Kjell Elfström

Linear Algebra

Linear Algebra

Kjell Elfström

Third edition, September 2024

Contents

1 Matrices

1.1 Definition

A matrix is a rectangular array of real numbers:

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

The size of a matrix with m rows and n columns is $m \times n$, which is read as 'm by n'. The number a_{ik} in row i and column k is called an entry. We shall also use the more compact notation $A = [a_{ik}]_{m\times n}$, $A = [a_{ik}]$, $A = [A_{ik}]_{m\times n}$ or $A = [A_{ik}]$. If $m = n$, we say that A is a square matrix of order n, and then the entries a_{ii} are said to form the main diagonal of A. By a column matrix we shall mean an $m \times 1$ matrix with a single column, and a row matrix is a $1 \times n$ matrix having a single row.

Example 1.1.

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 1 & 1 & 1 \end{bmatrix}, \quad B = \begin{bmatrix} 1 & 2 \\ 2 & 5 \\ 3 & 7 \end{bmatrix}, \quad C = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}, \quad D = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 3 & 4 \\ 3 & 4 & 5 \end{bmatrix}.$$

The sizes of these matrices are 2×3, 3×2, 3×1 and 3×3, respectively. D is a square matrix of order 3, and its main diagonal comprises the entries 1, 3 and 5.

1.2 Addition and Multiplication of Matrices

Definition 1.2. When the matrices $A = [a_{ik}]_{m\times n}$ and $B = [b_{ik}]_{m\times n}$ are of the same size, we define their sum as

$$A + B = [a_{ik} + b_{ik}]_{m\times n} = \begin{bmatrix} a_{11} + b_{11} & a_{12} + b_{12} & \cdots & a_{1n} + b_{1n} \\ a_{21} + b_{21} & a_{22} + b_{22} & \cdots & a_{2n} + b_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} + b_{m1} & a_{m2} + b_{m2} & \cdots & a_{mn} + b_{mn} \end{bmatrix}.$$

The $m \times n$ zero matrix is the matrix

$$0_{m\times n} = [0]_{m\times n} = \begin{bmatrix} 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 0 \end{bmatrix}.$$

When the size of the zero matrix is clear from context, we shall simply denote it by 0.

Example 1.3.

$$\begin{bmatrix}1 & 2 & 3\\ 0 & 1 & 1\end{bmatrix} + \begin{bmatrix}1 & 1 & 1\\ 2 & 1 & 2\end{bmatrix} = \begin{bmatrix}1+1 & 2+1 & 3+1\\ 0+2 & 1+1 & 1+2\end{bmatrix} = \begin{bmatrix}2 & 3 & 4\\ 2 & 2 & 3\end{bmatrix},$$

$$0 + \begin{bmatrix}1 & 2\\ 2 & 3\\ 3 & 4\end{bmatrix} = \begin{bmatrix}0 & 0\\ 0 & 0\\ 0 & 0\end{bmatrix} + \begin{bmatrix}1 & 2\\ 2 & 3\\ 3 & 4\end{bmatrix} = \begin{bmatrix}0+1 & 0+2\\ 0+2 & 0+3\\ 0+3 & 0+4\end{bmatrix} = \begin{bmatrix}1 & 2\\ 2 & 3\\ 3 & 4\end{bmatrix},$$

whereas

$$\begin{bmatrix}1 & 2 & 3\\ 0 & 1 & 1\end{bmatrix} + \begin{bmatrix}1 & 2\\ 2 & 3\\ 3 & 4\end{bmatrix}$$

is not defined.

Definition 1.4. Let $A = [a_{ik}]_{m\times n}$ be a matrix and s a real number. We define the product of s and A as

$$sA = [sa_{ik}]_{m\times n} = \begin{bmatrix}sa_{11} & sa_{12} & \cdots & sa_{1n}\\ sa_{21} & sa_{22} & \cdots & sa_{2n}\\ \vdots & \vdots & & \vdots\\ sa_{m1} & sa_{m2} & \cdots & sa_{mn}\end{bmatrix}.$$

This operation is called scalar multiplication of matrices. We also define $-A$ as $(-1)A$.

Example 1.5.

$$3\begin{bmatrix}1 & 2 & 3\\ 0 & 1 & 1\end{bmatrix} = \begin{bmatrix}3 & 6 & 9\\ 0 & 3 & 3\end{bmatrix}.$$

The reader should be able to verify the following rules of calculation.

Theorem 1.6. Below, s and t are real numbers, and A, B and C are matrices of the same size.

(i) $A + B = B + A$.
(ii) $A + (B + C) = (A + B) + C$.
(iii) $0 + A = A$.
(iv) $A + (-A) = 0$.
(v) $s(A + B) = sA + sB$.
(vi) $(st)A = s(tA)$.

Our next objective is to define multiplication of matrices in a way that enables us to write systems of linear equations in a compact manner. Therefore, we begin by defining the product AB of an $m \times n$ matrix $A = [a_{ik}]_{m\times n}$ and an $n \times 1$ matrix $B = [b_{ik}]_{n\times 1}$. Note that the number of columns of A and the number of rows of B are equal. We define

$$AB = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix} = \begin{bmatrix} a_{11}b_1 + a_{12}b_2 + \cdots + a_{1n}b_n \\ a_{21}b_1 + a_{22}b_2 + \cdots + a_{2n}b_n \\ \vdots \\ a_{m1}b_1 + a_{m2}b_2 + \cdots + a_{mn}b_n \end{bmatrix}.$$

As we can see, the product AB is a column matrix of size $m \times 1$. Now let

$$X = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \quad \text{and} \quad Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix}.$$

Then

$$\begin{aligned} AX = Y \quad &\Leftrightarrow \quad \begin{bmatrix} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n \\ \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix} \\ &\Leftrightarrow \quad \begin{cases} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = y_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n = y_2 \\ \qquad\vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n = y_m \end{cases}. \end{aligned}$$

Thus we achieved the goal we set out to ourselves. We shall now extend the definition to $n \times p$ matrices B. Denote by B_k the kth column of B. The matrix AB is the $m \times p$ matrix comprising the columns AB_k, $k = 1, 2, \ldots, p$.

Definition 1.7. Let $A = [a_{ik}]_{m\times n}$ and $B = [b_{ik}]_{n\times p}$ be matrices of size $m \times n$ and $n \times p$, respectively. The product AB is the $m \times p$ matrix $C = [c_{ik}]_{m\times p}$ for which

$$c_{ik} = a_{i1}b_{1k} + a_{i2}b_{2k} + \cdots + a_{in}b_{nk} = \sum_{j=1}^{n} a_{ij}b_{jk}, \quad 1 \le i \le m,\ 1 \le k \le p.$$

Hence, the entry in position i, k of the product AB is the sum of the products of the corresponding entries of row i of A and column k of B.

Example 1.8.

$$\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 1 & 2 \\ 1 & 2 \end{bmatrix} = \begin{bmatrix} 1\cdot1+2\cdot1+3\cdot1 & 1\cdot2+2\cdot2+3\cdot2 \\ 4\cdot1+5\cdot1+6\cdot1 & 4\cdot2+5\cdot2+6\cdot2 \\ 7\cdot1+8\cdot1+9\cdot1 & 7\cdot2+8\cdot2+9\cdot2 \end{bmatrix} = \begin{bmatrix} 6 & 12 \\ 15 & 30 \\ 24 & 48 \end{bmatrix},$$

whereas

$$\begin{bmatrix} 1 & 2 \\ 1 & 2 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix}$$

is undefined. As we see, BA need not be defined even though AB is.

$$\begin{bmatrix} 1 & 1 \\ 2 & 2 \end{bmatrix} \begin{bmatrix} 2 & -1 \\ 2 & -1 \end{bmatrix} = \begin{bmatrix} 1\cdot 2 + 1\cdot 2 & 1\cdot(-1) + 1\cdot(-1) \\ 2\cdot 2 + 2\cdot 2 & 2\cdot(-1) + 2\cdot(-1) \end{bmatrix} = \begin{bmatrix} 4 & -2 \\ 8 & -4 \end{bmatrix},$$
$$\begin{bmatrix} 2 & -1 \\ 2 & -1 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 2 & 2 \end{bmatrix} = \begin{bmatrix} 2\cdot 1 + (-1)\cdot 2 & 2\cdot 1 + (-1)\cdot 2 \\ 2\cdot 1 + (-1)\cdot 2 & 2\cdot 1 + (-1)\cdot 2 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} = 0.$$

Here we see that the commutative law and the cancellation law fail for matrix multiplication.

Theorem 1.9. In the identities below, $s \in \mathbf{R}$ and A, B, C are matrices. All members of an identity are defined whenever any member is defined.

(i) $A(B+C) = AB + AC$.
(ii) $(A+B)C = AC + BC$.
(iii) $A(BC) = (AB)C$.
(iv) $s(AB) = (sA)B = A(sB)$.

Proof. (i) In order that any member, hence both members, be defined, it is necessary and sufficient that the sizes of A, B and C are $m \times n$, $n \times p$ and $n \times p$, respectively. We then have

$$\begin{aligned}(A(B+C))_{ik} &= \sum_{\nu=1}^{n} A_{i\nu}(B+C)_{\nu k} = \sum_{\nu=1}^{n} A_{i\nu}(B_{\nu k} + C_{\nu k}) = \sum_{\nu=1}^{n} A_{i\nu}B_{\nu k} + \sum_{\nu=1}^{n} A_{i\nu}C_{\nu k} \\ &= (AB)_{ik} + (AC)_{ik} = (AB + AC)_{ik}\end{aligned}$$

for all i and k such that $1 \le i \le m$ and $1 \le k \le p$. This proves the statement.

The proof of (ii) is similar and is left to the reader.

(iii) We may here assume that the sizes of A, B and C are $m \times n$, $n \times p$ and $p \times q$, respectively. Then

$$\begin{aligned}(A(BC))_{ik} &= \sum_{\nu=1}^{n} A_{i\nu}(BC)_{\nu k} = \sum_{\nu=1}^{n} A_{i\nu} \sum_{\mu=1}^{p} B_{\nu\mu}C_{\mu k} = \sum_{\nu=1}^{n}\sum_{\mu=1}^{p} A_{i\nu}B_{\nu\mu}C_{\mu k} \\ &= \sum_{\mu=1}^{p}\sum_{\nu=1}^{n} A_{i\nu}B_{\nu\mu}C_{\mu k} = \sum_{\mu=1}^{p}\left(\sum_{\nu=1}^{n} A_{i\nu}B_{\nu\mu}\right)C_{\mu k} = \sum_{\mu=1}^{p} (AB)_{i\mu}C_{\mu k} \\ &= ((AB)C)_{ik}.\end{aligned}$$

The simple proof of (iv) is also left to the reader. ■

Definition 1.10. The unit matrix of order n is the square matrix

$$I^{(n)} = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{bmatrix}.$$

The entries along the main diagonal are ones and the remaining entries are zeros. When the order is clear from context, we shall simply write this matrix as I.

The following theorem is an immediate consequence of the definition.

Theorem 1.11. Let A be an $m \times n$ matrix. Then $I^{(m)}A = AI^{(n)} = A$.

1.3 The Transpose of a Matrix

Definition 1.12. Let

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

be an $m \times n$ matrix. The transpose of A is the $n \times m$ matrix

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

Hence, the transpose of A is the matrix whose columns are the rows of A.

Theorem 1.13. Below, $s \in \mathbf{R}$ and A and B are matrices. Both sides of an identity are defined whenever any side is defined.

(i) $(A + B)^t = A^t + B^t$.
(ii) $(sA)^t = sA^t$.
(iii) $(AB)^t = B^t A^t$.

Proof. The first two statements are simple consequences of the definition. To prove the last statement, suppose that the sizes of A and B are $m \times n$ and $n \times p$, respectively. Then

$$((AB)^t)_{ik} = (AB)_{ki} = \sum_{j=1}^{n} A_{kj} B_{ji} = \sum_{j=1}^{n} (A^t)_{jk} (B^t)_{ij} = (B^t A^t)_{ik},$$

which proves the claim. ∎

Definition 1.14. A matrix A is said to be symmetric if $A^t = A$.

1.4 The Inverse of a Matrix

Definition 1.15. Let A be a square matrix of order n. We say that A is invertible if there exists a square matrix B of order n such that

$$AB = BA = I.$$

Suppose that $AB = BA = I$ and $AC = CA = I$. We then have

$$B = IB = (CA)B = C(AB) = CI = C.$$

This shows that the matrix B in the definition is unique when it exists.

Definition 1.16. Let A be an invertible square matrix. The inverse A^{-1} of A is the unique matrix B in the definition above.

Example 1.17. Let

$$A = \begin{bmatrix} 1 & 2 \\ 3 & 5 \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} -5 & 2 \\ 3 & -1 \end{bmatrix}.$$

Then

$$AB = \begin{bmatrix} 1(-5) + 2 \cdot 3 & 1 \cdot 2 + 2(-1) \\ 3(-5) + 5 \cdot 3 & 3 \cdot 2 + 5(-1) \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = I$$

and

$$BA = \begin{bmatrix} (-5) \cdot 1 + 2 \cdot 3 & (-5) \cdot 2 + 2 \cdot 5 \\ 3 \cdot 1 + (-1) \cdot 3 & 3 \cdot 2 + (-1) \cdot 5 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = I.$$

Hence A is invertible with inverse $A^{-1} = B$.

Example 1.18. Let

$$A = \begin{bmatrix} 2 & -1 \\ 2 & -1 \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} 1 & 1 \\ 2 & 2 \end{bmatrix}.$$

As we saw in Example 1.8, $AB = 0$. From this it follows that A cannot be invertible. If it were, we should have $0 = A^{-1}0 = A^{-1}(AB) = (A^{-1}A)B = IB = B$, which is false.

Theorem 1.19. Let A be a square matrix of order n. Then A is invertible if and only if, for every $n \times 1$ matrix Y, there is a unique $n \times 1$ matrix X such that $AX = Y$.

Proof. First assume that A is invertible with inverse A^{-1}. If $AX = Y$, then we have $X = IX = A^{-1}AX = A^{-1}Y$, which shows that the equation $AX = Y$ can have no more than one solution X. Since $AA^{-1}Y = IY = Y$, we see that $X = A^{-1}Y$ is, in fact, a solution. This proves the implication to the right.

To show the converse, we assume that, for every $n \times 1$ matrix Y, there is a unique $n \times 1$ matrix X such that $AX = Y$. Let I_k be the kth column of the unit matrix I. By assumption, there is an $n \times 1$ matrix B_k such that $AB_k = I_k$. Let B be the matrix comprising the columns B_k, $k = 1, 2, \ldots, n$. Then $AB = I$. It remains to show that $BA = I$. Let $C = BA$. Then we have $AC = A(BA) = (AB)A = IA = A = AI$. Hence $AC_k = AI_k$ for $k = 1, 2, \ldots, n$, and it follows from the uniqueness that $C_k = I_k$ for $k = 1, 2, \ldots, n$. From this we conclude that $BA = C = I$. ∎

Sometimes the following simple theorem proves useful.

Theorem 1.20. Let A and B be $m \times n$ matrices. Then $AX = BX$ for all $n \times 1$ matrices X if and only if $A = B$.

Proof. If $AX = BX$ for all $n \times 1$ matrices X, then $AI_k = BI_k$ for all columns I_k of the $n \times n$ unit matrix I. Hence, $A = AI = BI = B$. The reverse implication is immediate. ∎

Theorems 1.19 and 1.20 can be used to devise a method for finding the inverse when it exists or disclosing its non-existence. If we find that the system $AX = Y$ has a unique solution $X = BY$, then A is invertible by Theorem 1.19. Hence also $X = A^{-1}Y$ is a solution. It follows that $BY = A^{-1}Y$ for all $n \times 1$ matrices Y and so, by Theorem 1.20, $B = A^{-1}$. If, instead, we find that the system has not a unique solution for some right-hand side Y, then A is not invertible by Theorem 1.19.

Example 1.21. Let us determine whether the matrix

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 1 & 3 & 5 \\ 1 & 4 & 6 \end{bmatrix}$$

is invertible. We have

$$AX = Y \quad \Leftrightarrow \quad \begin{cases} x_1 + 2x_2 + 3x_3 = y_1 \\ x_1 + 3x_2 + 5x_3 = y_2 \\ x_1 + 4x_2 + 6x_3 = y_3 \end{cases} \quad \Leftrightarrow \quad \begin{cases} x_1 + 2x_2 + 3x_3 = y_1 \\ x_2 + 2x_3 = -y_1 + y_2 \\ 2x_2 + 3x_3 = -y_1 + y_3 \end{cases}$$

$$\Leftrightarrow \quad \begin{cases} x_1 + 2x_2 + 3x_3 = y_1 \\ x_2 + 2x_3 = -y_1 + y_2 \\ - x_3 = y_1 - 2y_2 + y_3 \end{cases} \quad \Leftrightarrow \quad \begin{cases} x_1 = 2y_1 - y_3 \\ x_2 = y_1 - 3y_2 + 2y_3 \\ x_3 = -y_1 + 2y_2 - y_3 \end{cases}.$$

From this it follows that A is invertible and that

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

Of course, it suffices to keep track of the coefficients of the x_i and the y_i. Thus the above computations can be written as

$$AX = Y \quad \Leftrightarrow \quad \left[\begin{array}{ccc|ccc} 1 & 2 & 3 & 1 & 0 & 0 \\ 1 & 3 & 5 & 0 & 1 & 0 \\ 1 & 4 & 6 & 0 & 0 & 1 \end{array}\right] \quad \Leftrightarrow \quad \left[\begin{array}{ccc|ccc} 1 & 2 & 3 & 1 & 0 & 0 \\ 0 & 1 & 2 & -1 & 1 & 0 \\ 0 & 2 & 3 & -1 & 0 & 1 \end{array}\right]$$

$$\Leftrightarrow \quad \left[\begin{array}{ccc|ccc} 1 & 2 & 3 & 1 & 0 & 0 \\ 0 & 1 & 2 & -1 & 1 & 0 \\ 0 & 0 & -1 & 1 & -2 & 1 \end{array}\right] \quad \Leftrightarrow \quad \left[\begin{array}{ccc|ccc} 1 & 0 & 0 & 2 & 0 & -1 \\ 0 & 1 & 0 & 1 & -3 & 2 \\ 0 & 0 & 1 & -1 & 2 & -1 \end{array}\right].$$

Example 1.22. The matrix

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 1 & 3 & 5 \\ 1 & 4 & 7 \end{bmatrix}$$

differs from the matrix in the previous example only in its lower right position. This time we get

$$\begin{aligned} AX = Y \quad & \Leftrightarrow \quad \left[\begin{array}{ccc|ccc} 1 & 2 & 3 & 1 & 0 & 0 \\ 1 & 3 & 5 & 0 & 1 & 0 \\ 1 & 4 & 7 & 0 & 0 & 1 \end{array}\right] \quad \Leftrightarrow \quad \left[\begin{array}{ccc|ccc} 1 & 2 & 3 & 1 & 0 & 0 \\ 0 & 1 & 2 & -1 & 1 & 0 \\ 0 & 2 & 4 & -1 & 0 & 1 \end{array}\right] \\ & \Leftrightarrow \quad \left[\begin{array}{ccc|ccc} 1 & 2 & 3 & 1 & 0 & 0 \\ 0 & 1 & 2 & -1 & 1 & 0 \\ 0 & 0 & 0 & 1 & -2 & 1 \end{array}\right]. \end{aligned}$$

We see that the system has an infinite number of solutions when $y_1 - 2y_2 + y_3 = 0$ and no solutions otherwise. Hence A is not invertible.

Theorem 1.23. Let A and B be invertible square matrices. Then AB and A^t are invertible, $(AB)^{-1} = B^{-1}A^{-1}$ and $(A^t)^{-1} = (A^{-1})^t$.

Proof. The statements follow from

$$\begin{aligned} (AB)(B^{-1}A^{-1}) &= A(BB^{-1})A^{-1} = AIA^{-1} = AA^{-1} = I, \\ (B^{-1}A^{-1})(AB) &= B^{-1}(A^{-1}A)B = B^{-1}IB = B^{-1}B = I \end{aligned}$$

and

$$\begin{aligned} A^t(A^{-1})^t &= (A^{-1}A)^t = I^t = I, \\ (A^{-1})^tA^t &= (AA^{-1})^t = I^t = I. \ \blacksquare \end{aligned}$$

Exercises

1.1. Let

$$A = \begin{bmatrix} 1 & 2 & 3 \\ -1 & 1 & 2 \\ 1 & 1 & -1 \end{bmatrix}, \quad B = \begin{bmatrix} 1 & 0 & 1 \\ 1 & 2 & 1 \\ 2 & 1 & 3 \end{bmatrix}, \quad C = \begin{bmatrix} 2 & 2 & 1 \\ 1 & 1 & 3 \end{bmatrix}.$$

Perform the following operations or explain why they are not defined.

(a) $A + B$, (b) $A + C$, (c) AB, (d) BA,
(e) AC, (f) CA, (g) $C(A + B)$, (h) $C(A + 2B)$.

1.2. Let

$$A = \begin{bmatrix} 1 & 2 \\ -1 & 1 \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} 1 & 0 \\ 1 & 2 \end{bmatrix}.$$

Compute $A^2 - B^2$ and $(A + B)(A - B)$ and explain what you observe.

1.3. Let
$$A = \begin{bmatrix} 2 & 4 \\ 1 & 2 \end{bmatrix}.$$
Find all 2×2 matrices B such that
$$AB = BA = 0.$$

1.4. (a) Let A and be B be $n \times n$ matrices such that $AB = BA$. Show that the binomial theorem
$$(A + B)^n = \sum_{k=0}^{n} \binom{n}{k} A^{n-k} B^k$$
holds for A and B. Explain why the commutativity condition is necessary.

(b) Compute $(I + A)^6$ where
$$A = \begin{bmatrix} 0 & 2 & 3 \\ 0 & 0 & 4 \\ 0 & 0 & 0 \end{bmatrix}.$$

1.5. Let A, B and C be the matrices in Exercise 1.1. Compute $A^t B^t$ and $(A^t + B^t)C^t$.

1.6. Show that $A^t A$ and AA^t are symmetric matrices for all matrices A.

1.7. A matrix A is said to be skew-symmetric if $A^t = -A$.

(a) Let A be a square matrix. Show that $A + A^t$ is symmetric and that $A - A^t$ is skew-symmetric.

(b) Show that, for every square matrix A, there exist unique matrices B and C such that $A = B + C$, B is symmetric and C is skew-symmetric.

1.8. Find the inverse of each matrix below or explain why it does not exists.
$$\text{(a)} \begin{bmatrix} 1 & 2 & 3 \\ 1 & 1 & 2 \\ 1 & 1 & 1 \end{bmatrix}, \qquad \text{(b)} \begin{bmatrix} 1 & 2 & 3 \\ 1 & 3 & 2 \\ 1 & 1 & 4 \end{bmatrix}, \qquad \text{(c)} \begin{bmatrix} 1 & 2 & 3 \\ 1 & 3 & 2 \\ 2 & 1 & 4 \end{bmatrix}.$$

1.9. Compute the inverses of A, A^t and A^2 where
$$A = \begin{bmatrix} 2 & 1 & 1 \\ 1 & 1 & 2 \\ 2 & 1 & 2 \end{bmatrix}.$$

1.10. For which values of a is the matrix
$$\begin{bmatrix} 2 & 0 & 1 \\ 1 & 3 & 2 \\ 2 & 4 & a \end{bmatrix}$$
invertible? Find the inverse for those values.

1.11. Solve the matrix equation

$$AXB = C$$

where

$$A = \begin{bmatrix} 2 & 3 \\ 3 & 4 \end{bmatrix}, \quad B = \begin{bmatrix} 2 & 1 & 2 \\ 1 & 1 & 2 \\ 2 & 1 & 1 \end{bmatrix}, \quad C = \begin{bmatrix} 2 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix}.$$

1.12. Let A and B be $n \times n$ matrices such that $I + AB$ is invertible. Show that $I + BA$ is invertible with inverse $I - B(I + AB)^{-1}A$.

1.13. A nilpotent matrix is a square matrix A such that $A^k = 0$ for some positive integer k. The smallest such k is called the index of A. Show that

$$I + A + A^2 + \cdots + A^k$$

is invertible if A is a nilpotent matrix of index k. What is the inverse?

2 Linear Spaces

2.1 Definition

Definition 2.1. Let V be a non-empty set, and let there be defined two operations $V \times V \to V$ and $\mathbf{R} \times V \to V$ called addition and scalar multiplication, respectively. We denote the value of the addition at $(\boldsymbol{u}, \boldsymbol{v})$ by $\boldsymbol{u} + \boldsymbol{v}$ and the value of the scalar multiplication at $(s, \boldsymbol{u})$ by $s\boldsymbol{u}$. We say that the set V, together with the addition and scalar multiplication, forms a linear space provided that the following conditions are met:

(i) $\boldsymbol{u} + \boldsymbol{v} = \boldsymbol{v} + \boldsymbol{u}$ for all $\boldsymbol{u}$ and $\boldsymbol{v}$ in V.
(ii) $(\boldsymbol{u} + \boldsymbol{v}) + \boldsymbol{w} = \boldsymbol{u} + (\boldsymbol{v} + \boldsymbol{w})$ for all $\boldsymbol{u}$, $\boldsymbol{v}$ and $\boldsymbol{w}$ in V.
(iii) There exists an element $\mathbf{0} \in V$ such that $\boldsymbol{u} + \mathbf{0} = \boldsymbol{u}$ for all $\boldsymbol{u} \in V$.
(iv) For some such element $\mathbf{0} \in V$, there exists, for every $\boldsymbol{u} \in V$, an element $-\boldsymbol{u} \in V$ such that $\boldsymbol{u} + (-\boldsymbol{u}) = \mathbf{0}$.
(v) $s(\boldsymbol{u} + \boldsymbol{v}) = s\boldsymbol{u} + s\boldsymbol{v}$ for all $s \in \mathbf{R}$ and all $\boldsymbol{u}$ and $\boldsymbol{v}$ in V.
(vi) $(s + t)\boldsymbol{u} = s\boldsymbol{u} + t\boldsymbol{u}$ for all s and t in $\mathbf{R}$ and all $\boldsymbol{u} \in V$.
(vii) $s(t\boldsymbol{u}) = (st)\boldsymbol{u}$ for all s and t in $\mathbf{R}$ and all $\boldsymbol{u} \in V$.
(viii) $1\boldsymbol{u} = \boldsymbol{u}$ for all $\boldsymbol{u} \in V$.

The eight conditions are called the axioms for a linear space, an element of V is called a vector and real numbers are usually called scalars. The following statements are immediate consequences of the definition.

Theorem 2.2. Let V be a linear space.

(i) There is only one vector $\mathbf{0} \in V$ satisfying Axiom (iii).
(ii) For each $\boldsymbol{u} \in V$, there is only one vector $-\boldsymbol{u} \in V$ satisfying Axiom (iv).
(iii) $0\boldsymbol{u} = \mathbf{0}$ for all $\boldsymbol{u} \in V$.
(iv) $s\mathbf{0} = \mathbf{0}$ for all $s \in \mathbf{R}$.
(v) $(-1)\boldsymbol{u} = -\boldsymbol{u}$ for all $\boldsymbol{u} \in V$.

Proof. (i) Suppose that $\mathbf{0}_1$ and $\mathbf{0}_2$ both satisfy Axiom (iii) in the definition. Then it follows from that axiom and Axiom (i) that $\mathbf{0}_1 = \mathbf{0}_1 + \mathbf{0}_2 = \mathbf{0}_2$.

(ii) Suppose that $\boldsymbol{u} + \boldsymbol{v} = \mathbf{0}$ and $\boldsymbol{u} + \boldsymbol{w} = \mathbf{0}$. Then it follows from Axioms (i), (ii) and (iii) that $\boldsymbol{v} = \boldsymbol{v} + \mathbf{0} = \boldsymbol{v} + (\boldsymbol{u} + \boldsymbol{w}) = (\boldsymbol{v} + \boldsymbol{u}) + \boldsymbol{w} = \mathbf{0} + \boldsymbol{w} = \boldsymbol{w} + \mathbf{0} = \boldsymbol{w}$.

(iii) $0\boldsymbol{u} = 0\boldsymbol{u} + \mathbf{0} = 0\boldsymbol{u} + 0\boldsymbol{u} + (-(0\boldsymbol{u})) = (0+0)\boldsymbol{u} + (-(0\boldsymbol{u})) = 0\boldsymbol{u} + (-(0\boldsymbol{u})) = \mathbf{0}$.

(iv) $s\mathbf{0} = s\mathbf{0} + \mathbf{0} = s\mathbf{0} + s\mathbf{0} + (-(s\mathbf{0})) = s(\mathbf{0} + \mathbf{0}) + (-(s\mathbf{0})) = s\mathbf{0} + (-(s\mathbf{0})) = \mathbf{0}$.

(v) $\boldsymbol{u} + (-1)\boldsymbol{u} = 1\boldsymbol{u} + (-1)\boldsymbol{u} = (1-1)\boldsymbol{u} = 0\boldsymbol{u} = \mathbf{0}$, and hence $(-1)\boldsymbol{u} = -\boldsymbol{u}$ by the uniqueness of the additive inverse. ■

We call $\mathbf{0}$ the zero vector, and $-\boldsymbol{u}$ is called the additive inverse of $\boldsymbol{u}$. The linear space is in fact a triple $(V, +, \cdot)$ where $+$ and $\cdot$ denote the addition and scalar multiplication, respectively. When it is clear from context which operations $+$ and $\cdot$ are intended, we shall, by abuse of language, use V to denote also the linear space.

Example 2.3. The set $\mathbf{R}^n$ of n-tuples $(x_1, x_2, \ldots, x_n)$ of real numbers, together with the addition and scalar multiplication defined by

$$\begin{aligned}(x_1, x_2, \ldots, x_n) + (y_1, y_2, \ldots, y_n) &= (x_1 + y_1, x_2 + y_2, \ldots, x_n + y_n),\\ s(x_1, x_2, \ldots, x_n) &= (sx_1, sx_2, \ldots, sx_n),\end{aligned}$$

is a linear space. The zero vector is $\mathbf{0} = (0, 0, \ldots, 0)$ and the additive inverse of the vector $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)$ is $-\boldsymbol{x} = (-x_1, -x_2, \ldots, -x_n)$. We leave it to the reader to verify the axioms for a linear space.

Example 2.4. The set of vectors in space, together with the ordinary vector addition and scalar multiplication, forms a linear space.

Example 2.5. The set P of polynomials over $\mathbf{R}$, together with addition of polynomials and multiplication by a number, forms a linear space. The zero vector is the zero polynomial.

Example 2.6. The set P_n of polynomials over $\mathbf{R}$ of degree at most n, together with addition of polynomials and multiplication by a number, forms a linear space.

Example 2.7. The set of polynomials of degree exactly n, together with addition of polynomials and multiplication by a number, does not form a linear space. One reason for this is that addition is not a function into the set. The sum of two polynomials of degree n need not be a polynomial of degree n. For example, the sum of the two polynomials $1 + x + x^2$ and $1 + x - x^2$ of degree 2 is the polynomial $2 + 2x$ of degree 1. Another reason is that the zero polynomial does not belong to the set.

Example 2.8. Let $C(I)$ be the set of continuous functions $\boldsymbol{u} : I \to \mathbf{R}$ where I is an interval. The sum of two functions $\boldsymbol{u}$ and $\boldsymbol{v}$ in $C(I)$ is defined by $(\boldsymbol{u}+\boldsymbol{v})(x) = \boldsymbol{u}(x)+\boldsymbol{v}(x)$, $x \in I$, and the scalar multiplication is defined by $(s\boldsymbol{u})(x) = s\boldsymbol{u}(x)$, $x \in I$. In this way we get a linear space. The zero vector is the function $\mathbf{0}$ defined by $\mathbf{0}(x) = 0$, $x \in I$, and $-\boldsymbol{u}$ is defined by $(-\boldsymbol{u})(x) = -\boldsymbol{u}(x)$, $x \in I$. Here it is essential that the sum of two continuous functions is continuous and that $s\boldsymbol{u}$ is continuous when $\boldsymbol{u}$ is.

Example 2.9. The set $M_{m\times n}$ of matrices of size $m \times n$ forms a linear space together with addition and scalar multiplication of matrices.

Definition 2.10. Let U and V be linear spaces. We say that U is a subspace of V if $U \subseteq V$ and the addition and scalar multiplication of U agree with those of V on the sets $U \times U$ and $\mathbf{R} \times U$, respectively.

Theorem 2.11. Let U be a subset of a linear space V. Then the addition and scalar multiplication of V restricted to $U \times U$ and $\mathbf{R} \times U$, respectively, make U a subspace of V if and only if

(i) $U \neq \varnothing$,
(ii) $\boldsymbol{u} + \boldsymbol{v} \in U$ whenever $\boldsymbol{u}$ and $\boldsymbol{v}$ belong to U,
(iii) $s\boldsymbol{u} \in U$ whenever $s \in \mathbf{R}$ and $\boldsymbol{u} \in U$.

Proof. First assume that U is a subspace of V. Then U is non-empty by definition. Hence (i) is satisfied. Since the restrictions of the operations $+$ and $\cdot$ to $U \times U$ and $\mathbf{R} \times U$ are functions into U, conditions (ii) and (iii) are satisfied.

To show the converse, we assume that the three conditions are met. From (ii) and (iii) it follows that the restrictions of the operations to $U \times U$ and $\mathbf{R} \times U$ are functions into U. Since vectors in U are also vectors in V, all the axioms for linear spaces hold for U except possibly Axioms (iii) and (iv). Since $U \neq \varnothing$, U contains at least one vector $\boldsymbol{u}$. By condition (iii), $\mathbf{0} = 0\boldsymbol{u} \in U$. Hence Axiom (iii) holds. It also follows from condition (iii) that $-\boldsymbol{u} = (-1)\boldsymbol{u} \in U$ if $\boldsymbol{u}$ is any vector in U, whence Axiom (iv) is satisfied. ■

Corollary 2.12. Let U be a non-empty subset of a linear space V. Then U is a subspace of V if and only if $s\boldsymbol{u} + t\boldsymbol{v} \in U$ for all $\boldsymbol{u}$ and $\boldsymbol{v}$ in U and all real numbers s and t.

Proof. Assume that U is a subspace of V. If $\boldsymbol{u}$ and $\boldsymbol{v}$ belong to U and s and t are real numbers, then, by (iii) of Theorem 2.11, $s\boldsymbol{u} \in U$ and $t\boldsymbol{v} \in U$. Hence, by (ii) of Theorem 2.11, $s\boldsymbol{u} + t\boldsymbol{v} \in U$.

Assume that $s\boldsymbol{u} + t\boldsymbol{v} \in U$ for all $\boldsymbol{u}$ and $\boldsymbol{v}$ in U and all real numbers s and t. If $\boldsymbol{u}$ and $\boldsymbol{v}$ belong to U, then it follows that $\boldsymbol{u} + \boldsymbol{v} = 1\boldsymbol{u} + 1\boldsymbol{v} \in U$. Assume that $\boldsymbol{u} \in U$ and $s \in \mathbf{R}$. Then $s\boldsymbol{u} = s\boldsymbol{u} + 0\boldsymbol{u} \in U$. Hence, the conditions of Theorem 2.11 are satisfied, which shows that U is a subspace. ■

Note that it follows from the proof of Theorem 2.11 that the zero vector of a linear space V is also the zero vector of its subspaces and that the inverse of a vector in a subspace of V is the inverse of that vector in V.

Example 2.13. Let V be any linear space. If $\mathbf{0}$ is the zero vector of V, then $U = \{\mathbf{0}\}$ is a subset of V. In fact, U is a subspace by Corollary 2.12. Firstly, U is non-empty. Secondly, if $\boldsymbol{u}$ and $\boldsymbol{v}$ belong to U, then both are the zero vector. Hence, $s\boldsymbol{u} + t\boldsymbol{v} = s\mathbf{0} + t\mathbf{0} = \mathbf{0} \in U$ for all real numbers s and t. We call U the zero subspace of V.

Example 2.14. Let $U = \{\boldsymbol{x} \in \mathbf{R}^n;\ a_1x_1 + a_2x_2 + \cdots + a_nx_n = 0\}$. Then U is a subspace of $\mathbf{R}^n$. To show this, we first observe that $\mathbf{0} = (0, 0, \ldots, 0) \in U$. Hence U is non-empty. Let $\boldsymbol{x}$ and $\boldsymbol{y}$ belong to U and s and t be real numbers. Then

$$
\begin{aligned}
0 &= a_1x_1 + a_2x_2 + \cdots + a_nx_n, \\
0 &= a_1y_1 + a_2y_2 + \cdots + a_ny_n.
\end{aligned}
$$

Hence

$$\begin{aligned}0 = s0 + t0 &= s(a_1x_1 + a_2x_2 + \cdots + a_nx_n) + t(a_1y_1 + a_2y_2 + \cdots + a_ny_n)\\ &= a_1(sx_1 + ty_1) + a_2(sx_2 + ty_2) + \cdots + a_n(sx_n + ty_n),\end{aligned}$$

and it follows that $s\boldsymbol{x} + t\boldsymbol{y} = (sx_1 + ty_1, sx_2 + ty_2, \dots, sx_n + ty_n) \in U$.

Example 2.15. The linear space P_n of polynomials of degree at most n is a subspace of the linear space P of all polynomials. If $m \le n$, then P_m is a subspace of P_n.

Example 2.16. If we regard polynomials over $\mathbf{R}$ as functions on $\mathbf{R}$, then we can regard P as a subspace of the space $C(\mathbf{R})$ of continuous functions on $\mathbf{R}$.

By abuse of notation, we shall frequently regard elements of $\mathbf{R}^n$ as row or column matrices and vice versa. For example, when A is an $m \times n$ matrix, $\boldsymbol{x} = (x_1, \dots, x_n) \in \mathbf{R}^n$ and $\boldsymbol{y} = (y_1, \dots, y_m) \in \mathbf{R}^m$, $A\boldsymbol{x} = \boldsymbol{y}$ means that

$$A \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} y_1 \\ \vdots \\ y_m \end{bmatrix}.$$

Definition 2.17. Let A be an $m \times n$ matrix. The kernel $\ker A$ of A is the set of solutions in $\mathbf{R}^n$ of the equation $A\boldsymbol{x} = \mathbf{0}$. The image $\operatorname{im} A$ of A is the set of vectors $\boldsymbol{y} \in \mathbf{R}^m$ for which the equation $A\boldsymbol{x} = \boldsymbol{y}$ has a solution.

Theorem 2.18. Let A be an $m \times n$ matrix. Then $\ker A$ and $\operatorname{im} A$ are subspaces of $\mathbf{R}^n$ and $\mathbf{R}^m$, respectively.

Proof. Obviously, $\mathbf{0} = (0, 0, \dots, 0) \in \ker A$. Hence $\ker A \ne \varnothing$. Suppose that $\boldsymbol{x}$ and $\boldsymbol{y}$ belong to $\ker A$. Then $A\boldsymbol{x} = A\boldsymbol{y} = \mathbf{0}$, and so $A(s\boldsymbol{x} + t\boldsymbol{y}) = sA\boldsymbol{x} + tA\boldsymbol{y} = s\mathbf{0} + t\mathbf{0} = \mathbf{0}$. Hence, $s\boldsymbol{x} + t\boldsymbol{y}$ belongs to $\ker A$. This shows that $\ker A$ is a subspace of $\mathbf{R}^n$.

The vector $\mathbf{0} = (0, 0, \dots, 0) \in \mathbf{R}^m$ belongs to $\operatorname{im} A$ since $\boldsymbol{x} = \mathbf{0} = (0, 0, \dots, 0) \in \mathbf{R}^n$ is a solution of the equation $A\boldsymbol{x} = \mathbf{0}$. Hence $\operatorname{im} A \ne \varnothing$. Suppose that $\boldsymbol{u}$ and $\boldsymbol{v}$ belong to $\operatorname{im} A$. Then there exist $\boldsymbol{x}$ and $\boldsymbol{y}$ in $\mathbf{R}^n$ such that $A\boldsymbol{x} = \boldsymbol{u}$ and $A\boldsymbol{y} = \boldsymbol{v}$. From this we get $A(s\boldsymbol{x} + t\boldsymbol{y}) = sA\boldsymbol{x} + tA\boldsymbol{y} = s\boldsymbol{u} + t\boldsymbol{v}$, and so $s\boldsymbol{u} + t\boldsymbol{v} \in \operatorname{im} A$. ■

Example 2.19. The plane $a_1x_1 + a_2x_2 + a_3x_3 = 0$ through the origin can be regarded as the kernel of the matrix $\begin{bmatrix} a_1 & a_2 & a_3 \end{bmatrix}$. This plane also has a parametric equation

$$\begin{cases} x_1 = u_1t_1 + v_1t_2 \\ x_2 = u_2t_1 + v_2t_2 \\ x_3 = u_3t_1 + v_3t_2 \end{cases} \Leftrightarrow \begin{bmatrix} u_1 & v_1 \\ u_2 & v_2 \\ u_3 & v_3 \end{bmatrix} \begin{bmatrix} t_1 \\ t_2 \end{bmatrix} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}.$$

Hence, the plane is the image of the matrix

$$\begin{bmatrix} u_1 & v_1 \\ u_2 & v_2 \\ u_3 & v_3 \end{bmatrix}.$$

The intersection of the planes $a_1x_1 + a_2x_2 + a_3x_3 = 0$ and $b_1x_1 + b_2x_2 + b_3x_3 = 0$ through the origin is the kernel of the matrix

$$\begin{bmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \end{bmatrix}.$$

If the two planes are not identical, their intersection is a line through the origin. Since lines have parametric equations, this line can also be regarded as the image of a matrix.

2.2 Bases

Definition 2.20. Let $\boldsymbol{u}$ and $\boldsymbol{u}_1, \boldsymbol{u}_2, \dots, \boldsymbol{u}_k$ be vectors in a linear space V. We say that $\boldsymbol{u}$ is a linear combination of the $\boldsymbol{u}_i$ if there exist real numbers $s_1, s_2, \dots, s_k$ such that

$$\boldsymbol{u} = s_1\boldsymbol{u}_1 + s_2\boldsymbol{u}_2 + \cdots + s_k\boldsymbol{u}_k.$$

Definition 2.21. We say that the vectors $\boldsymbol{u}_1, \boldsymbol{u}_2, \dots, \boldsymbol{u}_k$ in a linear space V span V if every vector of V is a linear combination of them. One can also say that they generate V and call them generators of V.

Definition 2.22. Let $\boldsymbol{u}_1, \boldsymbol{u}_2, \dots, \boldsymbol{u}_k$ be vectors in a linear space V. We call the set of all linear combinations of those vectors the span of them and denote it by $[\boldsymbol{u}_1, \boldsymbol{u}_2, \dots, \boldsymbol{u}_k]$.

Theorem 2.23. Let $\boldsymbol{u}_1, \boldsymbol{u}_2, \dots, \boldsymbol{u}_k$ be vectors in a linear space V. Then their span U is a subspace of V and U is spanned by $\boldsymbol{u}_1, \boldsymbol{u}_2, \dots, \boldsymbol{u}_k$.

Proof. $U \neq \varnothing$ since $\boldsymbol{0} = 0\boldsymbol{u}_1 + 0\boldsymbol{u}_2 + \cdots + 0\boldsymbol{u}_k \in U$. If $\boldsymbol{u}$ and $\boldsymbol{v}$ belong to U, then $\boldsymbol{u} = s_1\boldsymbol{u}_1 + s_2\boldsymbol{u}_2 + \cdots + s_k\boldsymbol{u}_k$ and $\boldsymbol{v} = t_1\boldsymbol{u}_1 + t_2\boldsymbol{u}_2 + \cdots + t_k\boldsymbol{u}_k$. Therefore,

$$s\boldsymbol{u} + t\boldsymbol{v} = (ss_1 + tt_1)\boldsymbol{u}_1 + (ss_2 + tt_2)\boldsymbol{u}_2 + \cdots + (ss_k + tt_k)\boldsymbol{u}_k \in U,$$

and hence, by Corollary 2.12, U is a subspace of V. The fact that the vectors span U follows directly from the definition. ■

Example 2.24. The plane in Example 2.19 is the span of the two vectors $\boldsymbol{u} = (u_1, u_2, u_3)$ and $\boldsymbol{v} = (v_1, v_2, v_3)$. The line there is the span of a single vector $\boldsymbol{w} = (w_1, w_2, w_3)$.

Example 2.25. Let A be an $m \times n$ matrix. A vector $\boldsymbol{y} \in \mathbf{R}^m$ belongs to $\operatorname{im} A$ if and only if there exists a vector $\boldsymbol{x} \in \mathbf{R}^n$ such that $A\boldsymbol{x} = \boldsymbol{y}$. Since this can be written as $x_1A_1 + x_2A_2 + \cdots + x_nA_n = \boldsymbol{y}$, we see that $\boldsymbol{y} \in \operatorname{im} A$ if and only if $\boldsymbol{y}$ is a linear combination of the columns of A. Hence, $\operatorname{im} A = [A_1, A_2, \dots, A_n]$ is spanned by the columns of the matrix.

Definition 2.26. We say that the vectors $\boldsymbol{u}_1, \boldsymbol{u}_2, \dots, \boldsymbol{u}_k$ in a linear space V are linearly dependent if there exist real numbers $s_1, s_2, \dots, s_k$, not all zero, such that

$$s_1\boldsymbol{u}_1 + s_2\boldsymbol{u}_2 + \cdots + s_k\boldsymbol{u}_k = \boldsymbol{0}.$$

If the vectors are not linearly dependent, we say that they are linearly independent.

Hence, the vectors $\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_k$ are linearly independent if and only if

$$s_1\boldsymbol{u}_1 + s_2\boldsymbol{u}_2 + \cdots + s_k\boldsymbol{u}_k = \mathbf{0} \quad \Rightarrow \quad s_1 = s_2 = \cdots = s_k = 0.$$

Also note that a single vector $\boldsymbol{u}$ is linearly dependent if and only if it is the zero vector.

Example 2.27. Here we want to find out if the vectors $\boldsymbol{u}_1 = (1,1,1)$, $\boldsymbol{u}_2 = (1,3,1)$, $\boldsymbol{u}_3 = (1,4,3)$ in $\mathbf{R}^3$ are linearly dependent. We solve the equation $s_1\boldsymbol{u}_1+s_2\boldsymbol{u}_2+s_3\boldsymbol{u}_3 = \mathbf{0}$:

$$\begin{cases} s_1 + s_2 + s_3 = 0 \\ s_1 + 3s_2 + 4s_3 = 0 \\ s_1 + s_2 + 3s_3 = 0 \end{cases} \Leftrightarrow \begin{cases} s_1 + s_2 + s_3 = 0 \\ \qquad 2s_2 + 3s_3 = 0 \\ \qquad\qquad 2s_3 = 0 \end{cases} \Leftrightarrow \quad s_1 = s_2 = s_3 = 0.$$

Hence, the vectors are linearly independent.

Example 2.28. Consider the vectors $\boldsymbol{u}_1 = (1,1,1)$, $\boldsymbol{u}_2 = (1,3,5)$, $\boldsymbol{u}_3 = (1,4,7)$ in $\mathbf{R}^3$. This time the equation $s_1\boldsymbol{u}_1 + s_2\boldsymbol{u}_2 + s_3\boldsymbol{u}_3 = \mathbf{0}$ is equivalent to

$$\begin{cases} s_1 + s_2 + s_3 = 0 \\ s_1 + 3s_2 + 4s_3 = 0 \\ s_1 + 5s_2 + 7s_3 = 0 \end{cases} \Leftrightarrow \begin{cases} s_1 + s_2 + s_3 = 0 \\ \qquad 2s_2 + 3s_3 = 0 \\ \qquad 4s_2 + 6s_3 = 0 \end{cases} \Leftrightarrow \begin{cases} s_1 + s_2 + s_3 = 0 \\ \qquad 2s_2 + 3s_3 = 0 \end{cases}.$$

Since this equation has non-trivial solutions, the vectors $\boldsymbol{u}_1, \boldsymbol{u}_2, \boldsymbol{u}_3$ are linearly dependent.

Theorem 2.29. Let $k \geq 2$. Then the vectors $\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_k$ in a linear space V are linearly dependent if and only if one of them is a linear combination of the others.

Proof. Suppose that the vectors are linearly dependent. Then there exist real numbers $s_1, \ldots, s_i, \ldots, s_k$, where $s_i \neq 0$, such that $s_1\boldsymbol{u}_1 + \cdots + s_i\boldsymbol{u}_i + \cdots + s_k\boldsymbol{u}_k = \mathbf{0}$. Dividing by s_i and moving terms, we find that

$$\boldsymbol{u}_i = \sum_{\substack{1\leq j\leq k \\ j\neq i}} \frac{-s_j}{s_i}\boldsymbol{u}_j$$

is a linear combination of the other vectors.

Now suppose that

$$\boldsymbol{u}_i = \sum_{\substack{1\leq j\leq k \\ j\neq i}} s_j\boldsymbol{u}_j$$

is a linear combination of the other vectors. Then $s_1\boldsymbol{u}_1 + \cdots + s_k\boldsymbol{u}_k = 0$ where the coefficient $s_i = -1$ of $\boldsymbol{u}_i$ is non-zero. Hence the vectors are linearly dependent. ■

Definition 2.30. Let $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_k$ be vectors in a linear space V. We say that these vectors form a basis for V if they span V and are linearly independent.

Example 2.31. The vectors

$$\begin{aligned} \boldsymbol{\varepsilon}_1 &= (1,0,0,\ldots,0) \\ \boldsymbol{\varepsilon}_2 &= (0,1,0,\ldots,0) \\ \boldsymbol{\varepsilon}_3 &= (0,0,1,\ldots,0) \\ &\vdots \\ \boldsymbol{\varepsilon}_n &= (0,0,0,\ldots,1) \end{aligned}$$

of $\mathbf{R}^n$ form a basis for $\mathbf{R}^n$. Firstly, they span $\mathbf{R}^n$ since any vector

$$\boldsymbol{x} = (x_1, x_2, x_3, \ldots, x_n) = x_1\boldsymbol{\varepsilon}_1 + x_2\boldsymbol{\varepsilon}_2 + x_3\boldsymbol{\varepsilon}_3 + \cdots + x_n\boldsymbol{\varepsilon}_n$$

of $\mathbf{R}^n$ is a linear combination of them. Secondly, they are linearly independent since

$$x_1\boldsymbol{\varepsilon}_1 + x_2\boldsymbol{\varepsilon}_2 + x_3\boldsymbol{\varepsilon}_3 + \cdots + x_n\boldsymbol{\varepsilon}_n = (x_1, x_2, x_3, \ldots, x_n) = \mathbf{0}$$

only if $x_1 = x_2 = x_3 = \cdots = x_n = 0$.

Definition 2.32. The basis $\boldsymbol{\varepsilon}_1, \ldots, \boldsymbol{\varepsilon}_n$ in Example 2.31 is called the standard basis for the linear space $\mathbf{R}^n$.

Theorem 2.33. Let $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_k$ be vectors in a linear space V. Then they form a basis for V if and only if every vector $\boldsymbol{u} \in V$ can be written as a linear combination

$$\boldsymbol{u} = x_1\boldsymbol{u}_1 + \cdots + x_k\boldsymbol{u}_k$$

with unique coefficients $x_1, \ldots, x_k$.

Proof. First assume that the vectors form a basis for V and let $\boldsymbol{u} \in V$. Then, by the definition of bases, $\boldsymbol{u} = x_1\boldsymbol{u}_1 + \cdots + x_k\boldsymbol{u}_k$ is a linear combination of the vectors. It remains to show that the coefficients are uniquely determined. Assume, to that end, that we also have $\boldsymbol{u} = y_1\boldsymbol{u}_1 + \cdots + y_k\boldsymbol{u}_k$. Then $\mathbf{0} = \boldsymbol{u} - \boldsymbol{u} = (x_1 - y_1)\boldsymbol{u}_1 + \cdots + (x_k - y_k)\boldsymbol{u}_k$, and it follows from the linear independence of $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_k$ that $x_1 - y_1 = \cdots = x_k - y_k = 0$, whence $x_i = y_i$ for $i = 1, \ldots, k$.

To show the converse, we assume that every vector of V has a unique representation as in the theorem. Then, certainly, every vector of V is a linear combination of $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_k$. If $x_1\boldsymbol{u}_1 + \cdots + x_k\boldsymbol{u}_k = \mathbf{0}$, then the uniqueness and the fact that $0\boldsymbol{u}_1 + \cdots + 0\boldsymbol{u}_k = \mathbf{0}$ imply that $x_1 = \cdots = x_k = 0$. Hence $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_k$ are also linearly independent, and therefore they form a basis for V. ■

Definition 2.34. Let $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_k$ be a basis for a linear space V and let $\boldsymbol{u}$ be a vector of V. If $\boldsymbol{u} = x_1\boldsymbol{u}_1 + \cdots + x_k\boldsymbol{u}_k$, we call $(x_1, \ldots, x_k)$ the coordinates of $\boldsymbol{u}$ with respect to the basis $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_k$.

Example 2.35. The coordinates of $\boldsymbol{x} = (x_1, \ldots, x_n) \in \mathbf{R}^n$ with respect to the standard basis in Example 2.31 are $(x_1, \ldots, x_n)$. Hence, with respect to that specific basis, the coordinates of a vector in $\mathbf{R}^n$ are its components.

Example 2.36. The polynomials $1, x, x^2, \ldots, x^n$ form a basis for the linear space P_n of polynomials of degree at most n. This is so because every polynomial in P_n can be written as $\boldsymbol{p} = a_0 + a_1x + a_2x^2 + \cdots + a_nx^n$ with unique coefficients $a_0, a_1, a_2, \ldots, a_n$. The coordinates of $\boldsymbol{p}$ with respect to this basis are $(a_0, a_1, a_2, \ldots, a_n)$.

Example 2.37. The linear space P of all polynomials has no basis. No finite collection $\boldsymbol{p}_1, \ldots, \boldsymbol{p}_k$ of polynomials span P since a polynomial of degree greater than the maximum degree of the $\boldsymbol{p}_i$ cannot be a linear combination of them.

Example 2.38. We set out to find bases for $\ker A$ and $\operatorname{im} A$ where

$$A = \begin{bmatrix} 1 & 2 & 3 & 4 & 5 \\ 1 & 3 & 4 & 5 & 6 \\ 2 & 5 & 7 & 9 & 11 \end{bmatrix}.$$

We begin by solving the equation $A\boldsymbol{x} = \boldsymbol{0}$:

$$\begin{cases} x_1 + 2x_2 + 3x_3 + 4x_4 + 5x_5 = 0 \\ x_1 + 3x_2 + 4x_3 + 5x_4 + 6x_5 = 0 \\ 2x_1 + 5x_2 + 7x_3 + 9x_4 + 11x_5 = 0 \end{cases} \quad\Leftrightarrow\quad \begin{cases} x_1 + 2x_2 + 3x_3 + 4x_4 + 5x_5 = 0 \\ x_2 + x_3 + x_4 + x_5 = 0 \\ x_2 + x_3 + x_4 + x_5 = 0 \end{cases}$$

$$\Leftrightarrow \begin{cases} x_1 = -r - 2s - 3t \\ x_2 = -r - s - t \\ x_3 = r \\ x_4 = s \\ x_5 = t \end{cases} \quad\Leftrightarrow\quad \boldsymbol{x} = r\boldsymbol{u} + s\boldsymbol{v} + t\boldsymbol{w}$$

where $\boldsymbol{u} = (-1, -1, 1, 0, 0)$, $\boldsymbol{v} = (-2, -1, 0, 1, 0)$ and $\boldsymbol{w} = (-3, -1, 0, 0, 1)$. This shows that $\boldsymbol{x} \in \ker A$ if and only if $\boldsymbol{x} = r\boldsymbol{u} + s\boldsymbol{v} + t\boldsymbol{w}$. Hence the vectors $\boldsymbol{u}$, $\boldsymbol{v}$ and $\boldsymbol{w}$ span $\ker A$. The generators obtained when solving a system in the usual way will always be linearly independent. The free variables x_3, x_4 and x_5 in the last system correspond to the parameters r, s and t in the solution, which in turn correspond to the patterns $(1, 0, 0)$, $(0, 1, 0)$ and $(0, 0, 1)$ in the third, fourth and fifth positions of the generators. Hence $r\boldsymbol{u} + s\boldsymbol{v} + t\boldsymbol{w} = \boldsymbol{0}$ if and only if $(*, *, r, s, t) = \boldsymbol{0}$ which implies that $r = s = t = 0$. Thus $(-1, -1, 1, 0, 0)$, $(-2, -1, 0, 1, 0)$ and $(-3, -1, 0, 0, 1)$ form a basis for $\ker A$.

We can use the same computations to find a basis for $\operatorname{im} A$. Since $\operatorname{im} A$ is spanned by the columns A_1, A_2, A_3, A_4, A_5 of A, every vector $\boldsymbol{y} \in \operatorname{im} A$ can be written as

$$\boldsymbol{y} = x_1A_1 + x_2A_2 + x_3A_3 + x_4A_4 + x_5A_5. \tag{2.1}$$

From the solution of the system we have

$$(-r - 2s - 3t)A_1 + (-r - s - t)A_2 + rA_3 + sA_4 + tA_5 = 0.$$

By setting $r = 1$, $s = 0$, $t = 0$, we see that $A_3 = r_1A_1 + r_2A_2$ is a linear combination of A_1, A_2. By setting $r = 0$, $s = 1$, $t = 0$, we see that also $A_4 = s_1A_1 + s_2A_2$ is a linear combination of A_1, A_2. Finally, by setting $r = 0$, $s = 0$, $t = 1$, we see that $A_5 = t_1A_1 + t_2A_2$ is a linear combination of A_1, A_2. Substituting these expressions for A_3, A_4 and A_5 into (2.1) and collecting terms, we find that $\boldsymbol{y}$ is a linear combination of A_1, A_2, which therefore span $\operatorname{im} A$. The computations also reveal that these vectors are linearly independent. In fact,

$$x_1A_1 + x_2A_2 = 0 \quad\Leftrightarrow\quad x_1A_1 + x_2A_2 + 0A_3 + 0A_4 + 0A_5 = 0$$

$$\Leftrightarrow \begin{cases} x_1 = -r - 2s - 3t \\ x_2 = -r - s - t \\ 0 = r \\ 0 = s \\ 0 = t \end{cases}$$

from which it follows that $x_1 = x_2 = 0$. Hence, A_1, A_2 form a basis for $\operatorname{im} A$.

Lemma 2.39. Let $\boldsymbol{u}_1, \dots, \boldsymbol{u}_j, \boldsymbol{u}_{j+1}, \dots, \boldsymbol{u}_k$ be vectors and assume that $1 \le j < k$ and that $\boldsymbol{u}_1, \dots, \boldsymbol{u}_j$ are linearly dependent. Then the vectors $\boldsymbol{u}_1, \dots, \boldsymbol{u}_j, \boldsymbol{u}_{j+1}, \dots, \boldsymbol{u}_k$ are also linearly dependent.

Proof. By the assumption, there exist real numbers $s_1, \dots, s_j$, not all zero, such that $s_1\boldsymbol{u}_1 + \cdots + s_j\boldsymbol{u}_j = 0$. But then $s_1\boldsymbol{u}_1 + \cdots + s_j\boldsymbol{u}_j + 0\boldsymbol{u}_{j+1} + \cdots + 0\boldsymbol{u}_k = \mathbf{0}$, and at least one of these coefficients is non-zero. ■

Theorem 2.40. If the vectors $\boldsymbol{v}_1, \dots, \boldsymbol{v}_k$ belong to the span $[\boldsymbol{u}_1, \dots, \boldsymbol{u}_j]$ and $1 \le j < k$, then $\boldsymbol{v}_1, \dots, \boldsymbol{v}_k$ are linearly dependent.

Proof. By Lemma 2.39, it is sufficient to prove the statement for $k = j + 1$. We use induction on j. If $\boldsymbol{v}_1$ and $\boldsymbol{v}_2$ belong to $[\boldsymbol{u}_1]$, then $\boldsymbol{v}_1 = s_1\boldsymbol{u}_1$ and $\boldsymbol{v}_2 = s_2\boldsymbol{u}_1$ for some real numbers s_1 and s_2. If $s_2 = 0$, then $\boldsymbol{v}_2 = 0$ and hence $0\boldsymbol{v}_1 + 1\boldsymbol{v}_2 = 0$, which shows that $\boldsymbol{v}_1, \boldsymbol{v}_2$ are linearly dependent in this case. If $s_2 \neq 0$, the linear dependence follows from the equality $s_2\boldsymbol{v}_1 - s_1\boldsymbol{v}_2 = 0$. This shows the statement for $j = 1$.

Let $p \ge 2$ and suppose that the statement holds for $j = p - 1$ and that $\boldsymbol{v}_1, \dots, \boldsymbol{v}_{p+1}$ belong to the span $[\boldsymbol{u}_1, \dots, \boldsymbol{u}_p]$. Then

$$\begin{cases} a_{11}\boldsymbol{u}_1 + \cdots + a_{1(p-1)}\boldsymbol{u}_{p-1} + a_{1p}\boldsymbol{u}_p = \boldsymbol{v}_1 \\ \qquad \vdots \\ a_{p1}\boldsymbol{u}_1 + \cdots + a_{p(p-1)}\boldsymbol{u}_{p-1} + a_{pp}\boldsymbol{u}_p = \boldsymbol{v}_p \\ a_{(p+1)1}\boldsymbol{u}_1 + \cdots + a_{(p+1)(p-1)}\boldsymbol{u}_{p-1} + a_{(p+1)p}\boldsymbol{u}_p = \boldsymbol{v}_{p+1} \end{cases}.$$

If $a_{1p} = \cdots = a_{pp} = a_{(p+1)p} = 0$, then $\boldsymbol{v}_1, \dots, \boldsymbol{v}_{p+1}$ belong to the span $[\boldsymbol{u}_1, \dots, \boldsymbol{u}_{p-1}]$ and are therefore linearly dependent by hypothesis and Lemma 2.39. Hence, renaming the vectors if needed, we may assume that $a_{(p+1)p} \neq 0$. We can then eliminate $\boldsymbol{u}_p$ from all equations except the last one by adding suitable multiples of that equation to them. Doing so, we obtain

$$\begin{cases} b_{11}\boldsymbol{u}_1 + \cdots + b_{1(p-1)}\boldsymbol{u}_{p-1} = \boldsymbol{v}_1 + c_1\boldsymbol{v}_{p+1} \\ \qquad \vdots \\ b_{p1}\boldsymbol{u}_1 + \cdots + b_{p(p-1)}\boldsymbol{u}_{p-1} = \boldsymbol{v}_p + c_p\boldsymbol{v}_{p+1} \end{cases}.$$

Hence, the vectors $\boldsymbol{v}_1 + c_1\boldsymbol{v}_{p+1}, \dots, \boldsymbol{v}_p + c_p\boldsymbol{v}_{p+1}$ belong to $[\boldsymbol{u}_1, \dots, \boldsymbol{u}_{p-1}]$ and are therefore linearly dependent by the induction hypothesis. Thus there exist scalars $s_1, \dots, s_p$, not all zero, such that

$$s_1(\boldsymbol{v}_1 + c_1\boldsymbol{v}_{p+1}) + \cdots + s_p(\boldsymbol{v}_p + c_p\boldsymbol{v}_{p+1}) = \mathbf{0},$$

and hence

$$s_1\boldsymbol{v}_1 + \cdots + s_p\boldsymbol{v}_p + (s_1c_1 + \cdots + s_pc_p)\boldsymbol{v}_{p+1} = \mathbf{0}.$$

This shows that $\boldsymbol{v}_1, \dots, \boldsymbol{v}_{p+1}$ are linearly dependent. ■

Theorem 2.41. Let $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_m$ and $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n$ be bases for the same linear space V. Then $m = n$.

Proof. By assumption, $V = [\boldsymbol{u}_1, \ldots, \boldsymbol{u}_m]$ and $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n$ are linearly independent. Hence $n \le m$ by Theorem 2.40. We also have $V = [\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n]$, and $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_m$ are linearly independent. Hence, by the same theorem, $m \le n$. The two inequalities together yield $m = n$. ■

Definition 2.42. Let V be a linear space. If $V = \{\mathbf{0}\}$, we say that the dimension of V is zero. If V has a basis consisting of $n \ge 1$ vectors, we say that the dimension of V is n. In these two cases we say that V is finite-dimensional and denote the dimension of V by $\dim V$. In the remaining case where $V \ne \{\mathbf{0}\}$ and has no basis, we say that V is infinite-dimensional.

Example 2.43. We saw in Example 2.31 that $\dim \mathbf{R}^n = n$. From Example 2.36 we have $\dim P_n = n + 1$, and by Example 2.37, P is infinite-dimensional. The dimensions of the kernel and image in Example 2.38 are 3 and 2, respectively.

Lemma 2.44. Let $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_k, \boldsymbol{u}_{k+1}$ be vectors in a linear space V. If $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_k$ are linearly independent and $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_k, \boldsymbol{u}_{k+1}$ are linearly dependent, then $\boldsymbol{u}_{k+1}$ is a linear combination of $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_k$.

Proof. By assumption, there exist real numbers $s_1, \ldots, s_k, s_{k+1}$, not all zero, such that $s_1\boldsymbol{u}_1 + \cdots + s_k\boldsymbol{u}_k + s_{k+1}\boldsymbol{u}_{k+1} = \mathbf{0}$. If $s_{k+1} = 0$, then $s_1\boldsymbol{u}_1 + \cdots + s_k\boldsymbol{u}_k = \mathbf{0}$ where at least one of the coefficients is non-zero. Since this contradicts the linear independence of $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_k$, we must have $s_{k+1} \ne 0$. Hence, dividing by s_{k+1} and moving terms, we can express $\boldsymbol{u}_{k+1}$ as a linear combination of $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_k$. ■

Theorem 2.45. Let $U \ne \{\mathbf{0}\}$ be a subspace of a finite-dimensional linear space V. If $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_k$ are linearly independent vectors of U, then there exists a basis for U containing those vectors.

Proof. Set $n = \dim V$ and consider any finite set S of linearly independent vectors of U containing the vectors $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_k$. Then, by Theorem 2.40, S cannot contain more than n vectors, for the vectors in S are also vectors of V and V is spanned by n vectors. Therefore, among all such sets S, there is a set $S_0 = \{\boldsymbol{u}_1, \ldots, \boldsymbol{u}_m\}$ with a maximum number of vectors. If $\boldsymbol{u}$ is any other vector of U, it follows from the maximality of S_0 that $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_m, \boldsymbol{u}$ must be linearly dependent. Hence, by Lemma 2.44, $\boldsymbol{u}$ is a linear combination of the vectors in S_0. This shows that $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_m$ form a basis for U. ■

Corollary 2.46. If U is a subspace of a finite-dimensional linear space V, then U is finite-dimensional and $\dim U \le \dim V$.

Proof. The claim is trivial when $U = \{\mathbf{0}\}$. Otherwise, U contains a non-zero vector $\boldsymbol{u}_1$, which by Theorem 2.45 can be extended to a basis for U. The inequality now follows from Theorem 2.40. ■

In particular, any subspace of $\mathbf{R}^n$ is finite-dimensional. The dimension of a subspace of $\mathbf{R}^3$ must be 0, 1, 2 or 3. Hence, the only subspaces of $\mathbf{R}^3$ are $\{\mathbf{0}\}$, lines and planes through the origin and $\mathbf{R}^3$ itself.

Example 2.47. The linear space $C(\mathbf{R})$ cannot be finite-dimensional, for if it were, then its subspace P of polynomials would also be finite-dimensional, which it is not.

Lemma 2.48. If $\boldsymbol{u}_{k+1}$ is a linear combination of the vectors $\boldsymbol{u}_1, \dots, \boldsymbol{u}_k$, then

$$[\boldsymbol{u}_1, \dots, \boldsymbol{u}_k, \boldsymbol{u}_{k+1}] = [\boldsymbol{u}_1, \dots, \boldsymbol{u}_k].$$

Proof. If $\boldsymbol{u} \in [\boldsymbol{u}_1, \dots, \boldsymbol{u}_k]$, then $\boldsymbol{u} = s_1\boldsymbol{u}_1 + \cdots + s_k\boldsymbol{u}_k + 0\boldsymbol{u}_{k+1} \in [\boldsymbol{u}_1, \dots, \boldsymbol{u}_{k+1}]$. The assumption means that $\boldsymbol{u}_{k+1} = t_1\boldsymbol{u}_1 + \cdots + t_k\boldsymbol{u}_k$. Therefore, if $\boldsymbol{u} \in [\boldsymbol{u}_1, \dots, \boldsymbol{u}_{k+1}]$, then $\boldsymbol{u} = s_1\boldsymbol{u}_1 + \cdots + s_k\boldsymbol{u}_k + s_{k+1}\boldsymbol{u}_{k+1} = (s_1 + t_1 s_{k+1})\boldsymbol{u}_1 + \cdots + (s_k + t_k s_{k+1})\boldsymbol{u}_k \in [\boldsymbol{u}_1, \dots, \boldsymbol{u}_k]$. ■

Theorem 2.49. Let $V \neq \{\mathbf{0}\}$ be a linear space and assume that the vectors $\boldsymbol{u}_1, \dots, \boldsymbol{u}_k$ span V. Then there exists a basis for V comprising the vectors in a subset of $\{\boldsymbol{u}_1, \dots, \boldsymbol{u}_k\}$.

Proof. Among all non-empty subsets S of $\{\boldsymbol{u}_1, \dots, \boldsymbol{u}_k\}$ with the property that the vectors in S span V there must be a set S_0 with a minimum number of vectors. Suppose that the vectors in S_0 are linearly dependent. Then S_0 contains more than one vector, and one vector $\boldsymbol{u} \in S_0$ is a linear combination of the other vectors in S_0. Hence, by Lemma 2.48, the vectors in $S_0 \backslash \{\boldsymbol{u}\}$ span V. Since this contradicts the minimality of S_0, the vectors in S_0 are linearly independent and hence form a basis for V. ■

Theorem 2.50. Let $\boldsymbol{u}_1, \dots, \boldsymbol{u}_n$ be n vectors in an n-dimensional linear space V. Then these vectors span V if and only if they are linearly independent.

Proof. Suppose that $\boldsymbol{u}_1, \dots, \boldsymbol{u}_n$ span V. If they are also linearly dependent, then, by using Theorem 2.49, we can obtain a basis for V by removing one or more of the vectors $\boldsymbol{u}_1, \dots, \boldsymbol{u}_n$. Since this contradicts the fact that all bases for an n-dimensional space consist of n vectors, $\boldsymbol{u}_1, \dots, \boldsymbol{u}_n$ must be linearly independent.

Now suppose that $\boldsymbol{u}_1, \dots, \boldsymbol{u}_n$ are linearly independent. If they do not span V, then, by using Theorem 2.45, we can obtain a basis for V with more than n vectors. We get the same contradiction as before. Hence $\boldsymbol{u}_1, \dots, \boldsymbol{u}_n$ span V. ■

Hence, in order to find out whether n vectors of an n-dimensional space form a basis for that space, it is enough to check if they are linearly independent or to check if they span the space.

Corollary 2.51. Let U be a subspace of V and assume that the two spaces have the same finite dimension. Then $U = V$.

Proof. Assume that the common dimension is n. If $n = 0$, the statement is trivial. Otherwise we can find a basis for U consisting of n vectors. These vectors are linearly independent and belong also to V. Hence, they span V. Therefore, every vector of V is a linear combination of vectors of U, and hence belongs to U. ■

Example 2.52. We want to show that the vectors $\boldsymbol{e}_1 = (1,2,1)$, $\boldsymbol{e}_2 = (1,1,2)$ and $\boldsymbol{e}_3 = (1,4,0)$ form a basis for $\mathbf{R}^3$ and find the coordinates of $\boldsymbol{u} = (3,7,3)$ with respect to that basis. The equation $s_1\boldsymbol{e}_1 + s_2\boldsymbol{e}_2 + s_3\boldsymbol{e}_3 = \boldsymbol{0}$ is equivalent to

$$\begin{cases} s_1 + \ \ s_2 + \ \ s_3 = 0 \\ 2s_1 + \ \ s_2 + 4s_3 = 0 \\ s_1 + 2s_2 \qquad\quad = 0 \end{cases} \Leftrightarrow \begin{cases} s_1 + s_2 + \ \ s_3 = 0 \\ \quad - s_2 + 2s_3 = 0 \\ \qquad s_2 - \ \ s_3 = 0 \end{cases} \Leftrightarrow \begin{cases} s_1 + s_2 + \ \ s_3 = 0 \\ \quad - s_2 + 2s_3 = 0 \\ \qquad\qquad s_3 = 0 \end{cases}.$$

From this we see that $s_1 = s_2 = s_3 = 0$. Hence the vectors are linearly independent. Since the number of vectors is 3 and $\dim \mathbf{R}^3 = 3$, they must form a basis for $\mathbf{R}^3$. In order to find the coordinates (x_1, x_2, x_3) of $\boldsymbol{u}$, we solve the equation $x_1\boldsymbol{e}_1 + x_2\boldsymbol{e}_2 + x_3\boldsymbol{e}_3 = \boldsymbol{u}$, which is equivalent to

$$\begin{cases} x_1 + \ \ x_2 + \ \ x_3 = 3 \\ 2x_1 + \ \ x_2 + 4x_3 = 7 \\ x_1 + 2x_2 \qquad\quad = 3 \end{cases} \Leftrightarrow \begin{cases} x_1 + x_2 + \ \ x_3 = 3 \\ \quad - x_2 + 2x_3 = 1 \\ \qquad x_2 - \ \ x_3 = 0 \end{cases}$$
$$\Leftrightarrow \begin{cases} x_1 + x_2 + \ \ x_3 = 3 \\ \quad - x_2 + 2x_3 = 1 \\ \qquad\qquad x_3 = 1 \end{cases} \Leftrightarrow \quad x_1 = x_2 = x_3 = 1.$$

This shows that the coordinates of $\boldsymbol{u}$ with respect to $\boldsymbol{e}_1, \boldsymbol{e}_2, \boldsymbol{e}_3$ are $(1,1,1)$. Note that the coordinates of $\boldsymbol{u}$ with respect to the ordinary basis $\boldsymbol{\varepsilon}_1, \boldsymbol{\varepsilon}_2, \boldsymbol{\varepsilon}_3$ for $\mathbf{R}^3$ are $(3,7,3)$. When more than one basis are involved, some care must be taken when stating the coordinates of a vector.

Example 2.53. We solve the problem in the previous example by, instead, showing that $\boldsymbol{e}_1, \boldsymbol{e}_2, \boldsymbol{e}_3$ span $\mathbf{R}^3$ and using the fact that $\dim \mathbf{R}^3 = 3$. We must then show that any vector $\boldsymbol{y} = (y_1, y_2, y_3)$ of $\mathbf{R}^3$ is a linear combination of $\boldsymbol{e}_1, \boldsymbol{e}_2, \boldsymbol{e}_3$. We check if this holds by solving the equation $x_1\boldsymbol{e}_1 + x_2\boldsymbol{e}_2 + x_3\boldsymbol{e}_3 = \boldsymbol{y}$.

$$\begin{cases} x_1 + \ \ x_2 + \ \ x_3 = y_1 \\ 2x_1 + \ \ x_2 + 4x_3 = y_2 \\ x_1 + 2x_2 \qquad\quad = y_3 \end{cases} \Leftrightarrow \begin{cases} x_1 + x_2 + \ \ x_3 = y_1 \\ \quad - x_2 + 2x_3 = y_2 - 2y_1 \\ \qquad x_2 - \ \ x_3 = y_3 - y_1 \end{cases}$$
$$\Leftrightarrow \begin{cases} x_1 + x_2 + \ \ x_3 = y_1 \\ \quad - x_2 + 2x_3 = y_2 - 2y_1 \\ \qquad\qquad x_3 = y_3 + y_2 - 3y_1 \end{cases} \Leftrightarrow \begin{cases} x_1 = 8y_1 - 2y_2 - 3y_3 \\ x_2 = -4y_1 + y_2 + 2y_3 \\ x_3 = -3y_1 + y_2 + y_3 \end{cases}.$$

Indeed, the equation has a solution (x_1, x_2, x_3) for every $\boldsymbol{y}$. Therefore, the vectors span $\mathbf{R}^3$ and thus form a basis for $\mathbf{R}^3$. We can now find the coordinates of $\boldsymbol{u}$ by substituting its components $y_1 = 3$, $y_2 = 7$ and $y_3 = 3$ in the last system above. This gives the same result as before, namely $(x_1, x_2, x_3) = (1,1,1)$.

Example 2.54. We use the corollary to prove that $[\boldsymbol{u}_1, \boldsymbol{u}_2] = [\boldsymbol{v}_1, \boldsymbol{v}_2]$ where $\boldsymbol{u}_1$, $\boldsymbol{u}_2$, $\boldsymbol{v}_1$ and $\boldsymbol{v}_2$ are $(1,1,1,1)$, $(2,3,1,-1)$, $(4,5,3,1)$ and $(1,0,2,4)$, respectively. It is plain that $\boldsymbol{u}_1, \boldsymbol{u}_2$ are linearly independent and that $\boldsymbol{v}_1, \boldsymbol{v}_2$ are linearly independent. Hence, both spaces have dimension 2. Therefore, by the corollary, it is sufficient to show that one of the spaces is a subspace of the other. To do so, we begin by solving the equation $x_1\boldsymbol{u}_1 + x_2\boldsymbol{u}_2 + y_1\boldsymbol{v}_1 + y_2\boldsymbol{v}_2 = \mathbf{0}$.

$$\begin{cases} x_1 + 2x_2 + 4y_1 + y_2 = 0 \\ x_1 + 3x_2 + 5y_1 = 0 \\ x_1 + x_2 + 3y_1 + 2y_2 = 0 \\ x_1 - x_2 + y_1 + 4y_2 = 0 \end{cases} \Leftrightarrow \begin{cases} x_1 + 2x_2 + 4y_1 + y_2 = 0 \\ x_2 + y_1 - y_2 = 0 \\ - x_2 - y_1 + y_2 = 0 \\ - 3x_2 - 3y_1 + 3y_2 = 0 \end{cases}$$

$$\Leftrightarrow \begin{cases} x_1 + 2x_2 + 4y_1 + y_2 = 0 \\ x_2 + y_1 - y_2 = 0 \end{cases}.$$

From this we see that we can choose any values for y_1 and y_2 and then solve for x_1 and x_2. By choosing $y_1 = -1$, $y_2 = 0$, we see that there are numbers x_1 and x_2 such that $\boldsymbol{v}_1 = x_1\boldsymbol{u}_1 + x_2\boldsymbol{u}_2$. Hence, $\boldsymbol{v}_1$ is a linear combination of $\boldsymbol{u}_1$ and $\boldsymbol{u}_2$ and therefore belongs to $[\boldsymbol{u}_1, \boldsymbol{u}_2]$. By choosing $y_1 = 0$, $y_2 = -1$, we see that also $\boldsymbol{v}_2$ is a linear combination of $\boldsymbol{u}_1$ and $\boldsymbol{u}_2$ and therefore belongs to $[\boldsymbol{u}_1, \boldsymbol{u}_2]$. Hence, every linear combination of $\boldsymbol{v}_1$ and $\boldsymbol{v}_2$, and therefore every vector of $[\boldsymbol{v}_1, \boldsymbol{v}_2]$, belongs to $[\boldsymbol{u}_1, \boldsymbol{u}_2]$. This shows that $[\boldsymbol{v}_1, \boldsymbol{v}_2]$ is a subspace of $[\boldsymbol{u}_1, \boldsymbol{u}_2]$ and therefore, by the corollary, $[\boldsymbol{u}_1, \boldsymbol{u}_2] = [\boldsymbol{v}_1, \boldsymbol{v}_2]$.

Suppose that the dimension of the linear space V is $n > 0$ and that $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n$ form a basis for V. If $\boldsymbol{u}$ and $\boldsymbol{v}$ are vectors of V having coordinates $\boldsymbol{x} = (x_1, \ldots, x_n)$ and $\boldsymbol{y} = (y_1, \ldots, y_n)$, respectively, then $\boldsymbol{u} = x_1\boldsymbol{e}_1 + \cdots + x_n\boldsymbol{e}_n$ and $\boldsymbol{v} = y_1\boldsymbol{e}_1 + \cdots + y_n\boldsymbol{e}_n$. Hence, the coordinates of

$$\boldsymbol{u} + \boldsymbol{v} = (x_1 + y_1)\boldsymbol{e}_1 + \cdots + (x_n + y_n)\boldsymbol{e}_n \quad \text{and} \quad s\boldsymbol{u} = sx_1\boldsymbol{e}_1 + \cdots + sx_n\boldsymbol{e}_n$$

are

$$\boldsymbol{x} + \boldsymbol{y} = (x_1 + y_1, \ldots, x_n + y_n) \quad \text{and} \quad s\boldsymbol{x} = (sx_1, \ldots, sx_n).$$

By means of a basis for V, we may therefore identify V with $\mathbf{R}^n$. The two spaces behave the same in every linear respect. We exploit this fact in the next example.

Example 2.55. Let V be the space P_2 of polynomials of degree at most 2 and consider the polynomials $\boldsymbol{p}_1 = 1 + 2x + x^2$, $\boldsymbol{p}_2 = 1 + x + 2x^2$ and $\boldsymbol{p}_3 = 1 + 4x$. We intend to show that these vectors form a basis for P_2 and find the coordinates with respect to that basis of $\boldsymbol{p} = 3 + 7x + 3x^2$. The three polynomials $\boldsymbol{\pi}_1 = 1$, $\boldsymbol{\pi}_2 = x$, $\boldsymbol{\pi}_3 = x^2$ form a basis for P_2. With respect to this basis, the coordinates of the vectors $\boldsymbol{p}_1$, $\boldsymbol{p}_2$, $\boldsymbol{p}_3$ and $\boldsymbol{p}$ are $\boldsymbol{e}_1 = (1, 2, 1)$, $\boldsymbol{e}_2 = (1, 1, 2)$, $\boldsymbol{e}_3 = (1, 4, 0)$ and $\boldsymbol{u} = (3, 7, 3)$, respectively. Hence, it is enough to show that $\boldsymbol{e}_1, \boldsymbol{e}_2, \boldsymbol{e}_3$ form a basis for $\mathbf{R}^3$ and find the coordinates of $\boldsymbol{u}$ with respect to this basis. We did this in Example 2.52 and found that the coordinates of $\boldsymbol{u}$ with respect to $\boldsymbol{e}_1, \boldsymbol{e}_2, \boldsymbol{e}_3$ are $(1, 1, 1)$. Hence, $\boldsymbol{p}_1, \boldsymbol{p}_2, \boldsymbol{p}_3$ form a basis for P_2 and the coordinates of $\boldsymbol{p}$ with respect to this basis are $(1, 1, 1)$. Indeed, $\boldsymbol{p} = \boldsymbol{p}_1 + \boldsymbol{p}_2 + \boldsymbol{p}_3$.

2.3 More on Matrices

Let A be an $n \times n$ matrix and denote its columns as usual by $A_1, \ldots, A_n$. By the definition of matrix multiplication, $A\boldsymbol{x} = \boldsymbol{y}$ is equivalent to $x_1A_1 + \cdots + x_nA_n = \boldsymbol{y}$ for all $\boldsymbol{x}$ and $\boldsymbol{y}$ in $\mathbf{R}^n$. We shall use this fact in the proof of the following theorem.

Theorem 2.56. The following statements are equivalent for an $n \times n$ matrix A.

(i) The columns of A are linearly independent.
(ii) The columns of A span $\mathbf{R}^n$.
(iii) The columns of A form a basis for $\mathbf{R}^n$.
(iv) The equation $A\boldsymbol{x} = \mathbf{0}$ has only the trivial solution $\boldsymbol{x} = \mathbf{0}$.
(v) The equation $A\boldsymbol{x} = \boldsymbol{y}$ has some solution $\boldsymbol{x} \in \mathbf{R}^n$ for every $\boldsymbol{y} \in \mathbf{R}^n$.
(vi) The equation $A\boldsymbol{x} = \boldsymbol{y}$ has a unique solution $\boldsymbol{x} \in \mathbf{R}^n$ for every $\boldsymbol{y} \in \mathbf{R}^n$.
(vii) A is invertible.
(viii) A^t is invertible.
(ix) The rows of A are linearly independent.
(x) The rows of A span $\mathbf{R}^n$.
(xi) The rows of A form a basis for $\mathbf{R}^n$.

Proof. The first three statements are equivalent by Theorem 2.50. The equivalences (i) $\Leftrightarrow$ (iv), (ii) $\Leftrightarrow$ (v) and (iii) $\Leftrightarrow$ (vi) follow from the observation made before the theorem. The equivalence of (vi) and (vii) is the content of Theorem 1.19. The equivalence of (vii) and (viii) follows from Theorem 1.23. The last three statements are equivalent to (viii) since the rows of A are the columns of A^t. ■

Corollary 2.57. Let A and B be $n \times n$ matrices. Then $AB = I$ if and only if $BA = I$.

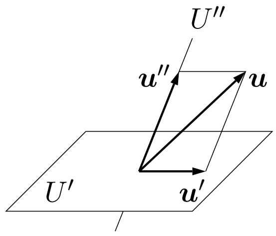
Proof. Assume that $AB = I$. It then follows that if $B\boldsymbol{x} = \mathbf{0}$, then $\boldsymbol{x} = AB\boldsymbol{x} = A\mathbf{0} = \mathbf{0}$. Hence, by the equivalence of (iv) and (vii) in Theorem 2.56, B has an inverse B^{-1}. Multiplying both sides of the equality $AB = I$ from the right by B^{-1}, we get $A = B^{-1}$ and hence $BA = BB^{-1} = I$. The converse now follows by interchanging A and B. ■

After this corollary it is sufficient to show one of the equalities $AB = I$ and $BA = I$ in order to show that a square matrix A is invertible with inverse B. In Example 1.17 we had to show both.

2.4 Direct Sums

Definition 2.58. Let U' and U'' be subspaces of a linear space V. We say that V is the sum of U' and U'' and write $V = U' + U''$ if every vector $\boldsymbol{u} \in V$ can be written as $\boldsymbol{u} = \boldsymbol{u}' + \boldsymbol{u}''$ where $\boldsymbol{u}' \in U'$ and $\boldsymbol{u}'' \in U''$. If $\boldsymbol{u}'$ and $\boldsymbol{u}''$ are uniquely determined by $\boldsymbol{u}$ for every vector $\boldsymbol{u} \in V$, we say that V is the direct sum of U' and U'' and write $V = U' \oplus U''$.

Definition 2.59. Assume that $V = U' \oplus U''$ is the direct sum of U' and U'' and let $\boldsymbol{u} \in V$. The unique vectors $\boldsymbol{u}'$ and $\boldsymbol{u}''$ such that $\boldsymbol{u} = \boldsymbol{u}' + \boldsymbol{u}''$ are called the projections of $\boldsymbol{u}$ on U' along U'' and of $\boldsymbol{u}$ on U'' along U', respectively.



Theorem 2.60. Using the same notation as in Definition 2.59, we have

$$(s\boldsymbol{u}+t\boldsymbol{v})' = s\boldsymbol{u}'+t\boldsymbol{v}' \quad \text{and} \quad (s\boldsymbol{u}+t\boldsymbol{v})'' = s\boldsymbol{u}''+t\boldsymbol{v}''.$$

Proof. The statement follows from the uniqueness and the fact that

$$\begin{aligned}(s\boldsymbol{u}+t\boldsymbol{v})' + (s\boldsymbol{u}+t\boldsymbol{v})'' &= s\boldsymbol{u}+t\boldsymbol{v} = s(\boldsymbol{u}'+\boldsymbol{u}'')+t(\boldsymbol{v}'+\boldsymbol{v}'') \\ &= (s\boldsymbol{u}'+t\boldsymbol{v}')+(s\boldsymbol{u}''+t\boldsymbol{v}''). \ \blacksquare\end{aligned}$$

Theorem 2.61. Let U' and U'' be subspaces of a linear space V. Then $V = U' \oplus U''$ if and only if $V = U' + U''$ and $U' \cap U'' = \{\boldsymbol{0}\}$.

Proof. Assume that $V = U' \oplus U''$. Then, by definition, $V = U' + U''$. If $\boldsymbol{u} \in U' \cap U''$, we use the fact that $\boldsymbol{u} = \boldsymbol{u} + \boldsymbol{0} = \boldsymbol{0} + \boldsymbol{u}$ and the uniqueness to conclude that $\boldsymbol{u} = \boldsymbol{0}$. Hence, $U' \cap U'' = \{\boldsymbol{0}\}$.

To prove the converse, we assume that $\boldsymbol{u} = \boldsymbol{u}' + \boldsymbol{u}'' = \boldsymbol{v}' + \boldsymbol{v}''$ where $\boldsymbol{u}'$ and $\boldsymbol{v}'$ belong to U' and $\boldsymbol{u}''$ and $\boldsymbol{v}''$ belong to U''. Then $\boldsymbol{u}' - \boldsymbol{v}' = \boldsymbol{v}'' - \boldsymbol{u}'' \in U' \cap U'' = \{\boldsymbol{0}\}$, and hence $\boldsymbol{u}' - \boldsymbol{v}' = \boldsymbol{v}'' - \boldsymbol{u}'' = \boldsymbol{0}$. This proves the uniqueness, and therefore $V = U' \oplus U''$. $\blacksquare$

Theorem 2.62. Let V be a finite-dimensional linear space and assume that $V = U' \oplus U''$ where U' and U'' are subspaces of V. Then $\dim U' + \dim U'' = \dim V$.

Proof. If one of the subspaces is the zero subspace, then the other subspace equals V, and the statement is trivial. We may therefore assume that none of the subspaces is the zero subspace. Since both subspaces are finite-dimensional, we can choose bases $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_k$ and $\boldsymbol{f}_1, \ldots, \boldsymbol{f}_m$ for U' and U'', respectively. We show that $\dim V = k+m = \dim U' + \dim U''$ by showing that $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_k, \boldsymbol{f}_1, \ldots, \boldsymbol{f}_m$ form a basis for V. If $\boldsymbol{u} \in V$, we can write $\boldsymbol{u} = \boldsymbol{u}' + \boldsymbol{u}''$ where $\boldsymbol{u}' \in U'$ and $\boldsymbol{u}'' \in U''$. Since $\boldsymbol{u}'$ and $\boldsymbol{u}''$ are linear combinations of $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_k$ and $\boldsymbol{f}_1, \ldots, \boldsymbol{f}_m$, respectively, it follows that $\boldsymbol{u}$ is a linear combination of the vectors $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_k, \boldsymbol{f}_1, \ldots, \boldsymbol{f}_m$. It remains to show that these vectors are linearly independent. Suppose that

$$s_1\boldsymbol{e}_1 + \cdots + s_k\boldsymbol{e}_k + t_1\boldsymbol{f}_1 + \cdots + t_m\boldsymbol{f}_m = \boldsymbol{0}.$$

Then $s_1\boldsymbol{e}_1 + \cdots + s_k\boldsymbol{e}_k = -(t_1\boldsymbol{f}_1 + \cdots + t_m\boldsymbol{f}_m) \in U' \cap U'' = \{\boldsymbol{0}\}$, and hence

$$s_1\boldsymbol{e}_1 + \cdots + s_k\boldsymbol{e}_k = \boldsymbol{0} \quad \text{and} \quad t_1\boldsymbol{f}_1 + \cdots + t_m\boldsymbol{f}_m = \boldsymbol{0}.$$

Since $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_k$ are linearly independent and $\boldsymbol{f}_1, \ldots, \boldsymbol{f}_m$ are linearly independent, it follows that $s_1 = \cdots = s_k = t_1 = \cdots = t_m = 0$. $\blacksquare$

Example 2.63. Let V be three-space and consider the plane U' and the line U'' defined by $x_1 + 2x_2 + 3x_3 = 0$ and $\boldsymbol{x} = t(2,1,2)$, respectively. It is easily checked that the intersection of the two subspaces is the zero space. Hence, their sum is direct, and therefore its dimension is $2+1=3$. Thus $V = U' \oplus U''$. Let $\boldsymbol{u} = (2,3,4)$. In order to find $\boldsymbol{u}'$ and $\boldsymbol{u}''$, we form the line $\boldsymbol{x} = t(2,1,2) + (2,3,4)$ through $\boldsymbol{u}$. Its intersection with the plane is given by

$$2t + 2 + 2(t+3) + 3(2t+4) = 10t + 20 = 0 \quad \Leftrightarrow \quad t = -2.$$

Hence, $\boldsymbol{u}' = -2(2,1,2) + (2,3,4) = (-2,1,0)$ and $\boldsymbol{u}'' = \boldsymbol{u} - \boldsymbol{u}' = (4,2,4)$.

2.5 The Rank-Nullity Theorem

Definition 2.64. Let U and V be linear spaces. A linear transformation F from U to V is a function $F : U \to V$ such that

$$F(s\boldsymbol{u} + t\boldsymbol{v}) = sF(\boldsymbol{u}) + tF(\boldsymbol{v})$$

for all $\boldsymbol{u}$ and $\boldsymbol{v}$ in U and all real numbers s and t. If $U = V$, we also say that F is a linear transformation on U.

The single condition in the definition can be replaced with the two conditions

$$F(\boldsymbol{u} + \boldsymbol{v}) = F(\boldsymbol{u}) + F(\boldsymbol{v}) \quad \text{and} \quad F(s\boldsymbol{u}) = sF(\boldsymbol{u}).$$

For if the condition in the definition holds, then

$$F(\boldsymbol{u} + \boldsymbol{v}) = F(1\boldsymbol{u} + 1\boldsymbol{v}) = 1F(\boldsymbol{u}) + 1F(\boldsymbol{v}) = F(\boldsymbol{u}) + F(\boldsymbol{v})$$

and

$$F(s\boldsymbol{u}) = F(s\boldsymbol{u} + 0\boldsymbol{u}) = sF(\boldsymbol{u}) + 0F(\boldsymbol{u}) = sF(\boldsymbol{u}).$$

Conversely, if the two conditions hold, then

$$F(s\boldsymbol{u} + t\boldsymbol{v}) = F(s\boldsymbol{u}) + F(t\boldsymbol{v}) = sF(\boldsymbol{u}) + tF(\boldsymbol{v}).$$

Also note that $F(\boldsymbol{0}) = F(0\boldsymbol{0}) = 0F(\boldsymbol{0}) = \boldsymbol{0}$.

Definition 2.65. Let F be a linear transformation from U to V. We define the kernel and image of F by

$$\ker F = \{\boldsymbol{u} \in U;\ F(\boldsymbol{u}) = \boldsymbol{0}\} \quad \text{and} \quad \operatorname{im} F = \{\boldsymbol{v} \in V;\ \boldsymbol{v} = F(\boldsymbol{u}) \text{ for some } \boldsymbol{u} \in U\}.$$

As in the proof of Theorem 2.18, one sees that $\ker F$ and $\operatorname{im} F$ are subspaces of U and V, respectively. We also note that if A is an $m \times n$ matrix, then $F(\boldsymbol{x}) = A\boldsymbol{x}$ defines a linear transformation F from $\mathbf{R}^n$ to $\mathbf{R}^m$ whose kernel and image agree with the kernel and image of A.

Example 2.66. Let $U = C^n(\mathbf{R})$ be the space of n times continuously differentiable functions on $\mathbf{R}$ and let $V = C(\mathbf{R})$. If $\boldsymbol{u} \in U$, define $F(\boldsymbol{u})$ by

$$F(\boldsymbol{u})(t) = \boldsymbol{u}^{(n)}(t) + a_1\boldsymbol{u}^{(n-1)}(t) + \cdots + a_{n-1}\boldsymbol{u}'(t) + a_n\boldsymbol{u}(t), \quad t \in \mathbf{R}.$$

Then F is a linear transformation from U to V. The kernel of F is the set of solutions of the homogenous linear differential equation

$$\boldsymbol{u}^{(n)}(t) + a_1\boldsymbol{u}^{(n-1)}(t) + \cdots + a_{n-1}\boldsymbol{u}'(t) + a_n\boldsymbol{u}(t) = 0, \quad t \in \mathbf{R}.$$

When $n = 2$ and the roots λ_1 and λ_2 of the characteristic equation are real and unequal, $\ker F$ is spanned by $\boldsymbol{e}_1$ and $\boldsymbol{e}_2$ defined by $\boldsymbol{e}_1(t) = e^{\lambda_1 t}$ and $\boldsymbol{e}_2(t) = e^{\lambda_2 t}$. The vectors $\boldsymbol{e}_1$ and $\boldsymbol{e}_2$ are linearly independent, for if

$$s_1 e^{\lambda_1 t} + s_2 e^{\lambda_2 t} = 0, \quad t \in \mathbf{R},$$

then, by taking the derivative,

$$\lambda_1 s_1 e^{\lambda_1 t} + \lambda_2 s_2 e^{\lambda_2 t} = 0, \quad t \in \mathbf{R}.$$

By inserting $t = 0$, we get

$$\begin{cases} \quad s_1 + \quad s_2 = 0 \\ \lambda_1 s_1 + \lambda_2 s_2 = 0 \end{cases}.$$

Since $\lambda_1 \neq \lambda_2$, this is possible only if $s_1 = s_2 = 0$. Thus we have shown that $\ker F$ is two-dimensional with basis $\boldsymbol{e}_1, \boldsymbol{e}_2$.

Theorem 2.67 (Rank-nullity theorem). Let F be a linear transformation from U to V where U and V are linear spaces and $\dim U = n$. Then $\operatorname{im} F$ is finite-dimensional and

$$\dim\ker F + \dim\operatorname{im} F = n.$$

Proof. If $\ker F = U$, then $\operatorname{im} F = \{\mathbf{0}\}$, and the statement holds trivially. Otherwise, $\ker F \neq U$, and hence $U \neq \{\mathbf{0}\}$. We can therefore choose a possibly empty set of basis vectors $\{\boldsymbol{e}_1, \ldots, \boldsymbol{e}_k\}$ for $\ker F$ and extend it to a basis $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_k, \boldsymbol{e}_{k+1}, \ldots, \boldsymbol{e}_n$ for U. Set $\boldsymbol{f}_i = F(\boldsymbol{e}_i)$ for $i = 1, \ldots, n$. Then $\boldsymbol{f}_i \in \operatorname{im} F$ for $i = 1, \ldots, n$ and $\boldsymbol{f}_i = 0$ for $i = 1, \ldots, k$. We show that $\dim\operatorname{im} F = n - k = n - \dim\ker F$ by showing that $\boldsymbol{f}_{k+1}, \ldots, \boldsymbol{f}_n$ form a basis for $\operatorname{im} F$.

First, we show that they are linearly independent. Suppose that

$$s_{k+1}\boldsymbol{f}_{k+1} + \cdots + s_n\boldsymbol{f}_n = \mathbf{0}.$$

Then, by the definition of the $\boldsymbol{f}_i$,

$$F(s_{k+1}\boldsymbol{e}_{k+1} + \cdots + s_n\boldsymbol{e}_n) = s_{k+1}F(\boldsymbol{e}_{k+1}) + \cdots + s_nF(\boldsymbol{e}_n) = \mathbf{0}.$$

Hence, $s_{k+1}\boldsymbol{e}_{k+1} + \cdots + s_n\boldsymbol{e}_n \in \ker F$. If $\ker F = \{\mathbf{0}\}$, this implies that

$$s_{k+1}\boldsymbol{e}_{k+1} + \cdots + s_n\boldsymbol{e}_n = \mathbf{0},$$

and consequently $s_{k+1} = \cdots = s_n = 0$ since the $\boldsymbol{e}_i$ are linearly independent. If $\ker F \neq \{\boldsymbol{0}\}$, then $s_{k+1}\boldsymbol{e}_{k+1} + \cdots + s_n\boldsymbol{e}_n$ is a linear combination of $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_k$. From this we get $s_{k+1} = \cdots = s_n = 0$ by once again using the fact that the $\boldsymbol{e}_i$ are linearly independent. Hence, $\boldsymbol{f}_{k+1}, \ldots, \boldsymbol{f}_n$ are linearly independent.

We complete the proof by showing that $\boldsymbol{f}_{k+1}, \ldots, \boldsymbol{f}_n$ span $\operatorname{im} F$. Let $\boldsymbol{v}$ be any vector of $\operatorname{im} F$. Then $\boldsymbol{v} = F(\boldsymbol{u})$ for some $\boldsymbol{u} \in U$. Since $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n$ form a basis for U, we can write $\boldsymbol{u} = x_1\boldsymbol{e}_1 + \cdots + x_k\boldsymbol{e}_k + x_{k+1}\boldsymbol{e}_{k+1} + \cdots + x_n\boldsymbol{e}_n$. Hence,

$$\begin{aligned}\boldsymbol{v} &= F(x_1\boldsymbol{e}_1 + \cdots + x_k\boldsymbol{e}_k + x_{k+1}\boldsymbol{e}_{k+1} + \cdots + x_n\boldsymbol{e}_n) \\ &= x_1\boldsymbol{f}_1 + \cdots + x_k\boldsymbol{f}_k + x_{k+1}\boldsymbol{f}_{k+1} + \cdots + x_n\boldsymbol{f}_n \\ &= x_{k+1}\boldsymbol{f}_{k+1} + \cdots + x_n\boldsymbol{f}_n\end{aligned}$$

is a linear combination of $\boldsymbol{f}_{k+1}, \ldots, \boldsymbol{f}_n$. ■

Example 2.68. We use the theorem to find the dimensions of $\ker A$ and $\operatorname{im} A$ where

$$A = \begin{bmatrix} 1 & -1 & 1 & -1 & 1 \\ 2 & 3 & 4 & 2 & 5 \\ 3 & 7 & 7 & 5 & 9 \\ 4 & 6 & 8 & 4 & 10 \end{bmatrix}.$$

We begin by solving the equation $A\boldsymbol{x} = \boldsymbol{0}$.

$$\begin{aligned} A\boldsymbol{x} = \boldsymbol{0} \quad &\Leftrightarrow \quad \left[\begin{array}{ccccc|c} 1 & -1 & 1 & -1 & 1 & 0 \\ 2 & 3 & 4 & 2 & 5 & 0 \\ 3 & 7 & 7 & 5 & 9 & 0 \\ 4 & 6 & 8 & 4 & 10 & 0 \end{array}\right] \quad \Leftrightarrow \quad \left[\begin{array}{ccccc|c} 1 & -1 & 1 & -1 & 1 & 0 \\ 0 & 5 & 2 & 4 & 3 & 0 \\ 0 & 10 & 4 & 8 & 6 & 0 \\ 0 & 10 & 4 & 8 & 6 & 0 \end{array}\right] \\ &\Leftrightarrow \quad \left[\begin{array}{ccccc|c} 1 & -1 & 1 & -1 & 1 & 0 \\ 0 & 5 & 2 & 4 & 3 & 0 \end{array}\right] \\ &\Leftrightarrow \quad \boldsymbol{x} = r(-7,-2,5,0,0) + s(1,-4,0,5,0) + t(-8,-3,0,0,5). \end{aligned}$$

Hence, $\dim\ker A = 3$, and by the rank-nullity theorem, $\dim\operatorname{im} A = 5 - 3 = 2$. In fact, in order to find the dimension of $\dim\ker A$ there is no need to find the solution $\boldsymbol{x}$. It is enough to establish that there are three free variables in the last system. Knowing that $\dim\operatorname{im} A = 2$, we can easily find a basis for $\operatorname{im} A$. We can choose any two non-proportional columns of A, for example the first two columns.

Theorem 2.69. Let $G : U \to V$ and $F : V \to W$ be linear transformations. Then their composition $F \circ G : U \to W$ is also a linear transformation.

Proof. Let $\boldsymbol{u}$ and $\boldsymbol{v}$ be vectors of U and s and t real numbers. Then

$$\begin{aligned}(F \circ G)(s\boldsymbol{u} + t\boldsymbol{v}) = F(G(s\boldsymbol{u} + t\boldsymbol{v})) &= F(sG(\boldsymbol{u}) + tG(\boldsymbol{v})) = sF(G(\boldsymbol{u})) + tF(G(\boldsymbol{v})) \\ &= s(F \circ G)(\boldsymbol{u}) + t(F \circ G)(\boldsymbol{v}). \ \blacksquare\end{aligned}$$

We shall usually omit the circle and write $F \circ G$ as FG.

Definition 2.70. A linear transformation $F : U \to V$ is said to be invertible if it is onto and one-to-one.

Theorem 2.71. Let $F : U \to V$ be an invertible linear transformation. Then its inverse function $F^{-1} : V \to U$ is a linear transformation.

Proof. Let $\boldsymbol{u}$ and $\boldsymbol{v}$ be vectors of V and s and t real numbers. Set $\boldsymbol{u}' = F^{-1}(\boldsymbol{u})$ and $\boldsymbol{v}' = F^{-1}(\boldsymbol{v})$. The linearity of F then yields $s\boldsymbol{u} + t\boldsymbol{v} = sF(\boldsymbol{u}') + tF(\boldsymbol{v}') = F(s\boldsymbol{u}' + t\boldsymbol{v}')$, and hence

$$sF^{-1}(\boldsymbol{u}) + tF^{-1}(\boldsymbol{v}) = s\boldsymbol{u}' + t\boldsymbol{v}' = F^{-1}(s\boldsymbol{u} + t\boldsymbol{v}). \blacksquare$$

Lemma 2.72. Let $F : U \to V$ be a linear transformation. Then F is one-to-one if and only if $\ker F = \{\boldsymbol{0}\}$.

Proof. If F is one-to-one, then the only solution of the equation $F(\boldsymbol{u}) = \boldsymbol{0}$ is $\boldsymbol{u} = \boldsymbol{0}$, and hence $\ker F = \{\boldsymbol{0}\}$. Conversely, suppose that $\ker F = \{\boldsymbol{0}\}$. If $F(\boldsymbol{u}) = F(\boldsymbol{v})$, then $F(\boldsymbol{u} - \boldsymbol{v}) = \boldsymbol{0}$. Therefore $\boldsymbol{u} - \boldsymbol{v} \in \ker F = \{\boldsymbol{0}\}$, whence $\boldsymbol{u} = \boldsymbol{v}$. This means that F is one-to-one. $\blacksquare$

Theorem 2.73. Let $F : U \to V$ be a linear transformation where U and V are finite-dimensional linear spaces of the same dimension. Then F is one-to-one if and only if F is onto.

Proof. Let n be the common dimension of U and V. Then the rank-nullity theorem and Corollary 2.51 yield

$$\begin{aligned} F \text{ is one-to-one} \quad &\Leftrightarrow \quad \ker F = \{\boldsymbol{0}\} \quad \Leftrightarrow \quad \dim\ker F = 0 \\ &\Leftrightarrow \quad \dim \operatorname{im} F = n \quad \Leftrightarrow \quad \operatorname{im} F = V \quad \Leftrightarrow \quad F \text{ is onto.} \blacksquare \end{aligned}$$

Theorem 2.74. Let $\boldsymbol{u}_1, \dots, \boldsymbol{u}_n$ be a basis for a linear space U and let $\boldsymbol{v}_1, \dots, \boldsymbol{v}_n$ be vectors in a linear space V. Then there exists a unique linear transformation $F : U \to V$ such that $F(\boldsymbol{u}_i) = \boldsymbol{v}_i$ for $i = 1, \dots, n$. Moreover, F is one-to-one if and only if $\boldsymbol{v}_1, \dots, \boldsymbol{v}_n$ are linearly independent.

Proof. Any such linear transformation F must satisfy

$$F(x_1\boldsymbol{u}_1 + \cdots + x_n\boldsymbol{u}_n) = x_1\boldsymbol{v}_1 + \cdots + x_n\boldsymbol{v}_n.$$

If we take this as a definition, then F is well-defined, for if

$$x_1\boldsymbol{u}_1 + \cdots + x_n\boldsymbol{u}_n = y_1\boldsymbol{u}_1 + \cdots + y_n\boldsymbol{u}_n,$$

then $x_i = y_i$ for $i = 1, \dots, n$, and hence

$$x_1\boldsymbol{v}_1 + \cdots + x_n\boldsymbol{v}_n = y_1\boldsymbol{v}_1 + \cdots + y_n\boldsymbol{v}_n.$$

F is a linear transformation, for if $\boldsymbol{u}$ and $\boldsymbol{u}'$ are vectors in U with coordinate vectors $\boldsymbol{x}$ and $\boldsymbol{x}'$ with respect to the basis $\boldsymbol{u}_1, \dots, \boldsymbol{u}_n$, then the coordinate vector of $s\boldsymbol{u} + s'\boldsymbol{u}'$ is $s\boldsymbol{x} + s'\boldsymbol{x}'$, and hence

$$F(s\boldsymbol{u} + s'\boldsymbol{u}') = sF(\boldsymbol{u}) + s'F(\boldsymbol{u}').$$

We have $x_1\boldsymbol{v}_1 + \cdots + x_n\boldsymbol{v}_n = \boldsymbol{0}$ if and only if $x_1\boldsymbol{u}_1 + \cdots + x_n\boldsymbol{u}_n \in \ker F$. This shows that $\ker F = \{\boldsymbol{0}\}$ if and only if $\boldsymbol{v}_1, \dots, \boldsymbol{v}_n$ are linearly independent and concludes the proof. $\blacksquare$

Exercises

2.1. Which of the following sets are subspaces of $\mathbf{R}^2$? Justify your answers.

(a) $\{\boldsymbol{x} \in \mathbf{R}^2;\ x_1 = 2x_2\}$, (b) $\{\boldsymbol{x} \in \mathbf{R}^2;\ (x_1, x_2) = t(1,2) + (1,1),\ t \in \mathbf{R}\}$,

(c) $\{\boldsymbol{x} \in \mathbf{R}^2;\ x_1 = x_2^2\}$, (d) $\{\boldsymbol{x} \in \mathbf{R}^2;\ (x_1, x_2) = t(1,2),\ t \in \mathbf{R}\}$.

2.2. Show that the set of symmetric $n \times n$ matrices is a subspace of $M_{n\times n}$.

2.3. Express the plane through the origin and the two points $(1,1,0)$ and $(2,0,1)$ as the image of a matrix and as the kernel of a matrix.

2.4. Which of the following sets of vectors are linearly dependent?

(a) $(1,2,3)$, $(2,3,3)$, $(2,5,7)$ in $\mathbf{R}^3$,

(b) $(1,2,3,1)$, $(2,3,2,3)$, $(1,1,-1,2)$ in $\mathbf{R}^4$,

(c) $(1,2,3,1,2)$, $(2,3,2,3,1)$, $(1,1,-1,2,3)$ in $\mathbf{R}^5$.

2.5. Consider the following vectors in $\mathbf{R}^4$.

$$\boldsymbol{u}_1 = (1,1,1,2), \quad \boldsymbol{u}_2 = (1,2,3,4), \quad \boldsymbol{u}_3 = (2,1,2,3), \quad \boldsymbol{u}_4 = (5,1,3,5).$$

Is $\boldsymbol{u}_1$ a linear combination of $\boldsymbol{u}_2$, $\boldsymbol{u}_3$ and $\boldsymbol{u}_4$? Are $\boldsymbol{u}_1$, $\boldsymbol{u}_2$, $\boldsymbol{u}_3$ and $\boldsymbol{u}_4$ linearly dependent?

2.6. Show that the vectors $\boldsymbol{u}_1$, $\boldsymbol{u}_2$ and $\boldsymbol{u}_3$ in $C(\mathbf{R})$ defined by

$$\boldsymbol{u}_1(t) = \sin t, \quad \boldsymbol{u}_2(t) = \sin 2t, \quad \boldsymbol{u}_3(t) = \sin 3t$$

are linearly independent.

2.7. Find bases for $\ker A$ and $\operatorname{im} A$ where

(a) $A = \begin{bmatrix} 1 & 1 & 3 & 3 \\ 1 & -1 & 1 & -1 \\ 2 & 1 & 5 & 4 \end{bmatrix}$, (b) $A = \begin{bmatrix} 1 & 3 & 4 & 1 \\ 1 & 1 & 2 & -1 \\ 2 & 5 & 7 & 2 \end{bmatrix}$.

2.8. Find a basis for the span of the vectors

$$(1,3,2,1), \quad (1,2,1,1), \quad (1,1,0,1), \quad (1,2,2,1), \quad (3,4,1,3).$$

2.9. Show that the vectors

$$(1,0,1,0), \quad (1,1,1,1), \quad (2,1,-1,2), \quad (1,-2,-2,1)$$

form a basis for $\mathbf{R}^4$. What are the coordinates of $(2,2,-1,1)$ with respect to this basis?

2.10. Find the dimensions of the spans of the following vectors in $\mathbf{R}^4$.

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

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

2.11. Show that $\boldsymbol{u}_1 = (1,1,1,-1)$, $\boldsymbol{u}_2 = (1,2,3,4)$ span the same subspace of $\mathbf{R}^4$ as $\boldsymbol{v}_1 = (-1,1,3,11)$, $\boldsymbol{v}_2 = (3,1,-1,-13)$.

2.12. Let $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_n$ be a basis for a linear space. What is the dimension of the subspace $[\boldsymbol{u}_1 - \boldsymbol{u}_2, \boldsymbol{u}_2 - \boldsymbol{u}_3, \ldots, \boldsymbol{u}_{n-1} - \boldsymbol{u}_n, \boldsymbol{u}_n - \boldsymbol{u}_1]$?

2.13. Let A and B be $n \times n$ matrices such that $A^2 - AB = I$. Show that $A^2 - BA = I$.

2.14. For which of the subspaces U and V of $\mathbf{R}^4$ is the sum $U + V$ direct?

(a) $U = \{\boldsymbol{x} \in \mathbf{R}^4;\ x_1 + x_2 + x_3 + x_4 = 0\}$, $V = \{\boldsymbol{x} \in \mathbf{R}^4;\ x_1 - x_2 + x_3 - x_4 = 0\}$.

(b) $U = \{(t, 0, -t, t) \in \mathbf{R}^4;\ t \in \mathbf{R}\}$, $V = \{\boldsymbol{x} \in \mathbf{R}^4;\ x_1 + 2x_2 + 3x_3 + 4x_4 = 0\}$.

2.15. Show that $\mathbf{R}^3 = U \oplus V$ where

$$U = \{\boldsymbol{x} \in \mathbf{R}^3;\ x_1 + x_2 + 3x_3 = 0\}, \quad V = \{t(1,1,2) \in \mathbf{R}^3;\ t \in \mathbf{R}\}.$$

Find the projections of $\boldsymbol{u} = (4,5,5)$ on U along V and on V along U.

2.16. Which of the following functions $F : \mathbf{R}^3 \to \mathbf{R}^2$ are linear transformations?

(a) $F(x_1, x_2, x_3) = (x_1 + 2x_2 + 1, x_2 - 2x_3 - 1)$,

(b) $F(x_1, x_2, x_3) = (x_1 + x_2, x_2 - x_3)$,

(c) $F(x_1, x_2, x_3) = (x_1 x_2, x_2 x_3)$.

Justify your answers.

2.17. Find the dimensions of $\ker A$ and $\operatorname{im} A$ where

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

2.18. Consider the linear transformation F from $\mathbf{R}^2$ to $\mathbf{R}^3$ defined by

$$F(x_1, x_2) = (x_1 + x_2, 2x_1 + x_2, 3x_1 + x_2).$$

Is F one-to-one? Is it onto?

2.19. Show that the linear transformation F on $\mathbf{R}^3$ defined by

$$F(x_1, x_2, x_3) = (x_1 + 2x_2 + 2x_3, 2x_1 + 2x_2 + x_3, 3x_1 + x_2 - x_3)$$

is invertible. Find the inverse transformation.

2.20. Let P be the space of polynomials over $\mathbf{R}$ and let F be the linear transformation on P defined by $F(\boldsymbol{p}) = \boldsymbol{p}'$ where $\boldsymbol{p}'$ denotes the derivative of the polynomial $\boldsymbol{p}$. Is F one-to-one? Is F onto? What are the kernel and image of F?

3 Inner Product Spaces

3.1 Definition

Definition 3.1. Let V be a linear space, and let there be defined a function $V \times V \to \mathbf{R}$ whose value at $(\boldsymbol{u}, \boldsymbol{v})$ is denoted by $\langle \boldsymbol{u}, \boldsymbol{v} \rangle$. We call the function an inner product on V if the following conditions are satisfied.

(i) $\langle s\boldsymbol{u} + t\boldsymbol{v}, \boldsymbol{w} \rangle = s\langle \boldsymbol{u}, \boldsymbol{w} \rangle + t\langle \boldsymbol{v}, \boldsymbol{w} \rangle$ for all $\boldsymbol{u}$, $\boldsymbol{v}$ and $\boldsymbol{w}$ in V and all s and t in $\mathbf{R}$.
(ii) $\langle \boldsymbol{u}, \boldsymbol{v} \rangle = \langle \boldsymbol{v}, \boldsymbol{u} \rangle$ for all $\boldsymbol{u}$ and $\boldsymbol{v}$ in V.
(iii) $\langle \boldsymbol{u}, \boldsymbol{u} \rangle \geq 0$ for all $\boldsymbol{u} \in V$ with equality only if $\boldsymbol{u} = \mathbf{0}$.

A linear space, furnished with an inner product, is called an inner product space. We call $\|\boldsymbol{u}\| = \sqrt{\langle \boldsymbol{u}, \boldsymbol{u} \rangle}$ the norm or length of a vector $\boldsymbol{u}$ in an inner product space.

When we talk about a subspace of an inner product space V, we shall assume that it is equipped with the same inner product as V.

Note that Axiom (i) means that $\boldsymbol{u} \mapsto \langle \boldsymbol{u}, \boldsymbol{w} \rangle$ is a linear transformation from V to $\mathbf{R}$ for every fixed $\boldsymbol{w} \in V$. Hence, Axiom (i) is equivalent to $\langle \boldsymbol{u} + \boldsymbol{v}, \boldsymbol{w} \rangle = \langle \boldsymbol{u}, \boldsymbol{w} \rangle + \langle \boldsymbol{v}, \boldsymbol{w} \rangle$ and $\langle s\boldsymbol{u}, \boldsymbol{w} \rangle = s\langle \boldsymbol{u}, \boldsymbol{w} \rangle$ for all $\boldsymbol{u}$, $\boldsymbol{v}$, $\boldsymbol{w}$ in V and all $s \in \mathbf{R}$.

Theorem 3.2.

(i) $\langle \boldsymbol{w}, s\boldsymbol{u} + t\boldsymbol{v} \rangle = s\langle \boldsymbol{w}, \boldsymbol{u} \rangle + t\langle \boldsymbol{w}, \boldsymbol{v} \rangle$ for all $\boldsymbol{u}$, $\boldsymbol{v}$ and $\boldsymbol{w}$ in V and all s and t in $\mathbf{R}$.
(ii) $\langle \mathbf{0}, \boldsymbol{u} \rangle = \langle \boldsymbol{u}, \mathbf{0} \rangle = 0$ for all $\boldsymbol{u} \in V$.
(iii) $\|s\boldsymbol{u}\| = |s| \, \|\boldsymbol{u}\|$ for all $\boldsymbol{u} \in V$ and $s \in \mathbf{R}$.

Proof. (i) We have

$$\langle \boldsymbol{w}, s\boldsymbol{u} + t\boldsymbol{v} \rangle = \langle s\boldsymbol{u} + t\boldsymbol{v}, \boldsymbol{w} \rangle = s\langle \boldsymbol{u}, \boldsymbol{w} \rangle + t\langle \boldsymbol{v}, \boldsymbol{w} \rangle = s\langle \boldsymbol{w}, \boldsymbol{u} \rangle + t\langle \boldsymbol{w}, \boldsymbol{v} \rangle$$

by Axioms (i) and (ii).

(ii) By the same two axioms,

$$\langle \boldsymbol{u}, \mathbf{0} \rangle = \langle \mathbf{0}, \boldsymbol{u} \rangle = \langle 0\boldsymbol{u}, \boldsymbol{u} \rangle = 0\langle \boldsymbol{u}, \boldsymbol{u} \rangle = 0.$$

(iii) $\|s\boldsymbol{u}\| = \sqrt{\langle s\boldsymbol{u}, s\boldsymbol{u} \rangle} = \sqrt{s^2\langle \boldsymbol{u}, \boldsymbol{u} \rangle} = |s| \, \|\boldsymbol{u}\|$. ■

Hence, also $\boldsymbol{u} \mapsto \langle \boldsymbol{w}, \boldsymbol{u} \rangle$ is a linear transformation from V to $\mathbf{R}$ for every $\boldsymbol{w} \in V$.

Definition 3.3. The dot product on $\mathbf{R}^n$ is defined by

$$(x_1, x_2, \ldots, x_n) \cdot (y_1, y_2, \ldots, y_n) = x_1y_1 + x_2y_2 + \cdots + x_ny_n.$$

We leave the simple proof of the following theorem to the reader.

Theorem 3.4. $\langle \boldsymbol{x}, \boldsymbol{y} \rangle = \boldsymbol{x} \cdot \boldsymbol{y}$ is an inner product on $\mathbf{R}^n$. The corresponding norm is

$$\|(x_1, \dots, x_n)\| = \sqrt{x_1^2 + \cdots + x_n^2}.$$

Example 3.5. Also $\langle \boldsymbol{x}, \boldsymbol{y} \rangle = x_1 y_1 + 2x_2 y_2 + 3x_3 y_3 + \cdots + n x_n y_n$ is an inner product on $\mathbf{R}^n$.

When we mention the inner product space $\mathbf{R}^n$ without further specification, it is always assumed that the inner product is the dot product.

Example 3.6. Let $V = C[0,1]$. Then $\langle \boldsymbol{u}, \boldsymbol{v} \rangle = \int_0^1 \boldsymbol{u}(x)\boldsymbol{v}(x)\,dx$ is an inner product on V. Axioms (i) and (ii) are easily verified. It is also clear that $\langle \boldsymbol{u}, \boldsymbol{u} \rangle = \int_0^1 (\boldsymbol{u}(x))^2\,dx \geq 0$ for all $\boldsymbol{u} \in V$. Suppose that $\boldsymbol{u} \neq \mathbf{0}$. Then $\boldsymbol{u}(a) \neq 0$ for some $a \in [0,1]$. Hence $(\boldsymbol{u}(a))^2 = b > 0$. Since $\boldsymbol{u}$ is continuous, $\boldsymbol{u}^2$ is continuous. Hence, there exists a real number $\delta > 0$ such that $(\boldsymbol{u}(x))^2 > \frac{b}{2}$ when $|x - a| < \delta$ and $x \in [0,1]$. We may clearly assume that $\delta < \frac{1}{2}$. At least one of the intervals $[a, a+\delta]$ and $[a-\delta, a]$ is contained in the interval $[0,1]$. Therefore $\int_0^1 (\boldsymbol{u}(x))^2\,dx \geq \frac{\delta b}{2} > 0$. This shows that also Axiom (iii) holds.

Example 3.7. Let V be the linear space of Riemann integrable functions in $[0,1]$. Then $\langle \boldsymbol{u}, \boldsymbol{v} \rangle = \int_0^1 \boldsymbol{u}(x)\boldsymbol{v}(x)\,dx$ is not an inner product. Let, for example, $\boldsymbol{u}$ be the function defined by $\boldsymbol{u}(0) = 1$ and $\boldsymbol{u}(x) = 0$ for $x \in (0,1]$. Then $\langle \boldsymbol{u}, \boldsymbol{u} \rangle = 0$ despite the fact that $\boldsymbol{u}$ is not the zero vector.

Example 3.8. Let l_2 be the set of all infinite sequences $\boldsymbol{x} = (x_n)_{n=1}^\infty$ of real numbers for which $\sum_{n=1}^\infty x_n^2$ is convergent. Let $\boldsymbol{x} = (x_n)_{n=1}^\infty$ and $\boldsymbol{y} = (y_n)_{n=1}^\infty$ be two elements of l_2 and s a real number. We define $\boldsymbol{x} + \boldsymbol{y} = (x_n + y_n)_{n=1}^\infty$ and $s\boldsymbol{x} = (sx_n)_{n=1}^\infty$. Clearly $\sum_{n=1}^\infty (sx_n)^2$ is convergent, which shows that $s\boldsymbol{x} \in l_2$. Since

$$(x_n + y_n)^2 \leq (x_n + y_n)^2 + (x_n - y_n)^2 = 2x_n^2 + 2y_n^2,$$

it follows from the comparison theorem for positive series that $\sum_{n=1}^\infty (x_n + y_n)^2$ is convergent. Hence, the two operations turn l_2 into a linear space. Since $|x_n|^2 = x_n^2$, we have $(|x_n|)_{n=1}^\infty \in l_2$. Hence, using the fact that

$$|x_n y_n| = \frac{(|x_n| + |y_n|)^2 - (|x_n| - |y_n|)^2}{4},$$

we find that $\sum_{n=1}^\infty x_n y_n$ is absolutely convergent, and hence convergent. Therefore, we can make l_2 an inner product space by defining $\langle \boldsymbol{x}, \boldsymbol{y} \rangle = \sum_{n=1}^\infty x_n y_n$.

Definition 3.9. We say that two vectors $\boldsymbol{u}$ and $\boldsymbol{v}$ in an inner product space are orthogonal if $\langle \boldsymbol{u}, \boldsymbol{v} \rangle = 0$. The vector $\boldsymbol{e}$ is called a unit vector if $\|\boldsymbol{e}\| = 1$.

If $\boldsymbol{u} \neq \boldsymbol{0}$, we can form the vector $\boldsymbol{e} = \frac{1}{\|\boldsymbol{u}\|}\boldsymbol{u}$. Then $\|\boldsymbol{e}\| = \left|\frac{1}{\|\boldsymbol{u}\|}\right| \|\boldsymbol{u}\| = 1$. Hence $\boldsymbol{e}$ is a unit vector. We say that we normalise $\boldsymbol{u}$ to unit norm.

Theorem 3.10 (Pythagorean theorem). If $\boldsymbol{u}$ and $\boldsymbol{v}$ are orthogonal vectors in an inner product space, then $\|\boldsymbol{u}+\boldsymbol{v}\|^2 = \|\boldsymbol{u}\|^2 + \|\boldsymbol{v}\|^2$.

Proof. Since $\langle \boldsymbol{u}, \boldsymbol{v} \rangle = 0$, we have

$$\|\boldsymbol{u}+\boldsymbol{v}\|^2 = \langle \boldsymbol{u}+\boldsymbol{v}, \boldsymbol{u}+\boldsymbol{v} \rangle = \langle \boldsymbol{u}, \boldsymbol{u} \rangle + 2\langle \boldsymbol{u}, \boldsymbol{v} \rangle + \langle \boldsymbol{v}, \boldsymbol{v} \rangle = \|\boldsymbol{u}\|^2 + \|\boldsymbol{v}\|^2. \blacksquare$$

Theorem 3.11 (Cauchy–Schwarz inequality). If $\boldsymbol{u}$ and $\boldsymbol{v}$ are vectors in an inner product space, then $|\langle \boldsymbol{u}, \boldsymbol{v} \rangle| \leq \|\boldsymbol{u}\|\|\boldsymbol{v}\|$.

Proof. If $\boldsymbol{v} = \boldsymbol{0}$, then both sides are zero and the inequality holds. Assume that $\boldsymbol{v} \neq \boldsymbol{0}$. Then we have

$$\begin{aligned} 0 \leq \|\boldsymbol{u} - t\boldsymbol{v}\|^2 &= \langle \boldsymbol{u} - t\boldsymbol{v}, \boldsymbol{u} - t\boldsymbol{v} \rangle = \|\boldsymbol{u}\|^2 - 2t\langle \boldsymbol{u}, \boldsymbol{v} \rangle + t^2\|\boldsymbol{v}\|^2 \\ &= \|\boldsymbol{v}\|^2 \left(t - \frac{\langle \boldsymbol{u}, \boldsymbol{v} \rangle}{\|\boldsymbol{v}\|^2} \right)^2 + \|\boldsymbol{u}\|^2 - \frac{\langle \boldsymbol{u}, \boldsymbol{v} \rangle^2}{\|\boldsymbol{v}\|^2} \end{aligned}$$

for any real number t. For $t = \frac{\langle \boldsymbol{u}, \boldsymbol{v} \rangle}{\|\boldsymbol{v}\|^2}$, this yields

$$\|\boldsymbol{u}\|^2 - \frac{\langle \boldsymbol{u}, \boldsymbol{v} \rangle^2}{\|\boldsymbol{v}\|^2} \geq 0.$$

Consequently, $\langle \boldsymbol{u}, \boldsymbol{v} \rangle^2 \leq \|\boldsymbol{u}\|^2\|\boldsymbol{v}\|^2$, and hence $|\langle \boldsymbol{u}, \boldsymbol{v} \rangle| \leq \|\boldsymbol{u}\|\|\boldsymbol{v}\|$. $\blacksquare$

In ordinary three-space one starts out with the notions of length and angle and defines the inner product of two non-zero vectors $\boldsymbol{u}$ and $\boldsymbol{v}$ by $\langle \boldsymbol{u}, \boldsymbol{v} \rangle = \|\boldsymbol{u}\|\|\boldsymbol{v}\| \cos\theta$. In this general setting we started out with an inner product and defined the length, or norm as we also call it. When $\boldsymbol{u}$ and $\boldsymbol{v}$ are non-zero vectors, the Cauchy–Schwarz inequality can be written as

$$-1 \leq \frac{\langle \boldsymbol{u}, \boldsymbol{v} \rangle}{\|\boldsymbol{u}\|\|\boldsymbol{v}\|} \leq 1.$$

This enables us to complete the situation by also defining angles.

Definition 3.12. Let $\boldsymbol{u}$ and $\boldsymbol{v}$ be non-zero vectors of an inner product space. The angle between $\boldsymbol{u}$ and $\boldsymbol{v}$ is the unique real number θ for which

$$\cos\theta = \frac{\langle \boldsymbol{u}, \boldsymbol{v} \rangle}{\|\boldsymbol{u}\|\|\boldsymbol{v}\|}, \quad 0 \leq \theta \leq \pi.$$

The inner product on ordinary three-space is an inner product in the sense of this chapter. By defining angles as we do here, we see that we get our old angles back in three-space.

Example 3.13. The lengths of the vectors $\boldsymbol{u} = (4, 3, -1, -1)$ and $\boldsymbol{v} = (1, 1, -1, -1)$ in $\mathbf{R}^4$ are

$$\|\boldsymbol{u}\| = \sqrt{4^2 + 3^2 + (-1)^2 + (-1)^2} = \sqrt{27} = 3\sqrt{3},$$
$$\|\boldsymbol{v}\| = \sqrt{1^2 + 1^2 + (-1)^2 + (-1)^2} = \sqrt{4} = 2.$$

Their inner product is

$$\langle \boldsymbol{u}, \boldsymbol{v} \rangle = 4 \cdot 1 + 3 \cdot 1 + (-1)(-1) + (-1)(-1) = 9.$$

Hence, the angle θ between the two vectors is given by

$$\cos\theta = \frac{\langle \boldsymbol{u}, \boldsymbol{v} \rangle}{\|\boldsymbol{u}\|\|\boldsymbol{v}\|} = \frac{9}{(3\sqrt{3}) \cdot 2} = \frac{\sqrt{3}}{2},$$

and is $\frac{\pi}{6}$.

Example 3.14. Consider the functions $\boldsymbol{u}$ and $\boldsymbol{v}$ in $C[0, 1]$ defined by $\boldsymbol{u}(x) = 1$ and $\boldsymbol{v}(x) = 6x - 2$. With the inner product $\langle \boldsymbol{u}, \boldsymbol{v} \rangle = \int_0^1 \boldsymbol{u}(x)\boldsymbol{v}(x)\, dx$ we have

$$\|\boldsymbol{u}\|^2 = \int_0^1 (\boldsymbol{u}(x))^2\, dx = \int_0^1 1\, dx = 1,$$
$$\|\boldsymbol{v}\|^2 = \int_0^1 (\boldsymbol{v}(x))^2\, dx = \int_0^1 (36x^2 - 24x + 4)\, dx = 4,$$
$$\langle \boldsymbol{u}, \boldsymbol{v} \rangle = \int_0^1 \boldsymbol{u}(x)\boldsymbol{v}(x)\, dx = \int_0^1 (6x - 2)\, dx = 1.$$

Consequently, the angle θ between $\boldsymbol{u}$ and $\boldsymbol{v}$ is given by

$$\cos\theta = \frac{1}{1 \cdot 2} = \frac{1}{2},$$

and is therefore $\frac{\pi}{3}$.

You cannot find the angle in the example in a figure depicting the graphs of the two functions. Instead, it is small if the functions are close to being directly proportional and large if they are close to being inversely proportional.

Theorem 3.15 (Triangle inequality). Let $\boldsymbol{u}$ and $\boldsymbol{v}$ be vectors in an inner product space. Then $\|\boldsymbol{u} + \boldsymbol{v}\| \le \|\boldsymbol{u}\| + \|\boldsymbol{v}\|$.

Proof. The Cauchy–Schwarz inequality gives that

$$\begin{aligned}\|\boldsymbol{u} + \boldsymbol{v}\|^2 = \langle \boldsymbol{u} + \boldsymbol{v}, \boldsymbol{u} + \boldsymbol{v} \rangle &= \|\boldsymbol{u}\|^2 + \|\boldsymbol{v}\|^2 + 2\langle \boldsymbol{u}, \boldsymbol{v} \rangle \le \|\boldsymbol{u}\|^2 + \|\boldsymbol{v}\|^2 + 2\|\boldsymbol{u}\|\|\boldsymbol{v}\| \\ &= (\|\boldsymbol{u}\| + \|\boldsymbol{v}\|)^2.\end{aligned}$$

The inequality now follows from the fact that both $\|\boldsymbol{u} + \boldsymbol{v}\|$ and $\|\boldsymbol{u}\| + \|\boldsymbol{v}\|$ are non-negative. ■

3.2 Orthonormal Bases

Theorem 3.16. If the vectors $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_k$ are pairwise orthogonal and non-zero, then they are linearly independent.

Proof. If $s_1\boldsymbol{u}_1 + \cdots + s_k\boldsymbol{u}_k = \boldsymbol{0}$, then

$$\begin{aligned} 0 = \langle \boldsymbol{0}, \boldsymbol{u}_i \rangle &= \langle s_1\boldsymbol{u}_1 + \cdots + s_i\boldsymbol{u}_i + \cdots + s_k\boldsymbol{u}_k, \boldsymbol{u}_i \rangle \\ &= s_1\langle \boldsymbol{u}_1, \boldsymbol{u}_i \rangle + \cdots + s_i\langle \boldsymbol{u}_i, \boldsymbol{u}_i \rangle + \cdots + s_k\langle \boldsymbol{u}_k, \boldsymbol{u}_i \rangle = s_i\langle \boldsymbol{u}_i, \boldsymbol{u}_i \rangle. \end{aligned}$$

Since $\boldsymbol{u}_i \neq \boldsymbol{0}$, we have $\langle \boldsymbol{u}_i, \boldsymbol{u}_i \rangle = \|\boldsymbol{u}_i\|^2 \neq 0$, and hence $s_i = 0$. ∎

Definition 3.17. Let V be an inner product space. We say that the vectors $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_k$ in V form an orthonormal set if

$$\langle \boldsymbol{e}_i, \boldsymbol{e}_j \rangle = \begin{cases} 1 & \text{when } i = j, \\ 0 & \text{when } i \neq j. \end{cases}$$

If the vectors also form a basis for V, we say that they form an orthonormal basis for V.

By Theorem 3.16, the vectors of an orthonormal set form a basis for V if and only if they span V.

Theorem 3.18. Let V be an inner product space with orthonormal basis $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n$ and assume that the coordinates of $\boldsymbol{u}$ and $\boldsymbol{v}$ with respect to this basis are $\boldsymbol{x} = (x_1, \ldots, x_n)$ and $\boldsymbol{y} = (y_1, \ldots, y_n)$, respectively. Then $\langle \boldsymbol{u}, \boldsymbol{v} \rangle = \boldsymbol{x} \cdot \boldsymbol{y}$ and $\|\boldsymbol{u}\|^2 = \|\boldsymbol{x}\|^2$ where the last norm is the ordinary norm in $\mathbf{R}^n$,

Proof. We have

$$\langle \boldsymbol{u}, \boldsymbol{v} \rangle = \langle x_1\boldsymbol{e}_1 + \cdots + x_n\boldsymbol{e}_n, y_1\boldsymbol{e}_1 + \cdots + y_n\boldsymbol{e}_n \rangle = \sum_{i=1}^{n}\sum_{j=1}^{n} x_i y_j \langle \boldsymbol{e}_i, \boldsymbol{e}_j \rangle = \sum_{i=1}^{n} x_i y_i = \boldsymbol{x} \cdot \boldsymbol{y},$$

and from this it follows that $\|\boldsymbol{u}\|^2 = \langle \boldsymbol{u}, \boldsymbol{u} \rangle = \boldsymbol{x} \cdot \boldsymbol{x} = \|\boldsymbol{x}\|^2$. ∎

On page 23, we remarked that a non-zero n-dimensional linear space V can be identified with the linear space $\mathbf{R}^n$ by means of a basis for V. This theorem allows us to identify a non-zero n-dimensional inner product space V with the inner product space $\mathbf{R}^n$ by means of an orthonormal basis for V. As will be seen in Corollary 3.22, every non-zero finite-dimensional inner product space has an orthonormal basis.

Theorem 3.19. If $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n$ form an orthonormal basis for an inner product space V, then the coordinates of a vector $\boldsymbol{u}$ are given by $x_i = \langle \boldsymbol{u}, \boldsymbol{e}_i \rangle$, $i = 1, \ldots, n$.

Proof. If $\boldsymbol{u} = x_1\boldsymbol{e}_1 + \cdots + x_n\boldsymbol{e}_n$, then

$$\langle \boldsymbol{u}, \boldsymbol{e}_i \rangle = \langle x_1\boldsymbol{e}_1 + \cdots + x_n\boldsymbol{e}_n, \boldsymbol{e}_i \rangle = x_1\langle \boldsymbol{e}_1, \boldsymbol{e}_i \rangle + \cdots + x_i\langle \boldsymbol{e}_i, \boldsymbol{e}_i \rangle + \cdots + x_n\langle \boldsymbol{e}_n, \boldsymbol{e}_i \rangle = x_i. \blacksquare$$

Theorem 3.20. Let $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_k$ be pairwise orthogonal non-zero vectors and let $\boldsymbol{v}$ be a vector in an inner product space. Then there exist unique numbers $s_1, \ldots, s_k$ such that

$$\boldsymbol{u} = s_1\boldsymbol{u}_1 + \cdots + s_k\boldsymbol{u}_k + \boldsymbol{v}$$

is orthogonal to the vectors $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_k$. The s_i are given by

$$s_i = -\frac{\langle \boldsymbol{v}, \boldsymbol{u}_i \rangle}{\|\boldsymbol{u}_i\|^2}.$$

Proof. The vector $\boldsymbol{u} = s_1\boldsymbol{u}_1 + \cdots + s_k\boldsymbol{u}_k + \boldsymbol{v}$ is orthogonal to $\boldsymbol{u}_i$ if and only if

$$0 = \langle \boldsymbol{u}, \boldsymbol{u}_i \rangle = s_i \langle \boldsymbol{u}_i, \boldsymbol{u}_i \rangle + \langle \boldsymbol{v}, \boldsymbol{u}_i \rangle$$

and this is in turn equivalent to

$$s_i = -\frac{\langle \boldsymbol{v}, \boldsymbol{u}_i \rangle}{\langle \boldsymbol{u}_i, \boldsymbol{u}_i \rangle} = -\frac{\langle \boldsymbol{v}, \boldsymbol{u}_i \rangle}{\|\boldsymbol{u}_i\|^2}. \blacksquare$$

Assume that the vectors $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n$ form a basis for an inner product space V. These vectors can then be used to construct an orthonormal basis for V. To do so, we first construct a basis for V consisting of pairwise orthogonal vectors $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_n$. First, set

$$\boldsymbol{u}_1 = \boldsymbol{v}_1.$$

If $n = 1$, we are done. Otherwise, by using Theorem 3.20, we can find a number s_{12} such that

$$\boldsymbol{u}_2 = s_{12}\boldsymbol{u}_1 + \boldsymbol{v}_2$$

is orthogonal to $\boldsymbol{u}_1$. Then $\boldsymbol{u}_2$ must be non-zero, for if $\boldsymbol{u}_2 = \boldsymbol{0}$, then $\boldsymbol{v}_1$ and $\boldsymbol{v}_2$ would be linearly dependent. If $n > 2$, we find numbers s_{13} and s_{23} such that

$$\boldsymbol{u}_3 = s_{13}\boldsymbol{u}_1 + s_{23}\boldsymbol{u}_2 + \boldsymbol{v}_3$$

is orthogonal to $\boldsymbol{u}_1$ and $\boldsymbol{u}_2$. Since $\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3$ are linearly independent, $\boldsymbol{u}_3 \neq \boldsymbol{0}$. When $k-1$ vectors $\boldsymbol{u}_i$ are constructed, and $k \leq n$, we form the next vector $\boldsymbol{u}_k$ by requiring that

$$\boldsymbol{u}_k = s_{1k}\boldsymbol{u}_1 + s_{2k}\boldsymbol{u}_2 + \cdots + s_{k-1k}\boldsymbol{u}_{k-1} + \boldsymbol{v}_k$$

be orthogonal to $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_{k-1}$. Owing to the linear independence of the $\boldsymbol{v}_i$, $\boldsymbol{u}_k$ must be non-zero. In this way we get n pairwise orthogonal non-zero vectors $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_n$. Next, we normalise these vectors by setting

$$\boldsymbol{e}_i = \frac{1}{\|\boldsymbol{u}_i\|}\boldsymbol{u}_i, \quad i = 1, \ldots, n.$$

By Theorem 3.16, the $\boldsymbol{e}_i$ are linearly independent, and since the $\boldsymbol{v}_i$ are linear combinations of the $\boldsymbol{e}_i$ and span V, the vectors $\boldsymbol{e}_i$ also span V. Hence, $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n$ form an orthonormal basis for V. We could instead have argued that the $\boldsymbol{e}_i$ are n linearly independent vectors in the n-dimensional space V. This algorithm for finding an orthonormal basis is called the Gram–Schmidt orthogonalisation process and can be summarised as follows.

Theorem 3.21. If $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n$ form a basis for an inner product space V, and

$$\begin{aligned}
\boldsymbol{u}_1 &= \boldsymbol{v}_1 \\
\boldsymbol{u}_2 &= s_{12}\boldsymbol{u}_1 + \boldsymbol{v}_2 \\
\boldsymbol{u}_3 &= s_{13}\boldsymbol{u}_1 + s_{23}\boldsymbol{u}_2 + \boldsymbol{v}_3 \\
&\vdots \\
\boldsymbol{u}_n &= s_{1n}\boldsymbol{u}_1 + s_{2n}\boldsymbol{u}_2 + \cdots + s_{n-1n}\boldsymbol{u}_{n-1} + \boldsymbol{v}_n
\end{aligned}$$

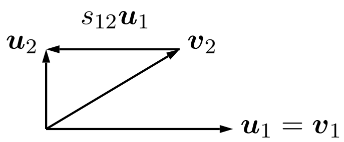
where

$$s_{ik} = -\frac{\langle \boldsymbol{u}_i, \boldsymbol{v}_k \rangle}{\|\boldsymbol{u}_i\|^2}, \quad 1 \le i < k \le n,$$

then the vectors

$$\boldsymbol{e}_i = \frac{1}{\|\boldsymbol{u}_i\|}\boldsymbol{u}_i, \quad i = 1, \ldots, n,$$

form an orthonormal basis for V.



Corollary 3.22. Every non-zero finite-dimensional inner product space has an orthonormal basis.

Example 3.23. We demonstrate the process by finding an orthonormal basis for the subspace of $\mathbf{R}^4$ spanned by $\boldsymbol{v}_1 = (1,1,1,1)$, $\boldsymbol{v}_2 = (1,2,2,1)$, $\boldsymbol{v}_3 = (2,3,1,6)$. We set $\boldsymbol{u}_1 = \boldsymbol{v}_1$. Then we determine the number r so that

$$\boldsymbol{u}_2 = r\boldsymbol{u}_1 + \boldsymbol{v}_2$$

is orthogonal to $\boldsymbol{u}_1$. By multiplying by $\boldsymbol{u}_1$, we get

$$\begin{aligned} 0 &= \langle \boldsymbol{u}_1, \boldsymbol{u}_2 \rangle = r\langle \boldsymbol{u}_1, \boldsymbol{u}_1 \rangle + \langle \boldsymbol{u}_1, \boldsymbol{v}_2 \rangle \\ &= (1 \cdot 1 + 1 \cdot 1 + 1 \cdot 1 + 1 \cdot 1)r + 1 \cdot 1 + 1 \cdot 2 + 1 \cdot 2 + 1 \cdot 1 = 4r + 6. \end{aligned}$$

Hence, $r = -\frac{3}{2}$, and we get

$$\boldsymbol{u}_2 = r\boldsymbol{u}_1 + \boldsymbol{v}_2 = -\tfrac{3}{2}(1,1,1,1) + (1,2,2,1) = \tfrac{1}{2}(-1,1,1,-1).$$

To avoid fractional numbers, we can replace $\boldsymbol{u}_2$ with $2\boldsymbol{u}_2$. This works, because also $\boldsymbol{u}_1$ and $2\boldsymbol{u}_2$ are orthogonal, and $2\boldsymbol{u}_2$ is a linear combination of $\boldsymbol{v}_1$ and $\boldsymbol{v}_2$. Hence, we set $\boldsymbol{u}_2 = (-1,1,1,-1)$. Next, we set out to find numbers s and t that make

$$\boldsymbol{u}_3 = s\boldsymbol{u}_1 + t\boldsymbol{u}_2 + \boldsymbol{v}_3$$

orthogonal to $\boldsymbol{u}_1$ and $\boldsymbol{u}_2$. Multiplying by $\boldsymbol{u}_1$, we get

$$\begin{aligned} 0 &= \langle \boldsymbol{u}_1, \boldsymbol{u}_3 \rangle = s\langle \boldsymbol{u}_1, \boldsymbol{u}_1 \rangle + t\langle \boldsymbol{u}_1, \boldsymbol{u}_2 \rangle + \langle \boldsymbol{u}_1, \boldsymbol{v}_3 \rangle \\ &= (1 \cdot 1 + 1 \cdot 1 + 1 \cdot 1 + 1 \cdot 1)s + 1 \cdot 2 + 1 \cdot 3 + 1 \cdot 1 + 1 \cdot 6 = 4s + 12, \end{aligned}$$

since $\langle \boldsymbol{u}_1, \boldsymbol{u}_2 \rangle = 0$. We get $s = -3$. Next, we multiply by $\boldsymbol{u}_2$ and get

$$\begin{aligned} 0 &= \langle \boldsymbol{u}_2, \boldsymbol{u}_3 \rangle = s\langle \boldsymbol{u}_2, \boldsymbol{u}_1 \rangle + t\langle \boldsymbol{u}_2, \boldsymbol{u}_2 \rangle + \langle \boldsymbol{u}_2, \boldsymbol{v}_3 \rangle \\ &= (-1 \cdot (-1) + 1 \cdot 1 + 1 \cdot 1 - 1 \cdot (-1))t - 1 \cdot 2 + 1 \cdot 3 + 1 \cdot 1 - 1 \cdot 6 = 4t - 4. \end{aligned}$$

Hence, $t = 1$, and we get

$$\boldsymbol{u}_3 = -3(1,1,1,1) + (-1,1,1,-1) + (2,3,1,6) = (-2,1,-1,2).$$

Now it only remains to normalise the vectors. Since $\|\boldsymbol{u}_1\| = 2$, $\|\boldsymbol{u}_2\| = 2$ and $\|\boldsymbol{u}_3\| = \sqrt{10}$, we obtain the orthonormal basis

$$\boldsymbol{e}_1 = \frac{1}{2}(1,1,1,1), \quad \boldsymbol{e}_2 = \frac{1}{2}(-1,1,1,-1), \quad \boldsymbol{e}_3 = \frac{1}{\sqrt{10}}(-2,1,-1,2).$$

Instead of deriving this result from scratch, we could of course have used the formulae of Theorem 3.21. Methods are, however, easier to remember than formulae.

The Gram–Schmidt process works also if the $\boldsymbol{v}_i$ merely span V. As long as the $\boldsymbol{u}_i$ constructed so far are non-zero, they are linearly independent and span the same subspace as the corresponding vectors $\boldsymbol{v}_i$. If $\boldsymbol{u}_i = \boldsymbol{0}$, then either $i = 1$ and $\boldsymbol{v}_1 = \boldsymbol{0}$ or $\boldsymbol{v}_i$ is a linear combination of $\boldsymbol{v}_1, \dots, \boldsymbol{v}_{i-1}$. Hence, $\boldsymbol{u}_i$ and $\boldsymbol{v}_i$ can be discarded. One can then repeat the last step using the next vector and proceed from there.

Definition 3.24. We say that an $n \times n$ matrix is orthogonal if its columns form an orthonormal set in $\mathbf{R}^n$.

The orthogonality means that $A^tA = I$. Since $A^tA = I$ if and only if $AA^t = I$, we get the following theorem.

Theorem 3.25. Let A be a square matrix. Then the following statements are equivalent.

(i) A is orthogonal.
(ii) $A^tA = I$.
(iii) $AA^t = I$.
(iv) The rows of A form an orthonormal set.

3.3 Orthogonal Complement

Definition 3.26. Let S be a subset of an inner product space V. The orthogonal complement of S is the set $S^\perp = \{\boldsymbol{v} \in V;\ \langle \boldsymbol{u}, \boldsymbol{v} \rangle = 0 \text{ for all } \boldsymbol{u} \in S\}$.

Theorem 3.27. If S is a subset of an inner product space V, then $S^\perp$ is a subspace of V.

Proof. Clearly, $\boldsymbol{0} \in S^\perp$, and hence $S^\perp \neq \varnothing$. Suppose that $\boldsymbol{v}$ and $\boldsymbol{w}$ belong to $S^\perp$. Then $\langle \boldsymbol{u}, \boldsymbol{v} \rangle = \langle \boldsymbol{u}, \boldsymbol{w} \rangle = 0$ for all $\boldsymbol{u} \in S$. Therefore, $\langle \boldsymbol{u}, s\boldsymbol{v} + t\boldsymbol{w} \rangle = s\langle \boldsymbol{u}, \boldsymbol{v} \rangle + t\langle \boldsymbol{u}, \boldsymbol{w} \rangle = 0$ for all $\boldsymbol{u} \in S$, which means that $s\boldsymbol{v} + t\boldsymbol{w} \in S^\perp$. ■

Lemma 3.28. Let $\boldsymbol{u}_1, \dots, \boldsymbol{u}_k$ be vectors of an inner product space. Then

$$[\boldsymbol{u}_1, \dots, \boldsymbol{u}_k]^\perp = \{\boldsymbol{u}_1, \dots, \boldsymbol{u}_k\}^\perp.$$

Proof. Since $\{\boldsymbol{u}_1, \ldots, \boldsymbol{u}_k\} \subseteq [\boldsymbol{u}_1, \ldots, \boldsymbol{u}_k]$, the inclusion $[\boldsymbol{u}_1, \ldots, \boldsymbol{u}_k]^\perp \subseteq \{\boldsymbol{u}_1, \ldots, \boldsymbol{u}_k\}^\perp$ is clear. Assume that $\boldsymbol{v} \in \{\boldsymbol{u}_1, \ldots, \boldsymbol{u}_k\}^\perp$. If $\boldsymbol{u} \in [\boldsymbol{u}_1, \ldots, \boldsymbol{u}_k]$, then $\boldsymbol{u} = s_1\boldsymbol{u}_1 + \cdots + s_k\boldsymbol{u}_k$, and hence $\langle \boldsymbol{u}, \boldsymbol{v}\rangle = s_1\langle \boldsymbol{u}_1, \boldsymbol{v}\rangle + \cdots + s_k\langle \boldsymbol{u}_k, \boldsymbol{v}\rangle = 0$. This shows that $\boldsymbol{v} \in [\boldsymbol{u}_1, \ldots, \boldsymbol{u}_k]^\perp$, and therefore also the reverse inclusion holds. ■

Theorem 3.29. Let U be a finite-dimensional subspace of an inner product space V. Then $V = U \oplus U^\perp$.

Proof. If $U = \{\boldsymbol{0}\}$, then $U^\perp = V$, and a vector $\boldsymbol{u} \in V$ can be written as $\boldsymbol{u} = \boldsymbol{0} + \boldsymbol{u}$. Assume that $U \neq \{\boldsymbol{0}\}$. Then, by Corollary 3.22, there exists an orthonormal basis $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n$ for U. We define $\boldsymbol{u}'$ and $\boldsymbol{u}''$ by

$$\boldsymbol{u}' = \langle \boldsymbol{u}, \boldsymbol{e}_1\rangle \boldsymbol{e}_1 + \cdots + \langle \boldsymbol{u}, \boldsymbol{e}_n\rangle \boldsymbol{e}_n, \quad \boldsymbol{u}'' = \boldsymbol{u} - \boldsymbol{u}'.$$

Then $\boldsymbol{u}' \in U$ and $\boldsymbol{u} = \boldsymbol{u}' + \boldsymbol{u}''$. By using Lemma 3.28 and the fact that

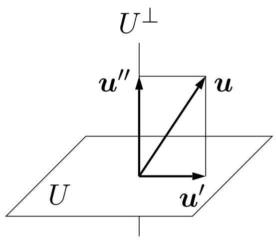
$$\langle \boldsymbol{e}_i, \boldsymbol{u}''\rangle = \langle \boldsymbol{e}_i, \boldsymbol{u}\rangle - \langle \boldsymbol{e}_i, \langle \boldsymbol{u}, \boldsymbol{e}_1\rangle \boldsymbol{e}_1 + \cdots + \langle \boldsymbol{u}, \boldsymbol{e}_n\rangle \boldsymbol{e}_n\rangle = \langle \boldsymbol{e}_i, \boldsymbol{u}\rangle - \langle \boldsymbol{e}_i, \boldsymbol{u}\rangle = 0, \quad i = 1, \ldots, n,$$

we also see that $\boldsymbol{u}'' \in U^\perp$. If $\boldsymbol{u} \in U \cap U^\perp$, then $\boldsymbol{u}$ is orthogonal to itself, and hence $\|\boldsymbol{u}\|^2 = \langle \boldsymbol{u}, \boldsymbol{u}\rangle = 0$. This shows that $U \cap U^\perp = \{\boldsymbol{0}\}$. The conclusion of this theorem now follows from Theorem 2.61. ■

Theorem 3.30. Let U be a subspace of an inner product space V. If $V = U \oplus U^\perp$, then $(U^\perp)^\perp = U$.

Proof. Let $\boldsymbol{u} \in U$. Then, by the definition of $U^\perp$, $\langle \boldsymbol{u}, \boldsymbol{v}\rangle = 0$ for every $\boldsymbol{v} \in U^\perp$. Hence $\boldsymbol{u} \in (U^\perp)^\perp$. This shows that $U \subseteq (U^\perp)^\perp$. Assume that $\boldsymbol{u} \in (U^\perp)^\perp$. By the assumption that $V = U \oplus U^\perp$, we can write $\boldsymbol{u} = \boldsymbol{u}' + \boldsymbol{u}''$ where $\boldsymbol{u}' \in U$ and $\boldsymbol{u}'' \in U^\perp$. The vector $\boldsymbol{u}''$ is orthogonal to both $\boldsymbol{u}$ and $\boldsymbol{u}'$. It follows that $\langle \boldsymbol{u}'', \boldsymbol{u}''\rangle = \langle \boldsymbol{u}'', \boldsymbol{u} - \boldsymbol{u}'\rangle = 0$, which shows that $\boldsymbol{u}'' = \boldsymbol{0}$. Hence $\boldsymbol{u} = \boldsymbol{u}' \in U$. This shows the reverse inclusion $(U^\perp)^\perp \subseteq U$. Hence $(U^\perp)^\perp = U$. ■

Definition 3.31. Assume that $V = U \oplus U^\perp$ and let $\boldsymbol{u} \in V$. The unique vectors $\boldsymbol{u}' \in U$ and $\boldsymbol{u}'' \in U^\perp$ for which $\boldsymbol{u} = \boldsymbol{u}' + \boldsymbol{u}''$ are called the orthogonal projections of $\boldsymbol{u}$ on U and $U^\perp$, respectively.



Note that the conclusion $V = U \oplus U^\perp$ of Theorem 3.29 need not be true if we drop the assumption that U be finite-dimensional, and if $V \neq U \oplus U^\perp$, the conclusion of

Theorem 3.30 need not be true. Consider the space l_2 defined in Example 3.8 and let U be the subspace of l_2 consisting of those sequences $(x_n)_{n=1}^{\infty}$ for which only a finite number of components are non-zero. Then the vector $\boldsymbol{\varepsilon}_n$ having a one in position n and zeros elsewhere belongs to U. If $\boldsymbol{x} \in U^{\perp}$, then $x_n = \langle \boldsymbol{x}, \boldsymbol{\varepsilon}_n \rangle = 0$ for all n, and hence $\boldsymbol{x} = \mathbf{0}$. This means that $U^{\perp} = \{\mathbf{0}\}$, whence $(U^{\perp})^{\perp} = \{\mathbf{0}\}^{\perp} = l_2 \neq U$.

Let U be a finite-dimensional subspace of an inner product space V and assume that $\boldsymbol{e}_1, \dots, \boldsymbol{e}_n$ form an orthonormal basis for U . Then the proof of Theorem 3.29 shows that the orthogonal projection $\boldsymbol{u}'$ on U of a vector $\boldsymbol{u} \in V$ is given by

$$\boldsymbol{u}' = \langle \boldsymbol{u}, \boldsymbol{e}_1 \rangle \boldsymbol{e}_1 + \cdots + \langle \boldsymbol{u}, \boldsymbol{e}_n \rangle \boldsymbol{e}_n. \tag{3.1}$$

Example 3.32. Let U be the subspace of $\mathbf{R}^4$ spanned by the vectors $\boldsymbol{u}_1 = (3, 0, 4, 0)$ and $\boldsymbol{u}_2 = (0, 3, 0, -4)$. We set out to find the orthogonal projection of $\boldsymbol{u} = (1, 1, -1, -1)$ on U. We note that the vectors $\boldsymbol{u}_1$ and $\boldsymbol{u}_2$ are orthogonal to each other. Hence, we can obtain an orthonormal basis for U by normalising these vectors to

$$\boldsymbol{e}_1 = \frac{1}{5}(3, 0, 4, 0), \quad \boldsymbol{e}_2 = \frac{1}{5}(0, 3, 0, -4).$$

Using the above formula, we get

$$\begin{aligned} \boldsymbol{u}' &= \langle \boldsymbol{u}, \boldsymbol{e}_1 \rangle \boldsymbol{e}_1 + \langle \boldsymbol{u}, \boldsymbol{e}_2 \rangle \boldsymbol{e}_2 = -\frac{1}{5}\boldsymbol{e}_1 + \frac{7}{5}\boldsymbol{e}_2 = -\frac{1}{25}(3, 0, 4, 0) + \frac{7}{25}(0, 3, 0, -4) \\ &= \frac{1}{25}(-3, 21, -4, -28). \end{aligned}$$

Example 3.33. This time we want to find the orthogonal projection of $\boldsymbol{u} = (1, 2, 1, 2)$ on the subspace U of $\mathbf{R}^4$ spanned by $\boldsymbol{v}_1 = (1, 1, 1, 1)$ and $\boldsymbol{v}_2 = (1, 1, -1, 3)$. We begin by applying the Gram–Schmidt process to $\boldsymbol{v}_1$ and $\boldsymbol{v}_2$. We set $\boldsymbol{u}_1 = \boldsymbol{v}_1$ and $\boldsymbol{u}_2 = s\boldsymbol{u}_1 + \boldsymbol{v}_2$, and we see that $\langle \boldsymbol{u}_1, \boldsymbol{u}_2 \rangle = 0$ if $s = -1$. Hence, we can choose $\boldsymbol{u}_1 = (1, 1, 1, 1)$ and $\boldsymbol{u}_2 = (0, 0, -2, 2)$. By normalising these vectors to

$$\boldsymbol{e}_1 = \frac{1}{2}(1, 1, 1, 1), \quad \boldsymbol{e}_2 = \frac{1}{\sqrt{2}}(0, 0, -1, 1),$$

we get an orthonormal basis for U. Hence,

$$\boldsymbol{u}' = 3\boldsymbol{e}_1 + \frac{1}{\sqrt{2}}\boldsymbol{e}_2 = \frac{3}{2}(1, 1, 1, 1) + \frac{1}{2}(0, 0, -1, 1) = \frac{1}{2}(3, 3, 2, 4).$$

Example 3.34. Here, we want to find the orthogonal projection of $\boldsymbol{u} = (5, 7, 3)$ on the plane U defined by $x_1 + 2x_2 + 3x_3 = 0$. One way of solving this problem would be to first find a basis $\boldsymbol{v}_1, \boldsymbol{v}_2$ for U, then apply the Gram–Schmidt process to $\boldsymbol{v}_1$ and $\boldsymbol{v}_2$ to get an orthonormal basis $\boldsymbol{e}_1, \boldsymbol{e}_2$ for U and finally use the formula.

To avoid this rather cumbersome procedure, we use the fact that $U^{\perp}$ is spanned by the single vector $(1, 2, 3)$. Hence

$$\boldsymbol{e} = \frac{1}{\sqrt{14}}(1, 2, 3)$$

constitutes an orthonormal basis for $U^\perp$. By using the formula, we find that the orthogonal projection of $\boldsymbol{u}$ on $U^\perp$ is

$$\boldsymbol{u}'' = \langle \boldsymbol{u}, \boldsymbol{e} \rangle \boldsymbol{e} = \frac{28}{\sqrt{14}}\,\boldsymbol{e} = (2,4,6).$$

The orthogonal projection of $\boldsymbol{u}$ on U is, therefore,

$$\boldsymbol{u}' = \boldsymbol{u} - \boldsymbol{u}'' = (5,7,3) - (2,4,6) = (3,3,-3).$$

As we saw in this example, it may be worthwhile to give some thought to which of the vectors $\boldsymbol{u}'$ and $\boldsymbol{u}''$ should be computed first.

Now assume that $V = U \oplus U^\perp$. Let $\boldsymbol{u} \in V$ and let $\boldsymbol{w}$ be any vector in U. Then $\boldsymbol{u}' - \boldsymbol{w}$ is orthogonal to $\boldsymbol{u}''$. Hence, by the Pythagorean theorem,

$$\|\boldsymbol{u} - \boldsymbol{w}\|^2 = \|\boldsymbol{u}' - \boldsymbol{w} + \boldsymbol{u}''\|^2 = \|\boldsymbol{u}' - \boldsymbol{w}\|^2 + \|\boldsymbol{u}''\|^2.$$

Therefore, $\|\boldsymbol{u} - \boldsymbol{w}\|$ is as small as possible when $\boldsymbol{w} = \boldsymbol{u}'$, and then $\|\boldsymbol{u} - \boldsymbol{w}\| = \|\boldsymbol{u}''\|$. Since $\|\boldsymbol{u} - \boldsymbol{w}\| > \|\boldsymbol{u} - \boldsymbol{u}'\|$ when $\boldsymbol{w} \in U$ and $\boldsymbol{w} \neq \boldsymbol{u}'$, $\boldsymbol{u}'$ is the unique closest vector to $\boldsymbol{u}$ in U. We call $\|\boldsymbol{u}''\|$ the distance from $\boldsymbol{u}$ to U.

Example 3.35. Consider once again the subspace U and the vector $\boldsymbol{u}$ in Example 3.33. We found there that the vector closest to $\boldsymbol{u}$ in U is $\boldsymbol{u}' = \frac{1}{2}(3,3,2,4)$. The distance from $\boldsymbol{u}$ to U is

$$\|\boldsymbol{u}''\| = \|\boldsymbol{u} - \boldsymbol{u}'\| = \left\|(1,2,1,2) - \frac{1}{2}(3,3,2,4)\right\| = \left\|\frac{1}{2}(-1,1,0,0)\right\| = \frac{1}{\sqrt{2}}.$$

Example 3.36. Consider the plane U defined by $ax + by + cz = 0$ and let $\boldsymbol{u} = (x,y,z)$ be a vector in $\mathbf{R}^3$. Then

$$\boldsymbol{e} = \frac{1}{\sqrt{a^2+b^2+c^2}}(a,b,c)$$

forms an orthonormal basis for $U^\perp$. Hence,

$$\boldsymbol{u}'' = \langle \boldsymbol{u}, \boldsymbol{e} \rangle \boldsymbol{e} = \frac{ax+by+cz}{\sqrt{a^2+b^2+c^2}}\,\boldsymbol{e}.$$

Consequently, the distance from $\boldsymbol{u}$ to the plane is

$$\|\boldsymbol{u}''\| = \left|\frac{ax+by+cz}{\sqrt{a^2+b^2+c^2}}\right| \|\boldsymbol{e}\| = \frac{|ax+by+cz|}{\sqrt{a^2+b^2+c^2}}.$$

Theorem 3.37. Let V be a finite-dimensional inner product space and let U be a subspace of V. Then $\dim U + \dim U^\perp = \dim V$.

Proof. The statement follows directly from Theorems 3.29 and 2.62. ∎

3.4 The Rank of a Matrix

Let A be an $m \times n$ matrix and denote its rows by $R_1, \ldots, R_m$. Since the rows of A are the columns of A^t, we have $\operatorname{im} A^t = [R_1, \ldots, R_m]$. By definition, $\boldsymbol{x} \in \ker A$ if and only if $A\boldsymbol{x} = \mathbf{0}$. Clearly, $A\boldsymbol{x} = \mathbf{0}$ means that $\boldsymbol{x}$ is orthogonal to all the rows R_i of A. By Lemma 3.28, this is equivalent to $\boldsymbol{x}$ being orthogonal to all vectors of $\operatorname{im} A^t = [R_1, \ldots, R_m]$, which by definition is equivalent to $\boldsymbol{x} \in (\operatorname{im} A^t)^\perp$. Thus, we have shown that

$$\ker A = (\operatorname{im} A^t)^\perp. \tag{3.2}$$

Replacing A with A^t and using the fact that $(A^t)^t = A$, we also get

$$\ker A^t = (\operatorname{im} A)^\perp. \tag{3.3}$$

By applying the rank-nullity theorem to A, Theorem 3.37 to $\mathbf{R}^n = \operatorname{im} A^t \oplus (\operatorname{im} A^t)^\perp$ and using (3.2), we get

$$\begin{aligned} \dim \operatorname{im} A &= n - \dim \ker A = n - \dim (\operatorname{im} A^t)^\perp, \\ \dim \operatorname{im} A^t &= n - \dim (\operatorname{im} A^t)^\perp. \end{aligned}$$

Hence, $\dim \operatorname{im} A = \dim \operatorname{im} A^t$. This means that the maximum number of linearly independent columns of A equals the maximum number of linearly independent rows of A.

Definition 3.38. The common value of $\dim \operatorname{im} A$ and $\dim \operatorname{im} A^t$ is called the rank of A.

3.5 The Method of Least Squares

Let A be an $m \times n$ matrix and $\boldsymbol{y} \in \mathbf{R}^m$. If $\boldsymbol{y} \notin \operatorname{im} A$, then the system $A\boldsymbol{x} = \boldsymbol{y}$ has no solutions. Often, however, it is interesting to find an approximate solution $\boldsymbol{x}$ by requiring that $A\boldsymbol{x}$ be as close as possible to $\boldsymbol{y}$. The distance from $A\boldsymbol{x}$ to $\boldsymbol{y}$ can be measured in several different ways. We choose to define this distance as $\|A\boldsymbol{x} - \boldsymbol{y}\|$ where the norm is the ordinary norm in $\mathbf{R}^m$.

We have $A\boldsymbol{x} \in \operatorname{im} A$ for all $\boldsymbol{x} \in \mathbf{R}^n$, and we know that the orthogonal projection $\boldsymbol{y}'$ of $\boldsymbol{y}$ on $\operatorname{im} A$ is the vector in $\operatorname{im} A$ that is closest to $\boldsymbol{y}$. Hence, we can find our vectors $\boldsymbol{x}$ by solving the system $A\boldsymbol{x} = \boldsymbol{y}'$. The method indicated here involves two steps: first finding $\boldsymbol{y}'$ and then solving $A\boldsymbol{x} = \boldsymbol{y}'$. We shall devise a method that involves only one step. Let $\boldsymbol{y}'$ and $\boldsymbol{y}''$ be the orthogonal projections of $\boldsymbol{y}$ on $\operatorname{im} A$ and $(\operatorname{im} A)^\perp$, respectively. By (3.3), $(\operatorname{im} A)^\perp = \ker A^t$, and hence $A^t\boldsymbol{y}'' = \mathbf{0}$. Since $\boldsymbol{y}'' = \boldsymbol{y} - \boldsymbol{y}'$, this yields $A^t\boldsymbol{y} = A^t\boldsymbol{y}'$. Therefore, $A\boldsymbol{x} = \boldsymbol{y}'$ implies that

$$A^t A\boldsymbol{x} = A^t\boldsymbol{y}' = A^t\boldsymbol{y}.$$

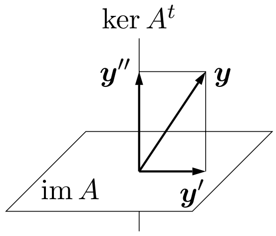
Conversely, if $A^t A\boldsymbol{x} = A^t\boldsymbol{y}$, then $A^t(A\boldsymbol{x} - \boldsymbol{y}') = A^t A\boldsymbol{x} - A^t\boldsymbol{y}' = A^t A\boldsymbol{x} - A^t\boldsymbol{y} = \mathbf{0}$. Hence, $A\boldsymbol{x} - \boldsymbol{y}' \in \ker A^t = (\operatorname{im} A)^\perp$. Since also $A\boldsymbol{x} - \boldsymbol{y}' \in \operatorname{im} A$, we must have $A\boldsymbol{x} - \boldsymbol{y}' = \mathbf{0}$ or, equivalently, $A\boldsymbol{x} = \boldsymbol{y}'$. Hence, the so-called normal equations

$$A^t A\boldsymbol{x} = A^t\boldsymbol{y}$$

give the same result as the two-step method. This method is called the method of least squares and its name stems from the fact that the approximate solutions minimise the sum

$$\|A\boldsymbol{x} - \boldsymbol{y}\|^2 = ((A\boldsymbol{x})_1 - \boldsymbol{y}_1)^2 + \cdots + ((A\boldsymbol{x})_m - \boldsymbol{y}_m)^2$$

of squares.



Note that $\boldsymbol{y} \in \operatorname{im} A$ if the system has solutions. Hence, in this case $\boldsymbol{y}' = \boldsymbol{y}$ and the solutions in the sense of least squares are the ordinary solutions.

Example 3.39. We seek the solution in the sense of least squares of the system

$$\begin{cases} x_1 + x_2 = 6 \\ 4x_1 - x_2 = 8 \\ 3x_1 + 2x_2 = 5 \end{cases}.$$

We set

$$A = \begin{bmatrix} 1 & 1 \\ 4 & -1 \\ 3 & 2 \end{bmatrix} \quad \text{and} \quad Y = \begin{bmatrix} 6 \\ 8 \\ 5 \end{bmatrix}.$$

Then

$$A^tA = \begin{bmatrix} 1 & 4 & 3 \\ 1 & -1 & 2 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 4 & -1 \\ 3 & 2 \end{bmatrix} = \begin{bmatrix} 26 & 3 \\ 3 & 6 \end{bmatrix} \quad \text{and} \quad A^tY = \begin{bmatrix} 1 & 4 & 3 \\ 1 & -1 & 2 \end{bmatrix} \begin{bmatrix} 6 \\ 8 \\ 5 \end{bmatrix} = \begin{bmatrix} 53 \\ 8 \end{bmatrix}.$$

The normal equations for this system are, therefore,

$$\begin{cases} 26x_1 + 3x_2 = 53 \\ 3x_1 + 6x_2 = 8 \end{cases} \Leftrightarrow \begin{cases} x_1 = 2 \\ x_2 = \frac{1}{3} \end{cases}.$$

Hence, the solution in the sense of least squares is given by $x_1 = 2$, $x_2 = \frac{1}{3}$.

Note that A^tA is a symmetric matrix and its entry in position i, k is the dot product $A_i \cdot A_k$. This observation might save you some time and effort.

The following example illustrates a common application of the method of least squares.

Example 3.40. We have reason to believe that some process is described by a linear model $y = at + b$ where y is some quantity and, for example, t is time. Measurements give the following data:

t	0	1	2	3
y	1.5	2.9	5.3	6.6

From these data we want to estimate the values of a and b. Ideally, we should be able to solve for a and b in the following system of equations.

$$\begin{cases} b = 1.5 \\ a + b = 2.9 \\ 2a + b = 5.3 \\ 3a + b = 6.6 \end{cases}.$$

This is, however, seldom possible owing to measure errors or perhaps to the fact that we are mistaken in our assumption about the model. We decide, instead, to minimise the distance in the sense of least squares between the vectors $(b, a + b, 2a + b, 3a + b)$ and $(1.5, 2.9, 5.3, 6.6)$. That is to say that we decide to solve the system in the same sense. We can write the system as $AX = Y$ where

$$A = \begin{bmatrix} 0 & 1 \\ 1 & 1 \\ 2 & 1 \\ 3 & 1 \end{bmatrix}, \quad Y = \begin{bmatrix} 1.5 \\ 2.9 \\ 5.3 \\ 6.6 \end{bmatrix}, \quad X = \begin{bmatrix} a \\ b \end{bmatrix}.$$

We get

$$A^tA = \begin{bmatrix} 14 & 6 \\ 6 & 4 \end{bmatrix}, \quad A^tY = \begin{bmatrix} 33.3 \\ 16.3 \end{bmatrix}$$

and find that the solution of the normal equations $A^tAX = A^tY$ is given by $a = 1.77$, $b = 1.42$.

The applicability of the method in the last example has nothing to do with the assumption that the model is linear. It would have worked equally well under the assumption that $y = at^2 + bt + c$. The important thing here is that the coefficients appear linearly in the expression. Thus, the method is not directly applicable to the exponential model $y = ce^{at}$. However, in this case there is a way to bypass this limitation. This is demonstrated in the next example.

Example 3.41. Assume that the model is $y = ce^{at}$ and that we have the following data:

t	0	1	2	3
y	3	3	5	9

Taking the logarithm, we get the equivalent relation $\ln y = at + \ln c$. Setting $z = \ln y$ and $b = \ln c$, we can write this as $z = at + b$. We compute the values of z and construct a new table:

t	0	1	2	3
z	$\ln 3$	$\ln 3$	$\ln 5$	$\ln 9$

The matrix A is the same matrix as in the previous example. Hence, also A^tA is the same. From now on, the problem is not well suited for manual calculations. By means of some numerical software, we should be able to find that

$$A^tZ = \begin{bmatrix} 10.90916185 \\ 6.003887068 \end{bmatrix}.$$

The solutions of the normal equations $A^tAX = A^tZ$ are now given by $a = 0.3806662490$, $b = 0.9299723935$. Hence, $c = e^b = 2.534439210$.

The approach in the above example usually serves its purpose well. Note, however, that minimising the distance between the vectors comprising the logarithmic values is not the same as minimising the distance between the vectors themselves.

The method of least squares also gives us a means to compute orthogonal projections. Let $U = [\boldsymbol{u}_1, \ldots, \boldsymbol{u}_n]$ be a subspace of an m-dimensional inner product space V. By introducing an orthonormal basis for V, we can regard V as $\mathbf{R}^m$ and the $\boldsymbol{u}_i$ as elements of $\mathbf{R}^m$. If A is the $m \times n$ matrix having the $\boldsymbol{u}_i$ as columns, then $U = \operatorname{im} A$. Hence, if $\boldsymbol{x}$ is any solution of the normal equations $A^tA\boldsymbol{x} = A^t\boldsymbol{u}$, then $\boldsymbol{u}' = A\boldsymbol{x}$ is the orthogonal projection of $\boldsymbol{u}$ on U. If $A^tA\boldsymbol{x} = \mathbf{0}$, then $A\boldsymbol{x} \in \operatorname{im} A \cap (\operatorname{im} A)^\perp = \{\mathbf{0}\}$, and hence $A\boldsymbol{x} = \mathbf{0}$. If the $\boldsymbol{u}_i$ are linearly independent, this implies that $A^tA\boldsymbol{x} = \mathbf{0}$ has the unique solution $\boldsymbol{x} = \mathbf{0}$ and therefore that A^tA is invertible. Hence, if the $\boldsymbol{u}_i$ are linearly independent, we have the following formula:

$$\boldsymbol{u}' = A\boldsymbol{x} = A(A^tA)^{-1}A^t\boldsymbol{u}. \tag{3.4}$$

If the basis $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_n$ above is an orthonormal basis for U, then $A^tA = I^{(n)}$, and (3.4) reads

$$\boldsymbol{u}' = AA^t\boldsymbol{u}.$$

This is Formula (3.1) written in the language of matrices.

Exercises

3.1. Find the angle between the two vectors $\boldsymbol{u} = (-1, 1, 1, -1, 0)$ and $\boldsymbol{v} = (0, 2, 1, 0, 2)$ in $\mathbf{R}^5$.

3.2. Show that the four points $(1, 1, 2, 2)$, $(2, 2, 3, 3)$, $(3, 1, 4, 2)$, $(2, 0, 3, 1)$ in $\mathbf{R}^4$ are the vertices of a square.

3.3. Show that the parallelogram identity

$$2\left(\|\boldsymbol{u}\|^2 + \|\boldsymbol{v}\|^2\right) = \|\boldsymbol{u} + \boldsymbol{v}\|^2 + \|\boldsymbol{u} - \boldsymbol{v}\|^2$$

holds in every inner product space.

3.4. Show that an equilateral triangle in an inner product space V is equiangular by showing that the angle between two non-zero vectors $\boldsymbol{u}$ and $\boldsymbol{v}$ of V is $\frac{\pi}{3}$ if $\|\boldsymbol{u}\| = \|\boldsymbol{v}\| = \|\boldsymbol{u} - \boldsymbol{v}\|$.

3.5. Find an orthonormal basis for the subspace of $\mathbf{R}^4$ spanned by $(2,1,1,1)$, $(1,2,3,0)$ and $(1,1,1,1)$.

3.6. Find an orthonormal basis for the subspace of $\mathbf{R}^4$ given by

$$x_1 + 2x_2 - 2x_3 - x_4 = 0.$$

3.7. Determine the constants a, b and c so that the matrix

$$\frac{1}{3}\begin{bmatrix} 2 & 1 & 2 \\ 2 & -2 & -1 \\ a & b & c \end{bmatrix}$$

is orthogonal.

3.8. Let A and B be orthogonal $n \times n$ matrices. Show that AB is an orthogonal matrix.

3.9. (a) Show that $\langle \boldsymbol{p}, \boldsymbol{q} \rangle = \int_0^1 \boldsymbol{p}(x)\boldsymbol{q}(x)\,dx$ defines an inner product on P_2.

(b) Use the Gram–Schmidt process to find an orthonormal basis for P_2 equipped with this inner product.

3.10. Find an orthonormal basis for the orthogonal complement of the subspace of $\mathbf{R}^4$ spanned by $(1,1,1,1)$ and $(0,1,2,1)$.

3.11. Find the orthogonal projections of $\boldsymbol{u} = (1,2,3,4)$ on the subspaces of $\mathbf{R}^4$ spanned by

(a) $(1,1,1,1)$ and $(1,-1,1,-1)$,

(b) $(1,1,1,1)$ and $(1,1,1,0)$,

and compute the distances from $\boldsymbol{u}$ to the two subspaces.

3.12. Show that the vectors $\boldsymbol{e}_1 = \frac{1}{\sqrt{3}}(1,1,1,0)$ and $\boldsymbol{e}_2 = \frac{1}{3}(0,2,-2,1)$ form an orthonormal set in $\mathbf{R}^4$. Extend this set to an orthonormal basis for $\mathbf{R}^4$ by first finding a basis $\boldsymbol{v}_1, \boldsymbol{v}_2$ for the orthogonal complement of $U = [\boldsymbol{e}_1, \boldsymbol{e}_2]$ and then applying the Gram–Schmidt process to $\boldsymbol{v}_1$ and $\boldsymbol{v}_2$.

3.13. Let $\boldsymbol{e}_1, \dots, \boldsymbol{e}_n$ be an orthonormal set in an inner product space V and let $\boldsymbol{u}$ be any vector in V.

(a) Show that

$$\|\boldsymbol{u}'\|^2 = \sum_{k=1}^{n} \langle \boldsymbol{u}, \boldsymbol{e}_k \rangle^2$$

where $\boldsymbol{u}'$ is the orthogonal projection of $\boldsymbol{u}$ on $[\boldsymbol{e}_1, \dots, \boldsymbol{e}_n]$.

(b) Show Bessel's inequality

$$\sum_{k=1}^{n} \langle \boldsymbol{u}, \boldsymbol{e}_k \rangle^2 \le \|\boldsymbol{u}\|^2.$$

3.14. (a) Show that the vectors $\boldsymbol{e}_1, \boldsymbol{e}_2, \boldsymbol{e}_3, \ldots$ in $C[0, \pi]$ defined by

$$\boldsymbol{e}_j(t) = \sqrt{\frac{2}{\pi}} \sin jt, \quad 0 \le t \le \pi, \quad j = 1, 2, 3, \ldots$$

are pairwise orthogonal and of norm 1 with respect to the inner product

$$\langle \boldsymbol{u}, \boldsymbol{v} \rangle = \int_0^\pi \boldsymbol{u}(t)\boldsymbol{v}(t)\, dt.$$

Hint: The identity

$$2 \sin kt \sin jt = \cos{(k-j)t} - \cos{(k+j)t}$$

might prove useful.

(b) Find the orthogonal projection of $\boldsymbol{u}$ defined by $\boldsymbol{u}(t) = t$, $0 \le t \le \pi$, on the span $[\boldsymbol{e}_1, \boldsymbol{e}_2]$.

3.15. Find the orthogonal projection of $\boldsymbol{u} = (3, 2, 1, 4, 5, 6)$ on the kernel of

$$\begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 & 1 & 1 \end{bmatrix}.$$

3.16. (a) Find the rank of the matrix

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

(b) Find bases for $\operatorname{im} A$ and $\operatorname{im} A^t$.

3.17. Find the column vector $X \in \mathbf{R}^3$ that minimises the distance between AX and Y where

$$A = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 2 \\ 2 & 1 & 2 \\ 2 & 1 & 0 \end{bmatrix} \quad \text{and} \quad Y = \begin{bmatrix} 1 \\ 0 \\ 1 \\ 0 \end{bmatrix}.$$

3.18. Find the polynomial $y = at + b$ that is the best least-squares fit to the following data.

t	0	1	2	3
y	5	2	1	0

3.19. Find the polynomial $y = at^2 + bt + c$ that is the best least-squares fit to the following data.

t	-1	0	1	2
y	2	2	4	6

3.20. Make use of Formula (3.4) to find the orthogonal projection of $\boldsymbol{u} = (1, 2, 3, 4)$ on the subspace of $\mathbf{R}^4$ spanned by $(1, 1, 1, 1)$ and $(1, 1, 1, -1)$.

4 Determinants

4.1 Multilinear Forms

Definition 4.1. Let V be a linear space. A function $F : V^n \to \mathbf{R}$, where the Cartesian product $V^n = V \times V \times \cdots \times V$ contains n copies of V, is said to be an n-multilinear form on V provided that the function $\boldsymbol{u} \mapsto F(\boldsymbol{u}_1, \ldots, \boldsymbol{u}_{i-1}, \boldsymbol{u}, \boldsymbol{u}_{i+1}, \ldots, \boldsymbol{u}_n)$ is a linear transformation $V \to \mathbf{R}$ for every index i and every $(n-1)$-tuple $(\boldsymbol{u}_1, \ldots, \boldsymbol{u}_{i-1}, \boldsymbol{u}_{i+1}, \ldots, \boldsymbol{u}_n)$. We say that F is alternating if $F(\boldsymbol{u}_1, \ldots, \boldsymbol{u}_n) = 0$ whenever there exists an index i, $1 \le i \le n-1$, such that $\boldsymbol{u}_i = \boldsymbol{u}_{i+1}$.

Hence, F is an n-multilinear form if

$$F(\ldots, \boldsymbol{u}_{i-1}, s\boldsymbol{u} + t\boldsymbol{v}, \boldsymbol{u}_{i+1}, \ldots) = sF(\ldots, \boldsymbol{u}_{i-1}, \boldsymbol{u}, \boldsymbol{u}_{i+1}, \ldots) + tF(\ldots, \boldsymbol{u}_{i-1}, \boldsymbol{v}, \boldsymbol{u}_{i+1}, \ldots)$$

and alternating if $F(\boldsymbol{u}_1 \ldots, \boldsymbol{u}_n) = 0$ whenever two adjacent vectors are equal.

Theorem 4.2. Let F be an n-multilinear alternating form on a linear space V. If $i \neq j$, then

$$F(\ldots, \boldsymbol{u}_i, \ldots, \boldsymbol{u}_j, \ldots) = -F(\ldots, \boldsymbol{u}_j, \ldots, \boldsymbol{u}_i, \ldots);$$

that is, F changes by a sign if two vectors are interchanged. If $i \neq j$ and $\boldsymbol{u}_i = \boldsymbol{u}_j$, then $F(\boldsymbol{u}_1, \ldots, \boldsymbol{u}_n) = 0$.

Proof. We begin by showing the first statement in the special case where $j = i+1$. Since F is alternating, we have

$$F(\ldots, \boldsymbol{u}_i + \boldsymbol{u}_j, \boldsymbol{u}_i + \boldsymbol{u}_j, \ldots) = 0.$$

Using the multilinearity and the alternating property, we get

$$\begin{aligned} 0 &= F(\ldots, \boldsymbol{u}_i + \boldsymbol{u}_j, \boldsymbol{u}_i + \boldsymbol{u}_j, \ldots) = F(\ldots, \boldsymbol{u}_i, \boldsymbol{u}_i + \boldsymbol{u}_j, \ldots) + F(\ldots, \boldsymbol{u}_j, \boldsymbol{u}_i + \boldsymbol{u}_j, \ldots) \\ &= F(\ldots, \boldsymbol{u}_i, \boldsymbol{u}_i, \ldots) + F(\ldots, \boldsymbol{u}_i, \boldsymbol{u}_j, \ldots) + F(\ldots, \boldsymbol{u}_j, \boldsymbol{u}_i, \ldots) + F(\ldots, \boldsymbol{u}_j, \boldsymbol{u}_j, \ldots) \\ &= F(\ldots, \boldsymbol{u}_i, \boldsymbol{u}_j, \ldots) + F(\ldots, \boldsymbol{u}_j, \boldsymbol{u}_i, \ldots). \end{aligned}$$

Hence, $F(\ldots, \boldsymbol{u}_i, \boldsymbol{u}_j, \ldots) = -F(\ldots, \boldsymbol{u}_j, \boldsymbol{u}_i, \ldots)$.

For the last statement, we interchange successively adjacent vectors until we obtain an n-tuple of vectors having two equal adjacent vectors. Since the resulting function value is zero and can differ from the original value only by a sign, also the original function value must be zero.

To show the first statement in general, we assume that $i \neq j$. It then follows from the last statement and the multilinearity that

$$\begin{aligned} 0 &= F(\dots, \boldsymbol{u}_i + \boldsymbol{u}_j, \dots, \boldsymbol{u}_i + \boldsymbol{u}_j, \dots) \\ &= F(\dots, \boldsymbol{u}_i, \dots, \boldsymbol{u}_i, \dots) + F(\dots, \boldsymbol{u}_i, \dots, \boldsymbol{u}_j, \dots) \\ &\quad + F(\dots, \boldsymbol{u}_j, \dots, \boldsymbol{u}_i, \dots) + F(\dots, \boldsymbol{u}_j, \dots, \boldsymbol{u}_j, \dots) \\ &= F(\dots, \boldsymbol{u}_i, \dots, \boldsymbol{u}_j, \dots) + F(\dots, \boldsymbol{u}_j, \dots, \boldsymbol{u}_i, \dots). \ \blacksquare \end{aligned}$$

Corollary 4.3. The value of F does not change if a multiple of the vector in one position is added to the vector in another position.

Proof. If $i \neq j$, we have

$$\begin{aligned} F(\dots, \boldsymbol{u}_i, \dots, \boldsymbol{u}_j + s\boldsymbol{u}_i, \dots) &= F(\dots, \boldsymbol{u}_i, \dots, \boldsymbol{u}_j, \dots) + sF(\dots, \boldsymbol{u}_i, \dots, \boldsymbol{u}_i, \dots) \\ &= F(\dots, \boldsymbol{u}_i, \dots, \boldsymbol{u}_j, \dots). \ \blacksquare \end{aligned}$$

Theorem 4.4. Let V be an n-dimensional linear space with basis $\boldsymbol{e}_1, \dots, \boldsymbol{e}_n$ and let F be an n-multilinear alternating form on V. If $F(\boldsymbol{e}_1, \dots, \boldsymbol{e}_n) = 0$, then $F(\boldsymbol{u}_1, \dots, \boldsymbol{u}_n) = 0$ for all n-tuples $(\boldsymbol{u}_1, \dots, \boldsymbol{u}_n)$ of vectors in V.

Proof. Let $i_1, \dots, i_n$ be any indices such that $1 \leq i_j \leq n$ for $j = 1, \dots, n$ and consider $F(\boldsymbol{e}_{i_1}, \dots, \boldsymbol{e}_{i_n})$. If two of the indices are equal, then $F(\boldsymbol{e}_{i_1}, \dots, \boldsymbol{e}_{i_n}) = 0$ by the second statement of Theorem 4.2. If the indices are distinct, we can successively interchange adjacent vectors until we get the n-tuple $(\boldsymbol{e}_1, \dots, \boldsymbol{e}_n)$. Hence, by the first statement of the same theorem, $F(\boldsymbol{e}_{i_1}, \dots, \boldsymbol{e}_{i_n}) = \pm F(\boldsymbol{e}_1, \dots, \boldsymbol{e}_n) = 0$. Let $\boldsymbol{u}_1, \dots, \boldsymbol{u}_n$ be vectors of V. We can then write $\boldsymbol{u}_i = x_{i1}\boldsymbol{e}_1 + \dots + x_{in}\boldsymbol{e}_n$. Using the linearity in the first argument, we get

$$F(\boldsymbol{u}_1, \boldsymbol{u}_2, \dots, \boldsymbol{u}_n) = F(x_{11}\boldsymbol{e}_1 + \dots + x_{1n}\boldsymbol{e}_n, \boldsymbol{u}_2, \dots, \boldsymbol{u}_n) = \sum_{i_1=1}^{n} x_{1i_1} F(\boldsymbol{e}_{i_1}, \boldsymbol{u}_2, \dots, \boldsymbol{u}_n).$$

By using the linearity in each of the remaining arguments, we eventually get

$$F(\boldsymbol{u}_1, \boldsymbol{u}_2, \dots, \boldsymbol{u}_n) = \sum x_{1i_1} x_{2i_2} \cdots x_{ni_n} F(\boldsymbol{e}_{i_1}, \boldsymbol{e}_{i_2}, \dots, \boldsymbol{e}_{i_n})$$

where the sum is taken over all n-tuples $(i_1, \dots, i_n)$ of indices. The assertion now follows from the fact that all terms of the sum are zero. $\blacksquare$

Corollary 4.5. Let V be an n-dimensional linear space with basis $\boldsymbol{e}_1, \dots, \boldsymbol{e}_n$ and let F and G be n-multilinear alternating forms on V. If $F(\boldsymbol{e}_1, \dots, \boldsymbol{e}_n) = G(\boldsymbol{e}_1, \dots, \boldsymbol{e}_n)$, then $F(\boldsymbol{u}_1, \dots, \boldsymbol{u}_n) = G(\boldsymbol{u}_1, \dots, \boldsymbol{u}_n)$ for all n-tuples $(\boldsymbol{u}_1, \dots, \boldsymbol{u}_n)$ of vectors in V.

Proof. It is plain that $F - G$ is an n-multilinear alternating form on V. Hence, the statement follows from the preceding theorem. $\blacksquare$

4.2 Definition of Determinants

Let M_n be the set of $n \times n$ matrices. By a determinant of order n we shall mean a mapping $D : M_n \to \mathbf{R}$ with the following properties:

(i) $D([\ldots, sA_k + tA'_k, \ldots]) = sD([\ldots, A_k, \ldots]) + tD([\ldots, A'_k, \ldots])$,
(ii) $D(A) = 0$ if two adjacent columns of A are equal,
(iii) $D(I) = 1$.

Consider the linear space V of $n \times 1$ columns. The columns I_k, $k = 1, \ldots, n$, of the unit matrix I form a basis for V. When D is viewed as a function $V^n \to \mathbf{R}$, conditions (i) and (ii) mean that D is an n-multilinear alternating form on V. By Corollary 4.5 and condition (iii), a determinant is uniquely determined if it exists.

We shall now define determinants of all orders recursively. Let A be an $n \times n$ matrix where $n \geq 2$. By A_{ik} we mean the $(n-1) \times (n-1)$ matrix obtained from A by deleting row i and column k. Hence, if

$$A = \begin{bmatrix} a_{11} & \cdots & a_{1(k-1)} & a_{1k} & a_{1(k+1)} & \cdots & a_{1n} \\ \vdots & & \vdots & \vdots & \vdots & & \vdots \\ a_{(i-1)1} & \cdots & a_{(i-1)(k-1)} & a_{(i-1)k} & a_{(i-1)(k+1)} & \cdots & a_{(i-1)n} \\ a_{i1} & \cdots & a_{i(k-1)} & a_{ik} & a_{i(k+1)} & \cdots & a_{in} \\ a_{(i+1)1} & \cdots & a_{(i+1)(k-1)} & a_{(i+1)k} & a_{(i+1)(k+1)} & \cdots & a_{(i+1)n} \\ \vdots & & \vdots & \vdots & \vdots & & \vdots \\ a_{n1} & \cdots & a_{n(k-1)} & a_{nk} & a_{n(k+1)} & \cdots & a_{nn} \end{bmatrix},$$

then

$$A_{ik} = \begin{bmatrix} a_{11} & \cdots & a_{1(k-1)} & a_{1(k+1)} & \cdots & a_{1n} \\ \vdots & & \vdots & \vdots & & \vdots \\ a_{(i-1)1} & \cdots & a_{(i-1)(k-1)} & a_{(i-1)(k+1)} & \cdots & a_{(i-1)n} \\ a_{(i+1)1} & \cdots & a_{(i+1)(k-1)} & a_{(i+1)(k+1)} & \cdots & a_{(i+1)n} \\ \vdots & & \vdots & \vdots & & \vdots \\ a_{n1} & \cdots & a_{n(k-1)} & a_{n(k+1)} & \cdots & a_{nn} \end{bmatrix}.$$

Theorem 4.6. Let $n \geq 2$ be an integer and i an integer such that $1 \leq i \leq n$. If D_{n-1} is a determinant of order $n-1$, then the mapping defined by

$$\begin{aligned} D_n(A) &= \sum_{j=1}^{n} (-1)^{i+j} a_{ij} D_{n-1}(A_{ij}) \\ &= (-1)^{i+1} a_{i1} D_{n-1}(A_{i1}) + (-1)^{i+2} a_{i2} D_{n-1}(A_{i2}) + \cdots + (-1)^{i+n} a_{in} D_{n-1}(A_{in}) \end{aligned}$$

is a determinant of order n.

Proof. We denote D_{n-1} by D in this proof. Assume that $A = [a_{ik}]_{n\times n}$ and let A' and B be the matrices obtained from A by replacing its kth column with

$$\begin{bmatrix} a'_{1k} \\ \vdots \\ a'_{ik} \\ \vdots \\ a'_{nk} \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} sa_{1k} + ta'_{1k} \\ \vdots \\ sa_{ik} + ta'_{ik} \\ \vdots \\ sa_{nk} + ta'_{nk} \end{bmatrix},$$

respectively. Plainly, $B_{ik} = A_{ik} = A'_{ik}$ and $b_{ik} = sa_{ik} + ta'_{ik}$. If $j \neq k$, then $b_{ij} = a_{ij} = a'_{ij}$. In this case we also have $D(B_{ij}) = sD(A_{ij}) + tD(A'_{ij})$ since D is assumed to satisfy condition (i). Therefore,

$$\begin{aligned} D_n(B) &= \sum_{j=1}^{n} (-1)^{i+j} b_{ij} D(B_{ij}) = (-1)^{i+k} b_{ik} D(B_{ik}) + \sum_{j\neq k} (-1)^{i+j} b_{ij} D(B_{ij}) \\ &= (-1)^{i+k}(sa_{ik} + ta'_{ik}) D(B_{ik}) + \sum_{j\neq k} (-1)^{i+j} b_{ij}(sD(A_{ij}) + tD(A'_{ij})) \\ &= (-1)^{i+k}(sa_{ik} D(A_{ik}) + ta'_{ik} D(A'_{ik})) + \sum_{j\neq k} (-1)^{i+j}(sa_{ij} D(A_{ij}) + ta'_{ij} D(A'_{ij})) \\ &= s\sum_{j=1}^{n} (-1)^{i+j} a_{ij} D(A_{ij}) + t\sum_{j=1}^{n} (-1)^{i+j} a'_{ij} D(A'_{ij}) = sD_n(A) + tD_n(A'). \end{aligned}$$

Hence, the mapping D_n meets condition (i).

Assume that the columns A_k and A_{k+1} of the $n \times n$ matrix A are equal. If $j \neq k$ and $j \neq k+1$, then two adjacent columns of A_{ij} are equal. Hence $D(A_{ij}) = 0$ in this case. It is also clear that $A_{ik} = A_{i(k+1)}$ and that $a_{ik} = a_{i(k+1)}$. Therefore,

$$D_n(A) = (-1)^{i+k} a_{ik} D(A_{ik}) + (-1)^{i+k+1} a_{i(k+1)} D(A_{i(k+1)}) = 0.$$

This shows that D_n satisfies condition (ii).

The ij-entry δ_{ij} of the unit matrix equals 1 when $i = j$ and 0 otherwise. We also see that I_{ii} is the unit matrix of order $n-1$. Hence,

$$D_n(I) = \sum_{j=1}^{n} (-1)^{i+j} \delta_{ij} D(I_{ij}) = (-1)^{i+i} \delta_{ii} D(I_{ii}) = D(I_{ii}) = 1$$

and, therefore, also condition (iii) is met. ■

Definition 4.7. We define the function $D_n : M_n \to \mathbf{R}$ recursively as follows. For 1×1 matrices $A = \begin{bmatrix} a \end{bmatrix}$, we set $D_1(A) = a$. When A is an $n \times n$ matrix where $n \geq 2$, we set

$$\begin{aligned} D_n(A) &= \sum_{j=1}^{n} (-1)^{1+j} a_{1j} D_{n-1}(A_{1j}) \\ &= a_{11} D_{n-1}(A_{11}) - a_{12} D_{n-1}(A_{12}) + \cdots + (-1)^{1+n} a_{1n} D_{n-1}(A_{1n}). \end{aligned}$$

Theorem 4.8. For every positive integer n, the function D_n is a determinant.

Proof. The function D_1 satisfies condition (ii) for the simple reason that a square matrix of order 1 has no adjacent columns. The other two conditions are trivially satisfied. Hence D_1 is a determinant. The statement now follows by induction on n, the induction step being supplied by Theorem 4.6. ∎

Hence, determinants exist of all orders and are unique. In order not to overload the notation, we shall from now on denote determinants of all orders by D. Other notations used for $D(A)$, where $A = [a_{ik}]_{n\times n}$, are $\det A$ and

$$\begin{vmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{vmatrix}.$$

For 2×2 matrices, the definition yields

$$\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = (-1)^{1+1} a_{11} \left|a_{22}\right| + (-1)^{1+2} a_{12} \left|a_{21}\right| = a_{11}a_{22} - a_{12}a_{21}.$$

Do not mistake $\left|a_{22}\right|$ and $\left|a_{21}\right|$ for absolute values here. To avoid this ambiguity, we shall never again use this notation for determinants of order 1.

Example 4.9. Using the definition and the above formula for determinants of order 2, we find that the determinant of the matrix

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix}$$

is

$$\begin{aligned} \det A = \begin{vmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{vmatrix} &= (-1)^2 \cdot 1 \cdot \begin{vmatrix} 5 & 6 \\ 8 & 9 \end{vmatrix} + (-1)^3 \cdot 2 \cdot \begin{vmatrix} 4 & 6 \\ 7 & 9 \end{vmatrix} + (-1)^4 \cdot 3 \cdot \begin{vmatrix} 4 & 5 \\ 7 & 8 \end{vmatrix} \\ &= 5 \cdot 9 - 6 \cdot 8 - 2(4 \cdot 9 - 6 \cdot 7) + 3(4 \cdot 8 - 5 \cdot 7) = 0. \end{aligned}$$

For a general 3×3 matrix

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix},$$

we obtain

$$\begin{aligned} D(A) &= a_{11} \begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} - a_{12} \begin{vmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{vmatrix} + a_{13} \begin{vmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{vmatrix} \\ &= a_{11}(a_{22}a_{33} - a_{23}a_{32}) - a_{12}(a_{21}a_{33} - a_{23}a_{31}) + a_{13}(a_{21}a_{32} - a_{22}a_{31}) \\ &= a_{11}a_{22}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} - (a_{11}a_{23}a_{32} + a_{12}a_{21}a_{33} + a_{13}a_{22}a_{31}). \end{aligned}$$

The reader probably recognises this as Sarrus's rule for determinants of order 3. Hence, for determinants of order 2 and 3, our definition agrees with the ones usually stated for such determinants.

4.3 Properties of Determinants

Theorem 4.10. Let A be an $n \times n$ matrix, where $n \geq 2$, and i an integer such that $1 \leq i \leq n$. Then

$$\begin{aligned} D(A) &= \sum_{j=1}^{n} (-1)^{i+j} a_{ij} D(A_{ij}) \\ &= (-1)^{i+1} a_{i1} D(A_{i1}) + (-1)^{i+2} a_{i2} D(A_{i2}) + \cdots + (-1)^{i+n} a_{in} D(A_{in}). \end{aligned}$$

Proof. By Theorem 4.6, the right-hand side is a determinant of order n. The equality therefore follows from the uniqueness of determinants. ■

Theorem 4.11. Let A be a square matrix. Then $D(A) = D(A^t)$.

Proof. Define $D'(A) = D(A^t)$ for $n \times n$ matrices A. We intend to show that D' is a determinant of order n. It will then follow from the uniqueness of the determinant that $D'(A) = D(A)$ for all square matrices A of order n. The statement holds for $n = 1$ since in this case $A^t = A$. We proceed by showing that each of the defining conditions is satisfied by D' for every $n \geq 2$.

To show condition (i) for D' we must show that D is linear in each row. Assume that $A = [a_{ik}]_{n \times n}$ and let A' and B be the matrices obtained from A by replacing its ith row with

$$\begin{bmatrix} a'_{i1} & \cdots & a'_{in} \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} sa_{i1} + ta'_{i1} & \cdots & sa_{in} + ta'_{in} \end{bmatrix},$$

respectively. Then $A_{ij} = A'_{ij} = B_{ij}$ for $j = 1, \ldots, n$, and hence, by Theorem 4.10,

$$\begin{aligned} D(B) &= \sum_{j=1}^{n} (-1)^{i+j} (sa_{ij} + ta'_{ij}) D(B_{ij}) \\ &= s \sum_{j=1}^{n} (-1)^{i+j} a_{ij} D(A_{ij}) + t \sum_{j=1}^{n} (-1)^{i+j} a'_{ij} D(A'_{ij}) = sD(A) + tD(A'). \end{aligned}$$

Hence, D' meets condition (i).

Condition (ii) for D' means that $D(A) = 0$ if two adjacent rows of A are equal. For $n = 2$, this follows from the formula $D(A) = a_{11}a_{22} - a_{12}a_{21}$. We proceed by induction on n. Assume that $n \geq 3$ and that the statement holds for orders less than n. Let i be the index of a row other than the two equal rows. Then, for every $j = 1, \ldots, n$, A_{ij} has two equal adjacent rows and hence, by hypothesis, $D(A_{ij}) = 0$. It now follows from Theorem 4.10 that $D(A) = 0$.

Condition (iii) is satisfied by D' since $D'(I) = D(I^t) = D(I) = 1$. ■

Since the columns of A are the rows of A^t, Theorems 4.10 and 4.11 yield the following theorem.

Theorem 4.12. Let A be an $n \times n$ matrix, where $n \geq 2$, and j an integer such that $1 \leq j \leq n$. Then

$$\begin{aligned} D(A) &= \sum_{i=1}^{n} (-1)^{i+j} a_{ij} D(A_{ij}) \\ &= (-1)^{1+j} a_{1j} D(A_{1j}) + (-1)^{2+j} a_{2j} D(A_{2j}) + \cdots + (-1)^{n+j} a_{nj} D(A_{nj}). \end{aligned}$$

Theorem 4.13. The value of a determinant does not change when a multiple of a column is added to another column or when a multiple of a row is added to another row.

Proof. The statement concerning columns is contained in Corollary 4.3. Since the rows of a matrix are the columns of its transpose, the statement about rows follows from Theorem 4.11. ∎

Theorem 4.14. $D(A) = 0$ if two columns of A are equal or two rows of A are equal. $D(A)$ changes by a sign if two columns of A are interchanged or two rows of A are interchanged.

Proof. The statements about columns are translations into the language of determinants of the statements of Theorem 4.2. Their row analogues follow from Theorem 4.11. ∎

Theorems 4.10, 4.12, 4.13 and 4.14 provide efficient tools for evaluation of determinants. The formulae of the first two theorems are called expansion along a row and column, respectively.

Using the recursive definition of determinants amounts to successive expansions along the first row. A determinant of order 4 first splits into 4 determinants of order 3, and then each of these splits into 3 determinants of order 2. Hence, we must evaluate 12 determinants of order 2.

By Theorem 4.13, a determinant can be transformed into one containing a row or column having at most one non-zero entry. The expansion along the transformed row or column contains at most one non-zero term. Correct use of the tools decreases the workload significantly.

Example 4.15. We demonstrate the tools by, once again, evaluating the determinant in Example 4.9. For the sake of easy calculations we choose to produce zeros in the third column. By subtracting twice the first row from the second row and thrice the first row from the last row, we obtain

$$\begin{vmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{vmatrix} = \begin{vmatrix} 1 & 2 & 3 \\ 2 & 1 & 0 \\ 4 & 2 & 0 \end{vmatrix}.$$

Expanding along column 3, we get the same value as before:

$$\begin{vmatrix} 1 & 2 & 3 \\ 2 & 1 & 0 \\ 4 & 2 & 0 \end{vmatrix} = (-1)^{1+3} \cdot 3 \begin{vmatrix} 2 & 1 \\ 4 & 2 \end{vmatrix} + (-1)^{2+3} \cdot 0 \begin{vmatrix} 1 & 2 \\ 4 & 2 \end{vmatrix} + (-1)^{3+3} \cdot 0 \begin{vmatrix} 1 & 2 \\ 2 & 1 \end{vmatrix} = 3(2 \cdot 2 - 1 \cdot 4) = 0.$$

The zero terms are written out for the convenience of the reader. In fact, the whole evaluation fits on a single line:

$$\begin{vmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{vmatrix} = \begin{vmatrix} 1 & 2 & 3 \\ 2 & 1 & 0 \\ 4 & 2 & 0 \end{vmatrix} = (-1)^{1+3} \cdot 3 \begin{vmatrix} 2 & 1 \\ 4 & 2 \end{vmatrix} = 3(2 \cdot 2 - 1 \cdot 4) = 0.$$

The reader is discouraged from using Sarrus's rule. The reason for this is twofold. Firstly, it applies only to determinants of order 3. Secondly, it requires unnecessarily long calculations.

Example 4.16. Consider the determinant

$$\begin{vmatrix} 3 & 8 & 4 & 6 \\ 5 & 3 & 2 & 4 \\ 7 & 11 & 4 & 3 \\ 11 & 13 & 6 & 10 \end{vmatrix}.$$

Column 3 is best suited for elimination since its entries are integral multiples of its entry in row 2. We subtract twice the second row from the first and third rows and thrice the same row from the last row and get

$$\begin{vmatrix} -7 & 2 & 0 & -2 \\ 5 & 3 & 2 & 4 \\ -3 & 5 & 0 & -5 \\ -4 & 4 & 0 & -2 \end{vmatrix}.$$

Expansion along the third column yields

$$(-1)^{2+3} \cdot 2 \begin{vmatrix} -7 & 2 & -2 \\ -3 & 5 & -5 \\ -4 & 4 & -2 \end{vmatrix} = -2 \begin{vmatrix} -7 & 2 & -2 \\ -3 & 5 & -5 \\ -4 & 4 & -2 \end{vmatrix}.$$

We now choose to eliminate in row 3 for the same reason as before. Hence, we subtract twice the last column from the first column and add twice the same column to the second column. Thus we get

$$-2 \begin{vmatrix} -3 & -2 & -2 \\ 7 & -5 & -5 \\ 0 & 0 & -2 \end{vmatrix}.$$

Expanding along the third row and then using the formula for determinants of order 2, we obtain

$$-2(-1)^{3+3}(-2) \begin{vmatrix} -3 & -2 \\ 7 & -5 \end{vmatrix} = 4((-3)(-5) - (-2) \cdot 7) = 116.$$

An upper triangular square matrix has zeros below its main diagonal. Likewise, a lower triangular square matrix has zeros above its main diagonal. A diagonal square matrix has zeros outside its main diagonal. Hence, a diagonal matrix is upper and lower triangular. The determinant of a matrix of any of these kinds equals the product of the diagonal entries of that matrix. For an upper triangular matrix, this can be seen by successively expanding along the first column:

$$\begin{vmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1n} \\ 0 & a_{22} & a_{23} & \cdots & a_{2n} \\ 0 & 0 & a_{33} & \cdots & a_{3n} \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & a_{nn} \end{vmatrix} = a_{11} \begin{vmatrix} a_{22} & a_{23} & \cdots & a_{2n} \\ 0 & a_{33} & \cdots & a_{3n} \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & a_{nn} \end{vmatrix} = \cdots = a_{11}a_{22}a_{33}\cdots a_{nn}.$$

Example 4.17. Here we evaluate the determinant

$$\begin{vmatrix} x & 1 & 1 & \cdots & 1 \\ 1 & x & 1 & \cdots & 1 \\ 1 & 1 & x & \cdots & 1 \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & 1 & 1 & \cdots & x \end{vmatrix}$$

of order n. Observing that all the row sums are equal, we add the last $n-1$ columns to the first column and obtain

$$\begin{vmatrix} x+n-1 & 1 & 1 & \cdots & 1 \\ x+n-1 & x & 1 & \cdots & 1 \\ x+n-1 & 1 & x & \cdots & 1 \\ \vdots & \vdots & \vdots & & \vdots \\ x+n-1 & 1 & 1 & \cdots & x \end{vmatrix}.$$

Now, subtracting the first row from the other rows, we get

$$\begin{vmatrix} x+n-1 & 1 & 1 & \cdots & 1 \\ 0 & x-1 & 0 & \cdots & 0 \\ 0 & 0 & x-1 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & x-1 \end{vmatrix}.$$

Since this matrix is triangular, the determinant equals $(x+n-1)(x-1)^{n-1}$.

Theorem 4.18 (Product theorem). Let A and B be square matrices of order n. Then $D(AB) = D(A)D(B)$.

Proof. Let, for a fixed matrix A, the mappings D' and D'' from M_n to $\mathbf{R}$ be defined by $D'(B) = D(A)D(B)$ and $D''(B) = D(AB)$. When viewed as functions $V^n \to \mathbf{R}$, where

V is the linear space of $n \times 1$ columns, these mappings are n-multilinear alternating forms on V. Since D is such a form, this statement is trivial for D'. For D'', it follows from the fact that $D''(B_1, \ldots, B_n) = D(AB_1, \ldots, AB_n)$. Therefore, and since $D'(I) = D''(I)$, Corollary 4.5 yields that $D'(B) = D''(B)$ for all $n \times n$ matrices B. ■

Theorem 4.19. A square matrix A is invertible if and only if $D(A) \neq 0$, and in that case $D(A^{-1}) = (D(A))^{-1}$.

Proof. Assume that A is invertible with inverse A^{-1}. Then it follows from Theorem 4.18 that $D(A)D(A^{-1}) = D(AA^{-1}) = D(I) = 1$, whence $D(A) \neq 0$ and $D(A^{-1}) = (D(A))^{-1}$.

The converse amounts to saying that $D(A) = 0$ if A is not invertible. Assume that A is not invertible and let n be the order of A. If $n = 1$, then $A = \begin{bmatrix} 0 \end{bmatrix}$ and hence $D(A) = 0$ in this case. Otherwise, the columns of A are linearly dependent, and therefore one column A_k is a linear combination of the other columns. Thus,

$$A_k = s_1 A_1 + \cdots + s_{k-1} A_{k-1} + s_{k+1} A_{k+1} + \cdots + s_n A_n$$

for some real numbers $s_1, \ldots, s_{k-1}, s_{k+1}, \ldots, s_n$, and therefore, by linearity in the kth argument,

$$D(A) = D(A_1, \ldots, A_{k-1}, A_k, A_{k+1}, \ldots, A_n) = \sum_{i \neq k} s_i D(A_1, \ldots, A_{k-1}, A_i, A_{k+1}, \ldots, A_n).$$

Each determinant in the sum has two equal columns, and is therefore equal to zero by Theorem 4.14. Hence, $D(A) = 0$ also in this case. ■

Example 4.20. Let us, for every $a \in \mathbf{R}$, find the dimensions of $\ker A$ and $\operatorname{im} A$ where

$$A = \begin{bmatrix} a & 2 & 1 & 2 \\ 2 & a & 2 & 1 \\ 1 & 2 & a & 2 \\ 2 & 1 & 2 & a \end{bmatrix}.$$

By Theorem 4.19, A is invertible if and only if $D(A) \neq 0$, and then $\dim \ker A = 0$ and $\dim \operatorname{im} A = 4$. We observe that the column sums are equal. We choose to add the first three rows to the last row. Thus we get

$$\begin{aligned}
D(A) &= \begin{vmatrix} a & 2 & 1 & 2 \\ 2 & a & 2 & 1 \\ 1 & 2 & a & 2 \\ a+5 & a+5 & a+5 & a+5 \end{vmatrix} = \begin{vmatrix} a & 2-a & 1-a & 2-a \\ 2 & a-2 & 0 & -1 \\ 1 & 1 & a-1 & 1 \\ a+5 & 0 & 0 & 0 \end{vmatrix} \\
&= -(a+5) \begin{vmatrix} 2-a & 1-a & 2-a \\ a-2 & 0 & -1 \\ 1 & a-1 & 1 \end{vmatrix} = -(a+5) \begin{vmatrix} 2-a & 1-a & 0 \\ a-2 & 0 & 1-a \\ 1 & a-1 & 0 \end{vmatrix} \\
&= (a+5)(1-a) \begin{vmatrix} 2-a & 1-a \\ 1 & a-1 \end{vmatrix} = (a+5)(a-1)^2(a-3).
\end{aligned}$$

If $a=-5$, $a=1$ or $a=3$, then $D(A)=0$ and A is not invertible. It remains to determine the dimensions in these three cases. We do this by solving the system $AX=0$.

When $a=-5$, this system is

$$\left[\begin{array}{cccc|c} -5 & 2 & 1 & 2 & 0 \\ 2 & -5 & 2 & 1 & 0 \\ 1 & 2 & -5 & 2 & 0 \\ 2 & 1 & 2 & -5 & 0 \end{array}\right] \Leftrightarrow \left[\begin{array}{cccc|c} -5 & 2 & 1 & 2 & 0 \\ 12 & -9 & 0 & -3 & 0 \\ -24 & 12 & 0 & 12 & 0 \\ 12 & -3 & 0 & -9 & 0 \end{array}\right]$$

$$\Leftrightarrow \left[\begin{array}{cccc|c} -5 & 2 & 1 & 2 & 0 \\ 12 & -9 & 0 & -3 & 0 \\ 0 & -6 & 0 & 6 & 0 \\ 0 & 6 & 0 & -6 & 0 \end{array}\right] \Leftrightarrow \left[\begin{array}{cccc|c} -5 & 2 & 1 & 2 & 0 \\ 4 & -3 & 0 & -1 & 0 \\ 0 & -1 & 0 & 1 & 0 \end{array}\right] \Leftrightarrow X = t\begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}.$$

Hence, $\dim\ker A=1$ and, by the rank-nullity theorem, $\dim\operatorname{im} A=3$.

When $a=1$, we get

$$\left[\begin{array}{cccc|c} 1 & 2 & 1 & 2 & 0 \\ 2 & 1 & 2 & 1 & 0 \\ 1 & 2 & 1 & 2 & 0 \\ 2 & 1 & 2 & 1 & 0 \end{array}\right] \Leftrightarrow \left[\begin{array}{cccc|c} 1 & 2 & 1 & 2 & 0 \\ 2 & 1 & 2 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right]$$

$$\Leftrightarrow \left[\begin{array}{cccc|c} 1 & 2 & 1 & 2 & 0 \\ 0 & -3 & 0 & -3 & 0 \end{array}\right] \Leftrightarrow X = s\begin{bmatrix} -1 \\ 0 \\ 1 \\ 0 \end{bmatrix} + t\begin{bmatrix} 0 \\ -1 \\ 0 \\ 1 \end{bmatrix}.$$

In this case, $\dim\ker A=\dim\operatorname{im} A=2$.

When $a=3$, the system reads

$$\left[\begin{array}{cccc|c} 3 & 2 & 1 & 2 & 0 \\ 2 & 3 & 2 & 1 & 0 \\ 1 & 2 & 3 & 2 & 0 \\ 2 & 1 & 2 & 3 & 0 \end{array}\right] \Leftrightarrow \left[\begin{array}{cccc|c} 3 & 2 & 1 & 2 & 0 \\ -4 & -1 & 0 & -3 & 0 \\ -8 & -4 & 0 & -4 & 0 \\ -4 & -3 & 0 & -1 & 0 \end{array}\right]$$

$$\Leftrightarrow \left[\begin{array}{cccc|c} 3 & 2 & 1 & 2 & 0 \\ -4 & -1 & 0 & -3 & 0 \\ 0 & -2 & 0 & 2 & 0 \\ 0 & -2 & 0 & 2 & 0 \end{array}\right] \Leftrightarrow \left[\begin{array}{cccc|c} 3 & 2 & 1 & 2 & 0 \\ -4 & -1 & 0 & -3 & 0 \\ 0 & -1 & 0 & 1 & 0 \end{array}\right] \Leftrightarrow X = t\begin{bmatrix} -1 \\ 1 \\ -1 \\ 1 \end{bmatrix}.$$

This time, $\dim\ker A=1$ and $\dim\operatorname{im} A=3$.

To sum up, we have found that $\dim\ker A=1$, $\dim\operatorname{im} A=3$ when $a=-5$ or $a=3$, $\dim\ker A=\dim\operatorname{im} A=2$ when $a=1$, and $\dim\ker A=0$, $\dim\operatorname{im} A=4$ for all other values of a.

Example 4.21. A plane is parallel to the vectors $\boldsymbol{u}=(1,2,3)$ and $\boldsymbol{v}=(1,-1,2)$ and passes through the point $Q=(2,1,4)$. A point $P=(x,y,z)$ lies in the plane if and only if the vectors $\overrightarrow{QP}=(x-2,y-1,z-4)$, $\boldsymbol{u}$ and $\boldsymbol{v}$ are linearly dependent. This in turn is equivalent to the determinant of the matrix with columns equal to the coordinate vectors being zero. Hence, the equation of the plane is

$$0 = \begin{vmatrix} x-2 & 1 & 1 \\ y-1 & 2 & -1 \\ z-4 & 3 & 2 \end{vmatrix} = (x-2)\begin{vmatrix} 2 & -1 \\ 3 & 2 \end{vmatrix} + (y-1)(-1)\begin{vmatrix} 1 & 1 \\ 3 & 2 \end{vmatrix} + (z-4)\begin{vmatrix} 1 & 1 \\ 2 & -1 \end{vmatrix}$$
$$= 7(x-2) + 1(y-1) - 3(z-4) = 7x + y - 3z - 3.$$

Theorem 4.22 (Cramer's rule). Let A be a square matrix of order n and assume that $AX = Y$ where X and Y are column vectors. Then

$$D(A_1, \ldots, A_{i-1}, Y, A_{i+1}, \ldots, A_n) = x_i D(A).$$

In particular, if A is invertible, then the unique solution of the system $AX = Y$ is given by

$$x_i = \frac{D(A_1, \ldots, A_{i-1}, Y, A_{i+1}, \ldots, A_n)}{D(A)}, \quad i = 1, \ldots, n.$$

Proof. Let $d = D(A_1, \ldots, A_{i-1}, Y, A_{i+1}, \ldots, A_n)$. The equality $AX = Y$ can be written as $Y = x_1 A_1 + \cdots + x_n A_n$. Hence,

$$d = D(A_1, \ldots, A_{i-1}, \sum_{k=1}^{n} x_k A_k, A_{i+1}, \ldots, A_n) = \sum_{k=1}^{n} x_k D(A_1, \ldots, A_{i-1}, A_k, A_{i+1}, \ldots, A_n)$$
$$= x_i D(A_1, \ldots, A_{i-1}, A_i, A_{i+1}, \ldots, A_n) = x_i D(A).$$

The second equality follows from the linearity in the ith argument and the third equality follows from the fact that the kth determinant in the sum has two equal columns when $k \neq i$. The last statement of the theorem should now be obvious. ■

Definition 4.23. Let A be an $n \times n$ matrix. The adjugate $\tilde{A}$ of A is I if $n = 1$ and otherwise

$$\tilde{A} = \Big[(-1)^{k+i} D(A_{ki})\Big]_{n \times n} = \begin{bmatrix} (-1)^{1+1} D(A_{11}) & \cdots & (-1)^{n+1} D(A_{n1}) \\ \vdots & & \vdots \\ (-1)^{1+n} D(A_{1n}) & \cdots & (-1)^{n+n} D(A_{nn}) \end{bmatrix}.$$

Note the reversal of indices.

Theorem 4.24. Let $A = [a_{ik}]_{n \times n}$ be a square matrix. Then $A\tilde{A} = \tilde{A}A = D(A)I$. In particular, if A is invertible, then

$$A^{-1} = \frac{1}{D(A)} \tilde{A}.$$

Proof. The assertions are trivial for $n = 1$. Otherwise, the ikth entry of $A\tilde{A}$ is

$$b_{ik} = \sum_{j=1}^{n} a_{ij}(-1)^{k+j} D(A_{kj}).$$

If $k = i$, this sum is the expansion of $D(A)$ along the ith row. Hence, $b_{ii} = D(A)$ for $i = 1, \ldots, n$. If $k \neq i$, let A' be the matrix obtained from A by replacing the kth row by the ith row and leaving all other rows unchanged. Then $A'_{kj} = A_{kj}$, and hence

$$b_{ik} = \sum_{j=1}^{n} a_{ij}(-1)^{k+j} D(A'_{kj}).$$

This is the expansion of $D(A')$ along the ith row. Since two rows of A' are equal, we have $D(A') = 0$. Hence, $b_{ik} = D(A') = 0$ if $i \neq k$. This shows that $A\tilde{A} = D(A)I$. It follows from the definition of the adjugate that $\tilde{A}^t = \widetilde{A^t}$. Hence

$$(\tilde{A}A)^t = A^t\tilde{A}^t = A^t\widetilde{A^t} = D(A^t)I = D(A)I,$$

and consequently, $\tilde{A}A = D(A)I$.

If A is invertible, then $D(A) \neq 0$. Therefore

$$A\left(\frac{1}{D(A)}\tilde{A}\right) = I,$$

and the last assertion follows. ■

The methods mentioned earlier for solving systems of equations and finding inverses usually involve much shorter calculations than the methods of the last two theorems. One exception to this is the following formula for the inverse of an invertible 2×2 matrix:

$$\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}^{-1} = \frac{1}{a_{11}a_{22} - a_{12}a_{21}} \begin{bmatrix} a_{22} & -a_{12} \\ -a_{21} & a_{11} \end{bmatrix}.$$

Example 4.25. Let A be an invertible square matrix with integer entries. Then A^{-1} has integer entries if and only if $D(A) = \pm 1$.

First assume that A^{-1} has integer entries. Then $D(A)$ and $D(A^{-1})$ are integers. By the product theorem, $D(A)D(A^{-1}) = 1$, and hence $D(A) = \pm 1$.

Since the entries of A are integers, also the entries of its adjugate are integers. If $D(A) = \pm 1$, it therefore follows from Theorem 4.24 that the entries of A^{-1} are integers.

Exercises

4.1. Evaluate the following determinants.

$$\text{(a)} \begin{vmatrix} 1 & 1 & 2 \\ 3 & 2 & 9 \\ 7 & 2 & 5 \end{vmatrix}, \qquad \text{(b)} \begin{vmatrix} 666 & 667 & 669 \\ 667 & 668 & 670 \\ 669 & 670 & 671 \end{vmatrix}, \qquad \text{(c)} \begin{vmatrix} 1 & a & b+c \\ 1 & b & c+a \\ 1 & c & a+b \end{vmatrix}.$$

4.2. Evaluate the following determinants.

$$\text{(a)} \begin{vmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 2 & 1 \\ 1 & 3 & 1 & 1 \\ 4 & 1 & 1 & 1 \end{vmatrix}, \qquad \text{(b)} \begin{vmatrix} 1 & 2 & 3 & 4 \\ 4 & 1 & 2 & 3 \\ 3 & 4 & 1 & 2 \\ 2 & 3 & 4 & 1 \end{vmatrix}, \qquad \text{(c)} \begin{vmatrix} 0 & 0 & a & b \\ 0 & a & b & b \\ a & b & b & b \\ b & b & b & b \end{vmatrix}.$$

4.3. Solve the following equations.

$$\text{(a)} \begin{vmatrix} x & 0 & 0 & 2 \\ 0 & x & 0 & 2 \\ 0 & 0 & x & 2 \\ 2 & 2 & 2 & x \end{vmatrix} = 0, \qquad \text{(b)} \begin{vmatrix} x & 1 & 1 & 1 \\ 1 & x & 1 & 1 \\ 1 & 1 & x & 1 \\ 1 & 1 & 1 & x \end{vmatrix} = 0.$$

4.4. The numbers 17887, 22041, 34503, 46159, 54777 are divisible by 31. Show that the determinant

$$\begin{vmatrix} 1 & 7 & 8 & 8 & 7 \\ 2 & 2 & 0 & 4 & 1 \\ 3 & 4 & 5 & 0 & 3 \\ 4 & 6 & 1 & 5 & 9 \\ 5 & 4 & 7 & 7 & 7 \end{vmatrix}$$

is divisible by 31.

4.5. Show that the determinant of a skew-symmetric matrix of odd order equals zero.

4.6. Evaluate the following determinants of order n.

$$\text{(a)} \begin{vmatrix} 1 & 1 & 1 & \cdots & 1 & 1 \\ 1 & 2 & 1 & \cdots & 1 & 1 \\ 1 & 1 & 3 & \cdots & 1 & 1 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 1 & 1 & 1 & \cdots & n-1 & 1 \\ 1 & 1 & 1 & \cdots & 1 & n \end{vmatrix}, \qquad \text{(b)} \begin{vmatrix} 1 & 1 & 1 & \cdots & 1 & 1 \\ 1 & 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & 0 \\ 0 & 0 & 0 & \cdots & 1 & 1 \end{vmatrix},$$

$$\text{(c)} \begin{vmatrix} 1 & 2 & 0 & 0 & \cdots & 0 & 0 & 0 \\ 1 & 3 & 2 & 0 & \cdots & 0 & 0 & 0 \\ 0 & 1 & 3 & 2 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 1 & 3 & 2 \\ 0 & 0 & 0 & 0 & \cdots & 0 & 1 & 3 \end{vmatrix}, \qquad \text{(d)} \begin{vmatrix} 0 & 1 & 0 & 0 & \cdots & 0 & 0 & 0 \\ 2 & 0 & 1 & 0 & \cdots & 0 & 0 & 0 \\ 0 & 2 & 0 & 1 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 2 & 0 & 1 \\ 0 & 0 & 0 & 0 & \cdots & 0 & 2 & 0 \end{vmatrix}.$$

4.7. Evaluate the determinant

$$\begin{vmatrix} 0 & 1 & 2 & 3 & \cdots & n-1 \\ 1 & 0 & 1 & 2 & \cdots & n-2 \\ 2 & 1 & 0 & 1 & \cdots & n-3 \\ \vdots & \vdots & \vdots & \vdots & & \vdots \\ n-1 & n-2 & n-3 & n-4 & \cdots & 0 \end{vmatrix}.$$

4.8. Show that
$$\begin{vmatrix} 1 & 2 & 3 & \cdots & n \\ 2 & 3 & 4 & \cdots & 1 \\ 3 & 4 & 5 & \cdots & 2 \\ \vdots & \vdots & \vdots & & \vdots \\ n & 1 & 2 & \cdots & n-1 \end{vmatrix} = (-1)^{\frac{n(n-1)}{2}} \frac{n^n + n^{n-1}}{2}.$$

4.9. Show that
$$\begin{vmatrix} x & 0 & 0 & 0 & \cdots & 0 & a_0 \\ -1 & x & 0 & 0 & \cdots & 0 & a_1 \\ 0 & -1 & x & 0 & \cdots & 0 & a_2 \\ \vdots & \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & -1 & a_{n-1}+x \end{vmatrix} = x^n + a_{n-1}x^{n-1} + \cdots + a_0.$$

4.10. (a) Let x_1, x_2, y_1, y_2 be real numbers. Show that
$$\begin{vmatrix} 1 & -(x_1+x_2) & x_1x_2 & 0 \\ 0 & 1 & -(x_1+x_2) & x_1x_2 \\ 1 & -(y_1+y_2) & y_1y_2 & 0 \\ 0 & 1 & -(y_1+y_2) & y_1y_2 \end{vmatrix} = (x_1-y_1)(x_1-y_2)(x_2-y_1)(x_2-y_2).$$

(b) Show that the quadratic polynomials $a_0x^2 + a_1x + a_2$ and $b_0x^2 + b_1x + b_2$ have a common zero if and only if
$$\begin{vmatrix} a_0 & a_1 & a_2 & 0 \\ 0 & a_0 & a_1 & a_2 \\ b_0 & b_1 & b_2 & 0 \\ 0 & b_0 & b_1 & b_2 \end{vmatrix} = 0.$$

Hint: Use the relationship between roots and coefficients.

4.11. The Vandermonde determinant of order n is
$$V_n(x_1, x_2, \ldots, x_n) = \begin{vmatrix} 1 & x_1 & x_1^2 & \cdots & x_1^{n-1} \\ 1 & x_2 & x_2^2 & \cdots & x_2^{n-1} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_n & x_n^2 & \cdots & x_n^{n-1} \end{vmatrix}.$$

Show by induction on n that
$$V_n(x_1, x_2, \ldots, x_n) = \prod_{j<k} (x_k - x_j).$$

Hint for the induction step: Perform column operations to obtain a determinant with zeros in all positions of the first row except the first position.

4.12. Suppose that there exist invertible $n \times n$ matrices A and B such that $AB = -BA$. Show that n is even.

4.13. Find, for every real constant a, the dimensions of $\operatorname{im} A$ and $\ker A$ where

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

4.14. Show that the matrix

$$A = \begin{bmatrix} 1 & 3 & 4 & 2 \\ 4 & 1 & 2 & 3 \\ 2 & 3 & 1 & 4 \\ 3 & 2 & 4 & 1 \end{bmatrix}$$

is invertible and compute $\det A^{-1}$.

4.15. For which values of x do the following vectors form a basis for $\mathbf{R}^4$?

$$(2,1,1,x), \quad (1,2,x,1), \quad (1,x,2,1), \quad (x,1,1,2).$$

4.16. Let

$$A = \begin{bmatrix} a & b & c & d \\ b & -a & d & -c \\ c & -d & -a & b \\ d & c & -b & -a \end{bmatrix}.$$

Show that $\det A = 0$ only if $a = b = c = d = 0$. Hint: Consider AA^t.

4.17. Let V be the volume of the parallelepiped spanned by the vectors $\boldsymbol{u}$, $\boldsymbol{v}$ and $\boldsymbol{w}$ in three-space. Show that

$$V^2 = \begin{vmatrix} \langle \boldsymbol{u}, \boldsymbol{u} \rangle & \langle \boldsymbol{u}, \boldsymbol{v} \rangle & \langle \boldsymbol{u}, \boldsymbol{w} \rangle \\ \langle \boldsymbol{v}, \boldsymbol{u} \rangle & \langle \boldsymbol{v}, \boldsymbol{v} \rangle & \langle \boldsymbol{v}, \boldsymbol{w} \rangle \\ \langle \boldsymbol{w}, \boldsymbol{u} \rangle & \langle \boldsymbol{w}, \boldsymbol{v} \rangle & \langle \boldsymbol{w}, \boldsymbol{w} \rangle \end{vmatrix}.$$

Hint: Choose an orthonormal basis and use the product theorem.

4.18. Use Theorem 4.24 to find the inverses of the following matrices.

$$\text{(a)} \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}, \qquad \text{(b)} \begin{bmatrix} 1 & 1 & 2 \\ 0 & 3 & 1 \\ 2 & 1 & 0 \end{bmatrix}.$$

5 Linear Transformations

According to Definition 2.64 on page 26, a linear transformation F from U to V is a function $F : U \to V$ such that

$$F(s\boldsymbol{u} + t\boldsymbol{v}) = sF(\boldsymbol{u}) + tF(\boldsymbol{v})$$

for all $\boldsymbol{u}$ and $\boldsymbol{v}$ in U and all real numbers s and t. If $U = V$, we also say that F is a linear transformation on V. We shall here study linear transformations on a linear space in more detail.

5.1 Matrix Representations of Linear Transformations

Definition 5.1. Let V be a finite-dimensional linear space with basis $\boldsymbol{e}_1, \dots, \boldsymbol{e}_n$ and let F be a linear transformation on V. The $n \times n$ matrix A whose columns are the coordinate vectors of the images $F(\boldsymbol{e}_1), \dots, F(\boldsymbol{e}_n)$ with respect to the basis $\boldsymbol{e}_1, \dots, \boldsymbol{e}_n$ is called the matrix of F with respect to the basis $\boldsymbol{e}_1, \dots, \boldsymbol{e}_n$.

Assume that the coordinates of $\boldsymbol{u}$, $F(\boldsymbol{u})$ and $F(\boldsymbol{e}_k)$, $k = 1, \dots, n$, are

$$X = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}, \quad Y = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} \quad \text{and} \quad A_k = \begin{bmatrix} a_{1k} \\ \vdots \\ a_{nk} \end{bmatrix},$$

respectively. By this assumption and the linearity of F, we have

$$F(\boldsymbol{u}) = F(x_1\boldsymbol{e}_1 + \cdots + x_n\boldsymbol{e}_n) = x_1F(\boldsymbol{e}_1) + \cdots + x_nF(\boldsymbol{e}_n).$$

Hence, by the uniqueness of coordinates,

$$Y = x_1A_1 + \cdots + x_nA_n = AX$$

where A is the matrix of F with respect to the basis $\boldsymbol{e}_1, \dots, \boldsymbol{e}_n$.

Definition 5.2. Let V be a linear space. The linear transformation I on V defined by $I(\boldsymbol{u}) = \boldsymbol{u}$ for $\boldsymbol{u} \in V$ is called the identity mapping on V.

Example 5.3. Let I be the identity mapping on an n-dimensional linear space V and let $\boldsymbol{e}_1, \dots, \boldsymbol{e}_n$ be any basis for V. Since $I(\boldsymbol{e}_k) = \boldsymbol{e}_k$ for $k = 1, \dots, n$, the matrix of I with respect to the basis is the unit matrix I of order n.

Example 5.4. Consider the linear transformation F on P_3 defined by

$$F(\boldsymbol{p}) = \frac{d^2\boldsymbol{p}}{dx^2} - \frac{d\boldsymbol{p}}{dx}.$$

The basis vectors $\boldsymbol{e}_1 = 1$, $\boldsymbol{e}_2 = x$, $\boldsymbol{e}_3 = x^2$, $\boldsymbol{e}_4 = x^3$ are mapped to

$$\begin{aligned}
F(\boldsymbol{e}_1) &= 0, \\
F(\boldsymbol{e}_2) &= -1 = -\boldsymbol{e}_1, \\
F(\boldsymbol{e}_3) &= 2 - 2x = 2\boldsymbol{e}_1 - 2\boldsymbol{e}_2, \\
F(\boldsymbol{e}_4) &= 6x - 3x^2 = 6\boldsymbol{e}_2 - 3\boldsymbol{e}_3.
\end{aligned}$$

Hence, the matrix of F with respect to the given basis is

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

Example 5.5. Let there be given an orthonormal basis $\boldsymbol{e}_1, \boldsymbol{e}_2, \boldsymbol{e}_3$ for three-space and let F be orthogonal projection on the plane $2x_1 + 2x_2 + x_3 = 0$. By Theorem 2.60, F is a linear transformation. The vector $\boldsymbol{e} = \frac{1}{3}(2, 2, 1)$ is a unit normal vector to the plane. The orthogonal projections of the basis vectors on the normal of the plane are

$$\begin{aligned}
\boldsymbol{e}_1'' &= \langle \boldsymbol{e}_1, \boldsymbol{e} \rangle \boldsymbol{e} = \frac{1 \cdot 2 + 0 \cdot 2 + 0 \cdot 1}{3} \cdot \frac{1}{3}(2, 2, 1) = \frac{2}{9}(2, 2, 1), \\
\boldsymbol{e}_2'' &= \langle \boldsymbol{e}_2, \boldsymbol{e} \rangle \boldsymbol{e} = \frac{0 \cdot 2 + 1 \cdot 2 + 0 \cdot 1}{3} \cdot \frac{1}{3}(2, 2, 1) = \frac{2}{9}(2, 2, 1), \\
\boldsymbol{e}_3'' &= \langle \boldsymbol{e}_3, \boldsymbol{e} \rangle \boldsymbol{e} = \frac{0 \cdot 2 + 0 \cdot 2 + 1 \cdot 1}{3} \cdot \frac{1}{3}(2, 2, 1) = \frac{1}{9}(2, 2, 1).
\end{aligned}$$

Hence, their projections on the plane are

$$\begin{aligned}
\boldsymbol{e}_1' &= \boldsymbol{e}_1 - \boldsymbol{e}_1'' = (1, 0, 0) - \frac{2}{9}(2, 2, 1) = \frac{1}{9}(5, -4, -2), \\
\boldsymbol{e}_2' &= \boldsymbol{e}_2 - \boldsymbol{e}_2'' = (0, 1, 0) - \frac{2}{9}(2, 2, 1) = \frac{1}{9}(-4, 5, -2), \\
\boldsymbol{e}_3' &= \boldsymbol{e}_3 - \boldsymbol{e}_3'' = (0, 0, 1) - \frac{1}{9}(2, 2, 1) = \frac{1}{9}(-2, -2, 8),
\end{aligned}$$

and the matrix of F with respect to the given basis is

$$\frac{1}{9}\begin{bmatrix} 5 & -4 & -2 \\ -4 & 5 & -2 \\ -2 & -2 & 8 \end{bmatrix}.$$

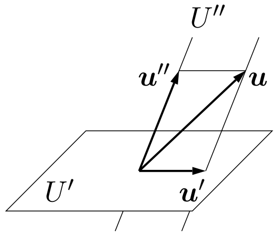
Let F be a linear transformation on V, $\boldsymbol{e}_1, \dots, \boldsymbol{e}_n$ a basis for V and A an $n \times n$ matrix. If $A\boldsymbol{x} = \boldsymbol{y}$ whenever

$$F(x_1\boldsymbol{e}_1 + \cdots + x_n\boldsymbol{e}_n) = y_1\boldsymbol{e}_1 + \cdots + y_n\boldsymbol{e}_n,$$

then A is the matrix of F with respect to the basis. In fact, by setting $\boldsymbol{x} = \boldsymbol{\varepsilon}_i$ we find that the ith column of A is the coordinate vector of $F(\boldsymbol{e}_i)$. We use this observation in the next example.

Example 5.6. Let us find the matrix of projection on the plane $U' : x_1 + 2x_2 + 3x_3 = 0$ along the line $U'' : \boldsymbol{x} = t(2, 1, 2)$ in Example 2.63. Let the coordinates of $\boldsymbol{u}$ be (ξ_1, ξ_2, ξ_3) and form the line $\boldsymbol{x} = (\xi_1, \xi_2, \xi_3) + t(2, 1, 2)$. The image $\boldsymbol{u}'$ of $\boldsymbol{u}$ is the intersection of this line with the plane. The intersection is given by

$$\xi_1 + 2t + 2(\xi_2 + t) + 3(\xi_3 + 2t) = 0 \quad \Leftrightarrow \quad t = -\frac{\xi_1 + 2\xi_2 + 3\xi_3}{10}.$$



Hence,

$$\begin{aligned}
\boldsymbol{u}' &= (\xi_1, \xi_2, \xi_3) + t(2, 1, 2) \\
&= \frac{1}{10}((10\xi_1, 10\xi_2, 10\xi_3) - (2\xi_1 + 4\xi_2 + 6\xi_3, \xi_1 + 2\xi_2 + 3\xi_3, 2\xi_1 + 4\xi_2 + 6\xi_3)) \\
&= \frac{1}{10}(8\xi_1 - 4\xi_2 - 6\xi_3, -\xi_1 + 8\xi_2 - 3\xi_3, -2\xi_1 - 4\xi_2 + 4\xi_3).
\end{aligned}$$

The matrix is therefore

$$\frac{1}{10}\begin{bmatrix} 8 & -4 & -6 \\ -1 & 8 & -3 \\ -2 & -4 & 4 \end{bmatrix}.$$

Theorem 5.7. Let F and G be linear transformations on a linear space V with basis $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n$ and assume that the matrices of F and G are A and B, respectively. Then the composition FG is a linear transformation on V with matrix AB.

Proof. FG is a linear transformation by Theorem 2.69. Assume that $\boldsymbol{w} = FG(\boldsymbol{u})$ where the coordinates of $\boldsymbol{w}$ and $\boldsymbol{u}$ are $\boldsymbol{z}$ and $\boldsymbol{x}$, respectively. Set $\boldsymbol{v} = G(\boldsymbol{u})$ and let the coordinates of $\boldsymbol{v}$ be $\boldsymbol{y}$. Then $\boldsymbol{w} = F(\boldsymbol{v})$, $\boldsymbol{z} = A\boldsymbol{y}$, $\boldsymbol{y} = B\boldsymbol{x}$, and therefore $\boldsymbol{z} = AB\boldsymbol{x}$. ■

From Theorem 2.73 we know that a linear transformation F on a finite-dimensional linear space V is one-to-one if and only if it is onto. In that case Theorem 2.71 yields that F^{-1} is a linear transformation on V.

Theorem 5.8. Let F be a linear transformation on a linear space V with basis $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n$ and let A be its matrix. Then F is invertible if and only if A is invertible, and in that case the matrix of F^{-1} is A^{-1}.

Proof. The invertibility of F means that the equation $F(\boldsymbol{u}) = \boldsymbol{v}$ has a unique solution for every $\boldsymbol{v} \in V$. This is equivalent to the equation $A\boldsymbol{x} = \boldsymbol{y}$ having a unique solution for every $\boldsymbol{y} \in \mathbf{R}^n$, which by Theorem 1.19 means that A is invertible. Assume that F is invertible and let B be the matrix of F^{-1}. Since $FF^{-1} = I$ is the identity mapping, Theorem 5.7 gives that $AB = I$, and hence $B = A^{-1}$. ■

5.2 Change of Basis

Let $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n$ and $\boldsymbol{e}'_1, \ldots, \boldsymbol{e}'_n$ be two bases for the linear space V and suppose that

$$\begin{aligned} \boldsymbol{e}'_1 &= t_{11}\boldsymbol{e}_1 + t_{21}\boldsymbol{e}_2 + \cdots + t_{n1}\boldsymbol{e}_n, \\ \boldsymbol{e}'_2 &= t_{12}\boldsymbol{e}_1 + t_{22}\boldsymbol{e}_2 + \cdots + t_{n2}\boldsymbol{e}_n, \\ &\vdots \\ \boldsymbol{e}'_n &= t_{1n}\boldsymbol{e}_1 + t_{2n}\boldsymbol{e}_2 + \cdots + t_{nn}\boldsymbol{e}_n. \end{aligned}$$

Assume that the coordinates of $\boldsymbol{u}$ with respect to the two bases are $\boldsymbol{x} = (x_1, \ldots, x_n)$ and $\boldsymbol{x}' = (x'_1, \ldots, x'_n)$, respectively. Then

$$\boldsymbol{u} = x_1\boldsymbol{e}_1 + \cdots + x_n\boldsymbol{e}_n \quad \text{and} \quad \boldsymbol{u} = x'_1\boldsymbol{e}'_1 + \cdots + x'_n\boldsymbol{e}'_n.$$

The second equality gives that

$$\begin{aligned} \boldsymbol{u} &= x'_1(t_{11}\boldsymbol{e}_1 + t_{21}\boldsymbol{e}_2 + \cdots + t_{n1}\boldsymbol{e}_n) + \cdots + x'_n(t_{1n}\boldsymbol{e}_1 + t_{2n}\boldsymbol{e}_2 + \cdots + t_{nn}\boldsymbol{e}_n) \\ &= (x'_1 t_{11} + \cdots + x'_n t_{1n})\boldsymbol{e}_1 + \cdots + (x'_1 t_{n1} + \cdots + x'_n t_{nn})\boldsymbol{e}_n. \end{aligned}$$

Since the coordinates of $\boldsymbol{u}$ are unique, the first equality now gives that

$$\begin{aligned} x_1 &= t_{11}x'_1 + t_{12}x'_2 + \cdots + t_{1n}x'_n, \\ x_2 &= t_{21}x'_1 + t_{22}x'_2 + \cdots + t_{2n}x'_n, \\ &\vdots \\ x_n &= t_{n1}x'_1 + t_{n2}x'_2 + \cdots + t_{nn}x'_n. \end{aligned}$$

Hence, $\boldsymbol{x} = T\boldsymbol{x}'$ where T is the matrix whose columns are the coordinate vectors of $\boldsymbol{e}'_1, \ldots, \boldsymbol{e}'_n$ with respect to the basis $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n$. Conversely, if $\boldsymbol{x} = T\boldsymbol{x}'$ whenever

$$x_1\boldsymbol{e}_1 + \cdots + x_n\boldsymbol{e}_n = x'_1\boldsymbol{e}'_1 + \cdots + x'_n\boldsymbol{e}'_n,$$

then the kth column of T must be the coordinate vector of $\boldsymbol{e}'_k$ with respect to $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n$. This can be seen by setting $\boldsymbol{x}' = \boldsymbol{\varepsilon}_k$. We call T the transition matrix from basis $\boldsymbol{e}'_1, \ldots, \boldsymbol{e}'_n$ to basis $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n$.

If S is the transition matrix from a basis $\boldsymbol{e}''_1, \ldots, \boldsymbol{e}''_n$ to $\boldsymbol{e}'_1, \ldots, \boldsymbol{e}'_n$ and if

$$x_1\boldsymbol{e}_1 + \cdots + x_n\boldsymbol{e}_n = x'_1\boldsymbol{e}'_1 + \cdots + x'_n\boldsymbol{e}'_n = x''_1\boldsymbol{e}''_1 + \cdots + x''_n\boldsymbol{e}''_n,$$

then $\boldsymbol{x} = T\boldsymbol{x}' = TS\boldsymbol{x}''$. Hence TS is the transition matrix from $\boldsymbol{e}''_1, \ldots, \boldsymbol{e}''_n$ to $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n$. If $\boldsymbol{e}''_k = \boldsymbol{e}_k$ for $k = 1, \ldots, n$, then $TS = I$ is the transition matrix from $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n$ to itself. This shows that the transition matrix T from $\boldsymbol{e}'_1, \ldots, \boldsymbol{e}'_n$ to $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n$ is invertible and that its inverse $S = T^{-1}$ is the transition matrix from $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n$ to $\boldsymbol{e}'_1, \ldots, \boldsymbol{e}'_n$.

Example 5.9. Consider the bases $\boldsymbol{e}_1 = (1, 2, 1)$, $\boldsymbol{e}_2 = (1, 1, 2)$, $\boldsymbol{e}_3 = (1, 4, 0)$ and $\boldsymbol{\varepsilon}_1, \boldsymbol{\varepsilon}_2, \boldsymbol{\varepsilon}_3$ for $\mathbf{R}^3$ in Example 2.52. Here

$$\begin{aligned} \boldsymbol{e}_1 &= 1\boldsymbol{\varepsilon}_1 + 2\boldsymbol{\varepsilon}_2 + 1\boldsymbol{\varepsilon}_3, \\ \boldsymbol{e}_2 &= 1\boldsymbol{\varepsilon}_1 + 1\boldsymbol{\varepsilon}_2 + 2\boldsymbol{\varepsilon}_3, \\ \boldsymbol{e}_3 &= 1\boldsymbol{\varepsilon}_1 + 4\boldsymbol{\varepsilon}_2 + 0\boldsymbol{\varepsilon}_3. \end{aligned}$$

The transition matrix from $\boldsymbol{e}_1, \boldsymbol{e}_2, \boldsymbol{e}_3$ to $\boldsymbol{\varepsilon}_1, \boldsymbol{\varepsilon}_2, \boldsymbol{\varepsilon}_3$ is therefore

$$T = \begin{bmatrix} 1 & 1 & 1 \\ 2 & 1 & 4 \\ 1 & 2 & 0 \end{bmatrix}.$$

In order to find the coordinates $\boldsymbol{x} = (x_1, x_2, x_3)$ of $\boldsymbol{u} = (3, 7, 3) = 3\boldsymbol{\varepsilon}_1 + 7\boldsymbol{\varepsilon}_2 + 3\boldsymbol{\varepsilon}_3$ with respect to $\boldsymbol{e}_1, \boldsymbol{e}_2, \boldsymbol{e}_3$, we can solve the system

$$T \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 3 \\ 7 \\ 3 \end{bmatrix}.$$

We did that in Example 2.52, where we obtained $\boldsymbol{x} = (1, 1, 1)$. Another possibility would be to compute the transition matrix

$$T^{-1} = \begin{bmatrix} 8 & -2 & -3 \\ -4 & 1 & 2 \\ -3 & 1 & 1 \end{bmatrix}$$

from $\boldsymbol{\varepsilon}_1, \boldsymbol{\varepsilon}_2, \boldsymbol{\varepsilon}_3$ to $\boldsymbol{e}_1, \boldsymbol{e}_2, \boldsymbol{e}_3$ and then get the coordinates

$$\begin{bmatrix} 8 & -2 & -3 \\ -4 & 1 & 2 \\ -3 & 1 & 1 \end{bmatrix} \begin{bmatrix} 3 \\ 7 \\ 3 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

by matrix multiplication.

Theorem 5.10. Let V be an inner product space. If T is the transition matrix from an orthonormal basis $\boldsymbol{e}'_1, \ldots, \boldsymbol{e}'_n$ to an orthonormal basis $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n$, then T is orthogonal.

Proof. Let $\boldsymbol{t}_i = (t_{1i}, \ldots, t_{ni})$ be the coordinates of $\boldsymbol{e}'_i$ with respect to $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n$. Since $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n$ is an orthonormal basis, it follows from Theorem 3.18 that $\langle \boldsymbol{e}'_i, \boldsymbol{e}'_j \rangle = \boldsymbol{t}_i \cdot \boldsymbol{t}_j$. Since also $\boldsymbol{e}'_1, \ldots, \boldsymbol{e}'_n$ is an orthonormal basis, we find that $\boldsymbol{t}_1, \ldots, \boldsymbol{t}_n$, and hence the columns of T, form an orthonormal set in $\mathbf{R}^n$. ■

Let F be a linear transformation on a linear space V and assume that the matrix of F is A with respect to a basis $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n$ for V. Then $A\boldsymbol{x} = \boldsymbol{y}$ whenever

$$F(x_1\boldsymbol{e}_1 + \cdots + x_n\boldsymbol{e}_n) = y_1\boldsymbol{e}_1 + \cdots + y_n\boldsymbol{e}_n.$$

Let us introduce a new basis $\boldsymbol{e}'_1, \ldots, \boldsymbol{e}'_n$ for V. Denote the transition matrix from $\boldsymbol{e}'_1, \ldots, \boldsymbol{e}'_n$ to $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n$ by T. If

$$F(x'_1\boldsymbol{e}'_1 + \cdots + x'_n\boldsymbol{e}'_n) = y'_1\boldsymbol{e}'_1 + \cdots + y'_n\boldsymbol{e}'_n,$$

then $AT\boldsymbol{x}' = T\boldsymbol{y}'$ and hence $T^{-1}AT\boldsymbol{x}' = \boldsymbol{y}'$. This shows that the matrix of F with respect to the new basis is $A' = T^{-1}AT$. Thus we have proved the following theorem.

Theorem 5.11. Let $F : V \to V$ be a linear transformation with matrix A with respect to a basis $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n$ and let T be the transition matrix from a basis $\boldsymbol{e}'_1, \ldots, \boldsymbol{e}'_n$ to $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n$. Then the matrix of F with respect to $\boldsymbol{e}'_1, \ldots, \boldsymbol{e}'_n$ is

$$A' = T^{-1}AT. \tag{5.1}$$

When V is an inner product space and both bases are orthonormal, it follows from Theorem 5.10 that (5.1) can also be written as $A' = T^tAT$.

If A and A' are the matrices of F with respect to two bases, then $A' = T^{-1}AT$ for some matrix T, and hence $\det A' = \det(T^{-1}AT) = \det T^{-1} \det A \det T = \det A$. This justifies the following definition.

Definition 5.12. Let F be a linear transformation on a finite-dimensional non-zero linear space V. We define the determinant $\det F$ as $\det A$ where A is the matrix of F with respect to any basis for V.

There is no natural way to define what it means for a basis for a general linear space to be positively oriented. We can, however, define what it means for two bases to have the same orientation:

Definition 5.13. Let $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n$ and $\boldsymbol{e}'_1, \ldots, \boldsymbol{e}'_n$ be two bases for the same linear space V, and let T be the transition matrix from $\boldsymbol{e}'_1, \ldots, \boldsymbol{e}'_n$ to $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n$. We say that $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n$ has the same orientation as $\boldsymbol{e}'_1, \ldots, \boldsymbol{e}'_n$ if $\det T > 0$. If $\det T < 0$, we say that the orientation of $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n$ is opposite to that of $\boldsymbol{e}'_1, \ldots, \boldsymbol{e}'_n$.

The following theorem follows directly from the fact that $\det I = 1$, $\det(T^{-1}) = (\det T)^{-1}$ and $\det(TS) = (\det T)(\det S)$ for transition matrices S and T.

Theorem 5.14.

(i) Every basis has the same orientation as itself.
(ii) If $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n$ has the same orientation as $\boldsymbol{e}'_1, \ldots, \boldsymbol{e}'_n$, then $\boldsymbol{e}'_1, \ldots, \boldsymbol{e}'_n$ has the same orientation as $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n$.
(iii) If $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n$ has the same orientation as $\boldsymbol{e}'_1, \ldots, \boldsymbol{e}'_n$ and $\boldsymbol{e}'_1, \ldots, \boldsymbol{e}'_n$ has the same orientation as $\boldsymbol{e}''_1, \ldots, \boldsymbol{e}''_n$, then $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n$ has the same orientation as $\boldsymbol{e}''_1, \ldots, \boldsymbol{e}''_n$.
(iv) If the orientation of $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n$ is opposite to that of $\boldsymbol{e}'_1, \ldots, \boldsymbol{e}'_n$ and the orientation of $\boldsymbol{e}'_1, \ldots, \boldsymbol{e}'_n$ is opposite to that of $\boldsymbol{e}''_1, \ldots, \boldsymbol{e}''_n$, then $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n$ has the same orientation as $\boldsymbol{e}''_1, \ldots, \boldsymbol{e}''_n$.

Let F be an invertible linear transformation on a finite-dimensional non-zero linear space V. If $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n$ form a basis for V, then also $F(\boldsymbol{e}_1), \ldots, F(\boldsymbol{e}_n)$ form a basis for V, and the transition matrix from $F(\boldsymbol{e}_1), \ldots, F(\boldsymbol{e}_n)$ to $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n$ equals the matrix of F with respect to $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n$. Hence, the two bases have the same orientation if $\det F > 0$ and opposite orientations if $\det F < 0$.

Definition 5.15. We say that a linear transformation F on a finite-dimensional non-zero linear space is orientation-preserving if $\det F > 0$.

5.3 Projections and Reflections

Suppose that $V = U' \oplus U''$ is the direct sum of two subspaces of V and let $\boldsymbol{u}' \in U'$ and $\boldsymbol{u}'' \in U''$ be the projections of $\boldsymbol{u}$ defined on page 24. The function $P : V \to V$ defined by $P(\boldsymbol{u}) = \boldsymbol{u}'$ for $\boldsymbol{u} \in V$ is called the projection of V on U' along U''. By Theorem 2.60, P is a linear transformation on V.

Since $P(\boldsymbol{u}' + \boldsymbol{u}'') = \boldsymbol{u}'$ for all $\boldsymbol{u}' \in U'$ and all $\boldsymbol{u}'' \in U''$, we see that $\operatorname{im} P = U'$ and $\ker P = U''$.

Theorem 5.16. Let P be a linear transformation on a linear space V. Then P is a projection if and only if $P^2 = P$.

Proof. Suppose that P is projection on U' along U'' and let $\boldsymbol{u} \in V$. Then

$$P^2(\boldsymbol{u}) = P(P(\boldsymbol{u})) = P(P(\boldsymbol{u}' + \boldsymbol{u}'')) = P(\boldsymbol{u}') = P(\boldsymbol{u}).$$

To show the converse, we assume that $P^2 = P$. Let $\boldsymbol{u}$ be any vector of V and set $\boldsymbol{u}' = P(\boldsymbol{u})$ and $\boldsymbol{u}'' = \boldsymbol{u} - \boldsymbol{u}'$. Then $\boldsymbol{u} = \boldsymbol{u}' + \boldsymbol{u}''$ and $\boldsymbol{u}' \in \operatorname{im} P$. Since

$$P(\boldsymbol{u}'') = P(\boldsymbol{u}) - P(\boldsymbol{u}') = P(\boldsymbol{u}) - P^2(\boldsymbol{u}) = P(\boldsymbol{u}) - P(\boldsymbol{u}) = \mathbf{0},$$

we also see that $\boldsymbol{u}'' \in \ker P$. Hence $V = \operatorname{im} P + \ker P$. We show that $V = \operatorname{im} P \oplus \ker P$ by showing that $\operatorname{im} P \cap \ker P = \{\mathbf{0}\}$. If $\boldsymbol{u} \in \operatorname{im} P \cap \ker P$, then $\boldsymbol{u} = P(\boldsymbol{v})$ for some $\boldsymbol{v} \in V$ and $\boldsymbol{u} = P(\boldsymbol{v}) = P^2(\boldsymbol{v}) = P(\boldsymbol{u}) = \mathbf{0}$. Hence, $V = \operatorname{im} P \oplus \ker P$ by Theorem 2.61. Now let $\boldsymbol{u} = \boldsymbol{u}' + \boldsymbol{u}''$ where $\boldsymbol{u}' \in \operatorname{im} P$ and $\boldsymbol{u}'' \in \ker P$. Then $\boldsymbol{u}' = P(\boldsymbol{v})$ for some $\boldsymbol{v} \in V$, and therefore $P(\boldsymbol{u}) = P(\boldsymbol{u}') + P(\boldsymbol{u}'') = P(P(\boldsymbol{v})) = P(\boldsymbol{v}) = \boldsymbol{u}'$. This shows that P is projection on $\operatorname{im} P$ along $\ker P$. ■

Corollary 5.17. Let V be a linear space with basis $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n$ and let A be the matrix of a linear transformation P on V. Then P is a projection if and only if $A^2 = A$.

Note that if P is projection on U' along U'', then $I - P$ is projection on U'' along U'. Hence $U' = \operatorname{im} P = \ker(I - P) = \ker(P - I)$. We use this observation in the next example.

Example 5.18. We set out to show that

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

is the matrix of a projection on U' along U'' where U' and U'' are subspaces of $\mathbf{R}^3$. By the corollary, it is sufficient to show that $A^2 = A$, and indeed,

$$A^2 = \frac{1}{9}\begin{bmatrix} 4 & 3 & 2 \\ -2 & -3 & -4 \\ 1 & 3 & 5 \end{bmatrix}\begin{bmatrix} 4 & 3 & 2 \\ -2 & -3 & -4 \\ 1 & 3 & 5 \end{bmatrix} = \frac{1}{9}\begin{bmatrix} 12 & 9 & 6 \\ -6 & -9 & -12 \\ 3 & 9 & 15 \end{bmatrix} = A.$$

Since $U' = \ker(P - I)$, we get U' by solving the system $(A-I)\boldsymbol{x} = \mathbf{0} \Leftrightarrow 3(A-I)\boldsymbol{x} = \mathbf{0}$.

$$\left[\begin{array}{ccc|c} 1 & 3 & 2 & 0 \\ -2 & -6 & -4 & 0 \\ 1 & 3 & 2 & 0 \end{array}\right] \quad \Leftrightarrow \quad \left[\begin{array}{ccc|c} 1 & 3 & 2 & 0 \end{array}\right].$$

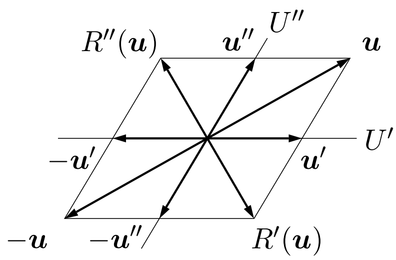
Hence, U' is the plane $x_1 + 3x_2 + 2x_3 = 0$. In order to find $U'' = \ker A$, we solve the system $A\boldsymbol{x} = \mathbf{0}$.

$$\left[\begin{array}{ccc|c} 4 & 3 & 2 & 0 \\ -2 & -3 & -4 & 0 \\ 1 & 3 & 5 & 0 \end{array}\right] \quad \Leftrightarrow \quad \left[\begin{array}{ccc|c} 4 & 3 & 2 & 0 \\ 2 & 0 & -2 & 0 \\ -3 & 0 & 3 & 0 \end{array}\right] \quad \Leftrightarrow \quad \boldsymbol{x} = t(1, -2, 1).$$

Thus we have shown that A is the matrix of projection on the plane $x_1 + 3x_2 + 2x_3 = 0$ along the line $\boldsymbol{x} = t(1, -2, 1)$.

Definition 5.19. Assume that $V = U' \oplus U''$ and let P' be projection on U' along U''. The linear transformation $R' = 2P' - I$ is called reflection in U' along U''.

Let P'' be projection on U'' along U' and let R'' be reflection in U'' along U'. Then $I = P' + P''$, and hence $R' = 2(I - P'') - I = I - 2P'' = -R''$. This means that R' also can be described as reflection in U'' along U' followed by reflection in the origin.



Theorem 5.20. Let R be a linear transformation on a linear space V. Then R is a reflection if and only if $R^2 = I$.

Proof. Assume that $R = 2P - I$ is a reflection. Then $R^2 = 4P^2 - 4P + I = I$ since $P^2 = P$. Assume that $R^2 = I$, and set $P = \frac{1}{2}(R + I)$. Then $R = 2P - I$ and P is a projection since $P^2 = \frac{1}{4}(R^2 + 2R + I) = \frac{1}{2}(R + I) = P$. Hence R is a reflection. ■

Corollary 5.21. Let V be a linear space with basis $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n$ and let A be the matrix of a linear transformation R on V. Then R is a reflection if and only if $A^2 = I$.

Theorem 5.22. Assume that $V = U' \oplus U''$ and let R be reflection in U' along U''. Then $U' = \ker(R - I)$ and $U'' = \ker(R + I)$.

Proof. The statement follows from the fact that $\ker(R-I) = \ker(-2P'') = \ker(P'') = U'$ and $\ker(R + I) = \ker(2P') = \ker(P') = U''$. ■

Theorem 5.23. Let $V = U' \oplus U''$ and assume that $n = \dim V > 0$ and $k = \dim U'$. The determinant of the reflection F in U' along U'' is then $(-1)^{n-k}$.

Proof. If $k = n$, then $F = I$, and the statement is true. If $k = 0$, then $F = -I$, and $\det F = \det(-I) = (-1)^n \det I = (-1)^n$. Otherwise, we can choose a basis $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_k$ for U' and a basis $\boldsymbol{e}_{k+1}, \ldots, \boldsymbol{e}_n$ for U''. Then $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n$ form a basis for V. We have $F(\boldsymbol{e}_i) = \boldsymbol{e}_i$ for $i = 1, \ldots, k$ and $F(\boldsymbol{e}_i) = -\boldsymbol{e}_i$ for $i = k+1, \ldots, n$. The matrix A of F with respect to this basis is a diagonal matrix with k diagonal entries 1 and $n-k$ entries -1. Therefore, $\det F = \det A = 1^k \cdot (-1)^{n-k} = (-1)^{n-k}$. ■

Definition 5.24. Let V be an inner product space and U a subspace. If $V = U \oplus U^\perp$, we call projection on U along $U^\perp$ orthogonal projection on U. We also call reflection in U along $U^\perp$ orthogonal reflection in U.

Lemma 5.25. Assume that $V = U \oplus W$ where V is an inner product space and U and W are subspaces of V. If $\langle \boldsymbol{u}, \boldsymbol{w} \rangle = 0$ for all $\boldsymbol{u} \in U$ and $\boldsymbol{w} \in W$, then $W = U^\perp$.

Proof. It is true that $W \subseteq U^\perp$, for if $\boldsymbol{w} \in W$, then $\langle \boldsymbol{u}, \boldsymbol{w} \rangle = 0$ for all $\boldsymbol{u} \in U$. It remains to show the reverse inclusion. Assume that $\boldsymbol{v} \in U^\perp$. By assumption, $\boldsymbol{v} = \boldsymbol{u} + \boldsymbol{w}$ for some vectors $\boldsymbol{u} \in U$ and $\boldsymbol{w} \in W$. Since $0 = \langle \boldsymbol{u}, \boldsymbol{v} \rangle = \langle \boldsymbol{u}, \boldsymbol{u} \rangle + \langle \boldsymbol{u}, \boldsymbol{w} \rangle = \langle \boldsymbol{u}, \boldsymbol{u} \rangle$, we have $\boldsymbol{u} = \boldsymbol{0}$ and, therefore, $\boldsymbol{v} = \boldsymbol{w} \in W$. Hence $U^\perp \subseteq W$. ■

Definition 5.26. Let F be a linear transformation on an inner product space V. We say that F is symmetric if

$$\langle F(\boldsymbol{u}), \boldsymbol{v} \rangle = \langle \boldsymbol{u}, F(\boldsymbol{v}) \rangle$$

for all vectors $\boldsymbol{u}$ and $\boldsymbol{v}$ in V.

Theorem 5.27. Let F be a linear transformation on an inner product space V. Then

(i) F is an orthogonal projection if and only if F is a projection and F is symmetric,
(ii) F is an orthogonal reflection if and only if F is a reflection and F is symmetric.

Proof. (i) Assume that F is orthogonal projection on U. If $\boldsymbol{u}$ and $\boldsymbol{v}$ are vectors of V, we write $\boldsymbol{u} = \boldsymbol{u}' + \boldsymbol{u}''$ and $\boldsymbol{v} = \boldsymbol{v}' + \boldsymbol{v}''$ where $\boldsymbol{u}'$ and $\boldsymbol{v}'$ belong to U and $\boldsymbol{u}''$ and $\boldsymbol{v}''$ belong to $U^\perp$. Since $F(\boldsymbol{u}) = \boldsymbol{u}'$ and $F(\boldsymbol{v}) = \boldsymbol{v}'$ are orthogonal to $\boldsymbol{u}''$ and $\boldsymbol{v}''$, we have

$$\begin{aligned}
\langle F(\boldsymbol{u}), \boldsymbol{v} \rangle &= \langle \boldsymbol{u}', \boldsymbol{v}' + \boldsymbol{v}'' \rangle = \langle \boldsymbol{u}', \boldsymbol{v}' \rangle + \langle \boldsymbol{u}', \boldsymbol{v}'' \rangle = \langle \boldsymbol{u}', \boldsymbol{v}' \rangle, \\
\langle \boldsymbol{u}, F(\boldsymbol{v}) \rangle &= \langle \boldsymbol{u}' + \boldsymbol{u}'', \boldsymbol{v}' \rangle = \langle \boldsymbol{u}', \boldsymbol{v}' \rangle + \langle \boldsymbol{u}'', \boldsymbol{v}' \rangle = \langle \boldsymbol{u}', \boldsymbol{v}' \rangle.
\end{aligned}$$

Hence, $\langle F(\boldsymbol{u}), \boldsymbol{v} \rangle = \langle \boldsymbol{u}, F(\boldsymbol{v}) \rangle$ for all $\boldsymbol{u}$ and $\boldsymbol{v}$ in V, and F is symmetric.

To prove the converse, we assume that F is a symmetric projection on $\operatorname{im} F$ along $\ker F$ where $V = \operatorname{im} F \oplus \ker F$. Assume that $\boldsymbol{u} = F(\boldsymbol{v}) \in \operatorname{im} F$ and $\boldsymbol{w} \in \ker F$. Then

$$\langle \boldsymbol{u}, \boldsymbol{w} \rangle = \langle F(\boldsymbol{v}), \boldsymbol{w} \rangle = \langle \boldsymbol{v}, F(\boldsymbol{w}) \rangle = \langle \boldsymbol{v}, \boldsymbol{0} \rangle = 0,$$

and it follows from Lemma 5.25 that $\ker F = (\operatorname{im} F)^\perp$. Consequently, F is an orthogonal projection.

(ii) The statement about reflections follows from the fact that $R = 2P - I$ is symmetric if and only if P is symmetric. ■

If we think of vectors $\boldsymbol{x} \in \mathbf{R}^n$ as column matrices, then $\boldsymbol{x} \cdot \boldsymbol{y} = \boldsymbol{x}^t \boldsymbol{y}$.

Lemma 5.28. Let A and B be $n \times n$ matrices. Then $A = B$ if and only if $\boldsymbol{x}^t A \boldsymbol{y} = \boldsymbol{x}^t B \boldsymbol{y}$ for all $\boldsymbol{x}$ and $\boldsymbol{y}$ in $\mathbf{R}^n$.

Proof. If $\boldsymbol{x}^t A \boldsymbol{y} = \boldsymbol{x}^t B \boldsymbol{y}$ for all $\boldsymbol{x}$ and $\boldsymbol{y}$ in $\mathbf{R}^n$, this is true for $\boldsymbol{x} = \boldsymbol{\varepsilon}_i$ and $\boldsymbol{y} = \boldsymbol{\varepsilon}_k$. Hence, $A_{ik} = \boldsymbol{\varepsilon}_i^t A \boldsymbol{\varepsilon}_k = \boldsymbol{\varepsilon}_i^t B \boldsymbol{\varepsilon}_k = B_{ik}$ for all indices i, k. The reverse implication is trivial. ■

Theorem 5.29. Let F be a linear transformation on an inner product space V with orthonormal basis $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n$ and let A be the matrix of F with respect to that basis. Then F is symmetric if and only if A is symmetric.

Proof. If $\boldsymbol{x}$ and $\boldsymbol{y}$ are the coordinates of $\boldsymbol{u}$ and $\boldsymbol{v}$, respectively, then

$$\langle F(\boldsymbol{u}), \boldsymbol{v} \rangle = \langle \boldsymbol{u}, F(\boldsymbol{v}) \rangle \quad \Leftrightarrow \quad (A\boldsymbol{x})^t \boldsymbol{y} = \boldsymbol{x}^t A \boldsymbol{y} \quad \Leftrightarrow \quad \boldsymbol{x}^t A^t \boldsymbol{y} = \boldsymbol{x}^t A \boldsymbol{y}.$$

Hence, F is symmetric if and only if $\boldsymbol{x}^t A^t \boldsymbol{y} = \boldsymbol{x}^t A \boldsymbol{y}$ for all $\boldsymbol{x}$ and $\boldsymbol{y}$ in $\mathbf{R}^n$. By the lemma, this is equivalent to $A^t = A$. ■

Example 5.30. Let there be given an orthonormal basis $\boldsymbol{e}_1, \boldsymbol{e}_2, \boldsymbol{e}_3$ for a 3-dimensional inner product space V and consider the linear transformation F with matrix

$$A = \frac{1}{9}\begin{bmatrix} 7 & 4 & -4 \\ 4 & 1 & 8 \\ -4 & 8 & 1 \end{bmatrix}.$$

We wish to show that F is an orthogonal reflection. Since A is a symmetric matrix and the basis is orthonormal, F is a symmetric linear transformation. Since

$$A^2 = \frac{1}{81}\begin{bmatrix} 7 & 4 & -4 \\ 4 & 1 & 8 \\ -4 & 8 & 1 \end{bmatrix}\begin{bmatrix} 7 & 4 & -4 \\ 4 & 1 & 8 \\ -4 & 8 & 1 \end{bmatrix} = \frac{1}{81}\begin{bmatrix} 81 & 0 & 0 \\ 0 & 81 & 0 \\ 0 & 0 & 81 \end{bmatrix} = I,$$

F is also a reflection. Hence, F is an orthogonal reflection. We also want to find the subspace $U = \ker(F - I)$ of reflection. To do so, we solve the system $(A - I)\boldsymbol{x} = \mathbf{0}$:

$$\left[\begin{array}{ccc|c} -2 & 4 & -4 & 0 \\ 4 & -8 & 8 & 0 \\ -4 & 8 & -8 & 0 \end{array}\right] \quad \Leftrightarrow \quad \left[\begin{array}{ccc|c} 1 & -2 & 2 & 0 \end{array}\right].$$

Hence, F is orthogonal reflection in the plane $x_1 - 2x_2 + 2x_3 = 0$.

Note that the classes of linear transformations we have discussed so far are not exclusive. For example, the identity mapping I on an inner product space V is orthogonal projection on V as well as orthogonal reflection in V.

5.4 Isometries

Definition 5.31. Let F be a linear transformation on an inner product space V. We say that F is an isometry if $\|F(\boldsymbol{u})\| = \|\boldsymbol{u}\|$ for all $\boldsymbol{u} \in V$.

Theorem 5.32. F is an isometry if and only if $\langle F(\boldsymbol{u}), F(\boldsymbol{v})\rangle = \langle \boldsymbol{u}, \boldsymbol{v}\rangle$ for all $\boldsymbol{u}$ and $\boldsymbol{v}$ in V.

Proof. If $\langle F(\boldsymbol{u}), F(\boldsymbol{v})\rangle = \langle \boldsymbol{u}, \boldsymbol{v}\rangle$ for all $\boldsymbol{u}$ and $\boldsymbol{v}$, then in particular,

$$\|F(\boldsymbol{u})\|^2 = \langle F(\boldsymbol{u}), F(\boldsymbol{u})\rangle = \langle \boldsymbol{u}, \boldsymbol{u}\rangle = \|\boldsymbol{u}\|^2, \quad \boldsymbol{u} \in V,$$

and hence F is an isometry.

Suppose that F is an isometry. Then it follows from the identity

$$\|\boldsymbol{u}+\boldsymbol{v}\|^2 = \|\boldsymbol{u}\|^2 + \|\boldsymbol{v}\|^2 + 2\langle \boldsymbol{u}, \boldsymbol{v}\rangle$$

that

$$\begin{aligned} 2\langle F(\boldsymbol{u}), F(\boldsymbol{v})\rangle &= \|F(\boldsymbol{u}) + F(\boldsymbol{v})\|^2 - \|F(\boldsymbol{u})\|^2 - \|F(\boldsymbol{v})\|^2 \\ &= \|F(\boldsymbol{u}+\boldsymbol{v})\|^2 - \|F(\boldsymbol{u})\|^2 - \|F(\boldsymbol{v})\|^2 \\ &= \|\boldsymbol{u}+\boldsymbol{v}\|^2 - \|\boldsymbol{u}\|^2 - \|\boldsymbol{v}\|^2 = 2\langle \boldsymbol{u}, \boldsymbol{v}\rangle, \end{aligned}$$

and therefore $\langle F(\boldsymbol{u}), F(\boldsymbol{v})\rangle = \langle \boldsymbol{u}, \boldsymbol{v}\rangle$ for all $\boldsymbol{u}$ and $\boldsymbol{v}$. ■

Hence, an isometry preserves lengths and inner products. Since

$$\frac{\langle F(\boldsymbol{u}), F(\boldsymbol{v})\rangle}{\|F(\boldsymbol{u})\|\|F(\boldsymbol{v})\|} = \frac{\langle \boldsymbol{u}, \boldsymbol{v}\rangle}{\|\boldsymbol{u}\|\|\boldsymbol{v}\|}$$

for non-zero vectors $\boldsymbol{u}$ and $\boldsymbol{v}$, it also preserves angles.

Theorem 5.33. Let A be the matrix of a linear transformation F on an inner product space V with respect to an orthonormal basis $\boldsymbol{e}_1, \dots, \boldsymbol{e}_n$ for V. Then F is an isometry if and only if A is an orthogonal matrix.

Proof. Let $\boldsymbol{u}$ and $\boldsymbol{v}$ be vectors of V and $\boldsymbol{x}$ and $\boldsymbol{y}$ their coordinates. Since the basis is orthonormal, $\langle \boldsymbol{u}, \boldsymbol{v}\rangle = \boldsymbol{x}^t\boldsymbol{y} = \boldsymbol{x}^t I\boldsymbol{y}$ and $\langle F(\boldsymbol{u}), F(\boldsymbol{v})\rangle = (A\boldsymbol{x})^t A\boldsymbol{y} = \boldsymbol{x}^t A^t A\boldsymbol{y}$. Hence, by Theorem 5.32, F is an isometry if and only if $\boldsymbol{x}^t A^t A\boldsymbol{y} = \boldsymbol{x}^t I\boldsymbol{y}$ for all $\boldsymbol{x}$ and $\boldsymbol{y}$. By Lemma 5.28, this in turn is equivalent to $A^t A = I$. ■

Theorem 5.34. The determinant of an orthogonal matrix equals 1 or -1.

Proof. The assertion follows from the fact that

$$(\det A)^2 = \det A^t \det A = \det (A^t A) = \det I = 1$$

for orthogonal matrices A. ■

Note that A need not be orthogonal even if $\det A = \pm 1$. Hence, the converse of the statement of the theorem is not true.

Corollary 5.35. If F is an isometry on a finite-dimensional non-zero inner product space V, then $\det F = \pm 1$.

Proof. We can choose an orthonormal basis for V. The matrix A of F with respect to this basis is orthogonal. Hence, $\det F = \det A = \pm 1$. ■

Theorem 5.36. Let F be orthogonal reflection in a subspace U of an inner product space $V = U \oplus U^{\perp}$. Then F is an isometry.

Proof. We have $F = I - 2P$ where P is orthogonal projection on $U^{\perp}$. Hence

$$\|F(\boldsymbol{u})\|^2 = \|\boldsymbol{u} - P(\boldsymbol{u}) - P(\boldsymbol{u})\|^2 = \|\boldsymbol{u} - P(\boldsymbol{u})\|^2 + \|P(\boldsymbol{u})\|^2 = \|\boldsymbol{u}\|^2$$

by the Pythagorean theorem. ■

Theorem 5.37. Let F be an isometry on an inner product space V. Then F is an orthogonal reflection if and only if F is symmetric.

Proof. If F is an orthogonal reflection, then F is symmetric by Theorem 5.27. To prove the converse, we assume that F is a symmetric isometry. By the same theorem, it is then sufficient to show that F is a reflection. Hence, by Theorem 5.20, it is sufficient to show that $F^2 = I$, and this follows from the fact that

$$\begin{aligned}\|F^2(\boldsymbol{u}) - \boldsymbol{u}\|^2 &= \|F^2(\boldsymbol{u})\|^2 + \|\boldsymbol{u}\|^2 - 2\langle F^2(\boldsymbol{u}), \boldsymbol{u}\rangle = \|\boldsymbol{u}\|^2 + \|\boldsymbol{u}\|^2 - 2\langle F(\boldsymbol{u}), F(\boldsymbol{u})\rangle \\ &= 2\|\boldsymbol{u}\|^2 - 2\|F(\boldsymbol{u})\|^2 = 0, \quad \boldsymbol{u} \in V.\end{aligned}$$ ■

5.5 Rotations

Definition 5.38. We define a rotation on a two-dimensional inner product space V to be an orientation-preserving isometry on V.

Definition 5.39. If $\theta \in \mathbf{R}$, we define

$$R(\theta) = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}.$$

Theorem 5.40. If θ and φ are real numbers, then

(i) $R(\theta)R(\varphi) = R(\theta + \varphi)$,
(ii) $(R(\theta))^{-1} = R(-\theta)$,
(iii) $R(\theta)R(\varphi) = R(\varphi)R(\theta)$.

Proof. (i) follows from the trigonometric addition formulae:

$$\begin{aligned} R(\theta)R(\varphi) &= \begin{bmatrix} \cos\theta\cos\varphi - \sin\theta\sin\varphi & -\cos\theta\sin\varphi - \sin\theta\cos\varphi \\ \sin\theta\cos\varphi + \cos\theta\sin\varphi & -\sin\theta\sin\varphi + \cos\theta\cos\varphi \end{bmatrix} \\ &= \begin{bmatrix} \cos(\theta+\varphi) & -\sin(\theta+\varphi) \\ \sin(\theta+\varphi) & \cos(\theta+\varphi) \end{bmatrix} = R(\theta+\varphi). \end{aligned}$$

By (i), $R(\theta)R(-\theta) = R(0) = I$ and $R(\theta)R(\varphi) = R(\theta+\varphi) = R(\varphi)R(\theta)$, and this proves the other two statements. ■

Theorem 5.41. Let A be a 2×2 matrix. Then A is orthogonal and $\det A = 1$ if and only if $A = R(\theta)$ for some $\theta \in \mathbf{R}$.

Proof. The simple proof of the fact that $R(\theta)$ is orthogonal and $\det R(\theta) = 1$ for any $\theta \in \mathbf{R}$ is left to the reader. To show the converse, we assume that

$$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$$

is orthogonal and has determinant 1. Since A is orthogonal, $a_{11}^2 + a_{21}^2 = a_{12}^2 + a_{22}^2 = 1$, and hence there exist real numbers θ and φ such that $a_{11} = \cos\theta$, $a_{21} = \sin\theta$, $a_{12} = \sin\varphi$ and $a_{22} = \cos\varphi$. Since $\det A = 1$ and the two columns are orthogonal to each other, we get

$$\begin{aligned} \cos(\theta+\varphi) &= \cos\theta\cos\varphi - \sin\theta\sin\varphi = a_{11}a_{22} - a_{12}a_{21} = 1, \\ \sin(\theta+\varphi) &= \cos\theta\sin\varphi + \sin\theta\cos\varphi = a_{11}a_{12} + a_{21}a_{22} = 0. \end{aligned}$$

This shows that $\varphi = -\theta + 2\pi n$ for some $n \in \mathbf{Z}$, and hence $A = R(\theta)$. ■

Corollary 5.42. Let F be a rotation on a two-dimensional inner product space V, and let $\boldsymbol{e}_1, \boldsymbol{e}_2$ be an orthonormal basis for V. The matrix of F with respect to $\boldsymbol{e}_1, \boldsymbol{e}_2$ then equals $R(\theta)$ for some $\theta \in \mathbf{R}$.

Proof. The matrix A is orthogonal by Theorem 5.33. Since $\det A = \pm 1$ by Theorem 5.34 and $\det A > 0$ by Definition 5.15, we must have $\det A = 1$. The statement therefore follows from Theorem 5.41. ■

Theorem 5.43. Let F be a rotation on a two-dimensional inner product space V, and assume that the matrix of F with respect to some orthonormal basis $\boldsymbol{e}_1, \boldsymbol{e}_2$ for V is $R(\theta)$. Let $\boldsymbol{e}_1', \boldsymbol{e}_2'$ be any orthonormal basis for V. Then the matrix of F with respect to $\boldsymbol{e}_1', \boldsymbol{e}_2'$ is $R(\theta)$ if $\boldsymbol{e}_1', \boldsymbol{e}_2'$ has the same orientation as $\boldsymbol{e}_1, \boldsymbol{e}_2$ and $R(-\theta)$ otherwise.

Proof. Since both bases are orthonormal, the transition matrix T from $\boldsymbol{e}_1', \boldsymbol{e}_2'$ to $\boldsymbol{e}_1, \boldsymbol{e}_2$ is orthogonal. Hence, $\det T = 1$ if the bases have the same orientation. In this case, $T = R(\varphi)$ for some $\varphi \in \mathbf{R}$ by Theorem 5.41, and consequently the matrix of F with respect to $\boldsymbol{e}_1', \boldsymbol{e}_2'$ equals $R(-\varphi)R(\theta)R(\varphi) = R(-\varphi+\theta+\varphi) = R(\theta)$. If they have opposite orientations, then $\boldsymbol{e}_1, \boldsymbol{e}_2$ and $\boldsymbol{e}'_2, \boldsymbol{e}'_1$ have the same orientation. The matrix with respect to $\boldsymbol{e}'_2, \boldsymbol{e}'_1$ is therefore equal to $R(\theta)$. Since the transition matrix from $\boldsymbol{e}'_1, \boldsymbol{e}'_2$ to $\boldsymbol{e}'_2, \boldsymbol{e}'_1$ is

$$S = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix},$$

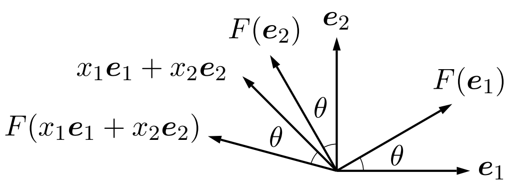
we find that the matrix with respect to $\boldsymbol{e}'_1, \boldsymbol{e}'_2$ is

$$S^{-1}R(\theta)S = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} R(\theta) \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix} = R(-\theta). \blacksquare$$

Theorem 5.44. Let F be a rotation on a two-dimensional inner product space V. Then there exists a unique angle $\theta \in [0, \pi]$ such that the matrix of F with respect to some orthonormal basis for V is $R(\theta)$.

Proof. To prove the existence of θ, we choose any orthonormal basis $\boldsymbol{e}_1, \boldsymbol{e}_2$ for V. By Corollary 5.42 and Theorem 5.43, there exists a $\varphi \in \mathbf{R}$ such that the matrices of F with respect to $\boldsymbol{e}_1, \boldsymbol{e}_2$ and $\boldsymbol{e}_2, \boldsymbol{e}_1$ are $R(\varphi)$ and $R(-\varphi)$, respectively. Either $\theta_1 = \varphi + 2\pi n$ or $\theta_2 = -\varphi + 2\pi n$ belongs to the interval $[0, \pi]$ for some $n \in \mathbf{Z}$, and the existence follows from the fact that $R(\theta_1) = R(\varphi)$ and $R(\theta_2) = R(-\varphi)$. If $R(\theta_1) = R(\theta_2)$ and θ_1 and θ_2 belong to $[0, \pi]$, then $\theta_1 = \theta_2$. This proves the uniqueness. $\blacksquare$

Definition 5.45. Let F be a rotation on a two-dimensional inner product space. The unique angle θ in Theorem 5.44 is called the angle of rotation. If $R(\theta)$ is the matrix of F with respect to the orthonormal basis $\boldsymbol{e}_1, \boldsymbol{e}_2$, we shall say that F is rotation through the angle θ in the direction from $\boldsymbol{e}_1$ towards $\boldsymbol{e}_2$.



Theorem 5.46. Let F be a rotation on a two-dimensional inner product V space, and let θ be its angle of rotation. If $\boldsymbol{u}$ is any non-zero vector in V, then the angle between $\boldsymbol{u}$ and $F(\boldsymbol{u})$ equals θ. If $0 < \theta < \pi$ and the matrix of F with respect to the orthonormal basis $\boldsymbol{e}_1, \boldsymbol{e}_2$ is $R(\theta)$, then $\boldsymbol{u}, F(\boldsymbol{u})$ form a basis for V with the same orientation as $\boldsymbol{e}_1, \boldsymbol{e}_2$.

Proof. We choose an orthonormal basis $\boldsymbol{e}_1, \boldsymbol{e}_2$ for V such that the matrix of F with respect to $\boldsymbol{e}_1, \boldsymbol{e}_2$ is $R(\theta)$. If (x_1, x_2) are the coordinates of $\boldsymbol{u}$ with respect to $\boldsymbol{e}_1, \boldsymbol{e}_2$, then the coordinates of $F(\boldsymbol{u})$ are $(x_1\cos\theta - x_2\sin\theta, x_1\sin\theta + x_2\cos\theta)$. The angle φ between $\boldsymbol{u}$ and $F(\boldsymbol{u})$ is given by

$$\cos\varphi = \frac{\langle \boldsymbol{u}, F(\boldsymbol{u})\rangle}{\|\boldsymbol{u}\|\|F(\boldsymbol{u})\|} = \frac{\langle \boldsymbol{u}, F(\boldsymbol{u})\rangle}{\|\boldsymbol{u}\|^2} = \frac{(x_1^2 + x_2^2)\cos\theta}{x_1^2 + x_2^2} = \cos\theta,$$

and hence $\varphi = \theta$.

Let T be the matrix whose columns are the coordinate vectors of $\boldsymbol{u}$ and $F(\boldsymbol{u})$ with respect to $\boldsymbol{e}_1, \boldsymbol{e}_2$. Then

$$\det T = \begin{vmatrix} x_1 & x_1\cos\theta - x_2\sin\theta \\ x_2 & x_1\sin\theta + x_2\cos\theta \end{vmatrix} = (x_1^2 + x_2^2)\sin\theta > 0.$$

This shows that $\boldsymbol{u}$ and $F(\boldsymbol{u})$ are linearly independent and thus form a basis. Since T then is the transition matrix from $\boldsymbol{u}, F(\boldsymbol{u})$ to $\boldsymbol{e}_1, \boldsymbol{e}_2$, it also shows that these two bases have the same orientation. ■

Note that if $\theta = 0$ or $\theta = \pi$, then $F = \pm I$ where I is the identity mapping. In this case, $\boldsymbol{u}, F(\boldsymbol{u})$ are linearly dependent.

Example 5.47. Let $\boldsymbol{e}_1, \boldsymbol{e}_2$ be an orthonormal basis for two-space and let F be rotation through the angle $\theta = \frac{\pi}{3}$ in the direction from $\boldsymbol{e}_1$ towards $\boldsymbol{e}_2$. We shall here compute the coordinates of the image under F of the vector with coordinates $(1, 2)$. The matrix of F with respect to $\boldsymbol{e}_1, \boldsymbol{e}_2$ is

$$A = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} = \frac{1}{2}\begin{bmatrix} 1 & -\sqrt{3} \\ \sqrt{3} & 1 \end{bmatrix},$$

and thus the coordinates of the image are

$$A\begin{bmatrix} 1 \\ 2 \end{bmatrix} = \frac{1}{2}\begin{bmatrix} 1 - 2\sqrt{3} \\ \sqrt{3} + 2 \end{bmatrix}.$$

Definition 5.48. A linear transformation F on an inner product space V is said to be a planar rotation if there exists a two-dimensional subspace U of V such that the restriction of F to U is a rotation on U and the restriction of F to $U^\perp$ is the identity mapping on $U^\perp$.

A rotation F on a two-dimensional inner product space V is a planar rotation since the restriction of F to $V^\perp = \{\boldsymbol{0}\}$ is the identity mapping. It should also be clear that the identity mapping on any inner product space of dimension greater than 1 is a planar rotation.

Theorem 5.49. Let F be a planar rotation on an inner product space V. Then F is an isometry. If V is finite-dimensional, then $\det F = 1$.

Proof. Let U be a two-dimensional subspace of V as in the definition. If $\boldsymbol{u} \in V$, we can write $\boldsymbol{u} = \boldsymbol{u}' + \boldsymbol{u}''$ where $\boldsymbol{u}' \in U$ and $\boldsymbol{u}'' \in U^\perp$. By the definition, $F(\boldsymbol{u}') \in U$, $\|F(\boldsymbol{u}')\| = \|\boldsymbol{u}'\|$ and $F(\boldsymbol{u}'') = \boldsymbol{u}'' \in U^\perp$. It therefore follows from the Pythagorean theorem that

$$\|F(\boldsymbol{u})\|^2 = \|F(\boldsymbol{u}') + \boldsymbol{u}''\|^2 = \|F(\boldsymbol{u}')\|^2 + \|\boldsymbol{u}''\|^2 = \|\boldsymbol{u}'\|^2 + \|\boldsymbol{u}''\|^2 = \|\boldsymbol{u}' + \boldsymbol{u}''\|^2 = \|\boldsymbol{u}\|^2,$$

and this shows that F is an isometry. The proof of the statement about the determinant is left as an exercise. ■

Theorem 5.50. Let F be a planar rotation on an inner product space V. Unless $F = I$, the subspace U in the definition is uniquely determined.

Proof. Let U_1 and U_2 be two such subspaces. Assume that $\boldsymbol{u} \in U_1^{\perp}$. Then $\boldsymbol{u} = \boldsymbol{u}' + \boldsymbol{u}''$ where $\boldsymbol{u}' \in U_2$ and $\boldsymbol{u}'' \in U_2^{\perp}$. Since F is the identity mapping on $U_1^{\perp}$ and on $U_2^{\perp}$, we get

$$\boldsymbol{u}' + \boldsymbol{u}'' = \boldsymbol{u} = F(\boldsymbol{u}) = F(\boldsymbol{u}') + F(\boldsymbol{u}'') = F(\boldsymbol{u}') + \boldsymbol{u}'',$$

and it follows that $\boldsymbol{u}' = F(\boldsymbol{u}')$. Since the restriction of F to U_2 is a rotation that is not the identity mapping, this is possible only if $\boldsymbol{u}' = \boldsymbol{0}$. Hence, $\boldsymbol{u} = \boldsymbol{u}'' \in U_2^{\perp}$. It follows from this that $U_1^{\perp} \subseteq U_2^{\perp}$. By interchanging the roles of U_1 and U_2, we get the reverse inclusion. Hence, $U_1^{\perp} = U_2^{\perp}$, and it follows from Theorem 3.30 that $U_1 = (U_1^{\perp})^{\perp} = (U_2^{\perp})^{\perp} = U_2$. ■

Definition 5.51. Let F be a planar rotation on an inner product space V. If $F \neq I$, the unique subspace U is called the plane of rotation of F and the angle θ of rotation of the restriction of F to U is called the angle of rotation of F. If $F = I$, we say that the angle of rotation of F is 0. If $\dim V = 3$ and $F \neq I$, we say that F is rotation through the angle θ about the line $U^{\perp}$.

Let F be a rotation on ordinary three-space about the line L and assume that the angle θ of rotation satisfies $0 < \theta < \pi$. If $\boldsymbol{e}_1, \boldsymbol{e}_2, \boldsymbol{e}_3$ form an orthonormal, positively oriented basis such that L is spanned by $\boldsymbol{e}_3$ and the rotation is in the direction from $\boldsymbol{e}_1$ towards $\boldsymbol{e}_2$, the rotation appears anticlockwise when viewed in the opposite direction of $\boldsymbol{e}_3$. By Theorem 5.46, this is the case if, for some non-zero vector $\boldsymbol{u} \in L^{\perp}$, $\boldsymbol{u}, F(\boldsymbol{u}), \boldsymbol{e}_3$ form a positively oriented basis for three-space.

Example 5.52. Let $\boldsymbol{e}_1, \boldsymbol{e}_2, \boldsymbol{e}_3$ be an orthonormal, positively oriented basis for three-space. Let F be rotation about the x_3-axis through an angle θ, and assume that the rotation appears anticlockwise when viewed in the opposite direction of $\boldsymbol{e}_3$. Then

$$\begin{aligned} F(\boldsymbol{e}_1) &= (\cos\theta)\boldsymbol{e}_1 + (\sin\theta)\boldsymbol{e}_2, \\ F(\boldsymbol{e}_2) &= (-\sin\theta)\boldsymbol{e}_1 + (\cos\theta)\boldsymbol{e}_2, \\ F(\boldsymbol{e}_3) &= \boldsymbol{e}_3, \end{aligned}$$

and so the matrix of F with respect to $\boldsymbol{e}_1, \boldsymbol{e}_2, \boldsymbol{e}_3$ is

$$\begin{bmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Example 5.53. Let $\boldsymbol{e}_1, \boldsymbol{e}_2, \boldsymbol{e}_3$ be an orthonormal, positively oriented basis for three-space. Let F be rotation about the line $\boldsymbol{x} = t(2,1,2)$ through the angle $\frac{\pi}{3}$ and suppose that the rotation appears anticlockwise on looking from the point $(2,1,2)$ towards the origin. We set out to find the matrix A of F with respect to the given basis. To that end we introduce a new orthonormal, positively oriented basis $\boldsymbol{e}_1', \boldsymbol{e}_2', \boldsymbol{e}_3'$. We choose $\boldsymbol{e}_3'$ as the unit direction vector $\frac{1}{3}(2,1,2)$ of the line. Then we take $\boldsymbol{e}_2'$ as any unit vector orthogonal to $\boldsymbol{e}_3'$, for example $\boldsymbol{e}_2' = \frac{1}{3}(1,2,-2)$. Finally, we set $\boldsymbol{e}_1' = \boldsymbol{e}_2' \times \boldsymbol{e}_3' = \frac{1}{3}(2,-2,-1)$. By Example 5.52, the matrix of F with respect to $\boldsymbol{e}_1', \boldsymbol{e}_2', \boldsymbol{e}_3'$ is

$$A' = \begin{bmatrix} \cos\frac{\pi}{3} & -\sin\frac{\pi}{3} & 0 \\ \sin\frac{\pi}{3} & \cos\frac{\pi}{3} & 0 \\ 0 & 0 & 1 \end{bmatrix} = \frac{1}{2}\begin{bmatrix} 1 & -\sqrt{3} & 0 \\ \sqrt{3} & 1 & 0 \\ 0 & 0 & 2 \end{bmatrix}.$$

The transition matrix from $\boldsymbol{e}_1', \boldsymbol{e}_2', \boldsymbol{e}_3'$ to $\boldsymbol{e}_1, \boldsymbol{e}_2, \boldsymbol{e}_3$ is

$$T = \frac{1}{3}\begin{bmatrix} 2 & 1 & 2 \\ -2 & 2 & 1 \\ -1 & -2 & 2 \end{bmatrix}.$$

Since both bases are orthonormal, $A' = T^t A T$ and hence

$$A = TA'T^t = \frac{1}{18}\begin{bmatrix} 13 & 2-6\sqrt{3} & 4+3\sqrt{3} \\ 2+6\sqrt{3} & 10 & 2-6\sqrt{3} \\ 4-3\sqrt{3} & 2+6\sqrt{3} & 13 \end{bmatrix}.$$

5.6 Isometries in Two and Three Dimensions

5.6.1 Isometries in Two Dimensions

Let F be an isometry on a two-dimensional inner product space V, and choose an orthonormal basis $\boldsymbol{e}_1, \boldsymbol{e}_2$ for V. Let

$$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$$

be the matrix of F with respect to $\boldsymbol{e}_1, \boldsymbol{e}_2$. Then A is orthogonal and $\det A = \pm 1$ by Theorem 5.33 and Theorem 5.34. If $\det A = 1$, then F is a rotation by Definition 5.38. Assume that $\det A = -1$. The matrix

$$A' = \begin{bmatrix} a_{11} & -a_{12} \\ a_{21} & -a_{22} \end{bmatrix}$$

is then orthogonal and $\det A' = 1$. By Theorem 5.41, $A' = R(\theta)$ for some real number θ. It follows that

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

for this θ. Since $A^2 = I$ and A is symmetric, F is an orthogonal reflection, and since $\det A = -1$, it follows from Theorem 5.23 that the reflection is in a line. We summarise our findings in the following theorem.

Theorem 5.54. An isometry F on a two-dimensional inner product space is either a rotation or orthogonal reflection in a line. In the first case, $\det F = 1$, and in the second case, $\det F = -1$.

5.6.2 Isometries in Three Dimensions

Theorem 5.55. An isometry F on a three-dimensional inner product space V is either rotation about a line or rotation about a line followed by reflection in the origin. In the first case, $\det F = 1$, and in the second case, $\det F = -1$.

Proof. Let A be the matrix of F with respect to a basis for V and λ a real number. The determinant

$$\det(F - \lambda I) = \det(A - \lambda I) = \begin{vmatrix} a_{11} - \lambda & a_{12} & a_{13} \\ a_{21} & a_{22} - \lambda & a_{23} \\ a_{31} & a_{32} & a_{33} - \lambda \end{vmatrix}$$

is a polynomial of degree 3 in λ. Hence, it has a real zero λ. For this λ there exists a non-zero $\boldsymbol{x} \in \mathbf{R}^3$ such that $(A - \lambda I)\boldsymbol{x} = \mathbf{0}$ or, equivalently, $A\boldsymbol{x} = \lambda\boldsymbol{x}$. Thus there exists a non-zero vector $\boldsymbol{v} \in V$ such that $F(\boldsymbol{v}) = \lambda\boldsymbol{v}$. Since F is an isometry, we have $|\lambda|\,\|\boldsymbol{v}\| = \|\lambda\boldsymbol{v}\| = \|F(\boldsymbol{v})\| = \|\boldsymbol{v}\|$, whence $\lambda = \pm 1$. It follows that $\dim(\ker(F - I)) > 0$ or $\dim(\ker(F + I)) > 0$. Let $U = \ker(F - I)$ and $n = \dim U$.

If $n = 3$, then $F(\boldsymbol{u}) = \boldsymbol{u}$ for all $\boldsymbol{u} \in V$. Hence, in this case, $F = I$ is the identity mapping, which can be regarded as rotation about any line through the angle 0.

Assume that $n = 2$ and let $\boldsymbol{u}$ be a non-zero vector orthogonal to U. Then $F(\boldsymbol{u})$ is orthogonal to U and $\|F(\boldsymbol{u})\| = \|\boldsymbol{u}\|$ since F preserves inner products and norms. Since $\boldsymbol{u} \notin U$, we have $F(\boldsymbol{u}) \neq \boldsymbol{u}$, and it follows that $F(\boldsymbol{u}) = -\boldsymbol{u}$. Hence F is orthogonal reflection in the plane U. In this case, F can also be regarded as rotation through the angle π about the line orthogonal to U followed by reflection in the origin.

Now assume that $n = 1$ so that U is a line. If $\boldsymbol{u} \in U^\perp$, then $F(\boldsymbol{u}) \in U^\perp$ since F preserves inner products. The restriction of F to the two-dimensional space $U^\perp$ is therefore an isometry on $U^\perp$. This restriction is therefore either a rotation or reflection in a line contained in $U^\perp$. In the latter case, there would exist a non-zero vector $\boldsymbol{u} \in U^\perp$ such that $F(\boldsymbol{u}) = \boldsymbol{u}$, contradicting the assumption that $n = 1$. Hence F is rotation about the line U in this case.

If $n = 0$, then $\dim(\ker(-F - I)) = \dim(\ker(F + I)) > 0$. It follows from what we have proved that $-F$ is rotation about a line or rotation about a line followed by reflection in the origin, and so is $F = -(-F)$.

When F is a rotation, its matrix with respect to an orthonormal basis for V consisting of two vectors $\boldsymbol{e}_1$ and $\boldsymbol{e}_2$ orthogonal to the line and one vector $\boldsymbol{e}_3$ parallel to the line is of the form

$$A = \begin{bmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Hence, $\det F = \det A = 1$ when F is a rotation. In the other case, $-F$ is a rotation, and therefore $\det F = (-1)^3 \det(-F) = (-1) \cdot 1 = -1$. ■

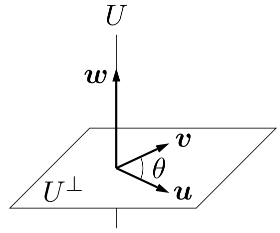
Suppose that we are given the matrix A of a linear transformation F on a 3-dimensional inner product space V with respect to an orthonormal basis. If we are asked to show that F is an isometry and to find its geometric meaning, we can proceed as follows:

- Establish that F is an isometry by showing that A is orthogonal.
- Find $U = \ker(F - I)$ by solving the system $(A - I)\boldsymbol{x} = \boldsymbol{0}$.
- Let $n = \dim U$ and take the appropriate action.

  $n = 3$: $F = I$, and we have completed the task.

  $n = 2$: F is orthogonal reflection in the plane U.

  $n = 1$: F is rotation about the line U. Let $\boldsymbol{w}$ be a direction vector of U, take any non-zero vector $\boldsymbol{u}$ orthogonal to $\boldsymbol{w}$ and compute $\boldsymbol{v} = F(\boldsymbol{u})$. The angle θ of rotation then equals the angle between $\boldsymbol{u}$ and $\boldsymbol{v}$. If $\theta = \pi$, F can also be characterised as orthogonal reflection in U. Otherwise, the vectors $\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w}$ form a basis for V. If V is ordinary three-space, this basis can be positively or negatively oriented. In the first case, the rotation appears anticlockwise on looking in the opposite direction of the vector $\boldsymbol{w}$, and in the second case, it appears clockwise.



  $n = 0$: Set $G = -F$, find $U = \ker(G - I) = \ker(F + I)$ and let $n = \dim U$. Then $n = 3$ or $n = 1$. If $n = 3$, then $F = -I$. If $n = 1$, then F is the rotation G followed by reflection in the origin.

Example 5.56. Consider the linear transformation F with matrix

$$A = \frac{1}{3}\begin{bmatrix} 2 & -1 & 2 \\ 2 & 2 & -1 \\ -1 & 2 & 2 \end{bmatrix}$$

with respect to an orthonormal, positively oriented basis for three-space. Since $AA^t = I$, A is orthogonal, and therefore F is an isometry. We solve the system $A\boldsymbol{x} = \boldsymbol{x}$.

$$\begin{cases} 2x_1 - x_2 + 2x_3 = 3x_1 \\ 2x_1 + 2x_2 - x_3 = 3x_2 \\ -x_1 + 2x_2 + 2x_3 = 3x_3 \end{cases} \Leftrightarrow \left[\begin{array}{ccc|c} -1 & -1 & 2 & 0 \\ 2 & -1 & -1 & 0 \\ -1 & 2 & -1 & 0 \end{array}\right] \Leftrightarrow \boldsymbol{x} = t(1,1,1).$$

Hence, F is a rotation about the line $U = [(1,1,1)]$. We set $\boldsymbol{w} = (1,1,1)$. A vector orthogonal to $\boldsymbol{w}$ is, for example, $\boldsymbol{u} = (1,-1,0)$. We have $\boldsymbol{v} = F(\boldsymbol{u}) = (1,0,-1)$, and the angle θ between $\boldsymbol{u}$ and $\boldsymbol{v}$ is given by

$$\cos\theta = \frac{\langle \boldsymbol{u}, \boldsymbol{v}\rangle}{\|\boldsymbol{u}\|\|\boldsymbol{v}\|} = \frac{1}{2}.$$

The angle of rotation is therefore $\theta = \frac{\pi}{3}$. The determinant of the matrix having columns $\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w}$ is

$$\begin{vmatrix} 1 & 1 & 1 \\ -1 & 0 & 1 \\ 0 & -1 & 1 \end{vmatrix} = 3 > 0.$$

Therefore, the rotation appears anticlockwise on looking from the point $(1,1,1)$ towards the origin.

Example 5.57. Consider a linear transformation F with matrix

$$A = \frac{1}{3}\begin{bmatrix} 2 & -1 & a \\ 2 & 2 & b \\ -1 & 2 & c \end{bmatrix}$$

with respect to an orthonormal basis for a 3-dimensional inner product space. Let us determine the values of a, b and c for which the transformation is a rotation. A necessary condition is that A be orthogonal. Since the first two columns are orthogonal and of length 1, we need only require that the last column be orthogonal to the first two columns and of length 1. The orthogonality condition means that

$$\begin{cases} 2a + 2b - c = 0 \\ -a + 2b + 2c = 0 \end{cases} \quad \Leftrightarrow \quad (a,b,c) = t(2,-1,2).$$

The length condition is now satisfied if and only if $\|t(2,-1,2)\| = 3\,|t| = 3$, which is equivalent to $t = \pm 1$. Hence,

$$A = \frac{1}{3}\begin{bmatrix} 2 & -1 & 2t \\ 2 & 2 & -t \\ -1 & 2 & 2t \end{bmatrix}$$

where $t = \pm 1$. We have

$$\det A = \left(\frac{1}{3}\right)^3 \begin{vmatrix} 2 & -1 & 2t \\ 2 & 2 & -t \\ -1 & 2 & 2t \end{vmatrix} = \frac{t}{27}\begin{vmatrix} 2 & -1 & 2 \\ 2 & 2 & -1 \\ -1 & 2 & 2 \end{vmatrix} = \frac{t}{27}\cdot 27 = t.$$

By Theorem 5.55, F is a rotation if and only if $t = 1$. Hence $(a,b,c) = (2,-1,2)$.

Example 5.58. We wish to determine a, b and c so that

$$A = \frac{1}{9}\begin{bmatrix} a & b & c \\ 4 & 7 & -4 \\ 8 & -4 & 1 \end{bmatrix}$$

becomes the matrix of an orthogonal reflection F in a plane in $\mathbf{R}^3$ with respect to the basis $\boldsymbol{\varepsilon}_1, \boldsymbol{\varepsilon}_2, \boldsymbol{\varepsilon}_3$. As in the previous example, we first make sure that A is orthogonal. This time we use the fact that the matrix is orthogonal if and only if its rows form an orthonormal basis for $\mathbf{R}^3$ and get $(a,b,c) = \pm(1,4,8)$. By Theorem 5.37, F is an orthogonal reflection if and only if A is symmetric. Hence $(a,b,c) = (1,4,8)$. We still do not know whether the subspace of reflection is a plane. However, by solving the system $A\boldsymbol{x} = \boldsymbol{x}$, we find that F actually is reflection in the plane $2x_1 - x_2 - 2x_3 = 0$.

We could instead have used Theorem 5.23. We have $\det A = -1$. Hence, F is reflection in a plane or in $\{\mathbf{0}\}$. Since $A \neq -I$, the subspace of reflection must be a plane.

Exercises

5.1. Let (x_1, x_2) be coordinates with respect to an orthonormal basis for two-space. Find the matrix with respect to that basis of orthogonal projection on the line $3x_1 = 4x_2$.

5.2. Let (x_1, x_2, x_3) be coordinates with respect to an orthonormal basis for three-space. Find the matrix of orthogonal projection on the plane $x_1 + 2x_2 - 2x_3 = 0$.

5.3. Let $\boldsymbol{e}_1, \boldsymbol{e}_2$ be a basis for two-space. Find the matrix with respect to that basis of projection on the line $x_1 = x_2$ along the x_1-axis.

5.4. Let $\boldsymbol{e}_1, \boldsymbol{e}_2, \boldsymbol{e}_3$ be a basis for three-space.

(a) Find the matrix of projection on the plane $x_1 + x_2 + x_3 = 0$ along the line $\boldsymbol{x} = t(1, 2, 3)$.

(b) Find the matrix of projection on the line along the plane.

5.5. Find, with respect to the basis $\boldsymbol{e}_1 = 1$, $\boldsymbol{e}_2 = x$, $\boldsymbol{e}_3 = x^2$, $\boldsymbol{e}_4 = x^3$, the matrix of the linear transformation F on P_3 defined by $F(\boldsymbol{p}) = \frac{d}{dx}((x-1)\boldsymbol{p})$.

5.6. Let F be a linear transformation with matrix

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

with respect to a basis $\boldsymbol{e}_1, \boldsymbol{e}_2, \boldsymbol{e}_3$. Find the matrix of F with respect to the basis

$$\begin{aligned} \boldsymbol{e}'_1 &= \boldsymbol{e}_1 + \boldsymbol{e}_2 - \boldsymbol{e}_3, \\ \boldsymbol{e}'_2 &= 2\boldsymbol{e}_1 + \boldsymbol{e}_2 + \boldsymbol{e}_3, \\ \boldsymbol{e}'_3 &= 2\boldsymbol{e}_1 + \boldsymbol{e}_2 + 2\boldsymbol{e}_3. \end{aligned}$$

5.7. Let $\boldsymbol{e}_1, \boldsymbol{e}_2, \boldsymbol{e}_3$ be a basis for a linear space and introduce new bases by

$$\begin{aligned} \boldsymbol{e}'_1 &= \boldsymbol{e}_1 + 2\boldsymbol{e}_2 + 3\boldsymbol{e}_3, \\ \boldsymbol{e}'_2 &= \boldsymbol{e}_1 + \boldsymbol{e}_2 + \boldsymbol{e}_3, \\ \boldsymbol{e}'_3 &= 2\boldsymbol{e}_1 + \boldsymbol{e}_2 + \boldsymbol{e}_3, \end{aligned} \qquad \text{and} \qquad \begin{aligned} \boldsymbol{e}''_1 &= \boldsymbol{e}_1 + \boldsymbol{e}_2 + 3\boldsymbol{e}_3, \\ \boldsymbol{e}''_2 &= \boldsymbol{e}_1 + \boldsymbol{e}_2 + 2\boldsymbol{e}_3, \\ \boldsymbol{e}''_3 &= 2\boldsymbol{e}_1 + \boldsymbol{e}_2 + 3\boldsymbol{e}_3. \end{aligned}$$

(a) Do the bases $\boldsymbol{e}_1, \boldsymbol{e}_2, \boldsymbol{e}_3$ and $\boldsymbol{e}'_1, \boldsymbol{e}'_2, \boldsymbol{e}'_3$ have the same orientation?

(b) Do the bases $\boldsymbol{e}'_1, \boldsymbol{e}'_2, \boldsymbol{e}'_3$ and $\boldsymbol{e}''_1, \boldsymbol{e}''_2, \boldsymbol{e}''_3$ have the same orientation?

5.8. Let F be the linear transformation on three-space with matrix

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

with respect to a basis $\boldsymbol{e}_1, \boldsymbol{e}_2, \boldsymbol{e}_3$. Show that F is projection on a plane along a line and find the plane and the line.

5.9. Show that the linear transformation with the following matrix with respect to a basis for $\mathbf{R}^4$ is projection on a subspace U' along a subspace U''. Find U' and U''.

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

5.10. Show that the linear transformation with matrix

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

with respect to a basis for $\mathbf{R}^3$ is reflection in a subspace U' along a subspace U'' and find U' and U''.

5.11. Let $\boldsymbol{e}_1, \boldsymbol{e}_2, \boldsymbol{e}_3$ be an orthonormal basis for $\mathbf{R}^3$ and consider the plane with equation $x_1 + 2x_2 - 2x_3 = 0$. Find the matrix of orthogonal reflection in that plane with respect to the given basis.

5.12. Let $\boldsymbol{e}_1, \boldsymbol{e}_2, \boldsymbol{e}_3$ be a basis for $\mathbf{R}^3$. Find the matrix with respect to that basis of reflection in the plane $2x_1 - x_2 - 3x_3 = 0$ along the line $\boldsymbol{x} = t(1, -2, 1)$.

5.13. Let I be the unit matrix of order n and B an $n \times 1$ column vector of unit length. Explain the geometric meaning of the so-called Householder matrix $I - 2BB^t$.

5.14. Let (x_1, x_2) be coordinates with respect to an orthonormal basis for two-space. Find the matrix with respect to that basis of rotation a quarter-turn in the direction from the x_1-axis towards the x_2-axis.

5.15. Let (x_1, x_2, x_3) be coordinates with respect to an orthonormal, positively oriented basis for three-space. Find the matrix of anticlockwise rotation about the x_2-axis through the angle $\frac{\pi}{6}$.

5.16. Let $\boldsymbol{e}_1, \boldsymbol{e}_2, \boldsymbol{e}_3$ be an orthonormal, positively oriented basis for three-space. Let F be rotation about the line $\boldsymbol{x} = t(1, 1, 0)$ through the angle $\frac{\pi}{6}$ and suppose that the rotation appears anticlockwise on looking from the point (1,1,0) towards the origin. Find the matrix of F with respect to the given basis.

5.17. Show that the determinant of a planar rotation on a finite-dimensional inner product space equals 1. Hint: Introduce a suitable basis.

5.18. The matrices below are matrices of linear transformations on two-space with respect to an orthonormal basis. Show that the linear transformations are isometries and find their geometric meaning.

(a) $\dfrac{1}{2}\begin{bmatrix} \sqrt{3} & 1 \\ -1 & \sqrt{3} \end{bmatrix}$, (b) $\dfrac{1}{2}\begin{bmatrix} -1 & \sqrt{3} \\ \sqrt{3} & 1 \end{bmatrix}$, (c) $\dfrac{1}{\sqrt{2}}\begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}$.

5.19. The matrices below are matrices of linear transformations on three-space with respect to an orthonormal, positively oriented basis. Show that the linear transformations are isometries and find their geometric meaning.

(a) $\frac{1}{9}\begin{bmatrix} 8 & -4 & 1 \\ -1 & -4 & -8 \\ 4 & 7 & -4 \end{bmatrix}$, (b) $\frac{1}{9}\begin{bmatrix} 1 & 4 & 8 \\ 4 & 7 & -4 \\ 8 & -4 & 1 \end{bmatrix}$, (c) $\begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}$,

(d) $\frac{1}{7}\begin{bmatrix} -6 & 2 & 3 \\ 2 & -3 & 6 \\ 3 & 6 & 2 \end{bmatrix}$, (e) $\frac{1}{3}\begin{bmatrix} -2 & 1 & -2 \\ -2 & -2 & 1 \\ 1 & -2 & -2 \end{bmatrix}$.

5.20. (a) Let F and G be rotations about lines in three-space. Show, for example by using determinants, that FG and GF are also rotations about lines.

(b) Let F and G be orthogonal reflections in two different planes through the origin in three-space. Show that FG and GF are rotations about the line of intersection between the planes. What are the angles of rotation?

6 Eigenvalues and Eigenvectors

6.1 Some Algebraic Preliminaries

We state here without proofs some algebraic results that will be used in this chapter. By the fundamental theorem of algebra, every non-constant polynomial with complex coefficients has a complex zero. As a consequence thereof, every polynomial

$$f = a_n x^n + a_{n-1} x^{n-1} + \cdots + a_1 x + a_0$$

over $\mathbf{C}$ of degree $n \geq 1$ can be written as

$$f = a_n (x - \alpha_1)(x - \alpha_2) \cdots (x - \alpha_n) \tag{6.1}$$

for some complex numbers α_i, and this factorisation is unique up to the order of the factors.

The α_i are zeros of f, and by the cancellation law, they are the only zeros. If α is a zero of f, we define its multiplicity as the number of indices i for which $\alpha = \alpha_i$.

By expanding the right-hand side of (6.1) and identifying coefficients, we can express the coefficients of f in terms of the zeros of f. The identities thus obtained are collectively called the relationship between roots and coefficients. In this chapter, we shall only use the following two identities:

$$\begin{aligned} \frac{a_{n-1}}{a_n} &= -(\alpha_1 + \alpha_2 + \cdots + \alpha_n), \\ \frac{a_0}{a_n} &= (-1)^n \alpha_1 \alpha_2 \cdots \alpha_n. \end{aligned}$$

6.2 Definition

Definition 6.1. Let F be a linear transformation on a linear space V. We say that $\lambda \in \mathbf{R}$ is an eigenvalue of F if there exists a non-zero vector $\boldsymbol{u} \in V$ such that $F(\boldsymbol{u}) = \lambda \boldsymbol{u}$. A non-zero vector $\boldsymbol{u} \in V$ for which $F(\boldsymbol{u}) = \lambda \boldsymbol{u}$ is called an eigenvector of F belonging to the eigenvalue λ. By an eigenvalue and an eigenvector of an $n \times n$ matrix A we shall mean an eigenvalue and an eigenvector of the linear transformation $\boldsymbol{x} \mapsto A\boldsymbol{x}$ on $\mathbf{R}^n$.

Definition 6.2. Let $A = [a_{ik}]$ be an $n \times n$ matrix. The trace of A is defined by

$$\operatorname{tr} A = \sum_{i=1}^{n} a_{ii}.$$

Theorem 6.3. Let A be an $n \times n$ matrix. Then $\det(A - \lambda I)$ is a polynomial in λ of the form

$$\det(A - \lambda I) = (-\lambda)^n + b_{n-1}(-\lambda)^{n-1} + \cdots + b_0.$$

If the complex zeros of $\det(A - \lambda I)$ are $\lambda_1, \ldots, \lambda_n$, where each zero is counted as many times as its multiplicity, then $b_{n-1} = \operatorname{tr} A = \lambda_1 + \cdots + \lambda_n$ and $b_0 = \det A = \lambda_1 \cdots \lambda_n$.

Proof. Let $A^{(k)}$ be the matrix obtained from I by replacing the kth column of I with A_k. Then, by multilinearity,

$$\begin{aligned}\det(A - \lambda I) &= \det[A_1 - \lambda I_1, \ldots, A_n - \lambda I_n] \\ &= (-\lambda)^n \det I + (-\lambda)^{n-1}\left(\det A^{(1)} + \cdots + \det A^{(n)}\right) + \cdots + \det A \\ &= (-\lambda)^n + (-\lambda)^{n-1} \operatorname{tr} A + \cdots + \det A.\end{aligned}$$

This shows that $\det(A - \lambda I)$ is of the form stated and that $b_{n-1} = \operatorname{tr} A$ and $b_0 = \det A$. The equalities $b_{n-1} = \lambda_1 + \cdots + \lambda_n$ and $b_0 = \lambda_1 \cdots \lambda_n$ follow from the relationship between roots and coefficients. ■

Definition 6.4. Let A be a square matrix. The polynomial $\det(A - \lambda I)$ is called the characteristic polynomial of A.

Theorem 6.5. Let A be an $n \times n$ matrix. Then $\lambda \in \mathbf{R}$ is an eigenvalue of A if and only if $\det(A - \lambda I) = 0$. The eigenvectors of A belonging to the eigenvalue λ are the non-zero vectors $\boldsymbol{x}$ given by $(A - \lambda I)\boldsymbol{x} = \mathbf{0}$.

Proof. We have $\det(A - \lambda I) = 0$ if and only if the system $A\boldsymbol{x} - \lambda\boldsymbol{x} = (A - \lambda I)\boldsymbol{x} = \mathbf{0}$ has a non-zero solution $\boldsymbol{x}$. The second statement follows directly from the definition. ■

Example 6.6. The eigenvalues of

$$A = \begin{bmatrix} 1 & 2 \\ 2 & -2 \end{bmatrix}$$

are given by

$$0 = \det(A - \lambda I) = \begin{vmatrix} 1 - \lambda & 2 \\ 2 & -2 - \lambda \end{vmatrix} = \lambda^2 + \lambda - 6 = (\lambda + 3)(\lambda - 2).$$

Hence, the eigenvalues of A are $\lambda_1 = -3$ and $\lambda_2 = 2$. The eigenvectors belonging to the eigenvalue λ_1 are the non-zero vectors $\boldsymbol{x}$ given by

$$\left[\begin{array}{cc|c} 1 - \lambda_1 & 2 & 0 \\ 2 & -2 - \lambda_1 & 0 \end{array}\right] \quad \Leftrightarrow \quad \left[\begin{array}{cc|c} 4 & 2 & 0 \\ 2 & 1 & 0 \end{array}\right] \quad \Leftrightarrow \quad \boldsymbol{x} = t(1, -2),$$

and those belonging to λ_2 are the non-zero vectors given by

$$\left[\begin{array}{cc|c} 1 - \lambda_2 & 2 & 0 \\ 2 & -2 - \lambda_2 & 0 \end{array}\right] \quad \Leftrightarrow \quad \left[\begin{array}{cc|c} -1 & 2 & 0 \\ 2 & -4 & 0 \end{array}\right] \quad \Leftrightarrow \quad \boldsymbol{x} = t(2, 1).$$

Theorem 6.7. Let F be a linear transformation on a linear space V with basis $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n$ and let A be the matrix of F with respect to that basis. Then λ is an eigenvalue of F if and only if λ is an eigenvalue of A. Moreover, if $\boldsymbol{x} \in \mathbf{R}^n$ is the coordinate vector of $\boldsymbol{u} \in V$, then $\boldsymbol{u}$ is an eigenvector of F belonging to λ if and only if $\boldsymbol{x}$ is an eigenvector of A belonging to λ.

Proof. If $\boldsymbol{x} \in \mathbf{R}^n$ is the coordinate vector of $\boldsymbol{u} \in V$, then $A\boldsymbol{x} = \lambda\boldsymbol{x}$ if and only if $F(\boldsymbol{u}) = \lambda\boldsymbol{u}$. ■

Example 6.8. Let V be a 2-dimensional linear space endowed with a basis $\boldsymbol{e}_1, \boldsymbol{e}_2$, and consider the linear transformation F on V whose matrix is

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

The eigenvalues of F are given by

$$0 = \det(A - \lambda I) = \begin{vmatrix} -\lambda & -1 \\ -1 & -\lambda \end{vmatrix} = \lambda^2 - 1 = (\lambda + 1)(\lambda - 1),$$

and are therefore $\lambda_1 = -1$ and $\lambda_2 = 1$. The coordinate vectors of the eigenvectors belonging to λ_1 satisfy the system

$$\left[\begin{array}{cc|c} 1 & -1 & 0 \\ -1 & 1 & 0 \end{array}\right] \quad \Leftrightarrow \quad \boldsymbol{x} = t(1, 1).$$

Hence, the eigenvectors $\boldsymbol{u}$ belonging to that eigenvalue are $\boldsymbol{u} = t(\boldsymbol{e}_1 + \boldsymbol{e}_2)$, $t \neq 0$. In the same way, we find that the eigenvectors belonging to λ_2 are $\boldsymbol{u} = t(\boldsymbol{e}_1 - \boldsymbol{e}_2)$, $t \neq 0$.

6.3 Diagonalisability

Let F be a linear transformation on a linear space V and suppose there exists a basis $\boldsymbol{f}_1, \ldots, \boldsymbol{f}_n$ for V consisting of eigenvectors of F belonging to the eigenvalues $\lambda_1, \ldots, \lambda_n$. Since $F(\boldsymbol{f}_i) = \lambda_i \boldsymbol{f}_i$ for $i = 1, \ldots, n$, the matrix of F with respect to that basis is the diagonal matrix

$$D = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix}.$$

Let A be the matrix of F with respect to any basis $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n$ for V. If T is the $n \times n$ matrix whose ith column is the coordinate vector of $\boldsymbol{f}_i$ with respect to $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n$, then $T^{-1}AT = D$. According to the following definition, A is diagonalisable.

Definition 6.9. Let A be an $n \times n$ matrix. We say that A is diagonalisable if there exists an invertible $n \times n$ matrix T such that $T^{-1}AT$ is a diagonal matrix.

Theorem 6.10. Let F be a linear transformation on a linear space V and let A be the matrix of F with respect to a basis $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n$ for V. Then A is diagonalisable if and only if there exists a basis for V consisting of eigenvectors of F.

Proof. We have already shown that A is diagonalisable if such a basis exists. To show the converse, we assume that A is diagonalisable, which means that there exist an invertible matrix T and a diagonal matrix D such that $T^{-1}AT = D$. Since T is invertible, its columns $\boldsymbol{t}_1, \ldots, \boldsymbol{t}_n$ form a basis for $\mathbf{R}^n$. Hence, the vectors $\boldsymbol{f}_1, \ldots, \boldsymbol{f}_n$ with coordinates $\boldsymbol{t}_1, \ldots, \boldsymbol{t}_n$ with respect to $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n$ form a basis for V. Let the diagonal entries of D be $\lambda_1, \ldots, \lambda_n$. Since $AT = TD$, we have $A\boldsymbol{t}_i = \lambda_i \boldsymbol{t}_i$, and hence $F(\boldsymbol{f}_i) = \lambda_i \boldsymbol{f}_i$, for $i = 1, \ldots, n$. This shows that $\boldsymbol{f}_1, \ldots, \boldsymbol{f}_n$ are also eigenvectors of F. ■

Corollary 6.11. Let A be an $n \times n$ matrix. Then A is diagonalisable if and only if there exists a basis $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n$ for $\mathbf{R}^n$ consisting of eigenvectors of A.

Example 6.12. The eigenvectors $\boldsymbol{e}_1 = (1, -2)$ and $\boldsymbol{e}_2 = (2, 1)$ of the matrix

$$A = \begin{bmatrix} 1 & 2 \\ 2 & -2 \end{bmatrix}$$

in Example 6.6 form a basis for $\mathbf{R}^2$. Hence, A is diagonalisable, and we have $T^{-1}AT = D$ where

$$T = \begin{bmatrix} 1 & 2 \\ -2 & 1 \end{bmatrix} \quad \text{and} \quad D = \begin{bmatrix} -3 & 0 \\ 0 & 2 \end{bmatrix}.$$

Example 6.13. Consider the matrix

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

We have

$$\det(A - \lambda I) = \begin{vmatrix} 1-\lambda & 0 \\ 1 & 1-\lambda \end{vmatrix} = (1-\lambda)^2.$$

The only eigenvalue of A is therefore $\lambda = 1$. The eigenvectors $\boldsymbol{x}$ belonging to this eigenvalue are given by

$$\left[\begin{array}{cc|c} 0 & 0 & 0 \\ 1 & 0 & 0 \end{array}\right] \quad \Leftrightarrow \quad \boldsymbol{x} = t(0, 1).$$

Hence, no basis for $\mathbf{R}^2$ consisting of eigenvectors of A exists, and A is not diagonalisable.

Example 6.14. Let

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

The characteristic polynomial $\det(A - \lambda I) = (1-\lambda)^2 + 1$ has no real zeros at all. Hence, A has no eigenvectors, and is therefore not diagonalisable.

Theorem 6.15. Let F be a linear transformation on a linear space V, and suppose that $\boldsymbol{e}_1, \dots, \boldsymbol{e}_k$ are eigenvectors belonging to distinct eigenvalues $\lambda_1, \dots, \lambda_k$ of F. Then $\boldsymbol{e}_1, \dots, \boldsymbol{e}_k$ are linearly independent.

Proof. We use induction on k. The statement holds for $k = 1$ since eigenvectors are non-zero. Assume that the statement holds for k eigenvectors and let $\boldsymbol{e}_1, \dots, \boldsymbol{e}_k, \boldsymbol{e}_{k+1}$ be eigenvectors belonging to distinct eigenvalues $\lambda_1, \dots, \lambda_k, \lambda_{k+1}$. Suppose that

$$s_1\boldsymbol{e}_1 + \cdots + s_k\boldsymbol{e}_k + s_{k+1}\boldsymbol{e}_{k+1} = \boldsymbol{0}. \tag{6.2}$$

Multiplying both sides of (6.2) by λ_{k+1}, we get

$$s_1\lambda_{k+1}\boldsymbol{e}_1 + \cdots + s_k\lambda_{k+1}\boldsymbol{e}_k + s_{k+1}\lambda_{k+1}\boldsymbol{e}_{k+1} = \boldsymbol{0}, \tag{6.3}$$

and taking F of both sides of (6.2) yields

$$s_1\lambda_1\boldsymbol{e}_1 + \cdots + s_k\lambda_k\boldsymbol{e}_k + s_{k+1}\lambda_{k+1}\boldsymbol{e}_{k+1} = \boldsymbol{0}. \tag{6.4}$$

Subtracting corresponding sides of (6.3) from (6.4) , we now get

$$s_1(\lambda_1 - \lambda_{k+1})\boldsymbol{e}_1 + \cdots + s_k(\lambda_k - \lambda_{k+1})\boldsymbol{e}_k = \boldsymbol{0}.$$

By hypothesis, $\boldsymbol{e}_1, \dots, \boldsymbol{e}_k$ are linearly independent. Therefore, $s_i(\lambda_i - \lambda_{k+1}) = 0$ for $i = 1, \dots, k$. Since $\lambda_i \neq \lambda_{k+1}$ for $i = 1, \dots, k$, we must have $s_1 = \cdots = s_k = 0$, and hence by (6.2), $s_{k+1} = 0$. ■

Corollary 6.16. Let F be a linear transformation on a non-zero n-dimensional linear space V. If $\boldsymbol{e}_1, \dots, \boldsymbol{e}_n$ are eigenvectors of F belonging to n distinct eigenvalues, then $\boldsymbol{e}_1, \dots, \boldsymbol{e}_n$ form a basis for V.

Example 6.17. Let

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

We set out to compute A^n for every positive integer n. We have

$$\begin{vmatrix} 6-\lambda & -2 & -2 \\ 1 & 2-\lambda & -1 \\ 3 & -2 & 1-\lambda \end{vmatrix} = \begin{vmatrix} 6-\lambda & -2 & -2 \\ 1 & 2-\lambda & -1 \\ \lambda-3 & 0 & 3-\lambda \end{vmatrix} = \begin{vmatrix} 6-\lambda & -2 & 4-\lambda \\ 1 & 2-\lambda & 0 \\ \lambda-3 & 0 & 0 \end{vmatrix}$$
$$= -(\lambda-2)(\lambda-3)(\lambda-4).$$

Hence, the eigenvalues of A are $\lambda_1 = 2$, $\lambda_2 = 3$ and $\lambda_3 = 4$. The eigenvectors belonging to λ_1 are given by

$$\left[\begin{array}{ccc|c} 4 & -2 & -2 & 0 \\ 1 & 0 & -1 & 0 \\ 3 & -2 & -1 & 0 \end{array}\right] \quad \Leftrightarrow \quad \left[\begin{array}{ccc|c} 1 & 0 & -1 & 0 \\ 1 & 0 & -1 & 0 \\ 3 & -2 & -1 & 0 \end{array}\right] \quad \Leftrightarrow \quad \boldsymbol{x} = t(1,1,1).$$

For λ_2, we have

$$\left[\begin{array}{ccc|c} 3 & -2 & -2 & 0 \\ 1 & -1 & -1 & 0 \\ 3 & -2 & -2 & 0 \end{array}\right] \quad\Leftrightarrow\quad \left[\begin{array}{ccc|c} 1 & 0 & 0 & 0 \\ 1 & -1 & -1 & 0 \end{array}\right] \quad\Leftrightarrow\quad \boldsymbol{x} = t(0,1,-1),$$

and for λ_3,

$$\left[\begin{array}{ccc|c} 2 & -2 & -2 & 0 \\ 1 & -2 & -1 & 0 \\ 3 & -2 & -3 & 0 \end{array}\right] \quad\Leftrightarrow\quad \left[\begin{array}{ccc|c} 0 & 2 & 0 & 0 \\ 1 & -2 & -1 & 0 \\ 0 & 4 & 0 & 0 \end{array}\right] \quad\Leftrightarrow\quad \boldsymbol{x} = t(1,0,1).$$

By Corollary 6.16, the eigenvectors $(1,1,1), (0,1,-1), (1,0,1)$ form a basis for $\mathbf{R}^3$. Thus $T^{-1}AT = D$, and hence $A = TDT^{-1}$, where

$$T = \begin{bmatrix} 1 & 0 & 1 \\ 1 & 1 & 0 \\ 1 & -1 & 1 \end{bmatrix} \quad\text{and}\quad D = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 4 \end{bmatrix}.$$

We compute the inverse T^{-1} and get

$$T^{-1} = \begin{bmatrix} -1 & 1 & 1 \\ 1 & 0 & -1 \\ 2 & -1 & -1 \end{bmatrix}.$$

It follows that

$$\begin{aligned} A^n &= (TDT^{-1})^n = TDT^{-1}TDT^{-1}\cdots TDT^{-1} = TD^nT^{-1} \\ &= \begin{bmatrix} 1 & 0 & 1 \\ 1 & 1 & 0 \\ 1 & -1 & 1 \end{bmatrix} \begin{bmatrix} 2^n & 0 & 0 \\ 0 & 3^n & 0 \\ 0 & 0 & 4^n \end{bmatrix} \begin{bmatrix} -1 & 1 & 1 \\ 1 & 0 & -1 \\ 2 & -1 & -1 \end{bmatrix} \\ &= \begin{bmatrix} -2^n + 2\cdot 4^n & 2^n - 4^n & 2^n - 4^n \\ -2^n + 3^n & 2^n & 2^n - 3^n \\ -2^n - 3^n + 2\cdot 4^n & 2^n - 4^n & 2^n + 3^n - 4^n \end{bmatrix}. \end{aligned}$$

Definition 6.18. Let F be a linear transformation on a linear space V. If λ is an eigenvalue of F, the eigenspace of F associated with λ is

$$\{\boldsymbol{u} \in V;\ F(\boldsymbol{u}) = \lambda\boldsymbol{u}\} = \ker(F - \lambda I).$$

Note that $\mathbf{0}$, albeit not an eigenvector, belongs to an eigenspace.

Definition 6.19. Let F be a linear transformation on a finite-dimensional space V, and let λ be an eigenvalue of F. The algebraic multiplicity of λ is the multiplicity of $\mu = \lambda$ as a zero of the polynomial $\det(F - \mu I)$, and its geometric multiplicity is the dimension of $\ker(F - \lambda I)$.

Theorem 6.20. Let F be a linear transformation on a finite-dimensional space V, and let λ be an eigenvalue of F. Then the geometric multiplicity of λ is less than or equal to the algebraic multiplicity of λ.

Proof. Assume that $\dim V = n$. We choose a basis $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_k$ for $\ker(F - \lambda I)$ and extend it to a basis $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_k, \boldsymbol{e}_{k+1}, \ldots, \boldsymbol{e}_n$ for V. Since $F(\boldsymbol{e}_i) = \lambda \boldsymbol{e}_i$ for $i = 1, \ldots, k$, the matrix of F with respect to this basis is of the form

$$A = \begin{bmatrix} \lambda & 0 & \cdots & 0 & a_{1(k+1)} & \cdots & a_{1n} \\ 0 & \lambda & \cdots & 0 & a_{2(k+1)} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \lambda & a_{k(k+1)} & \cdots & a_{kn} \\ 0 & 0 & \cdots & 0 & a_{(k+1)(k+1)} & \cdots & a_{(k+1)n} \\ \vdots & \vdots & & \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 0 & a_{n(k+1)} & \cdots & a_{nn} \end{bmatrix}.$$

Successively expanding along the first column, we get

$$\det(F - \mu I) = \det(A - \mu I) = (\lambda - \mu)^k f(\mu)$$

where f is a polynomial of degree $n - k$. Hence, the algebraic multiplicity of λ is greater than or equal to its geometric multiplicity k. ■

Corollary 6.21. Let F be a linear transformation on a non-zero n-dimensional linear space V. If the polynomial $\det(F - \lambda I)$ has non-real zeros, then V has no basis consisting of eigenvectors of F. If all the distinct zeros $\lambda_1, \ldots, \lambda_k$ are real, then there exists a basis for V of eigenvectors of F if and only if the algebraic and geometric multiplicities of λ_i are equal for $i = 1, \ldots, k$.

Proof. We have $n = \deg(\det(F - \lambda I)) = \dim V$. By Theorem 6.15, an eigenvector basis exists if and only if the sum of the geometric multiplicities of the distinct eigenvalues equals n. By Theorem 6.20, this cannot happen if $\det(F - \lambda I)$ has non-real zeros, and if all zeros are real, then the sum of the geometric multiplicities of the distinct eigenvalues equals n if and only if the algebraic and geometric multiplicities of each eigenvalue are equal. ■

Example 6.22. Let

$$A = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 2 \end{bmatrix}.$$

We have

$$\det(A - \lambda I) = \begin{vmatrix} 1-\lambda & 0 & 0 \\ 1 & 1-\lambda & 0 \\ 0 & 0 & 2-\lambda \end{vmatrix} = (1-\lambda)^2(2-\lambda).$$

The algebraic multiplicities of the eigenvalues $\lambda_1 = 1$ and $\lambda_2 = 2$ are 2 and 1, respectively. Hence the geometric multiplicity of λ_2 is 1 since it cannot be less than 1. The geometric multiplicity of λ_1 is either 1 or 2. Since

$$\left[\begin{array}{ccc|c} 1-\lambda_1 & 0 & 0 & 0 \\ 1 & 1-\lambda_1 & 0 & 0 \\ 0 & 0 & 2-\lambda_1 & 0 \end{array}\right] \quad \Leftrightarrow \quad \left[\begin{array}{ccc|c} 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{array}\right] \quad \Leftrightarrow \quad \boldsymbol{x} = t(0,1,0),$$

the geometric multiplicity of λ_1 equals 1. Hence, A is not diagonalisable.

Example 6.23. Consider the matrix

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

We have

$$\det(A - \lambda I) = \begin{vmatrix} -\lambda & 1 & 0 \\ 1 & -\lambda & 0 \\ 0 & 0 & 1-\lambda \end{vmatrix} = -(\lambda-1)^2(\lambda+1).$$

The algebraic multiplicities of $\lambda_1 = 1$ and $\lambda_2 = -1$ are 2 and 1, respectively. The eigenvectors belonging to λ_1 are given by

$$\left[\begin{array}{ccc|c} -1 & 1 & 0 & 0 \\ 1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right] \quad \Leftrightarrow \quad \boldsymbol{x} = s(1,1,0) + t(0,0,1).$$

Hence, the geometric multiplicity of λ_1 equals 2. As a basis for the eigenspace associated with λ_1 we can choose $\boldsymbol{e}_1 = (1,1,0)$, $\boldsymbol{e}_2 = (0,0,1)$. This time it is worthwhile to find also the eigenvectors belonging to λ_2.

$$\left[\begin{array}{ccc|c} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \end{array}\right] \quad \Leftrightarrow \quad \boldsymbol{x} = t(1,-1,0).$$

We can, therefore, choose $\boldsymbol{e}_3 = (1,-1,0)$ as a basis for the eigenspace associated with λ_2. The eigenvectors $\boldsymbol{e}_1, \boldsymbol{e}_2, \boldsymbol{e}_3$ form a basis for $\mathbf{R}^3$ since eigenvectors belonging to different eigenvalues are linearly independent. Hence, A is diagonalisable, and $T^{-1}AT = D$ where

$$T = \begin{bmatrix} 1 & 0 & 1 \\ 1 & 0 & -1 \\ 0 & 1 & 0 \end{bmatrix} \quad \text{and} \quad D = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{bmatrix}.$$

6.4 Recurrence Equations

Let A be a diagonalisable $n \times n$ matrix, and suppose that we are given a recurrence equation

$$\boldsymbol{u}_{k+1} = A\boldsymbol{u}_k.$$

Since A is diagonalisable, $\mathbf{R}^n$ has a basis $\boldsymbol{e}_1, \dots, \boldsymbol{e}_n$ consisting of eigenvectors of A. Let $\lambda_1, \dots, \lambda_n$ be the corresponding eigenvalues. We can then write $\boldsymbol{u}_0 = c_1\boldsymbol{e}_1 + \cdots + c_n\boldsymbol{e}_n$, where $(c_1, \dots, c_n)$ are the coordinates of $\boldsymbol{u}_0$ with respect to the basis $\boldsymbol{e}_1, \dots, \boldsymbol{e}_n$. From this we get

$$\begin{aligned}
\boldsymbol{u}_1 &= A\boldsymbol{u}_0 = A(c_1\boldsymbol{e}_1 + \cdots + c_n\boldsymbol{e}_n) = c_1\lambda_1\boldsymbol{e}_1 + \cdots + c_n\lambda_n\boldsymbol{e}_n,\\
\boldsymbol{u}_2 &= A\boldsymbol{u}_1 = A(c_1\lambda_1\boldsymbol{e}_1 + \cdots + c_n\lambda_n\boldsymbol{e}_n) = c_1\lambda_1^2\boldsymbol{e}_1 + \cdots + c_n\lambda_n^2\boldsymbol{e}_n,\\
&\vdots\\
\boldsymbol{u}_k &= c_1\lambda_1^k\boldsymbol{e}_1 + \cdots + c_n\lambda_n^k\boldsymbol{e}_n.
\end{aligned}$$

Hence, if we know $\boldsymbol{u}_0$, the eigenvalues and the eigenvectors, we know $\boldsymbol{u}_k$ for all k.

Example 6.24. Let the numbers a_k and b_k be defined by

$$\begin{cases} a_{k+1} = a_k + 2b_k \\ b_{k+1} = 2a_k + b_k \end{cases}, \qquad \begin{cases} a_0 = 3 \\ b_0 = 1 \end{cases}.$$

If we set $\boldsymbol{u}_k = (a_k, b_k)$ and

$$A = \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix},$$

the recurrence relation can be written as $\boldsymbol{u}_{k+1} = A\boldsymbol{u}_k$, $\boldsymbol{u}_0 = (3, 1)$. Let us first find the eigenvalues and eigenvectors of A. We have

$$\begin{vmatrix} 1-\lambda & 2 \\ 2 & 1-\lambda \end{vmatrix} = (1-\lambda)^2 - 4 = (\lambda+1)(\lambda-3).$$

Thus the eigenvalues are $\lambda_1 = -1$ and $\lambda_2 = 3$. For λ_1, the eigenvectors are given by

$$\left[\begin{array}{cc|c} 2 & 2 & 0 \\ 2 & 2 & 0 \end{array}\right] \quad \Leftrightarrow \quad \boldsymbol{x} = t(1, -1),$$

and for λ_2 by

$$\left[\begin{array}{cc|c} -2 & 2 & 0 \\ 2 & -2 & 0 \end{array}\right] \quad \Leftrightarrow \quad \boldsymbol{x} = t(1, 1).$$

The eigenvectors $\boldsymbol{e}_1 = (1, -1)$ and $\boldsymbol{e}_2 = (1, 1)$ form a basis for $\mathbf{R}^2$. We have

$$c_1\boldsymbol{e}_1 + c_2\boldsymbol{e}_2 = \boldsymbol{u}_0 \quad \Leftrightarrow \quad \left[\begin{array}{cc|c} 1 & 1 & 3 \\ -1 & 1 & 1 \end{array}\right] \quad \Leftrightarrow \quad c_1 = 1,\ c_2 = 2.$$

Hence,

$$\boldsymbol{u}_k = \begin{bmatrix} a_k \\ b_k \end{bmatrix} = c_1\lambda_1^k\boldsymbol{e}_1 + c_2\lambda_2^k\boldsymbol{e}_2 = (-1)^k \begin{bmatrix} 1 \\ -1 \end{bmatrix} + 2\cdot 3^k \begin{bmatrix} 1 \\ 1 \end{bmatrix}.$$

Example 6.25. Consider the recurrence relation

$$a_{n+3} = -2a_{n+2} + a_{n+1} + 2a_n, \quad a_0 = -2,\ a_1 = -1,\ a_2 = 1.$$

Setting $b_n = a_{n+1}$ and $c_n = a_{n+2}$, we get

$$\begin{cases} a_{n+1} = & b_n \\ b_{n+1} = & c_n \\ c_{n+1} = 2a_n + b_n - 2c_n \end{cases}, \qquad \begin{cases} a_0 = -2 \\ b_0 = -1 \\ c_0 = 1 \end{cases}.$$

Provided that

$$A = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 2 & 1 & -2 \end{bmatrix}$$

is diagonalisable, we can now apply the method used in the previous example. It turns out that the eigenvectors $\boldsymbol{e}_1 = (1, -2, 4)$, $\boldsymbol{e}_2 = (1, -1, 1)$, $\boldsymbol{e}_3 = (1, 1, 1)$ belonging to the eigenvalues $\lambda_1 = -2$, $\lambda_2 = -1$, $\lambda_3 = 1$ form a basis for $\mathbf{R}^3$. Setting

$$(a_0, b_0, c_0) = d_1\boldsymbol{e}_1 + d_2\boldsymbol{e}_2 + d_3\boldsymbol{e}_3$$

and solving for d_1, d_2, d_3, we get $d_1 = 1$, $d_2 = -2$, $d_3 = -1$. Hence

$$\begin{bmatrix} a_n \\ b_n \\ c_n \end{bmatrix} = d_1\lambda_1^n\boldsymbol{e}_1 + d_2\lambda_2^n\boldsymbol{e}_2 + d_3\lambda_3^n\boldsymbol{e}_3 = (-2)^n \begin{bmatrix} 1 \\ -2 \\ 4 \end{bmatrix} - 2 \cdot (-1)^n \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix} - \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix},$$

and in particular,

$$a_n = (-2)^n - 2 \cdot (-1)^n - 1.$$

Example 6.26. The numbers a_n defined by $a_{n+2} = a_{n+1} + a_n$, $a_0 = 0$, $a_1 = 1$, are known as the Fibonacci numbers. As in the previous example, we set $b_n = a_{n+1}$ and get

$$\begin{bmatrix} a_{n+1} \\ b_{n+1} \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} a_n \\ b_n \end{bmatrix}.$$

We have

$$\begin{vmatrix} -\lambda & 1 \\ 1 & 1-\lambda \end{vmatrix} = \lambda^2 - \lambda - 1 = 0 \quad \Leftrightarrow \quad \lambda = \lambda_1 = \frac{1+\sqrt{5}}{2} \quad \text{or} \quad \lambda = \lambda_2 = \frac{1-\sqrt{5}}{2}.$$

The relationship between roots and coefficients gives that $\lambda_1\lambda_2 = -1$ and $\lambda_1 + \lambda_2 = 1$. Hence,

$$\left[\begin{array}{cc|c} -\lambda_1 & 1 & 0 \\ 1 & 1-\lambda_1 & 0 \end{array}\right] \Leftrightarrow \left[\begin{array}{cc|c} -\lambda_1 & 1 & 0 \\ 1 & \lambda_2 & 0 \end{array}\right] \Leftrightarrow \left[\begin{array}{cc|c} -\lambda_1 & 1 & 0 \\ \lambda_1 & -1 & 0 \end{array}\right] \Leftrightarrow \boldsymbol{x} = t(1, \lambda_1).$$

An eigenvector belonging to λ_1 is therefore $\boldsymbol{e}_1 = (1, \lambda_1)$. By symmetry, we find that $\boldsymbol{e}_2 = (1, \lambda_2)$ is an eigenvector belonging to λ_2. Setting $(a_0, b_0) = (0, 1) = c_1\boldsymbol{e}_1 + c_2\boldsymbol{e}_2$, we get

$$\left[\begin{array}{cc|c} 1 & 1 & 0 \\ \lambda_1 & \lambda_2 & 1 \end{array}\right] \quad \Leftrightarrow \quad c_1 = \frac{1}{\lambda_1 - \lambda_2} = \frac{1}{\sqrt{5}} \quad \text{and} \quad c_2 = -c_1 = -\frac{1}{\sqrt{5}}.$$

Therefore

$$\begin{bmatrix} a_n \\ b_n \end{bmatrix} = \frac{1}{\sqrt{5}}\lambda_1^n \begin{bmatrix} 1 \\ \lambda_1 \end{bmatrix} - \frac{1}{\sqrt{5}}\lambda_2^n \begin{bmatrix} 1 \\ \lambda_2 \end{bmatrix}.$$

In particular, we have

$$a_n = \frac{1}{\sqrt{5}}\left(\left(\frac{1+\sqrt{5}}{2}\right)^n - \left(\frac{1-\sqrt{5}}{2}\right)^n\right).$$

6.5 The Spectral Theorem

So far, we have only discussed real linear spaces. If complex scalars are allowed in Definition 2.1, we get a complex linear space. The set $\mathbf{C}^n$ of n-tuples of complex numbers together with the addition

$$(x_1, \ldots, x_n) + (y_1, \ldots, y_n) = (x_1 + y_1, \ldots, x_n + y_n)$$

and the multiplication

$$s(x_1, \ldots, x_n) = (sx_1, \ldots, sx_n), \quad s \in \mathbf{C},$$

then forms a complex linear space. The dot product on $\mathbf{C}^n$ is defined by

$$\boldsymbol{x} \cdot \boldsymbol{y} = x_1\overline{y_1} + \cdots + x_n\overline{y_n}$$

and the norm by

$$\|\boldsymbol{x}\| = \sqrt{\boldsymbol{x} \cdot \boldsymbol{x}} = \sqrt{|x_1|^2 + \cdots + |x_n|^2}.$$

If we replace condition (ii) in Definition 3.1 with the condition $\langle \boldsymbol{u}, \boldsymbol{v} \rangle = \overline{\langle \boldsymbol{v}, \boldsymbol{u} \rangle}$, we get a complex inner product. The dot product on $\mathbf{C}^n$ is then a complex inner product. We can also allow complex numbers in the theory of matrices and determinants. Let A be a complex $n \times n$ matrix. Then $A\boldsymbol{x} = \mathbf{0}$ has a non-zero solution $\boldsymbol{x} \in \mathbf{C}^n$ if and only if $\det A = 0$. In particular, if $\lambda \in \mathbf{C}$, then there exists a non-zero vector $\boldsymbol{x} \in \mathbf{C}^n$ such that $A\boldsymbol{x} = \lambda\boldsymbol{x}$ if and only if $\det(A - \lambda I) = 0$. Let A be a complex matrix. By $\overline{A}$ we shall mean the matrix obtained from A by taking the complex conjugates of the entries of A. We say that a square matrix A is Hermitian if $A^t = \overline{A}$.

Lemma 6.27. Let A be a Hermitian $n \times n$ matrix. Then all the zeros of the characteristic polynomial $\det(A - \lambda I)$ are real.

Proof. Let λ be a zero of the characteristic polynomial. Then there exists a non-zero vector $\boldsymbol{x} \in \mathbf{C}^n$ such that

$$A\boldsymbol{x} = \lambda\boldsymbol{x}. \tag{6.5}$$

Taking the complex conjugate and using the assumption on A, we get

$$A^t\,\overline{\boldsymbol{x}} = \overline{A}\,\overline{\boldsymbol{x}} = \overline{A\boldsymbol{x}} = \overline{\lambda\boldsymbol{x}} = \overline{\lambda}\,\overline{\boldsymbol{x}}. \tag{6.6}$$

By (6.5),

$$(A\boldsymbol{x}) \cdot \boldsymbol{x} = (A\boldsymbol{x})^t\, \overline{\boldsymbol{x}} = (\lambda \boldsymbol{x})^t\, \overline{\boldsymbol{x}} = \lambda \boldsymbol{x}^t\, \overline{\boldsymbol{x}} = \lambda \|\boldsymbol{x}\|^2,$$

and by (6.6),

$$(A\boldsymbol{x}) \cdot \boldsymbol{x} = (A\boldsymbol{x})^t\, \overline{\boldsymbol{x}} = \boldsymbol{x}^t A^t\, \overline{\boldsymbol{x}} = \boldsymbol{x}^t \overline{\lambda}\, \overline{\boldsymbol{x}} = \overline{\lambda} \boldsymbol{x}^t\, \overline{\boldsymbol{x}} = \overline{\lambda} \|\boldsymbol{x}\|^2.$$

Therefore $\lambda\|\boldsymbol{x}\|^2 = \overline{\lambda}\|\boldsymbol{x}\|^2$, and since $\boldsymbol{x} \neq \boldsymbol{0}$, we find that $\lambda = \overline{\lambda}$. Hence λ is real. ■

From now on, linear spaces and matrices are real.

Lemma 6.28. Let A be a symmetric $n \times n$ matrix. Then A has an eigenvalue.

Proof. Since A is real, A is Hermitian. Therefore, by Lemma 6.27, the characteristic polynomial $\det(A - \lambda I)$ has real zeros. Hence, A has at least one eigenvalue. ■

Theorem 6.29 (Spectral theorem). Let F be a symmetric linear transformation on a non-zero finite-dimensional inner product space V. Then there exists an orthonormal basis for V of eigenvectors of F.

Proof. We prove the statement by induction on $\dim V = n$. If $n = 1$, then V has an orthonormal basis $\boldsymbol{e}$. Since $F(\boldsymbol{e}) \in V$, we have $F(\boldsymbol{e}) = \lambda \boldsymbol{e}$ where λ is the coordinate of $F(\boldsymbol{e})$ with respect to $\boldsymbol{e}$. Hence, the statement holds for $n = 1$.

Suppose that $\dim V = n \geq 2$ and that the statement holds for symmetric linear transformations on inner product spaces of dimension less than n. We can choose an orthonormal basis for V. Since F is symmetric, the $n \times n$ matrix A of F with respect to that basis is symmetric by Theorem 5.29. Hence, by Lemma 6.28, F has an eigenvalue λ_n. Let $\boldsymbol{e}_n$ be an eigenvector of length 1 belonging to λ_n. Set $V' = [\boldsymbol{e}_n]^\perp$ and let F' be the restriction of F to V'. If $\boldsymbol{u} \in V'$, then $\langle \boldsymbol{u}, \boldsymbol{e}_n \rangle = 0$. Since F is symmetric, we have

$$\langle F(\boldsymbol{u}), \boldsymbol{e}_n \rangle = \langle \boldsymbol{u}, F(\boldsymbol{e}_n) \rangle = \langle \boldsymbol{u}, \lambda \boldsymbol{e}_n \rangle = \lambda \langle \boldsymbol{u}, \boldsymbol{e}_n \rangle = 0,$$

which means that $F(\boldsymbol{u}) \in V'$. Hence, F' is a symmetric linear transformation on V'. Since $\dim V' = n - 1$, it follows from the induction hypothesis that V' has an orthonormal basis $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_{n-1}$ of eigenvectors of F', hence also of F. Since these vectors are orthogonal to $\boldsymbol{e}_n$, we find that $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n$ form an orthonormal basis for V. ■

The converse of the spectral theorem also holds. For if there exists an orthonormal basis for V of eigenvectors of F, then the matrix with respect to that basis is a diagonal matrix D. Since D is symmetric and the basis is orthonormal, F is a symmetric linear transformation.

Corollary 6.30. Let A be an $n \times n$ matrix. Then A is symmetric if and only if $\mathbf{R}^n$ has an orthonormal basis of eigenvectors of A.

Theorem 6.31. Let A be a symmetric $n \times n$ matrix. Then there exist an orthogonal matrix T and a diagonal matrix D such that $T^tAT = D$.

Proof. A is the matrix of the linear transformation $\boldsymbol{x} \mapsto A\boldsymbol{x}$ on $\mathbf{R}^n$ with respect to the orthonormal basis $\boldsymbol{\varepsilon}_1, \ldots, \boldsymbol{\varepsilon}_n$. Let D be the diagonal matrix of A with respect to an orthonormal basis $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n$ for $\mathbf{R}^n$ of eigenvectors of A and let T be the transition matrix from $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n$ to $\boldsymbol{\varepsilon}_1, \ldots, \boldsymbol{\varepsilon}_n$. Then T is orthogonal. Hence, $D = T^{-1}AT = T^tAT$. ■

Theorem 6.32. Let F be a symmetric linear transformation on an inner product space and let $\boldsymbol{e}_1$ and $\boldsymbol{e}_2$ be eigenvectors belonging to two different eigenvalues λ_1 and λ_2. Then $\langle \boldsymbol{e}_1, \boldsymbol{e}_2 \rangle = 0$.

Proof. We have

$$\lambda_1\langle \boldsymbol{e}_1, \boldsymbol{e}_2 \rangle = \langle \lambda_1\boldsymbol{e}_1, \boldsymbol{e}_2 \rangle = \langle F(\boldsymbol{e}_1), \boldsymbol{e}_2 \rangle = \langle \boldsymbol{e}_1, F(\boldsymbol{e}_2) \rangle = \langle \boldsymbol{e}_1, \lambda_2\boldsymbol{e}_2 \rangle = \lambda_2\langle \boldsymbol{e}_1, \boldsymbol{e}_2 \rangle.$$

Since $\lambda_1 \neq \lambda_2$, it follows that $\langle \boldsymbol{e}_1, \boldsymbol{e}_2 \rangle = 0$. ■

Example 6.33. The matrix

$$A = \begin{bmatrix} 3 & 1 & 2 \\ 1 & 3 & 2 \\ 2 & 2 & 6 \end{bmatrix}$$

is symmetric. Our goal is to find an orthogonal matrix T and a diagonal matrix D so that $D = T^tAT$.

$$\begin{vmatrix} 3-\lambda & 1 & 2 \\ 1 & 3-\lambda & 2 \\ 2 & 2 & 6-\lambda \end{vmatrix} = \begin{vmatrix} 3-\lambda & 1 & 2 \\ \lambda-2 & 2-\lambda & 0 \\ 2 & 2 & 6-\lambda \end{vmatrix} = \begin{vmatrix} 3-\lambda & 4-\lambda & 2 \\ \lambda-2 & 0 & 0 \\ 2 & 4 & 6-\lambda \end{vmatrix}$$
$$= -(\lambda-2)^2(\lambda-8).$$

The eigenvalues are $\lambda_1 = 2$ and $\lambda_2 = 8$. The algebraic multiplicity of λ_1 is 2, and since A is symmetric, also its geometric multiplicity must be 2. We seek its eigenspace.

$$\left[\begin{array}{ccc|c} 1 & 1 & 2 & 0 \\ 1 & 1 & 2 & 0 \\ 2 & 2 & 4 & 0 \end{array}\right] \quad \Leftrightarrow \quad x_1 + x_2 + 2x_3 = 0 \quad \Leftrightarrow \quad \boldsymbol{x} = s(1,-1,0) + t(2,0,-1).$$

The vectors $\boldsymbol{v}_1 = (1,-1,0)$ and $\boldsymbol{v}_2 = (2,0,-1)$ form a basis for this eigenspace. We apply the Gram–Schmidt process to them. We set $\boldsymbol{u}_1 = \boldsymbol{v}_1$ and $\boldsymbol{u}_2 = s\boldsymbol{u}_1 + \boldsymbol{v}_2$ and get $s = -1$ and $\boldsymbol{u}_2 = -\boldsymbol{u}_1 + \boldsymbol{v}_2 = (1,1,-1)$. Normalising $\boldsymbol{u}_1$ and $\boldsymbol{u}_2$, we get the orthonormal basis $\boldsymbol{e}_1 = \frac{1}{\sqrt{2}}(1,-1,0)$, $\boldsymbol{e}_2 = \frac{1}{\sqrt{3}}(1,1,-1)$ for the eigenspace $\ker(A - \lambda_1 I)$. By Theorem 6.32, we know that the eigenvectors belonging to λ_2 are orthogonal to $\boldsymbol{e}_1$ and $\boldsymbol{e}_2$. Hence, the unit normal vector $\boldsymbol{e}_3 = \frac{1}{\sqrt{6}}(1,1,2)$ of the plane $x_1+x_2+2x_3 = 0$ must be an eigenvector belonging to λ_2. With

$$T = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{6}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{6}} \\ 0 & -\frac{1}{\sqrt{3}} & \frac{2}{\sqrt{6}} \end{bmatrix} = \frac{1}{\sqrt{6}} \begin{bmatrix} \sqrt{3} & \sqrt{2} & 1 \\ -\sqrt{3} & \sqrt{2} & 1 \\ 0 & -\sqrt{2} & 2 \end{bmatrix} \quad \text{and} \quad D = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 8 \end{bmatrix},$$

we therefore have $D = T^tAT$.

6.6 Systems of Linear Differential Equations

Consider the system

$$\begin{cases} x_1'(t) = a_{11}x_1(t) + a_{12}x_2(t) + \cdots + a_{1n}x_n(t) \\ x_2'(t) = a_{21}x_1(t) + a_{22}x_2(t) + \cdots + a_{2n}x_n(t) \\ \qquad \vdots \\ x_n'(t) = a_{n1}x_1(t) + a_{n2}x_2(t) + \cdots + a_{nn}x_n(t) \end{cases}$$

of first-order linear differential equations. For every $t \in \mathbf{R}$, $\boldsymbol{x}(t) = (x_1(t), \ldots, x_n(t))$ and $\boldsymbol{x}'(t) = (x_1'(t), \ldots, x_n'(t))$ are elements of $\mathbf{R}^n$, and the system can be written as $\boldsymbol{x}'(t) = A\boldsymbol{x}(t)$ where $A = [a_{ik}]_{n \times n}$. Suppose that $\mathbf{R}^n$ has a basis $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n$ consisting of eigenvectors of A and let $\boldsymbol{y}(t) = (y_1(t), \ldots, y_n(t))$ be the coordinates of $\boldsymbol{x}(t)$ with respect to that basis. Then

$$\boldsymbol{x}(t) = y_1(t)\boldsymbol{e}_1 + \cdots + y_n(t)\boldsymbol{e}_n,$$

and hence

$$\boldsymbol{x}'(t) = y_1'(t)\boldsymbol{e}_1 + \cdots + y_n'(t)\boldsymbol{e}_n.$$

If the eigenvalue associated with $\boldsymbol{e}_i$ is λ_i for $i = 1, \ldots, n$, then

$$\boldsymbol{x}'(t) = A\boldsymbol{x}(t) = A(y_1(t)\boldsymbol{e}_1 + \cdots + y_n(t)\boldsymbol{e}_n) = \lambda_1 y_1(t)\boldsymbol{e}_1 + \cdots + \lambda_n y_n(t)\boldsymbol{e}_n,$$

and therefore

$$y_1'(t)\boldsymbol{e}_1 + \cdots + y_n'(t)\boldsymbol{e}_n = \lambda_1 y_1(t)\boldsymbol{e}_1 + \cdots + \lambda_n y_n(t)\boldsymbol{e}_n.$$

Since the coordinates are unique, we get

$$y_i'(t) = \lambda_i y_i(t), \quad i = 1, \ldots, n. \tag{6.7}$$

Hence,

$$y_i(t) = c_i e^{\lambda_i t}, \quad i = 1, \ldots, n,$$

for some constants c_i, and thus

$$\boldsymbol{x}(t) = c_1 e^{\lambda_1 t}\boldsymbol{e}_1 + \cdots + c_n e^{\lambda_n t}\boldsymbol{e}_n.$$

Example 6.34. Let us solve the following system of differential equations.

$$\begin{cases} x_1'(t) = x_1(t) + 3x_2(t) + 2x_3(t) \\ x_2'(t) = -x_1(t) + 2x_2(t) + x_3(t) \\ x_3'(t) = 4x_1(t) - x_2(t) - x_3(t) \end{cases}.$$

Proceeding as usual, we find that the eigenvalues of

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

are $\lambda_1 = -2$, $\lambda_2 = 1$, $\lambda_3 = 3$ with associated eigenvectors $\boldsymbol{e}_1 = (1, 1, -3)$, $\boldsymbol{e}_2 = (1, -2, 3)$, $\boldsymbol{e}_3 = (1, 0, 1)$. Since the eigenvalues are distinct, the eigenvectors form a basis for $\mathbf{R}^3$. By applying the above method, we therefore get

$$\begin{bmatrix} x_1(t) \\ x_2(t) \\ x_3(t) \end{bmatrix} = c_1 e^{-2t} \begin{bmatrix} 1 \\ 1 \\ -3 \end{bmatrix} + c_2 e^{t} \begin{bmatrix} 1 \\ -2 \\ 3 \end{bmatrix} + c_3 e^{3t} \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}.$$

Given the initial conditions $x_1(0) = 6$, $x_2(0) = -3$ and $x_3(0) = 6$, we obtain

$$\begin{bmatrix} 6 \\ -3 \\ 6 \end{bmatrix} = c_1 \begin{bmatrix} 1 \\ 1 \\ -3 \end{bmatrix} + c_2 \begin{bmatrix} 1 \\ -2 \\ 3 \end{bmatrix} + c_3 \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}.$$

Solving for c_1, c_2 and c_3, we get $c_1 = 1$, $c_2 = 2$, $c_3 = 3$. Hence, the solution of this initial condition problem is

$$\begin{bmatrix} x_1(t) \\ x_2(t) \\ x_3(t) \end{bmatrix} = e^{-2t} \begin{bmatrix} 1 \\ 1 \\ -3 \end{bmatrix} + 2e^{t} \begin{bmatrix} 1 \\ -2 \\ 3 \end{bmatrix} + 3e^{3t} \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}.$$

Exactly the same method can be used to solve a system of the form

$$\begin{cases} x_1''(t) = a_{11}x_1(t) + a_{12}x_2(t) + \cdots + a_{1n}x_n(t) \\ x_2''(t) = a_{21}x_1(t) + a_{22}x_2(t) + \cdots + a_{2n}x_n(t) \\ \qquad \vdots \\ x_n''(t) = a_{n1}x_1(t) + a_{n2}x_2(t) + \cdots + a_{nn}x_n(t) \end{cases}.$$

(6.7) is now replaced by $y_i''(t) = \lambda_i y_i(t)$, $i = 1, \ldots, n$. We can solve these second-order equations and proceed as before.

Example 6.35. Consider the system

$$\begin{cases} x_1''(t) = x_1(t) + x_2(t) \\ x_2''(t) = 3x_1(t) - x_2(t) \end{cases}.$$

This time, the eigenvalues and associated eigenvectors are $\lambda_1 = -2$, $\lambda_2 = 2$, $\boldsymbol{e}_1 = (1, -3)$, $\boldsymbol{e}_2 = (1, 1)$. We have

$$y_1''(t) = -2y_1(t) \quad \Leftrightarrow \quad y_1(t) = c_1 \sin\sqrt{2}\,t + c_2 \cos\sqrt{2}\,t$$

and

$$y_2''(t) = 2y_2(t) \quad \Leftrightarrow \quad y_2(t) = d_1 e^{\sqrt{2}\,t} + d_2 e^{-\sqrt{2}\,t}.$$

Hence,

$$\begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix} = \left(c_1 \sin\sqrt{2}\,t + c_2 \cos\sqrt{2}\,t\right) \begin{bmatrix} 1 \\ -3 \end{bmatrix} + \left(d_1 e^{\sqrt{2}\,t} + d_2 e^{-\sqrt{2}\,t}\right) \begin{bmatrix} 1 \\ 1 \end{bmatrix}.$$

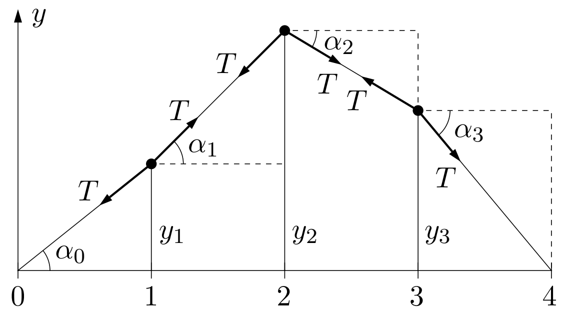
6.7 The Vibrating String

Consider a massless elastic string of length $n+1$ stretched between two points. Suppose that n point masses, each of mass m, are attached to the string at distances 1 from each other and from the ends. If the point masses perform small transverse oscillations, the tension T in the string is approximately constant. Disregarding gravity and using the notation in the figure below, we see that the mass at the point j is influenced in the y-direction by the force

$$-T\sin\alpha_{j-1} + T\sin\alpha_j, \quad j = 1, \dots, n.$$

Since $y_0 = y_{n+1} = 0$ and the displacements are supposed to be small, we have

$$\sin\alpha_j \approx \tan\alpha_j = y_{j+1} - y_j, \quad j = 0, \dots, n.$$



$n = 3$ point masses

Hence, the force exerted on the mass at j is approximately

$$-T(y_j - y_{j-1}) + T(y_{j+1} - y_j) = T(y_{j-1} - 2y_j + y_{j+1}), \quad j = 1, \dots, n.$$

According to Newton's second law, force equals mass times acceleration. Therefore,

$$T(y_{j-1}(t) - 2y_j(t) + y_{j+1}(t)) = my_j''(t), \quad j = 1, \dots, n.$$

Setting $q = \sqrt{\frac{T}{m}}$, we can write this as

$$y_j''(t) = q^2(y_{j-1}(t) - 2y_j(t) + y_{j+1}(t)), \quad j = 1, \dots, n,$$

and thus

$$\begin{bmatrix} y_1''(t) \\ y_2''(t) \\ y_3''(t) \\ y_4''(t) \\ \vdots \\ y_n''(t) \end{bmatrix} = q^2 \begin{bmatrix} -2 & 1 & 0 & 0 & \cdots & 0 \\ 1 & -2 & 1 & 0 & \cdots & 0 \\ 0 & 1 & -2 & 1 & \cdots & 0 \\ 0 & 0 & 1 & -2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & 0 & \cdots & -2 \end{bmatrix} \begin{bmatrix} y_1(t) \\ y_2(t) \\ y_3(t) \\ y_4(t) \\ \vdots \\ y_n(t) \end{bmatrix}.$$

Since this $n \times n$ matrix A is symmetric, $\mathbf{R}^n$ has an orthonormal basis $\boldsymbol{e}_1, \dots, \boldsymbol{e}_n$ consisting of eigenvectors of A. Let $\lambda_1, \dots, \lambda_n$ be the associated eigenvalues. If $(z_1(t), \dots, z_n(t))$ are the coordinates of $(y_1(t), \dots, y_n(t))$ with respect to $\boldsymbol{e}_1, \dots, \boldsymbol{e}_n$, we get

$$z_j''(t) = \lambda_j z(t), \quad j = 1, \dots, n.$$

We shall show in Section 7.6 that the eigenvalues are negative. Thus, with $k_j = \sqrt{-\lambda_j}$, we have

$$z_j''(t) + k_j^2 z_j(t) = 0, \quad j = 1, \dots, n.$$

The solutions of this second-order differential equation are given by

$$z_j(t) = a_j e^{ik_j t} + b_j e^{-ik_j t} = c_j \sin(k_j t + \delta_j), \quad j = 1, \dots, n.$$

Hence,

$$\boldsymbol{y}(t) = (y_1(t), \dots, y_n(t)) = c_1 \sin(k_1 t + \delta_1)\boldsymbol{e}_1 + \dots + c_n \sin(k_n t + \delta_n)\boldsymbol{e}_n.$$

The solution $\boldsymbol{y}(t) = c_j \sin(k_j t + \delta_j)\boldsymbol{e}_j$, for which all coefficients except c_j are zero, is called an eigenmode with eigenfrequency $\frac{k_j}{2\pi}$.

Example 6.36. Consider a string with two point masses. The corresponding matrix is

$$A = q^2 \begin{bmatrix} -2 & 1 \\ 1 & -2 \end{bmatrix}.$$

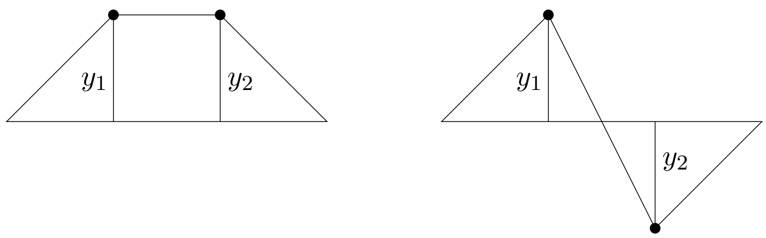
Its eigenvalues are $\lambda_1 = -q^2$ and $\lambda_2 = -3q^2$ with associated eigenvectors

$$\boldsymbol{e}_1 = (1, 1) \quad \text{and} \quad \boldsymbol{e}_2 = (1, -1).$$

The two eigenmodes

$$c_1 \sin(qt + \delta_1)\boldsymbol{e}_1 \quad \text{and} \quad c_2 \sin(\sqrt{3}\, qt + \delta_2)\boldsymbol{e}_2$$

are shown below.



Exercises

6.1. Find the eigenvalues and eigenvectors of the following matrices.

(a) $\begin{bmatrix} 5 & -2 \\ 6 & -2 \end{bmatrix}$, (b) $\begin{bmatrix} 3 & -2 \\ 4 & -1 \end{bmatrix}$,

(c) $\begin{bmatrix} -1 & 2 & 1 \\ 2 & -1 & 1 \\ -1 & 1 & 2 \end{bmatrix}$, (d) $\begin{bmatrix} 4 & -1 & 0 \\ 4 & 0 & 0 \\ 2 & -1 & 2 \end{bmatrix}$.

6.2. Let A be an invertible square matrix. Show that if λ is an eigenvalue of A, then λ^{-1} is an eigenvalue of A^{-1}.

6.3. (a) What eigenvalues can a projection have?

(b) What eigenvalues can a reflection have?

6.4. Let A be a square matrix such that the sum of the entries in each row equals λ. Show that λ is an eigenvalue of A.

6.5. Let F be a linear transformation on a linear space V and assume that every non-zero vector of V is an eigenvector of F. Show that there exists a real number λ such that $F = \lambda I$.

6.6. Determine, for each of the following matrices A, whether it is diagonalisable, and when it is, find a matrix T such that $T^{-1}AT$ is a diagonal matrix.

(a) $\begin{bmatrix} -3 & 3 & 1 \\ 4 & -1 & -2 \\ -14 & 9 & 6 \end{bmatrix}$, (b) $\begin{bmatrix} 3 & -1 & -1 \\ 4 & -1 & -2 \\ -2 & 1 & 2 \end{bmatrix}$, (c) $\begin{bmatrix} 1 & 4 & 6 \\ -3 & -7 & -7 \\ 4 & 8 & 7 \end{bmatrix}$.

6.7. Find a diagonalisable 3×3 matrix with eigenvalues 1, 2 and 4 and associated eigenvectors $(1, 1, 1)$, $(1, 2, -1)$ and $(1, 0, 2)$, respectively.

6.8. Compute A^n for every positive integer n, where

$$A = \begin{bmatrix} 2 & 2 \\ 3 & 1 \end{bmatrix}.$$

6.9. Let A be a diagonalisable square matrix with non-negative eigenvalues. Show that there exists a square matrix B such that $B^2 = A$.

6.10. Let A be a diagonalisable square matrix. Show that A^t is diagonalisable with the same eigenvalues as A.

6.11. Let V be an n-dimensional inner product space, where $n > 0$, and let F be the linear transformation on V defined by

$$F(\boldsymbol{u}) = \langle \boldsymbol{u}, \boldsymbol{c} \rangle \boldsymbol{b} - \langle \boldsymbol{b}, \boldsymbol{c} \rangle \boldsymbol{u}$$

where $\boldsymbol{b}$ and $\boldsymbol{c}$ are vectors of V for which $\langle \boldsymbol{b}, \boldsymbol{c} \rangle \neq 0$. Show that V has a basis consisting of eigenvectors of F and find the matrix of F with respect to some such basis.

6.12. Solve the recurrence problem

$$\begin{cases} a_{n+1} = 2a_n + 2b_n \\ b_{n+1} = 10a_n + 3b_n \end{cases}, \quad \begin{cases} a_0 = 4 \\ b_0 = 1 \end{cases}.$$

6.13. Solve the recurrence problem

$$\begin{cases} a_{n+1} = a_n + 2b_n + 2c_n \\ b_{n+1} = 3a_n + b_n + 9c_n \\ c_{n+1} = 2a_n + 2b_n + c_n \end{cases}, \quad \begin{cases} a_0 = 0 \\ b_0 = 5 \\ c_0 = 8 \end{cases}.$$

6.14. Solve the recurrence problem

$$a_{n+3} = a_{n+2} + 4a_{n+1} - 4a_n, \quad a_0 = 2,\ a_1 = 7,\ a_2 = 5.$$

6.15. Find, for each of the following matrices A, an orthogonal matrix T such that T^tAT is a diagonal matrix.

(a) $\begin{bmatrix} 1 & 0 & 3 \\ 0 & 1 & 4 \\ 3 & 4 & 1 \end{bmatrix}$, (b) $\begin{bmatrix} 13 & -4 & -2 \\ -4 & 13 & 2 \\ -2 & 2 & 10 \end{bmatrix}$.

6.16. Let A be an invertible square matrix.

(a) Show that the eigenvalues of the symmetric matrix A^tA are positive. Hint: $\boldsymbol{x}^tA^tA\boldsymbol{x} = \|A\boldsymbol{x}\|^2$.

(b) Show that there exists a unique symmetric matrix B with positive eigenvalues such that $B^2 = A^tA$. Hint: Show that $B\boldsymbol{x} = \sqrt{\lambda}\,\boldsymbol{x}$ if $A^tA\boldsymbol{x} = \lambda\boldsymbol{x}$.

(c) Show that $A = QB$ where Q is an orthogonal matrix and B a symmetric matrix with positive eigenvalues. Hint: Try $Q = (A^{-1})^tB$.

6.17. Solve the following initial value problem.

$$\begin{cases} x_1'(t) = x_1(t) + 3x_2(t) + 2x_3(t) \\ x_2'(t) = 3x_1(t) - 4x_2(t) + 3x_3(t) \\ x_3'(t) = 2x_1(t) + 3x_2(t) + x_3(t) \end{cases}, \quad \begin{cases} x_1(0) = 8 \\ x_2(0) = -5 \\ x_3(0) = 10 \end{cases}.$$

6.18. Find the general solution of the following system of differential equations.

$$\begin{cases} x_1''(t) = x_1(t) + 2x_2(t) + x_3(t) \\ x_2''(t) = 2x_1(t) + x_2(t) + x_3(t) \\ x_3''(t) = 3x_1(t) + 3x_2(t) + 4x_3(t) \end{cases}.$$

6.19. Find the eigenfrequencies and describe the corresponding eigenmodes for a string with three point masses.

7 Quadratic Forms

7.1 Bilinear Forms

Definition 7.1. Let V be a linear space. A bilinear form on V is a 2-multilinear form on V.

The definition means that a bilinear form b on V is a function $b : V \times V \to \mathbf{R}$ such that

$$b(s\boldsymbol{u} + t\boldsymbol{v}, \boldsymbol{w}) = sb(\boldsymbol{u}, \boldsymbol{w}) + tb(\boldsymbol{v}, \boldsymbol{w}) \quad \text{and} \quad b(\boldsymbol{w}, s\boldsymbol{u} + t\boldsymbol{v}) = sb(\boldsymbol{w}, \boldsymbol{u}) + tb(\boldsymbol{w}, \boldsymbol{v})$$

for all vectors $\boldsymbol{u}$, $\boldsymbol{v}$ and $\boldsymbol{w}$ of V and all real numbers s and t.

Example 7.2. The function $b : \mathbf{R}^n \times \mathbf{R}^n \to \mathbf{R}$ defined by

$$b(\boldsymbol{x}, \boldsymbol{y}) = \sum_{i=1}^{n} \sum_{k=1}^{n} b_{ik} x_i y_k$$

where b_{ik}, $i = 1, \ldots, n$, $k = 1, \ldots, n$, are real numbers is a bilinear form on $\mathbf{R}^n$ as is easily verified. If we think of $\boldsymbol{x}$ and $\boldsymbol{y}$ as columns, we can write

$$b(\boldsymbol{x}, \boldsymbol{y}) = \boldsymbol{x}^t B \boldsymbol{y}$$

where $B = [b_{ik}]$.

Theorem 7.3. Let b be a bilinear form on a linear space V with basis $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n$. If the coordinates of $\boldsymbol{u}$ and $\boldsymbol{v}$ with respect to that basis are $\boldsymbol{x}$ and $\boldsymbol{y}$, respectively, then

$$b(\boldsymbol{u}, \boldsymbol{v}) = \boldsymbol{x}^t B \boldsymbol{y}$$

where

$$B = \begin{bmatrix} b(\boldsymbol{e}_1, \boldsymbol{e}_1) & \cdots & b(\boldsymbol{e}_1, \boldsymbol{e}_n) \\ \vdots & & \vdots \\ b(\boldsymbol{e}_n, \boldsymbol{e}_1) & \cdots & b(\boldsymbol{e}_n, \boldsymbol{e}_n) \end{bmatrix}.$$

Proof. We have by assumption that

$$\boldsymbol{u} = \sum_{i=1}^{n} x_i \boldsymbol{e}_i \quad \text{and} \quad \boldsymbol{v} = \sum_{k=1}^{n} y_k \boldsymbol{e}_k.$$

Hence, by bilinearity,

$$\begin{aligned} b(\boldsymbol{u}, \boldsymbol{v}) &= b\left(\sum_{i=1}^{n} x_i \boldsymbol{e}_i, \sum_{k=1}^{n} y_k \boldsymbol{e}_k\right) = \sum_{i=1}^{n} x_i b\left(\boldsymbol{e}_i, \sum_{k=1}^{n} y_k \boldsymbol{e}_k\right) \\ &= \sum_{i=1}^{n} \sum_{k=1}^{n} x_i y_k b(\boldsymbol{e}_i, \boldsymbol{e}_k) = \boldsymbol{x}^t B \boldsymbol{y}. \ \blacksquare \end{aligned}$$

Definition 7.4. Let b be a bilinear form on a linear space V and let $\boldsymbol{e}_1, \dots, \boldsymbol{e}_n$ be a basis for V. The matrix B in Theorem 7.3 is called the matrix of b with respect to the basis $\boldsymbol{e}_1, \dots, \boldsymbol{e}_n$.

By using the basis $\boldsymbol{\varepsilon}_1, \dots, \boldsymbol{\varepsilon}_n$ for $\mathbf{R}^n$, we see that every bilinear form on $\mathbf{R}^n$ is of the form described in Example 7.2.

Definition 7.5. A bilinear form b on a linear space V is said to be symmetric if

$$b(\boldsymbol{u}, \boldsymbol{v}) = b(\boldsymbol{v}, \boldsymbol{u})$$

for all $\boldsymbol{u}$ and $\boldsymbol{v}$ in V.

Theorem 7.6. Let b be a bilinear form on a linear space V with basis $\boldsymbol{e}_1, \dots, \boldsymbol{e}_n$ and let B be the matrix of b with respect to that basis. Then b is symmetric if and only if B is symmetric.

Proof. Let $\boldsymbol{x}$ and $\boldsymbol{y}$ be the coordinates of $\boldsymbol{u}$ and $\boldsymbol{v}$, respectively. Then $b(\boldsymbol{u}, \boldsymbol{v}) = \boldsymbol{x}^t B \boldsymbol{y}$ and $b(\boldsymbol{v}, \boldsymbol{u}) = \boldsymbol{y}^t B \boldsymbol{x} = \boldsymbol{x}^t (\boldsymbol{y}^t B)^t = \boldsymbol{x}^t B^t \boldsymbol{y}$. Hence, b is symmetric if and only if $\boldsymbol{x}^t B \boldsymbol{y} = \boldsymbol{x}^t B^t \boldsymbol{y}$ for all $\boldsymbol{x}$ and $\boldsymbol{y}$ in $\mathbf{R}^n$, and this is equivalent to $B = B^t$. $\blacksquare$

Theorem 7.7. Let b be a bilinear form on a linear space V with bases $\boldsymbol{e}_1, \dots, \boldsymbol{e}_n$ and $\boldsymbol{e}_1', \dots, \boldsymbol{e}_n'$. If B is the matrix of b with respect to $\boldsymbol{e}_1, \dots, \boldsymbol{e}_n$ and T is the transition matrix from $\boldsymbol{e}_1', \dots, \boldsymbol{e}_n'$ to $\boldsymbol{e}_1, \dots, \boldsymbol{e}_n$, then the matrix of b with respect to $\boldsymbol{e}_1', \dots, \boldsymbol{e}_n'$ is

$$B' = T^t B T.$$

Proof. If the coordinates of $\boldsymbol{u}$ and $\boldsymbol{v}$ with respect to the bases are $\boldsymbol{x}$, $\boldsymbol{y}$ and $\boldsymbol{x}'$, $\boldsymbol{y}'$, then

$$b(\boldsymbol{u}, \boldsymbol{v}) = \boldsymbol{x}^t B \boldsymbol{y} = (T\boldsymbol{x}')^t B T \boldsymbol{y}' = (\boldsymbol{x}')^t T^t B T \boldsymbol{y}'.$$

This shows that the matrix with respect to $\boldsymbol{e}_1', \dots, \boldsymbol{e}_n'$ is $B' = T^t B T$. $\blacksquare$

7.2 Definition of Quadratic Forms

Definition 7.8. Let V be a linear space. A function $q : V \to \mathbf{R}$ is a quadratic form on V if there exists a bilinear form b on V such that

$$q(\boldsymbol{u}) = b(\boldsymbol{u}, \boldsymbol{u}), \quad \boldsymbol{u} \in V.$$

Example 7.9. The function $q : \mathbf{R}^3 \to \mathbf{R}$ defined by

$$q(\boldsymbol{x}) = x_1^2 + 2x_2^2 + x_3^2 + 2x_1x_2 + 4x_1x_3 + 6x_2x_3$$

is a quadratic form on $\mathbf{R}^3$ since $q(\boldsymbol{x}) = b(\boldsymbol{x}, \boldsymbol{x})$ where b is the bilinear form on $\mathbf{R}^3$ defined by

$$b(\boldsymbol{x}, \boldsymbol{y}) = x_1y_1 + 2x_2y_2 + x_3y_3 + 2x_1y_2 + 4x_1y_3 + 6x_2y_3.$$

We also have $q(\boldsymbol{x}) = c(\boldsymbol{x}, \boldsymbol{x})$ where c is the symmetric bilinear form on $\mathbf{R}^3$ defined by

$$c(\boldsymbol{x}, \boldsymbol{y}) = x_1y_1 + 2x_2y_2 + x_3y_3 + x_1y_2 + x_2y_1 + 2x_1y_3 + 2x_3y_1 + 3x_2y_3 + 3x_3y_2.$$

Theorem 7.10. Let q be a quadratic form on a linear space V. Then there exists a unique symmetric bilinear form c on V such that $q(\boldsymbol{u}) = c(\boldsymbol{u}, \boldsymbol{u})$ for all $\boldsymbol{u} \in V$.

Proof. Let b be any bilinear form on V such that $q(\boldsymbol{u}) = b(\boldsymbol{u}, \boldsymbol{u})$ for all $\boldsymbol{u} \in V$ and define c by $c(\boldsymbol{u}, \boldsymbol{v}) = \frac{1}{2}(b(\boldsymbol{u}, \boldsymbol{v}) + b(\boldsymbol{v}, \boldsymbol{u}))$. Then c is symmetric and $q(\boldsymbol{u}) = c(\boldsymbol{u}, \boldsymbol{u})$ for all $\boldsymbol{u} \in V$. To show the uniqueness, we let c be any symmetric bilinear form on V for which $q(\boldsymbol{u}) = c(\boldsymbol{u}, \boldsymbol{u})$ for all $\boldsymbol{u} \in V$. Then

$$\begin{aligned} q(\boldsymbol{u} + \boldsymbol{v}) &= c(\boldsymbol{u} + \boldsymbol{v}, \boldsymbol{u} + \boldsymbol{v}) = c(\boldsymbol{u}, \boldsymbol{u}) + c(\boldsymbol{v}, \boldsymbol{v}) + c(\boldsymbol{u}, \boldsymbol{v}) + c(\boldsymbol{v}, \boldsymbol{u}) \\ &= q(\boldsymbol{u}) + q(\boldsymbol{v}) + 2c(\boldsymbol{u}, \boldsymbol{v}). \end{aligned}$$

Consequently,

$$c(\boldsymbol{u}, \boldsymbol{v}) = \frac{q(\boldsymbol{u} + \boldsymbol{v}) - q(\boldsymbol{u}) - q(\boldsymbol{v})}{2}$$

is uniquely determined by q. ■

Definition 7.11. Let q be a quadratic form on a linear space V with basis $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n$ and let b be the symmetric bilinear form associated with q. The matrix of b with respect to $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n$ is called the matrix of q with respect to $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n$.

Hence, the matrix of a quadratic form is symmetric.

Example 7.12. The matrix with respect to the basis $\boldsymbol{\varepsilon}_1, \boldsymbol{\varepsilon}_2, \boldsymbol{\varepsilon}_3$ of the quadratic form in Example 7.9 is

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

7.3 The Spectral Theorem Applied to Quadratic Forms

Definition 7.13. Let q be a quadratic form on a linear space V with basis $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n$. If the matrix of q with respect to that basis is a diagonal matrix, the basis is said to diagonalise q.

If the basis $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n$ diagonalises q and

$$B = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix}$$

is the corresponding diagonal matrix, then

$$q(x_1\boldsymbol{e}_1 + x_2\boldsymbol{e}_2 + \cdots + x_n\boldsymbol{e}_n) = \boldsymbol{x}^t B\boldsymbol{x} = \lambda_1 x_1^2 + \lambda_2 x_2^2 + \cdots + \lambda_n x_n^2.$$

Theorem 7.14. If q is a quadratic form on a non-zero finite-dimensional linear space V, then there exists a basis for V that diagonalises q.

Proof. Let $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n$ be any basis for V and let B be the matrix of q with respect to that basis. Since B is a symmetric matrix, it follows from Theorem 6.31 that there exist an orthogonal matrix T and a diagonal matrix D such that $T^tBT = D$. Let $\boldsymbol{e}'_k$ be the vector in V whose coordinate vector with respect to $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n$ is the kth column of T. Then the matrix of q with respect to the basis $\boldsymbol{e}'_1 \ldots, \boldsymbol{e}'_n$ is D. ■

If V is an inner product space and we start out with an orthonormal basis $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n$ in the above proof, then also $\boldsymbol{e}'_1, \ldots, \boldsymbol{e}'_n$ is an orthonormal basis since T is orthogonal. Thus we get the following theorem.

Theorem 7.15. If q is a quadratic form on a non-zero finite-dimensional inner product space V, then there exists an orthonormal basis for V that diagonalises q.

Theorem 7.16. Let q be a quadratic form on an inner product space V with orthonormal bases $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n$ and $\boldsymbol{f}_1, \ldots, \boldsymbol{f}_n$. If

$$\begin{aligned} q(x_1\boldsymbol{e}_1 + x_2\boldsymbol{e}_2 + \cdots + x_n\boldsymbol{e}_n) &= \lambda_1 x_1^2 + \lambda_2 x_2^2 + \cdots + \lambda_n x_n^2, \\ q(x_1\boldsymbol{f}_1 + x_2\boldsymbol{f}_2 + \cdots + x_n\boldsymbol{f}_n) &= \mu_1 x_1^2 + \mu_2 x_2^2 + \cdots + \mu_n x_n^2 \end{aligned}$$

for all $\boldsymbol{x} \in \mathbf{R}^n$ where $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n$ and $\mu_1 \leq \mu_2 \leq \cdots \leq \mu_n$, then $\lambda_i = \mu_i$ for $i = 1, 2, \ldots, n$.

Proof. It suffices to show that the matrices B and C of q with respect to the two bases have the same characteristic polynomial. Since both bases are orthonormal, the transition matrix T from $\boldsymbol{f}_1, \ldots, \boldsymbol{f}_n$ to $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n$ is orthogonal. Hence, $C = T^tBT = T^{-1}BT$, and it follows that

$$\det(C - \lambda I) = \det(T^{-1}BT - T^{-1}\lambda IT) = \det(T^{-1}(B - \lambda I)T) = \det(B - \lambda I). \ ■$$

Definition 7.17. If q is a quadratic form on a non-zero finite-dimensional inner product space V, then the eigenvalues of q are the eigenvalues of the matrix of q with respect to any orthonormal basis for V.

Example 7.18. Consider the quadratic form

$$q(\boldsymbol{x}) = 3x_2^2 + 3x_3^2 + 4x_1x_2 + 4x_1x_3 - 2x_2x_3$$

on $\mathbf{R}^3$. The matrix of q with respect to the orthonormal basis $\boldsymbol{\varepsilon}_1, \boldsymbol{\varepsilon}_2, \boldsymbol{\varepsilon}_3$ is

$$B = \begin{bmatrix} 0 & 2 & 2 \\ 2 & 3 & -1 \\ 2 & -1 & 3 \end{bmatrix}.$$

The eigenvalues are $\lambda_1 = -2$, $\lambda_2 = 4$, $\lambda_3 = 4$, and

$$\boldsymbol{e}_1' = \frac{1}{\sqrt{6}}(2, -1, -1), \quad \boldsymbol{e}_2' = \frac{1}{\sqrt{2}}(0, 1, -1), \quad \boldsymbol{e}_3' = \frac{1}{\sqrt{3}}(1, 1, 1)$$

form an orthonormal basis for $\mathbf{R}^3$ of eigenvectors of B. Hence, if $\boldsymbol{x} = x_1'\boldsymbol{e}_1' + x_2'\boldsymbol{e}_2' + x_3'\boldsymbol{e}_3'$, then

$$q(\boldsymbol{x}) = -2(x_1')^2 + 4(x_2')^2 + 4(x_3')^2.$$

Theorem 7.19. Let q be a quadratic form on an n-dimensional inner product space V and let $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n$ be its eigenvalues ordered in ascending order. Then

$$\lambda_1 = \min_{\|\boldsymbol{u}\|=1} q(\boldsymbol{u}) \quad \text{and} \quad \lambda_n = \max_{\|\boldsymbol{u}\|=1} q(\boldsymbol{u}).$$

Proof. There exists an orthonormal basis that diagonalises q, and its vectors $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n$ can be ordered so that $\boldsymbol{e}_i$ corresponds to λ_i for $i = 1, \ldots, n$. Let $\boldsymbol{u}$ be any vector of length 1 in V. If $(x_1, \ldots, x_n)$ are the coordinates of $\boldsymbol{u}$ with respect to $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n$, then $x_1^2 + \cdots + x_n^2 = \|\boldsymbol{u}\| = 1$ and $q(\boldsymbol{u}) = \lambda_1 x_1^2 + \cdots + \lambda_n x_n^2$. Hence,

$$\lambda_1 = \lambda_1(x_1^2 + \cdots + x_n^2) \leq q(\boldsymbol{u}) \leq \lambda_n(x_1^2 + \cdots + x_n^2) = \lambda_n$$

with equality in the left inequality if $x_2 = \cdots = x_n = 0$ and in the right inequality if $x_1 = \cdots = x_{n-1} = 0$. ■

Corollary 7.20. Using the same notation as in Theorem 7.19, we have

$$\lambda_1 = \min_{\boldsymbol{u} \neq \boldsymbol{0}} \frac{q(\boldsymbol{u})}{\|\boldsymbol{u}\|^2} \quad \text{and} \quad \lambda_n = \max_{\boldsymbol{u} \neq \boldsymbol{0}} \frac{q(\boldsymbol{u})}{\|\boldsymbol{u}\|^2}.$$

Proof. When $\boldsymbol{u}$ ranges over all non-zero vectors, $\frac{1}{\|\boldsymbol{u}\|}\boldsymbol{u}$ ranges over all vectors of length 1. The statement of the corollary now follows from the fact that

$$q\left(\frac{1}{\|\boldsymbol{u}\|}\boldsymbol{u}\right) = \frac{q(\boldsymbol{u})}{\|\boldsymbol{u}\|^2}. \blacksquare$$

Example 7.21. Consider anew the quadratic form

$$q(\boldsymbol{x}) = 3x_2^2 + 3x_3^2 + 4x_1x_2 + 4x_1x_3 - 2x_2x_3$$

in Example 7.18 with eigenvalues $-2 \le 4 \le 4$. Its minimum and maximum values subject to the constraint $x_1^2 + x_2^2 + x_3^2 = 1$ are -2 and 4, respectively. When

$$x_1^2 + x_2^2 + x_3^2 = (x_1')^2 + (x_2')^2 + (x_3')^2 = 1,$$

we have equality in the inequality

$$-2 \le -2(x_1')^2 + 4(x_2')^2 + 4(x_3')^2$$

if and only if $(x_1', x_2', x_3') = \pm(1, 0, 0)$. Hence, the minimum value is attained at the two points $\pm\frac{1}{\sqrt{6}}(2, -1, -1)$. In the inequality

$$-2(x_1')^2 + 4(x_2')^2 + 4(x_3')^2 \le 4,$$

we have equality if and only if $x_1' = 0$ and $(x_2')^2 + (x_3')^2 = 1$. The maximum value is, therefore, attained at all points on the unit circle centred at the origin and lying in the plane $2x_1 - x_2 - x_3 = 0$ perpendicular to the vector $\boldsymbol{e}_1'$.

7.4 Conic Sections

We shall here discuss circles, ellipses, hyperbolae and parabolae. These are the conic sections. The name 'conic section' stems from the fact that these curves can be obtained by intersecting a cone with a plane. We shall, however, take a different approach.

7.4.1 Circles

A circle is the set of all points P on a plane M at a given distance $r > 0$ from a given point C on M. The point C is called the centre of the circle and r its radius. If $O\boldsymbol{e}_1\boldsymbol{e}_2$ is an orthonormal coordinate system for M and the coordinates of P and C with respect to this system are (x, y) and (a, b), respectively, then $\overrightarrow{CP} = (x - a, y - b)$ and the distance from P to C equals $\left\|\overrightarrow{CP}\right\| = \sqrt{(x-a)^2 + (y-b)^2}$. Hence, the equation of the circle is

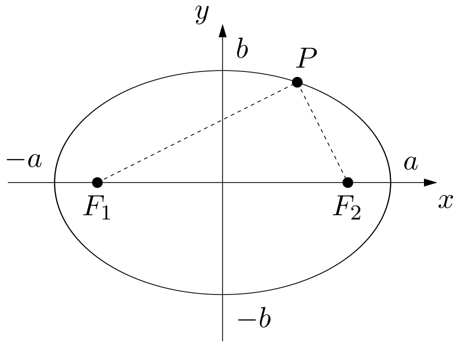
$$(x - a)^2 + (y - b)^2 = r^2.$$

7.4.2 Ellipses

Two different points F_1 and F_2 on a plane M and a real number a such that $2a > \left\|\overrightarrow{F_1F_2}\right\|$ define an ellipse on M. It is the set of points P on M for which

$$\left\|\overrightarrow{F_1P}\right\| + \left\|\overrightarrow{F_2P}\right\| = 2a.$$

Consider the orthonormal coordinate system for M whose x-axis passes through F_1 and F_2 and whose origin is the midpoint between these two points. With respect to this system, $F_1 = (-c, 0)$ and $F_2 = (c, 0)$ where $0 < |c| < a$.



With $b = \sqrt{a^2 - c^2}$, we shall show that the equation of the ellipse with respect to this system is

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1. \tag{7.1}$$

If $P = (x, y)$, then

$$\left\|\overrightarrow{F_1 P}\right\| = \sqrt{(x+c)^2 + y^2} = \sqrt{x^2 + y^2 + c^2 + 2xc}$$

and

$$\left\|\overrightarrow{F_2 P}\right\| = \sqrt{(x-c)^2 + y^2} = \sqrt{x^2 + y^2 + c^2 - 2xc}\,.$$

Hence, P lies on the ellipse if and only if

$$\sqrt{x^2 + y^2 + c^2 + 2xc} + \sqrt{x^2 + y^2 + c^2 - 2xc} = 2a.$$

Both sides are non-negative. By first squaring both sides and then dividing by 2, we therefore get the equivalent equation

$$x^2 + y^2 + c^2 + \sqrt{(x^2 + y^2 + c^2)^2 - 4x^2c^2} = 2a^2,$$

which in turn is equivalent to

$$\sqrt{(x^2 + y^2 + c^2)^2 - 4x^2c^2} = 2a^2 - (x^2 + y^2 + c^2). \tag{7.2}$$

Squaring both sides again yields

$$(x^2 + y^2 + c^2)^2 - 4x^2c^2 = 4a^4 + (x^2 + y^2 + c^2)^2 - 4a^2(x^2 + y^2 + c^2),$$

which simplifies to

$$(a^2 - c^2)x^2 + a^2y^2 = a^2(a^2 - c^2). \tag{7.3}$$

If (7.3) holds, then

$$a^2(a^2 - c^2) \geq (a^2 - c^2)(x^2 + y^2),$$

showing that the right-hand side of (7.2) is positive. Hence, (7.2) and (7.3) are equivalent. Dividing both sides of (7.3) by $a^2b^2 = a^2(a^2 - c^2)$, we find that (7.1) is the equation of the ellipse.

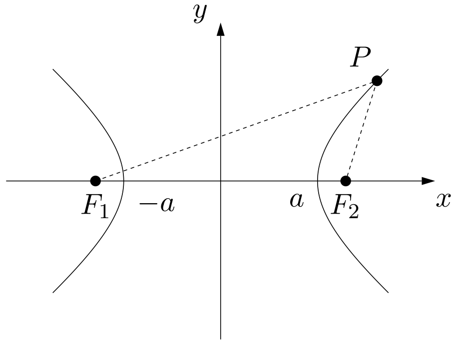
The points F_1 and F_2 are the foci of the ellipse. The line segment from $(0, 0)$ to $(a, 0)$ is called the semi-major axis and the line segment from $(0, 0)$ to $(0, b)$ the semi-minor axis of the ellipse. The line through the foci is called the transverse axis and the line perpendicular to this line and passing through the midpoint between the foci is called the conjugate axis.

7.4.3 Hyperbolae

Two different points F_1 and F_2 on a plane M and a real number a with $0 < 2a < \left\|\overrightarrow{F_1F_2}\right\|$ define a hyperbola on M. It is the set of points P on M for which

$$\left\|\overrightarrow{F_1P}\right\| - \left\|\overrightarrow{F_2P}\right\| = \pm 2a.$$

Consider the same kind of coordinate system as in the previous section. This time we assume that $F_1 = (-c, 0)$ and $F_2 = (c, 0)$ where $0 < a < |c|$.



With $b = \sqrt{c^2 - a^2}$, we shall show that the equation of the hyperbola with respect to this system is

$$\frac{x^2}{a^2} - \frac{y^2}{b^2} = 1. \tag{7.4}$$

$P = (x, y)$ lies on the hyperbola if and only

$$\sqrt{x^2 + y^2 + c^2 + 2xc} - \sqrt{x^2 + y^2 + c^2 - 2xc} = \pm 2a.$$

By first squaring both sides and then dividing by 2, we get the equivalent equation

$$x^2 + y^2 + c^2 - \sqrt{(x^2 + y^2 + c^2)^2 - 4x^2c^2} = 2a^2,$$

which is equivalent to

$$\sqrt{(x^2 + y^2 + c^2)^2 - 4x^2c^2} = x^2 + y^2 + c^2 - 2a^2. \tag{7.5}$$

Again squaring both sides, we get

$$(x^2 + y^2 + c^2)^2 - 4x^2c^2 = (x^2 + y^2 + c^2)^2 + 4a^4 - 4a^2(x^2 + y^2 + c^2)$$

which simplifies to

$$(c^2 - a^2)x^2 - a^2y^2 = a^2(c^2 - a^2). \tag{7.6}$$

If (7.6) holds, then $a^2(c^2 - a^2) \geq (c^2 - a^2)(x^2 + y^2)$, from which it follows that the right-hand side of (7.5) is positive. Thus, (7.5) and (7.6) are equivalent. We divide both sides of (7.6) by $a^2b^2 = a^2(c^2 - a^2)$, and find that (7.4) is the equation of the hyperbola.

The points F_1 and F_2 are the foci of the hyperbola. The line passing through the foci is called the transverse axis of the hyperbola. The conjugate axis is the line that passes through the midpoint between the foci and is perpendicular to the transverse axis.

7.4.4 Parabolae

Let F be a point on a plane M and L a line on M not passing through F. The set of all points P on M whose distance to L is equal to $\left\|\overrightarrow{FP}\right\|$ is a parabola on M with focus F and directrix L.

This time we choose an orthonormal coordinate system for M such that the y-axis is parallel to L and the coordinates of F are $(a, 0)$ where $2a$ is the distance between F and L.

$P = (x, y)$ lies on the parabola if and only if

$$x + a = \sqrt{(x-a)^2 + y^2},$$

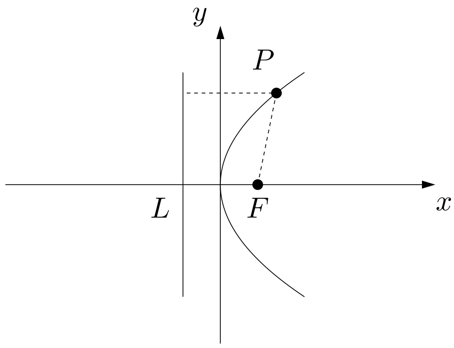
and this implies that

$$x^2 + 2ax + a^2 = (x-a)^2 + y^2 = x^2 - 2ax + a^2 + y^2 \quad \Leftrightarrow \quad y^2 = 4ax.$$

Conversely, if $y^2 = 4ax$, then $x \geq 0$ and hence $x + a \geq 0$. This shows that the implication is in fact an equivalence. The equation of the parabola is therefore

$$y^2 = 4ax$$

where $2a$ is the distance between F and L.



The line perpendicular to the directrix and passing through the focus is the transverse axis or symmetry axis of the parabola. The line parallel to the directrix and passing through the vertex $(0, 0)$ of the parabola is called the conjugate axis.

7.5 Quadratic Equations

7.5.1 Two Variables

The general quadratic equation in two variables is of the form

$$a_{11}x_1^2 + a_{12}x_1x_2 + a_{21}x_2x_1 + a_{22}x_2^2 + b_1x_1 + b_2x_2 = c$$

where the quadratic form $q(x_1, x_2) = a_{11}x_1^2 + a_{12}x_1x_2 + a_{21}x_2x_1 + a_{22}x_2^2$ is not identically zero.

Consider first the equation

$$q(x_1, x_2) = c.$$

By choosing an orthonormal basis $\boldsymbol{e}_1', \boldsymbol{e}_2'$ for $\mathbf{R}^2$ that diagonalises q, we get an equation

$$\lambda_1(x_1')^2 + \lambda_2(x_2')^2 = c.$$

Since q is not identically zero, we must have $\lambda_1 \neq 0$ or $\lambda_2 \neq 0$.

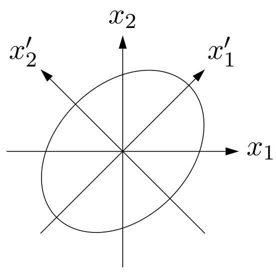
Suppose that $\lambda_1 > 0$ and $\lambda_2 > 0$. If $c < 0$, the solution set is empty. If $c = 0$, the only solution is $(0, 0)$. If $c > 0$, the equation can be written as

$$\frac{(x_1')^2}{a_1^2} + \frac{(x_2')^2}{a_2^2} = 1$$

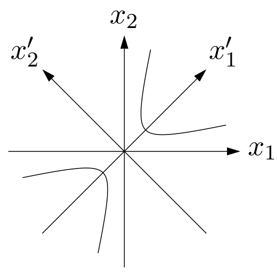
where

$$a_1 = \sqrt{\frac{c}{\lambda_1}} \quad \text{and} \quad a_2 = \sqrt{\frac{c}{\lambda_2}}.$$

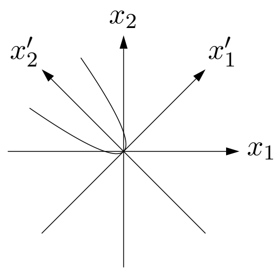
If $\lambda_1 = \lambda_2$, this is the equation of a circle centred at the origin and with radius a_1. Otherwise, it is the equation of an ellipse whose axes are the lines through the origin spanned by the vectors $\boldsymbol{e}_1'$ and $\boldsymbol{e}_2'$, respectively.



Ellipse



Hyperbola



Parabola

Suppose that $\lambda_1 > 0$ and $\lambda_2 < 0$. If $c = 0$, the solution set

$$\sqrt{\lambda_1}\, x_1' = \pm\sqrt{-\lambda_2}\, x_2'$$

consists of two lines intersecting at the origin. If $c > 0$, the equation can be written as

$$\frac{(x_1')^2}{a_1^2} - \frac{(x_2')^2}{a_2^2} = 1$$

where

$$a_1 = \sqrt{\frac{c}{\lambda_1}} \quad \text{and} \quad a_2 = \sqrt{\frac{c}{-\lambda_2}}.$$

Hence, the solution set is a hyperbola. Its transverse and conjugate axes are the lines through the origin spanned by $\boldsymbol{e}_1'$ and $\boldsymbol{e}_2'$, respectively. If $c < 0$, the equation can be written as

$$-\frac{(x_1')^2}{a_1^2} + \frac{(x_2')^2}{a_2^2} = 1.$$

This is also the equation of a hyperbola, but now the transverse and conjugate axes are the lines through the origin spanned by $\boldsymbol{e}_2'$ and $\boldsymbol{e}_1'$, respectively.

All the remaining cases with non-zero eigenvalues can be brought back to one of the previous cases by changing the signs of both sides of the equation or reindexing the eigenvalues and eigenvectors or both.

If one eigenvalue is zero, the solution set is empty or consists of one or two lines depending on the value of c.

Also in the general case

$$q(x_1, x_2) + b_1 x_1 + b_2 x_2 = c,$$

we diagonalise q. With respect to the new basis, the equation becomes

$$\lambda_1 (x_1')^2 + \lambda_2 (x_2')^2 + b_1' x_1' + b_2' x_2' = c'.$$

If both eigenvalues are non-zero, we can complete the two squares and get

$$\lambda_1 \left(x_1' + \frac{b_1'}{2\lambda_1} \right)^2 + \lambda_2 \left(x_2' + \frac{b_2'}{2\lambda_2} \right)^2 = c' + \frac{(b_1')^2}{4\lambda_1} + \frac{(b_2')^2}{4\lambda_2}.$$

By placing the origin at the point

$$\left(-\frac{b_1'}{2\lambda_1}, -\frac{b_2'}{2\lambda_2} \right),$$

we get an equation of the form

$$\lambda_1 (x_1'')^2 + \lambda_2 (x_2'')^2 = c''$$

bringing us back to the cases already discussed.

If $\lambda_1 \neq 0$ and $\lambda_2 = 0$, the equation is

$$\lambda_1 (x_1')^2 + b_1' x_1' + b_2' x_2' = c'.$$

Completing the square, we get

$$\lambda_1 \left(x_1' + \frac{b_1'}{2\lambda_1} \right)^2 + b_2' x_2' = c' + \frac{(b_1')^2}{4\lambda_1}.$$

If $b_2' = 0$, the solution set is empty or consists of one line or two parallel lines. Otherwise, it is a parabola with vertex at

$$\left(-\frac{b_1'}{2\lambda_1}, \frac{4\lambda_1 c' + (b_1')^2}{4\lambda_1 b_2'} \right)$$

and symmetry axis parallel to $\boldsymbol{e}_2'$.

The case where $\lambda_1 = 0$ and $\lambda_2 \neq 0$ can be brought back to the previous case by reindexing the eigenvectors and eigenvalues.

7.5.2 Three Variables

The general quadratic equation in three variables is of the form

$$q(x_1, x_2, x_3) + b_1x_1 + b_2x_2 + b_3x_3 = c$$

where q is a quadratic form on $\mathbf{R}^3$, not identically zero. Also now we begin by studying the equation

$$q(x_1, x_2, x_3) = c.$$

We can write this equation as

$$\lambda_1(x_1')^2 + \lambda_2(x_2')^2 + \lambda_3(x_3')^2 = c$$

where (x_1', x_2', x_3') are the coordinates of $\boldsymbol{x}$ with respect to an orthonormal basis $\boldsymbol{e}_1', \boldsymbol{e}_2', \boldsymbol{e}_3'$ that diagonalises q. At least one eigenvalue is non-zero. In the discussion below, we shall regard circles as ellipses.

Suppose first that the eigenvalues are positive. If $c < 0$, the solution set is empty, and if $c = 0$, the only solution is $(0, 0, 0)$. If $c > 0$, the surface is called an ellipsoid. The intersection between the surface and any of the coordinate planes $x_i' = 0$ is an ellipse. In fact, the intersection between the surface and the plane $x_i' = d$ is an ellipse if $\lambda_i d^2 < c$.

Assume that $\lambda_1 > 0$, $\lambda_2 > 0$ and $\lambda_3 < 0$. If $c > 0$, the surface is a hyperboloid of one sheet. The intersection between the surface and a plane $x_3' = d$ is an ellipse. The intersections between the surface and planes of the form $x_1' = d$ or $x_2' = d$ are hyperbolae. If $c = 0$, the surface is a cone. The intersection between the surface and the plane $x_3' = d$ is an ellipse when $d \neq 0$ and $(0, 0, 0)$ when $d = 0$. The intersection between the surface and one of the coordinate planes $x_1' = 0$ and $x_2' = 0$ consists of two intersecting lines. If $c < 0$, the surface is a hyperboloid of two sheets. The intersection between the surface and the plane $x_3' = d$ is empty when $\lambda_3 d^2 > c$, consists of one point when $\lambda_3 d^2 = c$ and is an ellipse when $\lambda_3 d^2 < c$. The intersection between the surface and one of the coordinate planes $x_1' = 0$ and $x_2' = 0$ is a hyperbola.

Suppose that $\lambda_1 > 0$, $\lambda_2 > 0$ and $\lambda_3 = 0$. If $c < 0$, the solution set is empty. If $c = 0$, the solution set consists of the x_3'-axis. If $c > 0$, the surface is an elliptic cylinder.

Assume that $\lambda_1 > 0$, $\lambda_2 < 0$ and $\lambda_3 = 0$. If $c = 0$, the solution set consists of two intersecting planes. Otherwise, the surface is a hyperbolic cylinder.

If $\lambda_1 \neq 0$ and $\lambda_2 = \lambda_3 = 0$, the solution set is empty, a plane or two parallel planes, depending on the value of c.

After diagonalisation, the general equation becomes

$$\lambda_1(x_1')^2 + \lambda_2(x_2')^2 + \lambda_3(x_3')^2 + b_1'x_1' + b_2'x_2' + b_3'x_3' = c'.$$

Completing squares takes us back to the previous cases when the eigenvalues are non-zero.

When $\lambda_1 \neq 0$, $\lambda_2 \neq 0$ and $\lambda_3 = 0$, we get an equation of the form

$$\lambda_1(x_1'')^2 + \lambda_2(x_2'')^2 + b_3''x_3'' = c''.$$

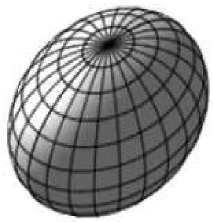
We need only consider the case where $b_3'' \neq 0$. If λ_1 and λ_2 have the same sign, the surface is an elliptic paraboloid, otherwise a hyperbolic paraboloid.

When $\lambda_1 \neq 0$ and $\lambda_2 = \lambda_3 = 0$, the equation becomes

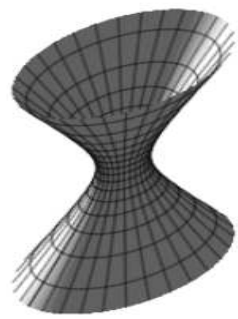
$$\lambda_1(x_1'')^2 + b_2''x_2'' + b_3''x_3'' = c''.$$

If at least one of b_2'' and b_3'' is non-zero, the surface is a parabolic cylinder.

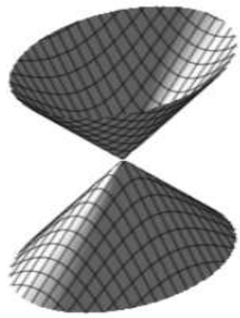
Above we have regarded spheres as ellipsoids. In general, if two or more eigenvalues are equal and the quadratic equation represents a surface, we call that surface a surface of revolution.



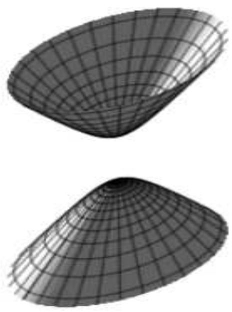
Ellipsoid



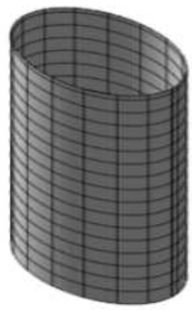
Hyperboloid of one sheet



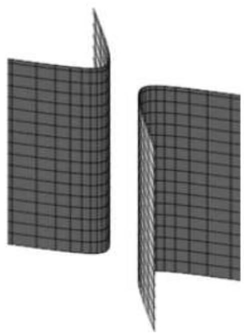
Cone



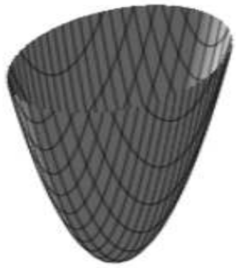
Hyperboloid of two sheets



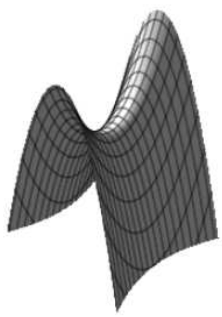
Elliptic cylinder



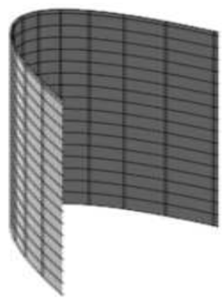
Hyperbolic cylinder



Elliptic paraboloid



Hyperbolic paraboloid



Parabolic cylinder

Example 7.22. We set out to find the type of the surface

$$q(x_1, x_2, x_3) = x_1^2 - 4x_2^2 + x_3^2 + 6x_1x_2 + 4x_1x_3 + 6_2x_3 = 110.$$

We also wish to find the points on the surface closest to the origin and the distance from those points to the origin. The eigenvalues of

$$B = \begin{bmatrix} 1 & 3 & 2 \\ 3 & -4 & 3 \\ 2 & 3 & 1 \end{bmatrix}$$

are $\lambda_1 = -6$, $\lambda_2 = -1$ and $\lambda_3 = 5$. The surface is therefore a hyperboloid of two sheets. Let $\boldsymbol{e}_1', \boldsymbol{e}_2', \boldsymbol{e}_3'$ be an orthonormal basis of eigenvectors associated with $\lambda_1, \lambda_2, \lambda_3$ and (x_1', x_2', x_3') the coordinates of $\boldsymbol{x} = (x_1, x_2, x_3)$ with respect to that basis. Then

$$q(\boldsymbol{x}) = -6(x_1')^2 - (x_2')^2 + 5(x_3')^2.$$

For $\boldsymbol{x}$ on the surface, we have

$$\|\boldsymbol{x}\|^2 = (x_1')^2 + (x_2')^2 + (x_3')^2 \geq \frac{-6(x_1')^2 - (x_2')^2 + 5(x_3')^2}{5} = \frac{q(\boldsymbol{x})}{5} = \frac{110}{5} = 22$$

with equality if and only if $x_1' = x_2' = 0$, $x_3' = \pm\sqrt{22}$. Hence, the minimum distance from a point on the surface to the origin is $\sqrt{22}$ and is attained at $(x_1', x_2', x_3') = \pm(0, 0, \sqrt{22})$. A unit eigenvector associated with λ_3 is $e_3' = (1/\sqrt{22})(3, 2, 3)$. The points on the surface closest to the origin are therefore $(x_1, x_2, x_3) = \pm\sqrt{22}\, e_3' = \pm(3, 2, 3)$.

7.6 Sylvester's Law of Inertia

Definition 7.23. Let q be a quadratic form on a linear space V.

- q is positive definite if $q(\boldsymbol{u}) > 0$ for all non-zero vectors $\boldsymbol{u} \in V$.
- q is positive semidefinite if $q(\boldsymbol{u}) \geq 0$ for all $\boldsymbol{u} \in V$ and $q(\boldsymbol{v}) = 0$ for some non-zero vector $\boldsymbol{v} \in V$.
- q is negative definite if $q(\boldsymbol{u}) < 0$ for all non-zero vectors $\boldsymbol{u} \in V$.
- q is negative semidefinite if $q(\boldsymbol{u}) \leq 0$ for all $\boldsymbol{u} \in V$ and $q(\boldsymbol{v}) = 0$ for some non-zero vector $\boldsymbol{v} \in V$.
- q is indefinite if $q(\boldsymbol{u}) > 0$ and $q(\boldsymbol{v}) < 0$ for some vectors $\boldsymbol{u}$ and $\boldsymbol{v}$ in V.

If a quadratic form q is diagonalised with respect to two orthonormal bases for an inner product space V, then the coefficients in the two representations are the same according to Theorem 7.16. For any two diagonalising bases, this need not be true. Let

$$q(\boldsymbol{u}) = c_1x_1^2 + \cdots + c_nx_n^2$$

be the diagonal representation of q with respect to a basis $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n$ that diagonalises q. Then q is positive definite if and only if $c_i > 0$ for $i = 1, \ldots, n$. Hence, if all coefficients are positive in one diagonal representation of q, then they are positive in every diagonal representation of q. Below, we shall see that this statement can be strengthened.

If q is a quadratic form on a linear space V and U is a subspace of V, then the restriction $q\restriction_U$ of q to U is clearly a quadratic form on U. We say that q is positive or negative definite on U if $q\restriction_U$ is positive or negative definite, respectively.

Definition 7.24. Let q be a quadratic form on a finite-dimensional linear space V. The positive index σ_+ of inertia of q is the maximum dimension of subspaces of V on which q is positive definite. The negative index σ_- of inertia of q is the maximum dimension of subspaces of V on which q is negative definite. The signature of q is the pair (σ_+, σ_-).

Theorem 7.25 (Sylvester's law of inertia). Let q be a quadratic form on a linear space V and suppose that the basis $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n$ for V diagonalises q. If

$$q(x_1\boldsymbol{e}_1 + x_2\boldsymbol{e}_2 + \cdots + x_n\boldsymbol{e}_n) = c_1x_1^2 + c_2x_2^2 + \cdots + c_nx_n^2$$

for all $\boldsymbol{x} \in \mathbf{R}^n$, then the number of positive c_i equals σ_+ and the number of negative c_i equals σ_-.

Proof. After reindexing the basis vectors and coefficients if necessary, we may assume that $c_1, \ldots, c_k$ are positive and $c_{k+1}, \ldots, c_n$ are non-positive. Set $U_+ = [\boldsymbol{e}_1, \ldots, \boldsymbol{e}_k]$ and $U_- = [\boldsymbol{e}_{k+1}, \ldots, \boldsymbol{e}_n]$. We use here the convention that the subspace spanned by no vectors is the zero space $\{\boldsymbol{0}\}$. Then q is positive definite on U_+. Let U be any subspace of V on which q is positive definite. Since $q(\boldsymbol{u}) > 0$ for all non-zero vectors $\boldsymbol{u} \in U$ and $q(\boldsymbol{u}) \le 0$ for all vectors $\boldsymbol{u} \in U_-$, we must have $U \cap U_- = \{\boldsymbol{0}\}$. Hence, by Theorem 2.62,

$$\dim U + n - k = \dim U + \dim U_- = \dim(U + U_-) \le \dim V = n,$$

and therefore $\dim U \le k$. This shows that k is the maximum dimension of subspaces on which q is positive definite. We now obtain the statement about σ_- by applying the statement about σ_+ to the quadratic form $-q$. ■

Example 7.26. We wish to find the type of the surface

$$x_1^2 + 6x_1x_2 - 4x_1x_3 + 7x_2^2 - 4x_2x_3 + 2x_3^2 = 1.$$

This time, as it turns out, the matrix

$$B = \begin{bmatrix} 1 & 3 & -2 \\ 3 & 7 & -2 \\ -2 & -2 & 2 \end{bmatrix}$$

of the quadratic form is not well suited for manual computation of eigenvalues. Instead, we set out to find the representation with respect to some diagonalising basis, not necessarily orthonormal. Then we can use Theorem 7.25 to find the number of positive and negative eigenvalues of the form. We find a diagonal representation by completing squares as follows.

$$\begin{aligned}
q(\boldsymbol{x}) &= x_1^2 + 6x_1x_2 - 4x_1x_3 + 7x_2^2 - 4x_2x_3 + 2x_3^2 \\
&= x_1^2 + (6x_2 - 4x_3)x_1 + 7x_2^2 - 4x_2x_3 + 2x_3^2 \\
&= (x_1 + (3x_2 - 2x_3))^2 - (3x_2 - 2x_3)^2 + 7x_2^2 - 4x_2x_3 + 2x_3^2 \\
&= (x_1 + 3x_2 - 2x_3)^2 - 2x_2^2 + 8x_2x_3 - 2x_3^2 \\
&= (x_1 + 3x_2 - 2x_3)^2 - 2(x_2 - 2x_3)^2 + 8x_3^2 - 2x_3^2 \\
&= (x_1 + 3x_2 - 2x_3)^2 - 2(x_2 - 2x_3)^2 + 6x_3^2.
\end{aligned}$$

Setting

$$\begin{cases} x_1' = x_1 + 3x_2 - 2x_3 \\ x_2' = x_2 - 2x_3 \\ x_3' = x_3 \end{cases},$$

we obtain

$$q(\boldsymbol{x}) = (x_1')^2 - 2(x_2')^2 + 6(x_3')^2.$$

The coefficient matrix

$$\begin{bmatrix} 1 & 3 & -2 \\ 0 & 1 & -2 \\ 0 & 0 & 1 \end{bmatrix}$$

is clearly invertible, whence it is the transition matrix from $\boldsymbol{\varepsilon}_1, \boldsymbol{\varepsilon}_2, \boldsymbol{\varepsilon}_3$ to some basis $\boldsymbol{e}_1, \boldsymbol{e}_2, \boldsymbol{e}_3$. Its inverse T is then the transition matrix from $\boldsymbol{e}_1, \boldsymbol{e}_2, \boldsymbol{e}_3$ to $\boldsymbol{\varepsilon}_1, \boldsymbol{\varepsilon}_2, \boldsymbol{\varepsilon}_3$. Hence, the matrix of q with respect to the basis $\boldsymbol{e}_1, \boldsymbol{e}_2, \boldsymbol{e}_3$ is

$$T^t B T = D = \begin{bmatrix} 1 & 0 & 0 \\ 0 & -2 & 0 \\ 0 & 0 & 6 \end{bmatrix}.$$

The representation of q with respect to the diagonalising basis $\boldsymbol{e}_1, \boldsymbol{e}_2, \boldsymbol{e}_3$ has two positive and one negative coefficients. Hence, q has two positive and one negative eigenvalues. Therefore and since the right-hand side of the equation of the surface is positive, the surface is a hyperboloid of one sheet. The reader should be aware that the coefficients 1, -2 and 6 are not eigenvalues of B and hence not of q.

When carried out correctly, the above method always yields an invertible coefficient matrix and hence a basis that diagonalises q. Sometimes, however, there are no squares to complete as in the following example.

Example 7.27. The quadratic form

$$q(\boldsymbol{x}) = x_1x_2 + x_1x_3 + x_2x_3$$

on $\mathbf{R}^3$ has no squares. As a remedy for this we begin with the following change of coordinates.

$$\begin{cases} x_1 = x_1' + x_2' \\ x_2 = x_1' - x_2' \\ x_3 = \qquad\quad x_3' \end{cases}.$$

Since the coefficient matrix is invertible, this yields a change of basis. We get

$$q(\boldsymbol{x}) = (x_1')^2 - (x_2')^2 + 2x_1'x_3'.$$

Now we can proceed as in the previous example.

$$q(\boldsymbol{x}) = (x_1')^2 - (x_2')^2 + 2x_1'x_3' = (x_1' + x_3')^2 - (x_2')^2 - (x_3')^2 = (x_1'')^2 - (x_2'')^2 - (x_3'')^2.$$

The method used in the above two examples works well for finding the type of a surface but is useless for exploring metric properties of the surface. For example, it cannot be used to find the points on the surface closest to the origin. The reason for this is that the diagonalising basis need not be orthonormal. Nor can it reveal whether the surface is a surface of revolution. Even if two coefficients happen to be equal in the diagonal representation, nothing says that two eigenvalues of the quadratic form must be equal.

The following result is an immediate consequence of Definition 7.24 but can also be regarded as a corollary to Theorem 7.25.

Corollary 7.28. Let q be a quadratic form on an n-dimensional linear space V. Then the following statements hold.

- q is positive definite if and only if $\sigma_+ = n$.
- q is positive semidefinite if and only if $\sigma_+ < n$ and $\sigma_- = 0$.
- q is negative definite if and only if $\sigma_- = n$.
- q is negative semidefinite if and only if $\sigma_+ = 0$ and $\sigma_- < n$.
- q is indefinite if and only if $\sigma_+ > 0$ and $\sigma_- > 0$.

Definition 7.29. The sign function on $\mathbf{R}$ is defined by

$$\operatorname{sgn}(x) = \begin{cases} 1, & \text{if } x > 0, \\ 0, & \text{if } x = 0, \\ -1, & \text{if } x < 0. \end{cases}$$

Let q be a quadratic form on an n-dimensional linear space V with matrix

$$B = \begin{bmatrix} b_{11} & \cdots & b_{1n} \\ \vdots & & \vdots \\ b_{n1} & \cdots & b_{nn} \end{bmatrix}$$

with respect to some basis $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n$ for V and let D be the matrix of q with respect to some basis for V that diagonalises q. Then $D = T^tBT$ for some invertible matrix T, whence

$$\det D = \det(T^tBT) = (\det T)^2 \det B.$$

Since $(\det T)^2 > 0$, it follows that $\operatorname{sgn}(\det D) = \operatorname{sgn}(\det B)$.

Assume that $\det B \neq 0$. Then $\sigma_+ + \sigma_- = n$. If $c_1, \dots, c_n$ are the diagonal entries of D, then σ_+ of the c_i are positive and σ_- of them are negative. Since $\det D = \prod_{i=1}^n c_i$, we get

$$\operatorname{sgn}(\det B) = (-1)^{\sigma_-}.$$

Set $U_0 = \{\mathbf{0}\}$ and $U_m = [\boldsymbol{e}_1, \dots, \boldsymbol{e}_m]$ for $m = 1, \dots, n$. Let, for $m = 0, \dots, n$, q_m be the restriction of q to U_m, and denote by $\sigma_+^{(m)}$ and $\sigma_-^{(m)}$ the indices of inertia of q_m. For $m = 1, \dots, n$, the matrix of q_m with respect to the basis $\boldsymbol{e}_1, \dots, \boldsymbol{e}_m$ is

$$B_m = \begin{bmatrix} b_{11} & \cdots & b_{1m} \\ \vdots & & \vdots \\ b_{m1} & \cdots & b_{mm} \end{bmatrix}.$$

Let $d_0 = 1$ and suppose that $d_m = \det B_m \neq 0$ for $m = 1, \dots, n$. Then

$$\sigma_+^{(m)} + \sigma_-^{(m)} = m \quad \text{and} \quad \operatorname{sgn}(d_m) = (-1)^{\sigma_-^{(m)}}, \qquad m = 0, \dots, n.$$

Suppose that $0 \leq m \leq n-1$. A subspace U of U_m is also a subspace of U_{m+1}. If q_m is positive definite or negative definite on U, then q_{m+1} is positive definite or negative definite, respectively, on U. Hence, $\sigma_+^{(m)} \leq \sigma_+^{(m+1)}$ and $\sigma_-^{(m)} \leq \sigma_-^{(m+1)}$, and therefore

$$\sigma_-^{(m+1)} = \sigma_-^{(m)} \quad \text{or} \quad \sigma_-^{(m+1)} = \sigma_-^{(m)} + 1.$$

In the first case, d_m and d_{m+1} have the same sign, and in the second case, d_m and d_{m+1} have opposite signs. Hence, σ_- equals the number of sign changes in the sequence $d_0, d_1, \dots, d_n$.

Summing up, we have the following theorem.

Theorem 7.30. Let q be a quadratic form on a non-zero n-dimensional linear space V and let B be its matrix with respect to a basis for V. Set $d_0 = 1$ and suppose that $d_m = \det B_m \neq 0$ for $m = 1, \dots, n$. Then σ_- equals the number of sign changes in the sequence $d_0, d_1, \dots, d_n$.

Example 7.31. Consider once again the quadratic form

$$x_1^2 + 6x_1x_2 - 4x_1x_3 + 7x_2^2 - 4x_2x_3 + 2x_3^2$$

in Example 7.26 with matrix

$$B = \begin{bmatrix} 1 & 3 & -2 \\ 3 & 7 & -2 \\ -2 & -2 & 2 \end{bmatrix}.$$

Here

$$d_0 = 1, \quad d_1 = 1, \quad d_2 = \begin{vmatrix} 1 & 3 \\ 3 & 7 \end{vmatrix} = 7 - 9 = -2, \quad d_3 = \begin{vmatrix} 1 & 3 & -2 \\ 3 & 7 & -2 \\ -2 & -2 & 2 \end{vmatrix} = -12.$$

Since all the determinants are non-zero and there is only one change of sign in the sequence $1, 1, -2, -12$, we see that $\sigma_+ = 2$ and $\sigma_- = 1$.

Example 7.32. We can now fulfil the promise made in Section 6.7. Let B_n be the symmetric $n \times n$ matrix

$$\begin{bmatrix} -2 & 1 & 0 & \cdots & 0 \\ 1 & -2 & 1 & \cdots & 0 \\ 0 & 1 & -2 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & -2 \end{bmatrix}.$$

Then $d_1 = \det B_1 = -2$, $d_2 = \det B_2 = 3$ and, for $n \geq 3$,

$$\begin{aligned} d_n = \det B_n = \begin{vmatrix} -2 & 1 & 0 & \cdots & 0 \\ 1 & -2 & 1 & \cdots & 0 \\ 0 & 1 & -2 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & -2 \end{vmatrix} &= -2 \begin{vmatrix} -2 & 1 & \cdots & 0 \\ 1 & -2 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & -2 \end{vmatrix} - \begin{vmatrix} 1 & 0 & \cdots & 0 \\ 1 & -2 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & -2 \end{vmatrix} \\ &= -2d_{n-1} - d_{n-2}. \end{aligned}$$

We can now prove by induction that $d_n = (-1)^n(n+1)$. The statement holds for $n = 1$ and $n = 2$, and if it holds for $k < n$ where $n \geq 3$, then

$$\begin{aligned} d_n &= -2d_{n-1} - d_{n-2} = -2(-1)^{n-1}n - (-1)^{n-2}(n-1) = (-1)^n(2n - (n-1)) \\ &= (-1)^n(n+1). \end{aligned}$$

Hence, there are n sign changes in the sequence $d_0, d_1, \ldots, d_n$, and therefore all the eigenvalues of B_n are negative.

Exercises

7.1. Find, for each of the following quadratic forms q on $\mathbf{R}^3$, an orthonormal basis for $\mathbf{R}^3$ that diagonalises q and find the corresponding diagonal representation.

(a) $q(\boldsymbol{x}) = 6x_1^2 + 3x_2^2 + 3x_3^2 - 4x_1x_2 + 4x_1x_3 - 2x_2x_3$,

(b) $q(\boldsymbol{x}) = x_1^2 + x_2^2 + x_3^2 - 2x_2x_3$.

7.2. Find the maximum and minimum values of

$$q(x_1, x_2, x_3) = 7x_1^2 + 3x_2^2 + 7x_3^2 + 2x_1x_2 + 4x_2x_3$$

subject to the constraint $x_1^2 + x_2^2 + x_3^2 = 1$. Also find the points where they occur.

7.3. Find the maximum and minimum values of

$$q(x_1, x_2, x_3) = x_1^2 + 2x_2^2 + 2x_3^2 + 8x_1x_2 + 8x_1x_3 + 6x_2x_3$$

subject to the constraint $x_1^2 + x_2^2 + x_3^2 \leq 9$.

7.4. (a) Find the minimum value of $r(x_1, x_2, x_3) = x_1^2 + x_2^2 + x_3^2$ subject to the constraint

$$q(x_1, x_2, x_3) = x_1^2 + 3x_2^2 + x_3^2 + 2x_1x_2 - 2x_1x_3 - 2x_2x_3 = 1.$$

(b) Does $r(x_1, x_2, x_3)$ have a maximum value in the set where $q(x_1, x_2, x_3) = 1$?

7.5. Find the least value of a for which

$$3x_1^2 + 5x_2^2 + 3x_3^2 - 2x_1x_2 + 2x_1x_3 - 2x_2x_3 \le a(x_1^2 + x_2^2 + x_3^2)$$

for all $\boldsymbol{x} \in \mathbf{R}^3$.

7.6. Let A be a square matrix. Show that if λ is the least eigenvalue of the symmetric matrix A^tA, then

$$\min_{\|\boldsymbol{x}\|=1} \|A\boldsymbol{x}\| = \sqrt{\lambda}\,.$$

7.7. Find the least possible distance between the points (x_1, x_2, x_3) and $(-x_3, x_1, x_2)$ on the unit sphere.

7.8. A circle passes through the three points whose coordinates are $(0, 2)$, $(6, 10)$ and $(3, 1)$ with respect to an orthonormal coordinate system for two-space. Find an equation of the circle with respect to this coordinate system.

7.9. In this exercise the coordinate system is supposed to be orthonormal.

(a) Find the foci of the ellipse

$$9x^2 + 25y^2 = 225.$$

(b) Find the foci of the hyperbola

$$16x^2 - 9y^2 = 144.$$

7.10. Also here, the coordinate system is supposed to be orthonormal.

(a) Find an equation of the ellipse that passes through the points $(0, \pm 2)$ and whose foci are located at the points $(\pm 2, 0)$.

(b) Find an equation of the hyperbola that passes through the points $(\pm 2, 0)$ and whose foci are located at the points $(\pm 3, 0)$.

(c) Find an equation of the parabola that passes through the points $(0, 0)$ and $(27, 18)$ and whose symmetry axis is the x-axis. Also find the location of the focus.

7.11. Show that the curve described by the equation

$$18x_1^2 + 12x_2^2 - 8x_1x_2 = 40$$

with respect to an orthonormal coordinate system for two-space is an ellipse. Find the lengths and directions of the semi-major and semi-minor axes.

7.12. A quadratic surface has, with respect to an orthonormal coordinate system for three-space, the equation

$$3x_1^2 + 3x_2^2 - 8x_1x_2 + 4x_1x_3 - 4x_2x_3 = 1.$$

Identify the type of surface and find the least distance from a point on the surface to the origin.

7.13. A quadratic surface has, with respect to an orthonormal coordinate system for three-space, the equation

$$11x_1^2 + 11x_2^2 + 14x_3^2 - 2x_1x_2 - 8x_1x_3 - 8x_2x_3 = 1.$$

Show that it is an ellipsoid and find the points on the surface closest to the origin and furthest from the origin.

7.14. A quadratic surface has, with respect to an orthonormal coordinate system for three-space, the equation

$$2x_1^2 - x_2^2 - x_3^2 + 4x_1x_2 - 4x_1x_3 + 8x_2x_3 = 1.$$

Show that it is a surface of revolution, identify its type and find the axis of revolution.

7.15. A quadratic surface has, with respect to an orthonormal coordinate system for three-space, the equation

$$x_1^2 + 6x_2^2 + x_3^2 + 6x_1x_2 + 4x_1x_3 + 6x_2x_3 = 81.$$

Identify the type of surface and determine the distance from the surface to the origin.

7.16. A quadratic surface has, with respect to an orthonormal coordinate system for three-space, the equation

$$2x_1^2 + 3x_2^2 + 3x_3^2 - 2x_1x_2 - 2x_1x_3 - 4x_2x_3 = 25.$$

Identify the type of surface and find the least distance from a point on the surface to the origin.

7.17. Find the signatures of the following quadratic forms on $\mathbf{R}^3$.

(a) $x_1^2 + 5x_2^2 + 11x_3^2 - 4x_1x_2 + 6x_1x_3 - 10x_2x_3$,

(b) $x_1^2 + 3x_2^2 + x_3^2 - 4x_1x_2 + 2x_1x_3 - 6x_2x_3$,

(c) $2x_1x_2 - 3x_1x_3 - x_2x_3$.

7.18. Which of the following quadratic forms on $\mathbf{R}^3$ are positive definite?

(a) $(2x_1 + x_2 + x_3)^2 + (x_1 + 2x_2 + 2x_3)^2 + (x_1 - x_2 - x_3)^2$,

(b) $(x_1 + 2x_2 + x_3)^2 + (x_1 + x_2 + x_3)^2 + (x_1 - x_2 + 2x_3)^2$.

7.19. Identify the types of the following surfaces.

(a) $x_1^2 + 3x_2^2 + 2x_3^2 - 2x_1x_2 - 2x_1x_3 + 6x_2x_3 = 1$,

(b) $x_1^2 + x_2^2 + x_3^2 + 2x_1x_2 - 4x_1x_3 - 4x_2x_3 = 1$.

7.20. Identify, for each value of the real constant a, the type of the surface

$$x_1^2 + (2a+1)x_2^2 + ax_3^2 + 2ax_2x_3 = 1.$$

7.21. Determine whether there is a change of coordinates that takes the quadratic form

$$q(x_1, x_2, x_3) = x_1^2 + 2x_1x_2 - 4x_1x_3 - 2x_2x_3$$

to the quadratic form r in the following two cases.

(a) $r(y_1, y_2, y_3) = y_1^2 + 3y_2^2 + y_3^2 - 4y_1y_2 - 4y_1y_3 + 6y_2y_3$,

(b) $r(y_1, y_2, y_3) = y_1^2 + 5y_2^2 + 3y_3^2 - 4y_1y_2 - 4y_1y_3 + 10y_2y_3$.

7.22. Let a, b and c be non-zero real numbers. Prove that the equation

$$ax_1x_2 + bx_1x_3 + cx_2x_3 = 1$$

is the equation of a hyperboloid. State conditions on a, b and c in order that the hyperboloid be of one sheet and two sheets, respectively.

7.23. Find the signature of the quadratic form

$$q(x_1, x_2, x_3) = 4x_1^2 + 10x_2^2 + 3x_3^2 + 4x_1x_2 + 8x_1x_3 + 10x_2x_3$$

by means of Theorem 7.30.

7.24. Show that the matrix

$$B = \begin{bmatrix} 3 & 2 & 1 \\ 2 & 4 & 1 \\ 1 & 1 & 1 \end{bmatrix}$$

has exactly one eigenvalue in the open interval $(1, 2)$ by studying the signatures of the quadratic forms with matrices $B - I$ and $B - 2I$.

Answers to Exercises

1.1. (a) $\begin{bmatrix} 2 & 2 & 4 \\ 0 & 3 & 3 \\ 3 & 2 & 2 \end{bmatrix}$, (b) not defined, (c) $\begin{bmatrix} 9 & 7 & 12 \\ 4 & 4 & 6 \\ 0 & 1 & -1 \end{bmatrix}$,

(d) $\begin{bmatrix} 2 & 3 & 2 \\ 0 & 5 & 6 \\ 4 & 8 & 5 \end{bmatrix}$, (e) not defined, (f) $\begin{bmatrix} 1 & 7 & 9 \\ 3 & 6 & 2 \end{bmatrix}$,

(g) $\begin{bmatrix} 7 & 12 & 16 \\ 11 & 11 & 13 \end{bmatrix}$, (h) $\begin{bmatrix} 13 & 17 & 23 \\ 19 & 16 & 24 \end{bmatrix}$.

1.2. $A^2 - B^2 = \begin{bmatrix} -2 & 4 \\ -5 & -5 \end{bmatrix}$, $(A+B)(A-B) = \begin{bmatrix} -4 & 2 \\ -6 & -3 \end{bmatrix}$.

1.3. $B = t \begin{bmatrix} 2 & -4 \\ -1 & 2 \end{bmatrix}$, $t \in \mathbf{R}$.

1.4. (b) $\begin{bmatrix} 1 & 12 & 138 \\ 0 & 1 & 24 \\ 0 & 0 & 1 \end{bmatrix}$.

1.5. $A^tB^t = \begin{bmatrix} 2 & 0 & 4 \\ 3 & 5 & 8 \\ 2 & 6 & 5 \end{bmatrix}$, $(A^t + B^t)C^t = \begin{bmatrix} 7 & 11 \\ 12 & 11 \\ 16 & 13 \end{bmatrix}$.

1.8. (a) $\begin{bmatrix} -1 & 1 & 1 \\ 1 & -2 & 1 \\ 0 & 1 & -1 \end{bmatrix}$, (b) not invertible, (c) $\dfrac{1}{5}\begin{bmatrix} -10 & 5 & 5 \\ 0 & 2 & -1 \\ 5 & -3 & -1 \end{bmatrix}$.

1.9. $A^{-1} = \begin{bmatrix} 0 & -1 & 1 \\ 2 & 2 & -3 \\ -1 & 0 & 1 \end{bmatrix}$, $\left(A^t\right)^{-1} = \begin{bmatrix} 0 & 2 & -1 \\ -1 & 2 & 0 \\ 1 & -3 & 1 \end{bmatrix}$, $\left(A^2\right)^{-1} = \begin{bmatrix} -3 & -2 & 4 \\ 7 & 2 & -7 \\ -1 & 1 & 0 \end{bmatrix}$.

1.10. $\dfrac{1}{6(a-3)} \begin{bmatrix} 3a-8 & 4 & -3 \\ 4-a & 2a-2 & -3 \\ -2 & -8 & 6 \end{bmatrix}$, $a \neq 3$.

1.11. $X = \begin{bmatrix} -2 & 6 & -5 \\ 1 & -4 & 4 \end{bmatrix}$.

2.1. The sets in (a) and (d) are subspaces, the sets in (b) and (c) are not.

2.3. E.g. $\operatorname{im}\begin{bmatrix}1 & 2\\ 1 & 0\\ 0 & 1\end{bmatrix} = \ker\begin{bmatrix}1 & -1 & -2\end{bmatrix}$.

2.4. Only the set in (b) is linearly dependent.

2.5. No. Yes.

2.7. (a) $\ker A$: E.g. $(-2,-1,1,0)$, $(-1,-2,0,1)$. $\operatorname{im} A$: E.g. $(1,1,2)$, $(1,-1,1)$.
(b) $\ker A$: E.g. $(-1,-1,1,0)$. $\operatorname{im} A$: E.g. $(3,1,5)$, $(4,2,7)$, $(1,-1,2)$.

2.8. E.g. $(1,1,0,1)$, $(1,2,2,1)$, $(3,4,1,3)$.

2.9. $(1,-2,2,-1)$.

2.10. (a) 2, (b) 3.

2.12. $n-1$.

2.14. Only the sum in (b) is direct.

2.15. The projection on U along V is $(1,2,-1)$. The projection on V along U is $(3,3,6)$.

2.16. Only the function in (b) is a linear transformation.

2.17. $\dim\ker A = 2$, $\dim\operatorname{im} A = 3$.

2.18. F is one-to-one but not onto.

2.19. $F^{-1}(x_1,x_2,x_3) = (3x_1-4x_2+2x_3, -5x_1+7x_2-3x_3, 4x_1-5x_2+2x_3)$.

2.20. F is not one-to-one but onto. The kernel is the set of constant polynomials. The image is P.

3.1. $\frac{\pi}{3}$.

3.5. E.g. $\frac{1}{\sqrt{7}}(2,1,1,1)$, $\frac{1}{\sqrt{7}}(-1,1,2,-1)$, $\frac{1}{\sqrt{14}}(-2,1,0,3)$.

3.6. E.g. $\frac{1}{\sqrt{5}}(-2,1,0,0)$, $\frac{1}{\sqrt{45}}(2,4,5,0)$, $\frac{1}{\sqrt{90}}(1,2,-2,9)$.

3.7. $(a,b,c) = \pm(1,2,-2)$.

3.9. (b) E.g. 1, $\sqrt{3}\,(2x-1)$, $\sqrt{5}\,(6x^2-6x+1)$.

3.10. E.g. $\frac{1}{\sqrt{6}}(1,-2,1,0)$, $\frac{1}{\sqrt{12}}(1,1,1,-3)$.

3.11. (a) $(2,3,2,3)$, 2.
(b) $(2,2,2,4)$, $\sqrt{2}$.

3.12. E.g. $\frac{1}{\sqrt{3}}(1,1,1,0)$, $\frac{1}{3}(0,2,-2,1)$, $\frac{1}{\sqrt{6}}(1,-1,0,2)$, $\frac{1}{\sqrt{18}}(3,-1,-2,-2)$.

3.14. (b) $2\sin t - \sin 2t$.

3.15. $\frac{1}{2}(-2,-4,-3,3,2,4)$.

3.16. (a) 2.

(b) E.g. $(1,2,3,4)$, $(1,-1,2,3)$ and $(1,1,1)$, $(2,-1,5)$, respectively.

3.17. $\begin{bmatrix}\frac{1}{3} & -\frac{23}{42} & \frac{1}{2}\end{bmatrix}^t$.

3.18. $y = -\frac{8}{5}t + \frac{22}{5}$.

3.19. $y = \frac{1}{2}t^2 + \frac{9}{10}t + \frac{23}{10}$.

3.20. $(2,2,2,4)$.

4.1. (a) 24, (b) 1, (c) 0.

4.2. (a) 6, (b) -160, (c) $b(b-a)^3$.

4.3. (a) $x = 0$, $x = 2\sqrt{3}$ or $x = -2\sqrt{3}$. (b) $x = 1$ or $x = -3$.

4.6. (a) $(n-1)!$, (b) 1 if n is odd, 0 if n is even,

(c) 1, (d) 0 if n is odd, $(-2)^{n/2}$ if n is even.

4.7. $(-1)^{n-1}(n-1)2^{n-2}$.

4.13. $\dim\ker A = 1$, $\dim\operatorname{im} A = 2$ if $a = -2$, $\dim\ker A = 2$, $\dim\operatorname{im} A = 1$ if $a = 2$, $\dim\ker A = 0$, $\dim\operatorname{im} A = 3$ if $a \neq -2$ and $a \neq 2$.

4.14. $-\frac{1}{50}$.

4.15. $x \neq -4$, $x \neq 0$ and $x \neq 2$.

4.18. (a) $\dfrac{1}{2}\begin{bmatrix}-4 & 2\\ 3 & -1\end{bmatrix}$, (b) $\dfrac{1}{11}\begin{bmatrix}1 & -2 & 5\\ -2 & 4 & 1\\ 6 & -1 & -3\end{bmatrix}$.

5.1. $\dfrac{1}{25}\begin{bmatrix}16 & 12\\ 12 & 9\end{bmatrix}$.

5.2. $\dfrac{1}{9}\begin{bmatrix}8 & -2 & 2\\ -2 & 5 & 4\\ 2 & 4 & 5\end{bmatrix}$.

5.3. $\begin{bmatrix}0 & 1\\ 0 & 1\end{bmatrix}$.

5.4. (a) $\dfrac{1}{6}\begin{bmatrix}5 & -1 & -1\\ -2 & 4 & -2\\ -3 & -3 & 3\end{bmatrix}$, (b) $\dfrac{1}{6}\begin{bmatrix}1 & 1 & 1\\ 2 & 2 & 2\\ 3 & 3 & 3\end{bmatrix}$.

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

5.6. $\begin{bmatrix} 2 & -7 & -12 \\ -4 & 13 & 23 \\ 3 & -6 & -12 \end{bmatrix}$.

5.7. (a) No. (b) Yes.

5.8. $U' : x_1 - 2x_2 - x_3 = 0$, $U'' : \boldsymbol{x} = t(2, -1, 3)$.

5.9. $U' : \boldsymbol{x} = s(0, 3, 3, -2) + t(1, -1, -2, 1)$, $U'' : \boldsymbol{x} = s(1, 0, 0, 1) + t(0, 1, 1, 0)$.

5.10. $U' : x_1 + x_2 + 2x_3 = 0$, $U'' : \boldsymbol{x} = t(1, 3, -1)$.

5.11. $\dfrac{1}{9}\begin{bmatrix} 7 & -4 & 4 \\ -4 & 1 & 8 \\ 4 & 8 & 1 \end{bmatrix}$.

5.12. $\begin{bmatrix} -3 & 2 & 6 \\ 8 & -3 & -12 \\ -4 & 2 & 7 \end{bmatrix}$.

5.14. $\begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$.

5.15. $\dfrac{1}{2}\begin{bmatrix} \sqrt{3} & 0 & 1 \\ 0 & 2 & 0 \\ -1 & 0 & \sqrt{3} \end{bmatrix}$.

5.16. $\dfrac{1}{4}\begin{bmatrix} 2+\sqrt{3} & 2-\sqrt{3} & \sqrt{2} \\ 2-\sqrt{3} & 2+\sqrt{3} & -\sqrt{2} \\ -\sqrt{2} & \sqrt{2} & 2\sqrt{3} \end{bmatrix}$.

5.18. (a) Rotation about the origin through the angle $\frac{\pi}{6}$ in the direction from $\boldsymbol{e}_2$ towards $\boldsymbol{e}_1$.

(b) Orthogonal reflection in the line $x_2 = \sqrt{3}\,x_1$.

(c) Rotation about the origin through the angle $\frac{\pi}{4}$ in the direction from $\boldsymbol{e}_1$ towards $\boldsymbol{e}_2$.

5.19. (a) Rotation about the line $\boldsymbol{x} = t(5, -1, 1)$ through the angle $\frac{2\pi}{3}$ in the anticlockwise direction when looking from the point $(5, -1, 1)$ towards the origin.

(b) Orthogonal reflection in the plane $2x_1 - x_2 - 2x_3 = 0$.

(c) Rotation about the line $\boldsymbol{x} = t(1, 1, 1)$ through the angle $\frac{2\pi}{3}$ in the clockwise direction when looking from the point $(1, 1, 1)$ towards the origin.

(d) Rotation about the line $\boldsymbol{x} = t(1, 2, 3)$ through the angle π.

(e) Rotation about the line $\boldsymbol{x} = t(1, 1, 1)$ through the angle $\frac{\pi}{3}$ in the anticlockwise direction when looking from the point $(1, 1, 1)$ towards the origin followed by reflection in the origin.

6.1. (a) 1: $t(1,2)$, $t \neq 0$, 2: $t(2,3)$, $t \neq 0$.

(b) No eigenvalues.

(c) -3: $t(9,-11,4)$, $t \neq 0$, 1: $t(1,1,0)$, $t \neq 0$, 2: $t(1,1,1)$, $t \neq 0$.

(d) 2: $s(0,0,1) + t(1,2,0)$, $s \neq 0$ or $t \neq 0$.

6.3. (a) 0 and 1. (b) -1 and 1.

6.6. (a) Diagonalisable, e.g.

$$T = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 2 \\ 2 & 1 & -1 \end{bmatrix}, \quad T^{-1}AT = \begin{bmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 2 \end{bmatrix}.$$

(b) Diagonalisable, e.g.

$$T = \begin{bmatrix} 0 & 1 & -1 \\ -1 & 2 & -2 \\ 1 & 0 & 1 \end{bmatrix}, \quad T^{-1}AT = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 2 \end{bmatrix}.$$

(c) Not diagonalisable.

6.7. $\begin{bmatrix} 12 & -7 & -4 \\ 4 & -1 & -2 \\ 16 & -11 & -4 \end{bmatrix}$.

6.8. $\dfrac{1}{5}\begin{bmatrix} 2 \cdot (-1)^n + 3 \cdot 4^n & -2 \cdot (-1)^n + 2 \cdot 4^n \\ -3 \cdot (-1)^n + 3 \cdot 4^n & 3 \cdot (-1)^n + 2 \cdot 4^n \end{bmatrix}$.

6.11. E.g.

$$\begin{bmatrix} 0 & 0 & 0 & \cdots & 0 \\ 0 & \lambda & 0 & \cdots & 0 \\ 0 & 0 & \lambda & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & \lambda \end{bmatrix}$$

where $\lambda = -\langle \boldsymbol{b}, \boldsymbol{c} \rangle$.

6.12. $\begin{bmatrix} a_n \\ b_n \end{bmatrix} = (-2)^n \begin{bmatrix} 2 \\ -4 \end{bmatrix} + 7^n \begin{bmatrix} 2 \\ 5 \end{bmatrix}$.

6.13. $\begin{bmatrix} a_n \\ b_n \\ c_n \end{bmatrix} = (-3)^n \begin{bmatrix} 3 \\ -9 \\ 3 \end{bmatrix} + (-1)^n \begin{bmatrix} -7 \\ 6 \\ 1 \end{bmatrix} + 7^n \begin{bmatrix} 4 \\ 8 \\ 4 \end{bmatrix}$.

6.14. $a_n = -(-2)^n + 1 + 2 \cdot 2^n$.

6.15. (a) E.g.

$$T = \frac{1}{5\sqrt{2}} \begin{bmatrix} 3 & -4\sqrt{2} & 3 \\ 4 & 3\sqrt{2} & 4 \\ -5 & 0 & 5 \end{bmatrix}, \quad T^t A T = \begin{bmatrix} -4 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 6 \end{bmatrix}.$$

(b) E.g.

$$T = \frac{1}{3} \begin{bmatrix} 1 & 2 & -2 \\ 2 & 1 & 2 \\ -2 & 2 & 1 \end{bmatrix}, \quad T^t A T = \begin{bmatrix} 9 & 0 & 0 \\ 0 & 9 & 0 \\ 0 & 0 & 18 \end{bmatrix}.$$

6.17. $\begin{bmatrix} x_1(t) \\ x_2(t) \\ x_3(t) \end{bmatrix} = 3e^{-6t} \begin{bmatrix} 1 \\ -3 \\ 1 \end{bmatrix} + e^{-t} \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix} + 2e^{5t} \begin{bmatrix} 3 \\ 2 \\ 3 \end{bmatrix}.$

6.18. $\begin{bmatrix} x_1(t) \\ x_2(t) \\ x_3(t) \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 \\ -1 & 1 & 1 \\ 0 & -2 & 3 \end{bmatrix} \begin{bmatrix} a_1 \cos t + b_1 \sin t \\ a_2 e^t + b_2 e^{-t} \\ a_3 e^{\sqrt{6}\,t} + b_3 e^{-\sqrt{6}\,t} \end{bmatrix}.$

6.19. Eigenfrequencies

$$\frac{q\sqrt{2-\sqrt{2}}}{2\pi}, \quad \frac{q\sqrt{2}}{2\pi}, \quad \frac{q\sqrt{2+\sqrt{2}}}{2\pi}$$

with associated eigenvectors

$$\begin{bmatrix} 1 \\ \sqrt{2} \\ 1 \end{bmatrix}, \quad \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}, \quad \begin{bmatrix} 1 \\ -\sqrt{2} \\ 1 \end{bmatrix},$$

respectively.

7.1. (a) E.g. $\frac{1}{\sqrt{2}}(0,1,1)$, $\frac{1}{\sqrt{3}}(1,1,-1)$, $\frac{1}{\sqrt{6}}(2,-1,1)$, $\quad 2(x_1')^2 + 2(x_2')^2 + 8(x_3')^2$.

(b) E.g. $\frac{1}{\sqrt{2}}(0,1,1)$, $(1,0,0)$, $\frac{1}{\sqrt{2}}(0,-1,1)$, $\quad (x_2')^2 + 2(x_3')^2$.

7.2. Minimum value 2 at $\pm\frac{1}{\sqrt{30}}(1,-5,2)$, maximum value 8 at $\pm\frac{1}{\sqrt{6}}(1,1,2)$.

7.3. Minimum value -27, maximum value 81.

7.4. (a) $\frac{1}{4}$. (b) No.

7.5. 6.

7.7. 1.

7.8. $(x-3)^2 + (y-6)^2 = 25$.

7.9. (a) $(\pm 4, 0)$, (b) $(\pm 5, 0)$.

7.10. (a) $x^2 + 2y^2 = 8$.

(b) $5x^2 - 4y^2 = 20$.

(c) $y^2 = 12x$. The focus is located at $(3,0)$.

7.11. The semi-major axis has length 2 and direction $(1, 2)$. The semi-minor axis has length $\sqrt{2}$ and direction $(-2, 1)$.

7.12. Hyperboloid of two sheets. $\frac{1}{\sqrt{8}}$.

7.13. The points closest to the origin are $\pm\frac{1}{6\sqrt{3}}(1, 1, -2)$. The points furthest from the origin are $\pm\frac{1}{3\sqrt{2}}(1, 1, 1)$.

7.14. Hyperboloid of one sheet. $(1, -2, 2)$.

7.15. Hyperbolic cylinder. 3.

7.16. Elliptic cylinder. $\sqrt{5}$.

7.17. (a) $(3, 0)$, (b) $(2, 1)$, (c) $(1, 2)$.

7.18. Only the quadratic form in (b) is positive definite.

7.19. (a) Hyperboloid of one sheet.
(b) Hyperbolic cylinder.

7.20. Ellipsoid if $a > 0$, elliptic cylinder if $a = 0$, hyperboloid of one sheet if $-1 < a < 0$, hyperbolic cylinder if $a = -1$ and hyperboloid of two sheets if $a < -1$.

7.21. (a) Yes. (b) No.

7.22. Two sheets if $abc > 0$, one sheet if $abc < 0$.

7.23. $(2, 1)$.

Index