Preface

This course has grown out of lectures going back well over 30 years. Names which come to mind are J.Nevin, R.Watson, G.Parish and J. Raftery. I have been involved with teaching linear algebra to engineers since the early 90s and some of my ideas have been incorporated in the course. What is new in these notes is mainly my approach to the theoretical side of the subject. Several of the numerical examples and exercises come from earlier notes, as well as a some theoretical ones.

I would like to thank Dr. Paddy Ewer for showing me how to use the program LatexCad for the diagrams and especially the Reverend George Parish for proof-reading and also for elucidating No.5 of Exercise 2.

About this course

The course is meant to be studied as a whole. Many examples and exercises in later Chapters refer the reader to earlier chapters. This is especially true of Chapter 3.

Chapter 1 motivates the idea of a vector through geometry and discusses lines and planes and transformations related to such geometric objects. A vector can be thought of as a displacement in space and an ordered triple of numbers.

Chapter 2 generalizes the idea of a triple to an \( n \)-tuple and motivates linear algebra through the problem of finding solutions to simultaneous linear equations in \( n \) unknowns. The coefficient matrix of such equations is known as a matrix. Simplification of such a matrix by row operations forms the major part of this chapter.

Chapter 3 considers matrices in detail and looks at them dynamically in the converse sense of Chapter 2: A matrix defines a transformation of points in \( n \)-dimensional space. Matrix multiplication is introduced in terms of the composition of such transformations and some other key concepts such as linear independence, the rank and inverse of a matrix are discussed. Abstract vector spaces are never mentioned, but the the proof of the basic theorem in 3.4.22 on linear independence goes through word-for-word for such spaces which also leads to the well-known Replacement Theorem (see No.14 of Exercise 93).

Chapter 4 is about determinants and the cross product (also called the vector product). The theory of determinants predates that of matrices, going back to Leibnitz in the 17th Century. One of the founders of linear algebra, the 19th Century mathematician Arthur Cayley, once remarked that many things about determinants should really come after the study of matrices, and this is the modern approach adopted by us. The cross product is used extensively in mechanics, in particular in the notes Dynamics for Mathematics 142. Algebraic properties of the cross product are derived from those of \( 3 \times 3 \) determinants, while the exercises can serve as an introduction to some of its applications.

Note 1 Some exercises (particularly those marked with an asterisk \(*\)) are harder and, at the discretion of the instructor, can be omitted or postponed to a later stage.
Bibliography

The following books can be consulted but they should be used with caution since different authors have a variety of starting points and use different notation. Many books also have a tendency to become too abstract too early. Unless you are a mature reader, books can confuse rather than help you.

1. A.E. Hirst: Vectors in 2 or 3 Dimensions (Arnold).
   This may be useful for our Chapter 1 since it makes 3-dimensional space its central theme.

   This is a quite elementary text.


4. K. Hardy: Linear Algebra for Engineers and Scientists (Prentice-Hall).

5. D. Lay: Linear Algebra and its Applications (3rd Edition, Pearson). This book treats matrix multiplication in much the same way as we do, but its treatment of geometric aspects is less thorough.
   It has over 560 pages, becomes abstract and advanced after p.230, but will probably be useful in later years of study.


I recommend especially (1) for Chapter 1 and (4) and (5) for later chapters of the notes. Because of its elementary nature, (2) is good for Chapter 2.
Some of the above books have been used on the Durban campus and, if consulted with care, can be helpful.

C. Zaverdinos
Pietermaritzburg, January 21, 2010
## Contents

1. Two and Three-Dimensional Analytic Geometry. Vectors 7
2. Matrices and the Solution of Simultaneous Linear Equations 41
3. Linear Transformations and Matrices 65
4. Determinants and the Cross Product 123
Chapter 1

Two and Three-Dimensional Analytic Geometry. Vectors

1.1 Points in three-dimensional space

The study of the geometry of lines and planes in space provides a good introduction to Linear Algebra. Geometry is a visual subject, but it also has an analytical aspect which is the study of geometry using algebra: how geometric problems can be expressed and solved algebraically. This geometry is also called Cartesian Geometry, after its founder, the 17th Century philosopher and mathematician René Descartes.

From school you are already familiar with the Cartesian plane. There are an $x$–axis and a $y$–axis meeting at right-angles at the origin $O$. Every point $A$ in the $x − y$ plane is uniquely represented by an ordered pair $(a, b)$ of real numbers, where $a$ is the $x$–coordinate and $b$ is the $y$–coordinate as in Figure 1.1. We write $A = (a, b)$. Here $M$ is the foot of the perpendicular from $A$ to the $x$–axis and likewise $N$ is the foot of the perpendicular from $A$ to the $y$–axis. Notice that $M = (a, 0)$ and $N = (0, b)$.

\[ \text{Figure 1.1} \]

In order to represent a point $A$ in space we add the $z$–axis which is perpendicular to both the
1.1.1 The Corkscrew Rule and right-handed systems

You wish to open a bottle of wine using a corkscrew. You first turn the corkscrew so that it enters the cork. As observed by the opener the turning is clockwise, but observed from the underside of the bottle the sense is reversed: it is anticlockwise. The direction in which a corkscrew moves as it is turned is given by the Corkscrew Rule. In Figure 1.2, rotating from the $x$–axis to the $y$–axis and applying this rule produces motion along the (positive) $z$–axis. Rotating from the $y$–axis to the $z$–axis and applying the same rule gives motion along the (positive) $x$–axis. Finally, rotating from the $z$–axis to the $x$–axis and applying the rule produces motion along the (positive) $y$–axis. The axes $x, y, z$ (in that order) are said to form a right-hand system. This can be represented by the scheme

$$x \rightarrow z \rightarrow y \rightarrow x$$

Have a look at yourself in the mirror while opening a bottle of wine and observe that what is clockwise for you is anticlockwise for the person in the mirror. Thus a right-hand system of axes observed in a mirror does not follow the corkscrew rule and is called a left-hand system. The usual convention is to use only right-hand systems.

1.1.2 Displacements and vectors

Consider two numbers $a$ and $b$ on the real line $\mathbb{R}$. The difference $c = b - a$ is the displacement from $a$ to $b$. Given any two numbers from $a, b$ and $c$, the third is uniquely determined by the
1.1. POINTS IN THREE-DIMENSIONAL SPACE

The equation $c = b - a$. For example, let $a = -2$ and $b = 3$. The displacement from $a$ to $b$ is $c = 3 - (-2) = 5$. The displacement from 6 to 11 is also 5.

Let $A = (a_1, a_2, a_3)$, and $B = (b_1, b_2, b_3)$ be two points in space. The displacement from $A$ to $B$ is denoted by $\vec{AB}$ and is given by

$$\vec{AB} = (b_1 - a_1, b_2 - a_2, b_3 - a_3)$$  \hspace{1cm} (1.1)

1.1.3 Units and distance

The unit can be any agreed upon distance, like the meter or kilometer. Units (of length, time) are important in applications but for the most part we won’t specify them.

The distance of $A = (a, b, c)$ from the origin $O$ is given by

$$|\vec{OA}| = \sqrt{a^2 + b^2 + c^2}$$

To see this, see Figure 1.2 and use the theorem of Pythagoras:

$$|\vec{OA}|^2 = |\vec{OR}|^2 + |\vec{RA}|^2$$

$$= |\vec{OL}|^2 + |\vec{LR}|^2 + |\vec{RA}|^2$$

$$= a^2 + b^2 + c^2$$

The distance from $A$ to $B$ is written as $|\vec{AB}|$ and is given by

$$|\vec{AB}| = \sqrt{(b_1 - a_1)^2 + (b_2 - a_2)^2 + (b_3 - a_3)^2}$$  \hspace{1cm} (1.2)

and is also called the length or magnitude of $\vec{AB}$.

Because the displacement $\vec{AB}$ has both direction and magnitude it is called a vector and pictured as an arrow going from $A$ to $B$. The tail of the arrow is at $A$ and its head is at $B$ as in Figure 1.3.

As an example, let $A = (-1, 2, 3)$ and $B = (2, 1, 3, 6, -2, 5)$. Then $\vec{AB} = (2.1 - (-1), 3.6 - 2.3, -2.5 - (-4.5)) = (3.1, 1.3, 2.0)$. m.

Two displacement vectors $\vec{AB}$ and $\vec{CD}$ are equal if they are equal as displacements. Hence if $C = (-3.5, 1.1, 3.3)$ and $D = (-4, 2.4, 5.3)$ and $A$, $B$ are as above then $\vec{CD} = (-4, 2.4, 5.3) - (-3.5, 1.1, 3.3) = (3.1, 1.3, 2.0) = \vec{AB}$.

The magnitude of this vector is $|\vec{AB}| = |\vec{CD}| = \sqrt{(3.1)^2 + (1.3)^2 + (2.0)^2} = 3.9115$ m.

Various vectors (displacement, velocity, acceleration, force) play an important role in mechanics and have their own units. For example, if positions are measured in meters and time is measured in seconds, a velocity vector will have units ms$^{-1}$. So if a particle moves in such a way that each second it displaces $(-1, 2, 3)$ meters, we say its velocity is constantly $(-1, 2, 3)$ ms$^{-1}$.

Depending on the geometric interpretation, the tail (head) of displacement vectors may be at any point. Vectors are often written as $\vec{u} = (u_1, u_2, u_3)$, $\vec{v} = (v_1, v_2, v_3)$, $\vec{a} = (a_1, a_2, a_3)$, $\vec{b} = (b_1, b_2, b_3)$ etc.

We now have three ways of expressing the position of a point $A$ in space. If $A$ has coordinates $\vec{a} = (a_1, a_2, a_3)$, then

$$A = (a_1, a_2, a_3) = \vec{a} = \vec{OA}$$  \hspace{1cm} (1.3)

Although equation (1.3) identifies a point with its position, geometrically speaking we like to distinguish the point itself from its position. The vector $\vec{OA}$ is called the position vector.
1.1.4 More systematic notational convention

Instead of speaking of the $x$-, $y$- and $z$-axis, we also refer to these as the $x_1$-, $x_2$- and $x_3$-axis respectively. Accordingly, given vectors $\mathbf{a}, \mathbf{b}, \mathbf{c}, \ldots$ it will be convenient to assume (unless otherwise stated) that $\mathbf{a} = (a_1, a_2, a_3), \mathbf{b} = (b_1, b_2, b_3), \mathbf{c} = (c_1, c_2, c_3), \ldots$. Also by convention, points $A, P, Q, \ldots$ have position vectors $\mathbf{a}, \mathbf{p}, \mathbf{q}, \ldots$, unless otherwise specified.

1.1.5 Addition of vectors

Let $\mathbf{u} = (u_1, u_2, u_3)$ and $\mathbf{v} = (v_1, v_2, v_3)$ be two vectors. Their sum $\mathbf{u} + \mathbf{v}$ is defined by the equation

$$\mathbf{u} + \mathbf{v} = (u_1 + v_1, u_2 + v_2, u_3 + v_3)$$

The vector $-\mathbf{v}$ is defined as $-\mathbf{v} = (-v_1, -v_2, -v_3)$ and the difference

$$\mathbf{u} - \mathbf{v} = \mathbf{u} + (-\mathbf{v}) = (u_1 - v_1, u_2 - v_2, u_3 - v_3)$$

The zero vector is $\mathbf{0} = (0, 0, 0)$.

1.1.6 Basic properties of addition of vectors

For all vectors $\mathbf{u}, \mathbf{v}$ and $\mathbf{w}$
1.1. POINTS IN THREE-DIMENSIONAL SPACE

1. \( \mathbf{u} + (\mathbf{v} + \mathbf{w}) = (\mathbf{u} + \mathbf{v}) + \mathbf{w} \) (associative law)

2. \( \mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u} \) (commutative law)

3. \( \mathbf{0} + \mathbf{u} = \mathbf{u} \) (the vector \( \mathbf{0} = (0, 0, 0) \) behaves like the number 0)

4. \( \mathbf{u} + (-\mathbf{u}) = \mathbf{0} \)

An important result for addition is the following:

1.1.7 Geometrical interpretation: The triangle law for addition of vectors

\[
\mathbf{AB} + \mathbf{BC} = \mathbf{AC}
\]

To see this, let the position vectors of \( A, B \) and \( C \) be \( \mathbf{a}, \mathbf{b} \) and \( \mathbf{c} \) respectively. Then \( \mathbf{AB} = \mathbf{b} - \mathbf{a} \) and \( \mathbf{BC} = \mathbf{c} - \mathbf{b} \). Hence by the properties of addition,

\[
\mathbf{AB} + \mathbf{BC} = (\mathbf{b} - \mathbf{a}) + (\mathbf{c} - \mathbf{b}) = (\mathbf{b} - \mathbf{b}) + (\mathbf{c} - \mathbf{a}) = \mathbf{0} + \mathbf{c} - \mathbf{a} = \mathbf{c} - \mathbf{a} = \mathbf{AC}
\]

Figure 1.4 illustrates this geometrical result in which it is understood that the head \( B \) of \( \mathbf{AB} \) is the tail of \( \mathbf{BC} \) etc. The point \( D \) is chosen so that \( \mathbf{AD} = \mathbf{BC} \). It follows that \( \mathbf{DC} = \mathbf{AB} \) (why?). We have a geometric interpretation of the above commutative law (known as the parallelogram law for addition of vectors):

\[
\mathbf{BC} + \mathbf{AB} = \mathbf{AD} + \mathbf{DC} = \mathbf{AC} = \mathbf{AB} + \mathbf{BC}
\]

![Image of geometric interpretation](image)

Figure 1.4 \( \mathbf{AC} = \mathbf{AB} + \mathbf{BC} \)

1.1.8 Multiplication of vectors by scalars. Direction of vectors. Parallel vectors

In contrast to a vector, a **scalar** is just a real number, i.e. an element of the real number system \( \mathbb{R} \). We use \( \alpha, \beta, \gamma, \delta, \rho, \sigma, \tau, \ldots \) to denote scalars. Note that \( \underline{f} \) (\( f \) underlined) denotes a vector while \( f \) is a scalar.
CHAPTER 1. TWO AND THREE-DIMENSIONAL ANALYTIC GEOMETRY. VECTORS

Let \( \mathbf{u} = (u_1, u_2, u_3) \) be a vector and \( \alpha \in \mathbb{R} \) be a scalar. The product \( \alpha \mathbf{u} \) is defined by

\[ \alpha \mathbf{u} = (\alpha u_1, \alpha u_2, \alpha u_3) \]

The vector \( \alpha \mathbf{u} \) is said to be a (scalar) multiple of \( \mathbf{u} \). We usually omit the qualification ‘scalar’.

1.1.9 Properties of the product \( \alpha \mathbf{u} \)

For all scalars \( \alpha, \beta \) and vectors \( \mathbf{u} \) and \( \mathbf{v} \),

1. \( \alpha (\mathbf{u} + \mathbf{v}) = \alpha \mathbf{u} + \alpha \mathbf{v} \)
2. \( (\alpha + \beta) \mathbf{u} = \alpha \mathbf{u} + \beta \mathbf{u} \)
3. \( \alpha (\beta \mathbf{u}) = (\alpha \beta) \mathbf{u} \)
4. \( 1 \mathbf{u} = \mathbf{u} \)
5. \( \alpha \mathbf{u} = 0 \) if, and only if, \( \alpha = 0 \) or \( \mathbf{u} = 0 \)
6. \( |\alpha \mathbf{u}| = |\alpha| |\mathbf{u}| \) (the length of \( \alpha \mathbf{u} \) is \( |\alpha| \) times the length of \( \mathbf{u} \))

The proofs of the first four are left to you. We prove the last:

\[
|\alpha \mathbf{u}| = |(\alpha u_1, \alpha u_2, \alpha u_3)| \\
= \sqrt{(\alpha u_1)^2 + (\alpha u_2)^2 + (\alpha u_3)^2} \\
= \sqrt{\alpha^2 u_1^2 + \alpha^2 u_2^2 + \alpha^2 u_3^2} \\
= \sqrt{\alpha^2(u_1^2 + u_2^2 + u_3^2)} \\
= |\alpha| \sqrt{u_1^2 + u_2^2 + u_3^2} \\
= |\alpha| |\mathbf{u}|
\]

1.1.10 Geometric interpretation of \( \alpha \mathbf{u} \). Parallel vectors

Since \( |\alpha \mathbf{u}| = |\alpha| |\mathbf{u}| \), multiplication of \( \mathbf{u} \) by the scalar \( \alpha \) multiplies the length of \( \mathbf{u} \) by \( |\alpha| \).

Let \( \mathbf{u} \) be non-zero, so that it has a definite direction. If \( \alpha > 0 \) the direction of \( \alpha \mathbf{u} \) is the same as that of \( \mathbf{u} \), while if \( \alpha < 0 \) the direction of \( \alpha \mathbf{u} \) is opposite to that of \( \mathbf{u} \).

Let \( \mathbf{u} \) and \( \mathbf{v} \) be non-zero vectors. We say they have the same direction if \( \mathbf{u} = \alpha \mathbf{v} \) for some scalar \( \alpha > 0 \) and opposite directions if \( \mathbf{u} = \alpha \mathbf{v} \) for some \( \alpha < 0 \). The vector \( \mathbf{u} \) is parallel to \( \mathbf{v} \) if \( \mathbf{v} \) is a multiple of \( \mathbf{u} \), that is, \( \mathbf{u} = \alpha \mathbf{v} \) for some scalar \( \alpha \). Necessarily \( \alpha \neq 0 \) and we write \( \mathbf{u} \parallel \mathbf{v} \) to express this relation.

Notice that we only speak of vectors being parallel if they are non-zero.

1.1.11 Properties of parallelism

For all non-zero \( \mathbf{u}, \mathbf{v} \) and \( \mathbf{w} \),

1. \( \mathbf{u} \parallel \mathbf{u} \)
2. \( \mathbf{u} \parallel \mathbf{v} \) implies \( \mathbf{v} \parallel \mathbf{u} \)
3. \( \mathbf{u} \parallel \mathbf{v} \) and \( \mathbf{v} \parallel \mathbf{w} \) together imply \( \mathbf{u} \parallel \mathbf{w} \).
These properties are geometrically evident, while analytic proofs are left as an exercise. For example, to see property (2) analytically, let \( \mathbf{u} = \alpha \mathbf{v} \). As \( \alpha \neq 0 \), \( \mathbf{v} = \frac{1}{\alpha} \mathbf{u} \) and \( \mathbf{v} \parallel \mathbf{u} \), as expected.

**Example 1.1.1** Let \( \mathbf{u} = \mathbf{AB} \), where \( A \) is the point \((-1, 7, -4)\) and \( B \) is the point \((-4, 11, 1)\), so that \( \mathbf{u} = (-3, 4, 5) \). Let \( \mathbf{v} = \mathbf{CD} \), where \( C = (2, 9, -11) \) and \( D = (-7, 21, 4) \), so \( \mathbf{CD} = (-9, 12, 15) = 3 \mathbf{u} \). Hence \( \mathbf{u} \) and \( \mathbf{v} \) have the same direction while \( \mathbf{u} \) and \(-\mathbf{v}\) have opposite directions but are parallel.

**1.1.12 The dot (scalar) product of two vectors**

Let \( \mathbf{u} = (u_1, u_2, u_3) \) and \( \mathbf{v} = (v_1, v_2, v_3) \) be any two vectors. Their **dot product** \( \mathbf{u} \cdot \mathbf{v} \) is defined as

\[
\mathbf{u} \cdot \mathbf{v} = u_1 v_1 + u_2 v_2 + u_3 v_3
\]

(1.4)

Note that the dot product of two vectors is a scalar and not a vector (that is why it is also called the **scalar product**).

**1.1.13 Properties of the dot product**

For all scalars \( \alpha, \beta \) and vectors \( \mathbf{u}, \mathbf{v} \) and \( \mathbf{w} \),

1. \( \mathbf{u} \cdot \mathbf{v} = \mathbf{v} \cdot \mathbf{u} \)
2. \( \alpha (\mathbf{u} \cdot \mathbf{v}) = (\alpha \mathbf{u}) \cdot \mathbf{v} = \mathbf{u} \cdot (\alpha \mathbf{v}) \)
3. \( (\alpha \mathbf{u} + \beta \mathbf{v}) \cdot \mathbf{w} = (\alpha \mathbf{u} \cdot \mathbf{w}) + (\beta \mathbf{v} \cdot \mathbf{w}) \)
4. \( \mathbf{u} \cdot \mathbf{u} = |\mathbf{u}|^2 \), so \( |\mathbf{u}| = +\sqrt{\mathbf{u} \cdot \mathbf{u}} \)
5. \( \mathbf{u} \cdot \mathbf{u} \geq 0 \) and \( \mathbf{u} \cdot \mathbf{u} = 0 \) if, and only if \( \mathbf{u} = \mathbf{0} \).

**1.1.14 Geometric interpretation of the dot product**

Let \( \mathbf{u} = \mathbf{AB} \) and \( \mathbf{v} = \mathbf{AC} \) be non-zero vectors. Suppose that \( \mathbf{AB} \) and \( \mathbf{AC} \) make an angle \( \theta \) between them at the vertex \( A \), where \( 0 \leq \theta \leq 180 \) in degrees, or \( 0 \leq \theta \leq \pi \) in radians (recall that 180 degrees = \( \pi \) radians). Then

\[
\mathbf{u} \cdot \mathbf{v} = |\mathbf{u}| |\mathbf{v}| \cos \theta
\]

(1.5)

To see this result, see Figure 1.5 and use \( \mathbf{AB} + \mathbf{BC} = \mathbf{AC} \), so that \( \mathbf{BC} = \mathbf{AC} - \mathbf{AB} = \mathbf{v} - \mathbf{u} \) and

\[
|\mathbf{BC}|^2 = (\mathbf{v} - \mathbf{u}) \cdot (\mathbf{v} - \mathbf{u})
= \mathbf{v} \cdot \mathbf{v} + (-\mathbf{u}) \cdot (-\mathbf{u}) - \mathbf{v} \cdot \mathbf{u} - \mathbf{u} \cdot \mathbf{v}
= |\mathbf{v}|^2 + |\mathbf{u}|^2 - 2 \mathbf{u} \cdot \mathbf{v}
\]

Then by the cosine rule (which applies whether \( \theta \) is acute or obtuse),

\[
|\mathbf{BC}|^2 = |\mathbf{AB}|^2 + |\mathbf{AC}|^2 - 2 |\mathbf{AB}| |\mathbf{AC}| \cos \theta
|\mathbf{v}|^2 + |\mathbf{u}|^2 - 2 \mathbf{u} \cdot \mathbf{v} = |\mathbf{u}|^2 + |\mathbf{v}|^2 - 2 |\mathbf{u}| |\mathbf{v}| \cos \theta \quad \text{(from the previous equation)}
\]

Cancelling \( |\mathbf{v}|^2 + |\mathbf{u}|^2 \) from both sides leads to the required result (1.5).

**Example 1.1.2** Let \( A = (1, 0, 1), \ B = (2, 0, 3), \ C = (4, \sqrt{10}, 2) \) and \( D = (2, 2, -2) \). Find the angle \( \theta \) between \( \mathbf{AB} \) and \( \mathbf{AC} \). Decide if the angle \( \phi \) between \( \mathbf{AB} \) and \( \mathbf{AD} \) is acute or obtuse.
CHAPTER 1. TWO AND THREE-DIMENSIONAL ANALYTIC GEOMETRY. VECTORS

Solution: \( \overrightarrow{AB} = (2, 0, 3) - (1, 0, 1) = (1, 0, 2) \) and \( \overrightarrow{AC} = (4, \sqrt{10}, 2) - (1, 0, 1) = (3, \sqrt{10}, 1) \). Hence

\[
\cos \theta = \frac{\overrightarrow{AB} \cdot \overrightarrow{AC}}{|\overrightarrow{AB}| |\overrightarrow{AC}|} = \frac{(1, 0, 2) \cdot (3, \sqrt{10}, 1)}{\sqrt{1^2 + 0^2 + 2^2} \sqrt{3^2 + (\sqrt{10})^2 + 1^2}} = \frac{1}{2}
\]

and \( \theta = 60 \) degrees or \( \frac{\pi}{3} \) radians.

Similarly, \( \overrightarrow{AD} = (1, 2, -3) \) and

\[
\cos \phi = \frac{\overrightarrow{AB} \cdot \overrightarrow{AD}}{|\overrightarrow{AB}| |\overrightarrow{AD}|} = \frac{(1, 0, 2) \cdot (1, 2, -3)}{\sqrt{1^2 + 0^2 + 2^2} \sqrt{1^2 + 2^2 + (-3)^2}} = -\sqrt{\frac{5}{14}}
\]

Since the dot product is negative the angle \( \phi \) is obtuse.

1.1.15 Some further properties of the dot product

1. The non-zero vectors \( \overrightarrow{AB} \) and \( \overrightarrow{AC} \) are at right angles (are perpendicular \( \overrightarrow{AB} \perp \overrightarrow{AC} \)) if, and only if, \( (\overrightarrow{AB}) \cdot (\overrightarrow{AC}) = 0 \).

Such vectors are also called mutually orthogonal.

2. \( |\overrightarrow{u} \cdot \overrightarrow{v}| \leq |\overrightarrow{u}| |\overrightarrow{v}| \) for any vectors \( \overrightarrow{u} \) and \( \overrightarrow{v} \) (Cauchy-Schwartz inequality).

3. \( |\overrightarrow{u} + \overrightarrow{v}| \leq |\overrightarrow{u}| + |\overrightarrow{v}| \) for any vectors \( \overrightarrow{u} \) and \( \overrightarrow{v} \) (Cauchy’s inequality).

The first result follows from the fact that \( \overrightarrow{AB} \perp \overrightarrow{AC} \) if, and only if, \( \cos \theta = 0 \).

The second result is left as an exercise.

To see the third result, use (2) and consider

\[
|\overrightarrow{u} + \overrightarrow{v}|^2 = (\overrightarrow{u} + \overrightarrow{v}) \cdot (\overrightarrow{u} + \overrightarrow{v}) = \overrightarrow{u} \cdot \overrightarrow{u} + \overrightarrow{v} \cdot \overrightarrow{v} + 2 \overrightarrow{u} \cdot \overrightarrow{v} \leq |\overrightarrow{u}|^2 + |\overrightarrow{v}|^2 + 2 |\overrightarrow{u}| |\overrightarrow{v}| = (|\overrightarrow{u}| + |\overrightarrow{v}|)^2
\]

As \( |\overrightarrow{u} + \overrightarrow{v}|^2 \leq (|\overrightarrow{u}| + |\overrightarrow{v}|)^2 \), it follows that \( |\overrightarrow{u} + \overrightarrow{v}| \leq |\overrightarrow{u}| + |\overrightarrow{v}| \).
Example 1.1.3 Find a non-zero vector perpendicular to both \( \mathbf{u} = (2, 1, -1) \) and \( \mathbf{v} = (3, 3, -1) \).

Solution:
Let \( \mathbf{x} = (x_1, x_2, x_3) \) satisfy \( \mathbf{x} \perp \mathbf{u} \) and \( \mathbf{x} \perp \mathbf{v} \). Then

\[
\mathbf{x} \cdot \mathbf{u} = 2x_1 + x_2 + (-1)x_3 = 0
\]
\[
\mathbf{x} \cdot \mathbf{v} = 3x_1 + 3x_2 + (-1)x_3 = 0
\]

By subtraction \(-x_1 - 2x_2 = 0\), or \(x_1 = -2x_2\). Thus from the first equation, \(x_3 = 2x_1 + x_2 = -3x_2\). We may let \(x_2\) be any non-zero number, e.g. \(x_2 = 1\). Then

\( \mathbf{x} = (-2, 1, -3) \)

is perpendicular to \( \mathbf{u} \) and to \( \mathbf{v} \).

We will see later why it is always possible to find a vector \( \mathbf{x} \neq 0 \) perpendicular to two non-zero vectors. This fact will also come into our study of planes (see 1.3.4).

1.1.16 Unit vectors. Components. Projections

A vector \( \mathbf{u} \) of length 1 is called a \textbf{unit vector}. For example, \( \mathbf{u} = \left( \frac{3}{5}, 0, -\frac{4}{5} \right) \) is a unit vector.

A vector \( \mathbf{c} \neq 0 \) defines the vector \( \hat{\mathbf{c}} = \frac{1}{|\mathbf{c}|} \mathbf{c} \) which is the unique \textbf{unit vector having the same direction as} \( \mathbf{c} \). To see that \( \hat{\mathbf{c}} \) has length 1 we note

\[
|\hat{\mathbf{c}}|^2 = \hat{\mathbf{c}} \cdot \hat{\mathbf{c}} = \left( \frac{1}{|\mathbf{c}|} \mathbf{c} \right) \cdot \left( \frac{1}{|\mathbf{c}|} \mathbf{c} \right) = \frac{1}{|\mathbf{c}|^2} \mathbf{c} \cdot \mathbf{c} = \frac{|\mathbf{c}|^2}{|\mathbf{c}|^2} = 1
\]

The \textbf{component} of the vector \( \mathbf{f} \) in the direction of (or along) the non-zero vector \( \mathbf{c} \) is defined as \( \mathbf{f} \cdot \hat{\mathbf{c}} \).

Suppose that \( \mathbf{f} \neq 0 \) and that the angle between \( \mathbf{f} \) and \( \mathbf{c} \) is \( \theta \). Then

\[
\mathbf{f} \cdot \hat{\mathbf{c}} = |\mathbf{f}| \cos \theta \tag{1.6}
\]

To see equation (1.6), use equation (1.5):

\[\text{Figure 1.6}\]
\[ f \cdot \hat{c} = f \cdot \frac{1}{|c|}c = \frac{f \cdot c}{|c|} = \frac{|f||c|\cos \theta}{|c|} = |f| \cos \theta \]

(See Figure 1.6). Notice that if \( \frac{\pi}{2} < \theta \leq \pi \) then the component (1.6) is negative (as is only natural).

The projection of \( f \) on \( c \) is the vector
\[ (f \cdot \hat{c}) \hat{c} = \frac{f \cdot c}{|c|^2} c \]

(1.7)

Figure 1.7 shows two cases of the projection of \( f \neq 0 \) on \( c \): when \( \cos \theta > 0 \) and when \( \cos \theta < 0 \).

Let \( d \parallel c \). Then the projection of \( f \) on \( c \) is the same as the projection of \( f \) on \( d \). See Exercise 2, No.13a.

The situation with components is slightly different: If \( d \) and \( c \) have the same direction then the components of \( f \) along \( c \) and \( d \) are the same. If \( d \) and \( c \) have opposite directions then the components of \( f \) along \( c \) and \( d \) are the same apart from sign - one is the negative of the other.

We return to projections in subsection 1.2.5.

**Example 1.1.4**

1. If we imagine that \( c \) is pointing due East and \( f \) is a displacement, then \((f \cdot \hat{c}) \hat{c}\) is the Easterly displacement, while \(|f| \cos \theta \) is the Easterly component and \(|f| \sin \theta \) is the Northerly component \( f \).

You may think of \( f \) as a vector representing a force that is dragging a box placed at its tail. Then \(|f| \sin \theta \) is the lifting effect of \( f \) while \(|f| \cos \theta \) is the horizontal dragging effect of the force.

2. Let \( f = (11, -3, 4) \) and \( c = (5, -1, -2) \). The length of \( c \) is \(|c| = \sqrt{5^2 + (-1)^2 + (-2)^2} = \sqrt{30} \).

The unit vector \( \hat{c} \) in the direction of \( c \) is
\[ \hat{c} = \frac{1}{\sqrt{30}} (5, -1, -2) \]

The component of \( f \) along \( c \) is
\[ f \cdot \hat{c} = (11, -3, 4) \cdot \frac{1}{\sqrt{30}} (5, -1, -2) = \frac{5}{3} \sqrt{30} \]
1.1. POINTS IN THREE-DIMENSIONAL SPACE

and the projection of \( \overline{f} \) on \( \overline{c} \) is

\[
(f \cdot \overline{c})\overline{c} = \frac{5}{3} \sqrt{30} \left( \frac{1}{\sqrt{30}} (5, -1, -2) \right) = \frac{5}{3} (5, -1, -2)
\]

1.1.17 The vectors \( \hat{i}, \hat{j}, \hat{k} \)

The vectors \( \hat{i}, \hat{j} \) and \( \hat{k} \) are defined as \( \hat{i} = (1, 0, 0) \), \( \hat{j} = (0, 1, 0) \), and \( \hat{k} = (0, 0, 1) \). They have unit length and are directed along the \( x-, y- \) and \( z- \)-axis respectively. It is clear that they are also mutually orthogonal and that any vector \( \overrightarrow{a} = (a_1, a_2, a_3) \) has a unique expression

\[
\overrightarrow{a} = a_1\hat{i} + a_2\hat{j} + a_3\hat{k}
\]  \hfill (1.8)

The scalars \( a_1 \), \( a_2 \) and \( a_3 \) are the components of \( \overrightarrow{a} \) in the directions \( \hat{i}, \hat{j} \) and \( \hat{k} \) respectively. See also No.15 of Exercise 2 below.

1.1.18 Linear combinations

Let \( \overrightarrow{a} \) and \( \overrightarrow{b} \) be two vectors and \( \lambda \) and \( \mu \) two scalars. The vector \( \overrightarrow{z} = \lambda \overrightarrow{a} + \mu \overrightarrow{b} \) is called a linear combination of \( \overrightarrow{a} \) and \( \overrightarrow{b} \) with coefficients \( \lambda \) and \( \mu \). Similarly, for scalars \( r, s \) and \( t \), the vector \( r\overrightarrow{a} + s\overrightarrow{b} + t\overrightarrow{c} \) is a linear combination of \( \overrightarrow{a}, \overrightarrow{b}, \) and \( \overrightarrow{c} \) with coefficients \( r, s \) and \( t \).

Equation (1.8) shows that every vector is a linear combination of \( \hat{i}, \hat{j} \) and \( \hat{k} \). If \( \overrightarrow{a} = (-1, 2, 3), \overrightarrow{b} = (5, -7, 1), \overrightarrow{c} = (-4, 8, -1) \) then

\[
\frac{1}{2}\overrightarrow{a} + \frac{2}{3}\overrightarrow{b} + \left(-\frac{5}{6}\right)\overrightarrow{c} = \left(\frac{37}{6}, \frac{-31}{3}, 3\right)
\]

is a linear combination of \( \overrightarrow{a}, \overrightarrow{b}, \) and \( \overrightarrow{c} \).

Exercise 2

1. In Figure 1.2 express \( ON, MP, OMQ, LP, NR, NM \) and \( OP \) as vectors involving \( \overrightarrow{a}, \overrightarrow{b} \) and \( \overrightarrow{c} \). Write down these vectors as linear combinations of \( \hat{i}, \hat{j} \) and \( \hat{k} \) and find their lengths in terms of \( \overrightarrow{a}, \overrightarrow{b} \) and \( \overrightarrow{c} \).

2. Interpret the equation \((\overrightarrow{r} - (2, -1, 5)) \cdot (\overrightarrow{r} - (2, -1, 5)) = 49 \) geometrically.

3. Let \( \overrightarrow{u} = (1, 3, -2), \overrightarrow{v} = (2, 0, 1) \) and \( \overrightarrow{w} = (3, -3, 4) \).

(a) Find the linear combinations \( \overrightarrow{u} + 3\overrightarrow{w} - 2\overrightarrow{v}, (5)\overrightarrow{u} + 2\overrightarrow{w} + 3\overrightarrow{v} \) and \( \frac{1}{3}\overrightarrow{u} - \frac{1}{2}\overrightarrow{w} + \frac{2}{3}\overrightarrow{v} \).

(b) If the tail of \( \overrightarrow{u} \) is the point \( A = (-2, -11, 13) \), what is the head of \( \overrightarrow{u} \)? If the head of \( \overrightarrow{w} \) is \( B = (2, -9, 10) \), what is the tail of \( \overrightarrow{w} \)?

(c) Find coefficients \( \alpha, \beta \) such that \( \overrightarrow{w} = \alpha\overrightarrow{u} + \beta\overrightarrow{v} \).

(d) Is \( 3(3, -1, 3) \) a linear combination of \( \overrightarrow{u}, \overrightarrow{v} \) and \( \overrightarrow{w} \)? This means you must look for scalars \( p, q, r \) such that \( p\overrightarrow{u} + q\overrightarrow{v} + r\overrightarrow{w} = 3(3, -1, 3) \), and this involves three equations in three unknowns.

4. (Generalization of No.3b). Let \( \overrightarrow{u} \) be a given vector. Show that if the tail \( A \) of \( \overrightarrow{u} \) is given then the head \( B \) of \( \overrightarrow{u} \) is uniquely determined. State and prove a similar result if the head of \( \overrightarrow{u} \) is given. Draw a picture.

5. See Figure 1.8. Imagine that in each case the axes are part of a room in which you are standing and looking at the corner indicated. Decide which of the systems of mutually orthogonal axes are left-handed and which right-handed. If you are outside the room, what would your answer be?

6. Let \( A_0, \ldots, A_9 \) be ten points in space and put \( \overrightarrow{a}_m = A_{m-1}A_m \) for \( m = 1, \ldots, 9 \) and \( \overrightarrow{a}_{10} = A_0A_9 \).

Find the sum \( \overrightarrow{a}_1 + \cdots + \overrightarrow{a}_{10} \).
7. Complete the proofs in statements 1.1.6, 1.1.9 and 1.1.11 above.

**Hint 3** These properties just reflect similar properties of numbers. For example, the commutative law of addition of vectors is proved by

\[ u + v = (u_1 + v_1, u_2 + v_2, u_3 + v_3) \]
\[ = (v_1 + u_1, v_2 + u_2, v_3 + u_3) \]
\[ = v + u \]

The proof of the first item of 1.1.9 is

\[ \alpha(u + v) = \alpha(u_1 + v_1, u_2 + v_2, u_3 + v_3) \]
\[ = (\alpha(u_1 + v_1), \alpha(u_2 + v_2), \alpha(u_3 + v_3)) \]
\[ = (\alpha u_1 + \alpha v_1, \alpha u_2 + \alpha v_2, \alpha u_3 + \alpha v_3) \]
\[ = \alpha u + \alpha v \]

8. Complete the proofs of subsections 1.1.13 and 1.1.15.

9. Deduce that a dot product \((\alpha r + \beta s) \cdot (\gamma u + \delta v)\) of linear combinations multiplies out just like in ordinary arithmetic and conclude that \(|\alpha u + \beta v|^2 = |\alpha|^2 |u|^2 + |\beta|^2 |v|^2 + 2\alpha\beta u \cdot v|.

10. Show that if the vectors \(u\) and \(v\) in equation (1.5) are unit vectors, then \(\cos \theta = u \cdot v\).  

11. Show that the non-zero vectors \(u\) and \(v\) are parallel if, and only if, \(|u \cdot v| = |u| |v|\).

**Solution:** The vectors are parallel if, and only if, the angle between them is 0 or \(\pi\) radians.

12. Let \(a = (3, -1, 2)\) and \(b = (-7, 3, -1)\). Find

(a) The unit vectors \(\hat{a}\) and \(\hat{b}\).

(b) The component of \(a\) along \(b\) and the component of \(b\) along \(a\).

(c) The projection of \(a\) on \(b\) and the projection of \(b\) on \(a\).

13. (a) If \(\gamma \neq 0\), show that the projections of \(f\) on \(c\) and \(f\) on \(\gamma c\) are equal.

(b) Show that if \(f \neq 0\), then the projection of \(f\) on itself is \(f\).
1.2. THE STRAIGHT LINE

14. Let \( u \) and \( v \) be two vectors. The set of all linear combinations of \( u \) and \( v \) is called the span of the vectors and is denoted by \( \text{sp} (u, v) \). (This is discussed more fully in item 2 of subsection 3.1.2 of Chapter 3).

Let \( u_1, u_2, v_1, v_2 \) be four vectors. Show that 
\[
\text{sp} (u_1, v_1) \subseteq \text{sp} (u_2, v_2)
\]
if, and only if that each of the vectors \( u_1 \) and \( v_1 \) are linear combinations of the vectors \( u_2 \) and \( v_2 \).

15. Let \( l, m, n \) be three mutually orthogonal unit vectors. Placing the tails of these vectors at the origin, it is visually clear that we may use these vectors as a frame of reference in place of \( i, j, k \). (An analytic proof of this will follow from work done in Chapter 3 - See Exercise 78, No.9). Assuming this result, any vector \( a \) can be expressed as a linear combination 
\[
a = \gamma_1 l + \gamma_2 m + \gamma_3 n
\]
Show that necessarily the coefficients \( \gamma_1, \gamma_2, \gamma_3 \) are the components (in the above sense) of \( a \) in the directions \( l, m, n \) respectively. In fact, \( \gamma_1 l, \gamma_2 m, \gamma_3 n \) are the projections of the vector \( a \) on \( l, m, n \) respectively. Conclude that 
\[
|a|^2 = \gamma_1^2 + \gamma_2^2 + \gamma_3^2
\]
Show that \( l = \frac{1}{\sqrt{6}} (1, -1, 2), m = \frac{1}{\sqrt{3}} (-1, 1, 1) \) and \( n = \frac{1}{\sqrt{2}} (1, 1, 0) \) is such a system of three mutually orthogonal unit vectors and find the coefficients \( \gamma_1, \ldots, \gamma_3 \), if \( a = (-3, 2, 1) \).

Hint 4 Take the dot product of both sides of the above equation with each of \( l, m, n \). The coefficients for the specific example are the components \( a \cdot l = \gamma_1 \) etc. A longer procedure to find the coefficients for the specific example is to use the method of 3d above.

16. What familiar result from Euclidean Geometry does \( |b - a| \leq |a| + |b| \) represent?

17. Let \( u \) and \( v \) have the same length: \( |u| = |v| \). Show that \( (u - v) \cdot (u + v) = 0 \). In other words, if \( u - v \) and \( u + v \) are not zero they are at right angles.

18. How are our general results affected (and simplified) when one of the coordinates is restricted to be zero for all vectors under consideration?

1.2 The straight line

Two distinct points determine a line. Points that lie on one and the same line they are called collinear.

Let \( A, B \) and \( C \) be three distinct points. Then it is geometrically obvious that \( C \) lies on the (infinite) line through \( A \) and \( B \) if, and only if, \( AC \parallel AB \). We need to show that this axiom is symmetric with regard to \( A, B \) and \( C \). In other words, that the statements \( AC \parallel AB, \quad CB \parallel CA \) and \( BA \parallel BC \) are equivalent. It is left as an exercise to prove this equivalence algebraically. See Exercise 5, No.2. This axiom for the collinearity of \( A, B, \) and \( C \) gives the clue on how to describe algebraically the line \( l \) through \( A \) and \( B \).

1.2.1 Generic (parametric) representations of \( l \)

Let \( A \) and \( B \) be two distinct points. Then as \( t \in \mathbb{R} \) varies, the point \( R \) with position vector 
\[
\mathbf{r} = \mathbf{r} (t) = \mathbf{OA} + t\mathbf{AB}
\]
varies over the entire straight line $\ell$ through $A$ and $B$. As in Figure 1.9, from $O$ jump onto the line at $A$ and then go parallel to $AB$ to $R$. The point $R = A$ is obviously on $\ell$ and corresponds with $t = 0$. If $R \neq A$ we have $AR = tAB$ for some $t \neq 0$, so in general

$$\mathbf{r} = \mathbf{OR} = \mathbf{OA} + \mathbf{AR} = \mathbf{OA} + t\mathbf{AB}$$

When $t = 1$ then we are at $R = B$.

As $\mathbf{r} = \mathbf{r}(t)$ represents a general point on the line $\ell$, the equation (1.9) is called a generic or parametric equation of the line through $A$ and $B$. The variable $t$ is called the parameter.

1.2.2 Some general observations

In what follows $\mathbf{a} = \mathbf{OA}$, $\mathbf{b} = \mathbf{OB}$, etc.

1. As $t$ varies over the set $\mathbb{R}$ of real numbers the point $\mathbf{r} = \mathbf{OR} = \mathbf{a} + t(\mathbf{b} - \mathbf{a})$ of equation (1.9) varies continuously over the line $\ell$. In this representation of the line $\ell$ we consider $\mathbf{b} - \mathbf{a} = \mathbf{AB}$ as the direction of $\ell$. In Figure 1.9 as $t$ increases $R$ moves to the right, as $t$ decreases $R$ moves to the left.

2. We remarked that $t = 0$ corresponds to $R = A$. When $t = 1$ we have $R = B$. If $t$ is between 0 and 1, say $t = \frac{1}{2}$ we get

$$\mathbf{OR} = \mathbf{a} + \frac{1}{2}(\mathbf{b} - \mathbf{a}) = \frac{1}{2}(\mathbf{a} + \mathbf{b})$$

Thus $\frac{1}{2}(\mathbf{a} + \mathbf{b})$ is the position vector of the point midway between $A$ and $B$.

3. More generally, a point

$$\mathbf{r} = \mathbf{r}(t) = \mathbf{a} + t(\mathbf{b} - \mathbf{a}) = (1 - t)\mathbf{a} + t\mathbf{b}$$

where $0 \leq t \leq 1$ lies between $A$ and $B$ and the set $AB$ of these points is called the line segment joining $A$ and $B$. If $A \neq B$ and $0 < t < 1$, the point $\mathbf{r}(t)$ is said to lie strictly between $A$ and $B$.

The line segment $AB$ is a set of points and it must be distinguished from the vector $\mathbf{AB}$ and also from the distance $|\mathbf{AB}|$. 

\[\text{Figure 1.9}\]
4. Let \( \mathbf{u} \neq \mathbf{0} \) define a direction and suppose \( \mathbf{a} \) is the position vector of the point \( A \). Then a generic equation of the line through \( A \) with direction \( \mathbf{u} \) is

\[
\mathbf{r}(t) = \mathbf{a} + t\mathbf{u}
\]  

(1.10)

A typical point \( \mathbf{a} + t\mathbf{u} \) on the line is called a **generic point**.

The set \( \ell \) of points \( \mathbf{R} = \mathbf{r}(t) \) satisfying a generic equation such as (1.10) as \( t \) varies over the real numbers is the **analytic** definition of a straight line.

5. A generic equation is not unique.

(a) If \( C \neq D \) are two points on the line given by equation (1.9), then since these points also determine the line, another generic equation (now with parameter \( s \)) is

\[
\mathbf{r}(s) = \mathbf{OC} + s\mathbf{CD}
\]

The parameters \( t \) and \( s \) are of course related. See Example 1.2.1 of subsection 1.2.3 and No.1b of Exercise 5.

(b) When do two generic equations like (1.10) define the same line? Let \( \mathbf{u} \) and \( \mathbf{v} \) be non-zero vectors and suppose that \( \mathbf{a} + t\mathbf{u} \) is a generic equation of line \( \ell_1 \) and \( \mathbf{b} + s\mathbf{v} \) a generic equation of line \( \ell_2 \).

Then \( \ell_1 = \ell_2 \) if, and only if, \( \mathbf{u} \parallel \mathbf{v} \) and \( \mathbf{a} - \mathbf{b} \) is a multiple of \( \mathbf{u} \). (See No.5 of Exercise 5).

### 1.2.3 Some illustrative Examples

**Example 1.2.1** Find two generic equations for the line \( \ell \) passing through \( A = (1, 2, 3) \) and \( B = (-1, 1, 4) \).

**Solution:** One generic equation is

\[
\mathbf{r} = (1, 2, 3) + t((-1, 1, 4) - (1, 2, 3)) \n= (1, 2, 3) + t(-2, -1, 1) = (1 - 2t, 2 - t, 3 + t)
\]

Another is

\[
\mathbf{r} = (-1, 1, 4) + s((1, 2, 3) - (-1, 1, 4)) \n= (-1, 1, 4) + s(2, 1, -1) = (-1 + 2s, 1 + s, 4 - s)
\]

In the second representation we have used \( \mathbf{r} = \mathbf{OB} + s\mathbf{BA} \), with parameter \( s \). The relationship between \( s \) and \( t \) is \( s = 1 - t \), as can be easily seen.

**Example 1.2.2** Find the position vector of the point \( R \) lying on the line \( \ell \) of Example 1.2.1 that is a distance \( \frac{1}{3} |\mathbf{AB}| \) from \( A \) but on the side of \( A \) opposite to that of \( B \).

**Solution:** To find \( R \) put \( t = -\frac{1}{3} \) in \( \mathbf{OR} = \mathbf{OA} + t\mathbf{AB} \) to get

\[
\mathbf{OR} = \mathbf{OA} + \left(-\frac{1}{3}\right) \mathbf{AB} \n= (1, 2, 3) + \left(-\frac{1}{3}\right)(-2, -1, 1) \n= \left(\frac{5}{3}, \frac{7}{3}, \frac{8}{3}\right)
\]
Example 1.2.3 Find the foot $Q$ of the perpendicular from the point $P = (2, -1, 4)$ to the line $\ell$ of Example 1.2.1. Hence find the shortest distance between $P$ and $\ell$.

Solution:

Let $Q = (1 - 2t, 2 - t, 3 + t)$ be the required foot of the perpendicular from $P = (2, -1, 4)$ to the line $\ell$. Then $PQ = (-1 - 2t, 3 - t, -1 + t)$ must be perpendicular to the direction $(-2, -1, 1)$ of $\ell$:

$$(−1−2t,3−t,−1+t)⋅(−2,−1,1)=0$$

This gives $t = \frac{1}{3}$ and so $Q = q = \left(\frac{1}{3}, \frac{5}{3}, \frac{10}{3}\right)$. The shortest distance of $P$ from $\ell$ is thus $\|PQ\| = \left\|(-\frac{5}{3}, \frac{8}{3}, -\frac{2}{3})\right\| = \frac{1}{3}\sqrt{93}$. (See Figure 1.10).

Example 1.2.4 Find the foot $Q$ of the perpendicular from a general point $P = (p_1, p_2, p_3)$ to the line $r = a + tu$ if $a = (1, 2, 3)$ and $u = (-2, -1, 1)$.

Solution:

Let $t$ be such that $Q$ has position vector $\mathbf{a} + t\mathbf{u}$.

As before, we consider $PQ = \mathbf{a} + t\mathbf{u} - \mathbf{p}$ and must solve $PQ \cdot \mathbf{u} = 0$ for $t$ and then substitute this value in $\mathbf{a} + t\mathbf{u}$ to obtain $Q$. We have

$$(\mathbf{a} + t\mathbf{u} - \mathbf{p}) \cdot \mathbf{u} = 0 \quad (1.11)$$

Since $PQ = (1 - 2t - p_1, 2 - t - p_2, 3 + t - p_3)$, the condition $PQ \cdot \mathbf{u} = 0$ becomes

$$(1 - 2t - p_1, 2 - t - p_2, 3 + t - p_3) \cdot (-2, -1, 1) = -1 + 6t + 2p_1 + p_2 - p_3 = 0$$

This gives $t = \frac{1}{6}(1 - 2p_1 - p_2 + p_3)$. Substituting this value of $t$ in $(1 - 2t, 2 - t, 3 + t)$ gives

$$q = \left(\frac{2}{3} + \frac{2}{3}p_1 + \frac{1}{3}p_2 - \frac{1}{3}p_3; \frac{11}{6} + \frac{1}{3}p_1 + \frac{1}{6}p_2 - \frac{1}{3}p_3; \frac{19}{6} - \frac{1}{3}p_1 - \frac{1}{6}p_2 + \frac{1}{6}p_3\right)$$
1.2.4 A general result for the foot of a perpendicular to a line

In Exercise 5, No.11 you are asked to show that the foot Q of the perpendicular from P to the line \( a + tu \) has position vector

\[
q = a + \left( \frac{1}{|u|^2} (p - a) \cdot u \right) u
\]

(1.12)

Example 1.2.5 Let us apply this formula to the above example 1.2.3 to find the foot of the perpendicular from \( P = (2, -1, 4) \) to the line \((1, 2, 3) + t(-2, -1, 1)\).

Solution:

We have \( |u|^2 = 6 \) and the foot has position vector

\[
q = (1, 2, 3) + \left( \frac{1}{6} \right)(2, -1, 4) - (1, 2, 3) \cdot (-2, -1, 1) (-2, -1, 1)
\]

\[
= \left( \frac{1}{3}, \frac{5}{3}, -\frac{10}{3} \right)
\]

This is the same result as before.

1.2.5 Projections on a line through the origin

When a line \( \ell \) passes through the origin, the foot Q of the perpendicular from P to \( \ell \) is called the projection of P on \( \ell \).

The projection of \( p \) on the line \( tu \) is just the projection of \( p \) on the vector \( u \). This is equation (1.7) with \( f = p \) and \( q = u \):

\[
q = \left( \frac{p \cdot u}{|u|^2} \right) u
\]

(1.13)

Example 1.2.6 Find the projection of \( P = (p_1, p_2, p_3) \) on the line \( t(-2, -1, 1) \).

Solution:

In this case \( PQ = (-2t - p_1, -t - p_2, t - p_3) \) and we can proceed as before by solving \((-2t - p_1, -t - p_2, t - p_3) \cdot (-2, -1, 1) = 0 \) for \( t \) and substituting into \((-2, -1, 1) \). It amounts to the same thing to use (1.13) and the required projection is

\[
q = \frac{(p_1, p_2, p_3) \cdot (-2, -1, 1)}{|(-2, -1, 1)|^2} (-2, -1, 1) = \frac{-2p_1 - p_2 + p_3}{6} (-2, -1, 1)
\]

\[
= \left( \frac{2}{3}p_1 + \frac{1}{3}p_2 - \frac{1}{3}p_3 + \frac{1}{3}p_1 + \frac{1}{6}p_2 - \frac{1}{6}p_3, -\frac{1}{3}p_1 - \frac{1}{6}p_2 + \frac{1}{6}p_3 \right)
\]

(1.14)

We will return to this example in Chapter 3 (See Example 3.3.1).

Example 1.2.7 A straight line lies in the \( x - y \) plane and passes through the origin. The line makes an angle of 30° with the positive \( x \)-axis. Find the foot Q of the projection from \( P = (p_1, p_2) \) on the line.

Solution:

Let \( u = (\cos 30°, \sin 30°) = \left( \frac{\sqrt{3}}{2}, \frac{1}{2} \right) \). The parametric equation of the line is \( \ell = t \left( \frac{\sqrt{3}}{2}, \frac{1}{2} \right) \) and the foot Q of the perpendicular from P to the line is the projection of \( p = (p_1, p_2) \) on the unit vector \( u \):

\[
q = \left( \frac{\sqrt{3}p_1}{2} + \frac{p_2}{2} \right) \left( \frac{\sqrt{3}}{2}, \frac{1}{2} \right) = \left( \frac{1}{4} \left( 3p_1 + \sqrt{3}p_2 \right), \frac{1}{4} \sqrt{3}p_1 + \frac{1}{4}p_2 \right)
\]

(1.15)

We shall return to this example in subsection 1.4 below and again in Chapter 3 Example 3.2.1.
A well-known result on parallelograms

Example 1.2.8 Show that the diagonals of a parallelogram bisect each other.

Solution:
In Figure 1.11 $ABCD$ is a parallelogram with $AB = DC$ and $BC = AD$. It is understood that the parallelogram is non-degenerate in that no three of $A, B, C, D$ are collinear.

Let $X$ be the midpoint of $AC$ so that $AX = XC$. Then $AB + BX = AX = XC$, so

$$BX = XC - AB = XC - DC = XC + CD = XD.$$
A familiar example of skew lines is provided by a road passing under a railway bridge. A train and car can only crash at a level crossing!
In the case of skew lines there will be points $P$ on $\ell_1$ and $Q$ on $\ell_2$ such that $PQ$ is perpendicular to the directions of both lines and the distance $|PQ|$ is minimum. See Figure 1.12.

**Example 1.2.9** Let $\ell_1$ be the line through $A = (0,0,0)$ and $B = (1,2,-1)$ and suppose $\ell_2$ is the line through $C = (1,1,1)$ and $D = (2,4,3)$. Are the lines skew? In any case, find the shortest distance between them.

**Solution:** Generic equations for $\ell_1$ and $\ell_2$ are $t(1,2,-1)$ and $(1,1,1) + s(1,3,2)$ respectively. Since $(1,2,-1) \not\parallel (1,3,2)$, the lines are not parallel (and so cannot possibly be identical). They will meet if for some $t$ and $s$

$$t(1,2,-1) = (1,1,1) + s(1,3,2)$$

This means that the three equations $t = 1 + s$, $2t = 1 + 3s$ and $-t = 1 + 2s$ must hold. From the first and last equation, $1 + s = -(1 + 2s)$ and thus $s = -\frac{2}{3}$, $t = \frac{1}{3}$. But then $2t = \frac{2}{3} \neq 1 + 3(-\frac{2}{3})$ and the second equation fails. It follows that $\ell_1$ and $\ell_2$ are skew.

Let $P = t(1,2,-1)$ and $Q = (1,1,1) + s(1,3,2)$ be the points where $PQ$ is minimum. Then $PQ = (1,1,1) + s(1,3,2) - t(1,2,-1)$ is perpendicular to both directions $(1,2,-1)$ and $(1,3,2)$. Hence:

$$(1,1,1) + s(1,3,2) - t(1,2,-1)$$

Solving these two equations gives $s = -\frac{26}{59}$ and $t = -\frac{2}{59}$. Hence,

$$PQ = (1,1,1) + s(1,3,2) - t(1,2,-1) = \left(\frac{35}{59}, \frac{-15}{59}, \frac{5}{59}\right)$$

The shortest distance between the lines is $\left|\left(\frac{35}{59}, \frac{-15}{59}, \frac{5}{59}\right)\right| = \frac{5}{\sqrt{59}}$.

**Exercise 5**

1. Let $A = (1,2,-3)$, $B = (-1,1,2)$, $C = (5,4,-13)$, $D = (-5,-1,12)$ and $E = (-1,-2,3)$. Answer the following questions.

(a) Find a generic equation with parameter $t$ of the straight line $\ell$ through $A$ and $B$ in the form of equation (1.9).

(b) Is $E$ on $\ell$? Show that $C$ and $D$ are on $\ell$ and find a corresponding generic equation of the line with parameter $s$. Find the relationship between $s$ and $t$.

(c) Find two points on $\ell$ at a distance $\sqrt{24}$ from $E$.

(d) Find an equation describing the line segment joining $A$ and $B$. In particular find the midpoint of the segment and also the three points strictly between $A$ and $B$ dividing the segment into four equal parts.

2. * Let $A$, $B$ and $C$ be three distinct points. Show that any of the conditions $AB \parallel AC$, $BC \parallel AC$, and $AB \parallel BC$ are equivalent.

**Hint 6** Since $A$, $B$ and $C$ are distinct, $AB = b - a$, $AC = c - a$ and $BC = c - b$ are all non-zero.

Let $p$ be the statement $AB \parallel AC$, and $q$ the statement $BC \parallel AC$ and $r$ the statement $AB \parallel BC$.

The statement $p$ says there is $s$ with $b - a = s(c - a)$. From this we get $c - b = c - (a + s(c - a)) = c - a - s(c - a) = (1 - s)(c - a)$. We have just shown $p \Rightarrow q$ ($p$ implies $q$).

Now substitute $C$ for $A$, $A$ for $B$ and $B$ for $C$ in $p \Rightarrow q$. You should get $q \Rightarrow r$. Making the same substitutions in $q \Rightarrow r$ you should get $r \Rightarrow p$. 

3. Use the previous criterion to show that any three points satisfying equation (1.10) lie on a straight line (are collinear) and hence that the equation does indeed define a straight line.

**Hint 7** Use the results of the previous problem.

4. Let the generic equation \( \mathbf{r}(t) = \mathbf{a} + t\mathbf{u} \) represent a line \( \ell \), where \( \mathbf{u} \neq \mathbf{0} \) and suppose that \( P \) with position-vector \( \mathbf{p} \) is a point. Show that \( P \) lies on \( \ell \)

(a) if, and only if, \( \mathbf{p} - \mathbf{a} \) is a multiple of \( \mathbf{u} \).

(b) if, and only if, the equation \( \mathbf{a} + tu = \mathbf{p} \)

has a unique solution for \( t \).

(Compare this with No.14 below for another criterion).

5. * Let \( \mathbf{u} \) and \( \mathbf{v} \) be non-zero vectors and suppose that \( a + tu \) and \( b + sv \) are generic equations of lines \( \ell_1 \) and \( \ell_2 \) respectively. Show that necessary and sufficient conditions for the generic equations to represent the same set of points (i.e. \( \ell_1 = \ell_2 \)) are

(i) \( \mathbf{u} \parallel \mathbf{v} \) and

(ii) \( a - b \) is a multiple of \( \mathbf{v} \).

Using this result show that two lines sharing two distinct points are identical.

6. * Use the results of the previous exercise to show that the four conditions in subsection 1.2.6 are in fact mutually exclusive and exhaustive.

7. Find - but not by simply substituting in equation (1.12) - the point \( Q \) of the line \( \ell \) with parametric equation \((1 - 2t, 2 - t, -3 + 5t)\) which is closest to

(a) the origin,

**Solution.** Solve \((1 - 2t, 2 - t, -3 + 5t) \cdot (-2, -1, 5) = -19 + 30t = 0 \). So \( t = \frac{19}{30} \) and \( Q = (1 - 2t, 2 - t, -3 + 5t) = (-\frac{4}{15}, \frac{41}{30}, \frac{1}{6}) \)

(b) the point \( P = (5, 8, 4) \),

(c) the point \( P = (-3, 2, 1) \).

8. Find the projections of the following points \( P \) on the line \((-2t, -t, 5t)\).

(a) \( P = (4, 6, 7) \)

**Solution:** We have \((-2t, -t, +5t) - (4, 6, 7) = (-2, -1, 5) = -21 + 30t = 0 \) and \( t = \frac{7}{10} \). The projection is \((-2t, -t, +5t) = (-\frac{7}{5}, -\frac{7}{10}, \frac{7}{2}) \). Alternatively, the projection of \( P \) on \((-2, -1, +5)\) is 

\[
\frac{(4, 6, 7) \cdot (-2, -1, +5)}{|(-2, -1, +5)|^2} (-2, -1, +5) = \frac{7}{10} (-2, -1, +5)
\]

(b) \( P = (-1, -2, 3) \).

**Hint 8** These can be done in the same way you handled the previous question, but realize that you can use the projection of \( P \) on the vector \((-2, -1, 5)\).

9. Check your results of 7a and 7b using equation (1.12). Can you see any connection between 7a and 8b and also between between 7b and 8a?
10. Find the projection of the point $P$ with position vector $p = (p_1, p_2, p_3)$ on the line $(-2t, -t, 5t)$.

11. Prove the formula (1.12) that finds the foot $Q$ of the perpendicular from $P = (p_1, p_2, p_3)$ onto the line $a + tu$. Deduce the projection formula (1.13).

**Hint 9** Show that the solution to $(a + tu - p) \cdot u = 0$ (equation (1.11)) is

$$t = -\frac{(a - p) \cdot u}{|u|^2}$$

12. Show that the closest line $\ell$ with parametric equation $a + tu$ comes to the point $P$ is

$$d = \frac{1}{|u|} \sqrt{(|p - a| u)^2 - ((p - a) \cdot u)^2} \quad (1.16)$$

13. Show that $|p(t)|^2$ in equation (1.10) is a quadratic in $t$. How does your knowledge of quadratics tie up with the formula (1.12) and equation (1.16)?

14. Deduce from equation (1.16) that $\ell$ passes through the point $P$ if, and only if,

$$(p - a) \cdot u = \pm|p - a||u|. \quad \text{Note that geometrically this is clear in view of the formula for the angle between} \ n p - a \ \text{and} \ u \ \text{if} \ p - a \neq 0.$$ 

15. * If $t$ is time in seconds and $u$ is the displacement in meters per second (the constant velocity), then equation (1.10) can be interpreted as the position of (say) an aircraft in meters at time $t$.

(a) In a situation like this we would be interested in how close the aircraft comes to a certain point $P$ for $t \geq 0$. Find a condition for (1.16) to hold at some time $t \geq 0$. If this condition fails, what is the closest the aircraft gets to $P$ and when does this occur?

(b) How can your findings be used to find the closest distance two aircraft travelling with constant velocities get to each other, assuming that as time $t \geq 0$ increases the distance between the aircraft decreases?

**Hint 10** Let $\alpha + tu$ and $\beta + tv$ describe the positions of the two aircraft at time $t$. Consider the line $\beta - \alpha + t(v - u)$. You want its closest distance from the origin. By assumption, why is $(\beta - \alpha) \cdot (v - u) < 0$?

16. A rhombus is a (non-degenerate) parallelogram all of whose sides have the same length. Show that the diagonals of a rhombus intersect at right angles.

**Hint 11** In Example 1.2.8 (Figure 1.11) use $|AB| = |BC|$ as well as $AC = AB + BC$ and $AB + BD = AD$, so $BD = AD - AB = BC - AB$. Now use the result of Exercise 2 No.17.

17. *(Generalization of previous exercise).* Let $a + tu$ and $b + sv$ represent two straight lines and suppose that $A = a + t_1 u$, $B = b + s_1 v$, $C = a + t_2 u$, and $D = b + s_2 v$ are points on these lines with $A \neq C$ and $B \neq D$.

Show that $|AB|^2 + |CD|^2 = |BC|^2 + |DA|^2$ if, and only if, the lines are perpendicular.

**Hint 12** Expand $|AB|^2 = (b + s_1 v - (a + t_1 u)) \cdot (b + s_1 v - (a + t_1 u))$ etc. and show that $|AB|^2 + |CD|^2 - |BC|^2 - |DA|^2 = 2u \cdot v(t_1 - t_2)(s_2 - s_1)$. 


18. Let $A$, $B$ and $C$ be three points not on a line and so forming the distinct vertices of a triangle. The median from $A$ is the line joining $A$ to the midpoint of the opposite side $BC$. Show that all three medians intersect in a single point $M$. This point is called the \textbf{centroid} of the triangle and is the \textbf{centre of mass} of three equal masses placed at the vertices of the triangle.

\textbf{Hint 13} Let $\mathbf{a}$, $\mathbf{b}$, and $\mathbf{c}$ be the position vectors of $A$, $B$ and $C$ respectively. Then the midpoint of $BC$ has position vector $\frac{1}{2}(\mathbf{b} + \mathbf{c})$. A generic equation of the line through $A$ and the midpoint of $BC$ is

$$\mathbf{a} + t \left( \frac{1}{2}(\mathbf{b} + \mathbf{c}) - \mathbf{a} \right)$$

Put $t = \frac{2}{3}$ and simplify the result. You should conclude that the centroid lies two thirds of the way from any vertex along its median toward the opposite side.

19. (Cartesian equations for a straight line). Let equation (1.10) define a straight line, where $\mathbf{u} = (u_1, u_2, u_3)$ and each $u_i \neq 0$. Show that $\mathbf{r} = (x, y, z)$ is on the line if, and only if,

$$\frac{x - a_1}{u_1} = \frac{y - a_2}{u_2} = \frac{z - a_3}{u_3}$$

What happens if one or two of the $u_i = 0$? Find Cartesian equations for the line of Question 1a above. Find generic and Cartesian equations for the $x-$, $y-$ and $z-$axis.

20. Let $\ell_1$ be the line through $A = (-1, -2, 6)$ and $B = (3, 2, -2)$ and $\ell_2$ the line through $C = (1, 3, 1)$ and $D = (1, 1, 5)$. Solve the following problems.

(a) Show that $\ell_1$ and $\ell_2$ are skew.

(b) Find points $P$ on $\ell_1$ and $Q$ on $\ell_2$ such that $PQ$ is perpendicular to the directions of both lines. Hence find the shortest distance between $\ell_1$ and $\ell_2$.

21. Let $\ell_1$ be the line through $A = (2, -1, 4)$ and $B = (3, 0, 5)$ and $\ell_2$ the line through $C = (-1, 0, 1)$ and $D = (-2, 1, 0)$. Show that $\ell_1$ and $\ell_2$ meet and find the point of intersection.

22. How are our general results affected (and simplified) when one of the coordinates (say the $z-$ coordinate) is restricted to be zero? In particular what does equation (1.12) reduce to?

1.3 Planes

1.3.1 Generic Equation for a plane

Our first object is to find a generic equation for the plane passing through three \textbf{non-collinear points} $A = (a_1, a_2, a_3)$, $B = (b_1, b_2, b_3)$ and $C = (c_1, c_2, c_3)$.

If four points lie on a plane, they are called \textbf{coplanar}. So if $R$ is a point, our first question is “when are $A$, $B$, $C$ and $R$ coplanar?” The answer is suggested by Figure 1.13:

Suppose that $R$ lies in the same plane as $A$, $B$ and $C$. Let the line through $R$ parallel to $\overrightarrow{AC}$ meet the line through $A$ and $B$ at the point $P$. Then $\overrightarrow{AP} = s\overrightarrow{AB}$ for some scalar $s$. Similarly, let the line through $R$ parallel to $\overrightarrow{AB}$ meet the line through $A$ and $C$ at the point $Q$, so that $\overrightarrow{AQ} = t\overrightarrow{AC}$ for some scalar $t$. The figure $APRQ$ forms a parallelogram and

$$\overrightarrow{OR} = \overrightarrow{OA} + \overrightarrow{AR} = \overrightarrow{OA} + (s\overrightarrow{AB} + \overrightarrow{PR}) = \overrightarrow{OA} + (s\overrightarrow{AB} + \overrightarrow{AQ}) = \overrightarrow{OA} + s\overrightarrow{AB} + t\overrightarrow{AC}$$

(1.17)

This equation describes the position vector $\mathbf{r} = \overrightarrow{OR}$ of a general point on the plane through $A$, $B$ and $C$ and is called a \textbf{generic equation} of the plane through $A$, $B$ and $C$. 
Example 1.3.1 Let \( A = (1, -1, 2) \), \( B = (1, 2, 3) \) and \( C = (-1, 0, 1) \) be three given points. They are not collinear since the vectors \( \overrightarrow{AP} = (0, 3, 1) \) and \( \overrightarrow{AC} = (-2, 1, -1) \) are not parallel. Hence a generic equation of the plane through \( A \), \( B \) and \( C \) is

\[
\mathbf{r} = (1, -1, 2) + s(0, 3, 1) + t(-2, 1, -1)
\]

In equation (1.17) the vectors \( \mathbf{u} = \overrightarrow{AB} \) and \( \mathbf{v} = \overrightarrow{AC} \) are non-zero and non-parallel and are thought of as parallel to the plane described. This suggests the following definition:

1.3.2 Analytic definition of a plane

Let \( \mathbf{u} \) and \( \mathbf{v} \) be two non-zero, non-parallel vectors and \( A = \mathbf{a} \) a given point. Then the set \( \Pi \) of points \( \mathbf{r} = \mathbf{r}(s, t) = \mathbf{a} + s\mathbf{u} + t\mathbf{v} \) (1.18)

describes a plane. The parameters are \( s \) and \( t \) and we call \( \mathbf{a} + s\mathbf{u} + t\mathbf{v} \) a generic point of the plane.

1.3.3 Equality of planes

As with lines, a generic equation for a plane is not unique. Let

\[
\mathbf{a}_1 + s_1\mathbf{u}_1 + t_1\mathbf{v}_1 \quad \text{and} \quad \mathbf{a}_2 + s_2\mathbf{u}_2 + t_2\mathbf{v}_2
\]

Represent planes \( \Pi_1 \) and \( \Pi_2 \) respectively. Then \( \Pi_1 = \Pi_2 \) if, and only if,

(i) \( \text{sp} (\mathbf{u}_1, \mathbf{v}_1) = \text{sp} (\mathbf{u}_2, \mathbf{v}_2) \)

and

(ii) \( \mathbf{a}_1 - \mathbf{a}_2 \) is in this common span.

For the meaning of ‘span’ see No.14 of Exercise 2 and compare No.5 of Exercise 5.

The proof of the above statement is left as an exercise. See No.11 of Exercise 16 below.
1.3.4 Cartesian Equation of a plane

A vector \( \mathbf{n} = (n_1, n_2, n_3) \neq 0 \) is called a normal of the plane defined by equation (1.18) if it is perpendicular to every line lying in the plane. So for a variable point \( R \) on the plane we must have \( \mathbf{n} \cdot \mathbf{AR} = 0 \). Thus as in Figure 1.14, all points on the plane with position vector \( \mathbf{r} = (x_1, x_2, x_3) \) must satisfy

\[
\mathbf{n} \cdot (\mathbf{r} - \mathbf{a}) = 0
\]

This is a Cartesian equation of the plane.

More fully this reads

\[
n_1x_1 + n_2x_2 + n_3x_3 - (n_1a_1 + n_2a_2 + n_3a_3) = 0
\]

(1.19)

where \( c = -(n_1a_1 + n_2a_2 + n_3a_3) \)

A Cartesian equation is unique up to non-zero constant multiples of \( (n_1, n_2, n_3, c) \) since evidently multiplying \( n_1, n_2, n_3 \) and \( c \) by \( \beta \neq 0 \) cannot change a solution \( (x_1, x_2, x_3) \).

1.3.5 Finding a normal \( \mathbf{n} \)

If \( \mathbf{n} \) is a normal to the plane given by equation (1.18), then in particular \( \mathbf{n} \perp n \) and \( \mathbf{n} \perp v \). In other words, simultaneously

\[
n_1u_1 + n_2u_2 + n_3u_3 = 0
\]

and

\[
n_1v_1 + n_2v_2 + n_3v_3 = 0
\]

(1.20)

Conversely, if equation (1.20) holds then \( \mathbf{n} \) must be a normal (see Exercise 16, No.4 below). Because equation (1.20) involves two equations for three unknowns, a solution \( \mathbf{n} \neq 0 \) can always be found. See Chapter 2, 2.2.15 for a proof of this.

Example 1.3.2 Find a normal \( \mathbf{n} \neq 0 \) to the plane of Example 1.3.1 and hence its Cartesian equation.
1.3. PLANES

Solution:

We solve
\[(n_1, n_2, n_3) \cdot (0, 3, 1) = 3n_2 + n_3 = 0\]
\[(n_1, n_2, n_3) \cdot (-2, 1, -1) = -2n_1 + n_2 - n_3 = 0\]
simultaneously for \(n_1\), \(n_2\) and \(n_3\). (We did much the same thing in Example 1.1.3). From the first equation, \(n_3 = -3n_2\). Substituting this into the second equation, we get \(n_1 = \frac{1}{2} (n_2 - n_3) = 2n_2\). Thus, provided \(n_2 \neq 0\), the vector
\[n = (2n_2, n_2, -3n_2) = n_2 (2, 1, -3)\]
is perpendicular (normal) to the plane. We may take \(n_2 = 1\) and \(n = (2, 1, -3)\). With \(A = (1, -1, 2)\) we get a Cartesian equation for the plane:
\[(2, 1, -3) \cdot ((x_1, x_2, x_3) - (1, -1, 2)) = 2x_1 + x_2 - 3x_3 + 5 = 0\]

1.3.6 Some remarks on planes

Remark 14 The following statements are geometrically obvious, but analytic proofs can be given after we have studied the subject more deeply, in particular after we have done Chapter 3. See No.10 in Exercise 78.

1. Provided \(n = (n_1, n_2, n_3) \neq 0\), an equation \(n_1x_1 + n_2x_2 + n_3x_3 + c = 0\) always represents a plane with normal \(n\). (If \(n = 0\) the equation represents nothing if \(c \neq 0\) and all of space if \(c = 0\)).

2. By definition, two non-parallel planes are parallel if their normals are parallel. Such planes are a constant positive distance apart and have no points in common.

3. Two non-parallel planes meet in a line.

4. By definition, a line is parallel to a plane if it is perpendicular to its normal (draw a picture to see that this is correct). A line not parallel to a plane meets the plane in exactly one point.

5. A line and a point not on it determine a unique plane.

6. There are an infinity of planes containing a given line.

Example 1.3.3 Find a generic equation for the line of intersection of the planes
\[x_1 + x_2 + x_3 + 1 = 0\]
\[2x_1 + x_2 - x_3 + 2 = 0\]

Solution:

Note that \((1, 1, 1) \nmid (2, 1, -1)\) so the planes are not parallel. Eliminating \(x_1\) gives \(x_2 + 3x_3 = 0\) or \(x_2 = -3x_3\). Substituting \(x_2 = -3x_3\) into the first (or second) equation gives \(x_1 = 2x_3 - 1\). Hence, with \(t = x_3\) the parameter,
\[r = (x_1, x_2, x_3) = (2t - 1, -3t, t)\]
is the line of intersection of the two planes.

Example 1.3.4 Find a generic equation of the plane with Cartesian equation
\[x_1 - 2x_2 + 5x_3 = 10\]
Solution:
By far the easiest solution is
\[ \mathbf{r} = (10 + 2x_2 - 5x_3, x_2, x_3) \]

Here the independent parameters are \( x_2 \) and \( x_3 \) and they can vary freely over \( \mathbb{R} \). We could of course let \( s = x_2 \) and \( t = x_3 \), but this is only an artificial device and changes nothing.

A more long-winded approach is to find three non-collinear points \( A, B \) and \( C \) on the plane and then proceed as in Example (1.3.1).

Letting \( x_2 = x_3 = 0 \) gives \( x_1 = 10 \) and \( A = (10, 0, 0) \) as one point. Putting \( x_1 = x_2 = 0 \) gives \( x_3 = 2 \) and \( B = (0, 0, 2) \). With \( x_1 = x_3 = 0 \) we have \( x_2 = -5 \) and \( C = (0, -5, 0) \). The points \( A, B \) and \( C \) do not lie on a line and another generic equation of the plane is

\[ \mathbf{r} = OA + sAB + tAC = (10 - 10s - 10t, -5t, 2s) \]

Remark 15 Instead of \( A, B \) and \( C \) we could have chosen any three non-collinear points satisfying equation (1.21). The parametric equation will be different but will represent the same plane.

Example 1.3.5 Find two non-parallel vectors \( \mathbf{u} \) and \( \mathbf{v} \) both perpendicular to \( \mathbf{n} = (1, -2, 5) \).

Solution:
Two such vectors are \( \mathbf{u} = \mathbf{AB} \) and \( \mathbf{v} = \mathbf{AC} \).

Example 1.3.6 Find the point of intersection of the line through \((1, -1, 0)\) and \((0, 1, 2)\) and the plane

\[ 3x_1 + 2x_2 + x_3 + 1 = 0 \]

Solution:
For some \( t \) the point

\[ \mathbf{r} = (1, -1, 0) + t \left((0, 1, 2) - (1, -1, 0)\right) \]

\[ = (1, -1, 0) + t (-1, 2, 2) = (1 - t, -1 + 2t, 2t) \]

must be on the given plane, i.e.

\[ 3(1 - t) + 2(-1 + 2t) + (2t) + 1 = 0 \]

Thus \( t = -\frac{3}{2} \) and the point of intersection is \((\frac{5}{4}, -\frac{7}{4}, -\frac{4}{3})\).

1.3.7 Perpendicular from a Point to a Plane. Projections

If \( \Pi \) is a plane and \( P \) a point, the problem in this section is to find the foot \( Q \) of the perpendicular from \( P \) to the plane. See Figure 1.15.

If the plane passes through the origin, the foot \( Q \) is called the projection of \( P \) on the plane.

Example 1.3.7 Find the foot \( Q \) of the perpendicular from the point \( P = (4, -1, 3) \) to the plane \( x + 2y + z = -4 \). Hence find the shortest distance from \( P \) to the given plane.

Solution:
The vector \((1, 2, 1)\) is normal to the plane so the line \( \ell \) defined by \((4, -1, 3) + t (1, 2, 1) = (4 + t, -1 + 2t, 3 + t) \) passes through \( P \) and has direction normal to the plane. Geometry tells
us that the line $\ell$ hits the plane at the required foot $Q$ of the perpendicular. Then $|PQ|$ will be the shortest distance of $P$ from the plane.

To find $t$, substitute $(4 + t, -1 + 2t, 3 + t)$ into the Cartesian equation of the plane

$$(4 + t) + 2(-1 + 2t) + (3 + t) = -4$$

and solve for $t$. This gives $t = -\frac{3}{2}$ and the position vector of the point $Q$ therefore has position vector

$$q = \left(4 + \left(-\frac{3}{2}\right), -1 + 2\left(-\frac{3}{2}\right), 3 + \frac{3}{2}\right) = \left(\frac{5}{2}, -4, \frac{3}{2}\right)$$

The required shortest distance of $P$ from the plane is

$$|PQ| = \left|\left(\frac{5}{2}, -4, \frac{3}{2}\right) - (4, -1, 3)\right|$$

$$= \left|\left(-\frac{3}{2}, -3, -\frac{3}{2}\right)\right| = \sqrt{\frac{27}{2}}$$

**Example 1.3.8** Find the foot $Q$ of the perpendicular from a general point $P = (p_1, p_2, p_3)$ to

1. the above plane $x + 2y + z = -4$, and
2. the parallel plane $x + 2y + z = 0$ which passes through the origin.

**Solution:**

1. Consider the line $(p_1, p_2, p_3) + t(1, 2, 1) = (p_1 + t, p_2 + 2t, p_3 + t)$. As above, we find where this line hits the given plane. The equation determining this point $Q$ is

$$(p_1 + t) + 2(p_2 + 2t) + (p_3 + t) = -4$$

This gives $t = -\frac{4 + p_1 + 2p_2 + p_3}{6}$. Hence $Q$ is

$$q = (p_1, p_2, p_3) - \frac{4 + p_1 + 2p_2 + p_3}{6} (1, 2, 1) \quad (1.22)$$
2. We must substitute \((p_1, p_2, p_3) + t(1, 2, 1) = (p_1 + t, p_2 + 2t, p_3 + t)\) into the equation 
\[ x + 2y + z = 0 \] of the plane:

\[(p_1 + t) + 2(p_2 + 2t) + (p_3 + t) = 0 \]

This gives 
\[ t = -\frac{p_1 + 2p_2 + p_3}{6} \]
and hence 
\[ q = (p_1, p_2, p_3) - \frac{p_1 + 2p_2 + p_3}{6} (1, 2, 1) \quad (1.23) \]

Because the plane \(x + 2y + z = 0\) passes through the origin the point \(Q\) given by equation (1.23) is the projection of \(P\) on the plane.

See Exercise 16, (5), where the theory is taken further.

**Exercise 16**  
1. Show that \(A = (1, 2, 1), B = (0, 1, 3)\) and \(C = (-1, 2, 5)\) are not collinear  
and so define a unique plane \(\Pi\). Find:

(a) A generic equation for \(\Pi\)  
(b) A normal to \(\Pi\) and hence a Cartesian equation for \(\Pi\)  
(c) The foot \(Q\) of the perpendicular from \(P = (1, 2, 3)\) to \(\Pi\) and so the shortest distance 
from \(P\) to the plane \(\Pi\)  
(d) The foot \(Q\) of the perpendicular from \(P = (p_1, p_2, p_3)\) to \(\Pi\) (see also No.5 below).  
(e) The intersection of the line \(\ell\) with parametric form \((1 + t, 3 + 2t, 4 - t)\) and \(\Pi\), as 
well as the acute angle between the line and the normal to \(\Pi\).  
(f) A plane containing the point \((4, 5, -3)\) and the line \((1 + t, 3 + 2t, 4 - t)\).

2. Find a generic equation for the plane \(2x + 3y - 4z + 5 = 0\).

3. Find generic and Cartesian equations for the \(x - y\) and \(y - z\) planes.

4. Prove that if \(\mathbf{n}\) satisfies equation (1.20) then it is normal to the plane given by the generic 
equation (1.18).

5. Follow the method of Examples (1.3.7) and (1.3.8) above to show that the foot \(Q\) of the 
perpendicular from the point \(P\) to the plane \(\Pi\) with Cartesian equation \(\mathbf{n} \cdot \mathbf{r} + c = 0\) is 
given by

\[ q = p - \left(\frac{c + \mathbf{n} \cdot p}{|\mathbf{n}|^2}\right) \mathbf{n} \quad (1.24) \]

Here \(p\) is the position vector of \(P\).  
Consider also the case \(c = 0\). Then \(Q\) is called the **projection** of \(P\) on the plane.

**Hint 17** Solve

\[ \mathbf{n} \cdot (p + t\mathbf{n}) + c = 0 \]

for \(t\).

6. Use equation (1.24) to show that the shortest distance of the point \(P\) from the plane 
\(\mathbf{n} \cdot \mathbf{r} + c = 0\) is

\[ \frac{|\mathbf{n} \cdot p + c|}{|\mathbf{n}|} \]

7. Write the equation of a plane \(\mathbf{n} \cdot \mathbf{r} + c = 0\) in the form \(\mathbf{m} \cdot \mathbf{r} + d = 0\) where \(\mathbf{m}\) is unit 
vector.
8. Let \( \mathbf{n} \cdot \mathbf{r} + c_1 = 0 \) and \( \mathbf{n} \cdot \mathbf{r} + c_2 = 0 \) represent planes \( \Pi_1 \) and \( \Pi_2 \) respectively. Show algebraically that \( \Pi_1 \) and \( \Pi_2 \) are either identical or have no point in common. In the latter case write down a formula for the shortest distance between \( \Pi_1 \) and \( \Pi_2 \) if \( P = (p_1, p_2, p_3) \) is a given point on the first plane.

9. Find a generic equation of the line of intersection of the planes \( 2x + 3y + z + 1 = 0 \) and \( x - 4y + 5z - 2 = 0 \).

10. (Hirst, p.36). In 10a - 10c find the intersection of the three given planes. In each case give a brief geometric description.

(a)  
\[
\begin{align*}
3x - 2y + 2z &= 5 \\
4x + y - z &= 3 \\
x + y + z &= 6
\end{align*}
\]

(b)  
\[
\begin{align*}
x + 2y - z &= 3 \\
2x + 4y - 2z &= 5 \\
x - y + z &= 1
\end{align*}
\]

(c)  
\[
\begin{align*}
x + 2y - z &= 2 \\
x - y + z &= 1 \\
3x + 3y - z &= 5
\end{align*}
\]

11. * Prove the statement of subsection 1.3.3.

**Hint 18** Imitate the procedure of No.5 of Exercise 5.

12. (See Figure 1.16) Let \( \mathbf{n} \cdot \mathbf{r} + c = 0 \) and \( \mathbf{m} \cdot \mathbf{r} + d = 0 \) represent non-parallel planes \( \Pi_1 \) and \( \Pi_2 \) respectively. Show that the Cartesian equation of a plane \( \Pi \) that contains the line of intersection of \( \Pi_1 \) and \( \Pi_2 \) has the form

\[
\lambda (\mathbf{n} \cdot \mathbf{r} + c) + \mu (\mathbf{m} \cdot \mathbf{r} + d) = 0
\]

where \( \lambda \) and \( \mu \) are two scalars not both of which are zero.

**Hint 19** First note that because \( \mathbf{n} \neq 0 \) and \( \mathbf{m} \neq 0 \) are not parallel, the linear combination \( \lambda \mathbf{n} + \mu \mathbf{m} \) is also non-zero if one or both of \( \lambda, \mu \) are non-zero. In that case equation (1.25) represents a plane \( \Pi \). Note that if (e.g.) \( \lambda \neq 0 \) and \( \mu = 0 \) then \( \Pi \) is just \( \Pi_1 \). Now check that a point on \( \Pi_1 \) and \( \Pi_2 \) is also on \( \Pi \).

13. Find the form of a Cartesian equation of the plane \( \Pi \) that contains the intersection of the planes

\[
2x_1 + 3x_2 + x_3 + 1 = 0 \quad \text{and} \quad x_1 - 4x_2 + 5x_3 - 2 = 0
\]

If the plane \( \Pi \) also contains the point \((1, -1, 2)\), find its Cartesian equation.
1.4 Reflections in a line or a plane

A line or plane can serve as a mirror so that a point $P$ is reflected in a point $P'$ in such a way that $P$ and $P'$ are symmetric with respect to the line or plane, as in Figure 1.17. From the figure the point $P'$ is given by

$$
\mathbf{p}' = \mathbf{OP}' = \mathbf{OP} + 2\mathbf{PQ} = \mathbf{p} + 2(\mathbf{q} - \mathbf{p}) = 2\mathbf{q} - \mathbf{p}
$$

(1.26)

where $Q$ is the foot of the perpendicular from $P$ to the line or plane and we have written $\mathbf{p}'$ for the position vector of $P'$.

Example 1.4.1 Consider example 1.2.7 of a line in the $x$-$y$ plane. Using equation (1.15) and equation (1.26), the reflection of $\mathbf{p} = (p_1, p_2)$ in the line is

$$
\mathbf{p}' = 2\mathbf{q} - \mathbf{p} = 2\left(\frac{1}{4}(3p_1 + \sqrt{3}p_2), \frac{1}{4}\sqrt{3}p_1 + \frac{1}{4}p_2\right) - (p_1, p_2)
$$

(1.27)

Example 1.4.2 Consider (1) of Example 1.3.8 where we found the foot of the perpendicular from $\mathbf{p}$ to the plane $x + 2y + z = -4$ given by equation (1.22):

$$
\mathbf{q} = (p_1, p_2, p_3) = \frac{4 + p_1 + 2p_2 + p_3}{6}(1, 2, 1)
$$

From this it follows that the reflection $P'$ of $P$ in the plane is the point with position

$$
\mathbf{p}' = 2\mathbf{q} - \mathbf{p} = p - \frac{4 + p_1 + 2p_2 + p_3}{3}(1, 2, 1)
$$

Exercise 20 1. Using equation (1.12) - which finds the foot $Q$ of the perpendicular from $P$ to the line $\ell$ with parametric equation $\mathbf{r} = \mathbf{a} + t\mathbf{u}$ - find a formula for the reflection $P'$ of $P$ in the line $\ell$.
2. Using equation (1.24) - which finds the foot $Q$ of the perpendicular from $P$ to the plane with Cartesian equation $\mathbf{n} \cdot \mathbf{r} + c = 0$ - find a formula for the reflection $P'$ of $P$ in the given plane.

3. Specialize your formulae when $a = 0$ and $c = 0$.

4. Using the answers to Questions (2) and (3), or otherwise, find formulae for reflections in the planes $x + 2y + z = 0$, $3x - y - 2z = 1$ and $3x - y - 2z = 0$.

5. In the $x - y$ plane find a formula for reflection in the line passing through the origin and making an angle of $60^\circ$ with the positive $x$-axis.

6. * In the $x - y$ plane find a formula for reflection in the line passing through the origin and making an angle of $\theta$ radians with the positive $x$-axis.

**Hint 21** All these questions use equation (1.26)
1.5 Summary of Chapter 1

1.5.1 Points in Space, Vectors

(See section 1.1).
A point \( P = (p_1, p_2, p_3) \) in space is identified with its position vector \( \overrightarrow{OP} = \mathbf{p} \). The displacement from \( A \) to \( B \) is \( \overrightarrow{AB} = \mathbf{b} - \mathbf{a} \).

Vectors \( \mathbf{u} = (u_1, u_2, u_3) \) are underlined.
Scalars \( \alpha, \beta, ..., s, t, ... \) are real numbers.

The product of a vector \( \mathbf{a} \) by a scalar \( \gamma \) is \( \gamma \mathbf{a} = (\gamma a_1, \gamma a_2, \gamma a_3) \). Non-zero vectors \( \mathbf{a} \) and \( \mathbf{b} \) are parallel if \( \mathbf{b} = s\mathbf{a} \) for a non-zero scalar \( s \).

Vectors \( \mathbf{a} \) and \( \mathbf{b} \) are added:
\[
\mathbf{a} + \mathbf{b} = (a_1 + b_1, a_2 + b_2, a_3 + b_3)
\]

Addition satisfies the triangle law \( \overrightarrow{AB} + \overrightarrow{BC} = \overrightarrow{AC} \).

The dot product of vectors \( \mathbf{a} \) and \( \mathbf{b} \) is
\[
\mathbf{a} \cdot \mathbf{b} = a_1b_1 + a_2b_2 + a_3b_3
\]

The length or magnitude of \( \mathbf{a} \) is
\[
|\mathbf{a}| = \sqrt{\mathbf{a} \cdot \mathbf{a}}
\]

The angle \( \theta \) between \( \mathbf{a} \) and \( \mathbf{b} \) is given by \( \mathbf{a} \cdot \mathbf{b} = |\mathbf{a}| |\mathbf{b}| \cos \theta \).

The unit vector in the direction of \( \mathbf{a} \neq \mathbf{0} \) is
\[
\mathbf{\hat{a}} = \frac{1}{|\mathbf{a}|} \mathbf{a}
\]

The component of \( \mathbf{a} \) in the direction \( \mathbf{b} \neq \mathbf{0} \) and the projection of \( \mathbf{a} \) on \( \mathbf{b} \) are respectively
\[
\mathbf{a} \cdot \mathbf{\hat{b}} \quad \text{and} \quad \left(\mathbf{a} \cdot \mathbf{\hat{b}}\right) \mathbf{\hat{b}}
\]

1.5.2 Straight lines

(See section 1.2).

A parametric equation of the straight line \( \ell \) through \( \mathbf{a} \) with direction \( \mathbf{u} \neq \mathbf{0} \) is
\[
\mathbf{r} = \mathbf{a} + t\mathbf{u} \quad (t \in \mathbb{R})
\]

The foot \( Q \) of the perpendicular from \( P \) to \( \ell \) is found by solving
\[
(\mathbf{a} + tu - \mathbf{p}) \cdot \mathbf{u} = 0
\]
for \( t \) and then substituting this value back into \( \mathbf{a} + tu \).

Lines \( \ell_1 \) and \( \ell_1 \) are parallel if they have parallel directions and are skew if they are not parallel and do not intersect.

1.5.3 Planes

(See section 1.3).

A generic equation of the plane containing two non-parallel vectors \( \mathbf{u} \) and \( \mathbf{v} \) and passing through the point \( A \) is
\[
\mathbf{r} = \mathbf{a} + su + tv \quad (s, t \in \mathbb{R})
\]
A vector \( \mathbf{n} \neq 0 \) is normal to the plane if it is perpendicular to every line lying in the plane.

A Cartesian equation of the plane has the form

\[ (\mathbf{r} - \mathbf{a}) \cdot \mathbf{n} + c = (x_1 - a_1)n_1 + (x_2 - a_2)n_2 + (x_3 - a_3)n_3 + c = 0 \]

where \( c \) is a constant.

The foot \( Q \) of the perpendicular from a point \( P \) to the plane is found by solving

\[ \mathbf{n} \cdot (\mathbf{p} + t\mathbf{n}) + c = 0 \]

for \( t \) and then substituting this into the line \( \mathbf{p} + t\mathbf{n} \).

### 1.5.4 Projections and reflections

(See equation 1.7, subsection 1.2.5, example 1.3.8 and subsection 1.4).

When a line or plane goes through the origin the foot \( Q \) of the perpendicular from \( P \) to the line or plane is called the projection of \( P \) on the plane.

The reflection \( P' \) of a point \( P \) in a line or plane is given by

\[ \mathbf{p}' = 2\mathbf{Q} - \mathbf{p} \]

where \( Q \) is the foot of the perpendicular from \( P \) to the line or plane.
Chapter 2

Matrices and the Solution of Simultaneous Linear Equations

Linear algebra is used in almost every branch of pure and applied mathematics. It is indispensable for all branches of engineering, the physical sciences (including physics, chemistry and biology), statistics and for operations research (mathematics applied to industry and commerce).

Although most results in this course are valid for various number systems (e.g. rational real or complex numbers), it is assumed that all numbers and number-variables dealt with are real, i.e. range over the real number system $\mathbb{R}$. Numbers will also be referred to as scalars.

2.1 The Solution of Linear Equations

At school you learned how to solve up to three simultaneous linear equations for three unknowns, say $x_1$, $x_2$, $x_3$. In Chapter 1 we saw how such equations arise in analytic geometry and used vector notation such as a row $\mathbf{z} = (x_1, x_2, x_3)$ to represent the variables $x_1$, $x_2$ and $x_3$. In this chapter we will learn some systematic techniques for solving $m$ equations for $n$ unknowns $x_1, x_2, \ldots, x_n$, and we will see the need to carefully distinguish rows from columns. While geometry often serves as an inspiration for various algebraic results, it is not possible to draw an actual picture representing a system involving $n$ variables when $n \geq 4$.

2.1.1 Some examples

Example 2.1.1 Find all solutions that satisfy simultaneously

\begin{align*}
2x_1 + 3x_2 &= -1 \quad \text{(a)} \\
x_1 - 4x_2 &= 5 \quad \text{(b)}
\end{align*}

Solution:

1. One way: Make $x_1$ “the subject of the formula” in (b), i.e. $x_1 = 5 + 4x_2$ and substitute $5 + 4x_2$ for $x_1$ in (a) to give $2(5 + 4x_2) + 3x_2 = -1$, so $x_2 = -1$ and $x_1 = 1$.

2. We prefer to replace the above system by a sequence of equivalent systems (i.e. that have exactly the same solutions), known as Gauss reduction.

(a) (First equivalent system). Multiply (b) by 2 and leave (a) unchanged:

\begin{align*}
2x_1 + 3x_2 &= -1 \quad \text{leave unchanged} \quad \text{(a)} \\
2x_1 - 8x_2 &= 10 \quad 2 \times \text{(b)} \quad \text{(b)}
\end{align*}
(b) (Second equivalent system). In the new system subtract (a) from (b), to obtain:

\[
\begin{align*}
2x_1 + 3x_2 &= -1 \quad \text{leave unchanged} \quad (a) \\
-11x_2 &= 11 \quad (-1) \times (a) + (b) \quad (b)
\end{align*}
\]

It is now obvious from (b) that \( x_2 = -1 \). By back-substitution in (a) we get \( x_1 = 1 \). For reasons that will become clear later we write this unique solution as a column vector

\[
\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 1 \\ -1 \end{bmatrix}
\]

**Example 2.1.2** Consider

\[
\begin{align*}
x_1 + x_2 + 3x_3 &= -1 \quad (a) \\
x_1 + x_2 + 4x_3 &= -3 \quad (b) \\
-x_1 + x_2 - 2x_3 &= 1 \quad (c)
\end{align*}
\]

**Solution:**

\[
\begin{align*}
x_1 + x_2 + 3x_3 &= -1 \quad \text{leave unchanged} \quad (a) \\
x_3 &= -2 \quad (-1) \times (a) + (b) \quad (b) \\
-x_1 + x_2 - 2x_3 &= 1 \quad \text{leave unchanged} \quad (c)
\end{align*}
\]

Finally, interchange the equations (b) and (c):

\[
\begin{align*}
x_1 + x_2 + 3x_3 &= -1 \quad \text{leave unchanged} \quad (a) \\
2x_2 + x_3 &= 0 \quad (a) + (c) \quad (b) \\
x_3 &= -2 \quad \text{leave unchanged} \quad (c)
\end{align*}
\]

Because the coefficients of \( x_1 \) in (b) and (c) are zero and also because the coefficient of \( x_2 \) in (c) is zero, we say the system is in (row) **echelon** or (upper) **triangular** form. From (c), \( x_3 = -2 \). Back-substitution into (b) gives \( x_2 = 1 \). Finally, back-substitution of \( x_3 = -2 \) and \( x_2 = 1 \) into (a) gives \( x_1 = 4 \). The solution is

\[
\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 4 \\ 1 \\ -2 \end{bmatrix}
\]

**Example 2.1.3** Find all solutions \( \mathbf{x} \) to the system

\[
\begin{align*}
x_1 + x_2 + 3x_3 + 2x_4 &= -1 \quad (a) \\
2x_2 + x_3 + x_4 &= 0 \quad (b) \\
x_3 - x_4 &= -2 \quad (c)
\end{align*}
\]

**Solution:**

There are only **three** equations for **four** unknowns and the equations are already in echelon form. One of the unknowns (say \( x_4 \)) can be any number and back-substitution gives in turn \( x_3 = x_4 - 2, \quad x_2 = -\frac{1}{2}(x_3 + x_4) = 1 - x_4 \quad \text{and} \quad x_1 = -1 - x_1 - x_2 - 3x_3 - 2x_4 = 4 - 4x_4 \). It is customary to let \( x_4 = t \), so
2.2 MATRICES

\[ x = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 4 - 4t \\ 1 - t \\ t - 2 \\ t \end{bmatrix} \] (2.5)

is the general solution. The parameter \( t \in \mathbb{R} \) can be any real number.

Remark 22 Here we have 3 equations for 4 unknowns and the solution shows that we have an infinity of solutions. In Example 1.3.2 of Chapter 1, we found a vector \( \mathbf{n} \neq 0 \) normal to a plane. For \( \beta \neq 0 \), \( \beta \mathbf{n} \) is also a normal, so again we have an infinite number of solutions, this time of two equations in three unknowns. Example 1.3.4 of Chapter 1 involved solving one equation for three unknowns and this involved two parameters.

These examples and also Example 2.1.3 illustrate a general result:
If a system of equations has more unknowns than equations and there is a solution, then the system has infinitely many solutions. See, for example, 2.2.15 below. In Exercise 52, No.5 you will get a more complete picture of this theorem.

Example 2.1.4 For which values of the variables does the following system have a solution?

\[
\begin{align*}
&x_1 + x_2 + 3x_3 = -1 \quad \text{(a)} \\
&x_1 + 3x_2 + 4x_3 = -1 \quad \text{(b)} \\
&2x_1 + 4x_2 + 7x_3 = 5 \quad \text{(c)}
\end{align*}
\]

Solution:

\[
\begin{align*}
&x_1 + x_2 + 3x_3 = -1 \quad \text{leave unchanged} \quad \text{(a)} \\
&2x_2 + x_3 = 0 \quad [(-1) \times \text{(a)}] + \text{(b)} \quad \text{(b)} \\
&2x_2 + x_3 = 7 \quad [(-2) \times \text{(a)}] + \text{(c)} \quad \text{(c)}
\end{align*}
\]

The system cannot have a solution since (b) and (c) are contradictory. The system is inconsistent (or the equations are incompatible) and the solution set is empty.

2.2 Matrices

A rectangular array \( A \) of numbers with \( m \) rows and \( n \) columns is called an \( m \times n \) ("m by n") matrix. Matrices will form a new kind of algebra and it is customary to place brackets (we will use square ones) around such arrays.

2.2.1 Examples of matrices

\[
\begin{bmatrix}
5 & -1 & 13 & -1 \\
-1 & 1 & -12 & \frac{1}{3} \\
1 & 0 & 40 & -3
\end{bmatrix}
\] is a \( 3 \times 4 \) matrix,

\[
\begin{bmatrix}
-\frac{1}{2} & 21 & -3 \\
-\frac{1}{2} & 5 & 0 \\
0 & -2 & 7 \\
1 & -2 & 6
\end{bmatrix}
\] is a \( 4 \times 3 \) matrix.

\[
\begin{bmatrix}
2 & \frac{3}{7} & -1 \\
1 & -6 & \frac{2}{5}
\end{bmatrix}
\] is a \( 2 \times 3 \) matrix;

\[
\begin{bmatrix}
3 & 2 & \frac{1}{2} \\
-1 & 0 & 5 \\
7 & 8 & -2
\end{bmatrix}
\] is a \( 3 \times 3 \) matrix. The column \( x = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \) is a \( 3 \times 1 \) matrix, while the row \( y = [ y_1 \ y_2 \ y_3 \ y_4 ] \) is a \( 1 \times 4 \) matrix.

One speaks of the row entries of a column and the column entries of a row.
2.2.2 Standard notation for an $m \times n$ matrix $A$

Given that $A$ is an $m \times n$ matrix, the standard notation is to write $A = [a_{ij}]$.

Here $a_{ij}$ is the entry of $A$ in the $i^{th}$ row and $j^{th}$ column. One calls $a_{ij}$ the $(i, j)$ entry of $A$. The first index $i$ labels the rows and has the range $i = 1, \ldots, m$, while the second index $j$ labels the columns and has the range $j = 1, \ldots, n$.

For example, if $m = 4$ and $n = 5$, 

$$A = \begin{bmatrix}
a_{11} & a_{12} & a_{13} & a_{14} & a_{15} \\
a_{21} & a_{22} & a_{23} & a_{24} & a_{25} \\
a_{31} & a_{32} & a_{33} & a_{34} & a_{35} \\
a_{41} & a_{42} & a_{43} & a_{44} & a_{45}
\end{bmatrix} \; \text{(2.6)}$$

is a $4 \times 5$ matrix. One reads $a_{12}$ as “a one-two”, not as “a twelve”. Likewise, $a_{34}$ is “a three-four”, not “a thirty four”. For example, consider the $2 \times 3$ matrix

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

The $(1, 1)$ entry is $a_{11} = 2$. It is the first entry in column 1 and also the first entry in row 1. The $(2, 1)$ entry $a_{21} = 1$ is the second row entry of column 1 and the first entry in row 2. The numbers $a_{12} = 3$, $a_{22} = -4$, are the first and second entries of column 2 respectively. Row 2 has entries $a_{21} = 1$, $a_{22} = -4$, $a_{23} = 5$ respectively. The $(1, 3)$ entry of $A$ is $a_{13} = -1$ the $(2, 3)$ entry is $a_{23} = 5$.

It is important to keep in mind that $a_{ij}$ is in row $i$ and in column $j$ of $A$.

Sometimes we write $A = A_{m \times n}$ to indicate that $A$ is $m \times n$. Thus if $A$ has two rows and three columns, we may write $A = A_{2 \times 3}$.

**Exercise 23**

1. For each matrix $A$ in 2.2.1 write down

   (a) each of its entries in the form $a_{ij} = ...$

   (b) each of its rows and columns. What notation do you think would be suitable for denoting row $i$ of a matrix $A$? What notation do you think would be suitable for denoting column $j$ of a matrix $A$?

   In Chapter 3, 3.1.1, we will develop a systematic notation for these; here you are only asked to use your imagination.

2. Find a formula for the entries $a_{ij}$ in the following $4 \times 4$ matrix

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

3. The $n \times n$ Hilbert matrix $H_n = [h_{ij}]$ is defined by

   $$h_{ij} = \frac{1}{i+j-1} \; (i = 1, \ldots, n; \; j = 1, \ldots, n)$$

   Write down the matrices $H_1$, $H_2$, $H_3$, $H_4$.

2.2.3 Expressing $m$ equations in $n$ unknowns as a matrix equation $Ax = b$

A system of linear equations has a matrix of coefficients, or coefficient matrix. For example, the coefficient matrix of Example 2.1.4 is

$$\begin{bmatrix}1 & 1 & 3 \\
1 & 3 & 4 \\
2 & 4 & 7\end{bmatrix}$$
If we are given \( m \) simultaneous equations in \( n \) unknowns \( x_1, x_2, \ldots, x_n \), there is a standard way to express the equations as a single matrix equation.

In the case of Example 2.1.4 this expression is

\[
\begin{bmatrix}
1 & 1 & 3 \\
1 & 3 & 4 \\
2 & 4 & 7 \\
\end{bmatrix}
\begin{bmatrix}
x_1 \\
x_2 \\
x_3 \\
\end{bmatrix}
= 
\begin{bmatrix}
-1 \\
-1 \\
5 \\
\end{bmatrix}
\] (2.7)

The system of Example 2.1.2 (equations (2.1)) has for coefficient matrix

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

The standard corresponding matrix equation is

\[
\begin{bmatrix}
1 & 1 & 3 \\
1 & 1 & 4 \\
-1 & 1 & -2 \\
\end{bmatrix}
\begin{bmatrix}
x_1 \\
x_2 \\
x_3 \\
\end{bmatrix}
= 
\begin{bmatrix}
-1 \\
-3 \\
1 \\
\end{bmatrix}
\] (2.8)

What this means is the equality of two column matrices:

\[
\begin{bmatrix}
x_1 + x_2 + 3x_3 \\
x_1 + x_2 + 4x_3 \\
-x_1 + x_2 - 2x_3 \\
\end{bmatrix}
= 
\begin{bmatrix}
-1 \\
-3 \\
1 \\
\end{bmatrix}
\] (2.9)

The system of equations of Example 2.1.3 has the matrix expression

\[
\begin{bmatrix}
1 & 1 & 3 & 2 \\
0 & 2 & 1 & 1 \\
0 & 0 & 1 & -1 \\
\end{bmatrix}
\begin{bmatrix}
x_1 \\
x_2 \\
x_3 \\
x_4 \\
\end{bmatrix}
= 
\begin{bmatrix}
-1 \\
0 \\
-2 \\
\end{bmatrix}
\] (2.10)

Using the general \( 4 \times 5 \) matrix of (2.6), the following matrix equation expresses a general system of 4 equations in 5 unknowns \( x_1, x_2, \ldots, x_5 \):

\[
\begin{bmatrix}
a_{11} & a_{12} & a_{13} & a_{14} & a_{15} \\
a_{21} & a_{22} & a_{23} & a_{24} & a_{25} \\
a_{31} & a_{32} & a_{33} & a_{34} & a_{35} \\
a_{41} & a_{42} & a_{43} & a_{44} & a_{45} \\
\end{bmatrix}
\begin{bmatrix}
x_1 \\
x_2 \\
x_3 \\
x_4 \\
x_5 \\
\end{bmatrix}
= 
\begin{bmatrix}
b_1 \\
b_2 \\
b_3 \\
b_4 \\
\end{bmatrix}
\] (2.11)

Written out, we again have the equality of two column matrices:

\[
\begin{bmatrix}
a_{11}x_1 + a_{12}x_2 + a_{13}x_3 + a_{14}x_4 + a_{15}x_5 \\
a_{21}x_1 + a_{22}x_2 + a_{23}x_3 + a_{24}x_4 + a_{25}x_5 \\
a_{31}x_1 + a_{32}x_2 + a_{33}x_3 + a_{34}x_4 + a_{35}x_5 \\
a_{41}x_1 + a_{42}x_2 + a_{43}x_3 + a_{44}x_4 + a_{45}x_5 \\
\end{bmatrix}
= 
\begin{bmatrix}
b_1 \\
b_2 \\
b_3 \\
b_4 \\
\end{bmatrix}
\] (2.12)

In general, let \( A \) be the matrix of coefficients of \( m \) simultaneous equations in \( n \) unknowns \( x_1, x_2, \ldots, x_n \). If the right-hand side of equation \( i \) is \( b_i \), the standard way to express the equations in matrix form is

\[
Ax = b
\] (2.13)

Here the unknowns \( x_j \) are written as an \( n \times 1 \) column matrix (a **column vector**) \( x \) and \( b \) is the \( m \times 1 \) column of the \( b_i \).

Observe that for the matrix equation (2.13) to be meaningful the number of rows of \( A \) must be the same as the number of rows of \( b \) and the number of columns of \( A \) must equal the number of rows of \( x \).
2.2.4 The dot product generalized

Consider the equations of the left-hand side of Example 2.1.2 or of equation (2.9). Each entry is a dot product, e.g. the third entry is \((-1, 1, -2) \cdot (x_1, x_2, x_3) = -x_1 + x_2 - 2x_3\). In the next chapter we will emphasize this as the product of a row and a column:

\[
\begin{bmatrix}
-1 & 1 & -2
\end{bmatrix}
\begin{bmatrix}
x_1 \\
x_2 \\
x_3
\end{bmatrix} = -x_1 + x_2 - 2x_3
\]

Likewise, the left-hand sides of the first and second equations are respectively

\[
\begin{bmatrix}
1 & 1 & 3
\end{bmatrix}
\begin{bmatrix}
x_1 \\
x_2 \\
x_3
\end{bmatrix} = x_1 + x_2 + 3x_3
\]

\[
\begin{bmatrix}
1 & 1 & 4
\end{bmatrix}
\begin{bmatrix}
x_1 \\
x_2 \\
x_3
\end{bmatrix} = x_1 + x_2 + 4x_3
\]

Similarly, the left-hand side of the first equation of Example 2.1.3 is the product

\[
\begin{bmatrix}
1 & 1 & 3 & 2
\end{bmatrix}
\begin{bmatrix}
x_1 \\
x_2 \\
x_3 \\
x_4
\end{bmatrix} = x_1 + x_2 + 3x_3 + 2x_4
\]

The left-hand side of equation (2.10) in Example 2.1.3 is the column with three products

\[
\begin{bmatrix}
1 & 1 & 3 & 2 \\
0 & 2 & 1 & 1 \\
0 & 0 & 1 & -1
\end{bmatrix}
\begin{bmatrix}
x \\
x \\
x \\
x
\end{bmatrix} = \begin{bmatrix}
x_1 + x_2 + 3x_3 + 2x_4 \\
2x_2 + x_3 + x_4 \\
x_3 - x_4
\end{bmatrix}
\]

Remark 24 The last examples show that we are dealing with a product which is a generalization of the dot product of Chapter 1 (subsection 1.1.12). As remarked earlier, we will later develop special notations for row \(i\) and column \(j\) of \(A\). What is more, we will emphasize \(Ax\) as the product of the matrices \(A\) and \(x\). See 3.1.6.

2.2.5 Detached Coefficients

We can also express the above problems compactly by considering the array formed by attaching the right-hand side \(b\) to the coefficient matrix \(A\) in the form \([A, b]\). Corresponding to the problem of solving the matrix equation (2.13) we call \(R = [A, b]\) the array of detached coefficients or the matrix \(A\) augmented by \(b\). Operations done on the equations have their exact counterpart when done on the rows of \(R\).

1. Example 2.1.1 above can be expressed as the matrix equation

\[
\begin{bmatrix}
2 & 3 \\
1 & -4
\end{bmatrix}
\begin{bmatrix}
x_1 \\
x_2
\end{bmatrix} = \begin{bmatrix}
-1 \\
5
\end{bmatrix}
\]

or as the array of detached coefficients:

\[
\begin{bmatrix}
2 & 3 & -1 \\
1 & -4 & 5
\end{bmatrix}
\]

We have omitted brackets around the array. Here \(x_1\) and \(x_2\) can be thought of as labels of the first two columns of the array. The third column of (2.15) is the right-hand side of
(2.14). It is clear that two equations in two unknowns determine and are determined by such an array. Instead of performing operations on the given equations, we perform the same operations on this array.

In the case of Example 2.1.1 these steps are:

\[
\begin{pmatrix}
2 & 3 & -1 \\
2 & -8 & 10
\end{pmatrix}
\begin{align*}
\text{multiply Row 2 by 2} & : & 2 & 3 & -1 \\
\text{add } (-1) \times \text{Row 1 to Row 2} & : & 0 & -11 & 11
\end{align*}
\]

From the last array we read off \( -11x_2 = 11 \), so \( x_2 = -1 \). The first row reads \( 2x_1 + 3x_2 = -1 \) and \( x_1 = 1 \). This is the same solution we found before.

2. Example 2.1.2 has an expression as the matrix equation (2.8) and as an array of detached coefficients the equations are:

\[
\begin{pmatrix}
1 & 1 & 3 & -1 \\
1 & 1 & 4 & -3 \\
-1 & 1 & -2 & 1
\end{pmatrix}
\]

Here \( x_1, x_2, \) and \( x_3 \) label the first three columns. The steps leading to the solution are:

\[
\begin{align*}
1 & 1 & 3 & -1 \\
0 & 0 & 1 & -2 & -R_1 + R_2 \\
-1 & 1 & -2 & 1
\end{align*}
\]

\[
\begin{align*}
1 & 1 & 3 & -1 \\
0 & 0 & 1 & -2 \\
0 & 2 & 1 & 0 & R_1 + R_3
\end{align*}
\]

Finally, we interchange rows 2 and 3:

\[
\begin{align*}
1 & 1 & 3 & -1 \\
0 & 2 & 1 & 0 & R_2 \leftrightarrow R_3 \\
0 & 0 & 1 & -2
\end{align*}
\]

The equations are in echelon form and we can read off the solution:

\[
x_3 = -2, \ 2x_2 + x_3 = 0, \ \text{so } x_2 = 1 \ \text{and finally from } x_1 + x_2 + 3x_3 = -1 \ \text{we get } x_1 = 4. \ \text{We have}
\]

\[
\begin{bmatrix}
x_1 \\
x_2 \\
x_3
\end{bmatrix}
= \begin{bmatrix} 4 \\ 1 \\ -2 \end{bmatrix}
\]

This is equation (2.3), found earlier.

**Observation:** It is important to observe that the right-hand side of the original set of equations is a linear combination of the columns of the coefficient matrix with coefficients \( x_1, x_2 \) and \( x_3 \):

\[
\begin{align*}
\begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix} & 4 + \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} + \begin{bmatrix} 3 \\ 4 \\ -2 \end{bmatrix} (-2) = \begin{bmatrix} -1 \\ -3 \\ 1 \end{bmatrix}
\end{align*}
\]

Linear combinations were introduced for row vectors in Chapter 1, subsection 1.1.18 and will be systematically developed in Chapter 3. See subsection 3.1.2 and especially 3.1.9.

Note that it is natural to write the coefficients on the right of the column vectors since \( \mathbf{z} \) appears on the right of \( \mathbf{A} \) when we express the equations in the form \( \mathbf{A} \mathbf{x} = \mathbf{b} \).
3. Example 2.1.3 (equations (2.4)) has the matrix expression (2.10). The corresponding array of detached coefficients is already in echelon form:

\[
\begin{pmatrix}
1 & 1 & 3 & 2 & -1 \\
0 & 2 & 1 & 1 & 0 \\
0 & 0 & 1 & -1 & -2
\end{pmatrix}
\] (2.18)

The general solution can be read off as before by putting \(x_4 = t\) and solving in turn for \(x_3\), \(x_2\) and \(x_1\):

\[
\begin{pmatrix}
x_1 \\
x_2 \\
x_3 \\
x_4
\end{pmatrix} = \begin{pmatrix} 4 - 4t \\ 1 - t \\ t - 2 \\ t \end{pmatrix}
\]

This is just the solution (2.5) found earlier. A particular solution is found by giving the parameter \(t\) some value, eg. \(t = 3\). In that case \(x_1 = -8\), \(x_2 = -2\), \(x_3 = 1\) and \(x_4 = 3\).

As a linear combination we have

\[
\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} (-8) + \begin{pmatrix} 1 \\ 2 \\ 0 \end{pmatrix} (-2) + \begin{pmatrix} 3 \\ 1 \\ 1 \end{pmatrix} 3 = \begin{pmatrix} -1 \\ 0 \\ -2 \end{pmatrix}
\] (2.19)

4. Example 2.1.4 has two expressions, one as equation (2.7), and also as the array of detached coefficients (augmented matrix)

\[
\begin{pmatrix}
1 & 1 & 3 & -1 \\
1 & 3 & 4 & -1 \\
2 & 4 & 7 & 5
\end{pmatrix}
\]

The steps leading to the solution are:

\[
\begin{array}{c}
1 & 1 & 3 & -1 \\
0 & 2 & 1 & 0 & -R_1 + R_2 \\
0 & 2 & 1 & 7 & -2R_1 + R_3
\end{array}
\]

Taking it a step further, we obtain the echelon (triangular) form:

\[
\begin{array}{c}
1 & 1 & 3 & -1 \\
0 & 2 & 1 & 0 \\
0 & 0 & 0 & 7 & -R_2 + R_3
\end{array}
\] (2.20)

The last row reads \(0x_1 + 0x_2 + 0x_3 = 7\), an impossibility. The given equations are inconsistent (incompatible); they do not have a solution.

5. As a detached array, the equations (2.11) appear as

\[
\begin{pmatrix}
a_{11} & a_{12} & a_{13} & a_{14} & a_{15} & b_1 \\
a_{21} & a_{22} & a_{23} & a_{24} & a_{25} & b_2 \\
a_{31} & a_{32} & a_{33} & a_{34} & a_{35} & b_3 \\
a_{41} & a_{42} & a_{43} & a_{44} & a_{45} & b_4
\end{pmatrix}
\] (2.21)

**Remark 25** Our convention is that an operation performed on row \(i\) is written next to row \(i\) after the operation has been done.
2.2.6 Permissible operations on equations: Elementary row operations. Equivalent row operations on the augmented matrix 

\[ R = [A, \mathbf{b}] \]

The row operations which do not alter the solution set of a system \( A\mathbf{x} = \mathbf{b} \) of \( m \) simultaneous linear equations for \( n \) unknowns are:

1. Interchange equations \( i \) and \( j \) (interchange rows \( R_i \) and \( R_j \) of the corresponding array of detached coefficients).
2. Multiply equation \( i \) (multiply row \( R_i \) of the corresponding array) by a scalar \( \beta \neq 0 \).
3. For any scalar \( \gamma \) and \( i \neq j \), add \( \gamma \) times equation \( i \) to equation \( j \) (add \( \gamma \) times row \( R_i \) of the corresponding array to row \( R_j \)).

The above three operations done on an array are called elementary row operations (abbreviated as eros). Note that forming \( \beta R_i \) is just like multiplying a vector in 3-space by a scalar, except that now the vector has \( n + 1 \) entries. The type (3) operation replaces \( R_j \) with \( \gamma R_i + R_j \), where addition of rows is addition of vectors.

What is obvious (and assumed at school) is that if \( \mathbf{x} \) is a simultaneous solution of the \( m \) given equations before any one of the above eros is performed, then \( \mathbf{x} \) remains a solution after the operation is performed.

To see that the reverse is true, suppose we have just performed the elementary row operation of type (3) and that \( \mathbf{x} \) is a solution to the new system of equations. Adding \(-\gamma\) times equation \( i \) \((-\gamma \) times row \( R_i \) of the augmented array) to equation \( j \) \((row R_j \) of the array) in the new system of equations (array) brings us back to the system before the type (3) operation was performed. This shows that \( \mathbf{x} \) is also a solution to the old system.

For example, consider example 2.1.2 again and the first step toward a solution, as in 2:

\[
\begin{pmatrix}
1 & 1 & 3 & -1 \\
1 & 1 & 4 & -3 \\
-1 & 1 & -2 & 1
\end{pmatrix}
\]

Adding row 1 to row 2 of this array restores the original array:

\[
\begin{pmatrix}
1 & 1 & 3 & -1 \\
0 & 0 & 1 & -2 \\
-1 & 1 & -2 & 1
\end{pmatrix}
\]

The following more general setting refers to the array (2.21). Suppose we add \( \gamma \) times Row 2 to Row 3, obtaining

\[
\begin{pmatrix}
a_{11} & a_{12} & a_{13} & a_{14} & a_{15} & b_1 \\
a_{21} & a_{22} & a_{23} & a_{24} & a_{25} & b_2 \\
\gamma a_{21} + a_{31} & \gamma a_{22} + a_{32} & \gamma a_{23} + a_{33} & \gamma a_{24} + a_{34} & \gamma a_{25} + a_{35} & \gamma b_2 + b_3 \\
a_{41} & a_{42} & a_{43} & a_{44} & a_{45} & b_4 \\
\end{pmatrix}
\]

Then,

\[
\begin{pmatrix}
a_{11} & a_{12} & a_{13} & a_{14} & a_{15} & b_1 \\
a_{21} & a_{22} & a_{23} & a_{24} & a_{25} & b_2 \\
a_{31} & a_{32} & a_{33} & a_{34} & a_{35} & b_3 \\
a_{41} & a_{42} & a_{43} & a_{44} & a_{45} & b_4 \\
\end{pmatrix}
\]

brings us back to the original array.
Remark 26 For convenience we have been using $R_i$ for row $i$ of an array $R$. In the next chapter we will use a slightly different notation for row $i$ of a matrix.

### 2.2.7 Arrays in row echelon form (row EF): Proper Definition

An array is in row echelon form if

1. All zero rows come after the non-zero rows.
2. For non-zero rows, the first non-zero entry in row $i+1$ comes later than the first non-zero entry in row $i$.

**Remark 27** Row echelon forms are also called upper triangular. In subsection 3.4.14 of Chapter 3 we will also discuss elementary column operations and column echelon forms. There, the distinction between the row and column forms will matter.

### 2.2.8 How an array gets reduced to row echelon form

At the start, and at any later stage, we arrange the rows so that

a. all zero rows come after non-zero rows and

b. for the non-zero rows, the first non-zero entry in row $i+1$ does not come earlier than the first non-zero entry in row $i$.

Suppose that only the first $r$ rows are non-zero and in row echelon form. The first non-zero entry in row $r$ is called the **pivot element** and row $r$ is the **pivot** or **reference** row. We assume that (a) and (b) hold. If the remaining rows are zero we are finished. Otherwise, multiples of row $r$ are added to rows below it in such a way that all entries below the pivot element become zeros. The first $r+1$ rows are then in echelon form and this remains the case after arranging rows $r+1$, $r+2$,... so that (a) and (b) hold. The process ends when the whole array is in row echelon form and then each non-zero row has a pivot element.

The columns of the array containing a pivot element are called **pivot columns**. (This will be used in 2.2.15 below and also in subsection 3.4.9).

### 2.2.9 Further examples

**Example 2.2.1** Reduce to row echelon form

\[
\begin{bmatrix}
-3 & -10 & 2 \\
7 & 5 & 6
\end{bmatrix}
\]

**Solution (method 1)**

1. $\frac{10}{3} - \frac{2}{3} = \left(-\frac{1}{3}R_1\right)$

\[
\begin{bmatrix}
1 & \frac{10}{3} & -\frac{2}{3} \\
7 & 5 & 6
\end{bmatrix}
\]

1. $\frac{10}{3} - \frac{2}{3} = \text{ref row, pivot1}$

\[
\begin{bmatrix}
1 & \frac{10}{3} & -\frac{2}{3} \\
0 & \frac{35}{3} & \frac{22}{3}
\end{bmatrix}
\]

\[\left(-7R_1 + R_2\right)
\]

**Remark 28** This illustrates how the pivot element can always be made 1.
Method 2 (Fraction-free solution)

\[
\begin{bmatrix}
-21 & -70 & 14 & 7R_1 \\
21 & 15 & 18 & 3R_2 \\
0 & -55 & 32 & R_1 + R_2 \\
\end{bmatrix}
\]

Remark 29 This example illustrates how it is possible to obtain a fraction-free answer when all entries are integers or even fractions.

Example 2.2.2 Reduce the array of detached coefficients to row echelon form and hence solve

\[
\begin{bmatrix}
1 & 0 & 2 \\
3 & 0 & 5 \\
-5 & 1 & -10 \\
\end{bmatrix}
\]

\[
x = \begin{bmatrix}
-1 \\
1 \\
2 \\
\end{bmatrix}
\]

Solution:
The augmented matrix (without brackets) is

\[
\begin{bmatrix}
1 & 0 & 2 & -1 \\
3 & 0 & 5 & 1 \\
-5 & 1 & -10 & 2 \\
\end{bmatrix}
\]

Elementary row operations (eros) bringing the detached array into echelon form are:

1. \[1 0 2 -1 \leftrightarrow R_1 \]
2. \[0 0 -1 4 \rightarrow -3R_1 + R_2 \]
3. \[0 1 0 -3 \rightarrow 5R_1 + R_3 \]
4. \[0 0 -1 4 \leftrightarrow R_3 \]

From this row echelon form we find the unique solution

\[
x = \begin{bmatrix}
7 \\
-3 \\
-4 \\
\end{bmatrix}
\]

Example 2.2.3 Triangulate and hence solve

\[
\begin{bmatrix}
3 & -3 & 1 & 10 \\
2 & -3 & -1 & 3 \\
-1 & 2 & 0 & 3 \\
1 & -1 & 1 & 2 \\
\end{bmatrix}
\]

\[
\begin{bmatrix}
x_1 \\
x_2 \\
x_3 \\
x_4 \\
\end{bmatrix} = \begin{bmatrix}
2 \\
0 \\
-6 \\
4 \\
\end{bmatrix}
\]

(2.22)

Solution:
The initial array of detached coefficients is

\[
\begin{bmatrix}
3 & -3 & 1 & 10 & 2 \\
2 & -3 & -1 & 3 & 0 \\
-1 & 2 & 0 & 3 & -6 \\
1 & -1 & 1 & 2 & 4 \\
\end{bmatrix}
\]

The steps giving equivalent systems are:

Step 1:

\[
\begin{bmatrix}
1 & -1 & 1 & 2 & 4 \\
2 & -3 & -1 & 3 & 0 \\
-1 & 2 & 0 & 3 & -6 \\
3 & -3 & 1 & 10 & 2 \\
\end{bmatrix} \leftrightarrow R_1 \leftrightarrow R_4 \text{ interchange row 1 and row 4} \text{ ref row}
\]
We have interchanged rows 1 and 4 as it is convenient to have a 1 as pivot element.

Step 2

\[
\begin{bmatrix}
1 & -1 & 1 & 2 & 4 \\
0 & -1 & -3 & -1 & -8 \\
0 & 1 & 1 & 5 & -2 \\
0 & 0 & -2 & 4 & 10
\end{bmatrix}
\]

\[-2R_1 + R_2 \text{ new ref row} \]
\[R_1 + R_3 \]
\[(-3)R_1 + R_4 \]

At this point the first two rows are in row echelon form (row EF).

Step 3. Use row 2 for reference row to obtain

\[
\begin{bmatrix}
1 & -1 & 1 & 2 & 4 \\
0 & -1 & -3 & -1 & -8 \\
0 & 0 & -2 & 4 & 10 \\
0 & 0 & 0 & 0 & 0
\end{bmatrix}
\]

Step 4. Row 3 is the new reference row:

\[
\begin{bmatrix}
1 & -1 & 1 & 2 & 4 \\
0 & -1 & -3 & -1 & -8 \\
0 & 0 & -2 & 4 & 10 \\
0 & 0 & 0 & 0 & 0
\end{bmatrix}
\]

\[-R_3 + R_4 \]

This is in echelon form but we can simplify still further:

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

\[-R_2 \]
\[\frac{1}{2}R_3 \]

\[(2.23)\]

The required general solution to Example 2.2.3 is with \(x_4 = t\) the parameter, \(x_3 = 2t + 5\), \(x_2 = -7t - 7\) and \(x_1 = -11t - 8\). Or,

\[
\begin{bmatrix}
x_1 \\
x_2 \\
x_3 \\
x_4
\end{bmatrix} =
\begin{bmatrix}
-11t - 8 \\
-7t - 7 \\
2t + 5 \\
t
\end{bmatrix}
\]

\[(2.24)\]

As a linear combination of the columns of the coefficient matrix,

\[
\begin{bmatrix}
3 \\
2 \\
1
\end{bmatrix}
\begin{bmatrix}
-11t - 8 \\
-7t - 7 \\
2t + 5
\end{bmatrix}
\]

\[
\begin{bmatrix}
\frac{3}{2} \\
\frac{2}{-1} \\
\frac{1}{0}
\end{bmatrix} (-11t - 8) +
\begin{bmatrix}
-3 \\
-3 \\
-1
\end{bmatrix} (-7t - 7) +
\begin{bmatrix}
1 \\
0 \\
1
\end{bmatrix} (2t + 5) +
\begin{bmatrix}
10 \\
3 \\
2
\end{bmatrix} t =
\begin{bmatrix}
2 \\
0 \\
-6
\end{bmatrix}
\]

\[(2.25)\]

Remark 30 Each column vector has 4 entries instead of 3 and so we cannot visualize these vectors in space. Nevertheless, this should not be a problem. In Chapter 3 we will consider \(n\)-dimensional vectors. See subsections 3.1.2 3.1.9.

Remark 31 Although we associate arrays with simultaneous equations, it is important to realize that such operations can be done on any arrays, i.e. matrices, without any reference to solving equations in the usual sense. Note however the following:

Reducing a matrix \(A\) to row echelon form amounts to reducing the augmented matrix \([A, \mathbf{0}]\) to row echelon form. This in turn amounts to solving \(A\mathbf{x} = \mathbf{0}\). See Remark 34.

Example 2.2.4 Reduce the following matrix to row echelon form:
2.2. MATRICES

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

Solution:

In order for (b) in 2.2.8 to hold, first interchange rows 1 and 4 of \( A \):

\[
\begin{align*}
2 & \quad -3 & \quad -4 & \quad 3 & \quad -3 & \quad \text{ref row} \\
0 & 1 & 2 & -6 & 1 & \quad R_1 \leftrightarrow R_4 \\
0 & 0 & 3 & -2 & -2 & \quad \text{new ref row} \\
0 & 0 & -3 & -3 & -1 & 
\end{align*}
\]

Next:

\[
\begin{align*}
2 & \quad -3 & \quad -4 & \quad 3 & \quad -3 & \quad \text{new ref row} \\
0 & 1 & 2 & -6 & 1 & \quad -R_1 + R_2 \\
0 & 0 & 3 & -2 & -2 & \\
0 & 0 & -3 & -3 & -1 & 
\end{align*}
\]

Finally,

\[
\begin{align*}
2 & \quad -3 & \quad -4 & \quad 3 & \quad -3 & \quad \text{ref row} \\
0 & 1 & 2 & -6 & 1 & \quad R_3 + R_4 \\
0 & 0 & 3 & -2 & -2 & \\
0 & 0 & 0 & -5 & -3 & (2.26) 
\end{align*}
\]

The matrix is now in row echelon form.

Exercise 32 1. Express the following simultaneous equations in the form \( Ax = b \) and use Gauss-reduction and detached coefficients to find solutions. Display the final array in row echelon form. Also, express \( b \) as a linear combination of the columns of \( A \) whenever this is possible.

\[(a) \quad i. \quad \begin{align*}
x_1 & \quad -2x_2 = -1 \\
3x_1 & \quad -x_2 = 1 
\end{align*} \]

\[(a) \quad ii. \quad \begin{align*}
x & \quad -2y = -3 \\
2x & \quad -4y = -1 
\end{align*} \]

\[(a) \quad iii. \quad \begin{align*}
x & \quad -2y = -3 \\
2x & \quad -4y = -6 
\end{align*} \]

\[(a) \quad iv. \quad \begin{align*}
x_1 & \quad -2x_2 = -3 \\
3x_1 & \quad -x_2 = 1 \\
4x_1 & \quad -3x_2 = 5 
\end{align*} \]

\[(a) \quad v. \quad \begin{align*}
x & \quad -2y = -3 \\
2x & \quad +y = 4 \\
3x & \quad -y = 1 
\end{align*} \]

\[(a) \quad vi. \quad 0x + 0y = 0 \]
vii. \[3x + 0y = -3\]

(b) i. 
\[
\begin{align*}
x_1 - 2x_2 - x_3 &= 0 \\
3x_2 + x_3 &= 1
\end{align*}
\]

ii. 
\[
\begin{align*}
x_1 + x_2 + x_3 - x_4 &= 1 \\
2x_2 - x_3 + 4x_4 &= -1 \\
x_4 &= -2
\end{align*}
\]

iii. 
\[
\begin{align*}
x_1 + 2x_2 + x_3 &= -1 \\
2x_2 - x_3 + 4x_4 &= -1
\end{align*}
\]

iv. 
\[
\begin{align*}
x_1 + 2x_2 + x_3 - 5x_4 &= 0
\end{align*}
\]

2. (a) In solving simultaneous equations, if we drop an equation the resulting solution set contains and is usually larger than the original solution set. Illustrate this with an example.

(b) Rule 3 of permissible row operations in 2.2.6 says: For any scalar \( \gamma \) and \( i \neq j \), add \( \gamma \) times row \( R_i \) of the array to row \( R_j \). Why must we have \( i \neq j \)?

(c) In an array with rows \( R_1, \ldots, R_m \), let \( i \neq j \). Suppose that at least one of the numbers \( \alpha, \beta \) is not zero and form the row \( \alpha R_i + \beta R_j \). Under what conditions can this new row replace row \( i \) or row \( j \)? Describe what is going on in terms of the strict rules of 2.2.6.

3. * (Simplified Elementary Row Operations). Consider the following simplified set of two elementary row operations that can be applied to an array. Show that each row operation in 2.2.6 can be obtained by applying a sequence of these two operations. In other words, the two row operations are equivalent to those in 2.2.6.

(a) Multiply row \( R_i \) of the array by a scalar \( \beta \neq 0 \).

(b) For \( i \neq j \) add row \( R_i \) to row \( R_j \).

Hint 33 In order to add \( \alpha R_i \) to \( R_j \) when \( i \neq j \), we may obviously assume \( \alpha \neq 0 \). Perform (a) with \( \beta = \alpha \) then do (b), i.e. add the resulting row \( i \) to row \( j \). What must be done next?

It still remains to show that we can interchange rows \( R_i \) and \( R_j \) of the array using operations like (a) and (b). Try adding row \( i \) to row \( j \), then changing the sign of row \( i \).

### 2.2.10 Homogeneous Equations

A system \( Ax = b \) of \( m \) equations in \( n \) unknowns is called homogeneous if in the right-hand side all \( b_i = 0 \). Such a system always has at least one solution, namely \( x = 0 \), the column vector with all \( n \) entries \( x_j = 0 \). This is known as the trivial solution. The existence of non-trivial solutions is the subject of subsection 2.2.15.

Every system defines a corresponding homogeneous system.

**Example 2.2.5** Consider the following homogeneous system of equations:

\[
\begin{bmatrix}
3 & -3 & 1 & 10 \\
2 & -3 & -1 & 3 \\
-1 & 2 & 0 & 3 \\
1 & -1 & 1 & 2
\end{bmatrix}
\begin{bmatrix}
x_1 \\
x_2 \\
x_3 \\
x_4
\end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}
\]  

(2.27)
This is the homogeneous system corresponding to Example 2.2.3. We found the row echelon form of the detached coefficients for that example as (2.23). The row EF for the homogeneous system (2.27) is therefore

\[
\begin{array}{cccccc}
1 & -1 & 1 & 2 & 0 \\
0 & 1 & 3 & 1 & 0 \\
0 & 0 & 1 & -2 & 0 \\
0 & 0 & 0 & 0 & 0 \\
\end{array}
\]

The general solution of (2.27) is with \( x_4 = t \in \mathbb{R} \) any real number,

\[
\begin{bmatrix}
x_1 \\
x_2 \\
x_3 \\
x_4 \\
\end{bmatrix} =
\begin{bmatrix}
-11t \\
-7t \\
2t \\
t \\
\end{bmatrix}
\]

(2.28)

**Remark 34** When solving a homogeneous system \( Ax = 0 \), the column \( 0 \) can simply be ignored in the array of detached coefficients.

**Exercise 35**

1. Find a row echelon form of the following matrices:

   (a) \( A = \begin{bmatrix}
1 & 2 & -3 & -1 \\
-3 & 4 & 1 & 2 \\
-2 & 6 & -2 & 1 \\
4 & -2 & -4 & -3 \\
\end{bmatrix} \)

   (b) \( B = \begin{bmatrix}
1 & -3 & -2 & 4 \\
2 & 4 & 6 & -2 \\
-3 & 1 & -2 & -4 \\
-1 & 2 & 1 & -3 \\
\end{bmatrix} \)

   (c) How are the matrices \( A \) and \( B \) related?

2. Let:

\[
A = \begin{bmatrix}
1 & 1 & 2 & 1 & -3 \\
2 & 1 & 5 & 2 & -7 \\
0 & -1 & 3 & 5 & -4 \\
-2 & 1 & -3 & 9 & 3 \\
\end{bmatrix}
\]

   (a) Reduce \( A \) to echelon form.

   (b) Solve the homogeneous system \( Ax = 0 \). Express \( 0 \) as a linear combination of the columns of \( A \) in a non-trivial way.

**Hint 36** For coefficients take the values of \( x_i \) in a homogeneous solution that are not all zero. (The trivial linear combination is the one with all coefficients \( 0 \).)

The columns of the above \( A \) are said to be **linearly dependent**.

Linear independence and dependence will be studied formally in Chapter 3, section 3.4. See also the theorem 2.2.15 below.

3. (a) Reduce the array of detached coefficients to echelon form and solve

\[
\begin{bmatrix}
1 & -1 & 1 & 2 \\
2 & -2 & -1 & 3 \\
-1 & 1 & 0 & 3 \\
1 & -2 & 2 & 5 \\
\end{bmatrix} x = \begin{bmatrix}
4 \\
0 \\
-6 \\
5 \\
\end{bmatrix}
\]
CHAPTER 2. MATRICES AND THE SOLUTION OF SIMULTANEOUS LINEAR EQUATIONS

(b) Show that whatever the RHS \( \mathbf{b} \) may be, the system always has a unique solution.

4. (a) Solve for \( \mathbf{x} \):
\[
\begin{bmatrix}
1 & 1 & 2 & 1 \\
2 & 1 & 5 & 2 \\
0 & -1 & 3 & 5 \\
3 & 1 & 10 & 8
\end{bmatrix} \mathbf{x} = \begin{bmatrix}
-3 \\
-7 \\
-4 \\
-14
\end{bmatrix}
\]

(b) Solve the corresponding homogeneous system for \( \mathbf{x} \):
\[
\begin{bmatrix}
1 & 1 & 2 & 1 \\
2 & 1 & 5 & 2 \\
0 & -1 & 3 & 5 \\
3 & 1 & 10 & 8
\end{bmatrix} \mathbf{x} = \begin{bmatrix}
0 \\
0 \\
0 \\
0
\end{bmatrix}
\]

5. (a) Solve for \( \mathbf{x} \):
\[
\begin{bmatrix}
1 & -1 & 1 & 2 \\
2 & -2 & -1 & 3
\end{bmatrix} \mathbf{x} = \begin{bmatrix}
4 \\
0
\end{bmatrix}
\]

(This consists of the first two equations of No.(3).

(b) Solve the homogeneous system for \( \mathbf{x} \).

6. (a) What are the solutions \( \mathbf{x} \) to
\[
\begin{bmatrix}
3 & 1 & 1 & 4 \\
2 & 5 & 1 & 7 \\
1 & 9 & 1 & 10
\end{bmatrix} \mathbf{x} = \begin{bmatrix}
-1 \\
-1 \\
2
\end{bmatrix}
\]

(b) Solve the corresponding homogeneous system.

7. If you have one solution to 4a and the general solution of 4b, show that you can write down the general solution of 4a.

8. The same for 5a and 5b.

9. Solve the following system for \( \mathbf{x} \) and relate your solution to that of its homogeneous counterpart.
\[
\begin{bmatrix}
2 & 1 & 3 & 5 \\
-2 & -1 & -1 & -2 \\
4 & 2 & 7 & 11 \\
6 & 3 & 10 & 16
\end{bmatrix} \mathbf{x} = \begin{bmatrix}
-1 \\
0 \\
-3 \\
-4
\end{bmatrix}
\]

10. Can you state a general theorem of which Nos.(7), (8) and (9) are special cases?
Note that one solution to Example 2.2.3 (matrix equation (2.22)) plus the general solution (2.28) to the homogeneous Example 2.2.5 gives the general solution (2.24) to Example 2.2.3. We will return to this in the next chapter. (See Exercise 52, No.5).

11. (a) Solve the system for \( x, y, z \):
\[
\begin{bmatrix}
1 & 1 & -1 \\
-2 & 3 & 4 \\
-1 & 4 & 1 \\
4 & -1 & -2
\end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix}
-3 \\
7 \\
-2 \\
2
\end{bmatrix}
\]

(b) Find the homogeneous solution to (a).
12. Reduce to row EF

\[
\begin{bmatrix}
2 & 1 & 3 & 5 \\
-2 & -1 & -1 & -2 \\
4 & 2 & 7 & 11 \\
6 & 3 & 10 & 16 \\
\end{bmatrix}
\]

13. * (Allenby, p.30) For any \(a, b, c\) and \(d\) discuss the solution \(x\) of

\[
\begin{bmatrix}
1 & 2 & -1 & 3 & -1 \\
3 & 4 & 2 & 5 & -2 \\
1 & 0 & 4 & -1 & 0 \\
3 & 2 & 7 & 1 & -1 \\
\end{bmatrix} x =
\begin{bmatrix}
a \\
b \\
c \\
d \\
\end{bmatrix}
\]

Find also the corresponding homogeneous solution.

14. For which scalars \(\alpha\) and \(\beta\) is the system

\[
\begin{bmatrix}
1 & 2 & -3 \\
3 & -1 & 2 \\
1 & -5 & 8 \\
\end{bmatrix} x =
\begin{bmatrix}
\alpha \\
\beta \\
-\alpha \\
\end{bmatrix}
\]

consistent? (recall that a system of equations is consistent (or compatible) if it has at least one solution). If so, find the general solution \(x\).

15. Show that the system

\[
\begin{bmatrix}
1 & 1 & \lambda \\
3 & 4 & 2 \\
2 & 3 & -1 \\
\end{bmatrix} x =
\begin{bmatrix}
2 \\
\lambda \\
1 \\
\end{bmatrix}
\]

is consistent for any scalar \(\lambda\). Find those values of \(\lambda\) for which

i. the system has a unique solution \(x\) and

ii. more than one solution \(x\).

16. Let \(Ax = b\) represent a system of \(m\) equations for \(n\) unknowns. Which of the following is true? Give reasons.

(a) The system always has a solution if \(m < n\).

(b) The system always has a solution if \(m = n\).

(c) The system never has a solution if \(m > n\) (more equations than unknowns).

(d) The system always has a solution if it is homogeneous.

2.2.11 Row Reduced Echelon Form (row REF)

A matrix \(A = [a_{ij}]\) is said to be in row reduced echelon form (row REF) if

1. It is in row echelon form.

2. If \(a_{ij}\) (the pivot entry) is the first non-zero entry in row \(i\), then

   (a) \(a_{ij} = 1\) and

   (b) all other entries in column \(j\) are zeros.

2.2.12 Some examples in row reduced echelon form

\[
\begin{bmatrix}
1 & -3 & 0 & 0 \\
0 & 0 & 1 & 0 \\
0 & 0 & 0 & 1 \\
\end{bmatrix},
\begin{bmatrix}
0 & 1 & 0 & 5 \\
0 & 0 & 1 & 6 \\
0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 \\
\end{bmatrix},
\begin{bmatrix}
1 & 3 & -5 & 0 & 0 & 7 \\
0 & 0 & 0 & 1 & 0 & -2 \\
0 & 0 & 0 & 0 & 0 & 1 \\
3 & 2 & 7 & 1 & -1 \\
\end{bmatrix}
\]
2.2.13 Any matrix can be brought into row reduced echelon form by suitable elementary row operations

Let \( A \) be in row echelon form. Suppose that \( A \) has \( r \) non-zero rows. We first divide each non-zero row by its leading (first non-zero) entry. So we may suppose that the leading entry of row \( r \) is \( a_{rj} = 1 \). Use row \( r \) for reference row and \( a_{rj} \) as a pivot element to reduce all entries above it to zeros. Now repeat the process with row \( r - 1 \) etc., the last reference row being row 2. The resulting array is in row reduced echelon form (row REF).

Note that in converting a row EF to a row REF all the pivot elements remain in the same positions but become 1s.

The process is known as the **Gauss-Jordan** method.

**Remark 37** Although a matrix does not in general have a unique row echelon form (which ones do?), it can be shown that the row reduced echelon form is unique.

2.2.14 Some examples of Gauss-Jordan reduction

**Example 2.2.6** For our first example let us reduce the array of Example 2.2.1 to row reduced echelon form. We found the row EF

\[
\begin{array}{ccc}
1 & 10 & -2 \\
0 & -\frac{5}{3} & \frac{32}{3}
\end{array}
\]

Step 1:

\[
\begin{array}{ccc}
1 & \frac{10}{3} & -\frac{2}{3} \\
0 & 1 & \frac{\frac{3}{2}}{5}
\end{array}
\] \( \text{ref row} \)

Step 2:

\[
\begin{array}{ccc}
1 & 0 & \frac{14}{11} - \frac{10}{3} R_2 + R_1 \\
0 & 1 & \frac{32}{35}
\end{array}
\]

**Example 2.2.7** As a second example consider Example 2.1.2. Its row echelon form was found in (2.16):

\[
\begin{array}{ccc}
1 & 1 & 3 & -1 \\
0 & 2 & 1 & 0 \\
0 & 0 & 1 & -2
\end{array}
\]

The following steps bring it into row REF.

Step 1:

\[
\begin{array}{ccc}
1 & 1 & 3 & -1 \\
0 & 1 & \frac{1}{2} & 0 & \frac{1}{2} R_2 \\
0 & 0 & 1 & -2 & \text{ref row, pivot } a_{33} = 1
\end{array}
\]

Step 2:

\[
\begin{array}{ccc}
1 & 1 & 0 & 5 & -3 R_3 + R_1 \\
0 & 1 & 0 & 1 & -\frac{1}{2} R_3 + R_2 \text{ new ref row, pivot } a_{22} = 1 \\
0 & 0 & 1 & -2
\end{array}
\]

Step 3

\[
\begin{array}{ccc}
1 & 0 & 0 & 4 & -R_2 + R_1 \\
0 & 1 & 0 & 1 \\
0 & 0 & 1 & -2
\end{array}
\]

We read off the same solution (2.3) as before, only that it is easier to do.
2.2. MATRICES

**Example 2.2.8** Consider Example 2.1.3 again. From (2.18) we already have the row EF of detached coefficients:

\[
\begin{bmatrix}
1 & 1 & 3 & 2 & -1 \\
0 & 2 & 1 & 1 & 0 \\
0 & 0 & 1 & -1 & -2
\end{bmatrix}
\]

The following steps bring it into row REF.

**Step 1**

\[
\begin{bmatrix}
1 & 1 & 3 & 2 & -1 \\
0 & 1 & \frac{1}{2} & 1 & 0 \\
0 & 0 & 1 & -1 & -2
\end{bmatrix}
\]

\[
\frac{1}{2}R_2
\]

**Step 2**

\[
\begin{bmatrix}
1 & 1 & 0 & 5 & 5 \\
0 & 1 & 0 & 1 & 1 \\
0 & 0 & 1 & -1 & -2
\end{bmatrix}
\]

\[-3R_3 + R_1
\]

\[-\frac{1}{2}R_3 + R_2
\]

**Step 3**

\[
\begin{bmatrix}
1 & 0 & 0 & 4 & 4 \\
0 & 1 & 0 & 1 & 1 \\
0 & 0 & 1 & -1 & -2
\end{bmatrix}
\]

Reading off the solution gives (unsurprisingly) the same solution (2.5) that we found before.

**Example 2.2.9** Example 2.1.4 (equation (2.7)) has a row echelon form (2.20):

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

The following steps bring it into row REF

**Step 1**

\[
\begin{bmatrix}
1 & 1 & 3 & -1 \\
0 & 1 & \frac{1}{2} & 0 \\
0 & 0 & 0 & 1
\end{bmatrix}
\]

\[
\frac{1}{2}R_2
\]

\[
R_3 + R_1
\]

**Step 2**

\[
\begin{bmatrix}
1 & 0 & 0 & 4 & 4 \\
0 & 1 & 0 & 1 & 1 \\
0 & 0 & 1 & -1 & -2
\end{bmatrix}
\]

\[-R_2 + R_1
\]

**Example 2.2.10** As a final example, consider Example 2.2.3. We found the row echelon form (2.23):

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

The following steps bring it into row REF

**Step 1**

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

\[-R_3 + R_1
\]

\[-3R_3 + R_2
\]

**Step 2**

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

\[-R_2 + R_1
\]

\[-3R_3 + R_2
\]
CHAPTER 2. MATRICES AND THE SOLUTION OF SIMULTANEOUS LINEAR EQUATIONS

Step 2

\[
\begin{array}{cccc}
1 & 0 & 0 & 11 \\
0 & 1 & 0 & 7 \\
0 & 0 & 1 & -2 \\
0 & 0 & 0 & 0 \\
\end{array}
\]

\[R_2 + R_1\]

\[
\begin{array}{cccc}
1 & 0 & 0 & 11 \\
0 & 1 & 0 & 7 \\
0 & 0 & 1 & -2 \\
0 & 0 & 0 & 0 \\
\end{array}
\]

The array of detached coefficients of Example 2.2.3 therefore has row REF

\[
\begin{array}{cccc}
1 & 0 & 0 & 11 \\
0 & 1 & 0 & 7 \\
0 & 0 & 1 & -2 \\
0 & 0 & 0 & 0 \\
\end{array}
\]

The required solution to Example 2.2.3 (matrix equation (2.22)) is (unsurprisingly) as before. We note that the row REF of the coefficient matrix of the same example is

\[
\begin{array}{cccc}
1 & 0 & 0 & 11 \\
0 & 1 & 0 & 7 \\
0 & 0 & 1 & -2 \\
0 & 0 & 0 & 0 \\
\end{array}
\]

(2.30)

The solution (2.28) to Example 2.2.5 can again be read off this array.

2.2.15 A homogeneous system with more unknowns than equations always has a non-trivial solution

More precisely, we will prove the following

Theorem: Let \( Ax = 0 \) be a homogeneous system of \( m \) equations in \( n \) unknowns and suppose that the row echelon form of \( A \) has \( r < n \) non-zero rows. Then there are solutions \( x \) in which the \( x_j \) corresponding to the \( k = n - r \) non-pivot columns can have arbitrary values. In other words, the solution set has \( k \) independent parameters.

Proof
Consider the row REF \( A'' \) of \( A \). The matrix \( A'' \) has \( r \) non-zero rows and \( r \) pivot columns. Thus there are \( k = n - r \) non-pivot columns. By the nature of the row REF, the \( k \) variables \( x_j \) corresponding to these columns can be given any values, and we can then solve uniquely for the remaining \( r \) variables. (See the illustrative examples below.)

Corollary
A homogeneous system with more unknowns than equations always has a non-trivial solution.

Proof
With the above notation, we have \( m < n \) and since the number \( r \) of non-zero rows in a row REF satisfies \( r \leq m \), we have \( r < n \).

Note 38 In Corollary 91 of section 3.6 in chapter 3 there is an alternative approach to the topic.

Some illustrative examples

1. Suppose that the row REF of a matrix \( A \) is

\[
A'' = \begin{bmatrix}
1 & 3 & -5 & 0 & 0 & 7 \\
0 & 0 & 0 & 1 & 0 & -2 \\
0 & 0 & 0 & 0 & 1 & 3 \\
\end{bmatrix}
\]
Here the non-pivot columns are columns 2, 3 and 6. The general solution to $Ax = 0$ is

$$x = \begin{bmatrix} -3x_2 + 5x_3 - 7x_6 \\ x_2 \\ x_3 \\ 2x_6 \\ -3x_6 \\ x_6 \end{bmatrix} = \begin{bmatrix} -3t_1 + 5t_2 - 7t_3 \\ t_1 \\ t_2 \\ 2t_3 \\ -3t_3 \\ t_3 \end{bmatrix}$$

We have renamed the parameters as $t_1 = x_2$, $t_2 = x_3$ and $t_3 = x_6$.

2. As a second example consider Example 2.2.5 (equation 2.27) which in REF reads (see array (2.30)):

$$\begin{bmatrix} 1 & 0 & 0 & 11 \\ 0 & 1 & 0 & 7 \\ 0 & 0 & 1 & -2 \\ 0 & 0 & 0 & 0 \end{bmatrix} x = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

This shows that we are actually dealing with 3 equations for 4 unknowns. Column 4 of the coefficient matrix is the only non-pivot column. Hence $x_4 = t$ can have any value and we get the same solution as before (equation (2.28).

**Exercise 39**

1. Prove that if the augmented array $[A, b]$ for a system of equations $Ax = b$ has been reduced to row (reduced) echelon form $[A', b']$, then $A'$ is also in (reduced) echelon form.

**Hint 40** A zero row of $A'$ comes after a non-zero row of $A'$.

2. Find the row reduced echelon form of the matrices and arrays of detached coefficients in Exercise (35), numbers 1a, 1b, 2, 3, 4, 9, 11, 12 and 13. Solve (once more) the problems involving systems of equations $Ax = b$ using the row REF. Write down the row REF of the corresponding coefficient matrix $A$. Illustrate the above theorem 2.2.15 in the homogeneous cases.

3. Consider the general problem of solving two linear equations for two unknowns $x_1$ and $x_2$:

$$\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}$$

(a) In (2.31) all the entries except $x_1$ and $x_2$ are supposed known. Using formal Gauss-reduction find formulae for the unknowns $x_1$ and $x_2$ in terms of the other symbols. For the purposes of reduction you may assume any number you like is non-zero. In fact, conclude that if only $d = a_{11}a_{22} - a_{21}a_{12} \neq 0$, then you have a formula for $x$. Can you see that the solution must be unique?

**Remark 41** The number $a_{11}a_{22} - a_{21}a_{12}$ is called the determinant of the matrix $A = [a_{ij}]$, and is denoted by $|A|$. For example,

$$\begin{vmatrix} 3 & -2 \\ 7 & 4 \end{vmatrix} = (3)(4) - (7)(-2) = 26$$

We will study determinants in some detail in Chapter 4.

**Hint 42** Consider the array of detached coefficients and, assuming $a_{11} \neq 0$, the step

$$a_{11} x_1 + a_{12} x_2 = b_1$$

Now assume

$$-\frac{a_{21}a_{12}}{a_{11}} + a_{22} \neq 0$$

and solve for $x_2$. Next, find $x_1$. 
(b) Use your formula to solve for \( \mathbf{z} \):

\[
\begin{bmatrix}
-5 & 11 \\
3 & -7
\end{bmatrix}
\mathbf{z} =
\begin{bmatrix}
-2 \\
13
\end{bmatrix}
\]
2.3 Summary of Chapter 2

2.3.1 Matrix equation of simultaneous equations

Given a system of \( m \) equations for \( n \) unknowns, this is expressed compactly by a matrix equation (See section 2.2).

\[ Ax = b \]

where \( A \) is the \( m \times n \) coefficient matrix, \( b \) is a column with \( m \) entries and \( x \) is a column with \( n \) entries.

Entry \( i \) of the column matrix \( Ax \) is the ‘dot’ product of row \( i \) of \( A \) and the column \( x \) (See subsection 2.2.3 and equation 2.13):

\[
\begin{bmatrix}
  a_{i1} & a_{i2} & \cdots & a_{in}
\end{bmatrix}
\begin{bmatrix}
  x_1 \\
  x_2 \\
  \vdots \\
  x_n
\end{bmatrix}
= \begin{bmatrix}
  a_{i1}x_1 + \cdots + a_{in}x_n
\end{bmatrix}
\]

2.3.2 Detached Coefficients

(See subsection 2.2.5).

The other way to express the system of equations \( Ax = b \) is as the augmented matrix, or array of detached coefficients

\[ C = [A, b] \]

where the last column of \( R \) is understood to be the column \( b \) of constants.

2.3.3 Elementary row operations (eros). Row echelon and row reduced echelon form

(See subsections 2.2.6, 2.2.8 and 2.2.11).

Elementary row operations (eros), known as Gauss reduction, are done on the array \( R \) reducing it to row echelon form (row EF):

\[ C' = [A', b'] \]

Further eros (Gauss-Jordan) done on \( C' \) reduce it to row reduced echelon form (row REF):

\[ C'' = [A'', b''] \]

The important fact about elementary row operations on a system of equations (or its detached array of coefficients) is that they leave the solution set unchanged. For any particular \( x \), the equation \( Ax = b \) is satisfied if, and only if, the equation \( A'x = b' \) is satisfied if, and only if, the equation \( A''x = b'' \) is satisfied.

2.3.4 Homogeneous equations

(See subsection 2.2.10).

A homogeneous system of equations has \( b = 0 \) and consequently \( b' = b'' = 0 \), since an elementary row operation done on \( 0 \) leaves it as \( 0 \).

For a particular \( x \), the matrix equation \( Ax = 0 \) holds if, and only if, \( A'x = 0 \) if, and only if, \( A''x = 0 \) holds.

A system of homogeneous equations with more unknowns than equations always has a non-trivial solution (see 2.2.15).
Chapter 3

Linear Transformations and Matrices

In this chapter we will learn how the solution to a set of linear equations is the inverse operation of the interpretation of a matrix $A$ as a transformation or mapping. In this sense $A$ transforms a vector $x$ into another vector $y = Ax$. Examples in ordinary space are projections on planes and lines through the origin, reflections in such objects and rotations about an axis through $O$ (important for mechanics).

Next, we deal with the difficult idea of linear independence, and it is followed by a section on subspaces that can be regarded as an introduction to the idea of a general 'vector space'. Finally, there is a section on inverses of matrices.

3.1 The Algebra of Vectors and Matrices

We begin by developing some systematic notation for $m \times n$ matrices, their rows and columns and for column and row vectors in general.

3.1.1 Systematic Notation For Matrices

1. Recall from section 2.2 that an $m \times n$ matrix is a rectangular $A$ with $m$ rows and $n$ columns. In standard notation, $A = [a_{ij}]$ where the $i - j$ entry $a_{ij} \in \mathbb{R}$ is at the intersection of row $i$ and column $j$.

The set of $m \times n$ matrices is denoted by $\mathbb{R}^{m \times n}$. Hence $A \in \mathbb{R}^{m \times n}$ means that $A$ has $m$ rows and $n$ columns and that the size of $A$ is $m \times n$. Some authors write $A = A_{m \times n}$ to indicate the size of $A$.

Alternative notation If $A = [a_{ij}]$, we will also find it most convenient to use the notation

$$[A]_{ij} = a_{ij}$$

(3.1)

to describe the $i - j$ entry $a_{ij}$ of $A$.

An $n \times n$ matrix is called square of size $n$. Such matrices deserve to be studied in their own right (see section 3.5).

2. From the definitions it follows that $\mathbb{R}^{n \times 1}$ is the set of columns with $n$ entries. We call
these column vectors and write
\[
\hat{b} = \begin{bmatrix}
b_1 \\
b_2 \\
\vdots \\
b_n
\end{bmatrix}
\] (3.2)

\(\mathbb{R}^{1 \times n}\) denotes the set of row vectors with \(n\) entries:
\[
a = \begin{bmatrix}a_1 & a_2 & \ldots & a_n\end{bmatrix}
\] (3.3)

For typographical reasons this row vector is sometimes written \([a_1, \ldots, a_n]\). It is traditional to use \(\mathbb{R}^n\) for both \(\mathbb{R}^{n \times 1}\) (column vectors) and \(\mathbb{R}^{1 \times n}\) (row vectors). We will only use the notation \(\mathbb{R}^n\) if the context makes it quite clear whether rows or columns are meant.

3. The \(m \times n\) zero matrix has 0 for all its entries and is written \(O_{m \times n}\), or simply as \(O\) when the size is understood.

4. Addition of matrices
Let \(A = [a_{ij}]\) and \(B = [b_{ij}]\) be of the same \(m \times n\) size. Their sum \(A + B = C\) where the \((i, j)\)-entry of \(C\) is \(c_{ij} = a_{ij} + b_{ij}\). In brief, using the notation or equation (3.1),
\[
[A + B]_{ij} = [A]_{ij} + [B]_{ij}
\] (3.4)

It is essential to note that the addition \(A + B\) makes sense only if \(A\) and \(B\) have the same size.

5. Multiplication of a matrix by a scalar
Let \(A = [a_{ij}]\) and \(s \in \mathbb{R}\). The product \(sA = As\) has the number \(sa_{ij}\) in the \((i, j)\)-position:
\[
[sA]_{ij} = [As]_{ij} = s[A]_{ij} = [A]_{ij}s
\]

Remark 43 Notice that it makes no difference if we write the scalar \(s\) on the left or on the right of \(A\). The matrix \(-A\) is defined as \((-1)A\) and has for \((i, j)\) entry the number \(-a_{ij}\) i.e., \([-1][A]_{ij} = [-A]_{ij}\). The difference \(A - B\) of two matrices of the same size means the same thing as \(A + (-B)\).

We will see that if \(a\) is a row vector (3.3) then it is more natural to write \(sa = [sa_1 \quad sa_2 \quad \ldots \quad sa_n]\) rather than \(as\).

On the other hand, for a column \(b\), as in equation (3.2), it is better to write \(bs\). See, for example, Equation (3.23) below.

6. Row \(i\) of the \(m \times n\) matrix \(A = [a_{ij}]\) is denoted by \(A_{i \bullet}\). Hence,
\[
A_{i \bullet} = \begin{bmatrix}a_{i1} & a_{i2} & \ldots & a_{in}\end{bmatrix} \quad (i = 1, \ldots, m)
\] (3.5)

7. Column \(j\) of the \(m \times n\) matrix \(A = [a_{ij}]\) is denoted by \(A_{\bullet j}\). Hence,
\[
A_{\bullet j} = \begin{bmatrix}a_{1j} \\
a_{2j} \\
\vdots \\
a_{mj}
\end{bmatrix} \quad (j = 1, \ldots, n)
\] (3.6)

Remark 44 The symbols for row \(i\) and column \(j\) are consistent with our other notation. Thus \([A]_{ij} = a_{ij}\) is at the intersection of row \(A_{i \bullet}\) and column \(A_{\bullet j}\).

Other notations in use for the \(i^{th}\) row of \(A\), are \(A_i\) and \([A]_i\), and for the \(j^{th}\) column of \(A\), \(A^{(j)}\) or \([A]^j\). However, we will not use these.
8. The unit column vectors $e_j \in \mathbb{R}^{n \times 1}$ ($j = 1, \ldots, n$) have a 1 in position $j$ and zeros elsewhere:

$$e_j = \begin{bmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{bmatrix} \text{ row } j \quad (3.7)$$

9. The $n \times n$ Identity matrix is the $n \times n$ matrix with columns $e_1, \ldots, e_n$ (in that order). It is denoted by $I_n$ (or simply $I$ if $n$ is understood):

$$I_n = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix} \quad (3.8)$$

As we will see, $I_n$ behaves like the number 1 when we multiply matrices. (See Exercise 63 No.5). We note that, writing $I_n = I$, by definition $I_{ij} = e_j$ for $j = 1, \ldots, n$. The unit row vectors are

$$I_i = \begin{bmatrix} a_{1} \\ a_{2} \\ \vdots \\ a_{n} \end{bmatrix} \quad (i = 1, \ldots, n)$$

The identity $I = I_n$ of (3.8) is equivalently defined by

$$[I]_{ij} = 1 \text{ if } i = j \text{ and } [I]_{ij} = 0 \text{ if } i \neq j \quad (i, j = 1, \cdots, n)$$

10. The transpose of the $m \times n$ matrix $A = [a_{ij}]$ is denoted by $A^T$ ("A transpose"). The transpose $A^T$ has for its $(i, j)$ entry the number $a_{ji}$. Thus $A^T$ is an $n \times m$ matrix and can also be vaguely defined as ‘the matrix $A$ with its rows and columns interchanged’. In fact, using the notation (3.1),

$$[A^T]_{ji} = [A]_{ij} \quad (i = 1, \ldots, n; j = 1, 2, \ldots, m) \quad (3.9)$$

11. The product $\mathbf{a}^T \mathbf{b}$ of the row $\mathbf{a}$ with $n$ entries and the column $\mathbf{b}$ with $n$ of entries is

$$\mathbf{a}^T \mathbf{b} = \begin{bmatrix} a_1 & a_2 & \cdots & a_n \end{bmatrix} \begin{bmatrix} b_1 \\ b_1 \\ \vdots \\ b_n \end{bmatrix} = a_1 b_1 + a_2 b_2 + \cdots + a_n b_n \quad (3.10)$$

**Remark 45** This was introduced in subsection 2.2.3 of Chapter 2. You can, if you wish, call this a ‘dot product’ since it is a generalization of the dot product Equation (1.4) of Chapter 1, but there is no need here for the ‘dot’. Note the order in $\mathbf{a}^T \mathbf{b}$: first $\mathbf{a}$, then $\mathbf{b}$. This is important as $\mathbf{b}^T \mathbf{a}$ will turn out to be something quite different: in fact, an $n \times n$ matrix. (See No.3 of Exercise 63).

### 3.1.2 Linear combinations of vectors, linear dependency, Span

1. The column vector $\mathbf{v} \in \mathbb{R}^{n \times 1}$ is a linear combination of the vectors $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_k$ with coefficients $s_1, s_2, \ldots, s_k$ if

$$\mathbf{v} = s_1 \mathbf{u}_1 + s_2 \mathbf{u}_2 + \cdots + s_k \mathbf{u}_k \quad (3.11)$$
We then say that \( v \) depends linearly on \( u_1, u_2, \ldots, u_k \). Of course, we can also write \( v = s_1 u_1 + s_2 u_2 + \cdots + s_k u_k \), with the coefficients on the left, but, as remarked, we prefer to write the coefficients of columns on the right. If \( v \) and \( u_1, u_2, \ldots, u_k \) are row vectors all of the same size our preference would be

\[
v = s_1 u_1 + s_2 u_2 + \cdots + s_k u_k
\]

2. The idea of 'span' was introduced in Exercise 2, No.14.

The span of (column or row) vectors \( u_1, u_2, \ldots, u_k \) is the set of all linear combinations of these vectors. It is also called the space spanned by the vectors and is denoted by \( sp(u_1, \ldots, u_k) \). Thus,

\[
v \in sp(u_1, u_2, \ldots, u_k)
\]

if, and only if, (3.11) holds for certain coefficients \( s_1, \ldots, s_k \).

The idea of 'span' first came up in No.14 of Exercise 2.

The column space of a matrix \( A \) is the span of its columns; its row space is the space spanned by its rows.

The ideas of 'span', column space and row space will only be systematically developed in the section 3.6 on vector spaces.

3. **Linear combinations of matrices**

We wish to extend the definition of 'linear combination' to matrices: Let \( A_1, A_2, \ldots, A_k \) be \( k \) matrices, all of the same size and suppose that \( s_1, s_2, \ldots, s_k \) are \( k \) scalars. The linear combination of \( A_1, A_2, \ldots, A_k \) with coefficients \( s_1, s_2, \ldots, s_k \) is

\[
s_1 A_1 + s_2 A_2 + \ldots + s_k A_k
\]

### 3.1.3 Some examples

1. \[
\begin{bmatrix}
-1 & 5 & 13 \\
2 & -4 & -7 \\
\end{bmatrix}
\in \mathbb{R}^{2\times3},
\begin{bmatrix}
-4 & 5 \\
11 & 2 \\
\end{bmatrix}
\in \mathbb{R}^{2\times2},
\begin{bmatrix}
0 & 0 \\
0 & 0 \\
\end{bmatrix}
\in \mathbb{R}^{2\times2}
\]

2. \[
\begin{bmatrix}
-3 \\
\frac{3}{2} \\
\frac{5}{6} \\
\end{bmatrix}
\in \mathbb{R}^{2\times1},
\begin{bmatrix}
\frac{3}{2} \\
\frac{5}{6} \\
\end{bmatrix}
\in \mathbb{R}^{3\times1}
\]

3. \[O_{23} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix},
O_{22} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}\]

4. \[
\begin{bmatrix}
-1 & 5 & 13 \\
2 & -4 & -7
\end{bmatrix} + \begin{bmatrix}
-7 & \frac{3}{2} & -7 \\
-3 & -\frac{3}{2} & -7
\end{bmatrix} = \begin{bmatrix}
1 & -\frac{1}{2} & 6 \\
-1 & -\frac{3}{2} & -14
\end{bmatrix}
\]

5. \[
\begin{bmatrix}
-1 & 5 & 13 \\
2 & -4 & -7
\end{bmatrix}(100) = (100)\begin{bmatrix}
-1 & 5 & 13 \\
2 & -4 & -7
\end{bmatrix} = \begin{bmatrix}
-100 & 125 & 1300 \\
200 & -400 & -700
\end{bmatrix}
\]

6. (a) Let

\[
A = \begin{bmatrix}
-1 & 2 & -\frac{2}{3} \\
\frac{5}{4} & -4 & \frac{5}{8} \\
13 & -7 & \frac{3}{8}
\end{bmatrix}
\]

Then

\[
A_{1\bullet} = \begin{bmatrix}
-1 & 2 & -\frac{2}{3}
\end{bmatrix}
\]

\[
A_{3\bullet} = \begin{bmatrix}
13 & -7 & \frac{3}{8}
\end{bmatrix}
\]
(b) Let

\[ X = \begin{bmatrix}
  7 \\
  -3 \\
  9 \\
  -5 \\
  6
\end{bmatrix} \]

then

\[ X_2 = \begin{bmatrix}
  -3 \\
  2
\end{bmatrix} \]

\[ X_4 = \begin{bmatrix}
  -5 \\
  6
\end{bmatrix} \]

**Remark 46** We usually use an ‘underline’ notation like (3.2) for column vectors but will occasionally write \( X_2 = X_{2\bullet} = -\frac{3}{2} \) if it is understood that \( X \) is a column matrix (i.e. a column vector). Similar remarks apply to row vectors (i.e. row matrices).

7. (a) If \( A \) is equation (3.12) above then

\[ A_{\bullet 2} = \begin{bmatrix}
  2 \\
  -4 \\
  -7
\end{bmatrix} \]

\[ A_{\bullet 1} = \begin{bmatrix}
  -1 \\
  5 \\
  13
\end{bmatrix} \]

(b) If

\[ Y = \begin{bmatrix}
  5 & -\frac{7}{2} & -9 & 10 & -23 \\
\end{bmatrix} \]

\[ Y_{\bullet 2} = \begin{bmatrix}
  -\frac{7}{2}
\end{bmatrix} \]

\[ Y_{\bullet 5} = [-23] \]

8. (a) The vectors (3 \times 1 column matrices) \( e_1 \), \( e_2 \) and \( e_3 \) in \( \mathbb{R}^{3\times1} \) are

\[ e_1 = \begin{bmatrix}
  1 \\
  0 \\
  0
\end{bmatrix} , \quad e_2 = \begin{bmatrix}
  0 \\
  1 \\
  0
\end{bmatrix} , \quad e_3 = \begin{bmatrix}
  0 \\
  0 \\
  1
\end{bmatrix} \]

We met these as **rows** in Chapter 1, 1.1.17, calling them \( \hat{i}, \hat{j}, \) and \( \hat{k} \) respectively.

(b) In \( \mathbb{R}^{4\times1} \)

\[ e_1 = \begin{bmatrix}
  1 \\
  0 \\
  0 \\
  0
\end{bmatrix} , \quad e_2 = \begin{bmatrix}
  0 \\
  1 \\
  0 \\
  0
\end{bmatrix} , \quad e_3 = \begin{bmatrix}
  0 \\
  0 \\
  1 \\
  0
\end{bmatrix} , \quad e_4 = \begin{bmatrix}
  0 \\
  0 \\
  0 \\
  1
\end{bmatrix} \]

9. We have

\[ I_3 = \begin{bmatrix}
  1 & 0 & 0 \\
  0 & 1 & 0 \\
  0 & 0 & 1
\end{bmatrix} \]

and

\[ I_4 = \begin{bmatrix}
  1 & 0 & 0 & 0 \\
  0 & 1 & 0 & 0 \\
  0 & 0 & 1 & 0 \\
  0 & 0 & 0 & 1
\end{bmatrix} \]
10. (a) Let

\[ A = \begin{bmatrix} -5 & 0 & 2 \\ 7 & 6 & -1 \\ \frac{1}{2} & -1 & 3 \\ -\frac{1}{13} & 9 & 0 \end{bmatrix} \]

then \( A \in \mathbb{R}^{4 \times 3} \), while the transpose \( A^T \in \mathbb{R}^{3 \times 4} \) and

\[ A^T = \begin{bmatrix} -5 & 7 & \frac{1}{2} & -\frac{1}{13} \\ 0 & 6 & -1 & 9 \\ 2 & -1 & 3 & 0 \end{bmatrix} \]

(b) The transpose of the column vector (3.2) is the row

\[ \mathbf{b}^T = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix} \]

and the transpose of the row vector (3.5) of \( A \) is a column vector.

(c) \([I_i]^T = I_\bullet \) and \([I_\bullet]^T = I_\bullet \) and \( I^T = I \), where \( I = I_n \).

11. (a) An example of the product \( a^T b \) of a row \( a \) and a column \( b \) is

\[ \begin{bmatrix} -3 & 5 & -7 & 4 \end{bmatrix} \begin{bmatrix} 2 \\ 3 \\ -2 \\ 9 \end{bmatrix} = (-3)(2) + (5)(3) + (-7)(-2) + (4)(9) = 59 \]

(b) Another, using a general 2 \( \times \) 3 matrix \( A = [a_{ij}] \), and a general 3 \( \times \) 4 matrix \( B = [b_{ij}] \) is

\[ A_{i\bullet}B_{\bullet j} = a_{i1}b_{1j} + a_{i2}b_{2j} + a_{i3}b_{3j} \quad (i = 1, 2; j = 1, 2, 3, 4) \]

Note that such a product is only possible when the number of columns of \( A \) equals the number of rows of \( B \). The significance of such products will be seen in subsection 3.3.4 below.

Referring to linear combinations (item (3) in 3.1.1) we have:

1. (a) In \( \mathbb{R}^{4 \times 1} \)

\[ \begin{bmatrix} -5 \\ \frac{1}{2} \\ -\frac{2}{3} \\ 2 \end{bmatrix} (-6) + \begin{bmatrix} \frac{1}{2} \\ -\frac{2}{9} \\ -\frac{5}{6} \end{bmatrix} (3) + \begin{bmatrix} 11 \\ 2 \\ -1 \end{bmatrix} \left( -\frac{1}{2} \right) = \begin{bmatrix} -35/2 \\ -47 \\ -6 \end{bmatrix} \]

(b) In \( \mathbb{R}^{1 \times 5} \)

\[ (-2) \begin{bmatrix} p & q & \frac{1}{2} & -3 & s \end{bmatrix} + 4 \begin{bmatrix} -1 \\ r + 2 \\ -\frac{1}{4} \\ \frac{5}{2} \\ 1 \end{bmatrix} = \begin{bmatrix} -2p - 4 \\ -2q + 4r + 8 \\ -2 \\ 16 \\ -2s + 4 \end{bmatrix} \]

2. The span of the rows of \( I_3 \) is clearly \( \mathbb{R}^{1 \times 3} \).

Consider the column space of matrix

\[ A = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix} \]
The span of the columns of $A$ is the whole of $\mathbb{R}^{4 \times 1}$, since for any choices of $b_1, b_2, b_3$ and $b_4$,
\[
\begin{bmatrix}
  b_1 \\
  b_2 \\
  b_3 \\
  b_4 
\end{bmatrix} = \begin{bmatrix}
  1 \\
  0 \\
  0 \\
  0 
\end{bmatrix} (b_1 - b_2) + \begin{bmatrix}
  1 \\
  0 \\
  0 \\
  0 
\end{bmatrix} (b_2 - b_3) + \begin{bmatrix}
  1 \\
  1 \\
  0 \\
  0 
\end{bmatrix} (b_3 - b_4) + \begin{bmatrix}
  1 \\
  1 \\
  1 \\
  1 
\end{bmatrix} b_4
\]

See also subsections 3.1.6 and 3.1.9 below.

3. In $\mathbb{R}^{2 \times 2}$
\[
(7) \begin{bmatrix}
  -1 & 2 \\
  3 & 6 
\end{bmatrix} + (-3) \begin{bmatrix}
  0 & 8 \\
 -2 & 5 
\end{bmatrix} = \begin{bmatrix}
  -1 & 2 \\
  3 & 6 
\end{bmatrix} + \begin{bmatrix}
  0 & 8 \\
 -2 & 5 
\end{bmatrix} (-3) \\
  = \begin{bmatrix}
  -7 & 10 \\
  27 & 27 
\end{bmatrix}
\]

3.1.4 Some basic algebraic properties
For all $m \times n$ matrices $A$, $B$ and $C$ and scalars $s$ and $t$,
1. $A + (B + C) = (A + B) + C$ (associative law for addition).
3. $O + A = A$ ($O = O_{m \times n}$ behaves like the number 0).
4. $A + (-A) = O$.
5. $sA = As$ (by definition).
6. $s(A + B) = sA + sB$.
7. $(s + t)A = sA + tA$.
8. $s(tA) = (st)A$.
9. $1A = A$.
10. Every vector $\mathbf{x} \in \mathbb{R}^{n \times 1}$ can be uniquely written as a linear combination
\[
\mathbf{x} = \mathbf{e}_1 x_1 + \mathbf{e}_2 x_2 + ... + \mathbf{e}_n x_n
\]
(Compare 1.1.17 in Chapter 1). Hence it is trivial that every $\mathbf{x}$ depends linearly on the unit vectors $\mathbf{e}_1, \mathbf{e}_2, ..., \mathbf{e}_n$, in other words that these unit vectors span $\mathbb{R}^{n \times 1}$. However, there are many other sets of vectors that have a similar property, as we shall see. In Chapter 1, No.15 of Exercise 2 we indicated that any three mutually orthogonal vectors in $\mathbb{R}^3$ span the whole of space. You are asked to show this rigorously in Exercise 78, No.9.
12. Referring to the product (3.10),
\[
(a \mathbf{b})^T = b^T a^T
\] (3.13)
13. (a) Let $\mathbf{u}$ and $\mathbf{v}$ be column vectors with $n$ entries and $\mathbf{a}$ a row with $n$ entries. Then
\[
\mathbf{a} (\mathbf{u} + \mathbf{v}) = \mathbf{a} \mathbf{u} + \mathbf{a} \mathbf{v}
\] (3.14)
(b) If $s$ is a scalar,
\[
\mathbf{a} (s \mathbf{u}) = (\mathbf{a} \mathbf{u}) s = (sa) \mathbf{u}
\] (3.15)
Hence, proof of equation (3.14) and (3.15) combine so that for all scalars \( s \) and \( t \):
\[
\alpha (u\mathbf{s} + v\mathbf{t}) = (\alpha \mathbf{u}) s + (\alpha \mathbf{v}) t \tag{3.16}
\]

(d) More generally, Equation (3.16) can be generalized as follows: Suppose \( u_1, u_2, \ldots, u_k \) are column vectors in \( \mathbb{R}^{n \times 1} \) and \( s_1, s_2, \ldots, s_k \) are scalars and \( \alpha \in \mathbb{R}^{1 \times n} \) is a row vector. Then
\[
\alpha (u_1 s_1 + u_2 s_2 + \ldots + u_k s_k) = (\alpha u_1) s_1 + (\alpha u_2) s_2 + \ldots + (\alpha u_k) s_k \tag{3.17}
\]

14. (a) Quite similarly, if \( \mathbf{a} \) and \( \mathbf{b} \) are row vectors with \( n \) entries and \( \mathbf{u} \) is a column with \( n \) entries,
\[
(\mathbf{a} + \mathbf{b}) \mathbf{u} = \mathbf{a} \mathbf{u} + \mathbf{b} \mathbf{u} \tag{3.18}
\]

(b) and for scalars \( t \),
\[
(t \mathbf{a}) \mathbf{u} = t (\mathbf{a} \mathbf{u}) \tag{3.19}
\]

(c) Equations (3.18) and (3.19) combine so that for all scalars \( s \) and \( t \):
\[
(s \mathbf{a} + t \mathbf{b}) \mathbf{u} = s (\mathbf{a} \mathbf{u}) + t (\mathbf{b} \mathbf{u}) \tag{3.20}
\]

(d) The result (3.20) can be generalized as follows: Suppose \( a_1, a_2, \ldots, a_k \) are rows in \( \mathbb{R}^{1 \times n} \) and \( t_1, t_2, \ldots, t_k \) are scalars and \( \mathbf{u} \in \mathbb{R}^{n \times 1} \) is a column, then
\[
(t_1 a_1 + t_2 a_2 + \ldots + t_k a_k) \mathbf{u} = t_1 (a_1 \mathbf{u}) + t_2 (a_2 \mathbf{u}) + \ldots + t_k (a_k \mathbf{u}) \tag{3.21}
\]

### 3.1.5 Some proofs

The above properties and their proofs are similar to those for vectors discussed in Chapter 1, subsections 1.1.6 and 1.1.9, equation (1.8). See No.7 of Exercise 47 below. We prove a few properties as illustrations.

**Proof of property (6)** that \( s (A + B) = sA + sB \).

The \((i, j)\) entry in \( s (A + B) \) is \( s (a_{ij} + b_{ij}) \). But \( s (a_{ij} + b_{ij}) = sa_{ij} + sb_{ij} \), which is the \((i, j)\)– entry in \( sA + sB \). Since \( s (A + B) \) and \( sA + sB \) have the same entries, they are equal.

Using the alternate notation (equation 3.1,
\[
[s (A + B)]_{ij} = s [A + B]_{ij} = s (A)_{ij} + (B)_{ij} \\
= s [A]_{ij} + s [B]_{ij} = [sA]_{ij} + [sB]_{ij} = [sA + sB]_{ij}
\]

**Proof of property (11)**

\[
[(A^T)^T]_{ij} = [A^T]_{ji} = [A]_{ij}
\]

**Proof of equation (3.14)** With obvious notation, entry \( i \) in the column \( \mathbf{u} + \mathbf{v} \) is \( u_i + v_i \). Hence,
\[
\alpha (\mathbf{u} + \mathbf{v}) = a_1 (u_1 + v_1) + \cdots + a_n (u_n + v_n) \\
= (a_1 u_1 + a_1 v_1) + \cdots + (a_n u_n + a_n v_n) \\
= (a_1 u_1 + \cdots + a_n u_n) + (a_1 v_1 + \cdots + a_n v_n) \\
= \alpha \mathbf{u} + \alpha \mathbf{v}
\]
3.1. THE ALGEBRA OF VECTORS AND MATRICES

Or, using $\sum$-notation,

$$a (u + v) = \sum_{i=1}^{n} a_i (u_i + v_i) = \sum_{i=1}^{n} (a_i u_i + a_i v_i)$$

$$= \sum_{i=1}^{n} a_i u_i + \sum_{i=1}^{n} a_i v_i$$

$$= a u + a v$$

Proof of equation (3.15). Again with obvious notation, the $j^{th}$ entry in the vector $us$ is $u_j s$. Hence

$$a (us) = a_1 (u_1 s) + \cdots + a_n (u_n s)$$

$$= (a_1 u_1 + \cdots + a_n u_n) s$$

$$= (a u) s$$

Proof of equation (3.16). By equation (3.14), $a (us + vt) = a (us) + a (vt)$. Using equation (3.15), $a (us) + a (vt) = (a u) s + (a v) t$.

**Exercise 47**

1. Consider 23 in Chapter 2 and the matrices in 2.2.1 referred to there. For each of these as well as for the following $A \in \mathbb{R}^{m \times n}$ find $m$ and $n$ and, using our notation in equation (3.5) and equation (3.6), write down all rows and columns of $A$.

   (a) $A = \begin{bmatrix} -2 & 5 \\ 3 & -4 \\ -1 & \frac{1}{2} \end{bmatrix}$
   
   (b) $A = \begin{bmatrix} 4 & 7 & 0 \\ 0 & 9 & -1 \\ -\frac{5}{3} & 1 & \frac{4}{3} \end{bmatrix}$
   
   (c) $A = \begin{bmatrix} p & f & c \\ b & q & p \\ d & f & r \end{bmatrix}$
   
   (d) $A = \begin{bmatrix} -3 & -1 & 8 & 0 \\ 2 & -7 & 9 & \frac{1}{2} \\ 0 & \frac{2}{3} & -2 & 0 \\ 1 & -4 & 5 & -6 \end{bmatrix}$

2. Write down $C$ if it is given that

   (a) $C$ has three columns $C_{1*} = \begin{bmatrix} 4 \\ -5 \end{bmatrix}$, $C_{2*} = \begin{bmatrix} 2 \\ 7 \end{bmatrix}$, $C_{3*} = \begin{bmatrix} 0 \\ \frac{3}{2} \end{bmatrix}$
   
   (b) $C$ has three rows and $C_{1*} = [ q \ 5 \ -z ]$, $C_{2*} = [ 0 \ a \ b ]$, $C_{3*} = [ -11 \ x \ \frac{2}{3} ]$.
   
   (c) $C = kI_3$ ($k$ a scalar)

3. Find:

   (a) $\begin{bmatrix} -1 \\ \frac{3}{2} \\ 0 \end{bmatrix} + \begin{bmatrix} \frac{1}{2} \\ -\frac{2}{3} \\ -5 \end{bmatrix} (-6) + \begin{bmatrix} 1 \\ -\frac{5}{12} \\ \frac{1}{10} \end{bmatrix} 5 + \begin{bmatrix} 0 \\ \frac{3}{1} \\ -1 \end{bmatrix} (-4) = (-1)$
   
   (b) $\begin{bmatrix} 6 & -2 \\ 7 & 2 \end{bmatrix} + 3 \begin{bmatrix} 11 & 0 \\ -4 & 5 \end{bmatrix} - 2 \begin{bmatrix} 0 & 6 \\ 9 & 4 \end{bmatrix}$
(c) \((\frac{1}{2} + p) \begin{bmatrix} 2 & -4 & 6 & 3 \end{bmatrix} - \frac{3}{2} (2p + q) \begin{bmatrix} -3 & 9 & 18 & -6 \end{bmatrix} + \frac{1}{2} (p - q) \begin{bmatrix} 0 & -24 & 8 & 6 \end{bmatrix}\)

4. Verify equation (3.16) in case \(a = \begin{bmatrix} -3 & \frac{5}{2} & 2 & -1 \end{bmatrix}, u = \begin{bmatrix} \frac{4}{3} & -1 & -3 & \frac{1}{4} \end{bmatrix}^T, v = \begin{bmatrix} -\frac{5}{6} & \frac{1}{3} & -\frac{1}{2} & 1 \end{bmatrix}^T, s = 2\) and \(t = -6\).

5. Find the transposes of the matrices
\[
\begin{bmatrix}
-\frac{5}{6} & -2 & \frac{5}{2} \\
1 & 8 & -9 \\
-7 & 2 & \frac{1}{3}
\end{bmatrix}, \begin{bmatrix} x & -u & c & d \end{bmatrix} \text{ and } \begin{bmatrix} v & -3 & 7 & a \\
-9 & w & 0 & z
\end{bmatrix}.
\]

6. For the matrices \(A\) in No.5 compare \(A_{ij}\) and \([A^T]_{ji}\), as well as \(A_{ij}\) and \([A^T]_{ji}\). Can we say that \([A_{ij}]^T = [A^T]_{ji}\) and \([A_{ij}]^T = [A^T]_{ji}\)? Are these true for any matrix \(A\)?

7. Prove the above basic properties in subsection 3.1.4. For equation (3.17) and equation (3.21) you may assume \(k=3\), since the proofs of the general cases are quite similar.

Hint 48 Study the above proofs in 3.1.5 as well as your own proofs of the statements in 1.1.6 and 1.1.9 of Chapter 1. Apart from the fact that Chapter 1 concentrated on vectors in \(\mathbb{R}^{1 \times 3}\) or \(\mathbb{R}^{1 \times 2}\) most of the proofs for general vectors go through almost word-for-word.

8. (a) i. In \(\mathbb{R}^m = \mathbb{R}^{n \times 1}\) let \(x\) be a linear combination of the vectors \(a_1, a_2, \text{ and } a_3\). Suppose that each of \(a_1, a_2, \text{ and } a_3\) is a linear combination of \(e, f, \text{ and } g\). Show that \(x\) is a linear combination of \(e, f, \text{ and } g\).

ii. Express the this result in terms of the ‘span’ concept.

(b) Is the above result 8(a) true for \(p \times q\) matrices \(X, A_1, A_2, A_3, E, F\) and \(G\) (in place of the vectors \(x, \ldots, g\))? Can you prove your answer?

(c) Find a statement that generalizes No.8a: Suppose that \(x \in \mathbb{R}^m\) depends linearly on \(a_1, a_2, \ldots, a_k\) and each vector \(a_j\) depends linearly on \(b_1, b_2, \ldots, b_n\), then ...? (You need not prove your statement yet, but see Exercise 63, No.13).

9. Show that, provided \(A\) and \(B\) are of the same size, \((A + B)^T = A^T + B^T\). (Use the notation of equation 3.1).

Solution:
\([A + B]^T_{ji} = [A + B]_{ij} = [A]_{ij} + [B]_{ij} = [A^T]_{ji} + [B^T]_{ji} = [A^T + B^T]_{ji}\).

10. For the following pairs \(A\), \(B\) of matrices find all possible products \(A_{ij}B_{ij}\) and arrange them in matrix form in what you think is a nice way.

(a)
\[
A = \begin{bmatrix} 2 & -9 \\
3 & 1
\end{bmatrix}, \ B = \begin{bmatrix} 8 & 2 & -1 \\
-5 & 4 & 7
\end{bmatrix}
\]

(b)
\[
A = \begin{bmatrix} -1 & 5 \\
3 & 0 \\
2 & -4
\end{bmatrix}, \ B = \begin{bmatrix} -1 & 2 & 1 & 0 \\
3 & 0 & -4 & 7
\end{bmatrix}
\]

Solution:
\[
\begin{bmatrix} 61 & -32 & -65 \\
19 & 10 & 4
\end{bmatrix}
\]
\[
A = \begin{bmatrix} -1 & 5 \\
3 & 0 \\
2 & -4
\end{bmatrix}, \ B = \begin{bmatrix} -1 & 2 & 1 & 0 \\
3 & 0 & -4 & 7
\end{bmatrix}
\]
3.1.6 \( Ax \) viewed in two equivalent ways

1. In terms of the products \( A_i \bullet x \):

Let \( A = [a_{ij}] \) be \( m \times n \) and \( x = [x_j] \) be \( n \times 1 \). In Chapter 2 we considered the system of \( m \) equations for \( n \) unknowns symbolized by \( Ax = b \). In 2.2.4 of the same chapter and in other examples we saw that \( Ax \) is a column with \( m \) entries. Its \( i^{th} \) entry is (compare equation (3.10)):

\[
A_i \bullet x = a_{i1}x_1 + a_{i2}x_2 + \ldots + a_{in}x_n
\]

Hence,

\[
Ax = \begin{bmatrix}
A_1 \bullet x \\
A_2 \bullet x \\
\vdots \\
A_m \bullet x
\end{bmatrix}
\]

We met such a product in 2.2.4 of Chapter 2. The entry in the \( i^{th} \) row of \( Ax \) is the ‘dot product’ \( A_i \bullet x \) of row \( A_i \) and column \( x \).

2. \( Ax \) as a linear combination of the columns of \( A \) with coefficients \( x_1, \ldots, x_n \):

\[
Ax = A_1x_1 + A_2x_2 + \ldots + A_nx_n
\]

This view was anticipated in Chapter 2, for example in equations (2.17, (2.19) and (2.25). Like the previous view, it can immediately be seen from the definitions. The following examples should make this clear.

Example 3.1.1 Let

\[
A = \begin{bmatrix}
a_{11} & a_{12} & a_{13} \\
a_{21} & a_{22} & a_{23}
\end{bmatrix}
\]

and \( x \in \mathbb{R}^{3 \times 1} \). Then

\[
Ax = \begin{bmatrix}
a_{11} & a_{12} & a_{13} \\
a_{21} & a_{22} & a_{23}
\end{bmatrix}
\begin{bmatrix}
x_1 \\
x_2 \\
x_3
\end{bmatrix}
= \begin{bmatrix}
a_{11}x_1 + a_{12}x_2 + a_{13}x_3 \\
a_{21}x_1 + a_{22}x_2 + a_{23}x_3
\end{bmatrix}
= \begin{bmatrix}
a_{11} \\
a_{21}
\end{bmatrix} x_1 + \begin{bmatrix}
a_{12} \\
a_{22}
\end{bmatrix} x_2 + \begin{bmatrix}
a_{13} \\
a_{23}
\end{bmatrix} x_3

So

\[
\begin{bmatrix}
-3 & 6 & 5 \\
2 & -4 & 7
\end{bmatrix}
\begin{bmatrix}
x_1 \\
x_2 \\
x_3
\end{bmatrix}
= \begin{bmatrix}
-3x_1 + 6x_2 + 5x_3 \\
2x_1 - 4x_2 + 7x_3
\end{bmatrix}
= \begin{bmatrix}
-3 \\
2
\end{bmatrix} x_1 + \begin{bmatrix}
6 \\
-4
\end{bmatrix} x_2 + \begin{bmatrix}
5 \\
7
\end{bmatrix} x_3
\]
Example 3.1.2 As a more general example, consider the general system of 4 equations for five unknowns, equation (2.11) of the previous chapter. We found there equation (2.12):

\[ A \overline{x} = \begin{bmatrix} A_{11} & A_{12} & A_{13} & A_{14} & A_{15} \\ A_{21} & A_{22} & A_{23} & A_{24} & A_{25} \\ A_{31} & A_{32} & A_{33} & A_{34} & A_{35} \\ A_{41} & A_{42} & A_{43} & A_{44} & A_{45} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix} = \begin{bmatrix} a_{11}x_1 + a_{12}x_2 + a_{13}x_3 + a_{14}x_4 + a_{15}x_5 \\ a_{21}x_1 + a_{22}x_2 + a_{23}x_3 + a_{24}x_4 + a_{25}x_5 \\ a_{31}x_1 + a_{32}x_2 + a_{33}x_3 + a_{34}x_4 + a_{35}x_5 \\ a_{41}x_1 + a_{42}x_2 + a_{43}x_3 + a_{44}x_4 + a_{45}x_5 \end{bmatrix} \]

As a more general example, consider the general system of 4 equations for five unknowns, equation (2.11) of the previous chapter. We found there equation (2.12): 

\[ A \overline{x} = \begin{bmatrix} A_{11} & A_{12} & A_{13} & A_{14} & A_{15} \\ A_{21} & A_{22} & A_{23} & A_{24} & A_{25} \\ A_{31} & A_{32} & A_{33} & A_{34} & A_{35} \\ A_{41} & A_{42} & A_{43} & A_{44} & A_{45} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix} = \begin{bmatrix} a_{11}x_1 + a_{12}x_2 + a_{13}x_3 + a_{14}x_4 + a_{15}x_5 \\ a_{21}x_1 + a_{22}x_2 + a_{23}x_3 + a_{24}x_4 + a_{25}x_5 \\ a_{31}x_1 + a_{32}x_2 + a_{33}x_3 + a_{34}x_4 + a_{35}x_5 \\ a_{41}x_1 + a_{42}x_2 + a_{43}x_3 + a_{44}x_4 + a_{45}x_5 \end{bmatrix} \]

Example 3.1.3 In the previous example let \( \overline{x} = e_j \) be the unit vector of equation (3.7) with \( n = 5 \) entries. Then \( A e_j = A_{*j} \) (the second column of \( A \)) since \( x_2 = 1 \) and \( x_1 = x_3 = x_4 = x_5 = 0 \). Similarly, for \( j = 1, \cdots, 5 \), we have \( A e_j = A_{*j} \).

3.1.7 \( A e_j = A_{*j} \) in general

From the above example it is clear that if \( A \) is an \( m \times n \) matrix and \( e_j \) is the \( j \)th unit vector as in equation (3.7), then

\[ A e_j = A_{*j} \quad (j = 1, \cdots, n) \quad (3.24) \]

Equivalently, if \( I \) is the \( n \times n \) identity matrix,

\[ A I = A_{*j} \quad (j = 1, \cdots, n) \]

3.1.8 \( A \overline{x} \) as a product of \( A \) and \( \overline{x} \)

Henceforth we will regard \( A \overline{x} \) as the product of the \( m \times n \) matrix \( A \) by the \( n \times 1 \) column matrix \( \overline{x} \).

3.1.9 Connection between linear combinations and solving equations, the column space

This was also anticipated in Chapter 2 (for example, in Exercise 32 and in No.2b from Exercise 35).

Solving a system of equations \( A \overline{x} = \overline{b} \) for \( \overline{x} \) is exactly the same as looking for coefficients \( x_j \) so that \( \overline{b} \) is a linear combination of the columns of \( A \):

\[ A \overline{x} = \overline{b} \text{ if, and only if } \overline{b} = A_{*1}x_1 + A_{*2}x_2 + \cdots + A_{*n}x_n \]

In terms of the concept of ‘span’ (see item 2 in subsection 3.1.1), this is saying that \( \overline{b} \) is in the span of the columns of \( A \):

\[ \overline{b} \in sp(A_{*1}, A_{*2}, \ldots, A_{*n}) \]

Example 3.1.4 Is \( a = [1, 5, 2, 7] \) a linear combination of \( u = [1, 3, 0, 5] \) and \( v = [0, 1, 1, 1] \)?

Another way to put it: 'Does \( a \) depend linearly on \( u \) and \( v \)?' In other terminology we have been using, 'is it true that \( a \in sp(u, v) \)?'

Solution:

To answer this, we change the vectors into column vectors (i.e. we use \( u^T \), \( v^T \), \( a^T \)) and try to solve the following matrix equation for \( \overline{x} \):

\[
\begin{bmatrix}
1 & 0 \\
3 & 1 \\
0 & 1 \\
5 & 1 \\
\end{bmatrix}
\begin{bmatrix}
x_1 \\
x_2 \\
\end{bmatrix}
=
\begin{bmatrix}
1 \\
5 \\
2 \\
7 \\
\end{bmatrix}
\]
In other words, we try to express the column on the right as a linear combination of the columns of the $4 \times 2$ coefficient matrix. Use the array of detached coefficients:

\[
\begin{bmatrix}
1 & 0 & 1 \\
3 & 1 & 5 \\
0 & 1 & 2 \\
5 & 1 & 7
\end{bmatrix}
\]

\[
\begin{align*}
1 & 0 & 1 \\
0 & 1 & 2 \\
0 & 1 & 2
\end{align*}
\]

\[-3R_1 + R_2
\]

\[-5R_1 + R_4
\]

This is equivalent to

\[
\begin{align*}
1 & 0 & 1 \\
0 & 1 & 2 \\
0 & 0 & 0
\end{align*}
\]

\[-R_2 + R_3
\]

\[-R_2 + R_4
\]

Hence $x_2 = 2$ and $x_1 = 1$, and $a^T = u^T + v^T$, so $a = u + 2v$ and $a$ depends linearly on $u$ and $v$.

**Remark 49** We are in the habit of solving equations in the form $Ax = b$ (so we are taking linear combinations of columns), but we could just as well have left the vectors as rows: Then we would be taking linear combinations of rows and using elementary column operations.

**Remark 50** Any result about columns has a corresponding result about rows and vice-versa. This is a fact you will appreciate more and more as we go ahead.

**Example 3.1.5** Consider from Chapter 2, Example 2.1.4, equation (2.7):

\[
\begin{bmatrix}
1 & 1 & 3 \\
1 & 3 & 4 \\
2 & 4 & 7
\end{bmatrix}
\begin{bmatrix}
x_1 \\
x_2 \\
x_3
\end{bmatrix}
= 
\begin{bmatrix}
-1 \\
-1 \\
5
\end{bmatrix}
\]

We found from the detached array in echelon form (2.20) in Chapter 2 that there is no solution. Thus the right-hand side $\begin{bmatrix}
-1 & -1 & 5
\end{bmatrix}^T$ is not a linear combination of the columns of the coefficient matrix $A$.

**Example 3.1.6** Let

\[
A = \begin{bmatrix}
 2 & 5 & -11 \\
-3 & -6 & 12 \\
 4 & 7 & -13
\end{bmatrix}
\]

(3.25)

What sort of solutions does the homogeneous system $Ax = 0$ have? Is some column of $A$ a linear combination of the other columns of $A$?

**Solution:**

Solve the homogeneous linear system

\[
\begin{bmatrix}
 2 & 5 & -11 \\
-3 & -6 & 12 \\
 4 & 7 & -13
\end{bmatrix}
\begin{bmatrix}
x_1 \\
x_2 \\
x_3
\end{bmatrix}
= 
\begin{bmatrix}
 0 \\
 0 \\
 0
\end{bmatrix}
\]

$x = 0$ is in any case the (trivial) solution. As usual, to find the general solution, use Gauss-reduction to find a row echelon form of the coefficient matrix $A$:
and the general solution with parameter \( t \) is

\[
\begin{bmatrix}
  x_1 \\
  x_2 \\
  x_3
\end{bmatrix} = \begin{bmatrix}
  -2t \\
  3t \\
  t
\end{bmatrix}
\]

(3.27)

(Check this). If we put \( t = 1 \), we get

\[
\begin{bmatrix}
  2 5 -11 \\
  -3 -6 12 \\
  4 7 -13
\end{bmatrix} \begin{bmatrix}
  -2 \\
  3 \\
  1
\end{bmatrix} = \begin{bmatrix}
  2 5 -11 \\
  -3 -6 12 \\
  4 7 -13
\end{bmatrix} \begin{bmatrix}
  -2 \\
  3 \\
  1
\end{bmatrix}
\]

\[
= \begin{bmatrix}
  -2 \\
  3 \\
  1
\end{bmatrix}
\]

(3.28)

Since each coefficient \(-2, 3\) and \(1\) is non-zero, every column of \( A \) is a linearly dependent on the other two columns. For example,

\[
A \cdot 1 = \begin{bmatrix}
  2 5 1 \\
  -3 -6 0 \\
  4 7 -2
\end{bmatrix} = \begin{bmatrix}
  2 5 -11 \\
  -3 -6 12 \\
  4 7 -13
\end{bmatrix} \begin{bmatrix}
  2 \\
  -3 \\
  4
\end{bmatrix} \begin{bmatrix}
  5 \\
  -6 \\
  7
\end{bmatrix} \begin{bmatrix}
  3 \\
  1
\end{bmatrix}
\]

Example 3.1.7 Let

\[
A = \begin{bmatrix}
  2 & 5 & 1 \\
  -3 & -6 & 0 \\
  4 & 7 & -2
\end{bmatrix}
\]

(3.29)

Can we find a non-trivial solution \( \bar{x} \) to \( A\bar{x} = 0 \)? Is any column of \( A \) linearly dependent on the other two?

Solution:

Reduce \( A \) to row-echelon form:

\[
\begin{align*}
2 & 5 & 1 & 2 & 5 & 1 & 2 & 5 & 1 \\
0 & 5 & 3 & 0 & 2 & 1 & 2 & 5 & 1 \\
0 & -3 & -4 & 0 & -1 & 2 & 1 & 2 & 1
\end{align*}
\]

So a row EF for \( A \) is

\[
C = \begin{bmatrix}
  2 & 5 & 1 \\
  0 & 1 & 1 \\
  0 & 0 & 1
\end{bmatrix}
\]

Thus \( A\bar{x} = 0 \) has only the trivial solution

\[
\bar{x} = \begin{bmatrix}
  0 \\
  0 \\
  0
\end{bmatrix}
\]

Hence no column of \( A \) can depend linearly on the other two. For example, should we have

\[
A \cdot 1 = A \cdot 2 + A \cdot 3, \text{ then}
\]

\[
\bar{x} = \begin{bmatrix}
  1 \\
  -\beta \\
  -\gamma
\end{bmatrix}
\]

would be a non-trivial solution to \( A\bar{x} = 0 \) because \( 1 \neq 0 \).
Remark 51 Questions of the sort “which columns of \( A \) are linearly dependent on other columns of \( A \)?” are further explored below in 3.4.22. When \( A x = 0 \) has only the trivial solution \( x = 0 \) we will call the columns of \( A \) linearly independent. See subsection 3.4.1.

3.1.10 Fundamental properties of the product \( A^T \)

Let \( A \) be an \( m \times n \) matrix and suppose \( v_1, v_2 \) are vectors in \( \mathbb{R}^n = \mathbb{R}^{n \times 1} \) and \( s_1, s_2 \) are scalars. Then

\[
A (v_1 s_1 + v_2 s_2) = (Av_1) s_1 + (Av_2) s_2
\]

(3.30)

Proof

Consider the \( i \)th entry of the left-hand side of (3.30). By equation (3.16) this is

\[
A_i (v_1 s_1 + v_2 s_2) = (A_i v_1) s_1 + (A_i v_2) s_2
\]

As \( (A_i v_1) s_1 + (A_i v_2) s_2 \) is by definition the \( i \)th entry of the right-hand side of equation (3.30), the result follows.

More generally, equation (3.30) easily extends to the following general result:

Let \( A \) be an \( m \times n \) matrix and suppose \( v_1, v_2, ..., v_k \) are vectors in \( \mathbb{R}^n = \mathbb{R}^{n \times 1} \) and \( s_1, s_2, ..., s_k \) are scalars. Then

\[
A (v_1 s_1 + v_2 s_2 + ... + v_k s_k) = (Av_1) s_1 + (Av_2) s_2 + ... + (Av_k) s_k
\]

(3.31)

Exercise 52

1. Find \( I_2 \) if \( I = I_n \) is the \( n \times n \) identity matrix as in equation (3.8) and \( x \in \mathbb{R}^{n \times 1} \).

2. Let \( A \) be \( m \times n \). Show that \( \beta A_2 + A_1 \) is a linear combination of \( A_2, \ldots, A_m \) if, and only if, \( A_i \) is a linear combination of \( A_2, \ldots, A_m \).

3. Suppose that \( v_1, v_2, ..., v_k \) are solutions to a system \( A \) = \( b \) of \( m \) equations in \( n \) unknowns \( x_1, ..., x_n \). Let \( \mu = v_1 \alpha_1 + v_2 \alpha_2 + \cdots + v_k \alpha_k \) be a linear combination of these vectors.

(a) Show that if the system is homogeneous i.e. \( b = 0 \), then \( \mu \) is also a solution, i.e. \( A \mu = 0 \).

Solution:

\[
A \mu = (A v_1) \alpha_1 + (A v_2) \alpha_2 + \cdots + (A v_k) \alpha_k = (Av_1) \alpha_1 + (Av_2) \alpha_2 + \cdots + (Av_k) \alpha_k = 0 \alpha_1 + \cdots 0 \alpha_k = 0
\]

(b) Find a simple example to show that (a) may fail if \( b \neq 0 \).

(c) If \( \alpha_1 + \alpha_2 + \cdots + \alpha_k = 1 \) show that \( A \mu = b \).

4. (a) Consider Exercise 35 of Chapter 2, Numbers 3, 4a, 5a and 6, 11a and 11b. These require solutions to a matrix equation \( A \) = \( b \). For which of these is \( b \) a linear combination of the columns of \( A \)?

Consider those for which \( b \) is a linear combination of the columns of \( A \). For which of these are the coefficients in the linear combination unique? For those which are not, express \( b \) in two ways as linear combinations of the columns of \( A \).

(b) * In Problem 13 from Exercise 35 of Chapter 2. For which values of \( a, b, c, d \) is

\[
\begin{bmatrix}
a & b \\
c & d
\end{bmatrix}^T
\]

a linear combination of the columns of \( A \)?

5. Prove the following theorem, which was anticipated in No.10 of Exercise 35 in Chapter 2.

Let \( A \) = \( b \) represent \( m \) equations in \( n \) unknowns. Suppose that the system has at least
one particular solution, say \( p \). Then if \( Ax = 0 \) so is \( x = p + v \) a solution and all solutions to \( Ax = b \) have this form.

Conclude that if \( m < n \) and \( Ax = b \) has at least one solution, then the system has infinitely many solutions. In fact, if there are \( k \) non-pivot columns then the general solution contains \( k \) independent parameters. (Use 2.2.15 from Chapter 2.)

### 3.2 Matrices as Mappings

Let \( A \) be an \( m \times n \) matrix. In what follows \( \mathbb{R}^n \) stands for the set of column vectors with \( n \) entries, i.e. \( \mathbb{R}^n = \mathbb{R}^{n \times 1} \). The matrix \( A \) converts or transforms the vector \( x \in \mathbb{R}^n \) into the vector \( Ax \in \mathbb{R}^m \). In other words, letting \( y = Ax \), have a function of the independent variable \( x \) with \( y = Ax \) being the dependent variable. In the language of set-theory, \( A \) is a function or transformation or mapping with domain \( \mathbb{R}^n \) and range in the set \( \mathbb{R}^m \):

\[
A : \mathbb{R}^n \to \mathbb{R}^m \quad (3.32)
\]

A maps the vector \( x \) to the vector \( y = Ax \). Symbolically,

\[ x \to Ax \]

#### 3.2.1 The range

The range of \( x \to Ax \) is the set of vectors \( y \) such that \( Ax = y \) for some \( x \). In other words, the range is just the set of all linear combinations of the columns of \( A \). The range is therefore the column space of \( A \) (see item 2 in subsections 3.1.1 and 3.1.9).

See also section 3.6, where some important properties of subspaces are developed.

**Remark 53** If \( c \) is a fixed column vector with \( m \) entries then \( x \to c + Ax \) is also a mapping from \( \mathbb{R}^n \) to \( \mathbb{R}^m \). However, we will concentrate mainly on the case \( c = 0 \) (linear transformations).

#### 3.2.2 Solving equations is the inverse process of the mapping \( x \to Ax \)

Given the mapping (3.32) and a vector \( y \in \mathbb{R}^m \), we can ask: Is there an \( x \in \mathbb{R}^n \) such that \( y = Ax \)? This is precisely the problem of searching for a solution \( x \) to the system of equations \( y = Ax \), as discussed in 3.1.9. In Chapter 2 we found that some systems have a solution, others not.

**Example 3.2.1** Consider the Cartesian plane \( \mathbb{R}^2 \). In equation (1.27) of Chapter 1 we found the reflection of the point \((p_1, p_2)\) in the line \( \ell \) lying in the \( x-y \) plane and which passes through the origin making an angle of \( 30^\circ \) with the positive \( x \)-axis:

\[
p' = \left( \frac{1}{2} p_1 + \frac{1}{2} \sqrt{3} p_2, \frac{1}{2} \sqrt{3} p_1 - \frac{1}{2} p_2 \right)
\]

Hence considered as a mapping, the matrix

\[
A = \begin{bmatrix}
\frac{1}{2} & \frac{1}{2} \\
\frac{1}{2} \sqrt{3} & -\frac{1}{2}
\end{bmatrix}
\]

transforms the point \( x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \) into its reflection in the above line:

\[
y = A \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix}
\frac{1}{2} x_1 + \frac{1}{2} \sqrt{3} x_2 \\
\frac{1}{2} \sqrt{3} x_1 - \frac{1}{2} x_2
\end{bmatrix}
\]
3.2. MATRICES AS MAPPINGS

(We have written \( p \) and \( p' \) as columns \( x \) and \( y \) respectively). See Figure 3.1. Geometrically, it is clear that the range of this reflection is \( \mathbb{R}^2 \). This can be seen algebraically by showing that, whatever the values of \( y_1 \) and \( y_2 \), the matrix equation

\[
\begin{bmatrix}
\frac{1}{2}x_1 + \frac{1}{2}\sqrt{3}x_2 \\
\frac{1}{2}\sqrt{3}x_1 - \frac{1}{2}x_2
\end{bmatrix}
= \begin{bmatrix}
y_1 \\
y_2
\end{bmatrix}
\]

always has a solution for \( x_1 \) and \( x_2 \). (Show this).

![Figure 3.1](image)

Example 3.2.2 The simpler matrix

\[
A = \begin{bmatrix}
-1 & 0 \\
0 & 1
\end{bmatrix}
\]

represents a reflection in the \( x_2 \)-axis. The range is again \( \mathbb{R}^2 \).

Example 3.2.3 The matrix

\[
A = \begin{bmatrix}
1 & 0 & 0 \\
0 & 1 & 0
\end{bmatrix}
\]

represents the mapping that projects the point \( \mathbf{p} = [x_1, x_2, x_3]^T \in \mathbb{R}^3 \) in space onto the point \( \mathbf{y} = A\mathbf{p} = [x_1, x_2]^T \in \mathbb{R}^2 \) in the \( x_1 - x_2 \) plane. The range is obviously \( \mathbb{R}^2 \).

Example 3.2.4 Consider a fixed column vector \( \mathbf{a} \) with three entries and non-zero column matrix

\[
A = \begin{bmatrix}
\mathbf{u}_1 \\
\mathbf{u}_2 \\
\mathbf{u}_3
\end{bmatrix}
\]

The mapping, \( A : \mathbb{R}^3 \to \mathbb{R}^1 = \mathbb{R} \), and \( t \to \mathbf{a} + At \) represents a parametric equation of a straight line in space passing through the point \( \mathbf{a} \), as in Chapter 1, 1.2. The only difference is that here we are using columns instead of rows.
Example 3.2.5 Consider the matrix
\[
A = \begin{bmatrix}
  u_1 & v_1 \\
  u_2 & v_2 \\
  u_3 & v_3 \\
\end{bmatrix}
\]
in which the columns are non-zero and non-parallel. The mapping
\[
\begin{bmatrix}
  s \\
  t \\
\end{bmatrix} \rightarrow A \begin{bmatrix}
  s \\
  t \\
\end{bmatrix}
\]
now represents a generic equation of a plane going through the origin. See Chapter 1, subsection 1.3.1.

Example 3.2.6 The matrix
\[
A = \begin{bmatrix}
  -4 & 0 & 0 \\
  0 & -4 & 0 \\
  0 & 0 & -4 \\
\end{bmatrix}
\]
transforms the point \( x \in \mathbb{R}^3 \) into the point \( y = Ax = -4x \in \mathbb{R}^3 \). This represents a stretching \( x \rightarrow 4x \) of \( x \) by a factor of 4 followed by an inversion \( x \rightarrow -x \) in the origin \( O \).

Example 3.2.7 The matrix \( A = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \) as mapping of \( \mathbb{R}^2 \) to \( \mathbb{R}^2 \) is
\[
y = \begin{bmatrix}
  -x_2 \\
  x_1 \\
\end{bmatrix} = \begin{bmatrix}
  0 & -1 \\
  1 & 0 \\
\end{bmatrix} \begin{bmatrix}
  x_1 \\
  x_2 \\
\end{bmatrix}
\]
and represents a rotation of the point with position vector \( x \) through 90° (anticlockwise looking down on the plane). Convince yourselves of this by looking at a few values of \( x \), but we will return to this below in 3.3.11.

Example 3.2.8 The matrix \( A = \begin{bmatrix} 1 & \alpha \\ 0 & 1 \end{bmatrix} \), interpreted as a mapping, leaves \( x_2 \) fixed and moves \( x_1 \) to \( x_1 + \alpha x_2 \). It is known as a shear transformation.
\[
Ax = \begin{bmatrix}
  x_1 + \alpha x_2 \\
  x_2 \\
\end{bmatrix}
\]
See Figure 3.2 where \( \alpha > 0 \). The box \( OBCD \) is transformed into the quadrilateral \( OEFD \). The range is again \( \mathbb{R}^2 \), as can very easily be seen. (Exercise 56 No.3).

3.2.3 \( Ax = Bx \) for all \( x \) if, and only if, \( A = B \)
It is obvious that if \( A = B \) then \( Ax = Bx \) for all \( x \).
Conversely, if \( Ax = Bx \) for \( x \in \mathbb{R}^n \) then \( A \) and \( B \) must have the same number \( n \) of columns and also the same number \( m \) of rows (the number of rows in \( Ax = Bx \)). Furthermore, if \( e_j \in \mathbb{R}^n \) is the \( j \)th unit vector, then by equation (3.24), \( A_{e_j} = A_{e_j} = B_{e_j} = B_{e_j} \) for \( j = 1, \ldots, n \). Hence \( A \) and \( B \) have the same columns and so \( A = B \).

3.3 Linear Transformations
Because of equation (3.31), the mapping (3.32) is said to be linear and is called a linear transformation.

In abstract terms, a mapping
\[
T : \mathbb{R}^n \rightarrow \mathbb{R}^m
\] (3.33)
is called a **linear transformation** if for all $\mathbf{u}_1$ and $\mathbf{u}_2$ in $\mathbb{R}^n$ and all scalars $s_1$ and $s_2$ it follows that

$$T (\mathbf{u}_1 s_1 + \mathbf{u}_2 s_2) = T (\mathbf{u}_1) s_1 + T (\mathbf{u}_2) s_2$$

(3.34)

An equivalent definition is the following.

The mapping (3.33) is a linear transformation if, and only if, for all $\mathbf{u}$ and $\mathbf{v}$ in $\mathbb{R}^n$ and all scalars $s$,

$$T (\mathbf{u} + \mathbf{v}) = T (\mathbf{u}) + T (\mathbf{v})$$

(3.35)

$$T (s \mathbf{u}) = s T (\mathbf{u})$$

(3.36)

For it is obvious that if (3.34) holds then so will (3.35) and (3.36). On the other hand, suppose that these last two conditions hold for the mapping $T$. Then for any $\mathbf{u}_1$ and $\mathbf{u}_2$ scalars $s_1$ and $s_2$,

$$T (\mathbf{u}_1 s_1 + \mathbf{u}_2 s_2) = T (\mathbf{u}_1 s_1) + T (\mathbf{u}_2 s_2) = T (\mathbf{u}_1) s_1 + T (\mathbf{u}_2) s_2$$

Remark 54 By putting $s = 0$ in (3.36) we see that a necessary condition for $T$ to be linear is that it maps the zero vector of $\mathbb{R}^n$ to the zero vector of $\mathbb{R}^m$. Most transformations are decidedly non-linear. Consider, for example, $T : \mathbb{R}^1 \to \mathbb{R}^1$ given by $T (x) = x^2$. Here $T (0) = 0$ but $T$ fails (3.35), since (e.g.) $T (1 + 1) = 2^2 \neq T (1) + T (1) = 2$.

### 3.3.1 An extension of equation (3.34)

The result expressed by equation (3.34) easily extends to the following:

Let $T$ be a linear transformation and suppose vectors $\mathbf{u}_1$, $\mathbf{u}_2$, ..., $\mathbf{u}_k$ in $\mathbb{R}^n$ are given as well as scalars $s_1$, $s_2$, ..., $s_k$, then

$$T (\mathbf{u}_1 s_1 + \mathbf{u}_2 s_2 + \ldots + \mathbf{u}_k s_k) = T (\mathbf{u}_1) s_1 + T (\mathbf{u}_2) s_2 + \ldots + T (\mathbf{u}_k) s_k$$

(3.37)

We use this result to show that
3.3.2 A linear transformation \( T : \mathbb{R}^n \to \mathbb{R}^m \) is defined by a unique matrix mapping \( x \to Ax \)

**Proof:**

We have seen that the mapping \( x \to Ax \) is a linear transformation. Conversely, let \( T : \mathbb{R}^n \to \mathbb{R}^m \) be a linear transformation. We show that there is a unique matrix \( A \) such that \( T(x) = Ax \) for all \( x \in \mathbb{R}^n \).

Let \( e_j \) be one of the \( n \) unit vectors (3.7). By assumption, \( T(e_j) \) is a column vector with \( m \) entries. Since \( x \in \mathbb{R}^n \) has the expression \( x = e_1 x_1 + e_2 x_2 + \ldots + e_n x_n \), we deduce from equation (3.37) that

\[
T(x) = T(e_1) x_1 + T(e_2) x_2 + \ldots + T(e_n) x_n \tag{3.38}
\]

In other words, by equation (3.23),

\[
T(x) = Ax \quad (x \in \mathbb{R}^n) \tag{3.39}
\]

Here the matrix \( A \) is defined by

\[
A \cdot j = T(e_j) \quad \text{for} \quad j = 1, \ldots, n.
\]

This is saying that \( A \) is the \( m \times n \) matrix

\[
A = \left[ T(e_1) \ T(e_2) \ \ldots \ T(e_n) \right]
\]

The matrix \( A \) in (3.39) is obviously unique.

3.3.3 A linear transformation \( T \) is completely determined by its effect \( T(e_j) \) on the unit vectors \( e_j \)

This is just a restatement of equation (3.38).

**Remark 55** Since a linear transformation is just a mapping determined by a matrix, why bother with the concept ‘linear transformation’? One reason is that our intuition often suggests that a transformation is linear.

**Example 3.3.1** In equation 1.14 of Chapter 1 we found the projection \( Q \) of a point \( P = (p_1, p_2, p_3) \) on the line \( t = (-2, -1, 1) \):

\[
q = \left( \frac{2}{3} p_1 + \frac{1}{3} p_2 - \frac{1}{3} p_3, \frac{1}{6} p_1 + \frac{1}{6} p_2 - \frac{1}{6} p_3, -\frac{1}{3} p_1 - \frac{1}{3} p_2 + \frac{1}{3} p_3 \right)
\]

Our geometric intuition strongly suggests that this should represent a linear transformation. Prove this by finding the matrix determining it.

**Solution:**

Writing \( x_1 = p_1, x_2 = p_2, x_3 = p_3 \) and using column vectors, the above projection formula becomes

\[
\begin{bmatrix}
q_1 \\
q_2 \\
q_3
\end{bmatrix} = \begin{bmatrix}
\frac{2}{3} & \frac{1}{3} & -\frac{1}{3} \\
\frac{1}{6} & \frac{1}{6} & -\frac{1}{6} \\
-\frac{1}{3} & -\frac{1}{3} & \frac{1}{3}
\end{bmatrix} \begin{bmatrix}
x_1 \\
x_2 \\
x_3
\end{bmatrix}
\]

This is the mapping \( x \to q = Ax \) and is indeed a linear transformation.

**Example 3.3.2** Find a matrix representation of the mapping which sends the point \( P \) to the foot of the perpendicular from \( P \) to the plane \( x_1 + 2x_2 + x_3 = -4 \). Show that this mapping is not linear but that the projection of \( P \) on the plane \( x_1 + 2x_2 + x_3 = 0 \) is a linear transformation. Find the matrix that reflects in this plane.
3.3. LINEAR TRANSFORMATIONS

Solution:

From equation (1.22) in Chapter 1, with \( P = (p_1, p_2, p_3) \) we found for the foot \( Q \)

\[
q = (p_1, p_2, p_3) - \frac{4 + p_1 + 2p_2 + p_3}{6} (1, 2, 1)
\]

Letting \( x_1 = p_1, x_2 = p_2, x_3 = p_3 \), and again using column vectors, this becomes

\[
\begin{bmatrix}
q_1 \\
q_2 \\
q_3
\end{bmatrix} = \begin{bmatrix}
x_1 \\
x_2 \\
x_3
\end{bmatrix} - \begin{bmatrix}
\frac{2}{3} \\
\frac{4}{3} \\
\frac{2}{3}
\end{bmatrix} + \begin{bmatrix}
-x_1 + 2x_2 + x_3 \\
-2x_1 + 2x_2 + x_3 \\
-x_1 + 2x_2 + x_3
\end{bmatrix}
\]

This cannot be linear since \( x = 0 \) does not map to \( q = 0 \).

However, from equation (1.23) we found the formula for the projection \( Q \) of \( P = (p_1, p_2, p_3) \) on the plane \( x_1 + 2x_2 + x_3 = 0 \) to be

\[
q = (p_1, p_2, p_3) - \frac{p_1 + 2p_2 + p_3}{6} (1, 2, 1)
\]

In matrix form with \( P = x \), this reads

\[
\begin{bmatrix}
q_1 \\
q_2 \\
q_3
\end{bmatrix} = \begin{bmatrix}
x_1 \\
x_2 \\
x_3
\end{bmatrix} - \begin{bmatrix}
\frac{2}{6}x_1 + \frac{2}{6}x_2 + \frac{1}{6}x_3 \\
\frac{4}{6}x_1 + \frac{2}{6}x_2 + \frac{1}{6}x_3 \\
\frac{2}{6}x_1 + \frac{2}{6}x_2 + \frac{1}{6}x_3
\end{bmatrix}
\]

\[
= \begin{bmatrix}
\frac{5}{6}x_1 - \frac{1}{3}x_2 - \frac{1}{3}x_3 \\
-\frac{1}{3}x_1 + \frac{1}{3}x_2 - \frac{1}{3}x_3 \\
-\frac{1}{3}x_1 - \frac{1}{3}x_2 + \frac{2}{3}x_3
\end{bmatrix}
\begin{bmatrix}
x_1 \\
x_2 \\
x_3
\end{bmatrix}
\]

(3.40)

The matrix that reflects in the plane \( x_1 + 2x_2 + x_3 = 0 \) is

\[
2 \begin{bmatrix}
\frac{5}{6} & -\frac{1}{3} & -\frac{1}{3} \\
-\frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\
-\frac{1}{3} & -\frac{1}{3} & \frac{2}{3}
\end{bmatrix} - \begin{bmatrix}
1 & 0 & 0 \\
0 & 1 & 0 \\
0 & 0 & 1
\end{bmatrix} = \begin{bmatrix}
\frac{2}{3} & -\frac{2}{3} & -\frac{1}{3} \\
-\frac{2}{3} & \frac{2}{3} & \frac{1}{3} \\
-\frac{1}{3} & -\frac{1}{3} & \frac{2}{3}
\end{bmatrix}
\]

This also defines a linear transformation.

Exercise 56 Partial Solutions

1. Find a matrix formula for the foot \( Q \) of the perpendicular from \( P = x \in \mathbb{R}^3 \) to the plane \( 2x + z - 3 = 0 \). Do the same for the plane \( 2x + z = 0 \), showing in the latter case that we get a linear transformation. Write down the matrix \( A \) of this transformation, as well as the matrix that reflects in the plane \( 2x + z = 0 \).

Hint 57 Refer to 1d from Exercise 16 in Chapter 1. There you found the foot \( Q \) of the perpendicular from \((p_1, p_2, p_3)\) to the plane \( 2x + z - 3 = 0 \):

\[
q = \left( \frac{1}{5}p_1 - \frac{2}{5}p_3 + \frac{6}{5}p_2, \frac{4}{5}p_3 - \frac{2}{5}p_1 + \frac{3}{5} \right)
\]

Now use appropriate columns in place of row vectors.

2. (a) Consider equation (1.13) in subsection 1.2.5 of Chapter 1 for the projection of a point on a line \( l \) that passes through the origin. Using column vectors find a matrix \( A \) such that the foot \( Q \) of the perpendicular from the point \( x \) is given by \( q = Ax \).
(b) Consider equation (1.24) from Exercise 16 in Chapter 1. Assume that the plane goes through the origin. Using column vectors find a matrix equation \( q = Ax \) for the foot \( Q \) of the perpendicular from the point \( x \) onto the plane (the projection of \( P \) onto the plane).

The matrices in 2a and 2b are called projection matrices.

(c) Specialize (a) in case the transformation is restricted to the Cartesian plane, i.e. is from \( \mathbb{R}^2 \) to \( \mathbb{R}^2 \).

3. Find the range of the shear transformation (Example 3.2.8) and prove that your statement is correct.

4. Use Exercise 20 from Chapter 1 to find linear transformations describing the following:

(a) The reflection in a line that passes through the origin.

Remark 58 Observe that this matrix actually represents a rotation through \( \pi \) radians about the given line. Contrast this with the situation when we restrict ourselves to the Cartesian plane in Exercise 63, No.8, where a reflection cannot be a rotation.

(b) The reflection in a plane that passes through the origin.

(c) Specialize (a) in case the transformation is restricted to the Cartesian plane, i.e. is from \( \mathbb{R}^2 \) to \( \mathbb{R}^2 \).

(d) What do you think the ranges of these mappings are?

5. In the \( x_1 - x_2 \) plane let \( A \) be the rotation matrix of Example 3.2.7 and suppose \( B \) is the \( 2 \times 2 \) matrix that reflects in the \( x_1 - \) axis.

(a) Find matrices \( C \) and \( D \) such that for all vectors \( x \in \mathbb{R}^2 \) we have \( Cx = B(Ax) \) and \( Dx = A(Bx) \).

(b) Show that \( C \) and \( D \) are reflection matrices and describe them geometrically.

Remark 59 In the next section, we will regard \( C \) as the product of \( B \) and \( A \), in that order: \( C = BA \). Similarly, \( D = AB \). See Example 3.3.3 and notice that \( AB \neq BA \).

6. Let \( A \) be one of the projection matrices from the previous exercises. Show geometrically why you expect that \( A(Ax) = Ax \) for all points \( x \). Now suppose that \( A \) is one of the reflection matrices. What should \( A(Ax) \) be? See also Exercise 63, No.10 below.

3.3.4 The product \( AB \) of two matrices

Let \( A \) be an \( m \times k \) matrix and \( B \) a \( k \times n \) matrix. As linear transformations,

\[
B : \mathbb{R}^n \to \mathbb{R}^k \quad \text{and} \quad A : \mathbb{R}^k \to \mathbb{R}^m.
\]

Consequently, for \( x \in \mathbb{R}^n \)

\[
x \to A(Bx) \quad \tag{3.41}
\]

defines a transformation from \( \mathbb{R}^n \) to \( \mathbb{R}^m \). See Figure 3.3. Symbolically

\[
x \quad B \quad Bx \quad A \quad A(Bx)
\]

It is essential to observe that (3.41) only makes sense if the number of columns of \( A \) equals the number of rows of \( B \) (here both are \( k \)).

Remark 60 We have to read this as: first apply \( B \) to \( x \), then apply \( A \) to the result \( Bx \). This “backward” reading is due to our traditional functional notation: For ordinary functions \( f \) and \( g \) we write the composition of \( f \) and \( g \) as \( (f \circ g)(x) = f(g(x)) \).
We will define the product $P = AB$ in such a way that

$$P \underline{x} = A (B \underline{x}) \quad (\underline{x} \in \mathbb{R}^n) \quad (3.42)$$

In other words, so that $AB$ defines the composition of the transformations $A$ and $B$.

### 3.3.5 There is one, and only one, matrix $P$ satisfying equation (3.42)

**Proof**

Suppose (3.42) holds. Put $\underline{x} = e_j$, one of the $n$ unit vectors (3.7). Then

$$P_e = P e_j = A (B e_j) = AB \bullet_j \quad (j = 1, \cdots, n) \quad (3.43)$$

In other words, for (3.42) to hold (3.43) is the only choice for $P$. On the other hand, (3.41) is a linear transformation: See No.15 in Exercise 63). Therefore its matrix can only have columns (3.43). We therefore define the matrix $AB$ as follows:

### 3.3.6 Definition

If $A$ is $m \times k$ and $B$ is $k \times n$, then

$$AB = \begin{bmatrix} AB_{\bullet_1} & AB_{\bullet_2} & \cdots & AB_{\bullet_n} \end{bmatrix} \quad (3.44)$$

Equivalently,

$$[AB]_{\bullet_j} = AB_{\bullet_j} \quad (j = 1, \cdots, n) \quad (3.45)$$

With this notation So writing $AB$ for $P$ we rewrite (3.42) as

$$(AB) \underline{x} = A (B \underline{x}) \quad (\underline{x} \in \mathbb{R}^n)$$
3.3.7 Alternative more symmetric definition of $AB$

Equation (3.44) seems biased towards columns, but that this is not the case can be seen by looking at the entries of $(AB)_{ij}$. Using (3.22) with $x = B_j$ and the notation of equation (3.1),

$$[AB]_{ij} = A_i B_j \quad (i = 1, \cdots, m; \ j = 1, \cdots, n) \quad (3.46)$$

The entry in row $i$ and column $j$ of $AB$ is the product of row $A_i$ and column $B_j$. Equation (3.46) can just as well be taken as the definition of $AB$ and in fact is the standard definition. This equation was anticipated in Exercise 47, No.10a above.

Whatever is true for columns (rows) has its exact counterpart for rows (columns). From (3.46) we see that

$$[AB]_{i\bullet} = A_i \bullet B \quad (i = 1, \cdots, m) \quad (3.47)$$

or more fully,

$$AB = \begin{bmatrix} A_1 \bullet B \\ A_2 \bullet B \\ \vdots \\ A_m \bullet B \end{bmatrix}$$

Expanding (3.46), we get using the usual convention for matrix entries and $\sum$-notation,

$$[AB]_{ij} = \sum_{r=1}^{k} a_{ir} b_{rj} = a_{i1} b_{1j} + \cdots + a_{ik} b_{kj}$$

As already remarked, this equation only makes sense if the number of columns of $A$ equals the number of rows of $B$. We may write

$$(AB)_{m \times n} = A_{m \times k} B_{k \times n}$$

**Example 3.3.3** Let $B$ be the reflection matrix of Example 3.2.1 and $A$ that of the rotation matrix $A$ of Example 3.2.7. Find the matrix of (i) first rotating then reflecting and (ii) the other way round, first reflecting then rotating.

See No.5 in Exercise 56.

**Solution (i)**

If we first rotate through $90^\circ$ then reflect in the $x_1$-axis we get the product

$$BA = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$$

**Solution (ii)**

If we first reflect and then rotate we obtain

$$AB = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

**Example 3.3.4** In Exercise 47, No.10a,

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

Find $AB$. 

3.3. LINEAR TRANSFORMATIONS

Solution:

From (3.46),

\[
AB = \begin{bmatrix}
2 & -9 \\
8 & -5
\end{bmatrix}
\begin{bmatrix}
2 & -9 \\
2 & 4
\end{bmatrix}
\begin{bmatrix}
2 & -9 \\
3 & 1
\end{bmatrix}
\begin{bmatrix}
1 \\
-1
\end{bmatrix}
= \begin{bmatrix}
61 & -32 & -65 \\
19 & 10 & 4
\end{bmatrix}
\]

Example 3.3.5 Find \( yB \), where \( y = \begin{bmatrix} y_1 & y_2 \end{bmatrix} \) and \( B \) is as in the previous example:

\[
y = \begin{bmatrix} y_1 & y_2 \end{bmatrix}
\begin{bmatrix}
8 & -1 \\
-5 & 7
\end{bmatrix}
\]

Solution:

\[
yB = \begin{bmatrix} y_1 & y_2 \end{bmatrix}
\begin{bmatrix}
8 & 2 & -1 \\
-5 & 4 & 7
\end{bmatrix}
= \begin{bmatrix}
8y_1 - 5y_2 & 2y_1 + 4y_2 & -y_1 + 7y_2
\end{bmatrix}
\]

Example 3.3.6 Express equation (3.48) as a linear combination of rows of \( B \).

Solution:

In (3.23) we found that the product \( Ax \) is a linear combination of the columns of \( A \) with coefficients \( x_k \). In exactly the same way we find the product (3.48) to be a linear combination of the rows of \( B \) with coefficients \( y_1 \) and \( y_2 \):

\[
y_1B_1 + y_2B_2 = y_1 \begin{bmatrix} 8 & 2 & -1 \end{bmatrix} + y_2 \begin{bmatrix} -5 & 4 & 7 \end{bmatrix}
= \begin{bmatrix}
8y_1 - 5y_2 & 2y_1 + 4y_2 & -y_1 + 7y_2
\end{bmatrix}
= yB
\]

3.3.8 For a row \( y \), the product \( yB \) is a linear combination of the rows of \( B \) with coefficients \( y_i \).

Let \( B \) be a \( k \times n \) matrix and \( y = [y_1, \cdots, y_k] \) a row vector with \( k \) entries. Then

\[
yB = y_1B_1 + \cdots + y_kB_k
\]

The reasoning is just as in the case of (3.23) except that now we are multiplying on the left of \( B \) by a row \( y \) and taking a linear combination of rows of \( B \). Like equation (3.22) we have

\[
yB = \begin{bmatrix} yB_1 & yB_2 & \cdots & yB_n \end{bmatrix}
\]

which is just a special case of equation (3.44).
3.3.9 The rows of $AB$ as linear combinations of rows of $B$ and the columns of $AB$ as linear combinations of columns of $A$

Let $A = [a_{ir}]$ be $m \times k$ and $B = [b_{rj}]$ be $k \times n$. In the product (3.46), we have

1. $[AB]_i = a_{i1}B_1 + a_{i2}B_2 + \cdots + a_{ik}B_k \quad (i = 1, \cdots, m) \quad (3.50)$

2. $[AB]_j = A_1b_{1j} + A_2b_{2j} + \cdots + A_kb_{kj} \quad (3.51)$

To see equation (3.50), use equations (3.47) and (3.49) with $y = A_i$.

In words: “Row $i$ of $AB$ is a linear combination of the rows of $B$ with coefficients taken from row $i$ of $A$.”

To see equation (3.51) use equations (3.45) and our earlier result equation (3.23) with $x = B_j$.

In words: “Column $j$ of $AB$ is a linear combination of the columns of $A$ with coefficients taken from column $j$ of $B$.”

The above examples illustrate these statements, as do the following.

Example 3.3.7 Find $AB$ in two ways: (i) in terms of linear combinations of rows of $B$ and (ii) in terms of linear combinations of columns of $A$, given that

$A = \begin{bmatrix} 2 & 1 \\ -1 & 3 \end{bmatrix}$ and $B = \begin{bmatrix} 5 & 4 \\ -7 & 0 \end{bmatrix}$

Solution:

(i) In terms of the rows of $AB$:

$AB = \begin{bmatrix} A_1B \\ A_2B \end{bmatrix} = \begin{bmatrix} 2 & 1 \\ -1 & 3 \end{bmatrix} \begin{bmatrix} 5 & 4 \\ -7 & 0 \end{bmatrix} = \begin{bmatrix} 2 \begin{bmatrix} 5 & 4 \end{bmatrix} + 1 \begin{bmatrix} -7 & 0 \end{bmatrix} \\ (-1) \begin{bmatrix} 5 & 4 \end{bmatrix} + 3 \begin{bmatrix} -7 & 0 \end{bmatrix} \end{bmatrix}$

$= \begin{bmatrix} 3 & 8 \\ -26 & -4 \end{bmatrix}$

(ii) In terms of the columns of $AB$:

$AB = \begin{bmatrix} AB_1 \\ AB_2 \end{bmatrix} = \begin{bmatrix} 2 & 1 \\ -1 & 3 \end{bmatrix} \begin{bmatrix} 5 \\ -7 \end{bmatrix} = \begin{bmatrix} 2 \begin{bmatrix} 5 \\ -7 \end{bmatrix} + 1 \begin{bmatrix} 1 \\ 0 \end{bmatrix} \\ (-1) \begin{bmatrix} 5 \\ -7 \end{bmatrix} + 3 \begin{bmatrix} 1 \\ 0 \end{bmatrix} \end{bmatrix}$

$= \begin{bmatrix} 3 & 8 \\ -26 & -4 \end{bmatrix}$

Example 3.3.8 If $A$ is a $3 \times 3$ matrix, find matrices $B$ and $C$ such that

$BA = \begin{bmatrix} -5A_3 + 67A_3 + 92A_3 \\ 9A_1 - 4A_1 \\ 8A_2 + 21A_1 \end{bmatrix}$
and
\[ AC = \begin{bmatrix} 7A_{\bullet 1} - 37A_{\bullet 2} & 19A_{\bullet 2} + 17A_{\bullet 1} - 7A_{\bullet 3} \end{bmatrix} \]

(\(AC\) has two columns).

**Solution:**

\[
B = \begin{bmatrix} 67 & -5 & 92 \\ -4 & 0 & 9 \\ 21 & 8 & 0 \end{bmatrix} \quad \text{and} \quad C = \begin{bmatrix} 7 & 17 \\ -37 & 19 \\ 0 & -7 \end{bmatrix}
\]

**Example 3.3.9** What is the matrix of the transformation \(T: \mathbb{R}^3 \rightarrow \mathbb{R}^3\) that first projects onto the plane \(x_1 + 2x_2 + x_3 = 0\) as in (3.40) and then projects onto the \(x_1 - x_2\) plane as in Example 3.2.3?

**Solution:**

The required matrix is the product
\[
\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 5 & -\frac{1}{3} & -\frac{1}{3} \\ -\frac{7}{6} & \frac{1}{3} & \frac{5}{3} \\ -\frac{7}{6} & \frac{1}{3} & -\frac{5}{3} \end{bmatrix} = \begin{bmatrix} 5 & -\frac{1}{3} & -\frac{1}{3} \\ -\frac{7}{3} & \frac{1}{3} & -\frac{7}{3} \end{bmatrix}
\]

### 3.3.10 The Associative law: \((AB)C = A(BC)\)

If \(A\) is \(m \times k\) and \(B\) is \(k \times n\), our whole inspiration (3.42) for the definition of the product \(P = AB\) is that \((AB)\mathbf{x} = A(B\mathbf{x})\) should hold for all \(\mathbf{x} \in \mathbb{R}^n\). This is a special case of the associative law for multiplication of matrices, but in fact gives us the general law directly.

**Statement and proof of the associative law**

Let \(A\) be an \(m \times k\) matrix, \(B\) a \(k \times n\) matrix and \(C\) an \(n \times p\) matrix. Then
\[
(AB)C = A(BC) \quad (3.52)
\]

Considered as mappings, both sides of (3.52) have the same meaning: First apply \(C\), then \(B\), then \(A\). Therefore both sides are equal as matrix products.

**Powers of \(A\) if \(A\) is square.** Let \(A\) be \(m \times m\). Then \((AA)A\) and \(A(AA)\) are both defined and are equal by (3.52), so we can write \(A^n\) for this product. Similarly, if \(n \geq 1\) is an integer we have \(A^n = AA \cdots A\ (n \text{ times})\), where the bracketing does not matter. By convention, \(A^0 = I_m\) (the \(m \times m\) identity matrix).

**Remark 61** We have given a motivated and especially simple proof of the associative law (3.52). The usual proof given in textbooks is the following:

\[
[A(BC)]_{ij} = \sum_{r=1}^{k} [A]_{ir} [BC]_{rj} = \sum_{r=1}^{k} [A]_{ir} \left( \sum_{s=1}^{n} [B]_{rs} [C]_{sj} \right)
\]

\[
= \sum_{r=1}^{k} \sum_{s=1}^{n} [A]_{ir} [B]_{rs} [C]_{sj} = \sum_{s=1}^{n} \left( \sum_{r=1}^{k} [A]_{ir} [B]_{rs} \right) [C]_{sj}
\]

\[
= \sum_{s=1}^{n} [AB]_{is} [C]_{sj} = [(AB)C]_{ij}
\]
3.3.11 Rotations are linear transformations

Let $X$ be a point in the Cartesian plane $\mathbb{R}^2 = \mathbb{R}^2 \times \mathbb{R}$ and suppose that $R$ rotates $OX = x$ through the angle $\theta$ about the origin \textit{anticlockwise} looking down the $x_3 -$ axis towards $O$. This is a \textit{clockwise} rotation of $\theta$ looking from the origin $O$ along the positive $x_3 -$axis. Let $R(x)$ be the effect of $R$ on $x$.

We will now see that $R : \mathbb{R}^2 \to \mathbb{R}^2$ is a linear transformation.

In Figure 3.5, $A$ and $B$ are any two points in the plane with position vectors $a$ and $b$. The point $C$ has position vector $c = a + b$ and the quadrilateral $OACB$ is a parallelogram.

The points $A'$, $B'$ and $C'$ are the respective results of rotating about $O$ the points $A$, $B$ and $C$ through the angle $\theta$. In other words, $A'$, $B'$ and $C'$ have position vectors $R(a)$, $R(b)$ and $R(a + b)$ respectively. The parallelogram $OACB$ and this shows that

$$R(a + b) = R(a) + R(b)$$

Thus the transformation $R$ satisfies the first condition in equation (3.35) for linearity. It is pretty obvious that if we multiply $a$ by a scalar $s$ that $R(a)$ will be multiplied by the same factor, i.e. that $R(as) = R(a)s$, so that that the second condition for linearity as expressed in equation (3.36) also holds. We have shown that $R : \mathbb{R}^2 \to \mathbb{R}^2$ is a linear transformation.

3.3.12 The matrix $R_\theta$ of a rotation through $\theta$

Let $R_\theta$ be the matrix of the above rotation transformation $R$ that rotates through the angle $\theta$. Then

$$R_\theta = \begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix}$$

To see this, note that from Figure 3.5,

$$R_\theta \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} \cos \theta \\ \sin \theta \end{bmatrix} \quad \text{and} \quad R_\theta \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} -\sin \theta \\ \cos \theta \end{bmatrix}$$
Since these are the first and second columns of $R_{\theta}$ respectively, this proves (3.53).

Example 3.3.10 Write down the linear transformations that rotate $\pi/3$ and $2\pi/3$ radians about the origin in the Cartesian plane.

Solution:

$$R_{\pi/3} = \begin{bmatrix} \cos \frac{\pi}{3} & -\sin \frac{\pi}{3} \\ \sin \frac{\pi}{3} & \cos \frac{\pi}{3} \end{bmatrix} = \begin{bmatrix} \frac{1}{2} & -\frac{\sqrt{3}}{2} \\ \frac{\sqrt{3}}{2} & \frac{1}{2} \end{bmatrix}$$

$$R_{2\pi/3} = \begin{bmatrix} \cos \frac{2\pi}{3} & -\sin \frac{2\pi}{3} \\ \sin \frac{2\pi}{3} & \cos \frac{2\pi}{3} \end{bmatrix} = \begin{bmatrix} -\frac{1}{2} & -\frac{\sqrt{3}}{2} \\ \frac{\sqrt{3}}{2} & -\frac{1}{2} \end{bmatrix}$$

3.3.13 The addition formulae for sine and cosine

We note that rotation through $\phi$ followed by the rotation through $\theta$ is equivalent to the rotation through $\theta + \phi$, that is,

$$R_{\theta}R_{\phi} = R_{\theta + \phi}$$

This means

$$\begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix} \begin{bmatrix} \cos \phi & -\sin \phi \\ \sin \phi & \cos \phi \end{bmatrix} = \begin{bmatrix} \cos (\theta + \phi) & -\sin (\theta + \phi) \\ \sin (\theta + \phi) & \cos (\theta + \phi) \end{bmatrix}$$

(3.54)

Multiplying the two matrices on the left gives the addition formulae for sine and cosine:

$$\begin{bmatrix} \cos \theta \cos \phi - \sin \theta \sin \phi & -\cos \theta \sin \phi - \sin \theta \cos \phi \\ \sin \theta \cos \phi + \cos \theta \sin \phi & \cos \theta \cos \phi - \sin \theta \sin \phi \end{bmatrix} = \begin{bmatrix} \cos (\theta + \phi) & -\sin (\theta + \phi) \\ \sin (\theta + \phi) & \cos (\theta + \phi) \end{bmatrix}$$

Remark 62 Let $X \neq O$ be a point in three-dimensional space. Looking along $OX$ we can rotate clockwise through $\theta$ and this defines a transformation $T : \mathbb{R}^3 \to \mathbb{R}^3$. Much the same argument shows that $T$ is linear. However, the formula for its matrix is more complicated.

Exercise 63 1. Find all possible products of the following matrices, whenever the product is defined, and when this is the case, express the rows and columns of the product as linear combinations, as in example 3.3.7.
CHAPTER 3. LINEAR TRANSFORMATIONS AND MATRICES

\[
A = \begin{bmatrix}
-1 & 3 \\
2 & 1
\end{bmatrix}, \quad B = \begin{bmatrix}
-2 & 5 \\
4 & -7 \\
0 & 3
\end{bmatrix}, \quad C = \begin{bmatrix}
2 & 0 & -5 \\
-4 & 3 & 11
\end{bmatrix},
\]

\[
D = \begin{bmatrix}
-3 & 4 \\
-1 & 6
\end{bmatrix}, \quad E = \begin{bmatrix}
2 & -1
\end{bmatrix}
\]

2. Given that \(B\) is a \(3 \times 3\) matrix, find by inspection matrices \(A\) and \(C\) such that

\[
AB = \begin{bmatrix}
B_{11} - 2B_{31} \\
2B_{21} + 5B_{31} \\
B_{11} + 4B_{21} + 7B_{31} \\
-3B_{11}
\end{bmatrix}
\]

\[
BC = \begin{bmatrix}
B_{11}(-2) + B_{12}(3) + B_{13}B_{21} \\
B_{21} - B_{11}
\end{bmatrix}
\]

Note that \(BC\) has two columns.

3. If \(a \in \mathbb{R}^{1 \times n}\) is a row and \(b \in \mathbb{R}^{m \times 1}\) is a column, is the product \(X = ba\) defined? If so, describe the entries \([X]_{ij}\).

4. If \(AB = C\) and two of \(A\), \(B\), \(C\) are square then so is the other. True or false?

5. Let \(A\) be an \(m \times n\) matrix and \(I_m\) and \(I_n\) identity matrices (see equation (3.8)). Show that

\[
I_m A = A = A I_n
\]

Conclude that if \(A\) is \(n \times n\) and \(I = I_n\) that \(IA = AI = A\).

6. In No.10 of Exercise 5 of Chapter 1 you found the formula for the projection of a point \((p_1, p_2, p_3)\) on the line \((-2t, -t, 5t)\). Find its matrix \(B\) as a linear transformation. In No.1 of Exercise 56 above you found the matrix \(A\) of the projection on the plane \(2x + z = 0\). Find

\(a\) the matrix of the transformation that first projects on the plane \(2x + z = 0\) then on the line \((-2t, -t, 5t)\).

\(b\) the matrix of the transformation that first projects on the line \((-2t, -t, 5t)\) then on the plane \(2x + z = 0\).

The answer is

\[
AB = \frac{1}{30} \begin{bmatrix}
\frac{1}{2} & 0 & -\frac{2}{5} \\
0 & 1 & 0 \\
-\frac{2}{5} & 0 & \frac{4}{5}
\end{bmatrix} \begin{bmatrix}
4 & 2 & -10 \\
2 & 1 & -5 \\
-10 & -5 & 25
\end{bmatrix} = \begin{bmatrix}
\frac{4}{25} & \frac{2}{25} & -\frac{2}{25} \\
\frac{1}{5} & \frac{3}{5} & \frac{2}{5} \\
-\frac{1}{25} & -\frac{1}{25} & \frac{4}{5}
\end{bmatrix}
\]

7. \(a\) complete the proof of the addition formulae for sine and cosine by multiplying out equation (3.54).

\(b\) Check by direct multiplication that \(R_\theta R_{-\theta} = I_2\).

8. \(a\) Let \(\ell\) be a line in the Cartesian plane making an angle \(\theta\) with the positive \(x_1\) axis. Find the projection matrix \(P_\theta\) on \(\ell\) and deduce that the matrix \(M_\theta\) which reflects in \(\ell\) is given by

\[
M_\theta = \begin{bmatrix}
\cos 2\theta & \sin 2\theta \\
\sin 2\theta & -\cos 2\theta
\end{bmatrix}
\]

Observe that, in spite of superficial appearances, \(M_\theta\) is not a rotation matrix. Why?

\(b\) Let \(\phi\) and \(\theta\) be two angles and \(\alpha = \phi - \theta\). Prove that \(M_\phi M_\theta = R_\alpha\) in two ways:
3.3. LINEAR TRANSFORMATIONS

![Figure 3.6](image)

\[ \alpha = \phi - \theta \]

9. Following the corkscrew rule, find matrices that represent rotations through \( \theta \) about the \( x-, y- \) and \( z- \) axes.

**Hint 65** Recall the corkscrew rule: looking along the specified axis, rotate clockwise through \( \theta \). The required rotation about the \( z- \) axis is

\[
\begin{bmatrix}
\cos \theta & -\sin \theta & 0 \\
\sin \theta & \cos \theta & 0 \\
0 & 0 & 1
\end{bmatrix}
\]

10. (Compare Exercise 56, No.6 above). Let \( M \) be the matrix of a reflection in a plane going through the origin and \( P \) the matrix of a projection on the same plane. Explain geometrically why you expect the following to be true. Draw diagrams!

(a) \( M^2 = I_3 \)
(b) \( P^2 = P \)
(c) \( MP = P = PM \)

Verify these results for the plane \( 2x + z = 0 \) of No.1 of Exercise 56 above.

For analytical proofs of these results see No.17 below.

11. If the matrix product \( AB \) is defined, show that so is \( B^T A^T \) and that \( (AB)^T = B^T A^T \).

**Hint 66** \( [(AB)^T]_{ji} = [AB]_{ij} = A_i B_j \). Now use equation (3.13) and Exercise 47 No.6.
12. Prove that the composite transformation \( \mathbf{x} \rightarrow A(B\mathbf{x}) \) of (3.41) is indeed linear. Can you see why a similar result holds for general linear transformations?

**Hint 67** Consider \( A[B(\mathbf{us} + \mathbf{vt})] = A[(B\mathbf{u})\mathbf{s} + (B\mathbf{v})\mathbf{t}] = \cdots \)

\[ A[B(\mathbf{us} + \mathbf{vt})] = A[(B\mathbf{u})\mathbf{s} + (B\mathbf{v})\mathbf{t}] = [A(B\mathbf{u})]\mathbf{s} + [A(B\mathbf{v})]\mathbf{t} \]

13. (“Linear combinations of linear combinations” theorem)

(a) In Exercise 47, No. 8c you anticipated the following result: Let \( \mathbf{x} \) be a column vector with \( m \) entries that depends linearly on \( a_1, \ldots, a_k \) and suppose that each vector \( a_j \) depends linearly on \( b_1, b_2, \ldots, b_n \). Then \( \mathbf{x} \) depends linearly on \( b_1, b_2, \ldots, b_n \). Using matrix multiplication give a short proof of this statement. It should be clear that the same result must be true if the vectors are rows, but give a modified matrix proof.

(b) Interpret this result in terms of the span concept.

**Hint 68** Let the \( m \times k \) matrix \( A \) have columns \( a_1, \ldots, a_k \) and let the \( m \times n \) matrix \( B \) have columns \( b_1, \ldots, b_n \). By assumption, there is an \( n \times k \) matrix \( C \) such that \( A = BC \). Since \( \mathbf{x} = A\mathbf{u} \) for some column vector \( \mathbf{u} \) with \( k \) entries, the result follows. Fill in the remaining details.

For the row case, think in terms of the transposes of the above matrices.

14. Show that the column case of No. 13 can also be proved using \( \sum \cdot \cdot \cdot \) notation as follows: Write \( \mathbf{x} = \sum_{j=1}^{k} a_j \mathbf{u}_j \) and \( \mathbf{x}_j = \sum_{i=1}^{n} b_{ij} \mathbf{v}_{ij} \) for \( j = 1, \ldots, n \). Then expand and rearrange the sum

\[ \mathbf{x} = \sum_{j=1}^{k} \left( \sum_{i=1}^{n} b_{ij} \mathbf{v}_{ij} \right) \mathbf{u}_j \]

15. Prove that the composite transformation \( \mathbf{x} \rightarrow A(B\mathbf{x}) \) of (3.41) is indeed linear. Can you see why a similar result holds for general linear transformations?

**Hint 69** Consider \( A[B(\mathbf{us} + \mathbf{vt})] = A[(B\mathbf{u})\mathbf{s} + (B\mathbf{v})\mathbf{t}] = \cdots \)

\[ A[B(\mathbf{us} + \mathbf{vt})] = A[(B\mathbf{u})\mathbf{s} + (B\mathbf{v})\mathbf{t}] = [A(B\mathbf{u})]\mathbf{s} + [A(B\mathbf{v})]\mathbf{t} \]

16. (a) If the matrix sum \( A + B \) and the matrix product \( C(A + B) \) are defined, show that \( C(A + B) = CA + CB \).

(b) If the matrix sum \( A + B \) and the matrix product \( (A + B)C \) are defined, show that \( (A + B)C = AC + BC \).

(c) If \( AB \) is defined, and \( s \) is a scalar, show \( s(AB) = (sA)B \) and \( (AB)s = A(Bs) \) and that both are equal. Conclude with statements and proofs of more general results than (a) and (b).

**Hint 70** To show (a) use the definition of matrix multiplication and equation (3.14):

\[ [C(A + B)]_{ij} = C_{ij} [A + B]_{ij} = C_{ij} (A_{ij} + B_{ij}) = C_{ij} A_{ij} + C_{ij} B_{ij} = \cdots \]

If \( C \) is \( m \times k \) and \( A \) and \( B \) are \( k \times n \), you may prefer to proceed directly:

\[ [C(A + B)]_{ij} = \sum_{r=1}^{k} c_{ir} [A + B]_{rj} = \sum_{r=1}^{k} c_{ir} (a_{rj} + b_{rj}) = \cdots \]
17. * 

(a) Let \(C\) be an \(n \times n\) matrix satisfying \(C^2 = I\) where \(I = I_n\).

Show that \((I - C)^2 = I - C\) and that \((2C - I)^2 = I\).

The rest of the question refers to No.10 above and to Exercise 56, No.2a, No.2b and No.4.

(b) Let \(P\) be the projection matrix of Exercise 56, No.2a and write the vector \(u\) as a column vector \(\begin{bmatrix} u_1 & u_2 & u_3 \end{bmatrix}^T\).

Show that

i. \(P = \frac{1}{||u||^2} uu^T\),

ii. \(P^2 = P\) using matrix multiplication,

iii. the matrix \(M = 2P - I\) reflecting in the line satisfies \(M^2 = I\).

(c) Let \(P\) be the projection matrix of Exercise 56, No.2b onto a plane through the origin with normal \(n = \begin{bmatrix} n_1 & n_2 & n_3 \end{bmatrix}^T\).

Show that

i. \(P = I - \frac{1}{||n||^2} nn^T\),

ii. \(P^2 = P\) using matrix multiplication,

iii. the matrix \(M = 2P - I\) reflecting in the plane satisfies \(M^2 = I\).

(d) Complete analytical proofs of No.10 above.

Hint 71 Note that we are forming matrix products of a column and a row as in No.3. Also make use of No.17a.

3.4 Linear independence

In \(\mathbb{R}^2\) let \(a\) and \(b\) be two column vectors. If if one depends linearly on the other, say \(a = bs\) for a scalar \(s\), we say they are linearly dependent. Otherwise the vectors are linearly independent: neither depends on the other. In that case, if some linear combination \(as + bt = 0\), then both \(s\) and \(t\) have to be zero. For if (say) \(t \neq 0\), then \(b = a \left( -\frac{s}{t} \right)\), and \(b\) would depend on \(a\).

Similarly, in \(\mathbb{R}^3\) three column vectors \(a, b\) and \(c\) are linearly dependent if one of them is a linear combination of the other two. This will be the case if there are numbers \(s, t, r\) not all zero such that \(as + bt + ct = 0\). For example, if \(s \neq 0\), then \(a\) will depend linearly on \(b\) and \(c\). They are linearly independent if no vector depends linearly on the other two. In that case we can only have \(as + bt + cr = 0\) if \(s = t = r = 0\).

Geometrically let \(a, b\) and \(c\) be linearly independent and let the tails of these vectors are placed at the origin \(O\). Then \(O\) and the heads of \(a, b\) and \(c\) do not lie on a plane. For more details on the geometric meaning of linear independence see subsection 3.4.11 below.

We now generalize and formalize these ideas as follows:

3.4.1 Column independence and column dependence

Definition

Let \(\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_n\) be a list of \(n\) column vectors each with \(m\) entries. Form the \(m \times n\) matrix \(A\) with these vectors as columns:

\[
A = \begin{bmatrix} a_1 & a_2 & \cdots & a_n \end{bmatrix}
\]
If there is a non-trivial solution \( \mathbf{x} \neq 0 \) to the homogeneous system

\[
A\mathbf{x} = 0
\]  

(3.55)

the vectors \( \mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_n \) are said to be linearly dependent, otherwise they are linearly independent.

To put it another way, the columns of the \( m \times n \) matrix \( A \) are linearly independent if, and only if for all column vectors \( \mathbf{x} \) with \( n \) entries

\[
A\mathbf{x} = 0 \quad \text{implies} \quad \mathbf{x} = 0
\]

(3.56)

To say the same thing once more, if we can find a linear combination such that

\[
\mathbf{a}_1x_1 + \mathbf{a}_2x_2 + \cdots + \mathbf{a}_nx_n = 0
\]

(3.57)

in which one or more of the coefficients \( x_j \) are not zero, then the vectors \( \mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_n \) are linearly dependent (there is a non-trivial linear relationship among them), otherwise they are independent.

3.4.2 An equivalent way to see independence/dependence

Let \( \mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_n \) be column vectors each with \( m \) entries.

If there is only one vector, i.e. \( n = 1 \), then \( \mathbf{a}_1 \) is dependent if, and only if it is the zero vector. If \( n \geq 2 \), the vectors are linearly dependent if, and only if, one of them depends linearly on the others.

Proof:
The proof of the case \( n = 1 \) is left to you. See Exercise 78, No.1a.

Let \( n \geq 2 \) and suppose the vectors are linearly dependent. Then (say) \( x_1 \neq 0 \) in equation 3.57.

In that case \( \mathbf{a}_1 \) depends linearly on \( \mathbf{a}_2, \ldots, \mathbf{a}_n \):

\[
\mathbf{a}_1 = \mathbf{a}_2 \left( -\frac{x_2}{x_1} \right) + \cdots + \mathbf{a}_n \left( -\frac{x_n}{x_1} \right)
\]

Similarly, if \( x_j \neq 0 \), then \( \mathbf{a}_j \) depends linearly on the vectors \( \mathbf{a}_i \) for \( i \neq j \).

Conversely, if (say) there are scalars \( s_2, \ldots, s_n \) such that

\[
\mathbf{a}_1 = \mathbf{a}_2s_2 + \cdots + \mathbf{a}_ns_n
\]

Then

\[
\mathbf{a}_1 + \mathbf{a}_2 (-s_2) + \cdots + \mathbf{a}_n (-s_n) = 0
\]

Since the coefficient of \( \mathbf{a}_1 \) is \( 1 \neq 0 \), the vectors are linearly dependent, as required.

For a slightly refined version of this theorem, see Exercise 78, No.1(c)v.

Independence/dependence of row vectors

The definition of linear independence for a list of row vectors is almost identical to that for column vectors. See 3.4.12 below.

3.4.3 Some examples of independence/dependence

The most elementary example of linear independence

The simplest example of linearly independent column vectors is afforded by one or more columns of the \( n \times n \) identity matrix \( \mathbf{I} = \mathbf{I}_n \). Since \( \mathbf{I}\mathbf{x} = \mathbf{x} \), we can only have \( \mathbf{I}\mathbf{x} = 0 \) if \( \mathbf{x} = 0 \). Slightly more complicated is
3.4. LINEAR INDEPENDENCE

Example 3.4.1 Consider the matrix (3.25) of example 3.1.6:

\[
A = \begin{bmatrix}
2 & 5 & -11 \\
-3 & -6 & 12 \\
4 & 7 & -13
\end{bmatrix}
\]

A row-echelon was found in equation (3.26):

\[
A' = \begin{bmatrix}
2 & 5 & -11 \\
0 & 1 & -3 \\
0 & 0 & 0
\end{bmatrix}
\]

The columns of \( A \) are linearly dependent, as can be seen from (3.27) by letting \( t \neq 0 \), say putting \( t = 1 \).

The columns of the matrix (3.29) in example 3.1.7 are linearly independent since \( A\mathbf{x} = 0 \) has only the trivial solution \( \mathbf{x} = 0 \).

The columns of \( A \) in Exercise 35 No.2b are linearly independent.

3.4.4 If a matrix has more columns than rows, its columns are linearly dependent; if more rows than columns then its rows are linearly dependent

This is just a rephrasing of the result of subsection 2.2.15 in Chapter 2: A homogeneous system with more unknowns than equations has a non-trivial solution.

The row case is quite similar.

3.4.5 Testing for linear independence of some columns of \( A \)

Suppose that \( A \) has seven columns and we wish to test if the columns of \([A_2, A_5, A_6]\) are linearly independent. This is the same as considering all vectors \( \mathbf{x} \) with \( x_1 = x_3 = x_4 = 0 \) and testing the implication (3.56). The same principle obviously applies to any selection of columns of \( A \).

3.4.6 Column-rank and row-rank

If \( \mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_n \) are column vectors, their column-rank is the maximum number of vectors \( \mathbf{a}_j \) that are linearly independent. This is also the column-rank of the matrix \( A = \begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \cdots & \mathbf{a}_n \end{bmatrix} \).

Row-rank The row-rank of a matrix \( A \) is defined similarly, except that ‘rows’ replace ‘columns’. See 3.4.12 below for more details.

Given a list of vectors of the same size (either rows or columns), it should be clear what is be meant by their rank: it is the maximum number of linearly independent vectors among them.

3.4.7 Elementary row operations preserve column independence and dependence

Consider a system \( A\mathbf{x} = \mathbf{b} \) of \( m \) equations for \( n \) unknowns. Let \([A', b']\) be the array after performing elementary row operations. We saw in the last chapter that \( A\mathbf{x} = \mathbf{b} \) holds if, and only if, \( A'\mathbf{x} = b' \) holds:

\[
b = A_1 x_1 + \cdots + A_n x_n \quad \text{if, and only if,} \quad b' = A'_1 x_1 + \cdots + A'_n x_n
\]
Now for the homogeneous system (3.55), \( \mathbf{b} = \mathbf{b}' = \mathbf{0} \), and any dependency relation among the columns of \( A \) also holds for \( A' \) and vice-versa:

\[
A_{\bullet 1}x_1 + \cdots + A_{\bullet n}x_n = 0 \quad \text{if, and only if,} \quad A'_{\bullet 1}x_1 + \cdots + A'_{\bullet n}x_n = 0
\]

In particular, certain columns of \( A \) are linearly independent if, and only if, the same columns of \( A' \) are linearly independent. In particular, the column ranks of \( A \) and \( A' \) are the same.

Also, column \( A_{\bullet j} \) is a linear combination of other columns of \( A \) if, and only if, \( A'_{\bullet j} \) is a linear combination of the corresponding columns of \( A' \) with the same coefficients.

For these reasons, questions regarding dependence or independence of the columns of \( A \) are best answered by considering the row reduced echelon form (row REF) of \( A \).

### 3.4.8 The column-rank of a matrix is the number of non-zero rows in a row echelon form

**Proof**

Let \( k \) be the number of non-zero rows in an echelon form of the \( m \times n \) matrix \( A \) and suppose that \( A'' \) is its row REF. Recall that the pivot columns (see 2.2.8) of \( A'' \) are columns of the identity matrix \( I_m \), which are certainly linearly independent. Hence the column-rank of \( A'' \) is at least \( k \). The \( k \times n \) matrix of the non-zero rows of \( A'' \) cannot have column-rank more than \( k \) by 3.4.4. Since the other rows of \( A'' \) are zero, the column rank of \( A'' \) is exactly \( k \). This is also the column-rank of \( A \).

### 3.4.9 Every column of \( A \) is a linear combination of the linearly independent pivot columns of \( A \)

This is because the pivot columns of \( A'' \), when restricted to the non-zero rows, are the columns of the identity matrix \( I_k \).

The statement says that the pivot columns span the column space.

**Example 3.4.2** Consider example 3.1.6 once more. Can we find a set of linearly independent columns of \( A \) in (3.25) such that every column of \( A \) is a linear combination of these independent columns? What is the column-rank of \( A \)? Illustrate the statements of 3.4.8 and 3.4.9.

**Solution:**

The original matrix (3.28) is

\[
A = \begin{bmatrix}
2 & 5 & -11 \\
-3 & -6 & 12 \\
4 & 7 & -13 \\
\end{bmatrix}
\]

The echelon form 3.26 is

\[
\begin{bmatrix}
2 & 5 & -11 \\
0 & 1 & -3 \\
0 & 0 & 0 \\
\end{bmatrix}
\]

shows that first two columns of \( A \) are linearly independent. The row REF shows things even more clearly:

\[
\begin{bmatrix}
1 & 0 & 2 \\
0 & 1 & -3 \\
0 & 0 & 0 \\
\end{bmatrix}
\]

Every column of \( A \) is a linear combination of the first two, which are pivot columns. The column rank is 2.
Example 3.4.3  Let

\[ A = \begin{bmatrix}
  -1 & -4 & -3 & 2 \\
  3 & 4 & 1 & 2 \\
  2 & 9 & 7 & -5 \\
  -5 & -1 & 4 & -9
\end{bmatrix} \]

Illustrate 3.4.7, 3.4.8 and 3.4.9.

Solution:

We first reduce the matrix \( A \) to row REF:

\[
\begin{align*}
-1 & -4 & -3 & 2 \\
0 & -8 & -8 & 8 & 3R_1 + R_2 \\
0 & 1 & 1 & -1 & 2R_1 + R_3 \\
0 & 19 & 19 & -19 & -5R_1 + R_4 \\
-1 & -4 & -3 & 2 \\
0 & -1 & -1 & 1 & \frac{1}{2}R_2 \\
0 & 1 & 1 & -1 & \frac{1}{2}R_4 \\
-1 & -4 & -3 & 2 \\
0 & 1 & 1 & -1 & -R_2 \\
0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 \\
-1 & 0 & 1 & -2 & 4R_2 + R_1 \\
0 & 1 & 1 & -1 \\
0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0
\end{align*}
\]

This gives the row REF \( A'' \) of \( A \) as

\[ E = \begin{bmatrix}
  1 & 0 & -1 & 2 \\
  0 & 1 & 1 & -1 \\
  0 & 0 & 0 & 0 \\
  0 & 0 & 0 & 0
\end{bmatrix} \]

The pivot columns of \( E \) are \( E_{\bullet 1} \) and \( E_{\bullet 2} \) and they are linearly independent, and therefore so are \( A_{\bullet 1} \) and \( A_{\bullet 2} \). The column-rank of \( A \) is 2. Clearly, every column of \( E \) is a linear combination of \( E_{\bullet 1} \) and \( E_{\bullet 2} \). Therefore, every column of \( A \) depends linearly on \( A_{\bullet 1} \) and \( A_{\bullet 2} \). For example, \( E_{\bullet 1} = E_{\bullet 1}(2) + E_{\bullet 2}(-1) \), and so \( A_{\bullet 1} = A_{\bullet 1}(2) + A_{\bullet 2}(-1) \), much as in the previous example.

3.4.10  Uniqueness of the solution to \( Ax = b \). One-to-one linear transformations

If the columns of the \( m \times n \) matrix \( A \) are linearly independent then for any \( m \times 1 \) column vector \( b \) the system of equations \( Ax = b \) has at most one solution.

To see this, consider the following:

Suppose that \( u \) and \( v \) are two solutions, i.e. \( Au = b \) and \( Av = b \). Then \( Au = Av \) and so \( Au - Av = A(u - v) = 0 \). As the columns of \( A \) are linearly independent, \( u - v = 0 \) and \( u = v \).
Consider the matrix mapping \( x \rightarrow Ax \) of (3.32). To say that this mapping is one-to-one means that for any two column vectors \( u \) and \( v \) with \( n \) entries, we can have \( Au = Av \) only if \( u = v \). In other words, two linear combinations of the columns of \( A \) are equal only if they have identical coefficients. The mapping \( x \rightarrow Ax \) is one-to-one if, and only if, the column-rank of \( A \) is \( n \).

Example 3.4.4 A row echelon form (2.2) of the coefficient matrix \( A \) of Example 2.1.2 in Chapter 2 is

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

The column-rank of \( A \) is 3 and the mapping

\[ x \rightarrow \begin{bmatrix}
1 & 1 & 3 \\
1 & 1 & 4 \\
-1 & 1 & -2 \\
\end{bmatrix} x \]

from \( \mathbb{R}^3 \) to \( \mathbb{R}^3 \) is one-to-one.

All this becomes even clearer if we look at the row-reduced echelon form in Example 2.2.7 of Chapter 2. This is the \( 3 \times 3 \) identity matrix and so \( Ax = b \) has a solution for any \( b \in \mathbb{R}^3 \): The mapping \( x \rightarrow Ax \) has range the whole of \( \mathbb{R}^3 \).

The example illustrates the following remark.

3.4.11 Geometric interpretation of column-rank

A proper definition of ‘dimension’ is given in section 3.6), but we can accept that a line in ordinary space through \( O \) has dimension 1, that a plane through \( O \) has dimension 2 and that the dimension of the whole of ordinary space is 3. The column-rank of a matrix \( A \) can be thought of as the dimension of its range when \( A \) is considered as a linear mapping. To better appreciate this, see Exercise 78, No.19 as well as No.8 in Exercise 93.

Consider the matrix of Example 3.3.1 that projects on a line. Each column is a multiple of any other and the column-rank is 1, as expected.

Again, consider the coefficient matrix \( A \) of equation (3.40). Because \( A_{11} + A_{21}(2) + A_{31} = 0 \), the column-rank of \( A \) is not 3. It is easy to see that any two columns are linearly independent, so \( r(A) = 2 \), in agreement with our comment.

The range of a reflection matrix is the whole of space, which has dimension 3. So we expect its rank to be 3, which is true.

3.4.12 Row Independence and row-rank

Everything that can be said about column independence has its counterpart for rows. Hence, as indicated earlier, the rows of the \( m \times n \) matrix \( A \) are linearly independent if, and only if for all row vectors \( y \) with \( m \) entries,

\[ yA = 0 \quad \text{implies} \quad y = 0 \]

In other words, if we can find a row vector \( y \neq 0 \) such that the linear combination

\[ y_1A_{1*} + y_2A_{2*} + \cdots + y_mA_{m*} = 0 \]

then the rows of \( A \) are linearly dependent, otherwise they are linearly independent.

The row-rank of \( A \) is the maximum number of linearly independent rows of \( A \).
3.4.13 The non-zero rows of a matrix in row echelon form are linearly independent.

To see this, let $C$ be an $m \times n$ matrix in row echelon form, and suppose that $C_1, C_2, \ldots, C_k$, are its non-zero rows. Let

$$y_1C_1 + y_2C_2 + \cdots + y_kC_k = \mathbf{0}$$

where $\mathbf{0}$ is the zero row with $n$ entries. The first non-zero entry in $C_1$ has only zeros below it. Therefore, $y_1 = 0$ and $y_2C_2 + \cdots + y_kC_k = \mathbf{0}$. As the first non-zero entry in $C_2$ has only zeros below it, $y_2 = 0$. Carrying on the argument, $y_1 = y_2 = \cdots = y_k = 0$. The claim is proved.

For example, the first two rows of

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

For more examples see Chapter 2, including Exercises 35 and 39.

3.4.14 Elementary column operations for equations of type $yA = \xi$

Let $A$ be an $m \times n$ matrix. In order to solve the matrix equation

$$yA = \xi$$

we can mimic the procedure followed to solve a system of the type $Ax = b$. To do this we must perform elementary column operations on the augmented array

$$\begin{bmatrix}
A \\
\xi
\end{bmatrix}$$

Such operations do not alter the solution set to equation 3.59 and are:

(i) interchange two columns of the array, (ii) multiply a column of the array by a non-zero constant and (iii) add a multiple of some column to another.

In remark 49 we mentioned that we are in the habit of rather doing elementary row operations on the system

$$A^T \begin{bmatrix} y \\ \xi \end{bmatrix} = \begin{bmatrix} b \\ c \end{bmatrix}$$

Every elementary row operation on this system has its exact counterpart as an elementary column operation on the system $yA = \xi$, and this is true in particular for $\xi = \mathbf{0}$.

Nevertheless, column operations on an array are important in their own right.

Example 3.4.5 Let

$$B = \begin{bmatrix}
-1 & 3 & 2 & -5 \\
-4 & 4 & 9 & -1 \\
-3 & 1 & 7 & 4 \\
2 & 2 & -5 & -9
\end{bmatrix}$$

Find the row-rank of $B$.

Solution:

$B$ is the transpose of the matrix $A$ in example 3.4.3. Thus its row-rank is the column-rank of $A$, which is 2.

Note 72 For further elementary properties of linear dependence/independence, see No.1 in Exercise 78 below.
3.4.15 A Column Lemma

Let $A$ and $B$ have the same number $n$ of columns. If for all column vectors $\mathbf{x}$ with $n$ entries,

$$Ax = 0 \implies Bx = 0$$

then $(\text{column-rank } A) \geq (\text{column-rank } B)$.

**Proof**: We may suppose that the first $k$ columns of $B$ are independent, where $k$ is the column-rank of $B$. Let $\mathbf{x}$ be any vector such that $x_{k+1} = \cdots = x_n = 0$ and $Ax = 0$. By assumption, $Bx = 0$. As the first $k$ columns of $B$ are independent, $\mathbf{x} = 0$. Hence the first $k$ columns of $A$ are independent.

**Remark 73** We have actually shown a bit more: if the condition (3.60) holds, and certain columns of $B$ are linearly independent, then the corresponding columns of $A$ are also linearly independent.

3.4.16 Corollary: column-rank of a matrix product

Let the matrix product $CA = B$ be defined. Then $(\text{column-rank } A) \geq (\text{column-rank } B)$ and if $C$ has independent columns, the two ranks are equal.

For, if $Ax = 0$, then certainly $Bx = 0$ and so $(\text{column-rank } A) \geq (\text{column-rank } B)$ by the Lemma 3.4.15.

Now let $C$ have independent columns and suppose $Bx = 0$. Then $C(Ax) = 0$ and therefore $Ax = 0$ as the columns of $C$ are independent. It follows, again by the Lemma, that $(\text{column-rank } B) \geq (\text{column-rank } A)$ and equality holds.

**Example 3.4.6** Let $C$ be the $3 \times 3$ matrix of Example 3.4.4 of rank 3 and let $A$ be the $3 \times 3$ matrix of example 3.4.2 with column-rank 2. From the corollary we can conclude that the column-rank of the product

$$CA = \begin{bmatrix} 1 & 1 & 3 \\ -1 & 1 & -2 \\ 1 & 1 & 4 \end{bmatrix} \begin{bmatrix} 2 & 5 & -11 \\ -3 & -6 & 12 \\ 4 & 7 & -13 \end{bmatrix} = \begin{bmatrix} 11 & 20 & -38 \\ -13 & -25 & 49 \\ 15 & 27 & -51 \end{bmatrix}$$

is also 2.

The counterpart for rows of the Column Lemma is

3.4.17 A Row Lemma

Let $A$ and $B$ have the same number $m$ of rows. If for all row vectors $\mathbf{y}$ with $m$ entries,

$$yA = 0 \implies yB = 0$$

then row-rank $A \geq$ row-rank $B$.

**Remark 74** As with the Column Lemma, we have shown that if the condition (3.4.17) holds and certain rows of $B$ are linearly independent, the corresponding rows of $A$ are also linearly independent.

3.4.18 Corollary: row-rank of a matrix product

Let $AD = B$ be a matrix product. Then row-rank $A \geq$ row-rank $B$ and equality holds if $D$ has independent rows.

The proof of this Lemma and its corollary are left as exercises. See No.13 in Exercise 78.
3.4. THE COLUMN-RANK AND ROW-RANK OF A MATRIX A ARE EQUAL

**Proof:**

Let the column-rank of the $m \times n$ matrix $A$ be $k$. By 3.4.8 the $k$ pivot columns of $A$ are linearly independent, forming an $m \times k$ matrix $B$. By 3.4.9 all columns of $A$ are linear combinations of the columns of $B$. Thus, for some some $k \times n$ matrix $D$,

$$A = BD$$

By Corollary 3.4.18, row-rank $B \geq$ row-rank $A$. But $B$ has $k$ columns, so its row-rank cannot exceed $k$:

$$\text{column-rank } A = k \geq \text{row-rank } B \geq \text{row-rank } A$$

Hence the column-rank of $A$ is at least the row-rank of $A$.

**By symmetry** (consider the transpose $A^T$), the row-rank of $A$ is at least the column-rank of $A$ and so the two are equal.

3.4.20 **Definition of rank** $r(A)$

The common row- and column-rank of $A$ is called the **rank** of $A$ and is denoted by $r(A)$.

By 3.4.8, the rank of $A$ is the number of non-zero rows in a row echelon form of $A$.

**Example 3.4.7** Find the rank of the matrix

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

**Solution:**

This is the matrix of Example 2.2.4 in Chapter 2. We found its row-echelon form in equation 2.26:

$$\begin{bmatrix} 2 & -3 & -4 & 3 & -3 \\ 0 & 1 & 2 & -6 & 1 \\ 0 & 0 & 3 & -2 & -2 \\ 0 & 0 & 0 & -5 & -3 \end{bmatrix}$$

Therefore, $r(A) = 4$

3.4.21 **Lemma: A basic property of linear independence**

A basic property concerning linear independence/dependence is the following.

Let $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_n$ be a list of $n$ linearly independent column vectors, each with $m$ entries and $\mathbf{b}$ a column vector with $m$ entries. Then the rank of $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_n, \mathbf{b}$ equals the rank of $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_n$ if, and only if, $\mathbf{b}$ depends linearly on $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_n$.

**Proof:**

The rank of $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_n, \mathbf{b}$ is either $n$ or $n + 1$. If it is $n$, then these vectors are linearly dependent, and there exist constants $x_1, \ldots, x_n, \beta$ not all zero such that

$$\mathbf{a}_1 x_1 + \mathbf{a}_2 x_2 + \cdots + \mathbf{a}_n x_n + b\beta = \mathbf{0}$$
We must have $\beta \neq 0$, otherwise one or more of the scalars $x_j$ would have to be non-zero, contradicting the linear independence of the vectors $a_1, a_2, \ldots, a_n$. Therefore,

$$b = a_1 (-x_1 \beta^{-1}) + a_2 (-x_2 \beta^{-1}) + \cdots + a_n (-x_n \beta^{-1})$$

and $b$ is a linear combination of $a_1, a_2, \ldots, a_n$.

Conversely, if $b$ is a linear combination of $a_1, a_2, \ldots, a_n$, then $a_1, a_2, \ldots, a_n, b$ cannot be linearly independent, and Theorem A has been proved.

The Lemma rephrased: Given linearly independent vectors $a_1, a_2, \ldots, a_n$, the vectors $a_1, a_2, \ldots, a_n, b$ are linearly independent if, and only if, $b$ does not depend linearly on $a_1, a_2, \ldots, a_n$.

3.4.22 Theorem: The fundamental property of linear independence

**Note 75** This theorem is a generalization of the Lemma. To understand what follows you should have a thorough grasp of the content of the “linear combinations of linear combinations theorem” contained in No.8c of Exercise 47 and No.13 of Exercise 63.

Let $a_1, a_2, \ldots, a_n, b$ be any list of $n$ column vectors, each with $m$ entries. Then the rank of $a_1, a_2, \ldots, a_n, b$ is not more than the rank of $a_1, a_2, \ldots, a_n$ if, and only if, $b$ depends linearly on $a_1, a_2, \ldots, a_n$.

Put slightly differently, the theorem says that

$$\text{column-rank } [A, b] = \text{column-rank } A$$

if, and only if, $Ax = b$ has a solution. Here, $A$ is the $m \times n$ matrix having the vectors $a_j$ as columns.

**Proof:**

Let the maximum number of linearly independent vectors among $a_1, a_2, \ldots, a_n$ be $s$.

1. Suppose that the rank of $a_1, a_2, \ldots, a_n, b$ is $s$. Then, by Theorem A, $b$ must be a linear combination of any $s$ linearly independent vectors among $a_1, a_2, \ldots, a_n$.

2. Conversely (and this is the only non-trivial case), suppose that the rank of $a_1, a_2, \ldots, a_n$, $b$ is $s + 1$.

Then we may assume (without loss in generality) that $a_1, \ldots, a_s, b$ are linearly independent. By Theorem A, every vector $a_j$ is a linear combination of $a_1, \ldots, a_s$. Then if $b$ were a linear combination of $a_1, a_2, \ldots, a_n$, it would also be a linear combination of $a_1, \ldots, a_s$, which is a contradiction, because $a_1, \ldots, a_s, b$ are assumed to be linearly independent.

**Corollary 76** Let the matrix $B$ consist of $k$ linearly independent columns of a matrix $A$ which cannot be enlarged by columns of $A$ without destroying independence. Then

1. Every column of $A$ is a linear combination of columns of $B$ (this follows from the Lemma)

2. $k$ is the column-rank of $A$.

**Remark 77** We emphasize that the above theorem 3.4.22 about column vectors is equally valid for row vectors. The statement 3.4.22 goes through word-for-word.

The full use of the theorem will be seen in section 3.6 when we study vector spaces.
3.4. LINEAR INDEPENDENCE

3.4.23 An Application of the Fundamental Theorem

We will use 3.4.22 to show directly that elementary row operations on an $m \times n$ matrix $A$ do not change its row-rank.

It is clear that interchanging two rows cannot change the row-rank. It is almost as obvious that multiplication of a row by a non-zero constant does not alter the row-rank (see No.1(c)iv in Exercise 78). It remains to show that replacing (say) $A_1$, by $\beta A_2 + A_1$, to form $A'$ does not change the row-rank.

Let the last $m-1$ rows $A_2, \ldots, A_m$ of $A$ have rank $k$. Since $\beta A_2 + A_1$ is a linear combination of these $m-1$ rows if, and only if, $A_1$ is (see No.2 in Exercise 52), it follows that the row-rank of $A'$ is $k+1$ if, and only if, the row-rank of $A$ is $k+1$.

For an alternative way to view this see Exercise 93 No. 9.

Exercise 78 1. In most of the elementary questions that follow, you may assume the given vectors are either rows or columns, each with the same number of entries. Prove the given statements.

(a) A single vector $\mathbf{a}$ is linearly independent if, and only if, $\mathbf{a} \neq \mathbf{0}$.

(b) Show directly that two vectors $\mathbf{a}$, $\mathbf{b}$ are linearly dependent if, and only if, one is a multiple of the other. Compare the idea of parallel vectors in subsection 1.1.10, Chapter 1.

(c) Let $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_n$ be a given list of vectors. Show

i. If some $\mathbf{a}_i = \mathbf{0}$, then $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_n$ are linearly dependent.

ii. If $n \geq 2$ and some subset, say $\mathbf{a}_2, \ldots, \mathbf{a}_m$, are linearly dependent, then are $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_n$.

iii. If $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_n$ are linearly independent, what can be said of any subset of $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_n$?

iv. The rank of $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_n$ is unaltered if one of the vectors $\mathbf{a}_j$ is multiplied by a non-zero scalar $k$.

v. Let $n \geq 2$ and suppose that $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_n$ are non-zero vectors. Prove that the vectors are linearly dependent if, and only if, one of the vectors $\mathbf{a}_j$ (for $j \geq 2$) depends linearly on $\mathbf{a}_1, \ldots, \mathbf{a}_{j-1}$. This statement is known as the principle of linear independence.

Hint 79 Suppose the vectors $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_n$ are linearly dependent. Since $\mathbf{a}_1 \neq \mathbf{0}$, there must be a first index $j \geq 2$ such that the vectors $\mathbf{a}_1, \ldots, \mathbf{a}_j$ are linearly dependent.

2. How does Lemma 3.4.21 show why 3.4.9 follows directly from 3.4.8?

3. Let $\mathbf{a}$, $\mathbf{b}$, $\mathbf{c}$ be the position vectors of the points $A$, $B$ and $C$ respectively. Show that the three vectors are linearly independent if, and only if, $A$, $B$, $C$ and $O$ do not lie on a plane. (This was mentioned in the preamble 3.4).

4. Let $A = [C, D]$ where $C$ and $D$ are matrices with the same number of rows. Show that if the rows of $C$ (or of $D$) are independent, so are those of $A$. Conclude that in any case $r(A) \geq r(C)$ and $r(A) \geq r(D)$.

5. For any matrix product $AB$, say why $r(AB) \leq r(A)$ and $r(AB) \leq r(B)$.

6. Let $C$ be $m \times n$. If $A$ is $m \times m$ of rank $m$, then $r(AC) = r(C)$. Let $B$ be $n \times n$ of rank $n$, then $r(CB) = r(C)$. Consequently, $r(ACB) = r(C)$. Prove these statements.
7. Without multiplying out, find the rank of

\[
\begin{bmatrix}
2 & 5 & -11 \\
-3 & -6 & 12 \\
4 & 7 & -13
\end{bmatrix}
\begin{bmatrix}
1 & 1 & 3 \\
-1 & 1 & -2 \\
1 & 1 & 4
\end{bmatrix}
= 
\begin{bmatrix}
-14 & -4 & -48 \\
15 & 3 & 51 \\
-16 & -2 & -54
\end{bmatrix}
\]

**Hint 80** Consider Example 3.4.6.

8. The following questions refer you to the matrices in Exercise 35 of Chapter 2. To answer them you should consult the row reduced echelon forms (row REF) found in Exercise 39. Occasionally you may need to find the new row reduced echelon forms.

(a) Consider the matrix \( A \) of No.1a. Show that the first two columns of \( A \) form a maximum independent set of columns of \( A \) and that every column of \( A \) is a linear combination of \( A_1 \) and \( A_2 \). Can you find three linearly independent columns of \( A \)? Find the rank \( r(A) \).

Show that the first two rows of \( A \) form a maximum number of independent rows of \( A \).

**Hint 81** Either show directly that \( A_1 \) and \( A_2 \) are linearly independent or use the matrix \( B \) of No.1b in Exercise 35.

(b) Let \( C \) be the row REF of matrix \( A \) of No.2 of Exercise 35.

Show that \( A_5 \) depends linearly on \( A_1, A_2, A_3 \) and that every maximum set of linearly independent columns must include \( A_4 \). Find a maximum set of linearly independent columns of \( A \) and the rank \( r(A) \).

Which of the following sets are linearly independent?

i. \( A_2, A_3, A_4, A_5 \)

ii. \( A_1, A_3, A_4, A_5 \)

iii. \( A_1, A_2, A_4, A_5 \)

Find a maximum number of linearly independent solutions to the homogeneous equation.

(c) Consider the coefficient matrix \( A \) in No.3 of Exercise 35.

i. Does any row vector \( \mathbf{c} = \begin{bmatrix} c_1 & c_2 & c_3 & c_4 \end{bmatrix} \) depend linearly on the rows of \( A \)?

**Solution:**

\( A \) is \( 4 \times 4 \) and the rows of \( A \) together with \( \mathbf{c} \) can have rank at most four. Since the rows of \( A \) are linearly independent (\( r(A) = 4 \) from the row REF), the answer is ‘yes’ by the row version of Lemma 3.4.21.

ii. Are the columns of the coefficient matrix \( A \) linearly independent?

iii. Does \( A_2 = \mathbf{b} \) have a unique solution for every column vector \( \mathbf{b} \) with 4 entries?

iv. Let \( C \) be the augmented matrix of No.3a in Exercise 35. Does every column vector \( \mathbf{b} \) with 4 entries depend linearly on \( C_1, C_2, C_3, C_5 \)?

9. Prove that in 3-space mutually orthogonal vectors are linearly independent. Hence complete a rigorous proof of No.15 of Exercise 2 in Chapter 1.

10. * Give rigorous proofs of the statements about planes in subsection 1.3.6.
11. Let $I$ be the $m \times m$ identity matrix. The matrix $J$ resulting from $I$ after performing an elementary row operation on $I$ is called an $m \times m$ **elementary matrix**. Hence $J$ is the result of doing one of the following: (i) multiplying row $r$ of $I$ by $\beta \neq 0$, (ii) interchanging rows $r$ and $s$ of $I$, and (iii) changing row $I_r$ to $\beta I_r + I_s$ if $s \neq r$.

(a) Show that an elementary matrix $J$ has rank $m$.

(b) Show furthermore that if $A$ is an $m \times n$ matrix, then $JA$ is the result of doing the corresponding elementary row operation on $A$.

**Hint 82** Use $[JA]_{i\bullet} = J_{i\bullet}A$. So, for example, let $J$ be the result of a type (iii) operation. Then for $i \neq s$, $[JA]_{i\bullet} = A_{i\bullet}$, while if $i = s$,

$$[JA]_{i\bullet} = J_{i\bullet}A = (\beta I_{i\bullet} + I_s)A = (\beta I_{i\bullet})A + I_sA = \beta A_{i\bullet} + A_{s\bullet}$$

(c) What are the corresponding results for elementary column operations?

(d) Give a description of "column echelon form" and "column reduced echelon form" of a matrix $A$.

12. Show that $A$, $A^T$, $A^T A$ and $A A^T$ all have the same rank.

**Hint 83** First of all, recall that if $a$ is a column vector and $a^T a = 0$, then $a = 0$ (See Exercise 47 No.11).

Now suppose $A^T A\underline{x} = 0$. Then $\underline{x}^T A^T A\underline{x} = (A\underline{x})^T A\underline{x} = 0$.

13. Prove the Lemma 3.4.17 and its corollary 3.4.18.

**Hint 84** One way: Use $A^T$ and $B^T$ in place of $A$ and $B$ in the previous Lemma 3.4.15. For a more pleasant direct proof, we may suppose that the first $k = r(B)$ rows of $B$ are independent. Let $y$ be any row vector such that $y_{k+1} = \cdots = y_m = 0$ and $yA = 0$. Now proceed as in the Column Lemma 3.4.15, except use rows in place of columns.

**Solution:**

By assumption, $yB = 0$. As the first $k = r(B)$ rows of $B$ are independent, $y = 0$. Hence the first $k$ rows of $A$ are independent.

14. Complete the following alternative proof of the corollary to the Column Lemma (Corollary 3.4.16) which says $r(BC) \leq r(C)$. If $k$ columns (say $[BC]_{i1}, \cdots, [BC]_{ik}$) of $BC$ are linearly independent then so are $C_{i1}, \cdots, C_{ik}$. Do the same for the row version.

15. In the following assume the relevant matrix products are defined. Prove the statements.

(a) If the columns of $A$ are independent and $AB = AC$, then $B = C$

(b) If the rows of $A$ are independent and $BA = CA$ then $B = C$.

16. Let $A$ be $n \times n$ and $A \neq I$, $A \neq O$ and $A^2 = I$, where $I = I_n$. Find all integers $k \geq 1$ such that $A^k = I$.

17. * Using (among other things) the Column Lemma 3.4.15, show that if the condition (3.60) holds then $B = CA$ for some $C$. (This is essentially the converse of the corollary 3.4.16). State and prove the corresponding result for the Row Lemma 3.4.17.

**Hint 85** Let $\underline{d} = \underline{d} = B_{i\bullet}$ be a row of $B$ and let $A'$ be the augmented matrix formed by adding row $\underline{d}$ to $A$. Then $A'\underline{x} = 0$ if, and only if, $A\underline{x} = 0$. Hence $A'$ and $A$ have the same column ranks and so the same row ranks. By the row version of the basic theorem 3.4.22 on linear independence, $\underline{d}$ is a linear combination of the rows of $A$. 
State and prove a similar result corresponding to the Row Lemma 3.4.17.

18. Find the ranks of the following matrices and prove your results correct.

(a) $A$ is the projection matrix on the plane $2x + z = 0$ of Exercise 56, No.1.

(b) $A$ is the reflection matrix in the plane $2x + z = 0$.

19. Let $A$ be a $3 \times 3$ matrix. In each of the following cases $A$ is considered as a linear transformation. Describe the rank of $A$ with a brief geometrical proof. Illustrate the comment 3.4.11.

(a) $A$ is a projection matrix onto a line that passes through the origin.

(b) $A$ is a projection matrix onto a plane that passes through the origin.

(c) $A$ is a matrix that reflects in a plane or line that passes through the origin. This again illustrates the remarks in 3.4.11.

20. * Use the results of Exercise 63, No.17 to give analytic proofs for No.19 above.

21. If $R_\theta$ is the $2 \times 2$ rotation matrix of equation (3.53), prove algebraically that $r(R_\theta) = 2$.

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

(a) What is the row REF $A^\prime\prime$ of $A$ if $r(A) = n$? What is $A^\prime\prime$ if $r(A) = m = n$?

(b) Let $A\mathbf{z} = b$ be a system of $m$ equations for $n$ unknowns. If the rows of $A$ are independent then the system always has a solution. Find a quick proof using the Fundamental Theorem 3.4.22.

(c) Let $A\mathbf{y} = c$ be a system of $n$ equations for $m$ unknowns $y_1, \ldots, y_m$. If the columns of $A$ are independent then the system always has a solution.

(d) What can be said of the solutions in 22b and 22c if $m = n$?

3.5 Square matrices

In this section we consider square matrices $A$ of size $n \times n$. Unless otherwise stated, $I$ stands for the identity matrix $I_n$. Observe that $r(I) = n$.

3.5.1 Invertible matrices

3.5.2 Definition of invertibility

An $n \times n$ matrix $A$ is said to be invertible (or non-singular or to have an inverse) if there is a matrix $B$ such that

$$AB = BA = I$$

(3.61)

It is clear that such a matrix $B$ must also be $n \times n$. From the definition it is also clear that if $B$ is an inverse of $A$ then $A$ is an inverse of $B$.

From Exercise 78 No.5 we see that a necessary condition for $A$ to have an inverse is that $r(A) = n$. The converse will be shown in 3.5.7.
3.5.3 Uniqueness of the inverse, definition of $A^{-1}$

Let $X$ and $Y$ be $n \times n$ matrices such that $AX = I$ and $YA = I$. Then $X = Y$ and this matrix is the unique inverse of $A$.

**Proof:** We have

$$Y = YI = Y(AX) = (YA)X = IX = X$$

(3.62)

Hence $X = Y$ is an inverse of $A$. But two inverses $X$ and $Y$ of $A$ must satisfy (3.62), and so $A$ can have only one inverse. It is denoted by $A^{-1}$.

**Example 3.5.1** Show directly that

$$\begin{bmatrix} 3 & 7 \\ 2 & 5 \end{bmatrix}^{-1} = \begin{bmatrix} 5 & -7 \\ -2 & 3 \end{bmatrix}$$

**Solution:**

By direct multiplication,

$$\begin{bmatrix} 3 & 7 \\ 2 & 5 \end{bmatrix} \begin{bmatrix} 5 & -7 \\ -2 & 3 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 5 & -7 \\ -2 & 3 \end{bmatrix} \begin{bmatrix} 3 & 7 \\ 2 & 5 \end{bmatrix}$$

3.5.4 Use of the inverse

3.5.5 Solving $Ax = b$

Consider a system

$$Ax = b$$

(3.63)

of $n$ equations for $n$ unknowns. If $A$ has an inverse, then the unique solution is

$$x = A^{-1}b$$

In the first place, we show that $x = A^{-1}b$ is a solution to (3.63):

$$A(A^{-1}b) = (AA^{-1})b = I_b = b$$

Conversely, let $x = u$ satisfy the equation (3.63), i.e. suppose $Au = b$. Multiply both sides of this equation on the left by $A^{-1}$:

$$A^{-1}(Au) = (A^{-1}A)u = I_u = u = A^{-1}b$$

Thus $x = A^{-1}b$ is the unique solution to (3.63).

3.5.6 Solving $yA = c$

If $A$ has an inverse, then for any row vector $c$ with $n$ entries, the unique solution to

$$yA = c$$

(3.64)

is

$$y = cA^{-1}$$

The proof can be derived from 3.5.5 using transposes (see equation 3.9). Alternatively, the proof parallels that of the previous section. The unique solution is found by multiplying both sides of equation (3.64) on the right by $A^{-1}$. 

3.5. SQUARE MATRICES
Example 3.5.2  Solve
\[
\begin{bmatrix}
3 & 7 \\
2 & 5
\end{bmatrix}
\begin{bmatrix}
x \\
y
\end{bmatrix}
= 
\begin{bmatrix}
9 \\
-11
\end{bmatrix}
\]

Solution:
Use the result of Example 3.5.1:
\[
\begin{bmatrix}
x \\
y
\end{bmatrix}
= 
\begin{bmatrix}
3 & 7 \\
2 & 5
\end{bmatrix}^{-1}
\begin{bmatrix}
9 \\
-11
\end{bmatrix}
= 
\begin{bmatrix}
5 & -7 \\
-2 & 3
\end{bmatrix}
\begin{bmatrix}
9 \\
-11
\end{bmatrix}
= 
\begin{bmatrix}
122 \\
-51
\end{bmatrix}
\]

3.5.7  Necessary and sufficient conditions for \( A \) to be invertible
The \( n \times n \) matrix \( A \) is invertible if, and only if, any of the following equivalent conditions hold.

If (14) holds, necessarily \( X = A^{-1} \). If (15) holds, necessarily \( Y = A^{-1} \).

1. \( r(A) = n \).
2. The columns of \( A \) are independent.
3. The rows of \( A \) are independent.
4. The homogeneous equation \( Ax = 0 \) has only the trivial solution.
5. The homogeneous equation \( yA = 0 \) has only the trivial solution.
6. The row REF of \( A \) is \( I \).
7. The column REF of \( A \) is \( I \) (the row REF of \( A^T \) is \( I \)).
8. For every column vector \( b \) with \( n \) entries \( Ax = b \) has a solution.
9. For every column vector \( b \) with \( n \) entries \( Ax = b \) has a unique solution.
10. There is a column vector \( b \) with \( n \) entries such that \( Ax = b \) has a unique solution.
11. For every row vector \( c \) with \( n \) entries \( yA = c \) has a solution.
12. For every row vector \( c \) with \( n \) entries \( yA = c \) has a unique solution.
13. There is a row vector \( c \) with \( n \) entries such that \( yA = c \) has a unique solution.
14. There is a matrix \( X \) such that \( AX = I \).
15. There is a matrix \( Y \) such that \( YA = I \).

Proof
The first seven properties are equivalent from the definition of linear independence, the fact that row-rank is equal to column-rank, and because the row (column) REF of \( A \) is \( I \) if, and only if, the columns (rows) of \( A \) are independent.

Now assume \( r(A) = n \). Since the addition of column \( b \) (row \( c \)) to \( A \) does not change the column (row) rank, \( Ax = b \) has a solution by Lemma 3.4.21, and similarly so does \( yA = c \). These solutions are necessarily unique on account of the independence of the columns (rows) of \( A \).

Assume next that (8) holds. Then we can solve the \( n \) matrix equations \( AX_{\bullet} = I_\bullet \) for the columns \( X_{\bullet} \), and obtain an \( n \times n \) matrix \( X \) such that \( AX = I \). Similarly, if (11) holds we can find an \( n \times n \) matrix \( Y \) such that \( YA = I \).

By No.5 of Exercise 78, if either \( AX = I \) for some \( X \) or \( YA = I \) for some \( Y \), necessarily
3.5. SQUARE MATRICES

Thus all the above properties (1) - (13) are equivalent.

Finally, suppose (14) holds. Then, as the properties are equivalent, \( YA = I \) for some \( Y \). By 3.5.3, \( X = Y \) and \( X \) is the inverse of \( A \).

A similar argument shows that if (15) holds, necessarily \( Y = A^{-1} \).

**Example 3.5.3** Show that none of the matrices

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

can have inverses.

**Solution:**

None of these has rank 2.

3.5.8 Reduction method finds \( A^{-1} \), or decides it does not exist

Reduce the augmented array \([A, I]\) to row REF \([A'', I'']\). Then \( A \) has an inverse if, and only if, \( A'' = I'' \), in which case \( I'' = A^{-1} \). We are using property (14) of 3.5.7.

Note that if at any stage say, \([A', I']\) of the reduction, \( A' \) has a zero row, then \( A^{-1} \) cannot exist.

**Example 3.5.4** Consider the matrix of example 3.1.7:

\[
A = \begin{bmatrix} 2 & 5 & 1 \\ -3 & -6 & 0 \\ 4 & 7 & -2 \end{bmatrix}
\]

We found there that an echelon form has three non-zero rows. Therefore, \( r(A) = 3 \) and \( A^{-1} \) exists. Find \( A^{-1} \) by row reduction.

**Solution:**

Start with

\[
\begin{array}{ccccccc}
2 & 5 & 1 & 1 & 0 & 0 \\
-3 & -6 & 0 & 0 & 1 & 0 \\
4 & 7 & -2 & 0 & 0 & 1 \\
\end{array}
\]

\[
\begin{array}{ccccccc}
2 & 5 & 1 & 1 & 0 & 0 \\
0 & \frac{3}{2} & \frac{3}{2} & \frac{3}{2} & 1 & 0 \\
0 & -3 & -4 & -2 & 0 & 1 \\
\end{array}
\]

\( \frac{3}{2} R_1 + R_2 \)

\( (-2) R_1 + R_3 \)

\[
\begin{array}{ccccccc}
2 & 5 & 1 & 1 & 0 & 0 \\
0 & \frac{3}{2} & \frac{3}{2} & \frac{3}{2} & 1 & 0 \\
0 & 0 & -1 & 1 & 2 & 1 \\
\end{array}
\]

\( 2R_2 + R_3 \)

\[
\begin{array}{ccccccc}
2 & 5 & 1 & 1 & 0 & 0 \\
0 & 1 & 1 & 2 & 0 & 0 \\
0 & 0 & 1 & -1 & -2 & -1 \\
\end{array}
\]

\( \frac{2}{3} R_2 \)

\( (-1) R_3 \)

\[
\begin{array}{ccccccc}
2 & 5 & 0 & 2 & 2 & 1 \\
0 & 1 & 0 & 2 & \frac{2}{3} & 1 \\
0 & 0 & 1 & -1 & -2 & -1 \\
\end{array}
\]

\( -R_3 + R_1 \)

\( -R_3 + R_2 \)

\( 0 \)

\[
\begin{bmatrix}
2 & 0 & 0 & -8 & -34 \\
0 & 1 & 0 & 2 & -1 \\
0 & 0 & 1 & -1 & -2 \\
1 & 0 & 0 & -4 & -17 \\
0 & 1 & 0 & 2 & -1 \\
0 & 0 & 1 & -1 & -2
\end{bmatrix} + \begin{bmatrix}
-5R_2 + R_1 \\
\frac{3}{7} \\
-2 \\
\frac{1}{7}R_1 \\
\frac{3}{7} \\
-1
\end{bmatrix}
\]

Hence
\[
A^{-1} = \begin{bmatrix}
-4 & -\frac{17}{3} & -2 \\
2 & \frac{3}{7} & 1 \\
-1 & -2 & -1
\end{bmatrix}
\]

**Exercise 86**

1. Find the inverses of the following matrices (where possible), by reduction to \( I \), and solve the related problems.

   (a) \[
   A = \begin{bmatrix}
   2 & 1 \\
   3 & 2
   \end{bmatrix}
   \]

   (b) i. \[
   B = \begin{bmatrix}
   3 & 2 & 1 \\
   0 & 2 & 2 \\
   0 & 0 & -1
   \end{bmatrix}
   \]

   ii. Find the linear transformation \( T \) that maps column \( B_{\bullet j} \) of \( B \) to the corresponding column \( Y_{\bullet j} \) of \( Y \) for \( j = 1, 2, 3 \), where

   \[
   Y = \begin{bmatrix}
   1 & 0 & 4 \\
   -1 & 1 & -2 \\
   2 & 3 & 1
   \end{bmatrix}
   \]

   **Hint 87** You want a \( 3 \times 3 \) matrix \( T \) such that \( Y_{\bullet j} = TB_{\bullet j} \). In fact, \( T = YB^{-1} \)

   (c) \[
   C = \begin{bmatrix}
   1 & 6 & 2 \\
   -2 & 3 & 5 \\
   7 & 12 & -4
   \end{bmatrix}
   \]

   (d) i. \[
   D = \begin{bmatrix}
   1 & 3 & -1 \\
   2 & 0 & -3 \\
   -1 & 4 & 2
   \end{bmatrix}
   \]

   ii. Use your inverse to solve

   \[
   \begin{align*}
   x_1 + 3x_2 - x_3 &= 3 \\
   2x_1 - 3x_3 &= -1 \\
   -x_1 + 4x_2 + 2x_3 &= 2
   \end{align*}
   \]

   iii. Use your inverse to solve

   \[
   \begin{align*}
   y_1 + 2y_2 - y_3 &= 4 \\
   3y_1 + 4y_3 &= 5 \\
   -y_1 - 3y_2 + 2y_3 &= -1
   \end{align*}
   \]

2. Prove that an invertible matrix maps linearly independent vectors to linearly independent vectors and state and prove the converse.
3.6. VECTOR SPACES: SUBSPACES OF $\mathbb{R}^m$

3. Let $A$ and $B$ be $n \times n$ matrices. Show that $AB$ is invertible if, and only if, both $A$ and $B$ are, in which case $(AB)^{-1} = B^{-1}A^{-1}$. Generalize this result to products of three or more matrices.

4. Let the $n \times n$ matrix $A$ be invertible. Show that $A^T$ is also invertible and that $(A^T)^{-1} = (A^{-1})^T$.

5. Let $A$ be $n \times n$. As a linear transformation, the mapping $A : \mathbb{R}^n \to \mathbb{R}^n$ is one-to-one if, and only if, $A$ maps $\mathbb{R}^n$ onto the whole of $\mathbb{R}^n$. Prove the statement. (Compare Example 3.4.4).

6. If $M$ is a matrix that reflects in a plane or line passing through the origin, show algebraically that $M^{-1} = M$.

**Hint 88** Use $M^2 = I$ from Exercise 63, No.17

7. * Let $A$ be an $m \times n$ matrix of rank $k$. Show that by applying elementary row operations and elementary column operations to the matrix $A$ we can find an invertible matrices $P$ and $Q$ such that $PAQ = \begin{bmatrix} I_k & O_{k \times (n-k)} \\ O_{(m-k) \times k} & O_{(m-k) \times (n-k)} \end{bmatrix}$

**Hint 89** First reduce $A$ to row REF. Now use elementary column operations and the pivot columns to reduce the non-pivot columns to zero columns. Apply suitable permutations to the columns and finally use the results of Exercise 78), No.11. (Compare No.6 of the same exercise).

8. In Chapter 2, Exercise 39), No.3 we considered the system $A\mathbf{x} = \mathbf{b}$ of two equations for two unknowns (2.32). Show directly that, provided the determinant defined by $|A| = a_{11}a_{22} - a_{21}a_{12}$ is not zero,

$$B = \frac{1}{|A|} \begin{bmatrix} a_{22} & -a_{12} \\ -a_{21} & a_{11} \end{bmatrix}$$

satisfies $AB = BA = I$, so that $B = A^{-1}$. Hence derive Cramer’s solution of the equations, as found in Chapter 2. (This theory will be further developed in Chapter 4).

9. Let $A$ be a $3 \times 3$ matrix such that $2I - 3A + 5A^2 - 7A^4 = 0$, where $0$ is the $3 \times 3$ zero matrix. Show that $A$ has an inverse and write it in terms of $A$.

For the solution see Chapter 4, subsection 4.1.1 and equation (4.1.3).

**Hint 90** $A(7A^3 - 5A + 3I) = 2I$

10. Find all $n \times n$ invertible $A$ such that $A^2 = A$.

### 3.6 Vector spaces: Subspaces of $\mathbb{R}^m$

Let $\mathbb{R}^m$ stand for $\mathbb{R}^{m \times 1}$ (the set of column vectors with $m$ entries) or $\mathbb{R}^{1 \times m}$ (the set of row vectors with $m$ entries). (The context will decide which is meant, or it may not matter, as in Theorem A below, subsection 3.6.4.)

A non-empty subset $S \subseteq \mathbb{R}^m$ is said to be a subspace of $\mathbb{R}^m$ (or a vector space in $\mathbb{R}^m$) if whenever $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_k$ are in $\mathbb{R}^m$ it follows that any linear combination of $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_k$ is also in $\mathbb{R}^m$. One says that $S$ is closed under linear combinations.

It is easy to see that $S$ is a subspace of $\mathbb{R}^m$ if $S$ is closed under linear combinations of any two vectors from $S$. (See Exercise 93, No.1.)
3.6.1 Examples

1. The set $S = \{0\}$ consisting of only the zero vector $0 \in \mathbb{R}^{m \times 1}$ is a subspace of $\mathbb{R}^{m \times 1}$. The whole space $\mathbb{R}^{m \times 1}$ is a subspace of itself. These are called trivial subspaces of $\mathbb{R}^{m \times 1}$.

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

3. The row space of $A$ is a subspace of $\mathbb{R}^{1 \times n}$. We denote this space by $\mathcal{R}(A)$. This is the span of the rows of $A$.

4. The column space of $A$ is a subspace of $\mathbb{R}^{m \times 1}$. It is denoted by $\mathcal{C}(A)$. This is the range of the mapping $x \rightarrow Ax$.

5. The set of solutions $x$ of the homogeneous equation $Ax = 0$ is a subspace of $\mathbb{R}^{n \times 1}$. It is called the null space of $A$ and is denoted by $\mathcal{N}(A)$.

These are important concepts, both for pure and applied mathematics.

3.6.2 Bases and dimension

We suppose for convenience here that subspaces are contained in $\mathbb{R}^{m} = \mathbb{R}^{m \times 1}$.

3.6.3 Definition of basis

Let $S \neq \{0\}$ be a subspace of $\mathbb{R}^{m}$. A set $a_1, a_2, \ldots, a_n$ of linearly independent vectors that span $S$ is called a basis of $S$.

The main results are the following

3.6.4 Theorem A

Let $a_1, a_2, \ldots, a_n$ be a basis for the subspace $S \neq \{0\}$ of $\mathbb{R}^{m}$. Then

1. The space $S$ contains at most $n$ linearly independent vectors.

2. If $b_1, b_2, \ldots, b_k$ are vectors in $S$ that span $S$ then $k \geq n$.

3. Any two bases of $S$ contain the same number of vectors.

Proof

1. Suppose that the vectors $b_1, b_2, \ldots, b_k$ are linearly independent. By the fundamental theorem in 3.4.22, since they are in the span of the vectors $a_1, a_2, \ldots, a_n$, adding the $b_j$ to this set cannot create more than $n$ independent vectors. Thus $k \leq n$. This proves 1.

2. A similar argument shows that if the vectors $b_1, b_2, \ldots, b_k$ in $S$ span the space $S$, then $k \geq n$, since the independent set $a_1, \ldots, a_n$ is contained in this span. This proves 2.

3. If the vectors $b_1, \ldots, b_k$ form a basis of $S$, then $k \geq n$ and $k \leq n$. Hence $k = n$. This is the statement of 3.

Corollary 91 A consequence of 1 is yet another proof that a homogeneous system $Ax = 0$ of $m$ equations in $n > m$ unknowns has a non-trivial solution. (See 2.2.15).

For, $\mathbb{R}^{m}$ has a basis consisting of the $m$ unit vectors $e_j$ and $A$ has $n > m$ columns in $\mathbb{R}^{m}$. 
3.6.5 Theorem B

A subspace $S \neq \{0\}$ of $\mathbb{R}^m$ always has a basis.

Proof:
Consider the following process. Let $a_1 \neq \{0\}$ be a member of $S$. If $a_1$ spans $S$ the process stops. Otherwise there is some $a_2 \notin \text{sp}(a_1)$. By the fundamental theorem on linear independence, 3.4.22, $a_1$ and $a_2$ are linearly independent. If $S = \text{sp}(a_1, a_2)$, stop. Otherwise, continuing in this way, the process must stop at some stage since $\mathbb{R}^m$ has at most $m$ linearly independent vectors. At which point we have linearly independent vectors $a_1, a_2, \ldots, a_n$ which span $S$. This proves the theorem.

Remark 92 Let the basis $a_1, a_2, \ldots, a_n$ for $S$ be the columns of the $m \times n$ matrix $A = [a_1, a_2, \ldots, a_n]$. What we have found is that $r(A) = n$ and the column space $C(A) = S$.

Definition
The number $k$ of vectors in any basis of $S$ is called the dimension of $S$, written $k = \text{dim}(S)$. By definition the dimension of the trivial space $\{0\}$ is 0.

Exercise 93
1. Let $S$ be a non-empty subset of $\mathbb{R}^m$. Show that it is a subspace if, and only if the following conditions hold: (i). If $a \in S$ and $\gamma$ is a scalar, then $\gamma a \in S$. (ii). If $a$ and $b$ are both in $S$ then $a + b \in S$.
2. Show that the items listed in 3.6.1 do satisfy the criteria of No.1 to be subspaces.
3. Show that the dimension of $\mathbb{R}^m$ is $m$.
4. Show that the columns of a matrix $A$ are linearly independent if, and only if, $N(A) = \{0\}$. (See 3.56.)
5. Let $A$ be an $n \times n$ matrix. Write down three equivalent necessary and sufficient on the dimensions of the row, column and null spaces of $A$ for $A$ to be invertible.
6. Let $S$ and $T$ be subspaces of $\mathbb{R}^m$. The sum $S + T$ consists of all vectors $a + b$, where $a \in S$ and $b \in T$. Show that both the intersection $S \cap T$ and sum $S + T$ are subspaces of $\mathbb{R}^m$.
7. Without any reference to elementary row operations use the fundamental theorem 3.4.22 to prove that the column - and row-ranks of an $m \times n$ matrix $A$ are equal.

Hint 94 Use results from Theorem A above and its Corollary 91 and suppose the columns of the $m \times k$ matrix $B$ form a maximum number of independent columns of $A$. Follow the proof in subsection 3.4.19. Alternatively, you can first assume the $m$ rows of $A$ are independent and show that the column-rank of $A$ is at least $m$. The general case, column-rank $A \geq$ row-rank $A$, follows from this.

8. Show that the column space $C(A)$ of an $m \times n$ matrix $A$ has a basis consisting of $r(A)$ columns of $A$. A similar result holds for the row space $R(A)$.
Conclude that the dimension of $C(A)$ and $R(A)$ are both $r(A)$.
9. Let $A$ be an $m \times n$ matrix. Suppose that $B$ is the result of performing an elementary row operation on $A$ (see subsection 2.2.6, Chapter 2). Show that $A$ and $B$ have the same row space. Conclude that if $E$ is an echelon form (or REF) of $A$, then $E$ and $A$ have the same row space.
For example, the rows of the coefficient matrix of example 2.2.5 in Chapter 2 are spanned
by the three rows $\begin{bmatrix} 1 & -1 & 1 & 2 \end{bmatrix}, \begin{bmatrix} 0 & 1 & 3 & 1 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 1 & -2 \end{bmatrix}$.

Find two rows which span the row space of the matrix $A$ in Exercise 35 of Chapter 2, No.1a.

**Hint 95** Suppose that $a_1, \ldots, a_m$ are the rows of matrix $A$ and let $y_1, \ldots, y_m$ and $k$ be scalars. Then,

$$y_1a_1 + \cdots + y_ma_m = y_1(a_1 + ka_2) + (y_2 - y_1k)a_2 + \cdots + y_ma_m$$

10. Let $b_1, b_2, \ldots, b_n$ span the space $S$ of dimension $n$. Show that the vectors $b_1, b_2, \ldots, b_n$ are necessarily independent.

11. Let independent vectors $a_1, a_2, \ldots, a_k$ be in the space $S$ of dimension $n$. Show that $S$ has a basis that includes the vectors $a_1, a_2, \ldots, a_k$. (This is implicit in the proof of the above theorem B).

12. Let $B$ be a matrix whose columns are in the column space of $A$, i.e., each $B_e \in \mathcal{C}(A)$. Then we know that $\mathcal{C}(B) \subseteq \mathcal{C}(A)$ and that $r(B) \leq r(A)$. (Why?) Show that $\mathcal{C}(B) = \mathcal{C}(A)$ if, and only if, $r(B) = r(A)$.

State and prove similar results for the row space $\mathcal{R}(A)$ of $A$.

13. * Show that if matrices $A$ and $B$ have the same size, then

$$r(A + B) \leq r[A, B] = r(A) + r(B)$$

**Hint 96** Use $\mathcal{C}(A + B) \subseteq \mathcal{C}[A, B] = \mathcal{C}(A) + \mathcal{C}(B)$

14. * Prove the following exchange theorem, first proved by the 19th Century mathematician H. Grassmann:

In $\mathbb{R}^m$ let $a_1, a_2, \ldots, a_n$ be linearly independent and suppose that each vector $a_j$ depends linearly on $b_1, b_2, \ldots, b_k$. Then

(i) $n \leq k$.

(ii) At least $n$ vectors among $b_1, b_2, \ldots, b_k$ can be dropped and replaced with the vectors $a_1, a_2, \ldots, a_n$ in such a way that the resulting set has the same span as that of $b_1, b_2, \ldots, b_k$.

**Hint 97** We may assume $a_1, a_2, \ldots, a_n, b_1, b_2, \ldots, b_k$ is a basis for $S = \text{sp}(b_1, b_2, \ldots, b_k)$. Use the fact that the dimension of this space is $n + r \leq k$.

15. * The following theorem is basic for vector space theory, and is closely related to the result proved in Exercise 52, No. 5:

Let $A$ be an $m \times n$ matrix. Show that the null and column spaces of $A$ are related by the equation

$$\dim \mathcal{N}(A) + \dim \mathcal{C}(A) = n$$

**Hint 98** Let $u_1, \ldots, u_k$ be a basis for the null space $\mathcal{N}(A)$ of $A$. Extend this to a basis $u_1, \ldots, u_k, u_{k+1}, \ldots, u_n$ for the whole space $\mathbb{R}^n$. Then $Au_{k+1} \ldots Au_n$ is a basis for the column space $\mathcal{C}(A)$.

16. Deduce the result of Exercise 52, No. 5 from this theorem.
3.7 Summary of Chapter 3

3.7.1 Matrix notation; columns and rows

(See subsection 3.1.1) An $m \times n$ matrix $A = [a_{ij}]$ is also described by $A_{ij} = a_{ij}$.

$\mathbb{R}^{m \times n}$ denotes the set of $m \times n$ matrices with real number entries.

Row $i$ of the $m \times n$ matrix $A$ is written $A_{i \bullet} = \begin{bmatrix} a_{i1} & a_{i2} & \ldots & a_{in} \end{bmatrix}$ $(i = 1, \cdots, m)$

Column $j$ is written $A_{\bullet j} = \begin{bmatrix} a_{1j} \\ a_{2j} \\ \vdots \\ a_{mj} \end{bmatrix}$ $(j = 1, \cdots, n)$

Addition of two matrices of the same size is defined naturally, as is multiplication by a scalar $\alpha$:

$[\alpha A]_{ij} = \alpha [A]_{ij}$

If $A$ is $m \times n$ and $x$ a column vector with $n$ entries $x_i$, we have

$A_{\bullet}x = \begin{bmatrix} A_{1 \bullet}x \\ A_{2 \bullet}x \\ \vdots \\ A_{m \bullet}x \end{bmatrix} = A_{1 \cdot}x_1 + A_{2 \cdot}x_2 + \cdots + A_{n \cdot}x_n$

Here

$A_{\bullet}x = \begin{bmatrix} a_{i1} & a_{i2} & \cdots & a_{in} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} a_{i1} x_1 + \cdots + a_{in} x_n \end{bmatrix}$

and

$A_{\cdot j}x = \begin{bmatrix} a_{1j}x_j \\ a_{2j}x_j \\ \vdots \\ a_{mj}x_j \end{bmatrix}$

Similarly, if $y$ is a row vector with $m$ entries $y_j$,

$yA = \begin{bmatrix} yA_{1 \bullet} \\ yA_{2 \bullet} \\ \cdots \\ yA_{n \bullet} \end{bmatrix} = y_1 A_{1 \bullet} + y_2 A_{2 \bullet} + \cdots + y_m A_{m \bullet}$

3.7.2 Span, row space, column space

(See item 2 in subsection 3.1.1)

The span of the vectors $\mathbf{a}_1, \ldots, \mathbf{a}_m$ (all of the same size) is the set of linear combinations of $\mathbf{a}_1, \ldots, \mathbf{a}_m$. It is denoted by $sp(\mathbf{a}_1, \ldots, \mathbf{a}_m)$.

The row space of a matrix $A$ is the span of its rows (see also exercise 9).

The column space of a matrix $A$ is the span of its columns.
3.7.3 Matrices as mappings

(See section 3.2. Let \( A \) be a fixed \( m \times n \) matrix. If \( \mathbf{x} \) is a variable column vector with \( n \) entries, then

\[
y = A\mathbf{x}
\]

expresses \( y \) as a function of \( \mathbf{x} \); \( A \) transforms the vector \( \mathbf{x} \) into the vector \( y \). The transformation is linear since for all vectors \( \mathbf{x}_1 \) and \( \mathbf{x}_2 \) and scalars \( \alpha_1 \) and \( \alpha_2 \),

\[
A(\alpha_1\mathbf{x}_1 + \alpha_2\mathbf{x}_2) = (A\mathbf{x}_1)\alpha_1 + (A\mathbf{x}_2)\alpha_2
\]

3.7.4 The product of two matrices

(See subsection 3.3.4) Let \( A \) be an \( m \times k \) matrix and \( B \) a \( k \times n \) matrix. Then the product \( AB \) is given by

\[
[AB]_{ij} = A_{i\bullet}B_{\bullet j} \quad (i = 1, \ldots, m; j = 1, \ldots, n)
\]

Note that \( AB \) is only defined when the number of columns of \( A \) equals the number of rows of \( B \).

Then \( AB \) is \( m \times n \) and

\[
[AB]_{i\bullet} = A_{i\bullet}B \quad (i = 1, \ldots, m)
\]

\[
[AB]_{\bullet j} = AB_{\bullet j} \quad (j = 1, \ldots, n)
\]

Furthermore,

\[
A_{i\bullet}B = a_{i1}B_{1\bullet} + a_{i2}B_{2\bullet} + \cdots + a_{ik}B_{k\bullet}
\]

\[
AB_{\bullet j} = A_{1\bullet}b_{1j} + A_{2\bullet}b_{2j} + \cdots + A_{k\bullet}b_{kj}
\]

If \( \mathbf{x} \) is a variable vector with \( n \) entries, then the composite transformation \( \mathbf{x} \rightarrow A(B\mathbf{x}) \) is given by

\[
A(B\mathbf{x}) = (AB)\mathbf{x}
\]

More generally, we have the associative law for multiplication: if \( AB \) and \( BC \) are defined, then so are \((AB)C\) and \(A(BC)\) and

\[
(AB)C = A(BC)
\]

3.7.5 Linear independence

For this see section 3.4.

The columns of a matrix \( A \) are linearly independent if the equation

\[
A\mathbf{x} = 0
\]

has only the trivial solution \( \mathbf{x} = 0 \).

The rows of a matrix \( A \) are linearly independent if the equation

\[
\mathbf{y}A = 0
\]

has only the trivial solution \( \mathbf{y} = 0 \).

Elementary row operations preserve linear relations among the columns of \( A \). See 3.4.7.

The column rank of any \( m \times n \) matrix \( A \) is the maximum number of linearly independent columns. This is the number of pivot columns in a row \( EF \) of \( A \). (See subsection 3.4.8.)

The row and column ranks of \( A \) are equal. (See 3.4.20.)

The main theorem on linear independence is 3.4.22): If we add a column \( \mathbf{b} \) with \( m \) entries to an \( m \times n \) matrix \( A \) to form \([A, \mathbf{b}]\), then \( A \) and \([A, \mathbf{b}]\) have the same column rank if, and only if \( A\mathbf{x} = \mathbf{b} \) has a solution, i.e., if, and only if, \( \mathbf{b} \) is linearly dependent on the columns of \( A \).
3.7. SUMMARY OF CHAPTER 3

3.7.6 Invertible square matrices

(See subsection 3.5.3.)
An $n \times n$ matrix $A$ is invertible if there is a matrix $B$ such that $AB = BA = I$. When this is the case, $B$ is unique and $B = A^{-1}$ is called the inverse of $A$.

$A$ has an inverse if, and only if, $r(A) = n$ and there are a number of equivalent conditions for this to be the case. See 3.5.7.

3.7.7 Vector spaces

(See section 3.6.)
A subspace is a set $S \subseteq \mathbb{R}^m$ closed under the taking of linear combinations. One of the main results is the dimension theorem: Any two bases of a space $S$ have the same number of vectors.
Chapter 4

Determinants and the Cross Product

4.1 Determinants

The determinant of a $1 \times 1$ matrix $A = [a]$ is simply $|A| = a$ (not $|a|$).

4.1.1 Determinants of $2 \times 2$ matrices

Starting in Chapter 2, Exercise 39 No.3 and continuing in Chapter 3, Exercise 86, No.8 we found that

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

provided that the determinant of $A$,

$$|A| = a_{11}a_{22} - a_{21}a_{12}$$

is not zero. We will now see that for $2 \times 2$ matrices $A$,

4.1.2 $A$ has an inverse if, and only if, $|A| \neq 0$.

Proof

Let $d = |A|$ and consider the matrix

$$C = \begin{bmatrix} a_{22} & -a_{12} \\ -a_{21} & a_{11} \end{bmatrix}$$

Then, by direct multiplication,

$$CA = AC = \begin{bmatrix} |A| & 0 \\ 0 & |A| \end{bmatrix} = dI$$

where $I = I_2$ is the $2 \times 2$ identity matrix. Therefore, if $|A| \neq 0$, the matrix

$$B = \frac{1}{|A|} C = \frac{1}{d} \begin{bmatrix} a_{22} & -a_{12} \\ -a_{21} & a_{11} \end{bmatrix}$$

satisfies

$$BA = AB = I$$
and so $B$ is the inverse of $A$.
Conversely, suppose that $A^{-1}$ exists. Multiply both sides of equation (4.3) by $A^{-1}$:

$$A^{-1} (AC) = (A^{-1}A) C = IC = C = dA^{-1}$$

Now if $d = 0$, then $C = O$, and therefore $A = O$. Obviously, the zero matrix cannot have an inverse, so $d \neq 0$.

### 4.1.3 Cramer’s Rule for two equations in two unknowns

Let us use

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

to solve the system of two equations for two unknowns:

$$A \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}$$

We have $A^{-1} (A \mathbf{z}) = (A^{-1}A) \mathbf{z} = I \mathbf{z} = \mathbf{z} = A^{-1} \mathbf{b}$, so

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \frac{1}{d} \begin{bmatrix} a_{22} & -a_{12} \\ -a_{21} & a_{11} \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}$$

$$= \frac{1}{d} \begin{bmatrix} a_{22}b_1 - a_{12}b_2 \\ -a_{21}b_1 + a_{11}b_2 \end{bmatrix}$$

Or,

$$x_1 = \frac{b_1 \cdot a_{12} - b_2 \cdot a_{11}}{|A|}$$
$$x_2 = \frac{a_{11} \cdot b_1 - a_{21} \cdot b_2}{|A|}$$

Notice how column $\mathbf{b}$ replaces column $A_{\bullet j}$ in the determinant of $A$ for $x_j$.
This is known as Cramer’s Rule, after the 18th Century Swiss mathematician G. Cramer.

### 4.1.4 Properties of determinants of $2 \times 2$ matrices

Although a determinant is a number, equation (4.1) is also called a $2 \times 2$ determinant. The following properties hold for $2 \times 2$ determinants

1. The determinant of $A$ is the same as the determinant of its transpose:

$$|A^T| = |A|$$

**Proof:** If $A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$, then $A^T = \begin{bmatrix} a_{11} & a_{21} \\ a_{12} & a_{22} \end{bmatrix}$, so

$$|A^T| = a_{11}a_{22} - a_{12}a_{21} = |A|.$$

2. Interchanging rows changes the sign of the determinant:

$$\begin{vmatrix} A_{2\bullet} \\ A_{1\bullet} \end{vmatrix} = -\begin{vmatrix} A_{1\bullet} \\ A_{2\bullet} \end{vmatrix} = -|A|$$

**Proof:**

$$\begin{vmatrix} A_{2\bullet} \\ A_{1\bullet} \end{vmatrix} = \begin{vmatrix} a_{21} & a_{22} \\ a_{11} & a_{12} \end{vmatrix} = a_{21}a_{12} - a_{11}a_{22} = -|A|$$

...
3. Multiplying a row by a scalar $\lambda$ multiplies the determinant by $\lambda$:

\[
\begin{vmatrix}
\lambda A_{1*} \\
A_{2*}
\end{vmatrix} = \begin{vmatrix}
A_{1*} \\
\lambda A_{2*}
\end{vmatrix} = \lambda |A|
\]

**Proof:**

\[
\begin{vmatrix}
\lambda A_{1*} \\
A_{2*}
\end{vmatrix} = \begin{vmatrix}
\lambda a_{11} & \lambda a_{12} \\
a_{21} & a_{22}
\end{vmatrix} = \lambda a_{11}a_{22} - a_{21}\lambda a_{12} = \lambda |A|
\]

Similarly,

\[
\begin{vmatrix}
A_{1*} \\
\lambda A_{2*}
\end{vmatrix} = |A|
\]

4. Let $C = [c_1 \ c_2]$, then we have the expansion

\[
\begin{vmatrix}
A_{1*} + C \\
A_{2*}
\end{vmatrix} = \begin{vmatrix}
A_{1*} \\
A_{2*}
\end{vmatrix} + \begin{vmatrix}
C
\end{vmatrix}
\]

**Proof:**

\[
\begin{vmatrix}
A_{1*} + C \\
A_{2*}
\end{vmatrix} = \begin{vmatrix}
a_{11} + c_1 & a_{12} + c_2 \\
a_{21} & a_{22}
\end{vmatrix} = (a_{11} + c_{11})a_{22} - a_{21}(a_{12} + c_{12})
\]

\[
= a_{11}a_{22} - a_{21}a_{12} + c_{1}a_{22} - a_{21}c_{2}
\]

\[
= \begin{vmatrix}
a_{11} & a_{12} \\
a_{21} & a_{22}
\end{vmatrix} + \begin{vmatrix}
c_1 & c_2 \\
a_{21} & a_{22}
\end{vmatrix} = \begin{vmatrix}
A_{1*} \\
A_{2*}
\end{vmatrix} + \begin{vmatrix}
C
\end{vmatrix}
\]

5. If one row is a multiple of another the determinant vanishes:

\[
\begin{vmatrix}
A_{1*} \\
\lambda A_{1*}
\end{vmatrix} = 0
\]

**Proof:**

\[
\begin{vmatrix}
A_{1*} \\
\lambda A_{1*}
\end{vmatrix} = \begin{vmatrix}
a_{11} & a_{12} \\
\lambda a_{11} & \lambda a_{12}
\end{vmatrix} = a_{11}\lambda a_{12} - a_{11}\lambda a_{12} = 0.
\]

6. Adding a multiple of one row to another row leaves the determinant invariant:

\[
\begin{vmatrix}
A_{1*} \\
\lambda A_{1*} + A_{2*}
\end{vmatrix} = |A|
\]

**Proof:** By (4) and (5),

\[
\begin{vmatrix}
A_{1*} \\
\lambda A_{1*} + A_{2*}
\end{vmatrix} = \begin{vmatrix}
A_{1*} \\
\lambda A_{1*}
\end{vmatrix} + \begin{vmatrix}
A_{1*} \\
A_{2*}
\end{vmatrix} = 0 + |A| = |A|.
\]

7. If $A$ and $B$ are $2 \times 2$ matrices, then

\[
|AB| = |A| |B|
\]

**Proof:**

To better see the argument let $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$. Then using properties (2) - (5) we have:

\[
|AB| = \begin{vmatrix}
a & b \\
c & d
\end{vmatrix} = \begin{vmatrix}
ab_{1*} + bb_{2*} \\
cb_{1*} + db_{2*}
\end{vmatrix}
\]

\[
= \begin{vmatrix}
ab_{1*} \\
bb_{2*}
\end{vmatrix} + \begin{vmatrix}
bb_{2*} \\
cb_{1*} + db_{2*}
\end{vmatrix}
\]

\[
= 0 + \begin{vmatrix}
ab_{1*} - cb_{2*} \\
bb_{2*} + db_{2*}
\end{vmatrix} + 0
\]

\[
= (ad - bc) |B| = |A| |B|
\]
8. The above properties hold for columns in place of rows. In No.1 of Exercise 101 you are asked to formally state and prove this.

4.1.5 \(3 \times 3\) determinants

4.1.6 The search for a formula solving simultaneous equations

Let us try to solve formally

\[
\begin{bmatrix}
  a_{11} & a_{12} & a_{13} \\
  a_{21} & a_{22} & a_{23} \\
  a_{31} & a_{32} & a_{33}
\end{bmatrix}
\begin{bmatrix}
  x_1 \\
  x_2 \\
  x_3
\end{bmatrix}
= \begin{bmatrix}
  b_1 \\
  b_2 \\
  b_3
\end{bmatrix}
\]

The detached array for this system is

\[
\begin{bmatrix}
  a_{11} & a_{12} & a_{13} \\
  a_{21} & a_{22} & a_{23} \\
  a_{31} & a_{32} & a_{33}
\end{bmatrix}
\begin{bmatrix}
  a_{11} \\
  a_{21} \\
  a_{31}
\end{bmatrix}
= b_1
\]

Assuming \(a_{11} \neq 0\) this is equivalent to

\[
\begin{bmatrix}
  a_{11} & a_{12} & a_{13} \\
  0 & -\frac{a_{11}}{a_{11}}a_{12} + a_{22} & -\frac{a_{11}}{a_{11}}a_{13} + a_{23} \\
  0 & -\frac{a_{11}}{a_{11}}a_{12} + a_{32} & -\frac{a_{11}}{a_{11}}a_{13} + a_{33}
\end{bmatrix}
\begin{bmatrix}
  x_2 \\
  x_3
\end{bmatrix}
= \begin{bmatrix}
  b_1 \\
  b_2 \\
  b_3
\end{bmatrix}
\]

In particular, this means

\[
\begin{bmatrix}
  -\frac{a_{11}}{a_{11}}a_{12} + a_{22} & -\frac{a_{11}}{a_{11}}a_{13} + a_{23} \\
  -\frac{a_{11}}{a_{11}}a_{12} + a_{32} & -\frac{a_{11}}{a_{11}}a_{13} + a_{33}
\end{bmatrix}
\begin{bmatrix}
  x_2 \\
  x_3
\end{bmatrix}
= \begin{bmatrix}
  -\frac{a_{11}}{a_{11}}b_1 + b_2 \\
  -\frac{a_{11}}{a_{11}}b_1 + b_3
\end{bmatrix}
\]

Assuming that

\[
\begin{vmatrix}
  -\frac{a_{11}}{a_{11}}a_{12} + a_{22} & -\frac{a_{11}}{a_{11}}a_{13} + a_{23} \\
  -\frac{a_{11}}{a_{11}}a_{12} + a_{32} & -\frac{a_{11}}{a_{11}}a_{13} + a_{33}
\end{vmatrix} \neq 0
\]

we can use Cramer’s rule to find

\[
x_2 = \frac{-\frac{a_{11}}{a_{11}}b_1 + b_2 - \frac{a_{11}}{a_{11}}a_{13} + a_{23}}{-\frac{a_{11}}{a_{11}}a_{12} + a_{22} - \frac{a_{11}}{a_{11}}a_{13} + a_{23}}
\]

Multiply numerator and denominator of (4.4) by \(a_{11}^2\) and use property 3 of subsection 4.1.4 above:

\[
x_2 = \frac{a_{11}b_2 - a_{21}b_1 - a_{23}a_{11} + a_{21}a_{13}}{a_{11}b_3 - a_{31}b_1 - a_{33}a_{11} + a_{31}a_{13}}
\]

Consider the denominator:

\[
(a_{11}a_{22} - a_{21}a_{12}) (a_{33}a_{11} - a_{31}a_{13}) - (a_{11}a_{32} - a_{31}a_{12}) (a_{23}a_{11} - a_{21}a_{13})
\]

\[
= a_{11}a_{11}a_{22}a_{33} - a_{11}a_{13}a_{22}a_{31} - a_{11}a_{12}a_{21}a_{33} + a_{12}a_{21}a_{31}a_{13}
\]

\[
- a_{11}a_{13}a_{23}a_{32} + a_{11}a_{13}a_{21}a_{32} + a_{11}a_{13}a_{23}a_{31} - a_{12}a_{13}a_{21}a_{31}
\]
4.1. DETERMINANTS

\[ x_2 = \frac{a_{11}b_2a_{33} - a_{13}b_2a_{31} - a_{12}b_1a_{33} + a_{13}a_{21}b_3 + a_{12}a_{23}a_{31}}{a_{11}a_{22}a_{33} - a_{13}a_{22}a_{31} - a_{12}a_{21}a_{33} - a_{11}a_{23}a_{32} + a_{13}a_{21}a_{32} + a_{12}a_{23}a_{31}} \] (4.5)

The denominator is defined as the determinant \(|A|\) of the 3 x 3 matrix A:

\[ |A| = a_{11}a_{22}a_{33} - a_{13}a_{22}a_{31} - a_{12}a_{21}a_{33} - a_{11}a_{23}a_{32} + a_{13}a_{21}a_{32} + a_{12}a_{23}a_{31} \] (4.6)

|A| is called a 3 x 3 determinant and it is sometimes written as det A.

Equation (4.5) can be written

\[ x_2 = \frac{\begin{vmatrix} a_{11} & b_1 & a_{13} \\ a_{21} & b_2 & a_{23} \\ a_{31} & b_3 & a_{33} \end{vmatrix}}{|A|} \]

Can you guess what \(x_1\) and \(x_3\) are, if \(|A| \neq 0\)? This is again Cramer’s rule. (See exercise No.7 below).

4.1.7 Discussion on the definition of the 3 x 3 determinant |A|

Each term of (4.6) is of the form

\[ \pm a_{ij}a_{jk}a_{ik} \]

The column indices \(i, j, \) and \(k\) form an arrangement (permutation) \(ijk\) of the numbers 1, 2, and 3. In fact all six permutations occur and each has its own sign according to the scheme:

\[ \begin{array}{cccccccc} +123 & -321 & -213 & -132 & +312 & +231 \\ \end{array} \]

4.1.8 The rule is for finding the sign of \(ijk\)

Start with 123 as +. Each time an interchange of two numbers occurs, change the sign as follows:

\[ +123 \rightarrow -132 \text{ (interchange 2 and 3)} \]
\[ -132 \rightarrow +231 \text{ (interchange 1 and 2)} \]
\[ +231 \rightarrow -213 \text{ (interchange 1 and 3)} \]
\[ -213 \rightarrow +312 \text{ (interchange 2 and 3)} \]
\[ +312 \rightarrow -321 \text{ (interchange 1 and 2)} \]
4.1.9 Formal definition of $|A|$

Let $\sigma = (\sigma_1, \sigma_2, \sigma_3)$ denote a typical permutation of $1, 2, 3$ and let $\text{sgn}(\sigma)$ be its sign. Then we can write

$$|A| = \sum_\sigma \text{sgn}(\sigma)a_{1\sigma_1}a_{2\sigma_2}a_{3\sigma_3}$$ (4.7)

Here the sum extends over all $3! = 6$ permutations $\sigma$ of $1, 2, 3$.

4.1.10 The cofactor $A_{ij}$ of $a_{ij}$

The cofactor of $a_{ij}$ is defined as

$$A_{ij} = (-1)^{i+j} \text{times determinant formed by crossing out row } i \text{ and column } j$$

Example 4.1.1

$$A_{1|1} = (-1)^{1+1} \begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} = \begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix},$$

$$A_{1|2} = (-1)^{1+2} \begin{vmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{vmatrix} = - \begin{vmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{vmatrix},$$

$$A_{2|2} = (-1)^{2+2} \begin{vmatrix} a_{11} & a_{13} \\ a_{31} & a_{33} \end{vmatrix} = \begin{vmatrix} a_{11} & a_{13} \\ a_{31} & a_{33} \end{vmatrix},$$

$$A_{2|3} = (-1)^{2+3} \begin{vmatrix} a_{11} & a_{12} \\ a_{31} & a_{32} \end{vmatrix} = - \begin{vmatrix} a_{11} & a_{12} \\ a_{31} & a_{32} \end{vmatrix},$$

$$A_{3|1} = (-1)^{3+1} \begin{vmatrix} a_{12} & a_{13} \\ a_{22} & a_{23} \end{vmatrix} = \begin{vmatrix} a_{12} & a_{13} \\ a_{22} & a_{23} \end{vmatrix}.$$  

4.1.11 The sign pattern for the cofactors $A_{ij}$

Each time we move by one row or by one column we get a change in sign:

$$+ \quad - \quad + \quad - \quad + \quad -$$

4.1.12 Expansion of $|A|$ by a row

Expansion by row 1:

$$|A| = a_{11}a_{22}a_{33} - a_{11}a_{23}a_{32} + a_{12}a_{23}a_{31} - a_{12}a_{21}a_{33} + a_{13}a_{21}a_{32} - a_{13}a_{22}a_{31}$$

$$= a_{11}(a_{22}a_{33} - a_{23}a_{32}) - a_{12}(a_{21}a_{33} - a_{23}a_{31}) + a_{13}(a_{21}a_{32} - a_{22}a_{31})$$

$$= a_{11} \begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} - a_{12} \begin{vmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{vmatrix} + a_{13} \begin{vmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{vmatrix}$$

Hence

$$|A| = a_{11}A_{1|1} + a_{12}A_{1|2} + a_{13}A_{1|3}$$

Expansion by row 2:

$$|A| = a_{11}a_{23}a_{31} - a_{11}a_{23}a_{32} + a_{12}a_{23}a_{31} - a_{12}a_{21}a_{33} + a_{13}a_{21}a_{32} - a_{13}a_{22}a_{31}$$
Let
\[-a_{21}(a_{12}a_{33} - a_{13}a_{32}) + a_{22}(a_{11}a_{33} - a_{13}a_{31}) - a_{23}(a_{11}a_{32} - a_{12}a_{31})\]
\[= -a_{21} \begin{vmatrix} a_{12} & a_{13} \\ a_{32} & a_{33} \end{vmatrix} + a_{22} \begin{vmatrix} a_{11} & a_{13} \\ a_{31} & a_{33} \end{vmatrix} - a_{23} \begin{vmatrix} a_{11} & a_{12} \\ a_{31} & a_{32} \end{vmatrix}\]
\[|A| = a_{21}A_{21} + a_{22}A_{22} + a_{23}A_{23}\]

Expansion by row 3:
\[|A| = a_{11}a_{22}a_{33} - a_{11}a_{23}a_{32} + a_{12}a_{23}a_{31} - a_{12}a_{21}a_{33} + a_{13}a_{21}a_{32} - a_{13}a_{22}a_{31}\]
\[= a_{31}(a_{12}a_{23} - a_{13}a_{22}) - a_{32}(a_{11}a_{23} - a_{13}a_{21}) + a_{33}(a_{11}a_{22} - a_{12}a_{21})\]
\[= a_{31} \begin{vmatrix} a_{12} & a_{13} \\ a_{22} & a_{23} \end{vmatrix} - a_{32} \begin{vmatrix} a_{11} & a_{13} \\ a_{21} & a_{23} \end{vmatrix} + a_{33} \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix}\]

Hence:
\[|A| = a_{31}A_{31} + a_{32}A_{32} + a_{33}A_{33}\]

4.1.13 Expansion by row \(i\)

\[|A| = a_{i1}A_{i1} + a_{i2}A_{i2} + a_{i3}A_{i3}\] (4.8)

Example 4.1.2 Let
\[A = \begin{bmatrix} 1 & 3 & -1 \\ 2 & 4 & 0 \\ -2 & 5 & -3 \end{bmatrix}\]

Find \(|A|\) by expanding by the second row.

Solution:
\[|A| = a_{21}A_{21} + a_{22}A_{22} + a_{23}A_{23}\]
\[= (2)(-1)^{2+1} \begin{vmatrix} 3 & -1 \\ 5 & -3 \end{vmatrix} + (4)(-1)^{2+2} \begin{vmatrix} 1 & -1 \\ -2 & -3 \end{vmatrix}\]
\[= (-2)(-4) + (4)(-5) = -12\]

4.1.14 Properties of 3 \(\times\) 3 determinants

1. The determinant of \(A\) and its transpose are equal:
\[|A| = |A^T|\]

Proof:
\[|A| = a_{11}a_{22}a_{33} - a_{11}a_{23}a_{32} - a_{12}a_{21}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} - a_{13}a_{22}a_{31}\]
\[= a_{11}a_{22}a_{33} - a_{31}a_{22}a_{13} - a_{21}a_{12}a_{33} - a_{11}a_{32}a_{23} + a_{21}a_{32}a_{13} + a_{31}a_{12}a_{23}\]

If \(B = A^T\) then \(b_{ij} = a_{ji}\) by definition, so that \(|A|\) is equal to
\[b_{11}b_{22}b_{33} - b_{11}b_{23}b_{31} - b_{12}b_{21}b_{33} - b_{11}b_{23}b_{32} + b_{13}b_{23}b_{31} + b_{13}b_{21}b_{32}\]
The idea of the proof can be seen as follows: Interchanging two rows changes the sign of the determinant. Let 

\[
\begin{vmatrix}
2 & 3 \\
-2 & 1
\end{vmatrix}
\]

By a scalar multiplies the determinant by the same scalar: \(A_{13} = \lambda A_{13}\), then

Hence, if we interchange \(A_A\) and \(A_B\), the \(2 \times 2\) determinants will change sign and so will \(|A|\).

3. Multiplying a row of \(A\) by a scalar multiplies the determinant by the same scalar:

\[
\begin{vmatrix}
\lambda A & \lambda A \\
A & A
\end{vmatrix}
\]

Proof:

Consider for example multiplying row 2 by \(\lambda\). Expand the result by row 2:

\[
\begin{vmatrix}
\lambda A & \lambda A \\
A & A
\end{vmatrix}
= (\lambda a_{21})A_{21} + (\lambda a_{22})A_{22} + (\lambda a_{23})A_{23} = \lambda (a_{21}A_{21} + a_{22}A_{22} + a_{23}A_{23}) = \lambda |A|
\]

Less elegantly, this can also be seen by expanding by row 1:

\[
\begin{vmatrix}
\lambda A & \lambda A \\
A & A
\end{vmatrix}
= \lambda |A|
\]

We have used the fact that each cofactor gets multiplied by \(\lambda\).

4. Let \(B = \left[ \begin{array}{rr} b_1 & b_2 & b_3 \end{array} \right]\), then

\[
\begin{vmatrix}
A_{11} + B \\
A_{21} + B \\
A_{31} + B
\end{vmatrix}
= \begin{vmatrix}
A_{11} + b_1 \\
A_{21} + b_2 \\
A_{31} + b_3
\end{vmatrix} = A + B
\]

Proof:

\[
\begin{vmatrix}
A_{11} + B \\
A_{21} + B \\
A_{31} + B
\end{vmatrix}
= (a_{11} + b_1)A_{11} + (a_{12} + b_2)A_{12} + (a_{13} + b_3)A_{13}
\]

\[
= a_{11}A_{11} + a_{12}A_{12} + a_{13}A_{13} + b_1A_{11} + b_2A_{12} + b_3A_{13}
\]

\[
= \begin{vmatrix} A_{11} + B \end{vmatrix}
\]

\[
= \begin{vmatrix} A_{21} + B \end{vmatrix}
\]

\[
= \begin{vmatrix} A_{31} + B \end{vmatrix}
\]
Similar results apply if the row $B$ is added to another row, e.g.

\[
\begin{vmatrix}
A_1^* \\
A_2^* + B \\
A_3^*
\end{vmatrix} = \begin{vmatrix}
A_1^* \\
A_2^* \\
A_3^*
\end{vmatrix} + \begin{vmatrix}
A_1^* \\
B \\
A_3^*
\end{vmatrix}
\]

Using property (2) this can be proved directly from the first result as follows

\[
\begin{vmatrix}
A_1^* \\
A_2^* + B \\
A_3^*
\end{vmatrix} = -\begin{vmatrix}
A_1^* \\
A_2^* \\
A_3^*
\end{vmatrix} + \begin{vmatrix}
A_1^* \\
B \\
A_3^*
\end{vmatrix} = -\begin{vmatrix}
A_2^* + B \\
A_1^* \\
A_3^*
\end{vmatrix} + 0 = |A|
\]

5. If one row of $A$ is a multiple of another then, $|A| = 0$.

**Proof**

First suppose that two rows are equal, say $A_1^* = A_2^*$. Then interchanging rows 1 and 2 of $A$ changes the sign of the determinant by property (2). Hence $|A| = -|A|$ and so $|A| = 0$. Now let $A_1^* = \lambda A_2^*$. By property (3), $|A|$ is $\lambda$ times a determinant with equal rows 1 and 2 and so vanishes.

6. Adding a multiple of one row to a different row leaves $|A|$ unchanged, e.g.

\[
\frac{\lambda A_1^* + A_2^*}{A_3^*} = |A|
\]

**Proof:**

Use property 5:

\[
\frac{A_1^*}{\lambda A_1^* + A_2^*} = \frac{A_1^*}{A_2^*} + \frac{A_1^*}{\lambda A_1^*} = A_1^* + 0 = |A|
\]

7. If $A$ and $B$ are $3 \times 3$ matrices,

\[
|AB| = |A| |B| \quad (4.9)
\]

* We will not prove this but note that the proof follows that for $2 \times 2$ determinants.

8. Any of the above properties (2) - (6) are true for columns in place of rows.

**Proof:**

This follows from property (1), namely that $|A| = |A^T|$.

In particular, we have:

(a) The expansion of $|A|$ by column $j$:

\[|A| = A_{1j}a_{1j} + A_{2j}a_{2j} + A_{3j}a_{3j}
\]

(b) Interchanging two columns changes the sign of the determinant.

(c) Multiplying a column of $A$ by a scalar multiplies the determinant by the same scalar.

(d) If one column of $A$ is a multiple of another, then $|A| = 0$.

(e) Adding a multiple of one column to another leaves $|A|$ unchanged.
Example 4.1.3 Let

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

Find \(|A|\) by expanding by the third column.

Solution:

\[
|A| = A_{13}a_{13} + A_{23}a_{23} + A_{33}a_{33}
\]

\[
= (-1)^{1+3} \begin{vmatrix}
2 & 4 \\
-2 & 5
\end{vmatrix} (-1) + A_{23}(0) + (-1)^{3+3} \begin{vmatrix}
1 & 3 \\
2 & 4
\end{vmatrix} (-3)
\]

\[
= 18(-1) + (-2)(-3) = -12
\]

Simplification:

We may also simplify the evaluation of the determinant by adding a multiple of a row to another:

Expanding by the third column gives

\[
|A| = (-1)^{1+3} \begin{vmatrix}
2 & 4 \\
-5 & -4
\end{vmatrix} (-3)R_1 + R_3
\]

\[
|A| = (-1)^{1+3} \begin{vmatrix}
2 & 4 \\
-5 & -4
\end{vmatrix} (-1) = -12
\]

4.1.15 Fundamental properties of cofactors

For rows:

\[
a_{i1}A_{j1} + a_{i2}A_{j2} + a_{i3}A_{j3} = |A| \text{ if } i = j \quad (4.10)
\]

\[
a_{i1}A_{j1} + a_{i2}A_{j2} + a_{i3}A_{j3} = 0 \text{ if } i \neq j
\]

For columns:

\[
A_{1i}a_{1j} + A_{2i}a_{2j} + A_{3i}a_{3j} = |A| \text{ if } i = j \quad (4.11)
\]

\[
A_{1i}a_{1j} + A_{2i}a_{2j} + A_{3i}a_{3j} = 0 \text{ if } i \neq j
\]

Proof

We know that if \( i = j \) in (4.10) the left-hand side is just the expansion of \(|A|\) by row \(i\). Now let \(i \neq j\), say \(i = 2\) and \(j = 1\). Then

\[
a_{21}A_{11} + a_{22}A_{12} + a_{23}A_{13}
\]

\[
= a_{21} \begin{vmatrix}
a_{22} & a_{23} \\
a_{32} & a_{33}
\end{vmatrix} - a_{22} \begin{vmatrix}
a_{21} & a_{23} \\
a_{31} & a_{33}
\end{vmatrix} + a_{23} \begin{vmatrix}
a_{21} & a_{22} \\
a_{31} & a_{32}
\end{vmatrix}
\]

This is the expansion of

\[
\begin{vmatrix}
A_{21} \\
A_{22} \\
A_{23}
\end{vmatrix}
\]

by its first row and so vanishes.

The result for columns can be derived from that for rows.
4.1. DETERMINANTS

4.1.16 The adjoint $\text{adj}(A)$ of $A$ and the inverse $A^{-1}$

The adjoint of the $n \times n$ matrix $A$ is the $n \times n$ matrix $\text{adj}(A)$, where

$$[\text{adj}(A)]_{ij} = A_{ji}$$

This means that we can first find the matrix $B$ with $[B]_{ij} = A_{ji}$ and then $\text{adj}(A) = B^T$.

4.1.17 Properties (4.10) and (4.11) as matrix products

Properties (4.10) and (4.11) just say

$$A \text{adj}(A) = |A| I = \text{adj}(A) A$$ (4.12)

where $I = I_3$ is the $3 \times 3$ identity matrix.

**Remark 100** For a $2 \times 2$ matrix $A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$, the adjoint is $\text{adj}(A) = \begin{bmatrix} a_{22} & -a_{12} \\ -a_{21} & a_{11} \end{bmatrix}$.

This was found in equation (4.2) and the property (4.12) was found in equation (4.3).

4.1.18 Corollary: condition for $A$ to be invertible. Formula for the inverse

The $3 \times 3$ matrix $A$ is invertible if, and only if, $|A| \neq 0$, in which case

$$A^{-1} = \frac{1}{|A|} \text{adj}(A)$$ (4.13)

(For the definition of ‘invertible’, see Chapter 3, subsection 3.5.2).

**Proof**

If $|A| \neq 0$, then equation (4.12) shows that

$$A \left( \frac{1}{|A|} \text{adj}(A) \right) = I = \left( \frac{1}{|A|} \text{adj}(A) \right) A$$

In other words, (4.13) holds.

Conversely, suppose that $A^{-1}$ exists. Let an elementary row operation be done on $A$ to obtain $A'$.

By properties (2), (3) and (6) we see that $|A| \neq 0$ if, and only if, $|A'| \neq 0$. As $A^{-1}$ exists, its row REF is $I$. Since $|I| = 1$, it follows that $|A| \neq 0$.

**Example 4.1.4** Let $A = \begin{bmatrix} 2 & 4 & 3 \\ 0 & 1 & -1 \\ 3 & 5 & 7 \end{bmatrix}$

If $A^{-1}$ exists, find it.

**Solution:**

$$|A| = a_{21}A_{21} + a_{22}A_{22} + a_{23}A_{23}$$

$$= 0 + \begin{vmatrix} 23 \\ 37 \end{vmatrix} - (-1) \begin{vmatrix} 24 \\ 35 \end{vmatrix} = 3$$

As $|A| \neq 0$, $A^{-1}$ exists. We find
\[
A_{11} = \begin{vmatrix} 1 & -1 \\ 5 & 7 \end{vmatrix} = 12, \quad A_{12} = -\begin{vmatrix} 0 & -1 \\ 3 & 7 \end{vmatrix} = -3, \quad A_{13} = \begin{vmatrix} 0 & 1 \\ 3 & 5 \end{vmatrix} = -3
\]
\[
A_{21} = -\begin{vmatrix} 4 & 3 \\ 5 & 7 \end{vmatrix} = -13, \quad A_{22} = \begin{vmatrix} 2 & 3 \\ 3 & 7 \end{vmatrix} = 5, \quad A_{23} = -\begin{vmatrix} 2 & 4 \\ 3 & 7 \end{vmatrix} = 2
\]
\[
A_{31} = \begin{vmatrix} 4 & 3 \\ 1 & -1 \end{vmatrix} = -7, \quad A_{32} = -(-2) = 2, \quad A_{33} = \begin{vmatrix} 2 & 4 \\ 0 & 1 \end{vmatrix} = 2
\]

Hence

\[
A^{-1} = \frac{1}{|A|} \text{adj}A = \frac{1}{|A|} \begin{bmatrix} A_{11} & A_{12} & A_{13} \\ A_{21} & A_{22} & A_{23} \\ A_{31} & A_{32} & A_{33} \end{bmatrix}^T
\]

\[
= \frac{1}{3} \begin{bmatrix} 12 & -13 & -3 \\ -13 & 5 & 2 \\ -7 & 2 & 2 \end{bmatrix}^T
\]

\[
= \frac{1}{3} \begin{bmatrix} 12 & -13 & -7 \\ -3 & 5 & 2 \\ -3 & 2 & 2 \end{bmatrix}
\]

### 4.2 Higher Order Determinants

Let \( A \) be an \( n \times n \) matrix. As with three numbers (see subsection 4.1.7), a permutation \((\sigma_1, \sigma_2, ..., \sigma_n)\) of \(1, 2, \cdots, n\) is called **even** if it can be obtained from \(1, 2, \cdots, n\) by an even number of interchanges, otherwise **odd**.

\[\text{sign}(\sigma) = +1\] if \(\sigma\) is even, and \[\text{sign}(\sigma) = -1\] if \(\sigma\) is odd.

Following the formal definition for \(3 \times 3\) determinants 4.7, for a general \(n \times n\) matrix \(A\), its determinant \(|A|\) is given by

\[
|A| = \sum_{\sigma} \text{sgn}(\sigma)a_{1\sigma_1}a_{2\sigma_2}...a_{n\sigma_n}
\]

where the sum extends over all \(n!\) permutations \(\sigma\) of the numbers \(1, 2, \cdots, n\).

Recall that \(n! = (1) \cdot (2) \cdot (3) \cdots (n - 1) \cdot (n)\) and is called \(n\) **factorial**.

As with \(2 \times 2\) and \(3 \times 3\) determinants (4.1.10 above), if \(A\) is an \(n \times n\) matrix, the cofactor \(A_{ij}\) of \(a_{ij}\) is given by

\[A_{ij} = (-1)^{i+j}\] times the determinant formed by crossing out row \(i\) and column \(j\)

For a \(4 \times 4\) determinant the sign pattern for the cofactors is

\[
\begin{pmatrix}
+&-&+&-\\
-&+&-&+\\
+&-&+&-\\
-&+&+-
\end{pmatrix}
\]

### 4.2.1 All the properties of \(2 \times 2\) and \(3 \times 3\) determinants hold for \(n \times n\) determinants

So properties in 4.1.14, 4.1.13 (equation 4.8) carry over to \(n \times n\) determinants. In particular, the formula \(|AB| = |A||B|\) and equation (4.12) hold as well as the statement on the inverse in Corollary 4.1.18 and equation (4.13).
Example 4.2.1 Find $|A|$ if

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

Solution:

$$|A| = \begin{vmatrix} 3 & -2 & -3 & 6 \\ 1 & -2 & 2 & 4 \\ -1 & 3 & 4 & -9 \\ 0 & 1 & 0 & 0 \end{vmatrix}$$

Here we have added column 2 to column 3 and $(-2)$(Column 2) to column 4. Expanding by row 4:

$$|A| = (-1)^{4+2} \begin{vmatrix} 3 & -3 & 6 \\ 1 & 2 & 4 \\ -1 & 4 & -9 \end{vmatrix} = \begin{vmatrix} 0 & -3 & 0 \\ 3 & 2 & 8 \\ 3 & 4 & -1 \end{vmatrix} = (-1)^{1+2}(-3) \begin{vmatrix} 3 & 8 \\ 3 & -1 \end{vmatrix} = -81$$

We have added column 2 to column 1 and 2(column 2) to column 3 then expanded by row 1.

Exercise 101 Partial Solutions

1. Using property (1) for $2 \times 2$ determinants, state and prove similar results for columns in (2)-(6).

2. Evaluate the determinants of the matrices in Chapter 3, Exercise 86, No. 1. Find the inverses of the matrices by the adjoint method if they exist.

The values of the the determinants are:

$$|A| = \begin{vmatrix} 2 & 1 \\ 3 & 2 \end{vmatrix} = 1$$

$$|B| = \begin{vmatrix} 3 & 2 & 1 \\ 0 & 2 & 2 \\ 0 & 0 & -1 \end{vmatrix} = -6$$

$$|Y| = \begin{vmatrix} 1 & 0 & 4 \\ -1 & 1 & -2 \\ 2 & 3 & 1 \end{vmatrix} = -13$$

$$|T| = \begin{vmatrix} \frac{1}{2} & -\frac{1}{3} & -\frac{13}{5} \\ \frac{3}{2} & \frac{5}{3} & \frac{10}{3} \\ \frac{1}{3} & \frac{5}{6} & \frac{2}{3} \end{vmatrix} = \frac{13}{6}$$

$$|C| = \begin{vmatrix} 1 & 6 & 2 \\ -2 & 3 & 5 \\ 7 & 12 & -4 \end{vmatrix} = 0$$

$$|D| = \begin{vmatrix} 1 & 3 & -1 \\ 2 & 0 & -3 \\ -1 & 4 & 2 \end{vmatrix} = 1$$

3. Evaluate the determinants
4. Verify that the adjoint of a $2 \times 2$ matrix is given by equation (4.2).

5. Let $A$ be a $3 \times 3$ matrix and $k$ a scalar. Show that $|kA| = k^3 |A|$. What do you think the general result is?

6. Suppose the $n \times n$ matrix $A$ is not invertible. Show that $A(\text{adj } A)$ is the zero matrix.

7. State and prove Cramer’s rule for 3 equations in 3 unknowns. Do the same for $n$ equations in $n$ unknowns using the fact that higher order determinants have the same properties as 2 and 3 determinants.

8. (a) Let $A = [a_{ij}]$ be a $3 \times 3$ matrix and $\beta$ a scalar. Replace the entry $a_{23}$ with $\beta$ and let the resulting matrix be $B$. Show that $|B| = |A| + (\beta - a_{23}) A_{23}$, and find a criterion for $B$ to be invertible.

(b) Generalize the previous result.

9. Let $A$ be a $3 \times 3$ matrix. A scalar $\lambda$ such that $A - \lambda I$ is not invertible is called an eigenvalue of $A$ (see No.5 in Examples 3.6.1). This means that $|A - \lambda I| = 0$. Show that $\lambda$ must satisfy a polynomial equation $f(\lambda) = 0$ of degree 3 (called the characteristic polynomial of $A$). Find this polynomial for the matrix

$$A = \begin{bmatrix} 10 & -12 & -1 \\ 5 & -6 & -1 \\ -11 & 12 & 0 \end{bmatrix}$$

as well as the eigenvalues and three corresponding eigenvectors.

**Solution:**

$f(\lambda) = \lambda^3 - 4\lambda^2 + \lambda + 6$. The eigenvalues are $\lambda = 2$, $\lambda = 3$ and $\lambda = -1$. Eigenvectors belonging to these eigenvalues are respectively

$$\begin{bmatrix} \frac{1}{2} & 1 & -\frac{3}{2} \end{bmatrix}^T$$

and

$$\begin{bmatrix} \frac{1}{2} & 1 & -\frac{3}{2} \end{bmatrix}^T$$

10. Let $|A|$ be a $3 \times 3$ determinant. We can permute the rows $A_{1\bullet}$, $A_{2\bullet}$ and $A_{3\bullet}$ in six ways. For which of these does the determinant keep its sign and for which does it change its sign?

State and prove a similar result for the columns $A_{\bullet 1}$, $A_{\bullet 2}$ and $A_{\bullet 3}$ of $A$.

### 4.3 The Cross Product $\mathbf{a} \times \mathbf{b}$

In this section all vectors have three entries and, as in the notation of Chapter 1, will normally be row vectors. The unit vectors $\mathbf{i} = (1,0,0)$, $\mathbf{j} = (0,1,0)$ and $\mathbf{k} = (0,0,1)$ were introduced in
subsection 1.1.17.
Let \( \mathbf{a} = (a_1, a_2, a_3) \) and \( \mathbf{b} = (b_1, b_2, b_3) \) are two vectors. Their cross product is defined as

\[
\mathbf{a} \times \mathbf{b} = \begin{vmatrix}
  a_2 & a_3 \\
  b_2 & b_3
\end{vmatrix} \mathbf{i} - \begin{vmatrix}
  a_1 & a_3 \\
  b_1 & b_3
\end{vmatrix} \mathbf{j} + \begin{vmatrix}
  a_1 & a_2 \\
  b_1 & b_2
\end{vmatrix} \mathbf{k}
\] (4.14)

This is sometimes called the vector product and is used constantly in mechanics.

Treating \( \mathbf{i}, \mathbf{j}, \mathbf{k} \) as though they are scalars (which they are not), the cross product can be remembered by the formula

\[
\mathbf{a} \times \mathbf{b} = \begin{vmatrix}
  \mathbf{i} & \mathbf{j} & \mathbf{k} \\
  a_1 & a_2 & a_3 \\
  b_1 & b_2 & b_3
\end{vmatrix}
\]

Expanding this ‘determinant’ by its first row gives (4.14).

If \( \mathbf{a} \) and \( \mathbf{b} \) are written as columns the convention is to write

\[
\mathbf{a} \times \mathbf{b} = \begin{vmatrix}
  \mathbf{i} & \mathbf{j} & \mathbf{k} \\
  a_1 & a_2 & a_3 \\
  b_1 & b_2 & b_3
\end{vmatrix}
\]

but we get the same result as before.

**Example 4.3.1**

\[
(-1, 2, 4) \times (2, -3, 1) = \begin{vmatrix}
  \mathbf{i} & \mathbf{j} & \mathbf{k} \\
  -1 & 2 & 4 \\
  2 & -3 & 1
\end{vmatrix}
\]

\[
= \begin{vmatrix}
  2 & 4 \\
  -3 & 1
\end{vmatrix} \mathbf{i} - \begin{vmatrix}
  -1 & 4 \\
  2 & 1
\end{vmatrix} \mathbf{j} + \begin{vmatrix}
  -1 & 2 \\
  2 & -3
\end{vmatrix} \mathbf{k}
\]

\[
= (14, 9, -1)
\]

**Example 4.3.2** A characteristic property of \( \mathbf{i}, \mathbf{j}, \mathbf{k} \)

\[
\mathbf{i} \times \mathbf{j} = \mathbf{k}, \quad \mathbf{j} \times \mathbf{k} = \mathbf{i}, \quad \mathbf{k} \times \mathbf{i} = \mathbf{j}
\]

To see these results, consider

\[
\mathbf{i} \times \mathbf{j} = \begin{vmatrix}
  \mathbf{i} & \mathbf{j} & \mathbf{k} \\
  1 & 0 & 0 \\
  0 & 1 & 0
\end{vmatrix}
\]

\[
= \begin{vmatrix}
  0 & 0 \\
  1 & 0
\end{vmatrix} \mathbf{i} - \begin{vmatrix}
  0 & 0 \\
  1 & 1
\end{vmatrix} \mathbf{j} + \begin{vmatrix}
  1 & 1 \\
  0 & 1
\end{vmatrix} \mathbf{k} = \mathbf{k}
\]

\[
\mathbf{j} \times \mathbf{k} = \begin{vmatrix}
  \mathbf{i} & \mathbf{j} & \mathbf{k} \\
  0 & 1 & 0 \\
  0 & 0 & 1
\end{vmatrix}
\]

\[
= \begin{vmatrix}
  1 & 0 \\
  0 & 1
\end{vmatrix} \mathbf{i} - \begin{vmatrix}
  1 & 0 \\
  0 & 1
\end{vmatrix} \mathbf{j} + \begin{vmatrix}
  1 & 1 \\
  1 & 0
\end{vmatrix} \mathbf{k} = \mathbf{i}
\]

\[
\mathbf{k} \times \mathbf{i} = \begin{vmatrix}
  \mathbf{i} & \mathbf{j} & \mathbf{k} \\
  0 & 0 & 1 \\
  1 & 0 & 0
\end{vmatrix}
\]

\[
= \begin{vmatrix}
  0 & 1 \\
  0 & 1
\end{vmatrix} \mathbf{i} - \begin{vmatrix}
  0 & 1 \\
  0 & 1
\end{vmatrix} \mathbf{j} + \begin{vmatrix}
  0 & 1 \\
  1 & 0
\end{vmatrix} \mathbf{k} = \mathbf{j}
\]
4.3.1 Properties of the cross product

The cross product (4.14) is quite different from the dot product \( \mathbf{a} \cdot \mathbf{b} \) in that it is a \textbf{vector} while the dot product is a \textbf{scalar}. Nevertheless, some properties are shared with the dot product.

For all vectors \( \mathbf{a}, \mathbf{b} \) and \( \mathbf{c} \) and scalars \( \beta \) and \( \gamma \):

1. \[ \mathbf{b} \times \mathbf{a} = - (\mathbf{a} \times \mathbf{b}) \]

   \textbf{Proof:}

   In \( \mathbf{b} \times \mathbf{a} \) we interchange the rows in the cofactors defining \( \mathbf{a} \times \mathbf{b} \). Since these all change sign, the result follows.

2. (The triple scalar product)

   \[ \mathbf{c} \cdot (\mathbf{a} \times \mathbf{b}) = \mathbf{b} \cdot (\mathbf{c} \times \mathbf{a}) = \mathbf{a} \cdot (\mathbf{b} \times \mathbf{c}) \]  
   \hspace{1cm} (4.15)

   \textbf{Proof:}

   First note that

   \[ \mathbf{c} \cdot (\mathbf{a} \times \mathbf{b}) = c_1 \begin{vmatrix} a_2 & a_3 \\ b_2 & b_3 \end{vmatrix} - c_2 \begin{vmatrix} a_1 & a_3 \\ b_1 & b_3 \end{vmatrix} + c_3 \begin{vmatrix} a_1 & a_2 \\ b_1 & b_2 \end{vmatrix} \]  
   \hspace{1cm} (4.16)

   Here we have expanded the determinant by its first row. The other properties follow by rearranging the rows of the above determinant according to the permutations \( bca \) and \( abc \) of \( cab \). These are all even and so the other determinants are all equal. (See Exercise 101, No.10).

3. \[ \mathbf{a} \cdot (\mathbf{a} \times \mathbf{b}) = \mathbf{b} \cdot (\mathbf{a} \times \mathbf{b}) = 0 \]

   \textbf{Proof:}

   The statement \( \mathbf{a} \cdot (\mathbf{a} \times \mathbf{b}) = 0 \) follows by putting \( \mathbf{c} = \mathbf{a} \) in the determinant of (4.16). The resulting determinant is zero as two of its rows are the same. The result \( \mathbf{b} \cdot (\mathbf{a} \times \mathbf{b}) = 0 \) is similar. Geometrically, this means that if \( \mathbf{a} \times \mathbf{b} \) is non-zero it is perpendicular to both \( \mathbf{a} \) and \( \mathbf{b} \).

4. \[ \beta(\mathbf{a} \times \mathbf{b}) = (\beta \mathbf{a}) \times \mathbf{b} = \mathbf{a} \times (\beta \mathbf{b}) \]

5. \[ (\beta + \gamma)(\mathbf{a} \times \mathbf{b}) = \beta(\mathbf{a} \times \mathbf{b}) + \gamma(\mathbf{a} \times \mathbf{b}) \]

6. \[ (\alpha \mathbf{a} + \beta \mathbf{b}) \times \mathbf{c} = (\alpha \mathbf{a} \times \mathbf{c}) + (\beta \mathbf{b} \times \mathbf{c}) \]

7. If either \( \mathbf{a} = \mathbf{0} \) or \( \mathbf{b} = \mathbf{0} \) then \( \mathbf{a} \times \mathbf{b} = \mathbf{0} \). Otherwise, \( \mathbf{a} \times \mathbf{b} = \mathbf{0} \) if, and only if, \( \mathbf{a} \) and \( \mathbf{b} \) are parallel. (See Exercise 102 No. 7)
4.3. THE CROSS PRODUCT $\mathbf{A} \times \mathbf{B}$

4.3.2 Geometric meaning of the cross product $\mathbf{a} \times \mathbf{b}$

Suppose that $0 < \theta < \pi$ is the angle between the non-zero and non-parallel vectors $\mathbf{a}$ and $\mathbf{b}$. Let $\mathbf{c}$ be the unit vector perpendicular to $\mathbf{a}$ and $\mathbf{b}$ with direction found by using the corkscrew rule: rotate from $\mathbf{a}$ to $\mathbf{b}$ through $\theta$, then the direction of $\mathbf{c}$ goes the way a corkscrew would go. See Chapter 1, subsection 1.1.1 and Figure 4.1.

Geometric Interpretation

\[
\mathbf{a} \times \mathbf{b} = (|\mathbf{a}| |\mathbf{b}| \sin \theta) \mathbf{c}
\]  

(4.17)

Proof of equation (4.17):

We first show that

\[
|\mathbf{a} \times \mathbf{b}|^2 + (\mathbf{a} \cdot \mathbf{b})^2 = |\mathbf{a}|^2 |\mathbf{b}|^2
\]

We have

\[
|\mathbf{a} \times \mathbf{b}|^2 + (\mathbf{a} \cdot \mathbf{b})^2 = \left( \begin{array}{c} a_2b_3 - a_3b_2 \\ b_2a_3 - b_3a_2 \\ c_2a_3 - c_3a_2 \\ \end{array} \right) \cdot \left( \begin{array}{c} a_2b_3 - a_3b_2 \\ b_2a_3 - b_3a_2 \\ c_2a_3 - c_3a_2 \\ \end{array} \right) + (a_1b_1 + a_2b_2 + a_3b_3)^2
\]

\[
= (a_1b_1 + a_2b_2 + a_3b_3)^2
\]

\[
= (a_1^2b_1^2 + a_2^2b_2^2 + a_3^2b_3^2 + a_1^2b_2^2 + a_2^2b_3^2 + a_1^2b_3^2 + a_2^2b_1^2 + a_3^2b_2^2 + a_3^2b_1^2 + a_2^2b_3^2 + a_3^2b_2^2 + a_3^2b_1^2)
\]

It follows that

\[
\frac{|\mathbf{a} \times \mathbf{b}|^2}{|\mathbf{a}|^2 |\mathbf{b}|^2} = 1 - \frac{(\mathbf{a} \cdot \mathbf{b})^2}{|\mathbf{a}|^2 |\mathbf{b}|^2} = 1 - \cos^2 \theta = \sin^2 \theta
\]

Hence $|\mathbf{a} \times \mathbf{b}| = |\mathbf{a}| |\mathbf{b}| \sin \theta$. Since $\mathbf{a} \times \mathbf{b}$ is parallel to $\mathbf{c}$, we have $\mathbf{a} \times \mathbf{b} = \pm (|\mathbf{a}| |\mathbf{b}| \sin \theta) \mathbf{c}$.

If $\mathbf{a} = \mathbf{i}$ and $\mathbf{b} = \mathbf{j}$, then from Example 4.3.2 we have from the Corkscrew Rule $\mathbf{c} = \mathbf{k}$ and $\mathbf{i} \times \mathbf{j} = \pm \mathbf{k}$. This suggests that the corkscrew rule applied to any cross product $\mathbf{a} \times \mathbf{b}$ gives the same direction as $\mathbf{c}$, and we will be satisfied with this. Equation (4.17) now follows.

From Example 4.3.2 we see how the geometric interpretation gives, besides $\mathbf{i} \times \mathbf{j} = \mathbf{k}$, also the other products $\mathbf{j} \times \mathbf{k} = \mathbf{i}$ and $\mathbf{k} \times \mathbf{i} = \mathbf{j}$.

Compare this with 1.1.1 on right-handed systems in Chapter 1.

**Exercise 102**

1. Find $(-1, 5, 7) \times (2, -3, 4)$ and $(-2, 4, -3) \times (6, -1, 2)$.

**Solution:**


140

CHAPTER 4. DETERMINANTS AND THE CROSS PRODUCT

\((-1, 5, 7) \times (2, -3, 4) = (41, 18, -7)\)
\((-2, 4, -3) \times (6, -1, 2) = (5, -14, -22)\)

2. Find formulae for \(i \times a\), \(j \times a\) and \(k \times a\).

3. Consider three points \(A\), \(B\) and \(C\) in space. If they are not collinear they form a triangle. Show that its area is \(\frac{1}{2} |AB \times AC|\).

4. Find the area of the triangle with vertices \(A = (3, 4, 5)\), \(B = (4, 4, 6)\), \(C = (5, 5, 4)\).

5. Show that if the non-zero vectors \(a\) and \(b\) are at right angles then \(|a \times b| = |a||b|\). What of the converse?

6. Let \(l\) and \(m\) be unit vectors and put \(n = l \times m\). Using the geometric interpretation of the cross product show that \(l\), \(m\) and \(n\) form a right-handed system (see 1.1.1). In fact,

\[l \times m = n\]
\[m \times n = l\]
\[n \times l = m\]

Compare this with Example 4.3.2.

7. Complete the proofs of the properties 4.3.1 of the cross product.

**Hint 103** For property 7, consider the case when \(a\) and \(b\) are both non-zero. If they are parallel, use properties of \(2 \times 2\) determinants in 4.1.4 to show that the cross product is zero. For the converse use the geometric interpretation (4.17) to show that \(a \times b \neq 0\) when the vectors are not parallel.

8. Let \(a\), \(b\) and \(c\) be three vectors in \(\mathbb{R}^3\). Show that they are linearly independent if, and only if, the triple scalar product \(a \cdot (b \times c)\) is not zero.

9. *Consider the parallelogram \(ABCD\) in Figure 4.2. The line \(EF\) passing through \(Q\) is parallel to \(AB\) and the line \(GH\) passing through \(Q\) is parallel to \(AD\). Using vector methods show that the parallelograms \(E\), \(Q\), \(H\) and \(G\), \(B\), \(F\) have equal areas if, and only if, \(Q\) lies on the diagonal \(AC\). Conclude that this is the case if, and only if, the parallelograms \(AGQE\) and \(QFCH\) are similar. (A theorem going back to Euclid).

**Hint 104** Consider the cross product \(AQ \times QC\).

10. A parallelepiped is a ‘squashed box’: it has six faces with opposite faces congruent parallelograms. Let \(\mathbf{u} = AD\), \(\mathbf{v} = AB\), and \(\mathbf{w} = AE\) be three of its linearly independent edges. Show that its volume is \(|(\mathbf{u} \times \mathbf{v}) \cdot \mathbf{w}|\).
4.3. THE CROSS PRODUCT $A \times B$

4.3.1. THE CROSS PRODUCT

$A \times B$

Figure 4.2 Parallelepiped

$h = \frac{w (u \times v)}{|u \times v|}$

**Hint 105** Let its base be $ABCD$. This has area $\alpha = |AD \times AB|$. The height $h$ of the parallelepiped is the absolute value of the component of $AE$ along $AD \times AB$. The volume is then $\alpha h$. Make a drawing.

11. * (The triple vector product). Let $a$, $b$ and $c$ be three given vectors. Show that

$$a \times (b \times c) = (a \cdot c)b - (a \cdot b)c$$

The result is not altogether surprising since $a \times (b \times c)$ is a linear combination of $b$ and $c$.

**Hint 106** Show the result is true if $a$ is $\hat{i}$, $\hat{j}$ or $\hat{k}$. Now use the result of No. 2.

12. * Application to torques (moments) of forces.

Let $\mathbf{F}$ be a force vector that is applied to a point $P$ in space. The **torque** (or **moment**) of $\mathbf{F}$ about the point $Q$ is defined as $\mathbf{M} = \mathbf{QP} \times \mathbf{F}$. The **line of action** of the force is the straight line $\ell$ parallel to $\mathbf{F}$ that passes through $P$. Show that the torques of $\mathbf{F}$ applied at any two points on its line of action about $Q$ are equal.

Let $H$ be the point on $\ell$ closest to $Q$. Then with $\mathbf{F}$ acting at $H$ we get the usual definition of 'moment', except that it now has a vector quality; its direction is perpendicular to both $\mathbf{QH}$ and $\mathbf{F}$.

**Hint 107** Express $\mathbf{QD}$ as a linear combination $\lambda \mathbf{l} + \mu \mathbf{m}$ and $\mathbf{F}$ as a linear combination $\alpha \mathbf{l} + \gamma \mathbf{n}$. Now multiply out the expression $\mathbf{M} = (\lambda \mathbf{l} + \mu \mathbf{m}) \times (\alpha \mathbf{l} + \gamma \mathbf{n})$ and find $\mathbf{M} \cdot \mathbf{l}$.

13. * (Continuation) Consider an axis $\xi$ passing through $Q$ and having the same direction as a given unit vector $\mathbf{l}$. Let $D$ and $E$ be points on the line of action of $\mathbf{F}$ and the axis $\xi$ respectively such that the distance $|ED|$ is minimum. (Thus $ED$, if not zero, is perpendicular to both lines.) Let $\mathbf{m}$ be the unit vector with direction $ED$, so that $ED = \mu \mathbf{m}$, where $\mu = |ED|$. Put $\mathbf{n} = \mathbf{l} \times \mathbf{m}$. In the sense of the right-hand rule, the torque of $\mathbf{F}$ about the **axis** $\xi$ is the component $\mathbf{F} \cdot \mathbf{n}$ times the distance $\mu = |ED|$ (draw a picture to see this).

Show that the torque about the axis $\xi$ is the **triple scalar product** $\mathbf{M} \cdot \mathbf{l}$.

Show further that if $P'$ and $Q'$ are any points on the line of action of $\mathbf{F}$ and the axis $\xi$ respectively, then

$$(Q'P' \times \mathbf{F}) \cdot \mathbf{l} = (QP \times \mathbf{F}) \cdot \mathbf{l}$$

This shows that the torque of a force about an axis is purely a property of the force with its line of action and the axis and does not depend on selected points on these lines.

**Hint 107** Express $\mathbf{QD}$ as a linear combination $\lambda \mathbf{l} + \mu \mathbf{m}$ and $\mathbf{F}$ as a linear combination $\alpha \mathbf{l} + \gamma \mathbf{n}$. Now multiply out the expression $\mathbf{M} = (\lambda \mathbf{l} + \mu \mathbf{m}) \times (\alpha \mathbf{l} + \gamma \mathbf{n})$ and find $\mathbf{M} \cdot \mathbf{l}$. 
4.4 Summary of Chapter 4

4.4.1 Determinants

Determinants are inspired by the search for a formula to solve \( n \) equations for \( n \) unknowns. (See, for example, subsection 4.1.6).

4.4.2 2 × 2 determinants

In Exercise 3a of Chapter 2 we introduced the idea of the determinant: \( |A| = a_{11}a_{22} - a_{21}a_{12} \).
In subsection 4.1.4 of 4.1.1 we develop some basic properties of 2 × 2 determinants.

4.4.3 3 × 3 determinants

For a 3 × 3 matrix \( A \), its determinant is given by equation (4.7):

\[
|A| = \sum_{\sigma} \text{sgn}(\sigma)a_{1\sigma_1}a_{2\sigma_2}a_{3\sigma_3}
\]

Here the sum extends over all \( \sigma = 3! = 6 \) permutations of 1, 2, 3 and \( \text{sgn}(\sigma) \) is the sign of the permutation \( \sigma = (\sigma_1, \sigma_2, \sigma_3) \).
A similar definition holds for an \( n \times n \) matrix \( A \).

4.4.4 Cofactors and properties of 3 × 3 determinants

(See subsections 4.1.10, 4.1.13, 4.1.14) The cofactor of \( a_{ij} \) is \(-1^{i+j}\) times the determinant formed by deleting row \( i \) and column \( j \).
A determinant can be expanded by any row:

\[
|A| = a_{i1}A_{i1} + a_{i2}A_{i2} + a_{i3}A_{i3}
\]  
(4.18)

(See equation (4.8).
Multiplying a row of \( A \) by a number \( k \) multiplies the determinant by \( k \).
Interchanging two rows (columns) changes the sign of the determinant.
Adding a row (column) of \( A \) to another row (column) does not change the determinant.
To each property of rows there is a corresponding property of columns.

4.4.5 The adjoint

(See subsection 4.1.16)
The adjoint of \( A \) is given by

\[
[\text{adj} (A)]_{ij} = A_{ji}
\]
This means that we can first find the matrix \( B \) with \( [B]_{ij} = A_{ji} \) and then \( \text{adj}(A) = B^T \).
The fundamental property is

\[
A \text{adj} (A) = |A| I = \text{adj} (A) A
\]
Consequently, \( A \) has an inverse if, and only if, \( |A| \neq 0 \), in which case

\[
A^{-1} = \frac{1}{|A|} \text{adj} (A)
\]

4.4.6 General determinants

All properties of 2 × 2 and 3 × 3 carry over to \( n \times n \) determinants. (See section 4.2).
4.4.7 The Cross Product

(See section 4.3).

The cross product is defined in equation 4.14

\[
\mathbf{a} \times \mathbf{b} = \begin{vmatrix}
  a_1 & a_2 & a_3 \\
  b_1 & b_2 & b_3 \\
  1   & 1   & 1 \\
\end{vmatrix}
\]

4.4.8 The Triple Scalar Product of \( \mathbf{a}, \mathbf{b} \) and \( \mathbf{c} \)

This is

\[
\mathbf{a} \cdot (\mathbf{b} \times \mathbf{c})
\]

Basic properties are given in equations (4.15) and (4.16).

4.4.9 Geometric interpretation

\[
\mathbf{a} \times \mathbf{b} = (|\mathbf{a}| |\mathbf{b}| \sin \theta) \mathbf{\hat{c}}
\]

Here \( 0 < \theta < \pi \) is the angle between the non-zero and non-parallel vectors \( \mathbf{a} \) and \( \mathbf{b} \) and \( \mathbf{\hat{c}} \) is the unit vector perpendicular to \( \mathbf{a} \) and \( \mathbf{b} \) with direction found by using the corkscrew rule: rotate from \( \mathbf{a} \) to \( \mathbf{b} \) through \( \theta \).

The cross product can also be seen geometrically as an area (exercise 3); the triple scalar product as a volume as a volume (exercise 10).