and \( v = (v_1, v_1) \), for some real numbers, \( u_1, v_1 \). Thus, \( u + v = (u_1 + v_1, u_1 + v_1) \), and hence, \( u + v \in U \).

(ii) Let \( c \) be any real scalar, and \( u \) an arbitrary vector in \( U \). Then, \( c u = c(u_1, u_1) = (cu_1, cu_1) \), and therefore, \( c u \in U \).

Hence, \( U \) is a subspace of \( V \). See Figure 1.3.

\[ \text{Example 1.30} \]

Let again \( V = \mathbb{R}^2 \), and let \( U = \{(x, y) \in \mathbb{R}^2 : y = x + 1\} \). Then, \( U \) is not a vector subspace of \( V \) (see Exercise 1.49).

Observe that because condition (ii) in Definition 1.27 must be true for \( c = 0 \), a vector subspace (and in general any vector space) must contain the zero vector. The subset \( U \) in Example 1.30 fails to be a vector subspace because it does not contain the origin.

\[ \text{Example 1.31} \]

Let \( V \) be the vector space of all continuous functions on an interval \([a, b]\). Then, the set \( U \) of all polynomials of degree at most \( n \), defined on the interval \([a, b]\), is a vector subspace of \( U \).
EXAMPLE 1.32

Let $V$ be the vector space of square matrices of order $n$. Then, the set $U$ of symmetric matrices of order $n$ is a vector subspace of $V$.

EXAMPLE 1.33

(Sum of subspaces). If $U$ and $W$ are subspaces of $V$, then the set

$$U + W = \{ u + w : u \in U, w \in W \}$$

(called the sum of $U$ and $W$) is a subspace of $V$.

Here is a specific example of a sum of subspaces.

EXAMPLE 1.34

Let $V$ be the vector space of square real matrices of order 2, and let

$$U = \left\{ \begin{bmatrix} 0 & a \\ 0 & b \end{bmatrix} : a, b \in \mathbb{R} \right\}, \quad W = \left\{ \begin{bmatrix} 0 & 0 \\ c & d \end{bmatrix} : c, d \in \mathbb{R} \right\}$$

Clearly, $U$ and $W$ are subspaces of $V$, and

$$U + W = \left\{ \begin{bmatrix} 0 & e \\ f & g \end{bmatrix} : e, f, g \in \mathbb{R} \right\},$$

which is a vector subspace of $V$.

EXAMPLE 1.35

A very important example of subspace is that of the column space of a matrix, which is introduced in Section 1.8.

1.6 LINEAR INDEPENDENCE AND BASIS

The concepts introduced in this section are related to a fundamental notion in linear algebra: a linear combination of a set of vectors $v_1, \ldots, v_k$, which is an expression of the form

$$c_1 v_1 + \cdots + c_k v_k,$$
where the \( c_1, \ldots, c_k \) are scalars.

**Definition 1.28** Let \( V \) be a vector space, and \( U \subseteq V \) a vector subspace of \( V \). A set of vectors \( S = \{v_1, \ldots, v_k\} \) in \( V \) is said to span \( U \) if any vector \( x \in U \) can be written as a linear combination of elements in \( S \); that is,

\[
x = c_1v_1 + \cdots + c_kv_k,
\]

for some scalars \( c_1, \ldots, c_k \). This linear combination does not need to be unique.

**Example 1.36**

Let \( V = \mathbb{R}^2 \). Then, the set

\[
S_1 = \begin{bmatrix} -1 & 1 \end{bmatrix}^T, \begin{bmatrix} 0 & 1 \end{bmatrix}^T, \begin{bmatrix} 2 & 2 \end{bmatrix}^T
\]

spans \( V \); that is, any vector in \( \mathbb{R}^2 \) can be expressed as a linear combination of the vectors in \( S_1 \). For instance,

\[
\begin{bmatrix} 3 \\ -4 \end{bmatrix} = -5 \begin{bmatrix} -1 \\ 1 \end{bmatrix} + 3 \begin{bmatrix} 0 \\ 1 \end{bmatrix} - \begin{bmatrix} 2 \\ 2 \end{bmatrix}.
\]

(1.33)

However, the combination in (1.33) is not unique. Indeed, instead of the scalars \(-5, 3, -1\) in (1.33), we could also use \(-7, 7, -2\) and the new combination would still give us \( \begin{bmatrix} 3 & -4 \end{bmatrix}^T \). This indicates again that in general the linear combination (1.32) need not be unique.

**Example 1.37**

Again, let \( V = \mathbb{R}^2 \), and let \( S_2 = \{\begin{bmatrix} -1 & 1 \end{bmatrix}^T, \begin{bmatrix} 2 & 2 \end{bmatrix}^T\} \). Observe that \( S_2 \) is formed with only two vectors of \( S_1 \) from Example 1.36, but it also spans \( V \). In particular, we can write

\[
\begin{bmatrix} 3 \\ -4 \end{bmatrix} = -\frac{7}{2} \begin{bmatrix} -1 \\ 1 \end{bmatrix} - \frac{1}{4} \begin{bmatrix} 2 \\ 2 \end{bmatrix}.
\]

(1.34)

Thus, it seems that although \( S_1 \) does span \( V \), it has one element too many. This is not the case with \( S_2 \). Furthermore, the combination (1.34) is unique.

**Example 1.38**

If the sets of vectors \( T_1, T_2 \) span the subspaces \( V \) and \( W \) respectively, then the set \( T_1 \cup T_2 \) spans the sum subspace \( V + W \).

In fact, let \( T_1 = \{v_1, \ldots, v_p\} \) and \( T_2 = \{w_1, \ldots, w_q\} \), and let \( x \in V + W \) be arbitrary. Then \( x = v + w \), for some \( v \in V \) and \( w \in W \). Then, we can write

\[x = (\alpha_1v_1 + \cdots + \alpha_pv_p) + (\beta_1w_1 + \cdots + \beta_qw_q),\]

which implies that \( x \) lies in the span of \( T_1 \cup T_2 \).
The fundamental difference between the sets $S_1$ and $S_2$ of Examples 1.36 and 1.37 is in a central concept in linear algebra given in the following definition.

**Definition 1.29** A set of vectors $\{v_1, \ldots, v_n\}$ is said to be **linearly independent** if

$$c_1 v_1 + \ldots + c_n v_n = 0 \quad \text{implies} \quad c_1 = \cdots = c_n = 0. \quad (1.35)$$

In simple words, this definition says that if $v_1, \ldots, v_n$ are linearly independent, none of them can be written as a combination of the others. (Otherwise, say $c_1 \neq 0$; then, we can write $v_1$ as the linear combination $v_1 = -\frac{c_2}{c_1} v_2 - \cdots - \frac{c_n}{c_1} v_n$.)

**Example 1.39**

Let $V = \mathbb{R}^2$ and let $v_1 = [2 \ 1]^T$ and $v_2 = [1 \ 3]^T$. To show linear independence, assume that $c_1 v_1 + c_2 v_2 = [0 \ 0]^T$. Then, this equality gives the system

$$
\begin{align*}
2c_1 + c_2 &= 0 \\
c_2 + 3c_2 &= 0,
\end{align*}
$$

whose solution is $c_1 = c_2 = 0$. Therefore, $v_1$ and $v_2$ are linearly independent. Observe that in this case, linear independence means the vectors are not parallel or multiple of each other. See Figure 1.4.

**Example 1.40**

The set $S_1$ in Example 1.36 is not linearly independent. In fact, any of the vectors in $S_1$ can be written as a combination of the other two. For instance,

$$
\begin{bmatrix}
0 \\
1
\end{bmatrix} = \frac{1}{2} \begin{bmatrix}
-1 \\
1
\end{bmatrix} + \frac{1}{4} \begin{bmatrix}
2 \\
2
\end{bmatrix}.
$$
However, the set $S_2$ in Example 1.37 is linearly independent.

**Example 1.41**

Let $V = P_2$ be the vector space of real polynomials of degree at most 2. The set \( \{ f_1 = x^2 + 3x - 1, f_2 = x + 3, f_3 = 2x^2 - x + 1 \} \) is linearly independent. In fact, if \( c_1 f_1 + c_2 f_2 + c_3 f_3 = 0 \), we get the system
\[
\begin{align*}
    c_1 + 2c_3 &= 0 \\
    3c_1 + c_2 - c_3 &= 0 \\
    -c_1 + 3c_2 + c_3 &= 0,
\end{align*}
\]
whose solution is $c_1 = c_2 = c_3 = 0$.

A set $S$ of vectors could span a vector space $V$ but not necessarily be linearly independent, and some other set may be linearly independent but not span $V$. If a given set of vectors is linearly independent and at the same time spans a given vector space, then it is a very special set.

**Definition 1.30** A set of vectors $B = \{ v_1, \ldots, v_n \}$ is said to be a basis of a vector space $V$ if it spans $V$ and is linearly independent.

The most important application of this definition is that if $x$ is an arbitrary element of $V$, then it can be written in a unique form as a linear combination of the elements of the basis. That is,
\[
x = c_1 v_1 + \cdots + c_n v_n, \quad (1.36)
\]
for some unique scalars $c_1, \ldots, c_n$. In other words, the scalars or coefficients $c_i$ in (1.36) uniquely determine the vector $x$ on a given basis. For a different basis, there will be different coefficients, but again this combination is unique, on that basis. From Definition 1.30 we can say that a basis $B$ is a genuine (although not unique) representative of a vector space $V$.

**Example 1.42**

The set of vectors
\[
B = \{ e_1 = [1 \ 0 \ \cdots \ 0]^T, e_2 = [0 \ 1 \ \cdots \ 0]^T, \ldots, e_n = [0 \ 0 \ \cdots \ 1]^T \}
\]
is a basis for $V = \mathbb{R}^n$. It is known as the canonical or standard basis of $\mathbb{R}^n$.

Observe for instance that
\[
\begin{bmatrix} 8 \\ 3 \\ 4 \end{bmatrix} = 8 \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} + 3 \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} + 4 \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}.
\]
Thus, in the basis $\mathcal{B} = \{e_1, e_2, e_3\}$, the (unique) coefficients in the linear combination are exactly the entries of the given vector. That is why it is called the canonical or standard basis.

- MATLAB commands: `orth(A)`.

**EXAMPLE 1.43**

The set of vectors

$$\mathcal{B} = \{v_1 = [1 \quad 0 \quad 2]^T, \quad v_2 = [-2 \quad 1 \quad 8]^T, \quad v_3 = [0 \quad 2 \quad 0]^T\}$$

is a basis of $\mathbb{R}^3$, so that any vector $x \in \mathbb{R}^3$ can be written as a unique combination of such basis vectors. By instance,

$$\begin{bmatrix} 8 \\ 3 \\ 4 \end{bmatrix} = 6 \begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix} - 1 \begin{bmatrix} -2 \\ 1 \\ 8 \end{bmatrix} + 2 \begin{bmatrix} 0 \\ 2 \\ 0 \end{bmatrix}.$$ 

This means that on the basis $\mathcal{B}$ above, the vector $[8 \quad 3 \quad 4]^T$ is fully represented by its coordinates $\{6, -1, 2\}$. Similarly, on the basis $\mathcal{B}$ of Example 1.42, the same vector is represented by its corresponding coordinates $\{8, 3, 4\}$.

**EXAMPLE 1.44**

The set of matrices

$$\mathcal{B} = \left\{ \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \cdots, \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \right\} \quad (1.37)$$

is a basis for the vector spaces of squares matrices of order 4, and just as in Example 1.42, it is known as the canonical or standard basis of such space. This basis is particularly important (among other things) in image compression, which we study in Chapter 5.

**EXAMPLE 1.45**

The set $\mathcal{B} = \{f_1 = 1, f_2 = x, f_3 = x^2\}$ is a basis for the vector space $P_2$ of the polynomials of degree at most two.

It is not difficult to see that the set $\mathcal{B}$ in Example 1.45 spans $P_2$. Let us see that it is also linear independent. Assume that $c_1 f_1 + c_2 f_2 + c_3 f_3 = 0$. This gives $c_1 \cdot 1 + c_2 x + c_3 x^2 = 0$. Comparing coefficients, we immediately get $c_2 = c_3 = 0$, leaving us with $c_1 \cdot 1 = 0$. Thus, $c_1$ must also be zero. This proves $\mathcal{B}$ is linearly independent.
The next is a particular case of a very interesting and useful class of polynomials called **Bernstein polynomials**, which form a basis for the vector space \( P_n \) of polynomial of degree at most \( n \).

**Example 1.46**

The following set of polynomials:

\[
B_{0,3}(t) = (1 - t)^3, \\
B_{1,3}(t) = 3t(1 - t)^2, \\
B_{2,3}(t) = 3t^2(1 - t), \\
B_{3,3}(t) = t^3.
\]

is a basis for the vector space \( P_3 \) of the real polynomials of degree at most three. In fact, any polynomial of degree at most three can be expressed as a unique combination of the Bernstein polynomials above. By instance, if \( p(t) = 25t^3 - 21t^2 - 3t + 2 \), then

\[
p(t) = 2B_{0,3}(t) + B_{1,3}(t) - 7B_{2,3}(t) + 3B_{3,3}(t).
\]

We will study Bernstein polynomials and their applications in detail later in Section 1.12.

Intuition tells us that the real line \( \mathbb{R} \) is one-dimensional, the plane \( \mathbb{R}^2 \) is two-dimensional, and the space \( \mathbb{R}^3 \) is three-dimensional. This coincides with the number of vectors in the bases of each of these spaces; e.g., \( B = \{ e_1, e_2 \} \) is the canonical basis of \( \mathbb{R}^2 \). These ideas generalize into the following definition.

**Definition 1.31** If a vector space \( V \) has a basis consisting of \( m \) vectors, then we say the **dimension** of \( V \) is \( m \), and we write \( \dim(V) = m \).

**Example 1.47**

In Example 1.29, \( \dim(V) = 2 \) and \( \dim(U) = 1 \).

**Remark 1.32** If \( U \) is a subspace of \( V \), then \( \dim(U) \leq \dim(V) \), and if \( \dim(U) = \dim(V) \), then we must have \( U = V \).

**Example 1.48**

The basis in Example 1.45 has 3 elements; therefore, the dimension of the vector space \( P_2 \) of real polynomials of degree at most 2, is 3. Similarly, the basis in
Example 1.46 indicates that the dimension of the vector space $P_3$ is 4. In general, the dimension of $P_n$ is $n + 1$.

The next result expresses the relationship between the dimensions of two vector subspaces and that of its sum.

**Theorem 1.33** If $V$ and $W$ are subspaces of $U$, then

\[
\dim(V + W) = \dim(V) + \dim(W) - \dim(V \cap W). \tag{1.38}
\]

**Proof.** Let $\mathcal{R} = \{u_1, \ldots, u_r\}$ be a basis of $V \cap W$. For some positive integers $m$ and $n$, extend this set $\mathcal{R}$ to form the set $B_V = \{u_1, \ldots, u_r, v_1, \ldots, v_m\}$, a basis of $V$, and the set $B_W = \{u_1, \ldots, u_r, w_1, \ldots, w_n\}$, a basis of $W$. In Example 1.38, we saw that the set $B = B_V \cup B_W$ spans $V + W$. Let us prove that $B$ is also linearly independent. Thus, assume that

\[
\sum_{i=1}^{r} a_i u_i + \sum_{j=1}^{m} b_j v_j + \sum_{k=1}^{n} c_k w_k = 0 \tag{1.39}
\]

for some scalars $a_i, b_j, c_k$. This implies that

\[
\sum_{i=1}^{r} a_i u_i + \sum_{j=1}^{m} b_j v_j = -\sum_{k=1}^{n} c_k w_k.
\]

Since the left-hand side is a vector in $V$, then the vector $\sum_{k=1}^{n} c_k w_k$ lies in $V \cap W$ and therefore can be written as $\sum_{k=1}^{n} c_k w_k = \sum_{i=1}^{r} d_i u_i$, for some scalars $d_i$. Thus,

\[
\sum_{k=1}^{n} c_k w_k - \sum_{i=1}^{r} d_i u_i = 0.
\]

But since $B_W$ is linearly independent, we must have $c_k = 0$ and $d_i = 0$, for all $k = 1, \ldots, n$ and $i = 1, \ldots, r$. Then from (1.39), we get

\[
\sum_{i=1}^{r} a_i u_i + \sum_{j=1}^{m} b_j v_j = 0.
\]

Similarly, since $B_V$ is linearly independent, we must have $a_i = 0$ and $b_j = 0$, for all $i = 1, \ldots, r$ and $j = 1, \ldots, m$. Thus, $B$ is linearly independent and therefore a basis of $V + W$. Finally, using the definition of dimension, we get

\[
\dim(V + W) = r + m + n = (r + m) + (r + n) - r = \dim V + \dim W - \dim(V \cap W). \]

\[\square\]
1.7 ORTHOGONALIZATION AND DIRECT SUMS

Given a set of linearly independent vectors, it is always possible to transform it into an orthonormal set. In particular, this means that given a basis of a vector space, we can always transform such basis into an orthonormal one. In fact, due to some technical advantages, several applications start by considering only orthonormal bases. Here we introduce a well-known procedure to orthonormalize a linearly independent set of vectors.

**Gram–Schmidt process:** Given $m$ linearly independent vectors

$$\{v_1, \ldots, v_m\}$$

in $\mathbb{R}^n$, $(m \leq n)$, an orthonormal set of vectors $\{q_1, \ldots, q_m\}$ can be obtained by defining

$$q_1 = \frac{v_1}{\|v_1\|},$$

$$w_2 = v_2 - (v_2^T q_1) q_1, \quad q_2 = \frac{w_2}{\|w_2\|},$$

$$w_3 = v_3 - (v_3^T q_1) q_1 - (v_3^T q_2) q_2, \quad q_3 = \frac{w_3}{\|w_3\|},$$

$$\vdots$$

$$w_m = v_m - \sum_{k=1}^{m-1} (v_m^T q_k) q_k, \quad q_m = \frac{w_m}{\|w_m\|}.$$  

(1.40)

Observe that in general the vector $v_2 - q_1$ need not be orthogonal to $q_1$. This is why in the definition of $w_2$, the vector $q_1$ is first rescaled by a factor $v_2^T q_1$ so that $w_2$ is orthogonal to $q_1$. A similar idea is applied to the remaining vectors $w_i$. See Figure 1.5.

- **MATLAB command:** `orth(A)`. 

---

**Figure 1.5** Gram-Schmidt: $w_2$ is orthogonal to $q_1$. 

---
EXAMPLE 1.49

Let $v_1 = [2 \ 1 \ 0]^T$, $v_2 = [1 \ 0 \ -1]^T$, and $v_3 = [3 \ 7 \ -1]^T$. Then, following (1.40),

$$
\|v_1\| = \sqrt{5}, \quad q_1 = \frac{1}{\sqrt{5}} [2 \ 1 \ 0]^T.
$$

$$
w_2 = v_2 - \frac{2}{\sqrt{3}} q_1 = \frac{1}{\sqrt{3}} [1 \ 2 \ -5]^T, \quad \|w_2\| = \frac{\sqrt{30}}{3},
$$

$$
q_2 = \frac{w_2}{\|w_2\|} = \frac{1}{\sqrt{30}} [1 \ 2 \ -5]^T.
$$

$$
w_3 = v_3 - \frac{2}{\sqrt{30}} q_1 - \frac{22}{\sqrt{30}} q_2 = \frac{1}{3} [8 \ 16 \ 8]^T, \quad \|w_3\| = \frac{\sqrt{6}}{3},
$$

$$
q_3 = \frac{w_3}{\|w_3\|} = \frac{1}{\sqrt{6}} [1 \ 2 \ 1]^T.
$$

The set \{q_1, q_2, q_3\} is orthonormal. See Figure 1.6.

Direct Sums: In some special cases, every element of a vector space can be expressed as a unique sum of two vectors, each one lying in one of two different vector subspaces. More formally, we have the following definition.

Definition 1.34 Let $V$ be a vector space, and let $U$ and $W$ be two subspaces of $V$. Then, we say that $U$ and $W$ form a direct sum of $V$, denoted by $V = U \oplus W$, if

$$
U \oplus W = \{v \in V : v = u + w, \quad u \in U, w \in W, \quad U \cap W = 0\}. \quad (1.41)
$$

This definition implies that $V$ is spanned by $U$ and $W$, and that every vector in $V$ can be uniquely expressed as $v = u + w$.

Remark 1.35 If $V = U \oplus W$, then $U$ and $W$ are called complementary spaces. The vectors $u$ and $w$ in (1.41) are called the projection of $v$ onto $U$ along $W$ and the projection of $v$ onto $W$ along $U$, respectively.
EXAMPLE 1.50

Let $V = \mathbb{R}^3$, and consider two subspaces: $U$ the $xy$ plane, and $W$ the line spanned by $w = [1 \ 1 \ 1]^T$. Then, $V = U \oplus W$. In fact, any vector $v = [x \ y \ z]^T$ can be expressed as a unique sum of a vector of $U$ and a vector of $W$:

\[
\begin{bmatrix}
    x \\
    y \\
    z
\end{bmatrix}
= \begin{bmatrix}
    x - z \\
    y - z \\
    0
\end{bmatrix}
+ \begin{bmatrix}
    z \\
    z \\
    z
\end{bmatrix}.
\]

Observe that $U \cap W = \{0\}$.

EXAMPLE 1.51

In Example 1.34, we can observe that $U \cap W \neq \{0\}$ and, therefore, this implies that $U + W$ is not a direct sum.

For a very useful example of direct sums, consider now an arbitrary vector subspace $U$ of a vector space $V$. We can define another vector subspace, the orthogonal complement of $U$, which consists of all vectors that are orthogonal to each vector in $U$ (and it is usually called the "perp" of $U$):

\[ U^\perp = \{ v \in V : v^T u = 0, \ \forall u \in U \}. \tag{1.42} \]

A direct sum of subspaces can be obtained by using the perp of a vector subspace. In fact, we have the following result.

**Theorem 1.36** Let $V$ be a vector space, and let $U$ be an arbitrary vector subspace of $V$. Then,

\[ V = U \oplus U^\perp. \]

**Proof.** Let $v$ be an arbitrary vector in $V$, and let $\{u_1, \ldots, u_m\}$ be an orthonormal basis of $U$. Then, similar to the way we defined the vectors in the Gram–Schmidt process (1.40), let

\[ u = (v^T u_1)u_1 + \cdots + (v^T u_m)u_m. \]

Then, the vector $w = v - u$ is orthogonal to each $u_i$, $i = 1, \ldots, m$, and therefore, it is in $U^\perp$. Hence, $v = u + w$, where $u \in U$ and $w \in U^\perp$. Also, it is clear that the only intersection of $U$ and $U^\perp$ is the zero vector: for if $u \in U$ and $w \in U^\perp$, then

\[ \|u\|^2 = u^T u = 0. \]

This proves that in fact, $V = U \oplus U^\perp$. \qed
**EXAMPLE 1.52**

If \( V = \mathbb{R}^3 \), and \( U \) is the \( XY \)-plane, then \( U^\perp \) is the \( Z \)-axis. Then, it is clear that \( V = U \oplus U^\perp \). In fact, every vector \( [x \ y \ z]^T \) can be uniquely written as \( [x \ y \ z]^T = [x \ y \ 0]^T + [0 \ 0 \ z]^T \).

**EXAMPLE 1.53**

A matrix \( A_{n \times n} \) is said to be skew symmetric if \( A = -A^T \). If we let \( S \) be the subspace of \( n \times n \) symmetric matrices and \( K \) the subspace of \( n \times n \) skew symmetric matrices, then \( \mathbb{R}^{n \times n} = S \oplus K \).

Here is an example of a skew symmetric matrix: \( A = \begin{bmatrix} 0 & 2 & 5 \\ -2 & 0 & -4 \\ -5 & 4 & 0 \end{bmatrix} \).

**Note.** In the complex case, we say that a matrix is \( A \) skew Hermitian if \( A = -A^* \).

Later on we will see some particular and useful examples of direct sums.

### 1.8 COLUMN SPACE, ROW SPACE, AND NULL SPACE

Associated with every \( m \times n \) matrix \( A \), there are three fundamental vector spaces: the column space of \( A \), \( \text{col}(A) \), the row space of \( A \), \( \text{row}(A) \), and the null space of \( A \), \( N(A) \), which are important in several applications, e.g. in the solution or least squares solution of a linear system \( Ax = b \), in web page ranking, and in information retrieval, topics that we consider in detail later on. Here we will use the concepts of linear independence, span, and so on, that we have introduced in the previous sections.

**Definition 1.37** The column space of a matrix \( A_{m \times n} \) is defined as the vector subspace of \( \mathbb{R}^m \) spanned by the columns of \( A \).

This definition says that \( \text{col}(A) \) is the vector subspace formed by the vectors obtained through all possible linear combinations of the columns of \( A \).

Another appropriate and useful definition of the column space is given as
\[
\text{col}(A) = \{ y \in \mathbb{R}^m : y = Ax, \text{ for } x \in \mathbb{R}^n \}. \tag{1.43}
\]

In other words, the column space of \( A \) is the set of vectors that can be expressed as \( Ax \), for some vector \( x \in \mathbb{R}^n \). In this sense, the column space is also known as the range of \( A \), or the image of \( A \), and it is usually denoted as \( R(A) \), or \( \text{Im}(A) \).

That is,

\[
\text{col}(A) = R(A).
\]

**Remark 1.38** Observe that each column of \( A_{m \times n} \) is an \( m \)-dimensional vector; that is why the \( \text{col}(A) \) is a subspace of \( \mathbb{R}^m \).

**Remark 1.39** Combining Definition 1.37 and (1.43), we conclude that for any \( x \in \mathbb{R}^n \), \( Ax \) is a combination of the columns of \( A \).

**Definition 1.40** The dimension of the column space of a matrix \( A_{m \times n} \) (the number of linearly independent columns of \( A \)) is called the rank of \( A \).

- **MATLAB command:** \texttt{rank(A)}.

**EXAMPLE 1.54**

Let \( A = \begin{bmatrix} 1 & -2 \\ 5 & 3 \\ -4 & 7 \end{bmatrix} \), and let \( x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \) be arbitrary. Then,

\[
Ax = \begin{bmatrix} x_1 - 2x_2 \\ 5x_1 + 3x_2 \\ -4x_1 + 7x_2 \end{bmatrix} = x_1 \begin{bmatrix} 1 \\ 5 \\ -4 \end{bmatrix} + x_2 \begin{bmatrix} -2 \\ 3 \\ 7 \end{bmatrix}.
\]

This clearly illustrates that in fact \( Ax \) is a combination of the columns of \( A \), for any \( x \), as noted in Remark 1.39. We also observe that in this case, since the two columns of \( A \) are linearly independent, the \( \text{col}(A) \) is 2-dimensional; that is, geometrically it is a plane in \( \mathbb{R}^3 \) spanned by the two columns of the matrix \( A \), and thus, \( \text{rank}(A) = 2 \).

**EXAMPLE 1.55**

Let \( A = \begin{bmatrix} 2 & -1 & 0 \\ 0 & 1 & -1 \\ 2 & 0 & -1 \end{bmatrix} \). By definition, the \( \text{col}(A) \) is the set of vectors of the form \( Ax \), for any \( x \). In this case,

\[
Ax = \begin{bmatrix} 2x_1 - x_2 \\ x_2 - x_3 \\ 2x_1 - x_3 \end{bmatrix} = x_1 \begin{bmatrix} 2 \\ 0 \\ 2 \end{bmatrix} + x_2 \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix} + x_3 \begin{bmatrix} 0 \\ 0 \\ -1 \end{bmatrix}.
\]
It is very important to observe in Example 1.55 that although it is true that the col(A) is spanned by (all) the columns of A, and A has three columns, they are not linearly independent. In fact, it is easy to verify that for instance, the third column is a combination of the first two, and that these first two columns are linearly independent. This means that the rank of A, and the dimension of col(A) is two. Thus, we can say that the col(A) is also spanned by just the first two columns of A, and hence it has dimension two. In other words, for Example 1.55, the column space of A is the subspace of vectors that can be obtained via linear combinations of the form

$$\text{col}(A) = c_1 \begin{bmatrix} 2 \\ 0 \\ 2 \end{bmatrix} + c_2 \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix},$$

where $c_1, c_2$ are arbitrary scalars. The col(A) is therefore a plane in $\mathbb{R}^3$.

**Remark 1.41** In Example 1.55, we have expressed col(A) as a combination of the first two columns, but it can also be expressed as a combination of the last two, or the first and the last. In any case, there are only two linearly independent columns.

- **MATLAB command:** orth

Naturally, we can also consider the subspace spanned by the columns of $A^T$, or in other words, by the rows of A, and we denote this space by row(A).

**Definition 1.42** The **row space** of a matrix $A_{m \times n}$ is defined as the vector subspace of $\mathbb{R}^n$ spanned by the rows of A.

**Example 1.56**

For the matrix $A$ of Example 1.54, we have seen that $\dim \text{col}(A) = 2$. If we now denote the first, second, and third rows of that matrix with $r_1, r_2,$ and $r_3$ respectively, then we can prove that $r_2 = -47 r_1 - 13 r_3$, that is, $r_2$ is a linear combination of the other two rows, which means that the three rows are not linearly independent. And in fact, we must have $\dim \text{row} (A) = 2$. In this case, the row space of A is the set of linear combinations of the form

$$\text{row}(A) = c_1 \begin{bmatrix} 1 \\ -2 \end{bmatrix} + c_2 \begin{bmatrix} -4 \\ 7 \end{bmatrix},$$

where $c_1, c_2$ are arbitrary scalars. Thus, row(A) spans the whole space $\mathbb{R}^2$.

It is a very remarkable fact that for any matrix $A_{m \times n}$, the row space has the same dimension as the column space. More formally,
Theorem 1.43 Let $A$ be an $m \times n$ matrix. Then,

$$\text{rank}(A) = \dim \text{col}(A) = \dim \text{row}(A).$$

(1.44)

Proof. Denote the rows of $A$ by $r_1, r_2, \ldots, r_m$, and let $\dim \text{row}(A) = k$. Assume that $B = \{v_1, \ldots, v_k\}$ is a basis of $\text{row}(A)$. Then, each row vector $r_i$, $i = 1, \ldots, m$ can be written as

$$r_i = c_{i1}v_1 + c_{i2}v_2 + \cdots + c_{ik}v_k,$$

for some scalars $c_{i1}, \ldots, c_{ik}$. But the $j$-th component of each row vector $r_i$ is nothing else than $a_{ij}$. Thus, with respect to each $j$-th component ($j = 1, \ldots, n$), the above equation can be written as

$$a_{ij} = c_{i1}v_{1j} + c_{i2}v_{2j} + \cdots + c_{ik}v_{kj}, \quad i = 1, \ldots, m,$$

which is the same as

$$\begin{bmatrix}
  a_{1j} \\
  a_{2j} \\
  \vdots \\
  a_{mj}
\end{bmatrix}
= v_{1j} \begin{bmatrix}
  c_{11} \\
  c_{21} \\
  \vdots \\
  c_{m1}
\end{bmatrix}
+ v_{2j} \begin{bmatrix}
  c_{12} \\
  c_{22} \\
  \vdots \\
  c_{m2}
\end{bmatrix}
+ \cdots + v_{kj} \begin{bmatrix}
  c_{1k} \\
  c_{2k} \\
  \vdots \\
  c_{mk}
\end{bmatrix}, \quad j = 1, \ldots, n.$$

This tells us that the columns of $A$ are linear combinations of the $k$ vectors $c$. Therefore,

$$\dim \text{col}(A) \leq k = \dim \text{row}(A).$$

By applying the ideas above to $A^T$, we can also show that $\dim \text{row}(A) \leq \dim \text{col}(A)$, and therefore conclude that indeed $\dim \text{col}(A) = \dim \text{row}(A)$.

We next state an important result on the rank of product of matrices.

Theorem 1.44 Let $A$ and $B$ be two arbitrary matrices whose product $AB$ is well defined. Then,

(a) $\text{rank}(AB) \leq \text{rank}(A)$, and $\text{rank}(AB) \leq \text{rank}(B)$.

(b) If $B$ is nonsingular, then $\text{rank}(AB) = \text{rank}(A)$.

If $A$ is nonsingular, then $\text{rank}(AB) = \text{rank}(B)$.

Proof. (a) If we denote the columns of $B$ by $b_1, \ldots, b_n$, then we have

$$AB = A[b_1 \cdots b_n] = [Ab_1 \cdots Ab_n].$$

By recalling Remark 1.39, we observe that the columns of $AB : Ab_1, \ldots, Ab_n$ are linear combinations of the columns of $A$, and therefore, we must have that $\text{rank}(AB) \leq \text{rank}(A)$. The second inequality is proved similarly using rows.

(b) If $B$ is nonsingular, we have

$$\text{rank}(A) = \text{rank}(ABB^{-1}) \leq \text{rank}(AB).$$
By combining this with the first inequality in (a), we conclude that \( \text{rank}(AB) = \text{rank}(A) \). The proof of the last statement is similar.

We now define the null space of \( A \), denoted as \( N(A) \).

**Definition 1.45** The null space of a matrix \( A_{m \times n} \) is defined as the vector subspace of \( \mathbb{R}^n \):

\[
N(A) = \{ x \in \mathbb{R}^n : A x = 0 \}.
\] (1.45)

**Example 1.57**

Let \( A = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 4 & 6 \\ 3 & 6 & 9 \end{bmatrix} \). From the definition, we see that the null space is the set \( \{ [x_1 \ x_2 \ x_3]^T : x_1 + 2x_2 + 3x_3 = 0 \} \). Geometrically, this is a plane (hence, \( N(A) \) is two-dimensional) passing through \((0,0,0)\) and with normal vector \([1 \ 2 \ 3]^T\). On the other hand, since the three columns of \( A \) are just multiples of each other, it is clear that the \( \text{col}(A) \) is one-dimensional and is spanned by the same vector \([1 \ 2 \ 3]^T\). So, for this particular example, \( \text{col}(A) \) and \( N(A) \) are orthogonal.

**Note:** It is very interesting to learn that in general, \( \text{col}(A) \) and \( N(A^T) \) are orthogonal. In the example above, the matrix \( A \) is symmetric \( (A^T = A) \); that is why \( \text{col}(A) \) is also orthogonal to \( N(A) \). See Theorem 1.48 below and the discussion after it.

Furthermore, there are some orthogonality relationships between these subspaces as the following theorem states.

**Theorem 1.46** Let \( A \) be an arbitrary \( m \times n \) matrix. Then,

\[
\text{col}(A)^\perp = N(A^T), \quad \text{and} \quad \text{col}(A^T)^\perp = N(A).
\] (1.46)

**Proof.** Let \( v \) be any vector in \( \text{col}(A)^\perp \). For an arbitrary vector \( x \in \mathbb{R}^n \), we know that \( Ax \in \text{col}(A) \); then we have

\[
(Ax)^T v = 0 \iff x^T (A^T v) = 0 \iff A^T v = 0 \iff v \in N(A^T).
\]

This implies that, in fact, \( \text{col}(A)^\perp = N(A^T) \). The second part of the theorem is proved similarly.

In general, an arbitrary matrix \( A_{m \times n} \) generates the four subspaces: \( \text{col}(A) \), \( \text{col}(A^T) \), \( N(A) \), and \( N(A^T) \). These vector subspaces decompose \( \mathbb{R}^m \) and \( \mathbb{R}^n \) in direct sums, and their dimensions are related by closed formulas.
We can now state the **Fundamental Theorem of Linear Algebra**.

**Theorem 1.47** Let $A_{m \times n}$ be an arbitrary matrix. Then,

\[ \mathbb{R}^m = \text{col}(A) \oplus \text{N}(A^T), \quad \mathbb{R}^n = \text{col}(A^T) \oplus \text{N}(A). \] (1.47)

**Proof.** The vector spaces $\text{col}(A)$ and $\text{N}(A^T)$ are subspaces of $\mathbb{R}^m$. Similarly, the vector spaces $\text{col}(A^T)$ and $\text{N}(A)$ are subspaces of $\mathbb{R}^n$. Now, to obtain the result, simply combine Theorems 1.36 and 1.46.

\[ \Box \]

An immediate consequence of Theorems 1.43 and 1.47 is the following corollary.

**Corollary 1.8.1** If $A_{m \times n}$ is a matrix with $\text{rank}(A) = r$, then

\[ \dim \text{col}(A) = \dim \text{col}(A^T) = r, \quad \dim \text{N}(A) = n - r, \quad \dim \text{N}(A^T) = m - r. \] (1.48)

At the same time, Corollary 1.8.1 implies the following important result.

**Theorem 1.48** (Dimension Formula). Consider an $m \times n$ matrix $A$. Then,

\[ \dim \text{col}(A) + \dim \text{N}(A) = n. \] (1.49)

**EXAMPLE 1.58**

Let $A = \begin{bmatrix} 1 & 4 & 7 \\ 2 & 5 & 8 \\ 3 & 6 & 9 \end{bmatrix}$. Observe that we can write the third column $a_3$ as a combination of the first two $a_1, a_2$ as $a_3 = 2a_2 - a_1$. And since these first two columns are linearly independent, this implies that $\text{col}(A)$ is 2-dimensional. Geometrically then, $\text{col}(A)$ is the plane $-x + 2y - z = 0$ passing through the origin, with normal vector $[-1 \ 2 \ -1]^T$. From Theorem 1.48, the null space $\text{N}(A)$ is one-dimensional. In fact, by solving the system $Ax = 0$, we observe that all solutions are of the form $[x_1 \ x_2 \ x_3]^T$, with $x_2 = -2x_1, x_3 = x_1$. That is, $\text{N}(A)$ is a line spanned by the vector $[-1 \ 2 \ -1]^T$. Using this vector, we observe that $\text{N}(A)$ is orthogonal to every row of $A$ and therefore $\text{N}(A) \perp \text{col}(A^T)$, in agreement with Theorem 1.46, and finally, because $\text{N}(A) = \text{col}(A^T)$, we also have $\mathbb{R}^3 = \text{col}(A^T) \oplus \text{N}(A)$, as in (1.47).
**Example 1.59**

Let \( A = \begin{bmatrix} 2 & 0 & 1 & 3 & 0 \\ 0 & 3 & 1 & 2 & 0 \\ 2 & 3 & 2 & 5 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \). In this case the columns 1, 2, and 5 are linearly independent and they form a basis of \( \text{col}(A) \). That is, any vector in \( \text{col}(A) \) (in particular, the third and fourth columns of \( A \)) can be expressed as a unique combination of columns 1, 2, and 5. This means that geometrically, \( \text{col}(A) \) is a 3-dimensional hyperplane (in \( \mathbb{R}^5 \)) spanned by those three columns, and therefore \( N(A) \) is a two-dimensional subspace of \( \mathbb{R}^5 \).

As stated before, one of the applications of column spaces, row spaces, and null spaces is in the solution of linear systems. We can now put together the links between the solution of \( Ax = b \) and the vector spaces \( \text{col}(A) \) and \( N(A) \) in a theorem.

**Theorem 1.49** Given an \( m \times n \) matrix \( A \), consider solving the system \( Ax = b \). Then,

1. **(Existence)** The system has a solution if and only if \( b \in \text{col}(A) \).
2. **(Uniqueness)** The system has at most one solution for every \( b \) if and only if \( N(A) = \{0\} \).

**Proof.** *Existence:* We have already seen that for arbitrary \( x \), the vector \( Ax \in \text{col}(A) \). Then, for \( Ax = b \) to have a solution, \( b \) must lie in the same subspace.

*Uniqueness:* If \( N(A) \neq \{0\} \), then besides \( x = 0 \), there is another solution to \( Ax = b \), with \( b = 0 \), which is a contradiction. On the other hand, assuming \( N(A) = \{0\} \), if there is a \( b \) for which \( Ax = b \) has more than one solution, that is \( Ax_1 = b \) and \( Ax_2 = b \), with \( x_1 \neq x_2 \), then \( A(x_1 - x_2) = Ax_1 - Ax_2 = b - b = 0 \), which means that \( x_1 - x_2 = 0 \), or \( x_1 = x_2 \).

**Remark 1.50** Observe that by combining Theorem 1.48 with the the uniqueness part of Theorem 1.49, we conclude that the system \( Ax = b \) has a unique solution if and only if \( \text{dim} \text{col}(A) = n \), which means \( A \) has to be a full rank matrix, or which is the same, all the columns of \( A \) must be linearly independent.

### 1.8.1 Linear transformations

Given two arbitrary vector spaces \( V \) and \( W \), we can define a function

\[ T : V \rightarrow W \]

that for any two vectors \( u, v \in V \) and any scalar \( c \), it satisfies
(a) $T(u + v) = T(u) + T(v)$,
(b) $T(cu) = cT(u)$.

Such a function $T$ is called a **linear transformation** from $V$ to $W$. In the case $V = W$, it is called a **linear operator**.

These transformations play an important role in the theory of linear algebra and we refer the reader to any standard textbook in linear algebra for a full discussion on general linear transformations. For completion, here we just give a few remarks on how some terminology for a particular class of linear transformations is closely related to the theory introduced in this chapter.

More precisely, we consider linear transformations

$$T : \mathbb{R}^n \longrightarrow \mathbb{R}^m. \quad (1.50)$$

Given a vector $x \in \mathbb{R}^n$, the linear transformation $T$ transform such vector $x$ into a vector $y \in \mathbb{R}^m$ so that $T(x) = y$.

The important fact here is that any linear transformation of the form (1.50) has a matrix representation and vice versa, any matrix $A_{m \times n}$ corresponds to a linear transformation (1.50).

**Example 1.60**

Define

$$T(x_1, x_2, x_3) = (2x_1 - x_2, 3x_1 + 4x_2 - x_3).$$

We can write this transformation as $T(x_1, x_2, x_3) = (y_1, y_2)$, where

$$y_1 = 2x_1 - x_2, \quad y_2 = 3x_1 + 4x_2 - x_3.$$

Or, in matrix-vector notation as

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} 2 & -1 & 0 \\ 3 & 4 & -1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}.$$  

That is,

$$y = Ax.$$

Just as the transformation $T$, the matrix $A$ takes a 3-dimensional vector $x$, and via multiplication, it transforms it into a 2-dimensional vector $y$.

The example above illustrates the fact that

To any linear transformation $T : \mathbb{R}^n \rightarrow \mathbb{R}^m$, there corresponds a matrix $A_{m \times n}$. 

**Remark 1.51** The correspondence between a linear transformation and a matrix is true not only for the spaces \( \mathbb{R}^n \) and \( \mathbb{R}^m \), but for any vector spaces \( V \) of dimension \( n \) and \( W \) of dimension \( m \).

Since there is a correspondence between linear transformations and matrices, then the same is true for some definitions in the two cases.

Given a linear transformation

\[
T : V \rightarrow W,
\]

we say that the vector space \( V \) is the **domain** and the vector space \( W \) is the **codomain** of \( T \).

The set

\[
R(T) = \{ y \in W : y = T(x) \}
\]

is called the **range** or **image** of \( T \).

The set

\[
\text{Ker}(T) = \{ x \in V : T(x) = 0 \}
\]

is called the **kernel** of \( T \).

From these definitions, we immediately see that the range of \( T \) is exactly the column space of the corresponding matrix \( A \). Similarly, the kernel of \( T \) is exactly the null space of the matrix \( A \).

<table>
<thead>
<tr>
<th>Linear Transformations</th>
<th>Matrices</th>
</tr>
</thead>
<tbody>
<tr>
<td>range</td>
<td>column space</td>
</tr>
<tr>
<td>kernel</td>
<td>null space</td>
</tr>
</tbody>
</table>

Accordingly, the **rank** of the linear transformation \( T \) is defined as the dimension of its range. The dimension of the kernel is called the **nullity**.

Some of the definitions and results on matrices we have introduced so far, and some others presented later, have their equivalent counterpart in the context of linear transformations. We speak of zero, identity linear transformations, similarity, eigenvalues, and so on. For a linear transformation \( T : V \rightarrow W \), where \( V \) is a vector space of dimension \( n \) and \( W \) a vector space of dimension \( m \), the dimension formula (1.49) now reads as

\[
\dim R(T) + \dim \text{Ker}(T) = n. \tag{1.51}
\]

**Remark 1.52** The set of all linear operators on \( \mathbb{R}^n \), denoted by \( L(\mathbb{R}^n) \), forms a normed vector space, with

\[
\| T \| = \max_{\| x \| = 1} \| T(x) \|.
\]
1.9 ORTHOGONAL PROJECTIONS

Definition 1.53 A projection matrix is a square matrix $P$ such that $P^2 = P$. A projection matrix onto a subspace $S$ is a projection matrix for which $\text{range}(P) = S$.

If $P$ is a projection matrix onto a subspace $S$, then the condition $P^2 = P$ says that for an arbitrary vector $x$, since $Px$ is already in $S$, a new application of $P$, that is $P(Px)$, should not move the vector $Px$ at all: $P(Px) = Px$.

**EXAMPLE 1.61**

Consider the vector subspace $S$ of $\mathbb{R}^2$ to be the line $y = x$. Then, a projection matrix onto $S$ is given by

$$P = \begin{bmatrix} 0 & 1 \\ 0 & 1 \end{bmatrix}.$$  

In fact, for an arbitrary vector $v = [x \ y]^T$, we get $Pv = [y \ y]^T$, which lies on $S$. Furthermore, it is easy to verify that $P^2 = P$.

**EXAMPLE 1.62**

Let $S$ be again the subspace of $\mathbb{R}^2$ determined by the line $y = x$, and let $P = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$. Then, for any vector $x = [x_1 \ x_2]^T$, the vector $Px$ lies on $S$; that is, $\text{ran}(P) = S$. However, $P$ is not a projection because $P^2 \neq P$. In fact, if we take say $x = [0 \ 1]^T$, then (see Figure 1.7)

$$Px = [1 \ 1]^T \quad \text{but} \quad P^2(x) = P(Px) = [2 \ 2]^T \neq [1 \ 1]^T.$$  

**EXAMPLE 1.63**

Let $S$ be the subspace of $\mathbb{R}^3$ spanned by $u_1 = [2 \ 1 \ -1]^T$ and $u_2 = [0 \ -1 \ 1]^T$. Then the matrix

$$P = \frac{1}{4} \begin{bmatrix} 4 & 0 & 0 \\ 2 & 1 & 1 \\ -2 & 3 & 3 \end{bmatrix}$$

is a projection matrix onto $S$. One can verify that for an arbitrary vector $v \in \mathbb{R}^3$, the vector $Pv$ can be written as a linear combination of $u_1$ and $u_2$.  

The projection matrices that probably have more applications are those that are also orthogonal.

**Definition 1.54** An orthogonal projection matrix is a projection matrix $P$ for which $P^T = P$.

**Example 1.64**

The matrix

$$P = \begin{bmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{bmatrix}$$

is an orthogonal projection. It clearly satisfies $P^2 = P$ and $P^T = P$. It is in fact an orthogonal projection onto the subspace $S$ determined by the line $y = x$ (see Example 1.62).

**Remark 1.55** It is important to note that an orthogonal projection matrix is not necessarily an orthogonal matrix. However, we will use orthogonal matrices to construct orthogonal projection matrices.

**Explicit formulas for orthogonal projections.** We want to see how to obtain an explicit expression for an orthogonal projection onto a subspace $S$, starting with the particular case when $S = \text{col}(A)$.

Later we will see that for a general $m \times n$ matrix $A$, with $m > n$, the least squares solution to the system $Ax = b$ is given by the solution of the so-called normal equations $A^T A x = A^T b$. This is nothing else but a consequence of projecting $b$ orthogonally onto $\text{col}(A)$, as it will be explained in detail in Chapter 4. Observe also that from the normal equations, if the columns of $A$ are linearly independent, we can write (see Exercise 1.54)

$$x = (A^T A)^{-1} A^T b.$$
The matrix
\[ A^\dagger = (A^T A)^{-1} A^T \]  
(1.52)
is called the **pseudo-inverse matrix** of \( A \).

Now, since \( Ax \) is in \( \text{col}(A) \), and \( Ax = A(A^T A)^{-1} A^T b \), then the matrix
\[ P = A(A^T A)^{-1} A^T \]  
(1.53)
has taken the vector \( b \) onto \( \text{col}(A) \), so we suspect such a matrix is an orthogonal projection onto \( \text{col}(A) \). In fact,
\[ P^2 = A(A^T A)^{-1} A^T A(A^T A)^{-1} A^T = A(A^T A)^{-1} A^T = P, \]
and since \( A^T A \) is symmetric,
\[ P^T = A(A^T A)^{-T} A^T = A(A^T A)^{-1} A^T = P. \]

**Example 1.65**

Consider the matrix \( A = \begin{bmatrix} 2 & 1 & 0 \\ 1 & 2 & 1 \\ 0 & 1 & 2 \\ 1 & 2 & 1 \end{bmatrix} \). Then, the matrix
\[ P = A(A^T A)^{-1} A^T = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1/2 & 0 & 1/2 \\ 0 & 0 & 1 & 0 \\ 0 & 1/2 & 0 & 1/2 \end{bmatrix} \]
will project an arbitrary vector in \( \mathbb{R}^4 \) orthogonally onto the \( \text{col}(A) \).

**Remark 1.56**

1) If the matrix \( A \) is square and nonsingular, then its pseudo-inverse coincides with its inverse matrix. In fact, in this case, we have that \( A^\dagger = (A^T A)^{-1} A^T = A^{-1} A^{-T} A^T = A^{-1} \).

2) \( A^\dagger \) satisfies \( A^\dagger A = I \) and \( A A^\dagger = P \).

- **Matlab command:** `pinv(A)`

The definition (1.53) of the orthogonal projection onto \( \text{col}(A) \), although theoretically valid and illustrative, is not useful in practice due especially to the fact that an inverse has to be computed, which is something that must always be avoided, if possible. A more practical, and by far more efficient, way to compute a projection matrix onto \( \text{col}(A) \) is through orthogonal matrices.
The idea is to compute an orthonormal basis \( \{ q_1, \ldots, q_n \} \) of \( \text{col}(A) \) and define the orthogonal matrix \( Q = [q_1 \cdots q_n] \). Then the projection is
\[
P = QQ^T. \tag{1.54}
\]

Furthermore, this approach is in fact true for a projection matrix onto any vector subspace \( S \). That is, the orthogonal projection matrix onto a vector subspace \( S \) can be defined as in (1.54), where \( Q \) is a matrix whose columns form an orthonormal basis of \( S \). See Exercise 1.62.

\section*{Example 1.66}

Let \( S \) be the subspace spanned by \( v_1 = [1 \ 2 \ 3]^T \), and \( v_2 = [1 \ 1 \ 1]^T \). We can orthonormalize \( \{v_1, v_2\} \), say, by using Gram–Schmidt (1.40) to obtain the vectors \( q_1 = [1 \ 2 \ 3]^T/\sqrt{14} \), \( q_2 = [4 \ 1 \ -2]^T/\sqrt{21} \). Then, we define the matrix \( Q \) with \( q_1 \) and \( q_2 \) as first column and second column respectively, so that the projection matrix onto \( S \) is given by
\[
P = QQ^T = \frac{1}{6} \begin{bmatrix} 5 & 2 & -1 \\ 2 & 2 & 2 \\ -1 & 2 & 5 \end{bmatrix},
\]

where \( Q = [q_1 \ q_2] \). Thus, given any vector \( x \in \mathbb{R}^3 \), the vector \( Px \) will be in the subspace \( S \).

\section*{Example 1.67}

Consider the matrix \( A \) of Example 1.65. The vectors
\[
v_1 = [1 \ 0 \ 0 \ 0]^T, \ v_2 = [0 \ 1 \ 0 \ 1]^T, \ v_3 = [0 \ 0 \ 1 \ 0]^T
\]
form a basis of \( \text{col}(A) \). They can be orthonormalized using Gram–Schmidt to get
\[
q_1 = [1 \ 0 \ 0 \ 0]^T, \ q_2 = [0 \ 1/\sqrt{2} \ 0 \ 1/\sqrt{2}]^T, \ q_3 = [0 \ 0 \ 1 \ 0]^T.
\]

Now define the matrix \( Q = [q_1 \ q_2 \ q_3] \). Then, the projection matrix onto \( \text{col}(A) \) is
\[
P = QQ^T = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1/2 & 0 & 1/2 \\ 0 & 0 & 1 & 0 \\ 0 & 1/2 & 0 & 1/2 \end{bmatrix},
\]

which is exactly what we had obtained before by using \( P = A(A^TA)^{-1}A^T \).

Given an orthogonal projection \( P \) onto a subspace \( S \), there is an obvious way of determining the orthogonal projection onto \( S^\perp \).
Theorem 1.57 If \( P \) is the orthogonal projection onto a subspace \( S \), then \( I - P \) is the orthogonal projection onto \( S^\perp \).

Through these two orthogonal projection matrices, an arbitrary vector \( v \in \mathbb{R}^n \) gets effectively decomposed into its orthogonal components. In other words,

\[
x = Px + (I - P)x.
\]

The orthogonality property can also be observed by multiplying both matrices:

\[
\begin{align*}
P(I - P) &= P - P^2 = P - P = 0, & \text{and} \\
(I - P)P &= P - P^2 = P - P = 0.
\end{align*}
\]  

(1.55)

We will find this property very useful when we study dynamical systems in Chapter 7.

A projection matrix onto a subspace \( S \) is in general not unique, but there is only one orthogonal projection.

Theorem 1.58 Let \( V \) be a vector space, and let \( S \) be a vector subspace of \( V \). Then, the orthogonal projection matrix \( P \) onto \( S \) is unique.

Proof. Assume there are two orthogonal projections \( P_1 \) and \( P_2 \) onto the subspace \( S \). Then, for any \( x \in V \),

\[
\begin{align*}
\| (P_1 - P_2)x \|^2 &= ((P_1x)^T - (P_2x)^T)(P_1x - P_2x) \\
&= x^T P_1 x - (P_1x)^T P_2 x - (P_2x)^T P_1 x + x^T P_2 x \\
&= (P_1x)^T (I - P_2)x + (P_2x)^T (I - P_1)x.
\end{align*}
\]

But \( P_1 x, P_2 x \in S \) and \( (I - P_1), (I - P_2) \in S^\perp \), so that the last expression above is zero and therefore \( P_1 = P_2 \).

\[\square\]

1.10 EIGENVALUES AND EIGENVECTORS

Theorem 1.49 gives conditions for the existence and uniqueness of solutions of linear systems of equations of the form

\[
Ax = b,
\]

(1.56)

where \( A \) is a general rectangular matrix. We learned that (1.56) has a solution if \( b \) is in \( \text{col}(A) \). For the special case when \( A \) is square, which is the case we are interested in, we ask instead for conditions under which the system has a unique solution. The answer can be given in terms of the singularity of the square matrix \( A \).
According to Theorem 1.49, a system with a square coefficient matrix $A$ of order $n$ has a unique solution if $\dim N(A) = 0$, or equivalently, $\dim \text{col}(A) = n$; that is, all the columns of $A$ must be linearly independent, and this is equivalent to saying that $A$ is nonsingular. More precisely.

**Theorem 1.59** Let $A$ be a square matrix of order $n$. The system $Ax = b$ has a unique solution if and only if the matrix $A$ is nonsingular.

**Remark 1.60** In particular, this theorem implies that if $A$ is nonsingular, then the system $Ax = 0$ has the unique solution $x = 0$.

Two essential concepts associated with every square matrix are eigenvalues and eigenvectors; they contain important information about the matrix, some of its associated subspaces, and about the structure of problems and phenomena whose modeling contain such a matrix. There are numerous applications of eigenvalues and eigenvectors, within and outside mathematics, e.g., differential equations, control theory, Markov chains, web page ranking, image compression, etc. We will study in detail some of these applications later on.

**Definition 1.61** Given a square matrix $A$ of order $n$, we say that $\lambda$ is an **eigenvalue** of $A$ with associated **eigenvector** $v \neq 0$ if

$$Av = \lambda v. \quad (1.57)$$

In general, eigenvalues are complex numbers, and eigenvectors are complex ($n$–dimensional) vectors, even though $A$ is a real matrix.

One first and immediate geometric observation is that given an eigenvector $v$, equation (1.57) says that the vector $Av$ is just a multiple of $v$, with larger or smaller magnitude. In fact, the line containing $v$ is a one-dimensional vector space and $Av$ lands on that same vector space. This is a simple instance of invariance, and we say that the space spanned by $v$ is **invariant** with respect to $A$. We will return to this concept when we study matrix factorizations in Chapter 3 and dynamical systems in Chapter 7.

Observe that equation (1.57) can be written as

$$(A - \lambda I)v = 0. \quad (1.58)$$

Since we are looking for eigenvectors $v \neq 0$, we need the matrix $(A - \lambda I)$ to be singular (see Remark 1.60). In other words, we need to require that

$$\det(A - \lambda I) = 0. \quad (1.59)$$

Equation (1.59) is known as the **characteristic equation**. The left–hand side is a polynomial on $\lambda$ of degree $n$, and the solutions to this equation are the eigenvalues of $A$. On the other hand, for each given eigenvalue $\lambda$, the system (1.58) is used to find a corresponding eigenvector.
Remark 1.62 The eigenvalues are uniquely determined as solutions of the characteristic equation (1.59), but eigenvectors are not unique, as they are solutions of the singular system (1.58).

**EXAMPLE 1.68**

Let \( A = \begin{bmatrix} 1 & -2 \\ 0 & 3 \end{bmatrix} \). Then, the characteristic equation is

\[
\det \begin{bmatrix} 1 - \lambda & -2 \\ 0 & 3 - \lambda \end{bmatrix} = (3 - \lambda)(1 - \lambda) = 0.
\]

Then, the eigenvalues of \( A \) are \( \lambda_1 = 3 \) and \( \lambda_2 = 1 \). To find the corresponding eigenvectors, we use equation (1.58) for each eigenvalue. First let us consider \( \lambda = 3 \). Then, equation (1.58) gives

\[
\begin{bmatrix} 1 - 3 & -2 \\ 0 & 3 - 3 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} -2 \\ 0 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix},
\]

from which \( v_2 = -v_1 \). Thus, any nonzero vector of the form \( v = [v_1 \ - v_1]^T \) is an eigenvector of \( A \) corresponding to \( \lambda = 3 \). In particular, we may take \( v = [1 \ - 1]^T \).

In a similar way, any eigenvector associated with \( \lambda = 1 \) is of the form \( [v_1 \ 0]^T \), so in particular we may take \( v = [1 \ 0]^T \).

Remark 1.63 In general, finding the eigenvalues of upper or lower triangular matrices like the one in Example 1.68 is immediate. The eigenvalues are nothing else but the diagonal entries.

**EXAMPLE 1.69**

There is a particular matrix that appears very often, and its eigenvalues can be seen directly from the matrix itself:

If \( A = \begin{bmatrix} a & b \\ -b & a \end{bmatrix} \), then \( \lambda_{1,2} = a \pm ib \). \hfill (1.60)
EXAMPLE 1.70

Let \( A = \begin{bmatrix} -2 & -1 & 0 \\ 1 & -2 & 0 \\ 0 & 0 & 3 \end{bmatrix} \). Then, the characteristic equation is

\[
\det \begin{bmatrix} -2 - \lambda & -1 & 0 \\ 1 & -2 - \lambda & 0 \\ 0 & 0 & 3 - \lambda \end{bmatrix} = \lambda^3 + \lambda^2 - 7\lambda - 15 = 0.
\]

The solution of this characteristic equation gives the eigenvalues: \( \lambda_{1,2} = -2 \pm i \) and \( \lambda_3 = 3 \). To find the corresponding eigenvectors, let us consider first \( \lambda_1 = -2 + i \). In this case, equation (1.58) gives

\[
\begin{bmatrix} -2 - (-2 + i) & -1 & 0 \\ 1 & -2 - (-2 + i) & 0 \\ 0 & 0 & 3 - (-2 + i) \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}.
\]

From the first or second row, we get that \( v_1 = i v_2 \), and from the third row, we get \( v_3 = 0 \). Then, any eigenvector associated with \( \lambda_1 \) is of the form \([i \ v_2 \ v_2 \ 0]^T \). Similarly, any eigenvector associated with \( \lambda_2 \) is of the form \([-i \ v_2 \ v_2 \ 0]^T \). Finally, it is easy to see that any eigenvector associated with \( \lambda_3 = 3 \) is of the form \([0 \ 0 \ v_3]^T \), and in particular, we can take \( w = [0 \ 0 \ 1]^T \).

In some applications, like in the solutions of differential equations, it is necessary to extract real solutions by using eigenvectors (even if they are complex). In this case, using for instance the complex eigenvectors associated with \( \lambda_{1,2} = -2 \pm i \), we can write

\[
u = \pm i \ v = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \pm i \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}.
\]

In this way, we can obtain two (linearly independent) real vectors \( u \) and \( v \) that span a 2-dimensional subspace.

Remark 1.64 By using the fact that the determinant of any matrix is the same as that of its transpose, one can easily show that any matrix \( A_{n \times n} \) and its transpose \( A^T \) have the same eigenvalues. However, their eigenvectors are not necessarily the same.

The matrix \( A \) in Example 1.70 has a particular form. Observe that we can represent \( A \) as

\[
A = \begin{bmatrix} -2 & -1 & 0 \\ 1 & -2 & 0 \\ 0 & 0 & 3 \end{bmatrix} = \begin{bmatrix} A_1 & 0 \\ 0 & A_2 \end{bmatrix}.
\]
Then, finding the eigenvalues of $A$ reduces itself to finding the eigenvalues of $A_1$ and $A_2$ separately. In this case, according to (1.60), the eigenvalues of $A_1$ are $\lambda_{1,2} = -2 \pm i$, and the eigenvalue of $A_2$ is $\lambda_3 = 3$.

This is actually a general result in linear algebra to compute the eigenvalues of block diagonal matrices: Let us denote with $\sigma(A)$ the set of eigenvalues of $A$ (this set is known as the spectrum of $A$). Then, we have the following theorem.

**Theorem 1.65** Let $A =$

\[
\begin{bmatrix}
A_1 & & \\
& A_2 & \\
& & \ddots \\
& & & A_k
\end{bmatrix}
\]

Then,

\[\sigma(A) = \sigma(A_1) \cup \sigma(A_2) \cup \cdots \cup \sigma(A_k).\]

A similar result applies if the matrix is block upper triangular. We have the following

**Theorem 1.66** Consider the block upper triangular matrix

\[
A = \begin{bmatrix} B & C \\ 0 & D \end{bmatrix},
\]

where $B$ and $D$ are square matrices of order say $p$ and $q$, respectively, for some positive integers $p, q$. Then, $\sigma(A) = \sigma(B) \cup \sigma(D)$, counting multiplicities.

**Proof.** Let $\lambda$ be an eigenvalue of $A$ with eigenvector $v = [v_1 \ v_2]^T$, where $v_1 \in \mathbb{R}^p$ and $v_2 \in \mathbb{R}^q$. Then,

\[
Av = \begin{bmatrix} B & C \\ 0 & D \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \lambda \begin{bmatrix} v_1 \\ v_2 \end{bmatrix}.
\]

We either have $v_2 = 0$ or $v_2 \neq 0$. If $v_2 = 0$, then $v_1 \neq 0$ and $Bv_1 = \lambda v_1$, so that $\lambda$ is an eigenvalue of $B$. If $v_2 \neq 0$, then $Dv_2 = \lambda v_2$ so that $\lambda$ is an eigenvalue of $D$. This proves that $\sigma(A) \subseteq \sigma(B) \cup \sigma(D)$. In addition, the two sets $\sigma(A)$ and $\sigma(B) \cup \sigma(D)$ have the same cardinality. Therefore, they must be equal.

\[\square\]

**Remark.** The result in Theorem 1.66 is also true if $A$ is block lower triangular.
EXAMPLE 1.71

The eigenvalues of \( A = \begin{bmatrix} 8 & -2 \\ 2 & 4 \end{bmatrix} \) are \( \lambda = 6, 6 \), and the eigenvalues of \( A = \begin{bmatrix} 5 & 3 & 0 \\ -3 & 5 & 0 \\ 1 & -2 & 7 \end{bmatrix} \) are \( \lambda = 7, 5 \pm 3i \). Therefore, the eigenvalues of

\[
A = \begin{bmatrix} 8 & -2 & 3 & 5 & 8 \\ 2 & 4 & 2 & -1 & 4 \\ 0 & 0 & 5 & 3 & 0 \\ 0 & 0 & -3 & 5 & 0 \\ 0 & 0 & 1 & -2 & 7 \end{bmatrix}
\]

are \( \lambda = 6, 6, 7, 5 \pm 3i \).

We have seen that in general the eigenvalues of a matrix \( A \) can be real or complex. However, if the matrix is symmetric, then its eigenvalues are real. Furthermore, its eigenvectors are orthogonal! The proof of the following theorem is left as an exercise.

**Theorem 1.67** Let \( A_{n \times n} \) be a symmetric matrix. Then all its eigenvalues are real, and eigenvectors corresponding to different eigenvalues are orthogonal.

EXAMPLE 1.72

Let \( A = \begin{bmatrix} 3 & -1 & 4 \\ -1 & 0 & 1 \\ 4 & 1 & 3 \end{bmatrix} \). By using the characteristic equation, we find that the eigenvalues are \( \lambda_1 = -2, \lambda_2 = 1, \) and \( \lambda_3 = 7 \). Let us compute the eigenvectors.

For \( \lambda_1 = -2 \), by letting \( u = [x \ y \ z]^T \), equation (1.58) gives

\[ 5x - y + 4z = 0, \quad -x + 2y + z = 0, \quad 4x + y + 5z = 0. \]

From the second equation, \( z = x - 2y \). By substituting this into the third equation, we get \( x = y \). Then, any eigenvector associated with \( \lambda_1 = -2 \) is of the form \( u = [x \ x - x]^T \).

For \( \lambda_2 = 1 \), equation (1.58) gives

\[ 2x - y + 4z = 0, \quad -x - y + z = 0, \quad 4x + y + 2z = 0. \]

From the second equation, \( z = x + y \). Substituting this into the first equation, we get \( y = -2x \). Then, any eigenvector associated with \( \lambda_2 = 1 \) is of the form \( v = [x - 2x \ -x]^T \).

For \( \lambda_3 = 7 \), equation (1.58) gives

\[ -4x - y + 4z = 0, \quad -x - 7y + z = 0, \quad 4x + y - 4z = 0. \]
From the second equation, \( z = x + 7y \). By substituting this into the third equation, we get \( y = 0 \). Then, any eigenvector associated with \( \lambda_3 = 7 \) is of the form \( w = [x \ 0 \ x]^T \). Now, it is simple to verify that in fact
\[
u^T v = 0, \quad u^T w = 0, \quad v^T w = 0,
\]
which means that the eigenvectors are mutually orthogonal.

**Note:** In Example 1.72, a general fact is illustrated: since by definition, \( \det(A - \lambda I) = 0 \), in each set of three equations to determine the eigenvectors, each of the individual equations can be written in terms of the other two; that is why the solution is not unique (eigenvectors are not unique). In other words, in each of the sets of three equations, one is redundant and we can solve each system by using only two of its equations.

The next question is whether repeated eigenvalues of a general matrix can give linearly independent eigenvectors.

**Example 1.73**

Let
\[
A = \begin{bmatrix}
9 & 4 & 0 \\
-6 & -1 & 0 \\
6 & 4 & 3
\end{bmatrix}.
\]
By solving the corresponding characteristic equation, the eigenvalues are \( \lambda_1 = 5 \), \( \lambda_2 = 3 \), and \( \lambda_3 = 3 \). Thus, we have repeated eigenvalues. The eigenvector corresponding to \( \lambda_1 \) is \( w_1 = [1 \ -1 \ 1]^T \). To find the eigenvector(s) corresponding to \( \lambda_2 = \lambda_3 \), equation (1.58) with \( v = [x \ y \ z]^T \) gives
\[
6x + 4y = 0, \quad -6x - 4y = 0, \quad 6x + 4y = 0.
\]
The three equations are the same. From the first equation, we get \( x = -\frac{2}{3}y \). The third entry \( z \) is free. Then, any eigenvector associated with \( \lambda_2 = \lambda_3 = 3 \) has the general form \( [-\frac{2}{3}y \ y \ z]^T \). We have two free variables, and in particular, by taking \( y = 0 \), \( z = 1 \) and then \( y = -3 \), \( z = 0 \), respectively, we get \( w_2 = [0 \ 0 \ 1]^T \), \( w_3 = [2 \ -3 \ 0]^T \). Thus, even though there are repeated eigenvalues, we are still able to find a complete set of three linearly independent eigenvectors.

There is something very important to observe in Example 1.73. The eigenvectors \( w_1 \), \( w_2 \), and \( w_3 \) not only are different but also they are linearly independent. This is something that becomes very important in several applications, especially in the solution of differential equations. We will study this in more detail in Chapter 6.

In the examples above, we have seen that it was possible to find the same number of linearly independent eigenvectors and eigenvalues, even when the eigenvalues are repeated. However this is not always the case. There are cases when repeated eigenvalues give a smaller number of linearly independent eigenvectors, so that it is not possible to form a complete set of linearly independent eigenvectors. In this case, the repeated eigenvalue is
called defective, and we need to generalize the concept of eigenvector, so that we complete our set of (generalized) eigenvectors.

More formally, if an eigenvalue $\lambda$ is repeated $k$ times, we say $\lambda$ has **algebraic multiplicity** $k$; e.g., in Example 1.73, $\lambda = 3$ has algebraic multiplicity two. The number of linearly independent eigenvectors associated with an eigenvalue $\lambda$ is called the **geometric multiplicity** of $\lambda$; e.g., in Example 1.73, the geometric multiplicity of $\lambda = 3$ is two, so in this particular case, algebraic and geometric multiplicity coincide. In general, we have (see Exercise 3.60)

$$\text{geometric multiplicity} \leq \text{algebraic multiplicity}.$$  

Thus, an eigenvalue is called defective if the above inequality is strict.

An eigenvalue with algebraic multiplicity 1 is called a **simple eigenvalue**.

For each eigenvalue $\lambda$ of a matrix $A_{m \times n}$, we can define a special set: the set of all the eigenvectors associated with it. If we add to this set the zero vector, then it becomes a subspace of $\mathbb{R}^n$.

**Definition 1.68** Let $A$ be a square matrix of order $n$, and $\lambda$ an eigenvalue of $A$. The set

$$E = \{ v \in \mathbb{R}^n : Av = \lambda v \}$$

is called the **eigenspace** of $A$ with respect to $\lambda$.

Several observations are in order.

1. We are considering all solutions of the equation $Av = \lambda v$. Thus, all eigenvectors of $A$ associated with $\lambda$ and also the zero vector are the elements of $E$.

2. The geometric multiplicity of $\lambda$ is the dimension of its eigenspace $E$.

3. The linearly independent eigenvectors of $A$ associated with $\lambda$ form a basis of $E$.

4. The eigenspace of $A$ can be defined as the nullspace of $A - \lambda I$.

When the geometric multiplicity of an eigenvector is strictly smaller than its algebraic multiplicity, that is, when the number of linearly independent eigenvectors is smaller than the number of times the corresponding eigenvalue is repeated, then the concept of an eigenvector needs to be generalized.

**Definition 1.69** (Generalized Eigenvector) If $\lambda$ is an eigenvalue of a matrix $A$, then a rank-$r$ **generalized eigenvector** associated with $\lambda$ is a vector $v$ such that

$$(A - \lambda I)^r v = 0, \quad \text{and} \quad (A - \lambda I)^{r-1} v \neq 0.$$  \hspace{1cm} (1.61)
Observe that a rank-1 generalized eigenvector is an ordinary eigenvector. That is,\[(A - \lambda I)v = 0, \quad \text{and} \quad v \neq 0.\]

**Example 1.74**

Let \(A = \begin{bmatrix} 2 & 1 & 0 \\ 0 & 2 & 1 \\ 0 & 0 & 3 \end{bmatrix}\). Then, according to Remark 1.63, the eigenvalues are \(\lambda_{1,2} = 2, \lambda_3 = 3\). For \(\lambda = 2\), equation (1.58) with \(v = [x \ y \ z]^T\) gives \(y = 0, z = 0, \) and \(x\) is a free variable. This means that any eigenvector associated with \(\lambda_{1,2} = 2\) is of the form \([x \ 0 \ 0]^T\). Thus, we can get only one (linearly independent) eigenvector. We need to find a generalized eigenvector associated with \(\lambda = 2\). Following (1.61), the equation

\[(A - \lambda I)^2 v = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}\]

gives \(z = 0\). Thus, any rank-2 generalized eigenvector of \(A\) associated with \(\lambda_1 = \lambda_2 = 2\) has the form \([x \ y \ 0]^T\), with two free variables. In particular, we can take \(v_1 = [1 \ 0 \ 0]^T\) and \(v_2 = [0 \ 1 \ 0]^T\). Observe that \(v_1\) is the eigenvector that was determined before, and \(v_2\) is a generalized eigenvector. Furthermore, we can verify that in this case

\[(A - \lambda I)v_2 = v_1 \neq 0,\]

thus satisfying the definition in (1.61). Finally, we can readily find out that any eigenvector associated with \(\lambda_3 = 3\) is of the form \([0 \ 0 \ z]\), and in particular we can take \(v_3 = [0 \ 0 \ 1]\).

We will see more about generalized eigenvectors in Chapter 6.

**Left eigenvectors.** It is possible to define eigenvectors when they are actually row vectors. Even though the eigenvalues of \(A\) would be the same, these new eigenvectors are in general different from the (right) eigenvectors introduced before. Suppose \(\lambda\) is an eigenvalue of \(A\) with associated eigenvector \(x\), so that \(Ax = \lambda x\). From Remark 1.64, we know that \(\lambda\) is also an eigenvalue of \(A^T\), for some eigenvector \(v\); that is,

\[A^Tv = \lambda v.\]

By applying the transpose operation to both sides of this equation, we obtain the following definition.

**Definition 1.70** We say a nonzero vector \(v \in \mathbb{R}^n\) is a **left eigenvector** of a matrix \(A_{n \times n}\) associated with an eigenvalue \(\lambda\) if

\[v^TA = \lambda v^T.\]  

(1.62)
In other words, a left eigenvector of \( A \) associated with an eigenvalue \( \lambda \) is a (right) eigenvector of \( A^T \) corresponding to the same eigenvalue \( \lambda \).

**Example 1.75**

Let \( A = \begin{bmatrix} 5 & 4 \\ 0 & 3 \end{bmatrix} \). For the eigenvalue \( \lambda = 3 \), any associated eigenvector \( x \) has the form \( x = \begin{bmatrix} x_1 \\ -\frac{1}{2}x_1 \end{bmatrix} \); e.g., \( x = \begin{bmatrix} 2 \\ -1 \end{bmatrix} \). However, for the same eigenvalue \( \lambda = 3 \), any left eigenvector of \( A \) has the form \( v = \begin{bmatrix} 0 \\ v_2 \end{bmatrix} \); e.g., \( v = \begin{bmatrix} 0 \\ 5 \end{bmatrix} \). Thus, we have

\[
\begin{bmatrix} 5 & 4 \\ 0 & 3 \end{bmatrix} \begin{bmatrix} 2 \\ -1 \end{bmatrix} = 3 \begin{bmatrix} 2 \\ -1 \end{bmatrix}, \quad \text{and} \\
\begin{bmatrix} 0 & 5 \\ 0 & 3 \end{bmatrix} \begin{bmatrix} 0 \\ 5 \end{bmatrix} = 3 \begin{bmatrix} 0 \\ 5 \end{bmatrix}.
\]

### 1.11 Similarity

We have seen that several sets of vectors can be chosen as a basis of a given vector space, and that arbitrary vectors in that space get different representations depending on the basis being used. In a similar way, given an arbitrary matrix \( A \), we would like to find a simpler matrix \( B \) that is related or is equivalent to \( A \) in the sense that it shares some of its properties (e.g., same eigenvalues). Although ideally we would like \( B \) to be diagonal (the simplest possible), for most applications, it suffices if \( B \) is triangular, block diagonal, or block triangular.

We start this topic with the following definition.

**Definition 1.71** Two square matrices \( A \) and \( B \) are said to be **similar** if there exists a nonsingular matrix \( P \) such that

\[
B = P^{-1} A P.
\]  

(1.63)

**Note:** Equation (1.63) is sometimes written as \( B = P A P^{-1} \).

The following theorem states one of the most important applications of similarity.

**Theorem 1.72** Similar matrices have the same eigenvalues, counting multiplicities.

**Proof.** The idea is to show that characteristic polynomials of the similar matrices are equal, for if that is true, then they will have the same eigenvalues, counting multiplicities. Let \( A \)
and $B$ be similar. Then,

\[
\det(B - \lambda I) = \det(P^{-1}AP - \lambda I) = \det(P^{-1}AP - P^{-1}\lambda I P) = \det(P^{-1}(A - \lambda I) P) = \frac{1}{\det(P)} \det(A - \lambda I) \det(P).
\]

\[
= \det(A - \lambda I).
\]

\[ \square \]

**Example 1.76**

The matrix $A = \begin{bmatrix} 1 & 2 \\ -1 & 3 \end{bmatrix}$ is similar to $B = \begin{bmatrix} 7 & 13 \\ -2 & -3 \end{bmatrix}$. In fact, you can verify that this is true by using the matrix $P = \begin{bmatrix} 2 & 5 \\ 1 & 3 \end{bmatrix}$. You can also verify that both $A$ and $B$ have the same eigenvalues $\lambda = 2 \pm i$.

Ideally, we would like matrices to be similar to diagonal matrices, the simplest matrices possible. If this is true, the matrix has a special name.

**Definition 1.73** A matrix $A_{n \times n}$ is called **diagonalizable** if it is similar to a diagonal matrix.

Given a matrix $A_{n \times n}$, how can we find the matrix $P$ that performs the similarity transformation (1.63)? One answer to this is the eigenvectors. If $A$ has $n$ linearly independent eigenvectors, then the columns of the matrix $P$ can be defined as the eigenvectors of $A$. Such a matrix $P$ is nonsingular because its columns are linearly independent. The existence of such matrix $P$ guarantees that $A$ is diagonalizable. The following results come in handy.

**Theorem 1.74** Let $A$ be a square matrix of order $n$. Then,

\[ A \text{ has distinct evlues} \implies A \text{ has } n \text{ l.i. eigenvectors } \iff A \text{ is diagonalizable}. \]

**Proof.** To prove the first part of the theorem, assume by contradiction that $A$ has $n$ distinct eigenvalues $\lambda_1, \ldots, \lambda_n$ but that the corresponding set of eigenvectors $V = \{v_1, \ldots, v_n\}$ is not linearly independent. For some $m < n$, let $\{v_1, \ldots, v_m\} \subset V$ be the set with the maximum number of linearly independent eigenvectors of $A$. Then,

\[ v_{m+1} = c_1 v_1 + \cdots + c_m v_m. \]

for some scalars $c_1, \ldots, c_m$. Using the general equation $(A - \lambda I)v = 0$ from (1.58), we have

\[
0 = (A - \lambda_{m+1} I)v_{m+1} = (A - \lambda_{m+1} I)(c_1 v_1 + \cdots + c_m v_m) = (c_1 \lambda_1 v_1 + \cdots + c_m \lambda_m v_m) - (c_1 \lambda_{m+1} v_1 + \cdots + c_m \lambda_{m+1} v_m) = c_1 (\lambda_1 - \lambda_{m+1}) v_1 + \cdots + c_m (\lambda_m - \lambda_{m+1}) v_m.
\]
The linear independence of $v_1, \ldots, v_m$ implies that

$$c_1(\lambda_1 - \lambda_{m+1}) = \cdots = c_m(\lambda_m - \lambda_{m+1}) = 0,$$

and since the eigenvalues are distinct, we get $c_1 = \cdots = c_m = 0$. But then $v_{m+1} = 0$, which is not possible because $v_{m+1}$ is an eigenvector.

The proof of the second part is left as an exercise.

\[\square\]

\textbf{Example 1.77}

Consider the matrix $A = \begin{bmatrix} 1 & 2 & 1 \\ 6 & -1 & 0 \\ -1 & -2 & -1 \end{bmatrix}$. The eigenvalues are $\lambda_1 = -4$, $\lambda_2 = 3$, $\lambda_3 = 0$. The eigenvectors are $v_1 = [1 - 2 - 1]^T$, $v_2 = [2 \ 3 \ -2]^T$, and $v_3 = [1 \ 6 \ -13]^T$, respectively, and they are linearly independent. Define

$$P = \begin{bmatrix} v_1 & v_2 & v_3 \end{bmatrix} = \begin{bmatrix} 1 & 2 & 1 \\ -2 & 3 & 6 \\ -1 & -2 & -13 \end{bmatrix}.$$ 

Then,

$$P^{-1}AP = \begin{bmatrix} -4 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$ 

Thus, the matrix $A$ is diagonalizable, and the diagonal matrix has the eigenvalues of $A$ on its diagonal entries.

\textbf{Invariant properties under similarity.} We have seen that two similar matrices have the same eigenvalues. But not only eigenvalues are invariant under a similarity transformation. The following properties are also invariant (the proof of each is left as an exercise to the reader):

rank, determinant, trace, invertibility.

This means, e.g., that if $A$ and $B$ are similar matrices and $A$ has rank $k$, then so does $B$ and if $A$ is invertible, so is $B$.

We finish this section by mentioning some more properties and applications of the eigenvalues of a matrix $A$.

1. The determinant of a matrix $A$ can be defined as the product of its eigenvalues. That is, if $\lambda_1, \ldots, \lambda_n$ are the eigenvalues of $A$, then

$$\det A = \lambda_1 \cdot \lambda_2 \cdots \lambda_n.$$  \hfill (1.64)
2. From above, we see that $A$ having one zero eigenvalue is necessary and sufficient to have a zero determinant. That is,

$$\det A = 0 \iff \lambda_i = 0, \text{ for some } i = 1, \ldots, n.$$  \hfill (1.65)

3. The matrix norm $\| \cdot \|_2$ given by (1.9) with $p = 2$, can also be defined through eigenvalues. More precisely, we compute the maximum eigenvalue of the symmetric matrix $A^T A$. That is,

$$\| A \|_2 = \sqrt{\max \lambda (A^T A)}. \hfill (1.66)$$

\subsection{Bezier Curves and Postscript Fonts}

As we mentioned in Section 1.6, one powerful and useful concept in vector spaces is that of a basis. Given a vector space $V$ of dimension $n$ and a basis $B = \{v_1, \ldots, v_n\}$ of $V$, then any element $x \in V$ can be uniquely expressed as

$$x = c_1 v_1 + \cdots + c_n v_n,$$ \hfill (1.67)

for some scalars $c_1, \ldots, c_n$. We say that $x$ is written as a linear combination of the elements of the basis $B$.

We want to see how this central concept in linear algebra can be used in a real-world application. Let $V$ be the vector space of all real polynomials of degree at most $n$. This space has dimension $n + 1$, and one basis is given by the set

$$\{1, t, t^2, \ldots, t^n\}.$$ 

Several other bases can be formed for this vector space, but here we consider one that is very useful in several practical applications. In particular, it has proved to be essential in the construction of special curves, such as those needed in defining letter fonts, the ones we use everyday in text editing. The goal is to construct the so-called Bezier curves, which are currently used in computer font rendering technologies such as Adobe PDF and Postscript.

The special basis we are interested in can be formed using the so-called Bernstein polynomials of degree $n$, which can be defined recursively as

$$B_{i,n}(t) = \begin{cases} 1, & i = 0, \cr (1 - t)B_{i,n-1}(t) + t B_{i-1,n-1}(t), & \text{for } i \geq 1. \end{cases}$$

where we take $B_{i,n}(t) = 0$, for $i < 0$, or $i > n$.

For example, the Bernstein polynomials of degrees 1, 2, and 3 are, respectively,
The cubic Bernstein polynomials.

\[
B_{0.1}(t) = 1 - t, \quad B_{1.1}(t) = t; \\
B_{0.2}(t) = (1 - t)^2, \quad B_{1.2}(t) = 2t(1 - t), \quad B_{2.2}(t) = t^2; \\
B_{0.3}(t) = (1 - t)^3, \quad B_{1.3}(t) = 3t(1 - t)^2, \quad B_{2.3}(t) = 3t^2(1 - t), \quad B_{3.3}(t) = t^3.
\]

(1.68)

A much nicer way of obtaining the polynomials in (1.68) is by just expanding 1. For instance, for the Bernstein cubic polynomials:

\[
1^3 = [(1 - t) + t]^3 = (1 - t)^3 + 3t(1 - t)^2 + 3t^2(1 - t) + t^3.
\]

Observe that each term in this expansion is one of the cubic Bernstein polynomials \(B_{i,3}, \ i = 0, \ldots, 3\). In general, we can write explicitly:

\[
B_{i,n}(t) = \binom{n}{i} t^i (1 - t)^{n-i}, \quad i = 0, \ldots, n.
\]

Here we are mostly interested in cubic Bernstein polynomials \(B_{i,3}(t)\); these are shown in Figure 1.8. They form a basis of the space of polynomials of degree at most three. Again, the most important fact is that any polynomial of degree at most three can be expressed as a unique linear combination of those four cubic Bernstein polynomials.

As an example, let \(P_3(t) = -5t^3 + 15t^2 - 9t + 2\). Then,

\[
P_3(t) = 2B_{0,3}(t) - B_{1,3} + B_{2,3} + 3B_{3,3}.
\]
Now let us see the usefulness of these special polynomials. Assume you are given four points

\[ P_i = (x_i, y_i), \ i = 0, \ldots, 3, \]
henceforth called control points, and that you are looking for a polynomial of degree at most three that passes through the first and last point, and whose graph roughly follows the shape of the polygonal path determined by the four points.

This is a problem quite different from that of interpolation, where the curve must pass through all the given points (a general Bezier curve always passes through the first and the last point, but not necessarily through the other points, which are mostly used to control the shape of the curve). The problem here has another application in mind: to obtain an arbitrary plane curve and be able to modify its shape by redefining (moving) a few of the control points. This Bezier curve is to be defined using polynomials of degree at most three, and following (1.67), it will be expressed as a combination of basis functions

\[ C(t) = \sum_{i=0}^{3} P_i B_{i,3}(t), \quad t \in [0, 1], \quad (1.69) \]

where \( B_{i,3} \) are the cubic Bernstein polynomials, and \( P_i \) are the given control points.

For general linear combinations of the type (1.67), the main problem is to find the coefficients \( c_i \), but in this present case, the coefficients \( P_i \) in (1.69) are the control points, which are always given.

An important fact is that we can express the Bezier curve (1.69) parametrically as

\[ C(t) = (x(t), y(t)), \]

where

\[ x(t) = \sum_{i=0}^{3} x_i B_{i,3}(t), \quad y(t) = \sum_{i=0}^{3} y_i B_{i,3}(t), \quad t \in [0, 1], \quad (1.70) \]

with \( P_i = (x_i, y_i) \). Thus, a Bezier curve is a differentiable curve in the \( xy \) plane.

This is a very useful and illustrative application of the concept of basis of a vector space. Given four points \( P_i, \ i = 0, 1, 2, 3, \) we construct a smooth curve in parametric form (1.70), where the coefficients of the linear combination in (1.70); that is, \( x_i \), and \( y_i \), are the coordinates of the given points, and \( B_{i,3} \) are the cubic Bernstein polynomials in (1.68), which as we know form a basis for the space of polynomials of degree \( \leq 3 \).

\subsection{1.12.1 Properties of Bezier curves}

\textbf{First and last point.} An important property of the uniquely determined curve \( C(t) \) is that it will pass through the first and last point. In fact, using (1.70), we have for the first coordinate of the curve at \( t = 0 \) and \( t = 1 \):

\[ x(0) = x_0 B_{0,3}(0) + x_1 B_{1,3}(0) + x_2 B_{2,3}(0) + x_3 B_{3,3}(0) \]
\[ x_0 + 0 + 0 + 0 = x_0, \quad \text{and} \]
\[ x(1) = x_0 B_{0,3}(1) + x_1 B_{1,3}(1) + x_2 B_{2,3}(1) + x_3 B_{3,3}(1) \]
\[ = 0 + 0 + 0 + x_3 = x_3. \]

A similar result applies for the second coordinate \( y(t) \) of \( C(t) \) at \( t = 0 \) and \( t = 1 \).

**Note:** The fact that the curve passes through the first and the last points is true in general for Bernstein polynomials of arbitrary degree \( n \), not only for cubic.

**Example 1.78**

As an illustration, let us find the Bernstein curve for the control points \( P_0 = (4, 1) \), \( P_1 = (3, 2) \), \( P_2 = (5, 5) \), and \( P_3 = (7, 3) \).

By substituting the first and second coordinates of the given control points into (1.70), we have

\[
\begin{align*}
x(t) &= 4B_{0,3}(t) + 3B_{1,3}(t) + 5B_{2,3}(t) + 7B_{3,3}(t) \\
&= 4(1-t)^3 + 9t(1-t)^2 + 15t^2(1-t) + 7t^3 \\
&= 4 - 3t + 9t^2 - 3t^3.
\end{align*}
\]

Similarly,

\[
\begin{align*}
y(t) &= 1B_{0,3}(t) + 2B_{1,3}(t) + 5B_{2,3}(t) + 3B_{3,3}(t) \\
&= 1(1-t)^3 + 6t(1-t)^2 + 15t^2(1-t) + 3t^3 \\
&= 1 + 3t + 6t^2 - 7t^3.
\end{align*}
\]

Thus,

\[ C(t) = (4 - 3t + 9t^2 - 3t^3, 1 + 3t + 6t^2 - 7t^3), \quad t \in [0, 1]. \]

The curve and the control points are shown in Figure 1.9. Observe how the curve somehow follows the polygonal path determined by the control points.

**Tangency property.** The main reason why a Bernstein curve roughly follows the shape determined by the control points and the segments joining them is that the slope of the curve at \( P_0 \) is the same as that of the segment joining \( P_0 \) and \( P_1 \). Similarly, the slope of the curve at \( P_3 \) coincides with the slope of the segment joining \( P_2 \) and \( P_3 \). This is apparent in Figure 1.9.

In fact, with \( P_0 = (x_0, y_0) \), and \( P_1 = (x_1, y_1) \), the slope of the segment joining these two points is

\[ m = \frac{y_1 - y_0}{x_1 - x_0}. \]

Now let us find the slope of the Bernstein curve at \( P_0 \). In parametric form, the Bernstein curve is

\[
\begin{align*}
x(t) &= (1-t)^3 x_0 + 3t(1-t)^2 x_1 + 3t^2(1-t)x_2 + t^3 x_3, \\
y(t) &= (1-t)^3 y_0 + 3t(1-t)^2 y_1 + 3t^2(1-t)y_2 + t^3 y_3.
\end{align*}
\]
The corresponding derivatives are
\[
\frac{dx}{dt} = -3(1-t)^2x_0 + 3(1-t)^2x_1 - 6t(1-t)x_1 + 6t(1-t)x_2 - 3t^2x_2 + 3t^2x_3,
\]
\[
\frac{dy}{dt} = -3(1-t)^2y_0 + 3(1-t)^2y_1 - 6t(1-t)y_1 + 6t(1-t)y_2 - 3t^2y_2 + 3t^2y_3
\]

Then, the slope of the curve at \(P_0\) is
\[
\frac{dy}{dx} \mid_{t=0} = \frac{dy/dt}{dx/dt} \mid_{t=0} = \frac{-3y_0 + 3y_1}{-3x_0 + 3x_1} = \frac{y_1 - y_0}{x_1 - x_0} = m.
\]

Similarly, it can be proved that the slope of the Bezier curve at \(P_3\) coincides with the slope of the segment joining \(P_2\) and \(P_3\).

**Bezier curve and convex hull.** The effectiveness and usefulness of Bezier curves lies in the great ease with which the shape of the curve can be modified, say by means of a mouse, by making adjustments to the control points. One good advantage is that the changes in the shape of the curve will be somewhat localized. The only term(s) in (1.71) that are modified are the ones involving the point(s) being moved.

**EXAMPLE 1.79**

Suppose that we change (or move) the control points \(P_1\) and \(P_2\) of Example 1.78 to \(P_1 = (3.7, 3)\) and \(P_2 = (6.5, 4.5)\). Then, the new Bezier curve will be
\[
C(t) = (4 - 3t + 9t^2 - 3t^3), \quad 1 + 3t + 6t^2 - 7t^3), \quad t \in [0, 1].
\]

In Figure 1.10, we show how by moving these two control points, the shape of the curve has been accordingly modified, the new curve being pulled toward the new control points.
The fact that the Bezier curve follows the shape of the polygonal path determined by the control points and stays within the quadrilateral is more than just a geometrical fact or coincidence. There is a concrete linear algebra concept behind this.

**Definition 1.75** Given a set of \( n \) points: \( S = \{ P_0, \ldots, P_{n-1} \} \), the convex hull of \( S \) is the smallest convex polygon that contains all the points in \( S \).

This definition is saying that any given point in the convex hull of \( S \) is either on the boundary of the polygon or in its interior. Intuitively, we can think of each point in \( S \) as a nail on a board; then, the convex hull is the shape formed by a tight rubber band that surrounds all the nails. Two clear examples are given in Figures 1.9 and 1.10, where the convex hull of the control points is the quadrilateral given by the lines and its interior. The relationship between convex hulls and Bezier curves, which is apparent in both Figures 1.9 and 1.10, is that

A Bezier curve always lies in the convex hull of its set of control points.

**Midpoint property.** We already know that \( P_0 = (x(0), y(0)) \) and that \( P_3 = (x(1), y(1)) \). All other points of the Bezier curve are \( (x(t), y(t)) \), for some \( t \in (0, 1) \). One interesting question is whether we can characterize the point on the curve for which \( t = \frac{1}{2} \).

If we substitute \( t = \frac{1}{2} \) in the parametric equations (1.71), we obtain

\[
x(\frac{1}{2}) = \frac{1}{8}x_0 + \frac{3}{8}x_1 + \frac{3}{8}x_2 + \frac{1}{8}x_3, \\
y(\frac{1}{2}) = \frac{1}{8}y_0 + \frac{3}{8}y_1 + \frac{3}{8}y_2 + \frac{1}{8}y_3.
\]
These are the coordinates of the point $M$ on the curve. However, the interesting thing is that this same point can be obtained in a different way, that geometrically is more intuitive (see Figure 1.11):

The point $a$ is the midpoint between $P_0$ and $P_1$. The point $b$ is the midpoint between $P_1$ and $P_2$, and the point $c$ is the midpoint between $P_2$ and $P_3$. Similarly, the point $d$ is the midpoint between $a$ and $b$, and the point $e$ is the midpoint between $b$ and $c$. Finally, the sought point $M = (x(\frac{1}{2}), y(\frac{1}{2}))$ is the midpoint between $d$ and $e$. We leave the verification of the statements above as an exercise.

Furthermore, it can be shown that the part of the Bezier curve that goes from $P_0$ to $M$ can be defined using the control points, $P_0$, $a$, $d$, and $M$. Similarly, the part of the Bezier curve that goes from $M$ to $P_3$ can be defined using the control points $M$, $e$, $c$, and $P_3$.

**Bezier curve and center of mass.** An interesting way of introducing Bezier curves is through the concept of center of mass of a set of point masses. Suppose we have four masses $m_0, m_1, m_2$ and $m_3$ which are located at points $P_0, P_1, P_2$, and $P_3$ respectively. Then, the center of mass of these four point masses is

$$P = \frac{m_0 P_0 + m_1 P_1 + m_2 P_2 + m_3 P_3}{m_0 + m_1 + m_2 + m_3}.$$

Assume also that the masses are not constant but that they vary as a function of a parameter $t$ according to the equations

$$m_0 = (1 - t)^3, \quad m_1 = 3t(1 - t)^2, \quad m_2 = 3t^2(1 - t), \quad m_3 = t^3,$$

for $t \in [0,1]$. 

Since for any value of $t$ we always have that $m_0(t) + m_1(t) + m_2(t) + m_3(t) = 1$, the center of mass reduces to

$$P = m_0P_0 + m_1P_1 + m_2P_2 + m_3P_3.$$ 

Observe that for $t = 0$, the center of mass $P$ is at $P_0$ and for $t = 1$ it is located at $P_3$. As $t$ varies between 0 and 1, the center of mass $P$ moves describing a curve, a cubic Bezier curve. The masses described above are the cubic Bernstein polynomials.

### 1.12.2 Composite Bezier curves

In practical applications, curves will be more sophisticated than a single Bezier curve, but these sophisticated curves can be produced by using a sequence of Bezier curves that share common end points and then are patched together ensuring the continuity of the final curve (but not necessarily differentiability). The final curve obtained is a composite Bezier curve, also known as Bezier spline.

Since a general Bezier curve always lies in the convex hull of its control points, oscillatory behavior will not be present. Also, changes in the curve will mean making just local changes of some control points, minimizing in this way the number of total modifications.

One more observation is in order. When performing interpolation, say with cubic splines, the resulting curve is smooth, with continuous first and second derivatives, so that sharp corners are out of the question. Composite Bezier curves are more flexible: A sharp corner will be well defined, because only continuity is required (in some applications, sharp corners will be needed, such as in creating Postscript fonts), and if we need to have smoothness at the points where two Bezier curves meet, it is sufficient to require (see Exercise 1.95) that the three control points (the point where they meet, the one before, and the one after) are collinear. We illustrate these ideas with the following example.

#### EXAMPLE 1.80

Let us find the composite Bezier curve for the following sets of control points:

\[
\begin{align*}
\{(−5.0,0.5), & \ (−12,1.5), \ (−12,4.5), \ (−5.0,5.0)\}, \\
\{(−5.0,5.0), & \ (−6.0,3.0), \ (−2.0,2.5), \ (−1.5,5.0)\}, \\
\{(−1.5,5.0), & \ (−1.5,5.0), \ (−1.0,4.4), \ (−1.0,4.4)\}, \\
\{(−1.0,4.4), & \ (−0.5,4.6), \ (0.5,4.6), \ (1.0,4.4)\}, \\
\{(1.0,4.4), & \ (1.0,4.4), \ (1.5,5.0), \ (1.5,5.0)\}, \\
\{(1.5,5.0), & \ (2.0,2.5), \ (6.0,3.0), \ (5.0,5.0)\}, \\
\{(5.0,5.0), & \ (12,4.5), \ (12,1.5), \ (5.0,5.0)\}.
\end{align*}
\]

Following the same process as in Example 1.78 to find one Bezier curve $C(t)$, we start by finding the following seven Bezier curves corresponding to the seven sets of
Figure 1.12 Composite Bezier curve of Example 1.80.

points above. This gives:

\[
\begin{align*}
C_1(t) &= (21t^2 - 21t - 5, -4.5t^3 + 6t^2 + 3t + 0.5), \\
C_2(t) &= (-8.5t^3 + 15t^2 - 3t - 5, 1.5t^3 + 4.5t^2 - 6t + 5), \\
C_3(t) &= (-t^3 + 1.5t^2 - 1.5, 1.2t^3 - 1.8t^2 + 5), \\
C_4(t) &= (0.6t^3 + 0.3t^2 + 2.7t - 1.4, -0.6t^3 + 0.6t + 4.4), \\
C_5(t) &= (-t^3 + 1.5t^2 + 1, -1.2t^3 + 1.8t^2 + 4.4), \\
C_6(t) &= (-8.5t^3 + 10.5t^2 + 1.5t + 1.5, -1.5t^3 + 9t^2 - 7.5t + 5), \\
C_7(t) &= (-21t^2 + 21t + 5, 9t^3 - 7.5t^2 - 1.5t + 5).
\end{align*}
\]

The resulting composite Bezier curve is shown in Figure 1.12 along with the control points. Observe that we can have sharp corners wherever needed. Also observe that if a particular Bezier curve is to be a line segment, then the corresponding four points are listed by repeating them. For instance, for \(C_3\), we have the control points

\[(-1.5, 5), (-1.5, 5), (-1, 4.4), (-1, 4.4).\]

One very important application of composite Bezier curves is the design of fonts. The polynomials used to create fonts could be quadratic or cubic (even linear), depending on the application. True Type typically uses quadratic composite Bezier curves, while Postscript uses cubic ones. In the next example, we want to illustrate this real-world application of Bezier curves (and hence of Bernstein polynomials and basis of a vector space) in creating a postscript font.

\section*{Example 1.81}

We can use the following set of points (listed without parenthesis for convenience) to generate the Times Roman character \(\text{R}\) as the composition of 22 Bezier curves. This
is illustrated in Figure 1.13.

\[
\begin{array}{cccccccc}
0.00 & 5.70 & 0.00 & 5.70 & 0.00 & 5.55 & 0.00 & 5.55 \\
0.00 & 5.55 & 0.60 & 5.55 & 0.80 & 5.35 & 0.80 & 4.80 \\
0.80 & 4.80 & 0.80 & 4.80 & 0.80 & 0.90 & 0.80 & 0.90 \\
0.80 & 0.90 & 0.80 & 0.30 & 0.60 & 0.15 & 0.00 & 0.15 \\
0.00 & 0.15 & 0.00 & 0.15 & 0.00 & 0.00 & 0.00 & 0.00 \\
0.00 & 0.00 & 0.00 & 0.00 & 2.40 & 0.00 & 2.40 & 0.00 \\
2.40 & 0.00 & 2.40 & 0.00 & 2.40 & 0.15 & 2.40 & 0.15 \\
2.40 & 0.15 & 1.80 & 0.15 & 1.65 & 0.30 & 1.65 & 0.90 \\
1.65 & 0.90 & 1.65 & 0.90 & 1.65 & 2.60 & 1.65 & 2.60 \\
1.65 & 2.60 & 1.65 & 2.60 & 2.65 & 2.62 & 2.65 & 2.62 \\
2.65 & 2.62 & 2.65 & 2.62 & 4.10 & 0.00 & 4.10 & 0.00 \\
4.10 & 0.00 & 4.10 & 0.00 & 5.50 & 0.00 & 5.50 & 0.00 \\
5.50 & 0.00 & 5.50 & 0.00 & 5.50 & 0.15 & 5.50 & 0.15 \\
5.50 & 0.15 & 5.15 & 0.15 & 4.95 & 0.20 & 4.76 & 0.50 \\
4.76 & 0.50 & 4.80 & 0.50 & 3.50 & 2.74 & 3.50 & 2.74 \\
3.50 & 2.74 & 5.38 & 3.00 & 5.10 & 5.70 & 3.05 & 5.70 \\
3.05 & 5.70 & 3.05 & 5.70 & 0.00 & 5.70 & 0.00 & 5.70 \\
1.65 & 3.00 & 1.65 & 3.00 & 1.65 & 5.00 & 1.65 & 5.00 \\
1.65 & 5.00 & 1.65 & 5.30 & 1.75 & 5.35 & 1.84 & 5.35 \\
1.84 & 5.35 & 1.84 & 5.35 & 2.30 & 5.35 & 2.30 & 5.35 \\
2.30 & 5.35 & 4.40 & 5.20 & 4.20 & 3.00 & 2.40 & 3.00 \\
2.40 & 3.00 & 2.40 & 3.00 & 1.65 & 3.00 & 1.65 & 3.00 \\
\end{array}
\]

1.13 FINAL REMARKS AND FURTHER READING

In this chapter, we have introduced those concepts and techniques of linear algebra and matrix analysis that are needed for some of the applications covered in this book. In particular, in Section 1.12, concepts like linear combination, basis of a vector space, convex
hull, and so on. have been illustrated as very useful tools for the construction of Bezier curves and postscript fonts. More applications are presented in the following chapters.

Linear algebra and matrix analysis offer several other very interesting concepts and techniques not covered here, but that are also very important in applications. An extensive study of matrix norms and their properties, matrix computations, and matrix analysis in general can be found in the classic books by Horn and Johnson [38] and Golub and Van Loan [25], the latter especially focused on the numerical and computational aspects of matrix algebra. An additional excellent reference for matrix computations, linear algebra, and their applications is the book by Meyer [49].

**EXERCISES**

1.1 Given two matrices $A$ and $B$, show that

$$(A + B)^T = A^T + B^T, \quad \text{and} \quad (AB)^T = B^TA^T,$$

whenever the sum or the product, respectively, are well defined.

1.2 Show that for two matrices $A$ and $B$ whose product is well defined,

$$\text{tr}(AB) = \text{tr}(BA).$$

1.3 Find a nonzero vector $x = [x_1 \ x_2]^T$ such that the sum, Euclidean, and maximum norm all coincide.

1.4 Show that for any $n$-dimensional vector $x$,

$$\|x\|_{\infty} \leq \|x\|_2 \leq \|x\|_1.$$

1.5 Prove the Cauchy–Schwarz inequality:

$$|x^Ty| \leq \|x\|_2 \|y\|_2,$$

for any two vectors $x, y \in \mathbb{R}^n$.

1.6 Show that the Frobenius norm of a matrix is consistent, in the sense that

$$\|AB\|_F \leq \|A\|_F \|B\|_F,$$

for any two matrices $A, B$ whose product is well-defined.

1.7 Let $A_{m \times n}$ be a matrix such that $\|Ax\|_2 \leq \|x\|_2$, for all vectors $x \in \mathbb{R}^n$. Show that also $\|ATy\|_2 \leq \|y\|_2$, for all vectors $y \in \mathbb{R}^m$.

1.8 Let $A$ be an $m \times n$ matrix, and $x \in \mathbb{R}^n$. Show that

$$\|Ax\|_2 = \|x\|_2 \quad \text{if and only if} \quad AT^TA = I.$$

1.9 Show that for any matrix norm induced from a vector norm as in (1.9), we have
for any two matrices $A$ and $B$ for which the product is well-defined.

1.10 Let $A = \begin{bmatrix} 3 & -2 & 5 \\ 0 & 8 & -1 \\ 2 & 2 & 7 \end{bmatrix}$. Find $\|A\|_F$ using the definitions given in (1.6), (1.7), and (1.8).

1.11 We may want to extend the vector norm $\|x\| = \max_i |x_i|$ into a matrix norm by defining

$$\|A\| = \max_{1 \leq i,j \leq n} |a_{ij}|.$$ 

However, this function cannot be considered a matrix norm because it does not satisfy the submultiplicativity inequality (1.72). Provide a counterexample to confirm this statement.

1.12 Show that for $p = 1$ and $p = \infty$, the $p$-norms in (1.9) coincide with the corresponding norms in (1.6).

1.13 Show that the Frobenius norm in (1.6) can also be defined in the following two ways:

(a) $\|A\|_F^2 = \|a_1\|_2^2 + \cdots + \|a_n\|_2^2$.

(b) $\|A\|_F^2 = \text{tr}(A^TA)$.

Here, $a_1, \ldots, a_n$ denote the columns of $A_{m \times n}$.

1.14 Let $A$ be an $m \times n$ matrix. Show that

$$\|A\|_2 \leq \sqrt{\|A\|_\infty \|A\|_1}.$$ 

1.15 Let $A_{n \times n}$ be an arbitrary matrix. We define its spectral radius as $\rho(A) = \max_{i=1,\ldots,n} |\lambda_i|$, where $\lambda_i$ are the eigenvalues of $A$. Show that

$$\rho(A) \leq \|A\|,$$

for any matrix norm $\| \cdot \|$.

1.16 Show that the product of two square lower triangular matrices is lower triangular.

1.17 True or False? A symmetric Hessenberg matrix is tridiagonal, that is, a matrix whose nonzero elements are on the diagonal, the diagonal below and the diagonal above the main diagonal.

1.18 Show that the inverse of a nonsingular matrix $A_{n \times n}$ is always unique.

1.19 Let $A$ be a nonsingular square matrix of order $n$. Show that if $AX = I$, for some matrix $X_{n \times n}$, then also $XA = I$.

1.20 Show that if $A_{n \times n}$ is nonsingular, then $A^T$ is also nonsingular and

$$(A^T)^{-1} = (A^{-1})^T.$$ 

(We usually write this matrix as $A^{-T}$.)
1.21 Let the matrices $A_1, A_2, \ldots, A_k$ be nonsingular. Show that 
\[(A_1 A_2 \cdots A_k)^{-1} = A_k^{-1} A_{k-1}^{-1} \cdots A_1^{-1}.\]

1.22 Show that a matrix $A_{n \times n}$ is nonsingular if there is a matrix norm such that $\|I - A\| < 1$. Then show that in such a case, 
\[A^{-1} = \sum_{k=0}^{\infty} (I - A)^k.\]

1.23 Let $A$ be a square matrix of order $n$ such that $\|I - A\| < 1$, for some matrix norm $\| \cdot \|$. Show that 
\[\|A^{-1}\| \leq \frac{1}{1 - \|I - A\|}.\]

1.24 Let $A_{n \times n}$ be an arbitrary matrix that can be expressed as $A = M - N$, for some matrices $M$ and $N$, with $M$ nonsingular, and let $B = M^{-1} N$. Show that if $\|B\| < 1$, for some matrix norm, then $A$ is nonsingular.

1.25 Let $A = \begin{bmatrix} 1 & a & a \\ 0 & 1 & a \\ 0 & 0 & 1 \end{bmatrix}$. Show that its inverse is given by the matrix 
\[A^{-1} = I + B + B^2,\]

where $B = I - A$. Find $A^{-1}$ and $B^3$ explicitly.

1.26 Let $A$ be a square matrix of order $n$ that is strictly diagonal dominant; that is, 
\[|a_{ii}| > \sum_{j \neq i} |a_{ij}|, \quad i = 1, 2, \ldots, n. \quad (1.73)\]

Show that $A$ is nonsingular.

**Note:** If the inequality in (1.73) is not strict, we say $A$ is weakly diagonal dominant.

1.27 Give an example of a matrix that is symmetric but not hermitian and an example of a matrix that is hermitian but not symmetric.

1.28 True or False? The diagonal entries of a Hermitian matrix must be real.

1.29 True or False? The sum of two symmetric matrices is a symmetric matrix.

1.30 Verify that the matrices in Example 1.13 are positive definite, and that the matrices in Example 1.11 are not.

1.31 Let $A = \begin{bmatrix} B & C \\ D & E \end{bmatrix}$ be symmetric positive definite (spd), with $B$ and $E$ square. Show that $B$ and $E$ are also spd.

1.32 Let $A$ be a square matrix of order $n$. Show that 
\[\frac{d}{dt} e^{At} = Ae^{At}.\]
1.33 Let $A_{m \times n}$ be nonsingular and assume that $e^{-At} \to 0$ as $t \to \infty$. Show that the inverse of $A$ is given by the integral

$$
\int_0^\infty e^{-At} \, dt.
$$

1.34 Show that if a matrix $A_{m \times n}$ has rank $n$, then there exists a matrix $X$ such that $XA = I$. Is the matrix $X$ unique? ($X$ is called a left inverse of $A$).

1.35 Let $v$ and $q$ be two $n$-dimensional vectors, with $\|q\| = 1$. We define the projection of $v$ along $q$ as

$$
\text{Proj}_q v = (v^T q) q,
$$

and therefore, the orthogonal projection of $v$ onto $q$ is $v - (v^T q) q$, just as in Gram–Schmidt process. Let $v = [1 \quad 0 \quad 3]^T$ and $u = [1 \quad 0 \quad 2]^T$. Find the orthogonal projection of $v$ along $q = u/\|u\|$.

1.36 The vectors $v_1 = [-1 \quad 0 \quad 1]^T$, $v_2 = [0 \quad -1 \quad 1]^T$, and $v_3 = [1 \quad 1 \quad 0]^T$ are linearly independent. Apply the Gram–Schmidt process to these vectors, but using the following dot product:

$$
x^T y := \frac{1}{3} x_1 y_1 + 2x_2 y_2 + x_3 y_3.
$$

1.37 The same row and column permutation of the matrix $A$ in Example 1.20 could have also been obtained by simply computing the product $PAP$. Is this true for any permutation matrix $P$?

1.38 Show that the columns of an orthogonal matrix $Q_{m \times n}$ are orthonormal vectors.

1.39 Let $P_{n \times n}$ and $Q_{n \times n}$ be two orthogonal matrices. Show that the matrix $PQ$ is also orthogonal. Show however that the sum $P + Q$ may fail to be orthogonal.

1.40 Let $x$, $y$ be two arbitrary vectors in $\mathbb{R}^n$ and $Q_{n \times n}$ an orthogonal matrix. Let $\theta_1$ be the angle between $x$ and $y$, and $\theta_2$ the angle between $Qx$ and $Qy$. Show that

$$
\cos \theta_1 = \cos \theta_2.
$$

1.41 Let $Q_{m \times n} = [q_1 \quad q_2 \cdots q_n]$ be an orthogonal matrix. What do you obtain when Gram–Schmidt is applied to the columns of $Q$?

1.42 A matrix norm is called **orthogonally invariant** if for an arbitrary matrix $A_{m \times n}$ and orthogonal matrices $Q$ and $U$ of appropriate dimensions, we have

$$
\|QAU^T\| = \|A\|.
$$

Show that the Frobenius and the 2-norm are orthogonally invariant.

1.43 Show that the orthogonal invariance of a norm implies that multiplication by an orthogonal matrix does not magnify errors, in the following sense: Let $A_{n \times n}$ be an arbitrary matrix and $Q_{n \times n}$ be orthogonal. If an error $E$ is introduced in $A$, then the error in $QAQ^T$ has the same norm as $E$. Also show that for a general nonsingular matrix $P$, the error in $PA P^{-1}$ is bounded by $\text{cond}(P)\|E\|$.
1.44 Let $A_{n \times n}$ be a nonsingular matrix and let $b$ be an arbitrary vector in $\mathbb{R}^n$. Let $x$ be the solution of $Ax = b$, and suppose there is a perturbation $\delta b$ on the vector $b$ so that the solution of the system is perturbed to $x + \delta x$, that is, $A(x + \delta x) = b + \delta b$. Show that

$$\frac{\|\delta x\|}{\|x\|} \leq \text{cond}(A) \frac{\|\delta b\|}{\|b\|}.$$  

1.45 Let $Q_{n \times n}$ be an orthogonal matrix. Prove that $\text{cond}(Q) = 1$.

1.46 Let $Q$ be an orthogonal matrix and $A$ be a nonsingular matrix. Show that

$$\text{cond}(QA) = \text{cond}(A).$$

1.47 Show that no two vectors in $\mathbb{R}^3$ can span all of $\mathbb{R}^3$.

1.48 Find the subspace spanned by the three vectors $[2 \ 3 \ 1]^T$, $[2 \ 1 \ -5]^T$, and $[2 \ 4 \ 4]^T$.

1.49 Show that the set $U$ in Example 1.30 is not a vector subspace of $V$.

1.50 Let $V$ be a vector space and $S = \{v_1, \ldots, v_k\}$ be an arbitrary subset of $V$. Show that the set of all linear combinations of vectors from $S$ is a subspace of $V$.

1.51 Let $B = \{v_1, \ldots, v_n\}$ be a basis of vector space $V$. Show that any subset of $V$ containing more than $n$ vectors is linearly dependent (not linearly independent).

1.52 Let $u$ and $v$ be arbitrary linearly independent vectors in $\mathbb{R}^n$. Show that for some values of the scalars $c_1$ and $c_2$, the vector $x = c_1 u + c_2 v$ has both positive and negative components.

1.53 Show that the set $\{f_1 = 2x^2 + 1, f_2 = x^2 + 4x, f_3 = x^2 - 4x + 1\}$ is linearly dependent.

1.54 Show that if the columns of a matrix $A_{m \times n}$ are linearly independent, then the matrix $A^T A$ has an inverse.

1.55 Let $S$ be a subspace of a vector space $V$. Show that

$$(S^\perp)^\perp = S.$$  

1.56 Let $B_1 = \{u_1, \ldots, u_n\}$ and $B_2 = \{w_1, \ldots, w_n\}$ be two orthonormal bases of $\mathbb{R}^n$, and let $a_1, \ldots, a_n$ and $b_1, \ldots, b_n$ be the coordinates of a vector $x \in \mathbb{R}^n$ in those bases, respectively. Show that

$$a_1^2 + \cdots + a_n^2 = b_1^2 + \cdots + b_n^2.$$  

1.57 $A = \begin{bmatrix} 4 & 0 & 1 & -1 & 1 \\ 0 & -5 & 0 & 5 & 0 \\ 0 & 0 & 3 & 4 & 1 \end{bmatrix}$. Find $\text{dim}(\text{col}(A))$ and $\text{dim}(\text{N}(A))$.

1.58 Let $A$ be an $m \times n$ matrix. Show that

$$\text{N}(A^T A) = \text{N}(A) \quad \text{and} \quad \text{rank}(A^T A) = \text{rank}(A).$$
Let \( V \) be the vector space of square real matrices of order 2, and let
\[
U = \left\{ \begin{bmatrix} 0 & a \\ 0 & b \end{bmatrix} : a, b \in \mathbb{R} \right\}, \quad W = \left\{ \begin{bmatrix} 0 & 0 \\ c & d \end{bmatrix} : c, d \in \mathbb{R} \right\}.
\]
Find \( \dim(U + W) \) and verify formula (1.38).

Consider the block matrix \( A = \begin{bmatrix} A_1 & A_2 \end{bmatrix} \), for some matrices \( A_1 \) and \( A_2 \). Show that
\[
\text{col}(A) = \text{col}(A_1) + \text{col}(A_2).
\]

Let \( P_1 \) and \( P_2 \) be orthogonal projection matrices onto two subspaces \( V_1 \) and \( V_2 \) of a vector space \( V \). Show that
\[
\text{col}(P_1 + P_2) = \text{col}(P_1) + \text{col}(P_2).
\]

Let the columns of a matrix \( Q_{m \times n} \) form an orthonormal basis of a vector subspace \( S \). Show that
\[
P = QQ^T
\]
is the orthogonal projection matrix onto \( S \).

Let \( A \) be a square matrix of order \( n \), and \( \lambda \) be an eigenvalue of \( A \). Show that the set of all eigenvectors associated with \( \lambda \), together with the zero vector, form a subspace of \( \mathbb{R}^n \).

Show that the determinant of a matrix \( A_{n \times n} \) is given by the product of its eigenvalues, that is,
\[
\det A = \lambda_1 \cdot \lambda_2 \cdots \lambda_n,
\]
where some of the eigenvalues may be repeated.

Consider the matrix
\[
A = \begin{bmatrix} 2 & 1 & 0 & 3 \\ 0 & 3 & 3 & 0 \\ 4 & 1 & 0 & 1 \\ 1 & 0 & 0 & 5 \end{bmatrix}.
\]
Without any computation of the characteristic equation, determine one of the eigenvalues of \( A \) as well as an associated eigenvector.

Suppose that the sum of the entries in each column of a square matrix \( A \) is 1. Show that 1 is an eigenvalue of \( A^T \).

True or False? The eigenvalues of a positive definite matrix are positive numbers.

Let \( u \) and \( v \) be right and left eigenvectors, respectively, of a matrix \( A_{n \times n} \) for a given eigenvalue \( \lambda \) of \( A \), with \( v^T u = 1 \). Let \( Q_{n \times (n-1)} \) be an orthogonal matrix whose columns form a basis of \( v^\perp \). Show that the matrix \( P = [u \quad Q] \) satisfies
\[
P^{-1}AP = \begin{bmatrix} \lambda & 0 \\ 0 & B \end{bmatrix},
\]
for some square matrix \( B \) of order \( n - 1 \).
1.69 Show by counterexample that the product of two symmetric matrices $A$ and $B$ is not necessarily symmetric. Show however that if $A$ and $B$ commute, then the product $AB$ is also symmetric.

1.70 True or False? The inverse of a nonsingular symmetric matrix is also symmetric.

1.71 Let $A_{n \times n}$ be a symmetric matrix. Show that all its eigenvalues are real, and that eigenvectors corresponding to distinct eigenvalues are orthogonal.

1.72 Let $U = V \oplus W$, and let $B_1, B_2$ be bases of $V$ and $W$, respectively. Prove that $B_1 \cap B_2 = \emptyset$ and that $B_1 \cup B_2$ is a basis for $U$.

1.73 For a matrix $A_{n \times n}$, define $r = Av - \lambda v \neq 0$, where $v$ is a unit vector (we say $v$ is an approximate eigenvector of $A$, with associated eigenvalue $\lambda$). Show that

$$(A + \delta A) v = \lambda v,$$

where $\delta A = -rv^T$.

1.74 Let $A_{n \times n}$ have $n$ distinct eigenvalues with associated eigenvectors $v_1, \ldots, v_n$ and left eigenvectors $w_1, \ldots, w_n$. Show that $w_i^Tv_j = 0$ if $i \neq j$, and $w_i^Tv_j \neq 0$ if $i = j$.

1.75 (Gershgorin) Let $A$ be a square matrix of order $n$, and for some nonsingular matrix $P$, let $P^{-1}AP = D + E$, with $D = \text{diag}(d_1, \ldots, d_n)$ and $E$ has zero diagonal entries. Show that

$$\sigma(A) \subseteq \bigcup_{i=1}^n D_i,$$

where $D_i = \{ z \in \mathbb{C} : |z - d_i| \leq \sum_{j=1}^n |e_{ij}| \}$.

Hint: Show $(D - \lambda I) + E$ is singular, and use Exercise 1.23 with $p = \infty$.

1.76 Let $D = \text{diag}(d_1, \ldots, d_n)$ with all $d_i$ distinct. Consider a scalar $c \neq 0$ and assume that all entries of a vector $u \in \mathbb{R}^n$ are nonzero. Define the matrix

$$A = D + cuu^T.$$ 

Show that $D$ and $A$ do not have any common eigenvalues.

1.77 Let $S$ and $K$ represent the subspaces of symmetric and skew-symmetric $n \times n$ matrices, respectively. Show that

$$R^{n \times n} = S \oplus K.$$ 

Hint: For any square matrix $A$ of order $n$, the matrix $A + A^T$ is symmetric and $A - A^T$ is skew-symmetric.

1.78 Let $A_{n \times n}$ be a symmetric matrix. Show that the matrix $iA$ is skew-hermitian.

1.79 Let $A_{n \times n}$ be a skew symmetric matrix. Show that $A$ is singular when $n$ is odd.

1.80 True or False? If $A$ is a skew symmetric matrix, then $e^A$ is orthogonal.

1.81 Let $P_{n \times n}$ be a projection matrix. Show that

$$\text{col}(P) \oplus N(P) = \mathbb{R}^n.$$
1.82 Let \( P \) be a projection matrix on a vector space \( V \). Show that for \( x \in V \),
\[
x \in \text{col}(P) \quad \text{if and only if} \quad x = Px.
\]

1.83 Let \( x, y \) be two vectors in \( \mathbb{R}^n \) such that \( y^T x = 1 \). Show that \( P = xy^T \) is a projection matrix and that
\[
\|P\|_2 = \|x\|_2 \|y\|_2.
\]

1.84 Show that if \( P \) is a projection matrix, then each eigenvalue of \( P \) is either 0 or 1.

1.85 Let \( u \) and \( v \) be right and left eigenvectors, respectively, of a matrix \( A_{n \times n} \), associated with a simple eigenvalue \( \lambda \) of \( A \). Show that
\[
P = \frac{u v^T}{v^T u}
\]
is a projection matrix onto \( N(A - \lambda I) \).

1.86 Show that if two square matrices \( A \) and \( B \) are similar, then they must have the same rank.

Hint. Use Theorem 1.44.

1.87 Let \( A \) and \( B \) be similar matrices. Show that for any positive integer \( k \),
\[
B^k = P^{-1} A^k P,
\]
for some nonsingular matrix \( P \).

1.88 True or False? Similar matrices have the same eigenvectors.

1.89 Let \( A \) be a square matrix of order \( n \). Show that \( A \) has a complete set of \( n \) linearly independent eigenvectors if and only if \( A \) is diagonalizable.

1.90 Show that two matrices \( A \) and \( B \) can be simultaneously diagonalized if and only if \( AB = BA \).

1.91 Let \( A_{n \times n} \) be a symmetric matrix. Show that the algebraic multiplicity of any eigenvalue of \( A \) is equal to its geometric multiplicity.

1.92 Let \( A \) and \( B \) be two diagonalizable matrices such that \( AB = BA \). Show that
\[
e^{A+B} = e^A e^B.
\]

1.93 Show that the matrix norm \( \| \cdot \|_2 \) in (1.9) with \( p = 2 \), can also be defined as
\[
\|A\|_2 = \sqrt{\max \lambda (A^T A)}.
\]

1.94 Is it possible to use quadratic Bernstein polynomials and still satisfy the tangential properties at both endpoints of a Bezier curve?

1.95 Show that to guarantee smoothness at a point \( P_i \) where two Bezier curves meet, the three points \( P_{i-1}, P_i, P_{i+1} \) must be collinear.
Figure 1.14  Control points for Exercise 1.97.

1.96  Verify the midpoint property of Bezier curves; that is, the point \( M = (x(\frac{1}{2}), y(\frac{1}{2})) \) can be obtained by first computing five other midpoints. Refer to Figure 1.11.

1.97  Sketch the Bezier curves corresponding to the control points in Figure 1.14.

1.98  Show that the slope of the Bezier curve at the midpoint \( M \) coincides with the slope of the segment joining \( d \) and \( M \).

1.99  True or False?  It is possible for the control points \( P_1 \) and \( P_2 \) to lie on opposite sides of the corresponding Bezier curve.

1.100 True or False?  It is not possible to draw an ellipse with one single Bezier curve.