UNIVERSITY OF PRIMORSKA
Faculty of Mathematics, Natural Sciences and Information Technologies

ALGEBRA I

Lecture notes

2016/2017

Table of contents

1 Vectors **2**
- 1.1 Three-dimensional coordinate system . . . 2
- 1.2 Vectors in $\mathbb{R}^3$ . . . 6
  - 1.2.1 Component form . . . 7
  - 1.2.2 Vector algebra operations . . . 8
  - 1.2.3 Unit vectors . . . 12
- 1.3 The dot product . . . 12
  - 1.3.1 Geometrical meaning of the dot product . . . 13
- 1.4 The cross product . . . 17
  - 1.4.1 Geometrical meaning of the cross product . . . 20
- 1.5 The box product . . . 22
  - 1.5.1 Geometrical meaning of the box product . . . 23

2 Lines and planes in $\mathbb{R}^3$ **24**
- 2.1 Intersections . . . 31
- 2.2 Distances . . . 33
- 2.3 Angles . . . 39

3 Matrices **42**
- 3.1 An Introduction to Gaussian Elimination . . . 43
- 3.2 The algebra of matrices . . . 47
  - 3.2.1 Matrix addition . . . 47
  - 3.2.2 Multiplication of a matrix by a scalar . . . 49
  - 3.2.3 Matrix multiplication . . . 50
  - 3.2.4 The transpose of a matrix . . . 52
- 3.3 Systems of linear equations . . . 54
- 3.4 Invertible matrices . . . 67
- 3.5 The determinant of a square matrix . . . 72
  - 3.5.1 Permutations and the definition of a determinant . . . 72
  - 3.5.2 Basic properties of determinants . . . 78
  - 3.5.3 Application of determinants . . . 81
- 3.6 Application of matrices . . . 86
  - 3.6.1 In analytic geometry - an example . . . 86

1
VECTORS

To apply calculus in many real-world situation and in higher mathematics, we need a mathematical description of three-dimensional space. In this chapter we introduce three-dimensional coordinate systems and vectors. Building on what we already know about the coordinates in the two-dimensional coordinate system (the xy-plane), we establish coordinates in space by adding a third axis that measures distance above and below the xy-plane. Vectors are used to study the analytic geometry of space, where they give simple ways to describe lines and planes in space, as we will see in Chapter 2.

1.1 Three-dimensional coordinate system

How do we locate a given point

- on a line?
  By equipping a line with a coordinate system, each point of the line corresponds to a number (coordinate) that tells us exactly, where the given point lays.
- on a plane?
  By equipping a plane with a coordinate system, consisting of two perpendicular lines equipped with a coordinate system and intersecting at 0.
- in space?
  By equipping the space with a coordinate system, consisting of three mutually perpendicular lines equipped with a coordinate system, all intersecting at 0.

To locate a point in space we use three mutually perpendicular lines that we call *coordinate axes*, arranged as in Figure 1.1.

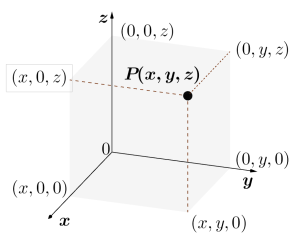
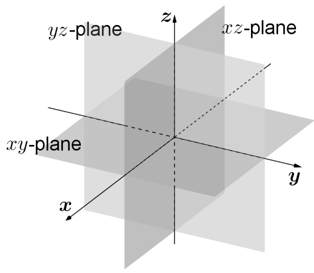


Figure 1.1: The Cartesian coordinate system in space

The *Cartesian coordinates* (x, y, z) of a point P in space are the values at which the planes through P perpendicular to the axes cut the axes. Cartesian coordinates for space are also called *rectangular coordinates* because the axes that define them meet at right angles. Points on the x-axes have y- and z-coordinates equal to 0. That is, they have coordinates of the form $(x, 0, 0)$. Similarly, points on the y-axes have coordinates of the form $(0, y, 0)$, and points on the z-axes have coordinates of the form $(0, 0, z)$.

The planes determined by the coordinates axes are the xy-plane, whose standard equation is $z = 0$; the yz-plane, whose standard equation is $x = 0$; and the xz-plane, whose standard equation is $y = 0$. They meet at the *origin* $(0, 0, 0)$. The origin is also identified by simply 0 or sometimes by the letter O.

The three coordinate planes, $x = 0$, $y = 0$ and $z = 0$ divide space into eight cells called *octants*. The octant in which the point coordinates are all positive is called the *first octant*.



In the following examples, we match coordinate equations and inequalities with the sets of points they define in space.

Example 1.1. *We interprete these equations and inequalities geometrically.*

(a) $z \geq 0$	The half-space consisting of the points on and above the xy-plane.
(b) $x = -3$	The plane perpendicular to the x-axes at $x = -3$. This plane lies parallel to the yz-plane and 3 units behind it.
(c) $z = 0,\ x \leq 0,\ y \geq 0$	The second quadrant of the xy-plane.
(d) $x \geq 0,\ y \geq 0,\ z \geq 0$	The first octant.
(e) $-1 \leq y \leq 1$	The slab between the planes $y = -1$ and $y = 1$ (planes included).
(f) $y = -2,\ z = 2$	The line in which the planes $y = -2$ and $z = 2$ intersect. Alternatively, the line through the point $(0, -2, 2)$ parallel to the x-axis.

Example 1.2. *What points $P(x, y, z)$ satisfy the equations*

$$x^2 + y^2 = 4 \quad \textit{and} \quad z = 3?$$

Solution. The points lie in the horizontal plane $z = 3$, and, in this plane, make up the circle $x^2 + y^2 = 4$. See Figure 1.2.

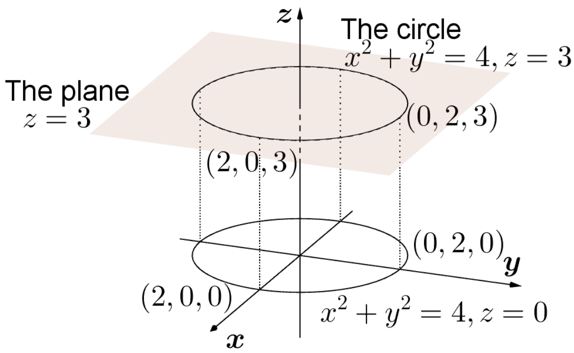


Figure 1.2: The circle $x^2 + y^2 = 4$ in the plane $z = 3$ (Example 1.2).

Distance in space

The formula for the distance between two points in the xy-plane extends to points in space.

The distance between $P_1(x_1, y_1, z_1)$ and $P_2(x_2, y_2, z_2)$ is

$$|P_1P_2| = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2}$$

Proof. Let $P_1(x_1, y_1, z_1)$ and $P_2(x_2, y_2, z_2)$ be two points in space. Construct a rectangular box with faces parallel to the coordinate planes and the points P_1 and P_2 at opposite corners of the box (see Figure 1.3).

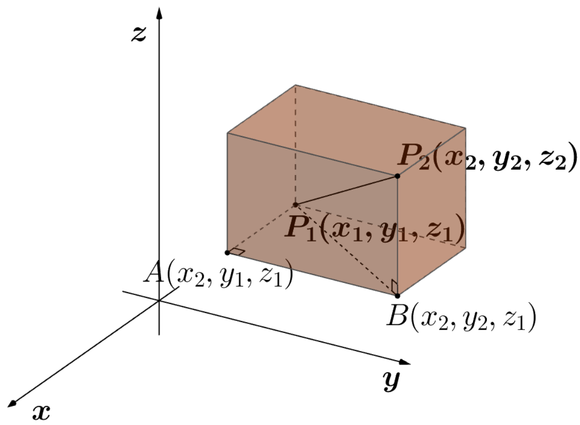


Figure 1.3: Finding the distance between points P_1 and P_2 by applying the Pythagorean theorem to the right triangles P_1AB and P_1BP_2.

If $A(x_2, y_1, z_1)$ and $B(x_2, y_2, z_1)$ are the vertices of the box indicated in Figure 1.3, then the three box edges P_1A, AB, and BP_2 have lengths

$$|P_1A| = |x_2 - x_1|, \quad |AB| = |y_2 - y_1|, \quad |BP_2| = |z_2 - z_1|.$$

Because the triangles P_1AB and P_1BP_2 are both right-angled, the application of the Pythagorean theorem gives us

$$|P_1P_2|^2 = |P_1B|^2 + |BP_2|^2 \quad \text{and} \quad |P_1B|^2 = |P_1A|^2 + |AB|^2.$$

So

$$\begin{aligned}|P_1P_2|^2 &= |P_1B|^2 + |BP_2|^2 \\ &= |P_1A|^2 + |AB|^2 + |BP_2|^2 \\ &= |x_2 - x_1|^2 + |y_2 - y_1|^2 + |z_2 - z_1|^2 \\ &= (x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2.\end{aligned}$$

Therefore

$$|P_1P_2| = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2}.$$

□

Example 1.3. *The distance between* $P_1(2, 1, 5)$ *and* $P_2(-2, 3, 0)$ *is*

$$\begin{aligned}|P_1P_2|^2 &= \sqrt{(-2-2)^2 + (3-1)^2 + (0-5)^2} \\ &= \sqrt{16 + 4 + 25} \\ &= \sqrt{45}.\end{aligned}$$

Spheres in space

We can use the distance formula to write equations for spheres in space (see Figure 1.4). A point $P(x, y, z)$ lies on the sphere of radius a centered at $P_0(x_0, y_0, z_0)$ precisely when $|P_0P| = a$ or $(x - x_0)^2 + (y - y_0)^2 + (z - z_0)^2 = a^2$.

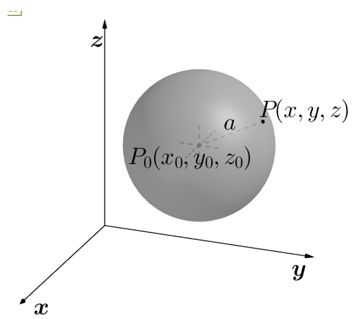


Figure 1.4: A sphere in space.

The standard equation for the sphere of radius a and center (x_0, y_0, z_0) is

$$(x - x_0)^2 + (y - y_0)^2 + (z - z_0)^2 = a^2.$$

Example 1.4. *Find the center and the radius of the sphere*

$$x^2 + y^2 + z^2 + 3x - 4z + 1 = 0.$$

Solution. We find the center and radius of a sphere the way we find the center and radius of a circle: complete the squares on the x-, y-, z-terms as necessary and write each quadratic as a squared linear expression. Then, from the equation in standard form, read of the center and radius. For the sphere here, we have

$$x^2 + y^2 + z^2 + 3x - 4z + 1 = 0$$
$$(x^2 + 3x) + y^2 + (z^2 - 4z) = -1$$
$$\left(x + \frac{3}{2}\right)^2 + y^2 + (z-2)^2 = -1 + \left(\frac{3}{2}\right)^2 + 2^2$$
$$\left(x + \frac{3}{2}\right)^2 + y^2 + (z-2)^2 = \frac{21}{4}$$

From this standard form, we read that $x_0 = -\frac{3}{2}$, $y_0 = 0$, $z_0 = 2$, and $a = \frac{\sqrt{21}}{2}$. The center is $(-\frac{3}{2}, 0, 2)$ and the radius is $\frac{\sqrt{21}}{2}$.

Example 1.5. *Here are some geometric interpretations of inequalities and equations involving spheres.*

(a) $x^2 + y^2 + z^2 < 4$	The interior of the sphere $x^2 + y^2 + z^2 = 4$.
(b) $x^2 + y^2 + z^2 \leq 4$	The solid ball bounded by the sphere $x^2 + y^2 + z^2 = 4$. Alternatively, the sphere $x^2 + y^2 + z^2 = 4$ together with its interior.
(c) $x^2 + y^2 + z^2 > 4$	The exterior of the sphere $x^2 + y^2 + z^2 = 4$.
(d) $x^2 + y^2 + z^2 = 4, z \leq 0$	The lower hemisphere cut from the sphere $x^2 + y^2 + z^2 = 4$ by the xy-plane (the plane $z = 0$).

1.2 Vectors in $\mathbb{R}^3$

A vector is represented by a *directed line segment* (Figure 1.5).

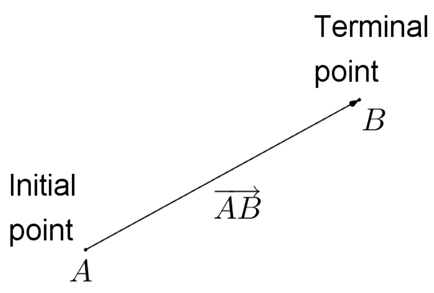


Figure 1.5: The directed line segment $\overrightarrow{AB}$ is called vector.

Definition 1.6. *The vector represented by the directed line segment $\overrightarrow{AB}$ has **initial point** A and **terminal point** B and its **length** is denoted by $|\overrightarrow{AB}|$. Two vectors are **equal** if they have the same length and direction.*

The arrows we use when we draw vectors are understood to represent the same vector if they have the same length, are parallel, and point in the same direction (see Figure 1.6) regardless of the initial point.

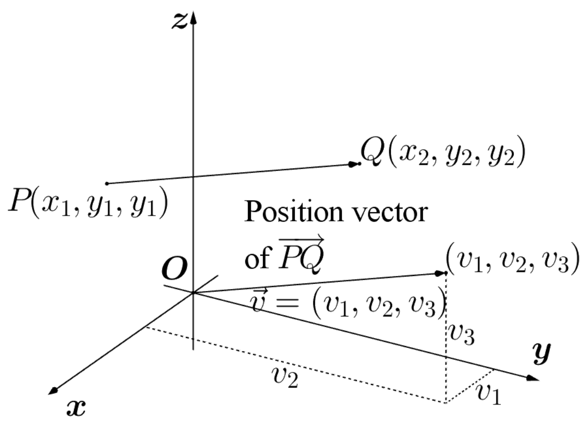


Figure 1.6: A vector $\overrightarrow{PQ}$ in standard position has its initial point at the origin.

1.2.1 Component form

We need a way to represent vectors algebraically so that we can be more precise about the direction of a vector. Let $\overrightarrow{v} = \overrightarrow{PQ}$. There is one directed line segment equal to $\overrightarrow{PQ}$ whose initial point is the origin (Figure 1.6). It is the representative of $\overrightarrow{v}$ in **standard position** and is the vector we normally use to represent $\overrightarrow{v}$. We can specify $\overrightarrow{v}$ by writing the coordinates of its terminal point (v_1, v_2, v_3) when $\overrightarrow{v}$ is in standard position.

Definition 1.7. *If $\overrightarrow{v}$ is a three-dimensional vector equal to the vector with initial point at the origin and terminal point (v_1, v_2, v_3), then the **component form** of $\overrightarrow{v}$ is*

$$\overrightarrow{v} = (v_1, v_2, v_3).$$

So a three-dimensional vector is an **ordered triple** $\overrightarrow{v} = (v_1, v_2, v_3)$ of real numbers. The numbers v_1, v_2 and v_3 are the *components* of $\overrightarrow{v}$.

If $\overrightarrow{v} = (v_1, v_2, v_3)$ is represented by the directed line segment $\overrightarrow{PQ}$, where the initial point is $P(x_1, y_1, z_1)$ and the terminal point is $Q(x_2, y_2, z_2)$, then $x_1 + v_1 = x_2$, $y_1 + v_2 = y_2$, and $z_1 + v_3 = z_2$ (see Figure 1.6). Thus, $v_1 = x_2 - x_1$, $v_2 = y_2 - y_1$, and $v_3 = z_2 - z_1$ are the components of $\overrightarrow{PQ}$.

In summary, given the points $P(x_1, y_1, z_1)$ and $Q(x_2, y_2, z_2)$, the standard position vector $\overrightarrow{v} = (v_1, v_2, v_3)$ equal to $\overrightarrow{PQ}$ is

$$\overrightarrow{v} = (x_2 - x_1, y_2 - y_1, z_2 - z_1).$$

Two vectors are equal if and only if their standard position vectors are identical. Thus (u_1, u_2, u_3) and (v_1, v_2, v_3) are *equal* if and only if $u_1 = v_1$, $u_2 = v_2$, and $u_3 = v_3$. The *magnitude* or *length* of the vector $\overrightarrow{PQ}$ is the length of any of its equivalent directed line segment representations. In particular, if $\overrightarrow{v} = (x_2 - x_1, y_2 - y_1, z_2 - z_1)$ is the standard position vector for $\overrightarrow{PQ}$, then the distance formula gives the magnitude or *length* of $\overrightarrow{v}$, denoted by $|\overrightarrow{v}|$ or $||\overrightarrow{v}||$ (also called *the norm*).

The magnitude or length of the vector $\overrightarrow{v} = \overrightarrow{PQ}$ is the non-negative number

$$|\overrightarrow{v}| = \sqrt{v_1^2 + v_2^2 + v_3^2} = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2}$$

The only vector with length 0 is the *zero vector* $\overrightarrow{0} = (0, 0, 0)$. This vector is also the only vector with no specific direction.

Example 1.8. *Find the (a) component form and (b) length of the vector with initial point* $P(-3, 4, 1)$ *and terminal point* $Q(-5, 2, 2)$.

Solution. (a) The standard position vector $\overrightarrow{v}$ representing $\overrightarrow{PQ}$ has components $v_1 = x_2 - x_1 = -5 - (-3) = -2$, $v_2 = y_2 - y_1 = 2 - 4 = -2$, and $v_3 = z_2 - z_1 = 2 - 1 = 1$. The component form of $\overrightarrow{PQ}$ is $\overrightarrow{v} = (-2, -2, 1)$.

(b) The length of $\overrightarrow{v} = \overrightarrow{PQ}$ is

$$|\overrightarrow{v}| = \sqrt{(-2)^2 + (-2)^2 + 1^2} = \sqrt{9} = 3.$$

1.2.2 Vector algebra operations

Two principal operations involving vectors are *vector addition* and *scalar multiplication.* Scalars can be positive, negative or zero and are used to "scale" a vector by multiplication.

Definition 1.9. *Let* $\overrightarrow{u} = (u_1, u_2, u_3)$ *and* $\overrightarrow{v} = (v_1, v_2, v_3)$ *be vectors and* k *a scalar.*

Vector Addition: $\overrightarrow{u} + \overrightarrow{v} = (u_1 + v_1, u_2 + v_2, u_3 + v_3)$
Scalar multiplication: $k\vec{u} = (ku_1, ku_2, ku_3)$

We add vectors by adding the corresponding components of the vectors. We multiply a vector by a scalar by multiplying each component of the vector by the scalar.

The definition of vector addition is illustrated geometrically (for planar vectors) in Figure 1.7 (left), where the initial point of one vector is placed at the terminal point of the other. Another interpretation is shown in Figure 1.7 (right), called the **parallelogram law** of addition, where the sum, called the *resultant vector*, is the diagonal of the parallelogram.

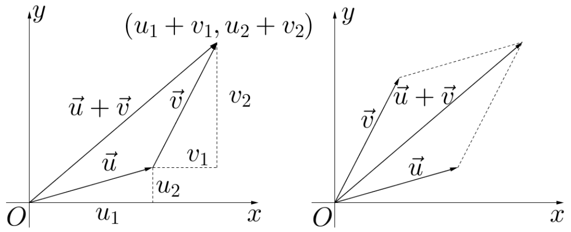


Figure 1.7: Geometric interpretation of the vector sum (left) and the parallelogram law of vector addition (right).

Figure 1.8 displays a geometric interpretation of the product $k\overrightarrow{u}$ of the scalar k and vector $\overrightarrow{u}$. If $k > 0$, then $k\overrightarrow{u}$ has the same direction as $\overrightarrow{u}$; if $k < 0$, then the direction of $k\overrightarrow{u}$ is opposite to that of $\overrightarrow{u}$. Comparing the lengths of $\overrightarrow{u}$ and $k\overrightarrow{u}$, we see that

$$\begin{aligned} |k\overrightarrow{u}| &= \sqrt{(ku_1)^2 + (ku_2)^2 + (ku_3)^2} = \sqrt{k^2(u_1^2 + u_2^2 + u_3^2)} \\ &= \sqrt{k^2}\sqrt{u_1^2 + u_2^2 + u_3^2} = |k||\overrightarrow{u}|. \end{aligned}$$

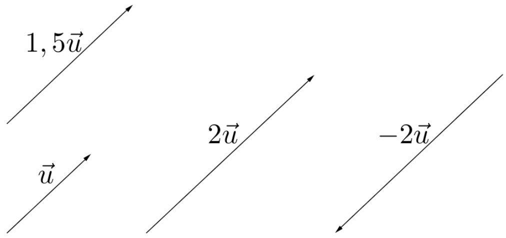


Figure 1.8: Scalar multiples of $\overrightarrow{u}$.

The length of $k\overrightarrow{u}$ is the absolute value of the scalar k times the length of $\overrightarrow{u}$.

The *difference* $\overrightarrow{u} - \overrightarrow{v}$ of two vectors is defined by

$$\overrightarrow{u} - \overrightarrow{v} = \overrightarrow{u} + (-\overrightarrow{v}).$$

Note that $(\overrightarrow{u} - \overrightarrow{v}) + \overrightarrow{v} = \overrightarrow{u}$, see Figure 1.9.

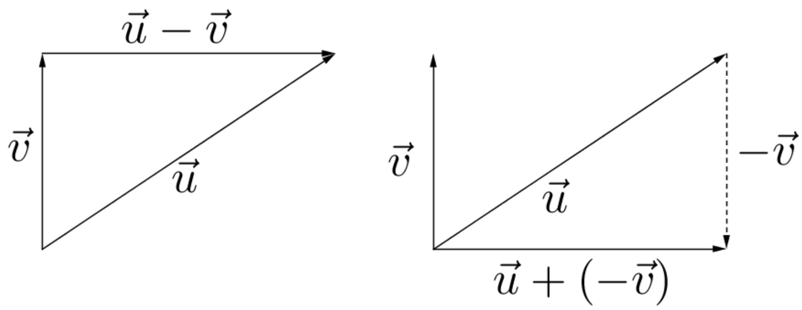


Figure 1.9: The vector $\overrightarrow{u} - \overrightarrow{v}$ when added to $\overrightarrow{v}$ gives $\overrightarrow{u}$ (lesft); and $\overrightarrow{u} - \overrightarrow{v} = \overrightarrow{u} + (-\overrightarrow{v})$ (right)

Example 1.10. *Let* $\overrightarrow{u} = (-1, 3, 1)$ *and* $\overrightarrow{v} = (4, 7, 0)$. *Find the components of*

(a) $2\overrightarrow{u} + 3\overrightarrow{v}$

(b) $\overrightarrow{u} - \overrightarrow{v}$

(c) $|\frac{1}{2}\overrightarrow{u}|$.

Solution. (a) $2\overrightarrow{u} + 3\overrightarrow{v} = 2(-1, 3, 1) + 3(4, 7, 0) = (-2, 6, 2) + (12, 21, 0) = (10, 27, 2)$

(b) $\overrightarrow{u} - \overrightarrow{v} = (-1, 3, 1) - (4, 7, 0) = (-5, -4, 1)$

(c) $|\frac{1}{2}\overrightarrow{u}| = \sqrt{(\frac{-1}{2})^2 + (\frac{3}{2})^2 + (\frac{1}{2})^2} = \frac{\sqrt{11}}{2}$

Vector operations have many of the properties of ordinary arithmetic.

Properties of vector operations
Let $\overrightarrow{u}$, $\overrightarrow{v}$, $\overrightarrow{w}$ be vectors and a, b be scalars.

1. Commutativity: $\overrightarrow{u} + \overrightarrow{v} = \overrightarrow{v} + \overrightarrow{u}$
2. Associativity: $(\overrightarrow{u} + \overrightarrow{v}) + \overrightarrow{w} = \overrightarrow{u} + (\overrightarrow{v} + \overrightarrow{w})$
3. Additive identity : $\overrightarrow{u} + \overrightarrow{0} = \overrightarrow{u}$, where $\overrightarrow{0} = (0, 0, 0)$
4. Additive inverse: $\overrightarrow{u} + (-\overrightarrow{u}) = 0$
5. Distributive properties: $a(\overrightarrow{u} + \overrightarrow{v}) = a\overrightarrow{u} + a\overrightarrow{v}$ and $(a+b)\overrightarrow{u} = a\overrightarrow{u} + b\overrightarrow{u}$
6. Multiplicative identity: $1\overrightarrow{u} = \overrightarrow{u}$
7. $a(b\overrightarrow{u}) = (ab)\overrightarrow{u}$

These properties can be verified using the definitions of vector addition and scalar multiplication, as example we will establish commutativity and the first one of the distributive properties. The rest of the proofs is left as an exercise.

Proof. First we prove commutativity:

$$\begin{aligned}
\overrightarrow{u} + \overrightarrow{v} &= (u_1, u_2, u_3) + (v_1, v_2, v_3) \\
&= (u_1 + v_1, u_2 + v_2, u_3 + v_3) \\
&= (v_1 + u_1, v_2 + u_2, v_3 + u_3) \\
&= (v_1, v_2, v_3) + (u_1, u_2, u_3) \\
&= \overrightarrow{v} + \overrightarrow{u}.
\end{aligned}$$

Now we prove the first of the distributive properties:

$$\begin{aligned}
a(\overrightarrow{u} + \overrightarrow{v}) &= a((u_1, u_2, u_3) + (v_1, v_2, v_3)) \\
&= a(u_1 + v_1, u_2 + v_2, u_3 + v_3) \\
&= (a(u_1 + v_1), a(u_2 + v_2), a(u_3 + v_3)) \\
&= (au_1 + av_1, au_2 + av_2, au_3 + av_3) \\
&= (au_1, au_2, au_3) + (av_1, av_2, av_3) \\
&= a(u_1, u_2, u_3) + a(v_1, v_2, v_3) \\
&= a\overrightarrow{u} + a\overrightarrow{v}.
\end{aligned}$$

□

Linear independence

Definition 1.11. *A set of vectors $\overrightarrow{v}_1, \overrightarrow{v}_2, \ldots, \overrightarrow{v}_n$ is* **linearly independent** *if $a_1\overrightarrow{v}_1 + a_2\overrightarrow{v}_2 + \cdots + a_n\overrightarrow{v}_n = \overrightarrow{0}$ if and only if $a_1 = a_2 = \cdots = a_n = 0$.*

If there exists some linear combination $a_1\overrightarrow{v}_1 + a_2\overrightarrow{v}_2 + \cdots + a_n\overrightarrow{v}_n = \overrightarrow{0}$ in which at least one of the integers a_i differs from 0, then the vectors $\overrightarrow{v}_1, \overrightarrow{v}_2, \ldots, \overrightarrow{v}_n$ are **linearly dependent**. This means that there exists at least one vector $\overrightarrow{v}_i$ that can be expressed as a linear combination of the others.

If, for example, $a_i \neq 0$, then

$$\overrightarrow{v}_i = \frac{1}{a_1}(-a_2\overrightarrow{v}_2 - \cdots - a_n\overrightarrow{v}_n) = -\frac{a_2}{a_1}\overrightarrow{v}_2 - \cdots - \frac{a_n}{a_1}\overrightarrow{v}_n.$$

Remark 1.12. *A set of vectors that contains $\overrightarrow{0}$ is linearly dependent.*

Example 1.13. *Determine whether the given sets of vectors are linearly independent:*

a) $\overrightarrow{u} = (-1, 1, 1)$, $\overrightarrow{v} = (1, 2, 3)$, *and* $\overrightarrow{w} = (0, 1, 8)$

b) $\overrightarrow{u} = (-1, 1, 21)$, $\overrightarrow{v} = (1, 2, 3)$, *and* $\overrightarrow{w} = (0, 1, 8)$

Solution. a) If $a\overrightarrow{v} + b\overrightarrow{u} + c\overrightarrow{w} = 0$, then

$$\begin{aligned} -a + b &= 0 \\ a + 2b + c &= 0 \\ a + 3b + 8c &= 0 \end{aligned}$$

thus

$$\begin{aligned} a &= b \\ c &= -3b \\ 20b &= 0 \Rightarrow b = 0 \Rightarrow c = 0 \Rightarrow a = 0 \end{aligned}$$

Therefore $\overrightarrow{v}$, $\overrightarrow{u}$, and $\overrightarrow{w}$ are linearly independent.

b) If $a\overrightarrow{v} + b\overrightarrow{u} + c\overrightarrow{w} = 0$, then

$$\begin{aligned} -a + b &= 0 \\ a + 2b + c &= 0 \\ 21a + 3b + 8c &= 0 \end{aligned}$$

thus

$$\begin{aligned} a &= b \\ c &= -3b \\ 21a + 3b + 8c = 0 &\Rightarrow 0b = 0 \end{aligned}$$

Note that b can be chosen arbitrarily. If, for example, $b = 1$, then $a = 1$ and $c = -3$. Therefore $\overrightarrow{v}$, $\overrightarrow{u}$, and $\overrightarrow{w}$ are linearly dependent.

1.2.3 Unit vectors

A vector $\overrightarrow{v}$ of length 1 is called **unit vector**. The standard unit vectors in $\mathbb{R}^3$ are

$$\overrightarrow{i} = (1,0,0), \quad \overrightarrow{j} = (0,1,0), \quad \text{and } \overrightarrow{k} = (0,0,1).$$

Any vector $\overrightarrow{v} = (v_1, v_2, v_3)$ can be written as a *linear combination* of the standard unit vectors as follows:

$$\begin{aligned} \overrightarrow{v} &= (v_1, v_2, v_3) = (v_1, 0, 0) + (0, v_2, 0) + (0, 0, v_3) \\ &= v_1(1,0,0) + v_2(0,1,0) + v_3(0,0,1) \\ &= v_1 \overrightarrow{i} + v_2 \overrightarrow{j} + v_3 \overrightarrow{k}. \end{aligned}$$

We call the scalar v_1 the i-component of the vector $\overrightarrow{v}$, v_2 the j-component, and v_3 the k-component. Whenever $\overrightarrow{v} \neq \overrightarrow{0}$, its length $|\overrightarrow{v}|$ is not 0 and

$$\left| \frac{1}{|\overrightarrow{v}|} \overrightarrow{v} \right| = \frac{1}{|\overrightarrow{v}|} |\overrightarrow{v}| = 1.$$

That is, $\frac{\overrightarrow{v}}{|\overrightarrow{v}|}$ is a unit vector in the direction of $\overrightarrow{v}$.

Example 1.14. *Find the unit vector $\overrightarrow{u}$ in the direction of the vector from $P(1,0,1)$ to $Q(3,2,0)$.*

Solution. We divide $\overrightarrow{PQ}$ by its length:
$\overrightarrow{PQ} = (3-1, 2-0, 0-1) = (2,2,-1)$
$|\overrightarrow{PQ}| = \sqrt{2^2 + 2^2 + (-1)^2} = \sqrt{4+4+1} = \sqrt{9} = 3$
$\overrightarrow{u} = \frac{\overrightarrow{PQ}}{|\overrightarrow{PQ}|} = \frac{1}{3}(2,2,-1) = (\frac{2}{3}, \frac{2}{3}, -\frac{1}{3})$

1.3 The dot product

Definition 1.15. *Let $\overrightarrow{u} = (u_1, u_2, u_3)$ and $\overrightarrow{v} = (v_1, v_2, v_3)$ be vectors. The **dot product** of $\overrightarrow{u}$ and $\overrightarrow{v}$ is the number $u_1v_1 + u_2v_2 + u_3v_3$. The dot product is denoted as $\langle u, v \rangle$.*

Remark 1.16. *In many textbooks the dot product of vectors $\overrightarrow{u}$ and $\overrightarrow{v}$ is denoted by $\overrightarrow{u} \cdot \overrightarrow{v}$, the $\cdot$ in the notation meant as a direct association to the name. The dot product is sometimes also called the* scalar product*, since it results in a scalar.*

Algebraic properties of the dot product
Let $\overrightarrow{u}$, $\overrightarrow{v}$, $\overrightarrow{w}$ be vectors and a be a scalar.

1. Additivity: $\langle u+v, w \rangle = \langle u, w \rangle + \langle v, w \rangle$
2. Homogeneity: $\langle au, v \rangle = a\langle u, v \rangle$
3. Symmetricity: $\langle u, v \rangle = \langle v, u \rangle$

Proof. Let $\overrightarrow{u} = (u_1, u_2, u_3)$, $\overrightarrow{v} = (v_1, v_2, v_3)$, and $\overrightarrow{w} = (w_1, w_2, w_3)$ be vectors and a a scalar.

Then:

1.
$$\begin{aligned}\langle u+v, w\rangle &= (u_1+v_1)w_1 + (u_2+v_2)w_2 + (u_3+v_3)w_3 \\ &= u_1w_1 + v_1w_1 + u_2w_2 + v_2w_2 + u_3w_3 + v_3w_3 \\ &= u_1w_1 + u_2w_2 + u_3w_3 + v_1w_1 + v_2w_2 + v_3w_3 \\ &= \langle u, w\rangle + \langle v, w\rangle.\end{aligned}$$

2.
$$\begin{aligned}\langle au, v\rangle &= (au_1)v_1 + (au_2)v_2 + (au_3)v_3 \\ &= a(u_1v_1) + a(u_2v_2) + a(u_3v_3) \\ &= a(u_1v_1 + u_2v_2 + u_3v_3)) \\ &= a\langle u, v\rangle.\end{aligned}$$

3.
$$\begin{aligned}\langle u, v\rangle &= u_1v_1 + u_2v_2 + u_3v_3 \\ &= v_1u_1 + v_2u_2 + v_3u_3 \\ &= \langle v, u\rangle.\end{aligned}$$

□

Corollary 1.17. *Let $\overrightarrow{u}$, $\overrightarrow{v}$, $\overrightarrow{w}$ be vectors and a be a scalar. Then the following holds*

4. $\langle u, v+w\rangle = \langle u, v\rangle + \langle u, w\rangle$

5. $\langle u, av\rangle = a\langle u, v\rangle$

Proof. Both properties are consequences of the first three given properties of the dot product. By consecutively using properties 3.), 1.) and 3.) we establish

4.), since $\langle u, v+w\rangle = \langle v+w, u\rangle = \langle v, u\rangle + \langle w, u\rangle = \langle u, v\rangle + \langle u, w\rangle$.

By consecutively using properties 3.), 2.) and 3.) we establish

5.), since $\langle u, av\rangle = \langle av, u\rangle = a\langle v, u\rangle = a\langle u, v\rangle$. □

1.3.1 Geometrical meaning of the dot product

Recall the *law of cosines* for a triangle, stating $c^2 = a^2 + b^2 - 2ab\cos\varphi$, where a, b, c are lengths of the triangle sides and φ is the angle between sides a and b. Now use the law of cosines with a triangle having sides $|\overrightarrow{u}|, |\overrightarrow{v}|$ and $|\overrightarrow{v} - \overrightarrow{u}|$:

$$|\overrightarrow{v} - \overrightarrow{u}|^2 = |\overrightarrow{u}|^2 + |\overrightarrow{v}|^2 - 2|\overrightarrow{u}||\overrightarrow{v}|\cos\varphi$$

or, in other words

$$\langle v-u, v-u\rangle = \langle u, u\rangle + \langle v, v\rangle - 2|\overrightarrow{u}||\overrightarrow{v}|\cos\varphi, \tag{1.1}$$

where φ represents the angle between vectors $\overrightarrow{u}$ and $\overrightarrow{v}$. Using properties of the dot product now we can deduce that

$$\begin{aligned}\langle v-u, v-u\rangle &= \langle v, v-u\rangle + \langle -u, v-u\rangle \\ &= \langle v, v\rangle + \langle v, -u\rangle + \langle -u, v\rangle + \langle -u, -u\rangle \\ &= \langle v, v\rangle - \langle v, u\rangle - \langle u, v\rangle + \langle u, u\rangle \\ &= \langle v, v\rangle - 2\langle u, v\rangle + \langle u, u\rangle.\end{aligned}$$

From equation 1.1 it follows that

$$-2\langle u, v\rangle = -2|\overrightarrow{u}||\overrightarrow{v}|\cos\varphi.$$

And from here we can get the geometrical meaning of the dot product (in $\mathbb{R}^2$).

$$\langle u, v\rangle = |\overrightarrow{u}||\overrightarrow{v}|\cos\varphi, \tag{1.2}$$

The geometrical meaning of the dot product in $\mathbb{R}^3$ (and in all higher dimensions) is no different. We can use a triangle defined by the vectors $\overrightarrow{u}$ and $\overrightarrow{v}$, aligned so, that their initial points coincide. Then the initial point together with both ending points, defines a plane in $\mathbb{R}^3$. Now we can use the law of cosine in this plane and do all the computations as we did in the $\mathbb{R}^2$ case, showing that

Theorem 1.18. *Let $\overrightarrow{u}$ and $\overrightarrow{v}$ be vectors. Then $\langle u, v\rangle = |\overrightarrow{u}||\overrightarrow{v}|\cos\varphi$, where φ is the angle between vectors $\overrightarrow{u}$ and $\overrightarrow{v}$.*

Remark 1.19. *The angle φ is always chosen from the interval $[0, \pi]$.*

Example 1.20. *Find the angle between vectors $\overrightarrow{u} = (2, -2, -1)$ and $\overrightarrow{v} = (-3, 0, 4)$.*

Solution. Compute:

$$|\overrightarrow{u}|^2 = \langle u, u\rangle = 4 + 4 + 1 = 9$$
$$|\overrightarrow{v}|^2 = \langle v, v\rangle = 9 + 0 + 16 = 25$$
$$\langle u, v\rangle = -6 + 0 - 4 = -10$$

Then $\cos\varphi = \frac{\langle u,v\rangle}{|\overrightarrow{u}||\overrightarrow{v}|} = \frac{-10}{3\cdot 5} = -\frac{2}{3}$. From here we get $\varphi \doteq 131,8°$.

Note that for $\overrightarrow{u} = (u_1, u_2, u_3)$, the dot product $\langle u, u\rangle = u_1^2 + u_2^2 + u_3^2$ is equal to 0 only if $\overrightarrow{u} = \overrightarrow{0}$. As we already mentioned, the zero vector is the only vector with length 0. Taking a closer look to the dot product formula $\langle u, v\rangle = |\overrightarrow{u}||\overrightarrow{v}|\cos\varphi$, one can notice that $\cos\varphi = 0$ if and only if $\varphi = \frac{\pi}{2}$. If $\overrightarrow{u}$ and $\overrightarrow{v}$ are both non-zero vectors, then $\langle u, v\rangle = 0$ if and only if $\cos\varphi = 0$, therefore showing the next Claim.

Claim 1.21. *Let $\overrightarrow{u}$ and $\overrightarrow{v}$ be vectors. Then $\overrightarrow{u}$ and $\overrightarrow{v}$ are **orthogonal** (or **perpendicular**) if and only if $\langle u, v\rangle = 0$.*

Remark 1.22. *We assume that the zero vector $\overrightarrow{0}$ is orthogonal to all other vectors.*

Example 1.23. *Check whether vectors $\overrightarrow{u} = (3, 2, -1)$ and $\overrightarrow{v} = (-1, 1, -1)$ are orthogonal. What about vectors $\overrightarrow{u_1} = (2, 1, 0)$ and $\overrightarrow{v_1} = (1, -1, 0)$?*

Solution. Since $\langle u, v\rangle = 3\cdot(-1) + 2\cdot 1 + (-1)\cdot(-1) = -3 + 2 + 1 = 0$ vectors $\overrightarrow{u}$ and $\overrightarrow{v}$ are orthogonal.

Since $\langle u_1, v_1\rangle = 2\cdot 1 + 1\cdot(-1) = 2 - 1 = 1 \neq 0$ vectors $\overrightarrow{u_1}$ and $\overrightarrow{v_1}$ are not orthogonal.

Recall that for $\varphi \in [0, \pi]$ the function $\cos\varphi$ takes value 1 precisely when $\varphi = 0$ and value -1 precisely when $\varphi = \pi$, showing the next Theorem.

Theorem 1.24. *Let $\overrightarrow{u}$ and $\overrightarrow{v}$ be non-zero vectors and φ the angle between them. Then $\cos\varphi = 1$ if and only if $\overrightarrow{u}$ and $\overrightarrow{v}$ point in the same direction, and $\cos\varphi = -1$ if and only if $\overrightarrow{u}$ and $\overrightarrow{v}$ point in the opposite direction.*

Recall also, that the range of the function $\cos\varphi$ is the interval $[-1, 1]$. Then $0 \leq |\cos\varphi| \leq 1$. From Theorem 1.18 we get the next Theorem:

Theorem 1.25 (Cauchy-Schwarz inequality)**.** *For any vectors $\overrightarrow{u}, \overrightarrow{v} \in \mathbb{R}^3$ it holds that*

$$|\langle u, v\rangle| \leq |\overrightarrow{u}| \cdot |\overrightarrow{v}|.$$

Proof. The proof is left as an exercise. □

The following results can be derived from properties of the dot product.

Theorem 1.26 (Pythagorean theorem)**.** *Let $\overrightarrow{u}, \overrightarrow{v} \in \mathbb{R}^3$ be orthogonal vectors. Then*

$$|\overrightarrow{u} + \overrightarrow{v}|^2 = |\overrightarrow{u}|^2 + |\overrightarrow{v}|^2.$$

Proof. Compute:

$$\begin{aligned} |\overrightarrow{u} + \overrightarrow{v}|^2 &= \langle u + v, u + v\rangle = \langle u, u + v\rangle + \langle v, u + v\rangle \\ &= \langle u, u\rangle + \langle u, v\rangle + \langle v, u\rangle + \langle v, v\rangle \\ &= \langle u, u\rangle + 2\langle u, v\rangle + \langle v, v\rangle. \end{aligned}$$

Since $\overrightarrow{u}, \overrightarrow{v}$ are orthogonal, $\langle u, v\rangle = 0$ and therefore

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

□

Theorem 1.27 (Triangle inequality)**.** *Let $\overrightarrow{u}, \overrightarrow{v} \in \mathbb{R}^3$. Then*

$$|\overrightarrow{u} + \overrightarrow{v}| \leq |\overrightarrow{u}| + |\overrightarrow{v}|.$$

Proof. From the proof of the Pythagorean theorem and using the Cauchy-Schwarz inequality we get

$$\begin{aligned} |\overrightarrow{u} + \overrightarrow{v}|^2 &= \langle u, u\rangle + 2\langle u, v\rangle + \langle v, v\rangle \\ &= |\overrightarrow{u}|^2 + 2|\overrightarrow{u}||\overrightarrow{v}|\cos\varphi + |\overrightarrow{v}|^2 \\ &\leq |\overrightarrow{u}|^2 + 2|\overrightarrow{u}||\overrightarrow{v}| + |\overrightarrow{v}|^2 = (|\overrightarrow{u}| + |\overrightarrow{v}|)^2. \end{aligned}$$

The inequality follows from the fact that $\cos\varphi \leq 1$. Since the length of a vector is non-negative, it follows that

$$|\overrightarrow{u} + \overrightarrow{v}| \leq |\overrightarrow{u}| + |\overrightarrow{v}|.$$

□

Vector projections

Let vectors $\overrightarrow{u}$ and $\overrightarrow{v}$ be such that their initial points coincide. The **vector projection** of $\overrightarrow{u} = \overrightarrow{PQ}$ onto a nonzero vector $\overrightarrow{v} = \overrightarrow{PS}$ is the vector $\overrightarrow{w} = \overrightarrow{PR}$, determined by dropping a perpendicular from Q to the line PS. The notation for this vector is

$$\mathrm{proj}_{\overrightarrow{v}}\overrightarrow{u} \quad \text{"the vector projection of } \overrightarrow{u} \text{ onto } \overrightarrow{v}\text{“}$$

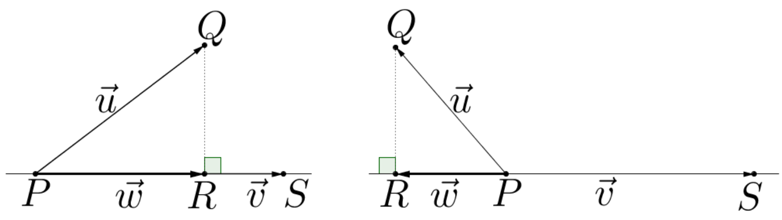


Figure 1.10: Vector projection.

If the angle φ between $\overrightarrow{u}$ and $\overrightarrow{v}$ is acute, $\mathrm{proj}_{\overrightarrow{v}}\overrightarrow{u}$ has length $|\overrightarrow{u}|\cos\varphi$ and direction $\frac{\overrightarrow{v}}{|\overrightarrow{v}|}$ (Figure 1.10). If the angle φ is obtuse, $\cos\varphi < 0$ and $\mathrm{proj}_{\overrightarrow{v}}\overrightarrow{u}$ has length $-|\overrightarrow{u}|\cos\varphi$ and direction $\frac{-\overrightarrow{v}}{|\overrightarrow{v}|}$. In both cases,

$$\overrightarrow{w} = \mathrm{proj}_{\overrightarrow{v}}\overrightarrow{u} = |\overrightarrow{u}|\cos\varphi\frac{\overrightarrow{v}}{|\overrightarrow{v}|} = \frac{|\overrightarrow{u}||\overrightarrow{v}|\cos\varphi}{|\overrightarrow{v}|^2}\overrightarrow{v} = \frac{\langle u,v\rangle}{\langle v,v\rangle}\overrightarrow{v}.$$

With this last equation we have shown the first part of the next Claim.

Claim 1.28. *Let $\overrightarrow{u}$ and $\overrightarrow{v}$ be two non-zero vectors. Then $\overrightarrow{w} = \frac{\langle u,v\rangle}{\langle v,v\rangle}\overrightarrow{v}$ is the vector projection of $\overrightarrow{u}$ onto $\overrightarrow{v}$. Vectors $\overrightarrow{u} - \overrightarrow{w}$ and $\overrightarrow{v}$ are orthogonal.*

Proof. We only have to show the second part of the Claim:

$$\begin{aligned}\langle u-w, v\rangle = \left\langle u - \frac{\langle u,v\rangle}{\langle v,v\rangle}\overrightarrow{v}, v\right\rangle &= \langle u,v\rangle + \left\langle -\frac{\langle u,v\rangle}{\langle v,v\rangle}\overrightarrow{v}, v\right\rangle \\ &= \langle u,v\rangle - \frac{\langle u,v\rangle}{\langle v,v\rangle}\langle v,v\rangle \\ &= \langle u,v\rangle - \langle u,v\rangle = 0.\end{aligned}$$

□

Example 1.29. *Find the vector projection of $\overrightarrow{u} = (3,-1,2)$ onto $\overrightarrow{v} = (-1,0,1)$.*

Solution. First compute

$$\langle u,v\rangle = 3\cdot(-1) + (-1)\cdot 0 + 2\cdot 1 = -3+0+2 = -1$$

$$\langle v,v\rangle = -1\cdot(-1) + 0\cdot 0 + 1\cdot 1 = 1+0+1 = 2$$

then,

$$\mathrm{proj}_{\overrightarrow{v}}\overrightarrow{u} = \frac{-1}{2}(-1,0,1) = \left(\frac{1}{2}, 0, -\frac{1}{2}\right)$$

1.4 The cross product

In studying lines in the plane, when we need to describe how a line was *tilting*, we used the notions of slope and angle of inclination. In space, we want a way to describe how a *plane* is tilting. We accomplish this by multiplying two vectors in the plane together to get a third vector perpendicular to the plane. The direction of this third vector tells us the "inclination" of the plane. The product we use to multiply the vectors together is the *cross product*.

We start with two non-zero vectors $\overrightarrow{u}$ and $\overrightarrow{v}$ in space. If $\overrightarrow{u}$ and $\overrightarrow{v}$ are not parallel, they determine a plane. We select a unit vector $\overrightarrow{n}$ orthogonal to the plane by the *right hand rule*. Then the **cross product** $u \times v$ is the *vector* defined as follows.

Definition 1.30. *Let $\overrightarrow{u}$ and $\overrightarrow{v}$ be two non-zero vectors in $\mathbb{R}^3$. Then the cross product $u \times v$ is the vector*

$$u \times v = (|\overrightarrow{u}||\overrightarrow{v}|\sin\varphi)\overrightarrow{n}.$$

Unlike the dot product, the cross product is a vector. For this reason it's also called the *vector product* of $\overrightarrow{u}$ and $\overrightarrow{v}$ and applies *only* to vectors in space. The vector $u \times v$ is orthogonal to both $\overrightarrow{u}$ and $\overrightarrow{v}$ because is a scalar multiple of $\overrightarrow{n}$. There is a straightforward way to compute the cross product of two vectors from their components. The method does not require that we know the angle between them (as suggested by the definition), but we postpone the calculation momentarily so we can focus first on the properties of the cross product.

Algebraic properties of the cross product
Let $\overrightarrow{u}$, $\overrightarrow{v}$, $\overrightarrow{w}$ be vectors in $\mathbb{R}^3$ and a be a scalar.

1. Additivity: $(u + v) \times w = (u \times w) + (v \times w)$ and $u \times (v + w) = (u \times v) + (u \times w)$
2. Homogeneity: $(au) \times v = a(u \times v) = u \times (av)$
3. Anticommutativity: $u \times v = -v \times u$

Proof. Proofs for additivity and anticommutativity will be given later. □

From the above properties we can deduce the following

4. $u \times u = \overrightarrow{0}$
5. $u \times 0 = \overrightarrow{0}$

Where property 4. follows from anticommutativity. Since $u \times u = -u \times u$ and $2(u \times u) = \overrightarrow{0}$. By multiplying both sides by $\frac{1}{2}$ we get $u \times u = \overrightarrow{0}$. The equality given in 5. can be verified using the definition of the cross product.

6. Linearity: $(au + bv) \times w = a(u \times w) + b(v \times w)$

Where $b \in \mathbb{R}$.

To prove linearity we use additivity and homogeneity. We get the following chain of equalities

$$(au + bv) \times w = (au \times w) + (bv \times w) = a(u \times w) + b(v \times w).$$

At last we give the formula for the double cross product:

7. $(u \times v) \times w = \langle u, w\rangle v - \langle v, w\rangle u$

The proof of this last equality is left as an exercise.

The following statement involving double cross products is known as **Jacobi identity**

For any vectors $\overrightarrow{u}$, $\overrightarrow{v}$, and $\overrightarrow{w} \in \mathbb{R}^3$ the following holds:

$$(u \times v) \times w + (v \times w) \times u + (w \times u) \times v = 0$$

Proof. $(u \times v) \times w + (v \times w) \times u + (w \times u) \times v = = \langle u, w\rangle v - \langle v, w\rangle u + \langle v, u\rangle w - \langle w, u\rangle v + \langle w, v\rangle u - \langle u, v\rangle w = 0$ □

Calculating the cross product as a determinant

As already mentioned, we can calculate the cross product of two vectors directly from their components and this method does not require us to know the angle between vectors.

Theorem 1.31. *Let* $\overrightarrow{u} = (u_1, u_2, u_3)$ *and* $\overrightarrow{v} = (v_1, v_2, v_3)$. *Then*

$$u \times v = (u_2v_3 - u_3v_2, u_3v_1 - u_1v_3, u_1v_2 - u_2v_1).$$

Proof. From the properties given in the previous section we first get

$$\begin{aligned} u \times v &= (u_1\overrightarrow{i} + u_2\overrightarrow{j} + u_3\overrightarrow{k}) \times (v_1\overrightarrow{i} + v_2\overrightarrow{j} + v_3\overrightarrow{k}) \\ &= u_1v_1 i \times i + u_1v_2 i \times j + u_1v_3 i \times k + \\ &+ u_2v_1 j \times i + u_2v_2 j \times j + u_2v_3 j \times k + \\ &+ u_3v_1 k \times i + u_3v_2 k \times j + u_3v_3 k \times k \end{aligned}$$

Taking into account that

$$\begin{array}{lll} i \times i = \overrightarrow{0}, & j \times j = \overrightarrow{0} & k \times k = \overrightarrow{0} \\ i \times j = \overrightarrow{k}, & j \times k = \overrightarrow{i}, & k \times i = \overrightarrow{j}, \\ j \times i = -\overrightarrow{k}, & i \times k = -\overrightarrow{j}, & k \times j = -\overrightarrow{i}, \end{array}$$

we get

$$\begin{aligned} u \times v &= (u_2v_3 - u_3v_2)\overrightarrow{i} + (u_3v_1 - u_1v_3)\overrightarrow{j} + (u_1v_2 - u_2v_1)\overrightarrow{k} \\ &= (u_2v_3 - u_3v_2, u_3v_1 - u_1v_3, u_1v_2 - u_2v_1). \end{aligned}$$

□

The component terms given in Theorem 1.31 are hard to remember. To this end we can gain help from matrices. A (real) **matrix** is a table of (real) numbers. Associated with every square matrix (that is, a matrix with the same number of rows and columns) is a scalar called the **determinant**.

Computing the determinant of a 2×2 matrix - the rule:

$$\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = a_{11}a_{22} - a_{12}a_{21}.$$

We do have a rule to compute also the determinant of a 3×3 matrix:

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

Now the component terms given in Theorem 1.31 can be easily recalled if written as:

$$u \times v = \begin{vmatrix} \overrightarrow{i} & \overrightarrow{j} & \overrightarrow{k} \\ u_1 & u_2 & u_3 \\ v_1 & v_2 & v_3 \end{vmatrix}$$

Since

$$\begin{vmatrix} \overrightarrow{i} & \overrightarrow{j} & \overrightarrow{k} \\ u_i & u_2 & u_3 \\ v_1 & v_2 & v_3 \end{vmatrix} = \begin{vmatrix} u_2 & u_3 \\ v_2 & v_3 \end{vmatrix}\overrightarrow{i} - \begin{vmatrix} u_1 & u_3 \\ v_1 & v_3 \end{vmatrix}\overrightarrow{j} + \begin{vmatrix} u_1 & u_2 \\ v_1 & v_2 \end{vmatrix}\overrightarrow{k}.$$

Example 1.32. *Find $u \times v$ if $\overrightarrow{u} = (2, 1, 1)$ and $\overrightarrow{v} = (-4, 3, 1)$.*

Solution.

$$\begin{aligned} u \times v &= \begin{vmatrix} \overrightarrow{i} & \overrightarrow{j} & \overrightarrow{k} \\ 2 & 1 & 1 \\ -4 & 3 & 1 \end{vmatrix} \\ &= \begin{vmatrix} 1 & 1 \\ 3 & 1 \end{vmatrix}\overrightarrow{i} - \begin{vmatrix} 2 & 1 \\ -4 & 1 \end{vmatrix}\overrightarrow{j} + \begin{vmatrix} 2 & 1 \\ -4 & 3 \end{vmatrix}\overrightarrow{k} \\ &= -2\overrightarrow{i} - 6\overrightarrow{j} + 10\overrightarrow{k} = (-2, -6, 10). \end{aligned}$$

We will use this approach to prove property 3. - anticommutativity.

Proof for anticommutativity. Let $\overrightarrow{u} = (u_1, u_2, u_3)$ and $\overrightarrow{v} = (v_1, v_2, v_3)$. Then

$$u \times v = \begin{vmatrix} \overrightarrow{i} & \overrightarrow{j} & \overrightarrow{k} \\ u_1 & u_2 & u_3 \\ v_1 & v_2 & v_3 \end{vmatrix} = (u_2v_3 - u_3v_2, u_3v_1 - u_1v_3, u_1v_2 - u_2v_1).$$

$$v \times u = \begin{vmatrix} \overrightarrow{i} & \overrightarrow{j} & \overrightarrow{k} \\ v_1 & v_2 & v_3 \\ u_1 & u_2 & u_3 \end{vmatrix} = (u_3v_2 - u_2v_3, u_1v_3 - u_3v_1, u_2v_1 - u_1v_2).$$

By comparing each coordinate we can see that $u \times v = -v \times u$. □

1.4.1 Geometrical meaning of the cross product

If $\overrightarrow{u}$ and $\overrightarrow{v}$ point either in the same either in the opposite direction, the parallelogram defined by $\overrightarrow{u}$ and $\overrightarrow{v}$ is degenerate and his area equals 0. Suppose that $\overrightarrow{u}$ and $\overrightarrow{v}$ do not point in the same nor in the opposite direction. Then the parallelogram defined by $\overrightarrow{u}$ and $\overrightarrow{v}$ is not degenerate and hence has a positive area. In this case $|u \times v|$ is the area of the parallelogram defined by $\overrightarrow{u}$ and $\overrightarrow{v}$.

Because $\overrightarrow{n}$ is a unit vector, the magnitude of $u \times v$ is

$$\boxed{|u \times v| = |\overrightarrow{u}||\overrightarrow{v}||\sin\varphi||\overrightarrow{n}| = |\overrightarrow{u}||\overrightarrow{v}|\sin\varphi.}$$

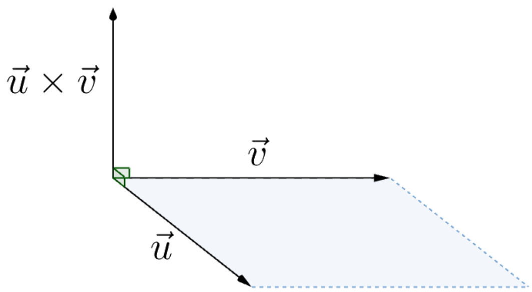
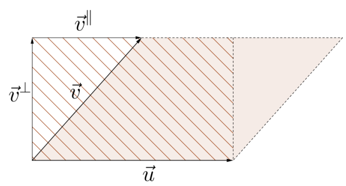


Figure 1.11: The parallelogram determined by $\overrightarrow{u}$ and $\overrightarrow{v}$.

We will use this fact to prove the second equality of property 1. - additivity.

Proof for additivity. To prove $u \times v + w = (u \times v) + (u \times w)$ we will first show that $u \times v = u \times v^{\perp}$, where $\overrightarrow{v} = v^{\perp} + v^{\parallel}$, and $v^{\perp} \perp u$, and $v^{\parallel} \parallel u$.



Note that the area of the parallelogram determined by $\vec{u}$ and $\vec{v}$ is the same as the area of the rectangle determined by $\vec{u}$ and $\vec{v}^{\perp}$ (recall that Area = length of the base $\cdot$ height). Furthermore, in both cases $\vec{n}$ has same direction. Therefore $u \times v = u \times v^{\perp}$.

Now we can use this fact to prove the main statement.

$$\begin{aligned} u \times (v + w) &= u \times (v + w)^{\perp} \\ &= u \times (v^{\perp} + w^{\perp}) \\ &= (u \times v^{\perp}) + (u \times w^{\perp}) \\ &= (u \times v) + (u \times w). \end{aligned}$$

□

Example 1.33. *Find the area of the triangle with vertices $P(1,-1,0)$, $Q(2,1,-1)$, and $R(-1,1,2)$. (see Figure 1.13)*

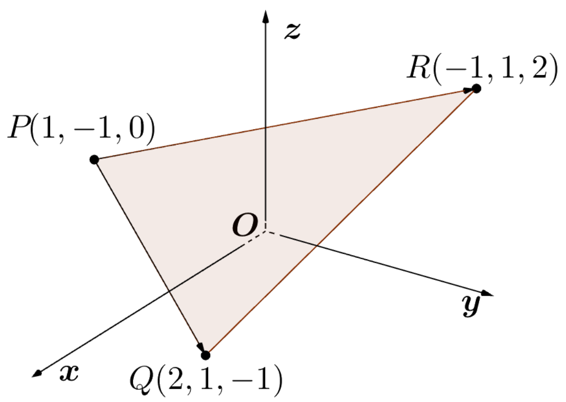


Figure 1.12: The triangle defined by P, Q, and R.

Solution. The area of the triangle determined by P, Q, and R equals to half of the area of the parallelogram determined by vectors $\overrightarrow{PQ}$ and $\overrightarrow{PR}$. Therefore, since

$$\overrightarrow{PQ} = (2,1,-1) - (1,-1,0) = (1,2,-1)$$
$$\overrightarrow{PR} = (-1,1,2) - (1,-1,0) = (-2,2,2)$$

and

$$\begin{aligned} PQ \times PR &= \begin{vmatrix} \vec{i} & \vec{j} & \vec{k} \\ 1 & 2 & -1 \\ -2 & 2 & 2 \end{vmatrix} \\ &= \begin{vmatrix} 2 & -1 \\ 2 & 2 \end{vmatrix} \vec{i} + \begin{vmatrix} 1 & -1 \\ -2 & 2 \end{vmatrix} \vec{j} + \begin{vmatrix} 1 & 2 \\ -2 & 2 \end{vmatrix} \vec{k} \\ &= 6\vec{i} + 6\vec{k} = (6,0,6). \end{aligned}$$

the area of the parallelogram determined by $\overrightarrow{PQ}$ and $\overrightarrow{PR}$ is

$$|PQ \times PR| = \sqrt{6^2 + 0^2 + 6^2} = \sqrt{2 \cdot 36} = 6\sqrt{2}.$$

And hence the area of the triangle determined by P, Q, and R is $3\sqrt{2}$ square units.

1.5 The box product

Definition 1.34. *Let $\overrightarrow{u}$, $\overrightarrow{v}$ and $\overrightarrow{w}$ be non-zero vectors in $\mathbb{R}^3$. Then the box product $\langle u \times v, w\rangle$ is the number*

$$\langle u \times v, w\rangle = |u \times v||\overrightarrow{w}|\cos\varphi,$$

where φ is the angle between $\overrightarrow{u} \times \overrightarrow{v}$ and $\overrightarrow{w}$.

The box product is also called *mixed product*, since it involves the dot and the cross product or *triple scalar product* since it involves 3 three-dimensional vectors and results in a scalar.

The box product can also be calculated with determinants.

Let $\overrightarrow{u} = (u_1, u_2, u_3)$, $\overrightarrow{v} = (v_1, v_2, v_3)$ and $\overrightarrow{w} = (w_1, w_2, w_3)$. Then

$$\langle u \times v, w\rangle = \begin{vmatrix} u_1 & u_2 & u_3 \\ v_1 & v_2 & v_3 \\ w_1 & w_2 & w_3 \end{vmatrix}.$$

Algebraical properties of the box product:
Let $\overrightarrow{u}$, $\overrightarrow{v}$, $\overrightarrow{w}$ and $\overrightarrow{z}$ be vectors in $\mathbb{R}^3$ and let $a \in \mathbb{R}$ be a scalar.

1. $\langle u \times v, w\rangle = \langle u, v \times w\rangle$
2. $\langle u \times v, w\rangle = -\langle v \times u, w\rangle = -\langle w \times v, u\rangle = -\langle u \times w, v\rangle$
3. $\langle au \times v, w\rangle = a\langle u \times v, w\rangle$
4. $\langle u \times u, w\rangle = 0$
5. $\langle u \times v, w + z\rangle = \langle u \times v, w\rangle + \langle u \times v, z\rangle$
6. $\langle u \times v, w \times z\rangle = \langle u, w\rangle\langle v, z\rangle - \langle u, z\rangle\langle v, w\rangle$ (**Lagrange identity**)

Proof. We will prove only the Lagrange identity, proofs of the other properties are left as an exercise.

$$\begin{aligned}\langle u \times v, w \times z\rangle &= \langle (u \times v) \times w, z\rangle = \langle \langle u, w\rangle v - \langle v, w\rangle u, z\rangle = \\ &= \langle u, w\rangle\langle v, z\rangle - \langle v, w\rangle\langle u, z\rangle.\end{aligned}$$

□

Claim 1.35. $\langle u \times v, w\rangle = 0$ *if and only if vectors $\overrightarrow{u}$, $\overrightarrow{v}$, and $\overrightarrow{w}$ are linearly dependent.*

Proof. $\langle u \times v, w\rangle = 0 \Leftrightarrow u \times v$ is orthogonal to $\overrightarrow{w} \Leftrightarrow \overrightarrow{w} = a\overrightarrow{u} + b\overrightarrow{v}$ for some $a, b \in \mathbb{R}$. □

1.5.1 Geometrical meaning of the box product

Claim 1.36. *The absolute value of the box product $\langle u \times v, w\rangle$ is the volume of the parallelepiped determined by $\overrightarrow{u}$, $\overrightarrow{v}$, and $\overrightarrow{w}$.*

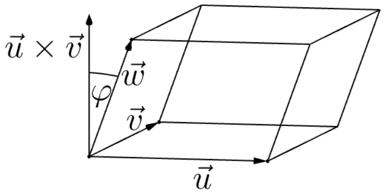


Figure 1.13: The parallelepiped determined by $\overrightarrow{u}$, $\overrightarrow{v}$, and $\overrightarrow{w}$.

Proof. Let φ be the angle between vectors $u \times v$ and $\overrightarrow{w}$. By Definition 1.34 $\langle u \times v, w\rangle = |u \times v||\overrightarrow{w}|\cos\varphi$. Note that $|u \times v|$ is the area of the base parallelogram and that $|\overrightarrow{w}||\cos\varphi|$ is the length of the vector projection $\text{proj}_{\overrightarrow{u \times v}}\overrightarrow{w}$ (this is discussed directly before Claim 1.28) and hence the height of the parallelepiped. Since

$$\begin{aligned}\text{Volume} &= \text{area of base} \cdot \text{height}\\ &= |u \times v| \cdot |\overrightarrow{w}||\cos\varphi|\\ &= \langle u \times v, w\rangle.\end{aligned}$$

□

Example 1.37. *Find the volume of the box (parallelepiped) determined by $\overrightarrow{u} = (1, 2, -1)$, $\overrightarrow{v} = (-2, 0, 3)$, and $\overrightarrow{w} = (0, 7, -4)$.*

Solution.

$$\begin{aligned}\langle u \times v, w\rangle &= \begin{vmatrix} 1 & 2 & -1 \\ -2 & 0 & 3 \\ 0 & 7 & -4 \end{vmatrix}\\ &= 1 \cdot \begin{vmatrix} 0 & 3 \\ 7 & -4 \end{vmatrix} - 2 \cdot \begin{vmatrix} -2 & 3 \\ 0 & -4 \end{vmatrix} - 1 \cdot \begin{vmatrix} -2 & 0 \\ 0 & 7 \end{vmatrix}\\ &= 1 \cdot (-21) - 2 \cdot 8 - 1 \cdot (-14) = -21 - 16 + 14 = -23.\end{aligned}$$

Therefore the volume is $|\langle u \times v, w\rangle| = |-23| = 23$ cubic units.

Example 1.38. *Find such $x \in \mathbb{R}$, that the vectors $\overrightarrow{a} = (1, 2, 3)$, $\overrightarrow{b} = (-2, x, 1)$, and $\overrightarrow{c} = (-1, 1, 0)$ will determine a parallelepiped (box) whose volume is 1.*

Solution. First we compute the box product:

$$\langle a \times b, c\rangle = \begin{vmatrix} 1 & 2 & 3 \\ -2 & x & 1 \\ -1 & 1 & 0 \end{vmatrix} = 1 \cdot (-1) - 2 \cdot 1 + 3 \cdot (-2 + x) = 3x - 9.$$

Since $|3x - 9| = 1$ we have two possibilities:

$$\begin{array}{lcl} 3x - 9 = 1 & \text{or} & 3x - 9 = -1 \\ 3x = 10 & & 3x = 8 \\ x = \frac{10}{3} & & x = \frac{8}{3} \end{array}$$

2
LINES AND PLANES IN $\mathbb{R}^3$

Now we will see how to use the previously mentioned products to write equations for lines and planes in space.

Lines in $\mathbb{R}^3$

In the plane, a line is determined by a point and a number giving the slope of the line. In space a line is determined by a point and a *vector* giving the direction of the line (also called the *direction vector*).

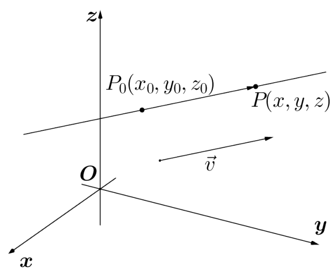


Figure 2.1: A line through P_0 parallel to $\overrightarrow{v}$.

Suppose that ℓ is a line in space passing through a point $P_0(x_0, y_0, z_0)$ parallel to a vector $\overrightarrow{v} = v_1\overrightarrow{i} + v_2\overrightarrow{j} + v_3\overrightarrow{k}$. Then ℓ is the set of all points $P(x, y, z)$ for which $\overrightarrow{P_0P}$ is paralell to $\overrightarrow{v}$ (see Fig. 2.1). Thus, $\overrightarrow{P_0P} = \lambda\overrightarrow{v}$ for some scalar parameter λ. The value of λ depends on the location of the point P along the line, and the domain of λ is $(-\infty, \infty)$. The expanded form of the equation $\overrightarrow{P_0P} = \lambda\overrightarrow{v}$ is

$$(x - x_0)\overrightarrow{i} + (y - y_0)\overrightarrow{j} + (z - z_0)\overrightarrow{k} = \lambda(v_1\overrightarrow{i} + v_2\overrightarrow{j} + v_3\overrightarrow{k}),$$

which can be rewritten as

$$x\overrightarrow{i} + y\overrightarrow{j} + z\overrightarrow{k} = x_0\overrightarrow{i} + y_0\overrightarrow{j} + z_0\overrightarrow{k} + \lambda(v_1\overrightarrow{i} + v_2\overrightarrow{j} + v_3\overrightarrow{k}).$$

If $\overrightarrow{r_1}$ is the position vector of a point $P(x, y, z)$ on the line and $\overrightarrow{r_0}$ is the position vector of the point $P_0(x_0, y_0, z_0)$, then the above equation gives the **vector form** of the equation for a line in space.

Vector form
A vector equation for the line ℓ through $P_0(x_0, y_0, z_0)$ parallel to $\overrightarrow{v}$ is

$$\ell = \overrightarrow{r_1} = \overrightarrow{r_0} + \lambda \overrightarrow{v}, \qquad -\infty < \lambda < \infty,$$

where $\overrightarrow{r_1}$ is the position vector of a point $P(x, y, z)$ on ℓ and $\overrightarrow{r_0}$ is the position vector of $P_0(x_0, y_0, z_0)$.

Equating the corresponding components of the two sides of the vector equation for a line, gives three scalar equations involving the parameter λ:

$$x = x_0 + \lambda v_1 \quad y = y_0 + \lambda v_2 \quad z = z_0 + \lambda v_3.$$

These equations give us the standard parametrization of the line for the parameter interval $-\infty < \lambda < \infty$.

Parametric form
The standard parametrisation of the line through $P_0(x_0, y_0, z_0)$ parallel to $\overrightarrow{v} = (v_1, v_2, v_3)$ is

$$x = x_0 + \lambda v_1 \quad y = y_0 + \lambda v_2 \quad z = z_0 + \lambda v_3, \quad -\infty < \lambda < \infty.$$

Finally, if we solve for λ in each of the three equations we get another way to describe a line.

Canonical form
The canonical equation of the line through $P_0(x_0, y_0, z_0)$ parallel to $\overrightarrow{v} = (v_1, v_2, v_3)$ is

$$\frac{x - x_0}{v_1} = \frac{y - y_0}{v_2} = \frac{z - z_0}{v_3} = \lambda.$$

Note that the canonical equation can be written only if $\overrightarrow{v}$ has only non-zero components.

Example 2.1. *Write the equation for the line through points $A(2, 1, 1)$, and $B(3, -1, 4)$ in all three forms.*

Solution. Since a line is defined by a point and a vector, we first have to compute $\overrightarrow{AB}$.
$\overrightarrow{v} = \overrightarrow{AB} = (3, -1, 4) - (2, 1, 1) = (1, -2, 3)$.
Now our line is determined by the point A and the vector $\overrightarrow{v}$.

- Vector form: $\overrightarrow{r_1} = (2, 1, 1) + \lambda(1, -2, 3)$
- Parametric form:

$$x = 2 + \lambda,$$
$$y = 1 - 2\lambda,$$
$$z = 1 + 3\lambda, \quad \lambda \in \mathbb{R}$$

- Canonical form: $x - 2 = \frac{y-1}{-2} = \frac{z-1}{3} = \lambda$

Notice that the forms are not unique. In the above example we could have choosen B as the "base point" and that would give us another set of equations for the same line. More precisely, any point that lays on the line can be taken as the "base point". So, for example, $\lambda = 2$ gives us the point $(4, -3, 7)$ and thus $(4, -3, 7) + \mu(1, -2, 3)$ is the same line.

The same is true if we take a (non-zero) multiple of $\overrightarrow{v}$ getting, for example, $(2, 1, 1)+ \rho(-2, 4, -6)$.

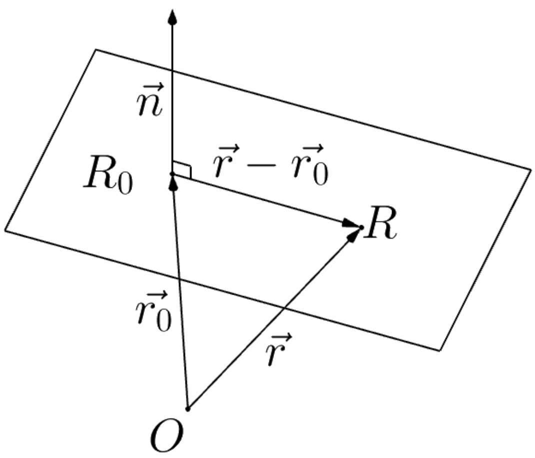
Planes in $\mathbb{R}^3$

Suppose V is a plane in $\mathbb{R}^3$. If you want to explain to someone else which plane it is, it suffices to give the following information:

- a point P_0 through which the plane passes and
- its "direction" (also "tilt" or orientation).

But how do we specify a direction of a plane? A smart move is to specify the plane's direction by giving the direction of a line which is orthogonal to the plane. The direction of that line can be given by a single (non-zero) vector (the length of the vector is irrelevant, only its direction matters).

Let us translate this into mathematics, first by an example and then in the general case.



Example 2.2. *Suppose the point $P_0(2, 1, 3)$ is on the plane U, and let $\overrightarrow{r}_0 = \overrightarrow{OP_0}$ be the vector pointing to P_0. Suppose moreover that $\overrightarrow{n} = (1, 2, 2)$ is orthogonal to U. Our task is to find those vectors $\overrightarrow{r} = (x, y, z)$ whose endpoints are in the plane. Now such an endpoint is in U if the direction from P_0 to this endpoint is orthogonal to $\overrightarrow{n}$. This direction is represented by the vector $\overrightarrow{r} - \overrightarrow{r_0} = (x - 2, y - 1, z - 3)$. So we require that the vector is orthogonal to $\overrightarrow{n} = (1, 2, 2)$, i.e.,*

$$\langle (x - 2, y - 1, z - 3), (1, 2, 2)\rangle = 0 \text{ or } (x - 2) + (y - 1) \cdot 2 + (z - 3) \cdot 2 = 0.$$

Of course we can rewrite this equation as $x + 2y + 2z = 10$*, but the form* $x - 2 + 2(y - 1) + 2(z - 3) = 0$ *shows clearly that (2,1,3) is on the plane. Note also that:*

- *For any non-zero scalar* t*, the equation* $tx + 2ty + 2tz = 10t$ *represents the same plane. For example* $6x + 12y + 12z = 60$.
- *The coefficients of* x, y, z *in the equation* $x + 2y + 2z = 10$ *form a vector which is (a multiple of) the vector* $\overrightarrow{n} = (1, 2, 2)$ *orthogonal to the plane we started with.*
- *The coordinates of* P_0 *satisfy the equation* $2 + 2 \cdot 1 + 2 \cdot 3 = 10$.

Here is now the **general story**.

Let $P_0(x_0, y_0, z_0)$ be a point and let $\overrightarrow{n} = (a, b, c)$ be a non-zero vector, which we suppose to be orthogonal to the plane we wish to describe. The plane through P and orthogonal to $\overrightarrow{n}$ consists of all the points $P(x, y, z)$ such that $\overrightarrow{P_0P}$ is orthogonal to $\overrightarrow{n}$, i.e., $(x - x_0, y - y_0, z - z_0)$ is orthogonal to (a, b, c). In vector form:

$$\langle (a, b, c), (x - x_0, y - y_0, z - z_0) \rangle = 0.$$

Vector form

A vector equation for the plane U through $P_0(x_0, y_0, z_0)$ orthogonal to the direction of the non-zero vector $\overrightarrow{n} = (a, b, c)$ is

$$\langle n, r - r_0 \rangle = 0,$$

where $\overrightarrow{r_0}$ corresponds to the position vector of $P_0(x_0, y_0, z_0)$.

By expanding the dot product we get:

$$a(x - x_0) + b(y - y_0) + c(z - z_0) = 0,$$

which is equivalent to

$$ax + by + cz = ax_0 + by_0 + cz_0.$$

Note that the right hand-side is a constant. If we replace the right hand side by a single symbol, we get the general form of the equation of a plane.

General form

A general form of the equation for the plane U through $P_0(x_0, y_0, z_0)$ orthogonal to the direction of the non-zero vector $\overrightarrow{n} = (a, b, c)$ is

$$ax + by + cz = d,$$

where $d = ax_0 + by_0 + cz_0$.

Another way of representing a plane is by its parametric vector equation, similar to the description of a line. First we discuss an example, where we use an intuitive approach, then we turn to a general strategy.

Example 2.3. *Suppose the plane passes through* $(0,0,0)$*, like for instance the plane* U *given by* $x+2y-3z=0$*. Take two vectors in the plane which are not multiples of each other, for example* $(2,-1,0)$ *and* $(3,0,1)$*. Geometrically, it is quite clear that any vector in the plane can be obtained by taking suitable multiples of the vectors and then adding them.*

So, any vector in the plane can be written as a linear combination

$$\lambda(2,-1,0)+\mu(3,0,1)$$

of (2,-1,0) and $(3,0,1)$ *for suitable* λ *and* μ*. Note that in this way we have described the vectors (points) of the plane explicitly: every value of* λ *and* μ *produces a point in the plane. For example, for* $\lambda=2$ *and* $\mu=-3$ *we find* $2\cdot(2,-1,0)-3(3,0,1)=(-5,-2,-3)$*.*

Next we turn to a general strategy for finding parametric vector descriptions. The way to find a parametric vector equation of a plane, starting from an equation, is to solve the equation (don't forget that there are infinitely many solutions; it's a plane after all). Suppose V is the plane with equation $x+2y-3z=4$; so V is parallel to U, since the vector $(1,2,-3)$ is orthogonal to both planes. If we rewrite this equation as $x=4-2y+3z$ then we clearly see, that for any given values of y and z, there is exactly one value of x such that the triple (x,y,z) is on the plane. So assign the value λ to y and μ to z. Then $x=4-2\lambda+3\mu$. We gain more insight to the solutions if we rewrite this explicit description in vector notation as follows:

$$(x,y,z)=(4-2\lambda+3\mu,\lambda,\mu)=(4,0,0)+\lambda(-2,1,0)+\mu(3,0,1).$$

Parametric vector description

A parametric vector description for the plane U is a description of the form

$$\overrightarrow{r_0}+\lambda\overrightarrow{u}+\mu\overrightarrow{v},$$

where $\overrightarrow{r_0}$ is a vector whose endpoint is in the plane and $\overrightarrow{u}$, $\overrightarrow{v}$ are two linearly independent vectors in the plane.

In this case you see (again) that V is parallel to U. How? The vector $(4,0,0)$ is usually called a *support vector*, while $(-2,1,0)$ and $(3,0,1)$ are called *direction vectors* (and they are the same as they were for U).

Note that the same plane can be described by parametric descriptions which may look quite different, as we will see in the following example. Equations of a plane (given by the vector form or by the general form), however, show less variation: if we only consider equations of the form $ax+by+cz=d$, then two such equations describe the same plane precisely when the coefficients differ by a common (non-zero) multiple.

Example 2.4. *Find a parametric vector description of the plane* $U: x+y+z=4$*.*

Solution. To find a parametric vector description of the plane $U: x+y+z=4$, any of the following approaches can be taken.

a) We solve for x, so we first rewrite as $x = 4 - y - z$. If we let $y = \lambda$ and $z = \mu$, then $x = 4 - \lambda - \mu$. In parametric vector form this becomes

$$(x, y, z) = (4 - \lambda - \mu, \lambda, \mu) = (4, 0, 0) + \lambda(-1, 1, 0) + \mu(-1, 0, 1).$$

Intuitively, the plane "rests" on $(4, 0, 0)$ and an arbitrary vector in the plane is described by adding any combination of the vectors $(-1, 1, 0)$ and $(-1, 0, 1)$ to $(4, 0, 0)$.

b) Alternatively, by looking carefully at the equation, pick any vector in U, say $(1, 1, 2)$. We use this as a support vector. Now take any two vectors which are orthogonal to $(1, 1, 2)$ and which are linearly independent, i.e., pick two independent solutions of $u + v + 2w = 0$. For example $(2, 2, -2)$ and $(0, 2, -1)$. Then the plane is described as

$$(x, y, z) = (1, 1, 2) + \lambda(2, 2, -2) + \mu(0, 2, -1).$$

Note that the description is quite different from the previous description. This illustrates the fact that the parametric vector description of planes are far from being unique.

c) If you follow the method as explained in *a)* but solve for y rather than for x, then the resulting parametric equation is again somewhat different: start by rewriting the equation as $y = 4 - x - z$, then set $x = \lambda$ and $z = \mu$. Finally, you get $(x, y, z) = (0, 4, 0) + \lambda(1, -1, 0) + \mu(0, -1, 1)$.

We have come across two ways of representing planes: by equations and by parametric descriptions. To go from the first to the second, comes down to solving an equation representing a plane and rewriting the solutions in the appropriate vector form, as we already discussed. Now we discuss how to transform parametric vector descriptions into equations. This is best illustrated by the next example.

Example 2.5. *Find the general form equation of the plane U given by*

$$(2, 1, 3) + \lambda(1, 2, 1) + \mu(1, 1, 3).$$

Solution. Our task is to find an equation of the form $ax + by + cz = d$ representing U, in particular, we have to find a, b, c, and d. This problem splits into two parts:

- To find a, b, and c note that the vector (a, b, c) is orthogonal to the plane, so is orthogonal to both direction vectors $(1, 2, 1)$ and $(1, 1, 3)$. This means:

$$\langle (a, b, c), (1, 2, 1) \rangle = 0 \text{ and } \langle (a, b, c), (1, 1, 3) \rangle = 0.$$

So we have to solve:

$$\begin{aligned} a + 2b + c &= 0 \\ a + b + 3c &= 0 \end{aligned}$$

for a, b, and c. By solving the system we get $a = -5c$ and $b = 2c$. So one vector perpendicular to $(1, 2, 1)$ and $(1, 1, 3)$ is, for example, the vector $(a, b, c) = (-5, 2, 1)$ obtained by taking $c = 1$. Thus our equation looks like $-5x + 2y + z = d$ and it remains to find d.

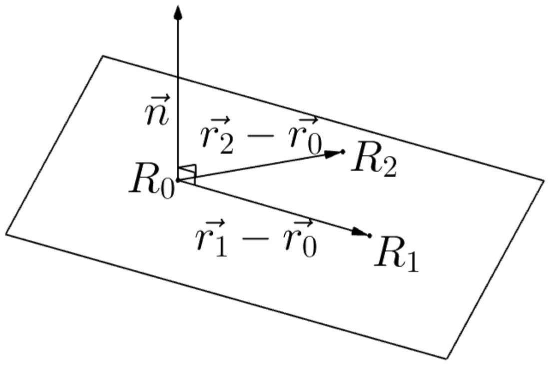
- To find d in $-5x+2y+z=d$ is simple: just substitute in the equation the x, y, z with any vector of U, for instance the support vector $(2,1,3)$:

$$(-5)\cdot 2 + 2\cdot 1 + 1\cdot 3 = d.$$

  In conclusion, a general form equation for U is $-5x+2y+z=-5$.

In this last part we have seen that two linearly independent vectors define a plane. If two linearly independent vectors define a plane, so do three noncollinear points, t.i., three points that do not lie on a single line.

Let R_0, R_1, and R_2 be three noncollinear points in space and denote their position vectors by $\overrightarrow{r_0}, \overrightarrow{r_1}$, and $\overrightarrow{r_2}$ (in this order). Then $\overrightarrow{R_0R_1} = \overrightarrow{r_1} - \overrightarrow{r_0}$ and $\overrightarrow{R_0R_2} = \overrightarrow{r_2} - \overrightarrow{r_0}$ are two linearly independent vectors and they define a plane.



By observing that in this case $\overrightarrow{n} = (\overrightarrow{r_1} - \overrightarrow{r_0}) \times (\overrightarrow{r_2} - \overrightarrow{r_0})$ (recall $\overrightarrow{u} \times \vec{v} = (|\vec{u}||\overrightarrow{v}|\sin\varphi)\vec{n}$) and putting that into the vector form equation of the plane $\langle n, \overrightarrow{r} - \overrightarrow{r_0}\rangle$ we get

$$(\overrightarrow{r_1} - \overrightarrow{r_0}) \times (\overrightarrow{r_2} - \overrightarrow{r_0}) \cdot (\overrightarrow{r} - \overrightarrow{r_0}) = 0.$$

Or, as a box product:

$$\langle (\overrightarrow{r_1} - \overrightarrow{r_0}) \times (\overrightarrow{r_2} - \overrightarrow{r_0}), (\overrightarrow{r} - \overrightarrow{r_0})\rangle = 0.$$

If the position vectors of the three points are $r_i = (x_i, y_i, z_i)$ (for $i = 0, 1, 2$) and $r = (x, y, z)$ is a point on the plane, then the expression

$$\begin{vmatrix} x - x_0 & y - y_0 & z - z_0 \\ x_1 - x_0 & y_1 - y_0 & z_1 - z_0 \\ x_2 - x_0 & y_2 - y_0 & z_2 - z_0 \end{vmatrix} = 0$$

will give us the vector form equation of a plane.

Example 2.6. *Find the vector form equation of the plane defined by points $A(1,0,0)$, $B(2,3,1)$, and $C(5,4,-1)$.*

Solution. Let $r = (x, y, z)$, $r_0 = (1,0,0)$, $r_1 = (2,3,1)$, and $r_2 = (5,4,-1)$. Then from the expression

$$\begin{vmatrix} x-1 & y & z \\ 1 & 3 & 1 \\ 4 & 4 & -1 \end{vmatrix} = 0$$

we get that

$$(x-1)(-3-4) - y(-1-4) + z(4-12) = 0,$$

or

$$-7x + 5y - 8z + 7 = 0.$$

2.1 The relative position of lines and planes: intersections

Intersecting two lines

Let $\ell_1 = \overrightarrow{r_1} + \lambda_1 \overrightarrow{v_1}$ and $\ell_2 = \overrightarrow{r_2} + \lambda_2 \overrightarrow{v_2}$ be two lines in $\mathbb{R}^3$ that are not parallel or skew lines. Then ℓ_1 and ℓ_2 intersect in exactly one point. To determine the point of intersection, we have to solve the equation

$$\overrightarrow{r_1} + \lambda_1 \overrightarrow{v_1} = \overrightarrow{r_2} + \lambda_2 \overrightarrow{v_2}$$

Claim 2.7. *Let $\ell_1 = \overrightarrow{r_1} + \lambda_1 \overrightarrow{v_1}$ and $\ell_2 = \overrightarrow{r_2} + \lambda_2 \overrightarrow{v_2}$ be two lines in space. Then ℓ_1 and ℓ_2 intersect if and only if there exists such λ_1 and λ_2, that the following equation holds:*

$$\overrightarrow{r_1} + \lambda_1 \overrightarrow{v_1} = \overrightarrow{r_2} + \lambda_2 \overrightarrow{v_2}.$$

Example 2.8. *Prove that $\ell_1 = (2,1,0) + \lambda_1(1,0,1)$ and $\ell_2 = (1,4,2) + \lambda_2(1,-1,0)$ intersect for some $\lambda_1, \lambda_2 \in \mathbb{R}$.*

Solution. By the above claim it should hold that $(2,1,0) + \lambda_1(1,0,1) = (1,4,2) + \lambda_2(1,-1,0)$. From the given equation we get a system of three equations:

$$\begin{aligned} 2 + \lambda_1 &= 1 + \lambda_2 \\ 1 &= 4 - \lambda_2 \\ \lambda_1 &= 2 \end{aligned}$$

From the second equation we get $\lambda_2 = 3$ and the third one gives us $\lambda_1 = 2$. If this solution is feasible also in the first equation, then we have found our λ_1 and λ_2, otherwise such λ does not exist.

Since $2 + 2 = 4 = 1 + 3$, lines ℓ_1 and ℓ_2 intersect, furthermore, we can tell that the intersection point is $P(4,1,2)$.

Intersecting a line and a plane

Let $\ell_1 = (r_1, r_2, r_3) + \lambda(v_1, v_2, v_3)$ ba a line and $ax + by + cz = d$ be a plane in $\mathbb{R}^3$. The intersection of a line and a plane usually consists of exactly one point. If the line and the plane happen to be parallel, the intersection may be empty or consist of the whole line. To determine the intersection we have to find the value(s) of λ for which the corresponding point(s) on ℓ belong to U as well. In particular, we have to solve the equation

$$a(r_1 + \lambda v_1) + b(r_2 + \lambda v_2) + c(r_3 + \lambda v_3) = d.$$

By substituting the obtained value for λ in the equation of ℓ we get the point of intersection.

Example 2.9. *Determine the point of intersection of the plane U, determined by $2x + 3y - 5z = 3$ and the line $\ell = (-1, 4, -3) + \lambda(1, -1, 2)$.*

Solution. To find the value of λ for which the corresponding point on ℓ belongs to U we have to solve the equation:

$$2(-1 + \lambda) + 3(4 - \lambda) - 5(-3 + 2\lambda) = 3.$$

This reduces to $25 - 11\lambda = 3$, so $\lambda = 2$.

Now we substitute this value of λ in the equation of ℓ to find the point of intersection: $(-1, 4, -3) + 2(1, -1, 2) = (1, 2, 1)$.

In the case when the line is parallel to the plane (but not laying on the plane), the above equation will result in a contradiction, if the line lies on the plane, the solution of the above equation are all $\lambda \in \mathbb{R}$.

Intersecting two planes

Let $a_1x + b_1y + c_1z = d_1$ and $a_2x + b_2y + c_2z = d_2$ be two planes in $\mathbb{R}^3$. If the two planes are not parallel, then it is geometrically quite clear that they meet along a line. To find the line of intersection we have to solve the system given by the equations representing the two planes. Since there are three unknowns and only two equations we have to express two of the unknowns x, y, z with the third one. By assigning an arbitrary value λ to this third unknown we get the equation for the line in vector form.

Example 2.10. *Find the intersection between planes $x+2y-z = 4$ and $2x+y-5z = 2$.*

Solution. As illustrated above we have to find the solutions of the set of equalities.

$$\begin{aligned} x + 2y - z &= 4 \\ 2x + y - 5z &= 2 \end{aligned}$$

Which comes down to

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

And from here we can express x and y in terms of z as follows

$$\begin{aligned} x &= 3z \\ y &= 2 - z \end{aligned}$$

The next step consists of assigning the value λ to z, resulting in the following equation:

$$(x, y, z) = (3\lambda, 2 - \lambda, \lambda) = (0, 2, 0) + \lambda(3, -1, 1)$$

which is exactly the equation of the line of intersection in vector form.

2.2 The relative position of points, lines and planes: distances

Computing the distance between two points is straightforward, but the computation of distances between points and lines or points and planes is more subtle. In this section we will look at strategies to compute distances and give some explicit formulas.

The distance between a point and a line

Suppose P is a point and ℓ is a line. The distance between P and a point on ℓ varies as this point varies through ℓ. So the question is: for which point on ℓ is this distance minimal? A bit of experimentation with the Pythagorean Theorem shows that the point Q, such that $\overrightarrow{PQ}$ is orthogonal to ℓ is the point we are looking for.

Let us work this out on a specific example, where P is the point $(7, 2, -3)$ and ℓ is the line given by the vector form $(2, -1, 1) + \lambda(1, -1, 2)$.

- First, an arbitrary point Q on ℓ is described by $(2+\lambda, -1-\lambda, 1+2\lambda)$. The vector $\overrightarrow{PQ}$ is then $(-5 + \lambda, -3 - \lambda, 4 + 2\lambda)$.
- The next step is to solve for λ from the condition that PQ is orthogonal to the direction vector $(1, -1, 2)$ of ℓ, i.e., solve
$$\langle(-5 + \lambda, -3 - \lambda, 4 + 2\lambda), (1, -1, 2)\rangle = 0.$$
Thus $(-5 + \lambda) - (-3 - \lambda) + 2(4 + 2\lambda) = 0$ or $6 + 6\lambda = 0$. So $\lambda = -1$ and the point Q on ℓ closest to P is therefore (by making a substitution in the parametric equation of ℓ) $(2, -1, 1) - (1, -1, 2) = (1, 0, -1)$.
- Finally, the distance between P and the line ℓ is calculated as the distance between P and Q:
$$\sqrt{(7 - 1)^2 + (2 - 0)^2 + (-3 + 1)^2} = \sqrt{44} = 2\sqrt{11}.$$

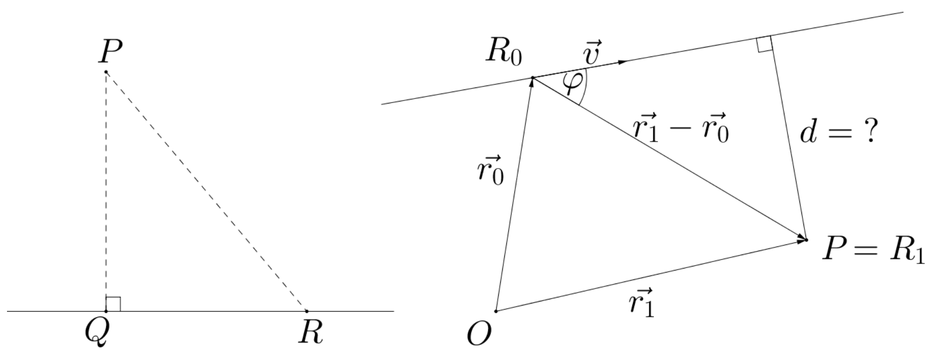
In the above procedure we used the dot product to ensure orthogonality, but how can the described procedure lead to an explicit formula for computing the distance between a point and a line?

In fact it can not, at least not directly. To gain an explicit formula we have to adopt a different point of view.

Let $P = R_1$ be a given point in space and ℓ a line given by the vector form. What we want to compute is the distance d from ℓ to R_1. The distance d can be also seen as the height of the parallelepiped determined by $\overrightarrow{r_1} - \overrightarrow{r_0}$ and $\overrightarrow{v}$.

So to obtain an explicit formula we start with the magnitude of the cross product:
$$|(\overrightarrow{r_1} - \overrightarrow{r_0}) \times v| = |(\overrightarrow{r_1} - \overrightarrow{r_0})| \cdot |\overrightarrow{v}| \cdot \sin\varphi.$$

Using the right triangle, we see
$$\sin\varphi = \frac{d}{|\overrightarrow{r_1} - \overrightarrow{r_0}|},$$



and therefore the magnitude of the cross product becomes

$$|(\overrightarrow{r_1} - \overrightarrow{r_0}) \times v| = |\overrightarrow{r_1} - \overrightarrow{r_0}| \cdot |\overrightarrow{v}| \cdot \frac{d}{|\overrightarrow{r_1} - \overrightarrow{r_0}|} = d \cdot |\overrightarrow{v}|.$$

To obtain an explicit formula for the distance we only have to solve for d.

The distance between a point and a line
Let R_1 be a point in space and ℓ a line. Then the distance between R_1 and ℓ is

$$d = \frac{|(\overrightarrow{r_1} - \overrightarrow{r_0}) \times v|}{|\overrightarrow{v}|},$$

where $\overrightarrow{r_1}$ is the position vector of the point R_1, and $\ell = \overrightarrow{r_0} + \lambda\vec{v}$.

Example 2.11. *Calculate the distance from point $P(1,2,1)$ to a line through $A(1,0,-1)$ and $B(3,-1,2)$.*

Solution. First we need to compute the direction vector $\overrightarrow{v} = (3,-1,2) - (1,0,-1) = (2,-1,3)$.

Then we proceed as follows:

- $\overrightarrow{r_P} - \overrightarrow{r_A} = (1,2,1) - (1,0,-1) = (0,2,2)$
- $(r_P - r_A) \times v = \begin{vmatrix} i & j & k \\ 0 & 2 & 2 \\ 2 & -1 & 3 \end{vmatrix} = (8,4,-4)$
- $|(r_P - r_A) \times v| = \sqrt{64+16+16} = \sqrt{96}$
- $|\overrightarrow{v}| = \sqrt{4+1+9} = \sqrt{14}$

Thus $d = \frac{\sqrt{96}}{\sqrt{14}} = \sqrt{\frac{48}{7}}$ units.

The distance between a point and a plane

The distance between a point P and a point Q in the plane U varies as Q varies. By the distance between a point P and a plane U we, again, mean the shortest possible distance between P and any of the points of U. But how do we find such a point in the plane? It is geometrically obvious that we can locate such a point by moving from P in the direction orthogonal to U, until we meet U, i.e., we need the line through P orthogonal to the plane U. This is, as in the previous case, based on the Pythagorean Theorem.

Let us use this strategy when P is the point $(5, -4, 6)$ and U is the plane given by the equation $x - 2y + 2z = 7$.

- First we look for the line through P and orthogonal to U. A vector orthogonal to U is easily extracted from its equation: $(1, -2, 2)$. So the line through P with direction vector $(1, -2, 2)$ is described by

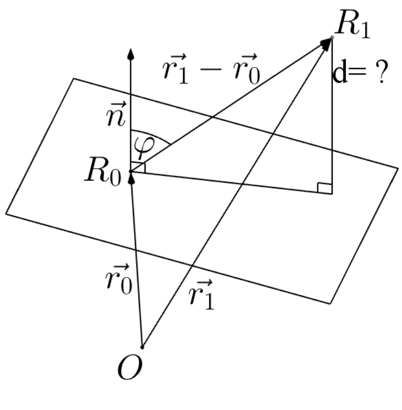
$$(5, -4, 6) + \lambda(1, -2, 2).$$

- The next step is to intersect this line with U. To this end we substitute $(5, -4, 6)+ \lambda(1, -2, 2)$ in the equation: $(5 + \lambda) - 2(-4 - 2\lambda) + 2(6 + 2\lambda) = 7$. This is easily written as $25 + 9\lambda = 7$, so that $\lambda = -2$. For $\lambda = -2$ we find the point $Q = (5, -4, 6) - 2(1, -2, 2) = (3, 0, 2)$. So $Q(3, 0, 2)$ is the point on U closest to P.

- Finally, the distance between P and U is calculated as the distance between P and Q:$\sqrt{4 + 16 + 16} = \sqrt{36} = 6$ units.

If the plane equation is not given by the general form but, for example, by the parametric vector description, we can proceed as above, we just first have to compute the general form (from the parametric vector description), as we did in Example 2.5.

We will now describe how to get an explicit formula to calculate the distance between a point and a plane.



Let R_1 be a given point in space and U a plane given by the general form. What we want to compute is the distance d from U to R_1. The distance d can be also seen as the length of the vector projection of $\overrightarrow{r_1} - \overrightarrow{r_0}$ onto $\overrightarrow{n}$. Therefore

The distance between a point and a plane

Let R_1 be a point in space and U a plane. Then the distance between R_1 and U is

$$d = \frac{|\langle \overrightarrow{r_1} - \overrightarrow{r_0}, n\rangle|}{|\overrightarrow{n}|},$$

where r_1 is the position vector of the point R_1, r_0 is the position vector of a point R_0 on the plane U and $\overrightarrow{n}$ is the direction vector orthogonal to U.

Example 2.12. *Find the distance between the point $P(4, 0, 1)$ and the plane U given by $2x - y + 2z = 3$.*

Solution. Let $\overrightarrow{r_1} = (4, 0, 1)$ and $\overrightarrow{n} = (2, -1, 2)$. Furthermore, choose a point on the plane, for example $R_0(1, -1, 0) = \overrightarrow{r_0}$. Then

$$d = \frac{|\langle (3, 1, 1), (2, -1, 2)\rangle|}{\sqrt{4 + 1 + 4}} = \frac{6 - 1 + 2}{\sqrt{9}} = \frac{7}{3} \text{ units.}$$

The distance between two lines

In $\mathbb{R}^2$ and in $\mathbb{R}^3$ the distance between two lines is the minimum distance between a point from the first line and a point from the second line.

If two lines intersect, then the distance between them is 0. In $\mathbb{R}^2$ the only other option is, that the two lines are parallel. If this is the case, then the distance between the two lines is equal to the distance between any point on the first line and a second line. Luckily we already know how to calculate that. In $\mathbb{R}^3$ we have three options:

1.) The two lines intersect.

Then the distance between them is 0.

2.) The two lines are parallel.

Then the distance between them can be calculated as the distance between a point on the first line and a second line (we already did that).

Claim 2.13. *Two lines in $\mathbb{R}^3$ are parallel if and only if the cross product of their direction vectors is $\overrightarrow{0}$.*

Proof. Two lines are parallel if we can obtain one from the other by a parallel displacement. This is only possible when their direction vector point either in the same or in the opposite direction. In this case the cross product of the direction vectors equals $\overrightarrow{0}$ as by definition $\overrightarrow{u} \times \overrightarrow{v}$ has a factor of $\sin\varphi$. □

Example 2.14. *Show that $\ell_1 = (3, 2, -2) + \lambda_1(2, 4, -2)$ and $\ell_2 = (1, 0, -1) + \lambda_2(-1, -2, 1)$ are parallel.*

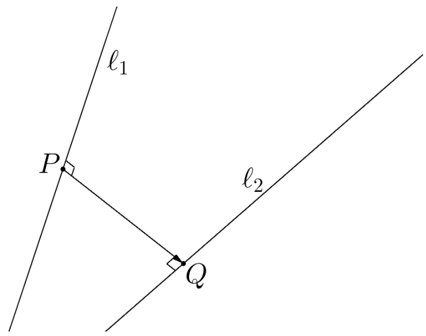
Solution. By the above claim we have to show that $v_1 \times v_2 = \overrightarrow{0}$. Indeed,

$$(2, 4, -2) \times (-1, -2, 1) = \begin{vmatrix} i & j & k \\ 2 & 4 & -2 \\ -1 & -2 & 1 \end{vmatrix} = (0, 0, 0).$$

Corollary 2.15. *Two lines in $\mathbb{R}^3$ are parallel if and only if their direction vectors are linearly dependent.*

3.) The two lines are *skew lines*, that is, they are not parallel and do not intersect.

In this case the computation of the distance turns out to be more complicated. Suppose ℓ_1 and ℓ_2 are two such lines. Among all points P on ℓ_1 and all points Q on ℓ_2 we need to find points for which the distance $|PQ|$ is minimal. As before this condition turns out to be related to right angles: we have to find P on ℓ_1 and Q on ℓ_2 in such a way that $\overrightarrow{PQ}$ is orthogonal to both ℓ_1 and ℓ_2.



The procedure is best illustrated with an example.

Suppose ℓ_1 is given by $(1,3,3)+\lambda_1(2,0,1)$ and ℓ_2 is given by $(1,6,-3)+\lambda_2(0,1,1)$. The lines are certainly not parallel, since the direction vectors are not multiples of one another. They may intersect, but then our distance computation will simply yield 0.

- Begin by taking an arbitrary point P on ℓ_1, for example $(1+2\lambda_1, 3, 3+\lambda_1)$ and an arbitrary point Q on ℓ_2, say $(1, 6+\lambda_2, -3+\lambda_2)$.
- Then we impose the condition that $\overrightarrow{PQ}$ has to be orthogonal to both direction vectors of the two lines, t.i., to $(2,0,1)$ and $(0,1,1)$.

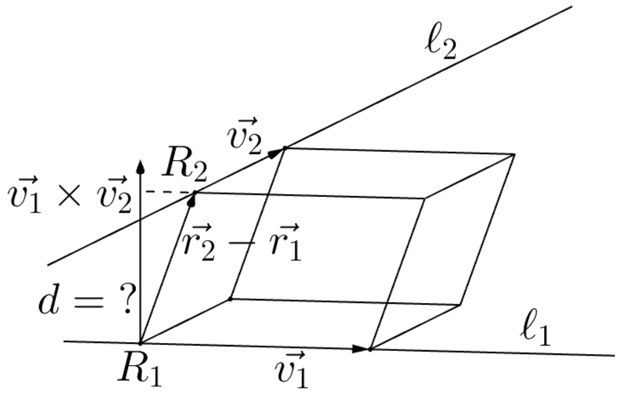
  Since $\overrightarrow{PQ} = (-2\lambda_1, 3+\lambda_2, -6-\lambda_1+\lambda_2)$, the orthogonality condition translates into two equations in the variables λ_1 and λ_2:

  $$\begin{aligned} -4\lambda_1 + (-6-\lambda_1+\lambda_2) &= 0 \\ (3+\lambda_2) + (-6-\lambda_1+\lambda_2) &= 0 \end{aligned}$$

  Which leads to a system that has exactly one solution: $\lambda_1 = -1$ and $\lambda_2 = 1$.
- At last, corresponding to this solution we find the desired points $P(-1,3,2)$, and $Q(1,7,-2)$. The distance between the lines is therefore

  $$\sqrt{2^2+4^2+(-4)^2} = \sqrt{36} = 6 \text{ units.}$$

To obtain en explicit formula note that if $\ell_1 = \overrightarrow{r_1} + \lambda_1\overrightarrow{v_1}$ and $\ell_2 = \overrightarrow{r_2} + \lambda_2\overrightarrow{v_2}$ are two skew lines, then $v_1 \times v_2 \neq 0$. Therefore $|\langle v_1 \times v_2, r_2 - r_1\rangle|$ is equal to the volume of the parallelepiped determined by $\overrightarrow{v_1}, \overrightarrow{v_2}$, and $\overrightarrow{r_2} - \overrightarrow{r_1}$. The vector $v_1 \times v_2$ is orthogonal to both direction vectors $\overrightarrow{v_1}$ and $\overrightarrow{v_2}$ and because of that the distance between ℓ_1 and ℓ_2 equals the length of the vector projection of $\overrightarrow{r_2} - \overrightarrow{r_1}$ onto $v_1 \times v_2$. From here we get the explicit formula:



The distance between two skew lines

Let $\ell_1 = \vec{r_1} + \lambda_1 \vec{v_1}$ and $\ell_2 = \vec{r_2} + \lambda_2 \vec{v_2}$ be two skew lines. Then the distance between them is

$$d = \frac{|\langle v_1 \times v_2, r_2 - r_1 \rangle|}{|v_1 \times v_2|}.$$

Example 2.16. *Find the distance between* $\ell_1 = (3, -4, 4) + \lambda(-2, 7, 2)$ *and* $\ell_2 = (-3, 4, 1) + \mu(1, -2, -1)$.

Solution. Compute
$\vec{r_2} - \vec{r_1} = (-6, 8, -3)$ and
$v_1 \times v_2 = (-3, 0, -3)$.
Then the distance is $d = \frac{|18+0+9|}{\sqrt{9+0+9}} = \frac{27}{\sqrt{18}} = \frac{3\sqrt{2}}{2}$ units.

The distance between a line and a plane

If the line and the plane intersect, then the distance between them is 0. If the line and the plane do not intersect, then they are parallel and therefore the distance between them can be calculated as the distance between any point on the line and the plane.

Claim 2.17. *Let* $\ell = \vec{r_0} + \lambda \vec{v}$ *be a line and* $ax + by + cz = d_2$ *be a plane, call it* U. *Then* ℓ *and* U *are parallel if and only if* $\langle (a, b, c), v \rangle = 0$.

Proof. The proof is left as an exercise. □

Example 2.18. *Find the distance between the line* $\ell = (2, -1, 0) + \lambda(2, 0, 1)$ *and the plane* $x - 2y - 2z = 0$.

Solution. Let $\vec{v} = (2, 0, 1)$, and $\vec{n} = (1, -2, -2)$. Furthermore let $A(2, -1, 0) = \vec{r_A}$ be a point on ℓ and $B(0, 0, 0) = \vec{r_B}$ be a point on the plane. Since $\langle n, v \rangle = 0$ the line and the plane are indeed parallel. Then $\vec{r_A} - \vec{r_B} = (2, -1, 0)$, $\langle r_A - r_B, n \rangle = 4$ and $|\vec{n}| = 3$. Therefore the distance is $d = \frac{4}{3}$ units.

The distance between two planes

As we observed in the previous case, there are only two options how can two planes lie in space. If the two planes intersect, then the distance between them is 0. If the two planes do not intersect, then they are parallel and therefore the distance between them can be calculated as the distance between any point on the first plane and the second plane.

Claim 2.19. *Let $a_1x + b_1y + c_1z = d_1$ and $a_2x + b_2y + c_2z = d_2$ be two planes, call them U and V, respectively. Then U and V are parallel (or possibly equal) if and only if $(a_1, b_1, c_1) \times (a_2, b_2, c_2) = 0$.*

Proof. The statement of the claim follows from the fact that two planes are parallel if and only if their direction vectors, call them $\overrightarrow{n_U}$ and $\overrightarrow{n_V}$, point either in the same or in the opposite direction. This is equivalent to the statement that $n_U \times n_V = 0$. □

Corollary 2.20. *Let $a_1x + b_1y + c_1z = d_1$ and $a_2x + b_2y + c_2z = d_2$ be two planes, call them U and V, and let $\overrightarrow{n_U} = (a_1, b_1, c_1)$ and $\overrightarrow{n_V} = (a_2, b_2, c_2)$ be their respective direction vectors. Then U and V are parallel if and only if $\overrightarrow{n_U} = \alpha\overrightarrow{n_V}$ for some $\alpha \in \mathbb{R}$.*

Example 2.21. *Find the distance between planes $x + y = 1$ and $x + y = 3$.*

Solution. Since $\overrightarrow{n_1} = \overrightarrow{n_2} = (1, 1, 0)$ the two planes are indeed parallel. Let $A(1, 0, 0) = \overrightarrow{r_A}$ be a point on the first plane and $B(2, 1, 0) = \overrightarrow{r_B}$ be a point on the second plane. Then $\overrightarrow{r_B} - \overrightarrow{r_A} = (1, 1, 0)$ and $\langle r_B - r_A, n\rangle = 2$. Since $|\overrightarrow{n}| = \sqrt{2}$ the distance between the two planes is: $d = \sqrt{2}$.

2.3 The relative position of lines and planes: angles

The angle between two lines

The angle between two lines $\ell_1 = \overrightarrow{r_1} + \lambda_1\overrightarrow{v_1}$ and $\ell_2 = \overrightarrow{r_2} + \lambda_2\overrightarrow{v_2}$ that are not parallel is defined as the acute angle between their direction vectors $\overrightarrow{v_1}$ and $\overrightarrow{v_2}$. To compute the angle between $\overrightarrow{v_1}$ and $\overrightarrow{v_2}$ we use the dot product, however this does not ensure that the obtained angle will be acute. To this end we slightly change the dot product formula.

The angle between two lines

If $\overrightarrow{v_1}$ and $\overrightarrow{v_2}$ are the direction vectors of lines ℓ_1 and ℓ_2, respectively, then the angle φ between ℓ_1 and ℓ_2 is computed from

$$\cos\varphi = \frac{|\langle v_1, v_2\rangle|}{|\overrightarrow{v_1}| \cdot |\overrightarrow{v_2}|}.$$

Using the absolute value in the enumerator ensures that the fraction is non-negative and consequently that the angle φ is acute.

Example 2.22. *Determine the angle between $\ell_1 = (2, 7, -1) + \lambda_1(1, 0, 1)$ and $\ell_2 = (2, 0, -1) + \lambda_2(1, 1, 2)$.*

Solution.

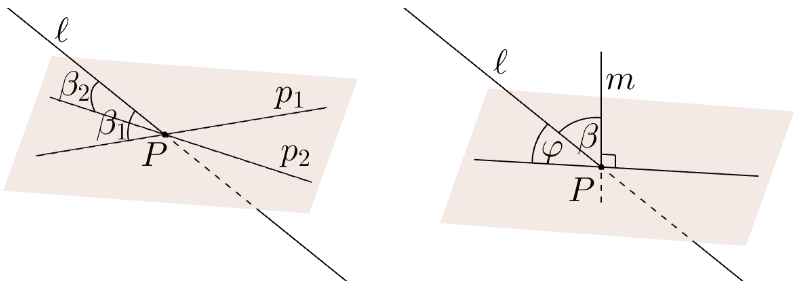
$$\cos\varphi = \frac{\langle v_1, v_2\rangle}{|\overrightarrow{v_1}|\cdot|\overrightarrow{v_2}|} = \frac{|1+0+2|}{|\sqrt{1+0+1}|\cdot|\sqrt{1+1+4}|} = \frac{3}{\sqrt{2\cdot 6}} = \frac{\sqrt{3}\sqrt{3}}{\sqrt{2}\sqrt{2}\sqrt{3}} = \frac{\sqrt{3}}{2}$$

Therefore $\varphi = 30°$.

The angle between a line and a plane

Consider a line ℓ and a plane U in space that are not parallel. Since they are not parallel there is some angle between them. But, if we think a moment about that angle we quickly realize that it is not clear what the angle between a line ℓ and a plane U should be precisely. The reason is that the plane contains so many directions. So which direction of the plane should we compare with the direction of the line?

Suppose the line ℓ and the plane U meet in a point P. The plane U contains a whole family of lines through P and any such line makes an angle with ℓ. Thus the angle with such a line in the plane varies as this line varies on U. Can you see the only exception?



The "way out" is to realize the the direction of a plane is also characterized by the direction of a line orthogonal to the plane. So let us take a line m through P which is orthogonal to U. It is easy to see what the angle between ℓ and m is, but this angle is not really what we want, $90°-$ *this angle* turns out to be the appropriate angle.

Since we are actually calculating the angle between two lines, the equation given before applies. And it can be further simplified by observing that $\cos(90-\varphi) = \sin\varphi$. To summarize,

> **The angle between a line and a plane**
> The angle φ between a line $\ell = \overrightarrow{r_0} + \lambda\overrightarrow{v}$ and a plane U is computed from
>
> $$\sin\varphi = \frac{|\langle v, n\rangle|}{|\overrightarrow{v}|\cdot|\overrightarrow{n}|},$$
>
> where $\overrightarrow{n}$ is the vector orthogonal to U.

Example 2.23. *Compute the angle φ between the line $\ell = (2,0,3) + \lambda(0,1,1)$ and the plane U given by $x + y + 2z = \sqrt{13}$.*

Solution. We can proceed in two ways:

a) By using directly the above equation:

$$\sin\varphi = \frac{\langle v, n\rangle}{|\overrightarrow{v}| \cdot |\overrightarrow{n}|} = \frac{|0+1+2|}{|\sqrt{0+1+1}| \cdot |\sqrt{1+1+4}|} = \frac{3}{\sqrt{2\cdot 6}} = \frac{\sqrt{3}}{2}$$

Therefore $\varphi = 60°$.

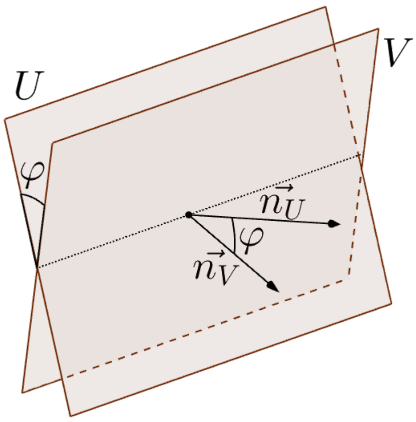
b) By using the formula for the angle β between two lines and then calculating the wanted angle:

$$\cos\beta = \frac{\langle v, n\rangle}{|\overrightarrow{v}| \cdot |\overrightarrow{n}|} = \frac{\sqrt{3}}{2}$$

Therefore $\beta = 30°$ and $\varphi = 90° - 30° = 60°$.

The angle between two planes

Suppose U and V are two planes which meet along a line ℓ. Of course they seem to make a definite angle with each other, but again, to get the right angle we need to look at the angle between the vectors orthogonal to each plane. Therefore the angle between two planes U and V with direction vector $\overrightarrow{n_U}$ and $\overrightarrow{n_V}$, respectively, can be obtained as the angle between $\overrightarrow{n_U}$ and $\overrightarrow{n_V}$.



In particular, two planes are orthogonal, if the direction vectors are orthogonal.

Example 2.24. *Compute the angle φ between planes U and V with equations $x-z=7$ and $x-y-2z=25$, respectively.*

Solution. We first need to determine $\overrightarrow{n_U}$ and $\overrightarrow{n_V}$. Let $\overrightarrow{n_U} = (1,0,-1)$ and $\overrightarrow{n_V} = (1,-1,-2)$. Then

$$\cos\varphi = \frac{\langle n_U, n_V\rangle}{|\overrightarrow{n_U}| \cdot |\overrightarrow{n_V}|} = \frac{1+0+2}{\sqrt{1+0+1}\cdot\sqrt{1+1+4}} = \frac{3}{\sqrt{2}\cdot\sqrt{6}} = \frac{\sqrt{3}}{2}$$

Therefore $\varphi = 30°$.

3
MATRICES

If m and n are positive integers then by a **matrix of size m by n**, or an $m \times n$ **matrix** we mean a rectangular array consisting of $m \cdot n$ numbers in a boxed display having m rows and n columns. Simple examples of such objects are the following:

$$\text{size } 1 \times 5: \quad \begin{bmatrix} 10 & 9 & 8 & 7 & 6 \end{bmatrix} \qquad \text{size } 3 \times 2: \quad \begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{bmatrix}$$

$$\text{size } 4 \times 4: \quad \begin{bmatrix} 1 & 2 & 3 & 4 \\ 2 & 3 & 4 & 5 \\ 3 & 4 & 5 & 6 \\ 4 & 5 & 6 & 7 \end{bmatrix} \qquad \text{size } 3 \times 1: \quad \begin{bmatrix} 0 \\ 2 \\ 4 \end{bmatrix}$$

In general we will display an $m \times n$ matrix as:

$$X = \begin{bmatrix} x_{11} & x_{12} & x_{13} & \cdots & x_{1n} \\ x_{21} & x_{22} & x_{23} & \cdots & x_{2n} \\ x_{31} & x_{32} & x_{33} & \cdots & x_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ x_{m1} & x_{m2} & x_{m3} & \cdots & x_{mn} \end{bmatrix}.$$

Note that in the given display the first subscript gives the number of the row and the second subscript that of the column, so that x_{ij} appears at the intersection of the i-th row and the j-th column. For convenience the given display can be shortly written as $X = (x_{ij})_{m \times n}$, meaning that X is the $m \times n$ matrix whose **(i,j)-th element** (or the **(i,j)-th entry** is x_{ij}.

Example 3.1. *The* 3×3 *matrix* $X = \begin{bmatrix} 1 & 1 & 1 \\ 2 & 2^2 & 2^3 \\ 3 & 3^2 & 3^3 \end{bmatrix}$ *can be expressed as* $X = (x_{ij})_{3 \times 3}$ *where* $x_{ij} = i^j$.

Example 3.2. *The* 3×3 *matrix whose entries are given by* $x_{ij} = (-1)^{i-j}$ *looks like* *this:*

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

.

Some special matrices

Some matrices have a special structure and therefore a "name" was given to them. The basic ones are:

- The **row matrix**, that is a matrix of size $1 \times n$ (that is often referred to as a vector).
- The **column matrix**, that is a matrix of size $n \times 1$ (that is also called a column vector).
- The **square matrix**, that is a matrix with same number of rows and columns, i.e. an $n \times n$ matrix.
- The **diagonal matrix**, that is a square matrix with $a_{ij} = 0$ for all $i \neq j$ (we usually give this matrices in the abbreviated form as *diag*(1,2,3))
- The **upper (or lower) triangular matrix**, that is a square matrix with elements $a_{ij} = 0$ for all $i > j$ (or $i < j$), in particular, all elements below (or above) the diagonal are equal to 0. A *triangular matrix* is either upper triangular or lower triangular.
- The **strictly triangular matrix**, that is an (upper or lower) triangular matrix with diagonal entries equal to 0.

We will encounter some more special kinds of matrices in the following sections.

Before we move on to the algebra of matrices it is essential that we decide what is meant by saying that two matrices are *equal*.

Definition 3.3. *If $A = (a_{ij})_{m \times n}$ and $B = (b_{ij})_{p \times q}$ then we say that A and B are* ***equal****, and write $A = B$, if and only if*

(a) $m = p$ and $n = q$

(b) $a_{ij} = b_{ij}$ for all i, j.

3.1 An Introduction to Gaussian Elimination

We wish to solve the system

$$\begin{array}{rcrcrcr} x_1 & - & 5x_2 & + & 4x_3 & = & -3 \\ 2x_1 & - & 7x_2 & + & 3x_3 & = & -2 \\ -2x_1 & + & x_2 & + & 7x_3 & = & -1 \end{array}$$

First, we may represent this system with a matrix:

$$\left[\begin{array}{rrr|r} 1 & -5 & 4 & -3 \\ 2 & -7 & 3 & -2 \\ -2 & 1 & 7 & -1 \end{array}\right]$$

The first column of the above matrix contains the coefficients of the variable x_1, the second column the coefficients of x_2, the third the coefficients of the variable x_3. The dotted line (written here are a solid for ease of typesetting), represents the '=' sign, and the last column the constants. Such a matrix is called the **augmented matrix** with respect to the system.

There are three, relatively obvious, ways of manipulating or rewriting the system to obtain another system with the same solutions:

1. Rearranging the order of the equations. As the order in which the equations in a system of equations is arbitrary, we may rearrange the equations in any way we choose. We will swap the second and third rows to obtain the system

   $$\begin{array}{rcrcrcr} x_1 & - & 5x_2 & + & 4x_3 & = & -3 \\ -2x_1 & + & x_2 & + & 7x_3 & = & -1 \\ 2x_1 & - & 7x_2 & + & 3x_3 & = & -2 \end{array} .$$

   The augmented matrix with respect to this system is obtained from the original matrix by swapping rows 2 and 3:

   $$\left[\begin{array}{rrr|r} 1 & -5 & 4 & -3 \\ -2 & 1 & 7 & -1 \\ 2 & -7 & 3 & -2 \end{array}\right].$$

2. Multiply a row by a constant. We will multiply the second equation by 3:

   $$\begin{array}{rcrcrcr} x_1 & - & 5x_2 & + & 4x_3 & = & -3 \\ -2x_1 & + & x_2 & + & 7x_3 & = & -1 \\ 2x_1 & - & 7x_2 & + & 3x_3 & = & -2 \end{array} .$$

   The augmented matrix with respect to this system is obtained from the original matrix by multiplying each entry in the second row by 3:

   $$\left[\begin{array}{rrr|r} 1 & -5 & 4 & -3 \\ 6 & -21 & 9 & -6 \\ -2 & 1 & 7 & -1 \end{array}\right].$$

3. Multiply a row by a constant and add to another row: Let's multiply Equation 1 by -2 and add to equation 2:

$$\begin{array}{rcrcrcr} -2x_1 & + & 10x_2 & - & 8x_3 & = & 6 \\ 2x_1 & - & 7x_2 & + & 3x_3 & = & -2 \\ \hline & & 3x_2 & - & 5x_3 & = & 4 \end{array}$$

The system of equations is then

$$\begin{array}{rcrcrcr} x_1 & - & 5x_2 & + & 4x_3 & = & -3 \\ & & 3x_2 & - & 5x_3 & = & -4 \\ 2x_1 & - & 7x_2 & + & 3x_3 & = & -2 \end{array}$$

with augmented matrix

$$\left[\begin{array}{rrr|r} 1 & -5 & 4 & -3 \\ 0 & 3 & -5 & -4 \\ 2 & -7 & 3 & -2 \end{array}\right].$$

We denote these changes by $-2R1 + R2 = R_2$ which means to multiply Row 1 by Row 2 and put the result in Row 2. Note that we 'eliminated' the variable x_1 from this equation.

Our goal: Use the elementary row operations 1), 2), and 3) to make the matrix look like:

$$\left[\begin{array}{ccc|c} 1 & 0 & 0 & a \\ 0 & 1 & 0 & b \\ 0 & 0 & 1 & c \end{array}\right].$$

if possible! Then the corresponding system is $x_1 = a$, $x_2 = b$, and $x_3 = c$. This process is called **Gaussian elimination**. Back to our original problem:

$$\left[\begin{array}{rrr|r} 1 & -5 & 4 & -3 \\ 2 & -7 & 3 & -2 \\ -2 & 1 & 7 & -1 \end{array}\right] \xrightarrow{-2R1+R2=R2} \left[\begin{array}{rrr|r} 1 & -5 & 4 & -3 \\ 0 & 3 & -5 & 4 \\ -2 & 1 & 7 & -1 \end{array}\right]$$

$$\begin{array}{rcrcrcr} 2x_1 & - & 10x_2 & - & 8x_3 & = & -6 \\ -2x_1 & + & x_2 & + & 7x_3 & = & -2 \\ \hline & & -9x_2 & + & 15x_3 & = & -8 \end{array} \quad \xrightarrow{-2R1+R3=R3} \left[\begin{array}{rrr|r} 1 & -5 & 4 & -3 \\ 0 & 3 & -5 & 4 \\ 0 & -9 & 15 & -8 \end{array}\right]$$

$$\begin{array}{rcrcr} 9x_2 & - & 15x_3 & = & 12 \\ -9x_2 & + & 15x_3 & = & -8 \\ \hline 9 & + & 0 & = & -4 \end{array} \quad \xrightarrow{3R2+R3=R3} \left[\begin{array}{rrr|r} 1 & -5 & 4 & -3 \\ 0 & 3 & -5 & 4 \\ 0 & 0 & 0 & -4 \end{array}\right]$$

This gives that $0 = 4$, which is not true. Hence the system has no solution or that the system is inconsistent.

Example Solve the system of linear equations

$$\begin{array}{rcrcrcr} x_1 & & & - & 3x_3 & = & 8 \\ 2x_1 & + & 2x_2 & + & 9x_3 & = & 7 \\ & & x_2 & + & 5x_3 & = & -2 \end{array}$$

We write the augmented matrix and reduce to reduced row echelon form.

$$
\left[\begin{array}{ccc|c} 1 & 0 & -3 & 8 \\ 2 & 2 & 9 & 7 \\ 0 & 1 & 5 & -2 \end{array}\right]
\xrightarrow{-2R1+R2}
\left[\begin{array}{ccc|c} 1 & 0 & -3 & 8 \\ 0 & 2 & 15 & -9 \\ 0 & 1 & 5 & -2 \end{array}\right]
$$

$$
\xrightarrow{Swap\ R2\ and\ R3}
\left[\begin{array}{ccc|c} 1 & 0 & -3 & 8 \\ 0 & 1 & 5 & -2 \\ 0 & 2 & 15 & -9 \end{array}\right]
$$

Note that the form to the right is echelon form. Intuitively, this means one can "see" a solution exists by inspection

$$
\xrightarrow{-2R2+R3=R3}
\left[\begin{array}{ccc|c} 1 & 0 & -3 & 8 \\ 0 & 1 & 5 & -2 \\ 0 & 0 & 5 & -5 \end{array}\right]
$$

$$
\xrightarrow{R3/5=R3}
\left[\begin{array}{ccc|c} 1 & 0 & -3 & 8 \\ 0 & 1 & 5 & -2 \\ 0 & 0 & 1 & -1 \end{array}\right]
$$

$$
\xrightarrow[-5R3+R2=R2]{R3+R1=R1}
\left[\begin{array}{ccc|c} 1 & 0 & 0 & 5 \\ 0 & 1 & 0 & 3 \\ 0 & 0 & 1 & -1 \end{array}\right]
$$

So $x = -5$, $y = 3$, and $z = -1$.

Example Solve the system

$$
\begin{array}{ccccccc} x & + & 2y & - & 5z & = & 0 \\ & & y & - & 3z & = & -2 \end{array}
$$

We write the augmented matrix and reduce to reduced row echelon form.

$$
\left[\begin{array}{ccc|c} 1 & 2 & -5 & 0 \\ 0 & 1 & -3 & -2 \end{array}\right]
\xrightarrow{-2R2+R1=R1}
\left[\begin{array}{ccc|c} 1 & 0 & 1 & 4 \\ 0 & 1 & -3 & -2 \end{array}\right]
$$

Now we return to the system and solve for x and y:

$$
\begin{array}{ccccccc} x & & & + & z & = & 4 \\ & & y & - & 3z & = & -2 \end{array}
\quad \text{or} \quad
\begin{array}{ccr} x & = & 4 - z \\ y & = & -2 + 3z \\ z & = & z \end{array}
$$

Then $(x, y, z) = (4 - z, -2 + 3z, z) = (4, -2, 0) + z(-1, 3, 1)$, which is the parametric vector form of a line! So geometrically, the two planes given by the two linear equations intersect in the line that is the solution!

Example

Solve the system of linear equations

$$
\begin{array}{ccccccc} x & + & y & + & z & = & 2 \\ & & 4y & + & 3z & = & 13 \\ 4x & + & 2y & & & = & 4 \end{array}
$$

We write the augmented matrix and reduce to reduced row echelon form.

$$\left[\begin{array}{ccc|c} 1 & 1 & 1 & 2 \\ 0 & 4 & 3 & 13 \\ 4 & 2 & 0 & 4 \end{array}\right] \xrightarrow{-4R1+R3=R3} \left[\begin{array}{ccc|c} 1 & 1 & 1 & 2 \\ 0 & 4 & 3 & 13 \\ 0 & -2 & -4 & -4 \end{array}\right]$$

$$\xrightarrow{-R3/2=R3} \left[\begin{array}{ccc|c} 1 & 1 & 1 & 2 \\ 0 & 4 & 3 & 13 \\ 0 & 1 & 2 & 2 \end{array}\right]$$

$$\xrightarrow[-4R3+R2=R2]{-R3+R1=R1} \left[\begin{array}{ccc|c} 1 & 0 & -1 & 0 \\ 0 & 0 & -5 & 5 \\ 0 & 1 & 2 & 2 \end{array}\right]$$

$$\xrightarrow[R3 \text{ becomes } R2]{-R2/5=R3} \left[\begin{array}{ccc|c} 1 & 0 & -1 & 0 \\ 0 & 1 & 2 & 2 \\ 0 & 0 & 1 & -1 \end{array}\right]$$

$$\xrightarrow[-2R3+R2=R2]{R3+R1=R1} \left[\begin{array}{ccc|c} 1 & 0 & 0 & -1 \\ 0 & 1 & 0 & 4 \\ 0 & 0 & 1 & -1 \end{array}\right]$$

So $x = -1$, $y = 4$, and $z = -1$.

3.2 The algebra of matrices

The algebra of matrices has many of the familiar properties enjoyed by the system of real numbers, however there are some important differences as well.

3.2.1 Matrix addition

First we will take a look at the sum of two matrices.

Definition 3.4. *Given $m \times n$ matrices $A = (a_{ij})$ and $B = (b_{ij})$ the* ***sum*** *$A + B$ is the $m \times n$ matrix whose (i, j)-th element is $a_{ij} + b_{ij}$.*

Note that the sum $A + B$ is defined only when A and B are of the same size and to obtain this sum we simply add corresponding entries therefore obtaining again a matrix of the same size.

Example 3.5. $\begin{bmatrix} -1 & 0 \\ 2 & -2 \end{bmatrix} + \begin{bmatrix} 1 & 2 \\ -1 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 2 \\ 1 & -2 \end{bmatrix}$

Theorem 3.6. *Let $A = (a_{ij})$, $B = (b_{ij})$ and $C = (c_{ij})$ be three $m \times n$ matrices. Then the addition of matrices is:*

(1) commutative: $A + B = B + A$

(2) associative: $A + (B + C) = (A + B) + C$

Proof. (1) First note that since A and B are of size $m \times n$ then $A+B$ and $B+A$ are also of size $m \times n$. By the definition of addition we have that $A+B = (a_{ij}+b_{ij})$ and $B+A = (b_{ij}+a_{ij})$. Since the addition of numbers is commutative we have $a_{ij}+b_{ij} = b_{ij}+a_{ij}$ for all i, j and so, by the definition of equality of matrices, we conclude that $A+B = B+A$.

(2) Can be proved using the same reasoning as for (1).

□

Definition 3.7. *The $m \times n$ matrix whose entries are all equal* 0 *is called the $m \times n$* zero matrix *and its denoted by $0_{m\times n}$ or simply by* 0.

Theorem 3.8. *There is a unique $m \times n$ matrix M such that, for every $m \times n$ matrix A, $A + M = A$.*

Proof. Let M be the zero matrix. Then for every matrix $A = (a_{ij})_{m\times n}$ we have

$$A + M = (a_{ij} + m_{ij})_{m\times n} = (a_{ij} + 0)_{m\times n} = (a_{ij})_{m\times n} = A.$$

To establish the uniqueness of matrix M, suppose that $B = (b_{ij})_{m\times n}$ is also such that $A + B = A$ for every $m \times n$ matrix A. Then in particular we have $M + B = M$. But, taking B instead of A in the property for M, we have $B + M = B$. It now follows by Theorem 3.6 that $B = M$. □

Because of this property the zero matrix is also called *the additive identity.*

Definition 3.9. *Let A be an $m\times n$ matrix. The the $m\times n$ matrix B for which $A+B = 0$ is called the* additive inverse *of A and is denoted by $-A$.*

Theorem 3.10. *For every $m \times n$ matrix A there is a unique $m \times n$ matrix B such that $A + B = 0$.*

Proof. Given $A = (a_{ij})_{m\times n}$, consider $B = (-a_{ij})_{m\times n}$, i.e. the matrix whose (i, j)-th element is the additive inverse of the (i, j)-th element of A. Clearly we have

$$A + B = (a_{ij} + (-a_{ij}))_{m\times n} = 0.$$

To establish the uniqueness of such a matrix B, suppose that $C = (c_{ij})_{m\times n}$ is also such that $A + C = 0$. Then for all i, j we have $a_{ij} + c_{ij} = 0$ and consequently $c_{ij} = -a_{ij}$ which means, by the above definition of equality, that $C = B$. □

Given numbers x and y the **difference** $x - y$ is defined to be $x + (-y)$. For matrices A, B of the same size we similarly write $A - B$ for $A + (-B)$, the operation "$-$" so defined being called **substraction** of matrices.

So far our matrix algebra has been confined to the operation of addition. This is a simple extension of the same notion for numbers, for we can think of 1×1 matrices as behaving essentially as numbers. We will now investigate how the notion of multiplication for numbers can be extended to matrices. This, however, is not quite so straightforward. There are in fact two distinct *multiplications* that can be defined. The first involving a matrix and a number, and the second involving two matrices.

3.2.2 Multiplication of a matrix by a scalar

Definition 3.11. *Given a matrix A and a number λ, the* ***product of*** *A* ***by*** *λ is the matrix denoted by λA, that is obtained from A by multiplying every element of A by λ. Thus, if $A = (a_{ij})_{m\times n}$, then $\lambda A = (\lambda a_{ij})_{m\times n}$.*

This operation is traditionally called **multiplying a matrix by a scalar**. Such multiplication by scalars may also be thought of as scalars *acting* on matrices. The principal properties of this action are as follows.

Theorem 3.12. *Let A, B be two $m \times n$ matrices and λ, μ two scalars. Then*

(1) $\lambda(A + B) = \lambda A + \lambda B$

(2) $(\lambda + \mu)A = \lambda A + \mu A$

(3) $\lambda(\mu A) = (\lambda\mu)A$

(4) $(-1)A = -A$

(5) $0A = 0_{m\times n}$

Proof. Let $A = (a_{ij})_{m\times n}$ and $B = (b_{ij})_{m\times n}$. Then the above equalities follow from the observations

(1) $\lambda(a_{ij} + b_{ij}) = \lambda a_{ij} + \lambda b_{ij}$

(2) $(\lambda + \mu)a_{ij} = \lambda a_{ij} + \mu a_{ij}$

(3) $\lambda(\mu a_{ij}) = (\lambda\mu)a_{ij}$

(4) $(-1)a_{ij} = -a_{ij}$

(5) $0a_{ij} = 0$

□

Example 3.13. *Given any $m \times n$ matrices A and B, solve the matrix equation*

$$3\left(X + \frac{1}{2}A\right) = 5\left(X - \frac{3}{4}B\right).$$

Solution.

$$\begin{aligned}
3\left(X + \frac{1}{2}A\right) &= 5\left(X - \frac{3}{4}B\right)\\
3X + \frac{3}{2}A &= 5X - \frac{15}{4}B\\
\frac{3}{2}A + \frac{15}{4}B &= 5X - 3X\\
2X &= \frac{3}{2}A + \frac{15}{4}B\\
X &= \frac{3}{4}A + \frac{15}{8}B
\end{aligned}$$

3.2.3 Matrix multiplication

We will now describe the operation that is called **matrix multiplication**. This is the multiplication of one matrix by another. At first sight this concept (due originally to Cayley) appears to be a most curious one.

Definition 3.14. *Let $A = (a_{ij})_{m\times n}$ and $B = (b_{ij})_{n\times p}$ (note the sizes!). Then we defne the* product AB *to be the $m \times p$ matrix whose (i, j)-th element is*

$$(AB)_{ij} = a_{i1}b_{1j} + a_{i2}b_{2j} + a_{i3}b_{3j} + \cdots + a_{in}b_{nj}.$$

In other words, the (i, j)-th element of the product AB is obtained by summing the products of the elements in the i-th **row** of A with the corresponding elements in the j-th **column** of B.

The above expression for $(AB)_{ij}$ can be written in abbreviated form using the so-called **Σ-notation**:

$$(AB)_{ij} = \sum_{k=1}^{n} a_{ik}b_{kj}.$$

It is important to note that, in forming these sums of products, there are no elements that are "left over" since in the definition of the product AB the number n of columns of A is the same as the number of rows of B.

Example 3.15. *Consider the matrices* $A = \begin{bmatrix} 0 & 1 & 0 \\ 2 & 3 & 1 \end{bmatrix}$ *and*

$B = \begin{bmatrix} 2 & 0 \\ 1 & 2 \\ 1 & 1 \end{bmatrix}$. *The product AB is defined since A is of size 2×3 and B is of size 3×2; moreover, AB is of size 2×2. We have:*

$$AB = \begin{bmatrix} 0\cdot 2 + 1\cdot 1 + 0\cdot 1 & 0\cdot 0 + 1\cdot 2 + 0\cdot 1 \\ 2\cdot 2 + 3\cdot 1 + 1\cdot 1 & 2\cdot 0 + 3\cdot 2 + 1\cdot 1 \end{bmatrix} = \begin{bmatrix} 1 & 2 \\ 8 & 7 \end{bmatrix}.$$

Note that in this case the product BA is also defined (since B has the same number of columns as A has rows). The product BA is of size 3×3:

$$BA = \begin{bmatrix} 2\cdot 0 + 0\cdot 2 & 2\cdot 1 + 0\cdot 3 & 2\cdot 0 + 0\cdot 1 \\ 1\cdot 0 + 2\cdot 2 & 1\cdot 1 + 2\cdot 3 & 1\cdot 0 + 2\cdot 1 \\ 1\cdot 0 + 1\cdot 2 & 1\cdot 1 + 1\cdot 3 & 1\cdot 0 + 1\cdot 1 \end{bmatrix} = \begin{bmatrix} 0 & 2 & 0 \\ 4 & 7 & 2 \\ 2 & 4 & 1 \end{bmatrix}.$$

The above example exhibits a curious fact concerning matrix multiplication, namely that if AB and BA are defined then these products are not necessarily equal. Indeed, as we have just seen, AB and BA are not even of the same size. It is also possible for AB and BA to be defined and of the same size and still not be equal.

Definition 3.16. *We say that matrices A, B* commute *if $AB = BA$.*

Note that for A, B to commute it is necessary that they are square and of the same size. The next example will show us that even this conditions are not enough.

Example 3.17. *The matrices* $A \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$ *and* $B = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$ *are such that $AB = 0$ and $BA = A$.*

We thus observe that in general *matrix multiplication is not commutative.* Another curious property of matrix multiplication arising from this example is that the product of two non-zero matrices can be the zero matrix.

We will now consider the basic properties of matrix multiplication.

Theorem 3.18. *Matrix multiplication is associative (in the sense that, when the products are defined, $A(BC) = (AB)C$).*

Proof. For $A(BC)$ to be defined we require the respective sizes to be $m \times n$, $n \times p$, $p \times q$ in which case the product $A(BC)$ is also defined, and conversely. Computing the (i, j)-th element of $A(BC)$, we obtain

$$\begin{aligned}(A(BC))_{ij} = \sum_{k=1}^{n} a_{ik}(BC)_{kj} &= \sum_{k=1}^{n} a_{ik}\Big(\sum_{t=1}^{p} b_{kt}c_{tj}\Big)\\ &= \sum_{k=1}^{n}\sum_{t=1}^{p} a_{ik}b_{kt}c_{tj}.\end{aligned}$$

If we now compute the (i, j)-th element of $(AB)C$, we obtain the same:

$$\begin{aligned}((AB)C)_{ij} = \sum_{t=1}^{p} (AB)_{it}c_{tj} &= \sum_{t=1}^{p}\Big(\sum_{k=1}^{n} a_{ik}b_{kt}\Big)c_{tj}\\ &= \sum_{t=1}^{p}\sum_{k=1}^{n} a_{ik}b_{kt}c_{tj}.\end{aligned}$$

Consequently we see that $A(BC) = (AB)C$. □

Matrix multiplication and matrix addition are connected by the following **distributive laws**.

Theorem 3.19. *When the relevant sums and products are defined, we have*

- $A(B + C) = AB + AC$ *and*
- $(B + C)A = BA + CA$

Proof. For the first equality we require A to be of size $m \times n$ and B, C to be of size $n \times p$, in which case

$$\begin{aligned}(A(B + C))_{ij} = \sum_{k=1}^{n} a_{ik}(B + C)_{kj} &= \sum_{k=1}^{n} a_{ik}(b_{kj} + c_{kj})\\ &= \sum_{k=1}^{n} a_{ik}b_{kj} + \sum_{k=1}^{n} a_{ik}c_{kj}\\ &= (AB)_{ij} + (AC)_{ij}\\ &= (AB + AC)_{ij}\end{aligned}$$

For the second equality, in which we require B, C to be of size $m \times n$ and A to be of size $n \times p$, a similar argument applies. □

Matrix multiplication is also connected with multiplication by scalars.

Theorem 3.20. *If AB is defined, then for all scalars λ we have*

$$\lambda(AB) = (\lambda A)B = A(\lambda B).$$

Proof. The (i,j)-th elements of the three products are

$$\lambda(\sum_{k=1}^{n} a_{ik}b_{kj}) = \sum_{k=1}^{n}(\lambda a_{ik})b_{kj} = \sum_{k=1}^{n} a_{ik}(\lambda b_{kj}),$$

from which the result follows. □

Definition 3.21. *The $n \times n$ diagonal matrix whose elements are all equal 1 is called the* identity matrix *and is denoted by I_n or simply by I.*

Our next result is the multiplicative analogue of Theorem 3.8 (about the unique additive identity), but we should note that it applies only in the case of square matrices.

Theorem 3.22. *There exists a unique $n \times n$ matrix M with the property that, for every $n \times n$ matrix A, $AM = A = MA$.*

Proof. Let M to be the identity matrix. If $A = (a_{ij})_{n\times n}$ then $(AM)_{ij} = \sum_{k=1}^{n} a_{ik}m_{kj} = a_{ij}$, the last equality following from the fact that every term in the summation is 0 except that in which $k = j$, and this term is $a_{ij} \cdot 1 = a_{ij}$. We deduce, therefore, that $AM = A$. Similarly, we can show that $MA = A$. This then establishes the existence of a matrix M with the stated property. To show that such a matrix M is unique, suppose that P is also an $n \times n$ matrix such that $AP = A = PA$ for every $n \times n$ matrix A. Then in particular we have $MP = M = PM$. But, by the same property for M, we have $PM = P = MP$. Thus we see that $P = M$. □

Because of this property (and by analogy withe the additive identity), this matrix is also called the *multiplicative identity.*

Note that there is no analogue of Theorem 3.10 (about the additive inverse); for example, if

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

then we have

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} \cdot \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 0 & a \\ 0 & c \end{bmatrix},$$

so there is no matrix M such that $MA = I_2$.

3.2.4 The transpose of a matrix

Definition 3.23. *If A is an $m \times n$ matrix then by the* transpose *of A we mean the $n \times m$ matrix A^T whose (i,j)-th element is the (j,i)-th element of A. More precisely, if $A = (a_{ij})_{m\times n}$ then $A^T = (a_{ji})_{n\times m}$.*

The principal properties of transposition of matrices are summarised in the following result.

Theorem 3.24. *Where the relevant sums and products are defined, we have*

(1) $(A^T)^T = A$

(2) $(A + B)^T = A^T + B^T$

(3) $(\lambda A)^T = \lambda A^T$

(4) $(AB)^T = B^T A^T$

Proof. The first three equalities follow immediately from the definitions. To prove that $(AB)^T = B^T A^T$, suppose that $A = (a_{ij})_{m\times n}$ and $B = (b_{ij})_{n\times p}$. Then $(AB)^T$ and $B^T A^T$ are each of size $p \times m$. Since

$$(B^T A^T)_{ij} = \sum_{k=1}^{n} b_{ki} a_{jk} = \sum_{k=1}^{n} a_{jk} b_{ki} = (AB)_{ji} = (AB)^T_{i,j}$$

we deduce that $(AB)^T = B^T A^T$. □

We will now define two more "special" types of matrices.

Definition 3.25. *A square matrix A is* symmetric *if $A = A^T$ and* skew-symmetric *if $A = -A^T$.*

Example 3.26. *For every square matrix A the matrix $A + A^T$ is symmetric, and the matrix $A - A^T$ is skew-symmetric. In fact, by the previous theorem we have*

$$(A + A^T)^T = A^T + (A^T)^T = A^T + A = A + A^T,$$

$$-(A - A^T)^T = -(A^T - (A^T)^T) = -(A^T - A) = -A^T + A = A - A^T.$$

Theorem 3.27. *Every square matrix can be expressed uniquely as the sum of a symmetric matrix and a skew-symmetric matrix.*

Before giving the proof of the theorem, let's take a look at an example.

Example 3.28. *Write the matrix $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ as the the sum of a symmetric matrix and a skew-symmetric matrix.*

Solution. First note that $A^T = \begin{bmatrix} a & c \\ b & d \end{bmatrix}$, and therefore that $A + A^T = \begin{bmatrix} 2a & b+c \\ c+b & 2d \end{bmatrix}$ is symmetric and that $A - A^T = \begin{bmatrix} 0 & b-c \\ c-b & 0 \end{bmatrix}$ is skew-symmetric. Therefore $\begin{bmatrix} a & b \\ c & d \end{bmatrix} = \alpha \begin{bmatrix} 2a & b+c \\ c+b & 2d \end{bmatrix} + \beta \begin{bmatrix} 0 & b-c \\ c-b & 0 \end{bmatrix}$.

From here we get a system of equalities:

$$\begin{aligned} a &= 2a\alpha + 0\beta \\ b &= \alpha(b+c) + \beta(b-c) \\ c &= \alpha(c+b) + \beta(c-b) \\ d &= 2d\alpha + 0\beta \end{aligned}$$

From the first (and the last) equation we get $\alpha = \frac{1}{2}$. Using this in the third equation gives us $\beta = \frac{1}{2}$. The obtained solution is feasible also in the second equation, therefore

$$A = \frac{1}{2}(A + A^T) + \frac{1}{2}(A - A^T).$$

Proof of Theorem 3.27. The equality

$$A = \frac{1}{2}(A + A^T) + \frac{1}{2}(A - A^T)$$

shows that such an expression exists. As for uniqueness, suppose that $A = B + C$, where B is symmetric and C is skew-symmetric. Then $A^T = B^T + C^T = B - C$. It follows from these equations that $B = \frac{1}{2}(A + A^T)$ and $C = \frac{1}{2}(A - A^T)$. □

Furthermore, note that the sum of two symmetric (or equivalently skew-symmetric) matrices is again a symmetric (equiv. skew-symmetric) matrix and that a scalar multiple of a symmetric (or. equiv. skew-symmetric) matrix is also a symmetric (equiv. skew-symmetric) matrix. But the product of two symmetric (or equiv. skew-symmetric) matrices is not necessary a symmetric (equiv. skew-symmetric) matrix, as we will see in the next example.

Example 3.29. *The product of matrices* $A = \begin{bmatrix} 2 & 1 \\ 1 & -1 \end{bmatrix}$ *and* $B = \begin{bmatrix} -1 & 1 \\ 1 & 0 \end{bmatrix}$ *is*

$AB = \begin{bmatrix} -1 & 2 \\ -2 & 1 \end{bmatrix}$ *which is not a symmetric matrix.*

Claim 3.30. *If A is an $m \times n$ matrix then the matrix AA^T is symmetric.*

Proof. From the properties of matrix transposition we have

$$(AA^T)^T = (A^T)^T A^T = AA^T.$$

Therefore AA^T is a symmetric matrix. □

3.3 Systems of linear equations

We will now consider in some detail a systematic method of solving systems of linear equations. In working with such systems, there are three basic operations involved:

(1) interchanging two equations (usually for convenience);

(2) multiplying an equation by a non-zero scalar;

(3) forming a new equation by adding one equation to another.

The operation of adding a multiple of one equation to another can be achieved by a combination of (2) and (3).

We begin by considering the following three examples.

Example 3.31. *To solve the system*

$$\begin{array}{rcrcrcrl} & & y & + & 2z & = & 1 & (e_1) \\ x & - & 2y & + & z & = & 0 & (e_2) \\ & & 3y & - & 4z & = & 23 & (e_3) \end{array}$$

We multiply equation (e_1) by 3 and subtract the new equation from equation (e_3) to obtain $-10z = 20$, whence we see that $z = -2$. It then follows from equation (e_1) that $y = 5$, and then by equation (e_2) that $x = 2y - z = 12$.

Example 3.32. *Consider the system*

$$\begin{array}{rcrcrcrl} x & - & 2y & - & 4z & = & 0 & \quad (e_1) \\ -2x & + & 4y & + & 3z & = & 1 & \quad (e_2) \\ -x & + & 2y & - & z & = & 1 & \quad (e_3). \end{array}$$

If we add together equations (e_1) *and* (e_2)*, we obtain equation* (e_3)*, which is therefore redundant. Thus we have only two equations in three unknowns. What do we mean by a solution in this case?*

Example 3.33. *Consider the system*

$$\begin{array}{rcrcrcrcrl} x & + & y & + & z & + & t & = & 1 & \quad (e_1) \\ x & - & y & - & z & + & t & = & 3 & \quad (e_2) \\ -x & - & y & + & z & - & t & = & 1 & \quad (e_3) \\ -3x & + & y & - & 3z & - & 3t & = & 4 & \quad (e_4). \end{array}$$

Adding equations (e_1) *and* (e_2)*, we obtain* $x+t=2$*, whence it follows that* $y+z=-1$*. Adding equations* (e_1) *and* (e_3)*, we obtain* $z=1$ *and consequently* $y=-2$*. Substituting in equation* (e_4)*, we obtain* $-3x-3t=9$ *so that* $x+t=-3$*, which is not consistent with* $x+t=2$*. This system therefore does not have a solution. Expressed in another way, given the* 4×4 *matrix*

$$A = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & -1 & 1 \\ -1 & -1 & 1 & -1 \\ -3 & 1 & -3 & -3 \end{bmatrix},$$

there are no numbers x, y, z, t *that satisfy the matrix equation*

$$A \cdot \begin{bmatrix} x \\ y \\ z \\ t \end{bmatrix} = \begin{bmatrix} 1 \\ 3 \\ 1 \\ 4 \end{bmatrix}.$$

The above three examples were chosen to provoke the question: is there a systematic method of tackling systems of linear equations that

- avoids a random manipulation of the equations;
- yields all the solutions when they exist;
- makes it clear when no solution exists?

In what follows our objective will be to obtain a complete answer to this questions.

We note first that in dealing with systems of linear equations the "unknowns" play a secondary role. It is in fact the coefficients (which are usually integers) that are important. Indeed, each such system is completely determined by its augmented matrix. In order to work solely with this, we consider the following **elementary row operations** on this matrix:

(1) interchange two rows;

(2) multiply a row by a non-zero scalar;

(3) add one row to another.

These elementary row operations clearly correspond to the basic operations on equations listed above.

It is important to observe that these elementary row operations do not affect the solutions (if any) of the system. In fact, if the original system of equations has a solution then this solution is also a solution of the system obtained by applying any of the operations (1), (2), (3); and since we can in each case perform the "inverse" operation and thereby obtain the original system, the converse is also true.

We begin by showing that the above elementary row operations have a fundamental interpretation in terms of matrix products.

Theorem 3.34. *Let P be the $n \times n$ matrix that is obtained from the identity matrix I_n by permuting its rows in some way. Then for any $m \times n$ matrix A the matrix PA is the matrix obtained from A by permuting its rows in precisely the same way.*

Proof. Suppose that the i-th row of P is the j-th row of I_n. Then we have

$$(k = 1, \dots, n) \quad p_{ik} = \delta_{jk}.$$

Here, δ_{jk} is the Kronecker delta function which is defined by $\delta_{jk} = 1$ if $j = k$ and $\delta_{jk} = 0$ otherwise. Consequently, for every value of k,

$$(PA)_{ik} = \sum_{t=1}^{n} p_{it}a_{tk} = \sum_{t=1}^{n} \delta_{jt}a_{tk} = a_{jk},$$

hence we see that the i-th row of PA is the j-th row of A. □

Example 3.35. *Consider the matrix*

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

Obtained from I_4 by permuting the second and third rows. If we consider any 4×2 matrix

$$A = \begin{bmatrix} a_1 & b_1 \\ a_2 & b_2 \\ a_3 & b_3 \\ a_4 & b_4 \end{bmatrix}$$

and we compute the product

$$PA = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} a_1 & b_1 \\ a_2 & b_2 \\ a_3 & b_3 \\ a_4 & b_4 \end{bmatrix} = \begin{bmatrix} a_1 & b_1 \\ a_3 & b_3 \\ a_2 & b_2 \\ a_4 & b_4 \end{bmatrix},$$

we see that the effect of multiplying A on the left by P is to permute the second and the third rows of A.

Theorem 3.36. *Let A be an $n \times n$ matrix and let D be the $n \times n$ diagonal matrix*

$$D = \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{bmatrix}.$$

Then DA is obtained from A by multiplying the i-th row of A by λ_i for $i = 1, \ldots, n$.

Proof. Clearly, we have $d_{ij} = \lambda_i \delta_{ij}$. Consequently,

$$(DA)_{ij} = \sum_{k=1}^{n} d_{ik} a_{kj} = \sum_{k=1}^{n} \lambda_i \delta_{ik} a_{kj} = \lambda_i a_{ij},$$

and so the i-th row of DA is simply λ_i times the i-th row of A. □

Example 3.37. *Consider the matrix*

$$D = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \alpha & 0 & 0 \\ 0 & 0 & \beta & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix},$$

Obtained from I_4 by multiplying the second row by α and the third row by β. If we consider any 4×2 matrix

$$A = \begin{bmatrix} a_1 & b_1 \\ a_2 & b_2 \\ a_3 & b_3 \\ a_4 & b_4 \end{bmatrix}$$

and we compute the product

$$DA = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \alpha & 0 & 0 \\ 0 & 0 & \beta & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} a_1 & b_1 \\ a_2 & b_2 \\ a_3 & b_3 \\ a_4 & b_4 \end{bmatrix} = \begin{bmatrix} a_1 & b_1 \\ \alpha a_2 & \alpha b_2 \\ \beta a_3 & \beta b_3 \\ a_4 & b_4 \end{bmatrix},$$

we see that the effect of multiplying A on the left by D is to multiply the second row of A by α and the third row by β.

Example

Consider the matrix $A = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$ obtained from the 3×3 identity matrix by adding Row 3 to Row 1. Set $B = \begin{bmatrix} a & b \\ c & d \\ e & f \end{bmatrix}$. Then

$$AB = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} a & b \\ c & d \\ e & f \end{bmatrix} = \begin{bmatrix} a+e & b+f \\ c & d \\ e & f \end{bmatrix}.$$

We see the effect of multiplying B by A is the add Row 3 of B to Row 1 of A.

Definition 3.38. *By an* ***elementary matrix*** *of size $n \times n$ we mean a matrix that is obtained from the identity matrix I_n by applying to it a single elementary row operation.*

In what follows we will use the notation ρ_i to mean "row i".

Example 3.39. *The following are examples of 3×3 elementary matrices:*

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} \ \rho_2 \leftrightarrow \rho_3; \qquad \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{bmatrix} \ 2\rho_2; \qquad \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \ \rho_1 + \rho_3$$

Definition 3.40. *In a product AB we shall say that B is* ***pre-multiplied*** *by A or, equivalently, that A is* ***post-multiplied*** *by B.*

Let us return to the system of equations described in matrix form by $Ax = b$. It is clear that when we perform a basic operation on the equations all we do is to perform an elementary row operation on the augmented matrix $A \mid b$. It therefore follows from the theorems we had so far, that performing a basic operation on the equations is the same as changing the system $Ax = b$ to the system $EAx = Eb$ where E is some elementary matrix. Moreover, the system of equations that corresponds to the matrix equation $EAx = Eb$ is equivalent to the original system in the sense that it has the same set of solutions (if any).

Proceeding in this way, we see that to every string of k basic operations there corresponds a string of elementary matrices $E_1, \dots, E_k$ such that the the resulting system is represented by

$$E_k \dots E_2 E_1 Ax = E_k \dots E_2 E_1 b,$$

which is of the form $Bx = c$ and is equivalent to the original system.

Now the whole idea of applying matrices to solve systems of linear equations is to obtain a simple systematic method of determining a convenient final matrix B so that the solutions (if any) of the system $Bx = c$ can be found easily, such solutions being precisely the solutions of the original system $Ax = b$.

Our objective is to develop such a method. We have to insist that the method

- will avoid having to write down explicitly the elementary matrices involved at each stage;
- will determine automatically whether or not a solution exists;
- will provide all the solutions.

In this connection, there are two main problems that we have to deal with, namely

1. what form should the matrix B have?;
2. can our method be designed to remove all equations that may be superfluous?

We begin by considering the following type of matrix.

Definition 3.41. *A* ***row-echelon*** *(or* ***stairstep****) matrix is a matrix of the general form in which all the entries "under the stairstep" are zero, all the "corner entries" (marked with a $\star$) are non-zero, and all other entries are arbitrary.*

$$\begin{bmatrix} 0 & \dots 0 & \star & & & & & \\ & & & \star & & & & \\ & & & & \star & & & \\ & & & & & & \star & \\ & & & & & & & \ddots \end{bmatrix}$$

Note that the "stairstep" descends one row at a time and that a "step" may traverse several columns.

Example 3.42. *Every diagonal matrix in which the diagonal entries are non-zero is a row-echelon matrix.*

Theorem 3.43. *By means of elementary row operations, a non-zero matrix can be transformed to a row-echelon matrix.*

Proof. The proof is omitted. □

The process that transforms an arbitrary matrix into a row-echelon matrix in known as **Gaussian elimination**.

Example 3.44.

$$\begin{bmatrix} 0 & 0 & 2 \\ 1 & -1 & 1 \\ -1 & 1 & -4 \end{bmatrix} \underset{\rho_1 \leftrightarrow \rho_2}{\sim} \begin{bmatrix} 1 & -1 & 1 \\ 0 & 0 & 2 \\ -1 & 1 & -4 \end{bmatrix} \underset{\rho_3 + \rho_1}{\sim} \begin{bmatrix} 1 & -1 & 1 \\ 0 & 0 & 2 \\ 0 & 0 & -3 \end{bmatrix} \underset{\rho_3 + \frac{3}{2}\rho_2}{\sim} \begin{bmatrix} 1 & -1 & 1 \\ 0 & 0 & 2 \\ 0 & 0 & 0 \end{bmatrix}$$

Note that the stairstep does not necessarily reach the bottom row (as the above case displays).

Definition 3.45. *An* ***Hermite*** *(or* ***reduced row-echelon****) matrix is a row-echelon matrix in which every corner entry is* 1 *and every entry lying above a corner entry is* 0*.*

$$\begin{bmatrix} 0 & \dots 0 & 1 & 0 & 0 & 0 \\ & & & 1 & 0 & 0 \\ & & & & 1 & 0 \\ & & & & & 1 \\ & & & & & \end{bmatrix}$$

Theorem 3.46. *Every non-zero matrix A can be transformed in an Hermite matrix by means of elementary row operations.*

Proof. Let Z be a row-echelon matrix obtained from A by the Gaussian elimination process. Divide each non-zero row of Z by the (non-zero) corner entry in that row. This makes each of the corner entries 1. Now subtract suitable multiples of every such non-zero row from every row above it to obtain an Hermite matrix. □

The systematic procedure that is described in the proof of the above theorem is best illustrated by an example.

Example 3.47.

$$\begin{bmatrix} 1 & 2 & 1 & 2 & 1 \\ 2 & 4 & 4 & 8 & 4 \\ 3 & 6 & 5 & 7 & 7 \end{bmatrix} \begin{matrix} \sim \\ \rho_2 - 2\rho_1 \\ \rho_3 - 3\rho_1 \end{matrix} \begin{bmatrix} 1 & 2 & 1 & 2 & 1 \\ 0 & 0 & 2 & 4 & 2 \\ 0 & 0 & 2 & 1 & 4 \end{bmatrix} \begin{matrix} \\ \sim \\ \rho_3 - \rho_2 \end{matrix} \begin{bmatrix} 1 & 2 & 1 & 2 & 1 \\ 0 & 0 & 2 & 4 & 2 \\ 0 & 0 & 0 & -3 & 2 \end{bmatrix}$$

We now have the row-echelon form and proceed with the procedure to obtain the Hermite matrix, first by multiplying the second row by $\frac{1}{2}$ and the third row by $-\frac{1}{3}$.

$$\begin{bmatrix} 1 & 2 & 1 & 2 & 1 \\ 0 & 0 & 2 & 4 & 2 \\ 0 & 0 & 0 & -3 & 2 \end{bmatrix} \begin{matrix} \sim \\ \cdot\frac{1}{2} \\ \cdot(-\frac{1}{3}) \end{matrix} \begin{bmatrix} 1 & 2 & 1 & 2 & 1 \\ 0 & 0 & 1 & 2 & 1 \\ 0 & 0 & 0 & 1 & -\frac{2}{3} \end{bmatrix} \begin{matrix} \rho_1 - \rho_2 \\ \rho_2 - 2\rho_3 \\ \sim \end{matrix} \begin{bmatrix} 1 & 2 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & \frac{7}{3} \\ 0 & 0 & 0 & 1 & -\frac{2}{3} \end{bmatrix}.$$

Note that we didn't talk about *the* row-echelon form of a matrix. In fact, there is no unique row-echelon form. To see this, it suffices to observe that we can begin the process of reduction to row-echelon form by moving any non-zero row to become the first row. However, we can talk of *the* Hermite form since, as we will see, this is unique. In fact, it is precisely because of this that such matrices are the focus of our attention. As far as the problem in hand is concerned, namely the solution of $Ax = b$, we can reveal that the Hermite form of A is precisely the matrix that will satisfy the requirements we have listed before. In order to establish these facts, however, we must develop some new ideas. For this purpose, we introduce the following notation.

Given a matrix $A = (a_{ij})_{m\times n}$ we will denote by $A_i = [a_{i1}\ a_{i2} \dots a_{in}]$ the i-th row of A and by $a_i = \begin{bmatrix} a_{1i} \\ a_{2i} \\ \vdots \\ a_{mi} \end{bmatrix}$ the i-th column of A.

Definition 3.48. *A* ***linear combination*** *of the rows (columns) of A is an expression of the form*

$$\lambda_1 x_1 + \lambda_2 x_2 + \cdots + \lambda_p x_p$$

where each x_i is a row (column) of A.

Example 3.49. *The row matrix $\begin{bmatrix} 2 & -3 & 1 \end{bmatrix}$ can be written in the form*

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

and so is a linear combination of the rows of I_3.

The column matrix

$$\begin{bmatrix} 0 \\ 4 \\ -2 \end{bmatrix}$$

can be written as

$$4\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} - 2\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

and so is a linear combination of the columns of I_3.

We recall the following from earlier in the course.

Definition 3.50. *If $x_1, \dots, x_p$ are rows (columns) of A then we say that $x_1, \dots, x_p$ are* ***linearly independent*** *if the only scalars $\lambda_1, \dots, \lambda_p$ which are such that*

$$\lambda_1 x_1 + \lambda_2 x_2 + \cdots + \lambda_p x_p = 0$$

are $\lambda_1 = \lambda_2 = \cdots = \lambda_p = 0$. If $x_1, \dots, x_p$ are not linearly independent then they are ***linearly dependent****.*

Expressed in an equivalent way, the rows (columns) $x_1, \dots, x_p$ are linearly independent if the only way that the zero row (column) can be expressed as a linear combination of $x_1, \dots, x_p$ is the trivial way, namely

$$0 = 0x_1 + 0x_2 + \cdots + 0x_p.$$

Theorem 3.51. *If the rows/columns $x_1, \dots, x_p$ are linearly dependent then none can be zero.*

Proof. If we had $x_i = 0$ then we could write

$$0x_1 + \cdots + 0x_{i-1} + 1x_i + 0x_{i+1} + \cdots + 0x_p = 0,$$

so that $x_1, \dots, x_p$ would not be independent. □

Theorem 3.52. *The following statements are equivalent:*

1. *$x_1, \dots, x_p$ $(p \geq 2)$ are linearly independent;*
2. *one of the x_i can be expressed as a linear combination of the others.*

Proof. (1) ⇒ (2) : Suppose that $x_1, \dots, x_p$ are linearly dependent, where $p \geq 2$. Then there exists $\lambda_1, \dots, \lambda_p$ such that

$$\lambda_1 x_1 + \cdots + \lambda_p x_p = 0$$

with at least one of the λ_i not zero. Suppose that $\lambda_k \neq 0$. Then the above equation can be rewritten in the form

$$x_k = -\frac{\lambda_1}{\lambda_k} x_1 - \cdots - \frac{\lambda_p}{\lambda_k} x_p,$$

i.e., x_k is a linear combination of $x_1, \dots, x_{k-1}, x_{k+1}, \dots, x_p$.

(2) ⇒ (1) : Suppose that

$$x_k = \alpha_1 x_1 + \cdots + \alpha_{k-1} x_{k-1} + \alpha_{k+1} x_{k+1} + \cdots + \alpha_p x_p.$$

Then this can be written in the form

$$x_k = \alpha_1 x_1 + \cdots + \alpha_{k-1} x_{k-1} + (-1) x_k + \alpha_{k+1} x_{k+1} + \cdots + \alpha_p x_p = 0$$

where the left-hand side is a non-trivial linear combination of $x_1, \dots, x_p$. Thus $x_1, \dots, x_p$ are linearly dependent. □

Corollary 3.53. *The rows of a matrix are linearly dependent if and only if one can be obtained from the others by means of elementary row operations.*

Example 3.54. *The rows of the matrix*

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

are linearly dependent. This follows from the fact that $A_3 = A_1 + 2A_2$.

We now consider the following important notion.

Definition 3.55. *The **row rank** of a matrix is the maximum number of linearly independent rows in the matrix.*

It turns out that the row rank of the augmented matrix $A \mid b$ of the system $Ax = b$ determines precisely how many of the equations in the system are not superfluous, so it is important to have a simple method of determining the row rank of a matrix. The next result provides the key to obtaining such a method.

Theorem 3.56. *Elementary row operations do not affect row rank.*

Proof. It is clear that the interchange of two rows has no effect on the maximum number of independent rows, i.e. the row rank. If the k-th row A_k is a linear combination of p other rows, which by the above we may assume to be the rows $A_1, \ldots, A_p$, then clearly so is λA_k for every non-zero scalar λ. It follows by Theorem 3.52 that multiplying a row by a non-zero scalar has no effect on the row rank. Finally, suppose that we add the i-th row to the j-th row to obtain a new j-th row, say $A'_j = A_i + A_j$. Since then

$$\lambda_1 A_1 + \cdots + \lambda_i A_i + \cdots + \lambda_j A'_j + \cdots + \lambda_p A_p = \lambda_1 A_1 + \cdots + (\lambda_i + \lambda_j) A_i + \cdots + \lambda_j A_j + \cdots + \lambda_p A_p$$

it is clear that if $A_1, \ldots, A_i, \ldots, A_j, \ldots, A_p$ are linearly independent then so also are $A_1, \ldots, A_i, \ldots, A'_j, \ldots, A_p$. Thus the addition of one row to another has no effect on row rank. □

Definition 3.57. *A matrix B is said to be **row-equivalent** to matrix A if B can be obtained from A by means of a finite sequence of elementary row operations.*

Since row operations are reversible, we have that if an $m \times n$ matrix B is row-equivalent to the $m \times n$ matrix A then A is row-equivalent to B. The relation of being row-equivalent is therefore a symmetric relation on the set of $m \times n$ matrices. It is trivially reflexive; and it is transitive since if F and G are each products of elementary matrices then clearly so is FG. So if $B = GA$ is row-equivalent to A and B is row-equivalent to $C = FB$ then $C = FGA$ is row-equivalent to A. Thus the relation of being row equivalent is an equivalence relation on the set of $m \times n$ matrices.

Theorem 3.58. *Row-equivalent matrices have the same row rank.*

We can now establish the relation we were aiming for.

Theorem 3.59. *Every non-zero matrix can be reduced by means of elementary row operations to a unique Hermite matrix.*

Proof. The proof is omitted. □

Corollary 3.60. *The row rank of a matrix is the number of non-zero rows in any row-echelon form of the matrix.*

Proof. Let B be a row-echelon form of A and let H be the Hermite form obtained from B. Since H is unique, the number of non-zero rows of B is precisely the number of non-zero rows of H, which is the row rank of A. □

The uniqueness of the Hermite form means that two given matrices are row-equivalent if and only if they have the same Hermite form. The Hermite form of A is therefore a particularly important "representative" in the equivalence class of A relative to the relation of being row-equivalent.

Similar to the concept of an elementary row operation is that of an **elementary column operation**. To obtain this we simply replace *row* by *column* in the definition. It should be noted immediately that such column operations cannot be used in the same way as row operations to solve systems of linear equations since they do not produce an equivalent system. However, there are results concerning column operations that are "dual" to those concerning row operations. This is because column operations on a matrix can be regarded as row operations on the transpose of the matrix.

Theorem 3.61. *An elementary column operation on an $m \times n$ matrix can be achieved by post-multiplication by a suitable elementary matrix, namely that obtained from I_n by applying to I_n precisely the same column operation.*

The notion of column-equivalence is dual to that of row-equivalence.

Definition 3.62. *The **column rank** of a matrix is defined to be the maximum number of linearly independent columns in the matrix.*

Theorem 3.63. *Column operations do not affect column rank.*

Since it is clear that column operations can have no effect on the independence of rows, it follows that column operations have no effect on row rank. We can therefore assert:

Theorem 3.64. *Row and column rank are invariant under both row and column operations.*

We now ask if there is any connection between the row rank and the column rank of a matrix; i.e. if the maximum number of linearly independent rows is connected in any way with the maximum number of linearly independent columns. The answer is yes.

Theorem 3.65. *Row rank and column rank are the same.*

Proof. Given a non-zero $m \times n$ matrix A, let $H(A)$ be its Hermite form. Since $H(A)$ is obtained from A by row operations it has the same row rank, p say, as A. Also, we can apply column operations to $H(A)$ without changing this row rank. Also, both A and $H(A)$ have the same column rank, since row operations do not affect column rank.

Now by suitable rearrangement of its columns $H(A)$ can be transformed into the general form

$$A' = \begin{bmatrix} I_p & ? \\ 0_{m-p,p} & 0_{m-p,n-p} \end{bmatrix},$$

in which the submatrix marked ? is unknown but can be reduced to 0 by further column operations using the first p columns. Thus $H(A)$ can be transformed by column operations into the matrix

$$A'' = \begin{bmatrix} I_p & 0 \\ 0 & 0 \end{bmatrix}.$$

Now by its construction this matrix has the same row rank and the same column rank as A. But clearly the row rank and the column rank of this matrix are each p. It therefore follows that the row rank and the column rank of A are the same. □

Because of that last Theorem we can talk simply of the **rank** of a matrix, meaning by this the row rank or the column rank, whichever is appropriate. The following is now immediate:

Corollary 3.66. $r(A) = r(A^T)$.

Definition A matrix equation $A\overrightarrow{x} = \overrightarrow{b}$ is called **homogenous** if $\overrightarrow{b} = 0$.

A homogenous system $A\overrightarrow{x} = \overrightarrow{0}$ always has $\overrightarrow{x} = \overrightarrow{0}$ as a solution. This solution is called **trivial** with all nonzero solutions being **nontrivial**.

We now have in hand enough machinery to solve the problem we were dealing with.

Theorem 3.67. *If A is an $m \times n$ matrix then the homogeneous system of equations $Ax = 0$ has a non-trivial solution if and only if $r(A) < n$.*

Proof. Let a_i be the i-th column of A. Then there is a non-zero column matrix $x = [x_1\ x_2\ \dots\ x_n]^T$ such that $Ax = 0$ if and only if

$$x_1 a_1 + x_2 a_2 + \cdots + x_n a_n = 0;$$

for, as is readily seen, the left-hand side of this equation is simply Ax. Hence a non-trivial (=non-zero) solution x exists if and only if the columns of A are linearly dependent. Since A has n columns in all, this is the case if and only if the (column) rank of A is less than n. □

Theorem 3.68. *A non-homogeneous system $Ax = b$ has a solution if and only if $r(A) = r(A \mid b)$.*

Proof. If A is of size $m \times n$ then there is an $n \times 1$ matrix x such that $Ax = b$ if and only if there are scalars $x_1, \dots, x_n$ such that

$$x_1 a_1 + x_2 a_2 + \cdots + x_n a_n = b.$$

This is the case if and only if b is linearly dependent on the columns of A, which is the case if and only if the augmented matrix $A \mid b$ is column-equivalent to A, i.e. has the same (column) rank as A. □

Definition 3.69. *A system of linear equations is **consistent** if it has a solution (which, in the homogeneous case, is non-trivial); otherwise it is* inconsistent.

Theorem 3.70. *Let a consistent system of linear equations have as coefficient matrix the $m \times n$ matrix A. If $r(A) = p$ then $n-p$ of the unknowns can be assigned arbitrarily and the equations can be solved in terms of them as parameters.*

Proof. Working with the augmented matrix $A \mid b$, or simply with A in the homogeneous case, perform row operations to transform A to Hermite form. We thus obtain a matrix of the form $H(A) \mid c$ in which, if the rank of A is p, there are p non-zero rows. The corresponding system of equations $H(A)x = c$ is equivalent to the original system, and its form allows us to assign $n-p$ of the unknowns as solution parameters. □

The final statement in the above proof depends on the form of $H(A)$.

We will now put the last three theorems in practice, in particular, we will (try to) solve (with matrices) the three examples given at the beginning of the section.

Example 3.71.

$$\begin{array}{rcrcrcr} & & y & + & 2z & = & 1 \\ x & - & 2y & + & z & = & 0 \\ & & 3y & - & 4z & = & 23 \end{array}$$

Since we already solved this system of equations, we already know that $x = 12$, $y = 5$ and $z = -2$. Let's see how the solution can be found using matrices.

We begin with the augmented matrix of the system and our goal is to get the Hermitian matrix.

$$\left[\begin{array}{rrr|r} 0 & 1 & 2 & 1 \\ 1 & -2 & 1 & 0 \\ 0 & 3 & -4 & 23 \end{array}\right] \underset{\sim}{\rho_1 \leftrightarrow \rho_2} \left[\begin{array}{rrr|r} 1 & -2 & 1 & 0 \\ 0 & 1 & 2 & 1 \\ 0 & 3 & -4 & 23 \end{array}\right] \underset{\rho_3 - 3\rho_2}{\sim} \left[\begin{array}{rrr|r} 1 & -2 & 1 & 0 \\ 0 & 1 & 2 & 1 \\ 0 & 0 & -10 & 20 \end{array}\right]$$

At this point we can see that $r(A) = 3 = r(A \mid b)$, therefore the given system has a unique solution. We can now proceed to find the solution by further transforming our matrix, first by multiplying the last row by $-\frac{1}{10}$, to obtain the corner 1's, and then:

$$\left[\begin{array}{rrr|r} 1 & -2 & 1 & 0 \\ 0 & 1 & 2 & 1 \\ 0 & 0 & 1 & -2 \end{array}\right] \begin{array}{c} \rho_1 + 2\rho_2 \\ \rho_2 - 2\rho_1 \\ \sim \end{array} \left[\begin{array}{rrr|r} 1 & 0 & 5 & 2 \\ 0 & 1 & 0 & 5 \\ 0 & 0 & 1 & -2 \end{array}\right] \underset{\sim}{\rho_1 - 5\rho_3} \left[\begin{array}{rrr|r} 1 & 0 & 0 & 12 \\ 0 & 1 & 0 & 5 \\ 0 & 0 & 1 & -2 \end{array}\right]$$

Thus we obtained a form from which we can directly read off the solution of our system. In particular:

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 12 \\ 5 \\ -2 \end{bmatrix}$$

directly implies that $x = 12$, $y = 5$ and $z = -2$.

Example 3.72.

$$\begin{array}{rcrcrcr} x & - & 2y & - & 4z & = & 0 \\ -2x & + & 4y & + & 3z & = & 1 \\ -x & + & 2y & - & z & = & 1. \end{array}$$

Was the example where we end up having only two equations in three unknowns. Let's put in matrix form, and see what happens.

$$\left[\begin{array}{rrr|r} 1 & -2 & -4 & 0 \\ -2 & 4 & 3 & 1 \\ -1 & 2 & -1 & 1 \end{array}\right] \underset{\substack{\rho_2+2\rho_1 \\ \rho_3+\rho_1}}{\sim} \left[\begin{array}{rrr|r} 1 & -2 & -4 & 0 \\ 0 & 0 & -5 & 1 \\ 0 & 0 & -5 & 1 \end{array}\right] \underset{\rho_3-\rho_2}{\sim} \left[\begin{array}{rrr|r} 1 & -2 & -4 & 0 \\ 0 & 0 & -5 & 1 \\ 0 & 0 & 0 & 0 \end{array}\right]$$

As we can see $r(A) = r(A \mid b)$, so there exists a solution. Since $r(A) = 2 < 3 = n$ the solution will be 1-parametric.

$$\left[\begin{array}{rrr|r} 1 & -2 & -4 & 0 \\ 0 & 0 & -5 & 1 \\ 0 & 0 & 0 & 0 \end{array}\right] \underset{\cdot(-1/5)}{\sim} \left[\begin{array}{rrr|r} 1 & -2 & -4 & 0 \\ 0 & 0 & 1 & -1/5 \\ 0 & 0 & 0 & 0 \end{array}\right] \overset{\rho_1+4\rho_2}{\sim} \left[\begin{array}{rrr|r} 1 & -2 & 0 & -4/5 \\ 0 & 0 & 1 & -1/5 \\ 0 & 0 & 0 & 0 \end{array}\right]$$

Therefore from the above we get that $z = -\frac{1}{5}$ and that $x - 2y = -\frac{4}{5}$ and thus that $x = -\frac{4}{5} + 2y$. We can see that the solution is not unique and that it depends on our choice of the parameter y.

Example 3.73. *The last given example was the system with no feasible solution.*

$$\begin{array}{rcrcrcrcr} x & + & y & + & z & + & t & = & 1 \\ x & - & y & - & z & + & t & = & 3 \\ -x & - & y & + & z & - & t & = & 1 \\ -3x & + & y & - & 3z & - & 3t & = & 4. \end{array}$$

Using Gaussian elimination on the augmented matrix of the system we get that:

$$\left[\begin{array}{rrrr|r} 1 & 1 & 1 & 1 & 1 \\ 1 & -1 & -1 & 1 & 3 \\ -1 & -1 & 1 & -1 & 1 \\ -3 & 1 & -3 & -3 & 4 \end{array}\right] \underset{\substack{\rho_2-\rho_1 \\ \rho_3+\rho_1 \\ \rho_4+3\rho_1}}{\sim} \left[\begin{array}{rrrr|r} 1 & 1 & 1 & 1 & 1 \\ 0 & -2 & -2 & 0 & 2 \\ 0 & 0 & 2 & 0 & 2 \\ 0 & 4 & 0 & 0 & 7 \end{array}\right] \underset{\substack{\cdot -1/2\rho_2 \\ \cdot 1/2\rho_3}}{\sim} \left[\begin{array}{rrrr|r} 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 0 & -1 \\ 0 & 0 & 1 & 0 & 1 \\ 0 & 4 & 0 & 0 & 7 \end{array}\right]$$

$$\underset{\rho_4-4\rho_2}{\sim} \left[\begin{array}{rrrr|r} 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 0 & -1 \\ 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & -4 & 0 & 11 \end{array}\right] \underset{\rho_4+4\rho_3}{\sim} \left[\begin{array}{rrrr|r} 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 0 & -1 \\ 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 15 \end{array}\right]$$

Since in this case $r(A) = 3 \neq r(A \mid b) = 2$ the system has no solution.

Example 3.74. *Determine for which values of $\alpha \in \mathbb{R}$ the system*

$$\begin{array}{rcrcrcr} x & + & y & + & z & = & 1 \\ 2x & - & y & + & 2z & = & 1 \\ x & + & 2y & + & z & = & \alpha \end{array}$$

has a solution.

Solution. Since the system will have a solution if and only if $r(A) = r(A \mid b)$ we proceed by applying Gaussian elimination to the augmented matrix $A \mid b$.

$$\left[\begin{array}{rrr|r} 1 & 1 & 1 & 1 \\ 2 & -1 & 2 & 1 \\ 1 & 2 & 1 & \alpha \end{array}\right] \underset{\substack{\rho_2-2\rho_1 \\ \rho_3-\rho_1}}{\sim} \left[\begin{array}{rrr|c} 1 & 1 & 1 & 1 \\ 0 & -3 & 0 & -1 \\ 0 & 1 & 0 & \alpha-1 \end{array}\right] \underset{\rho_3+\frac{1}{3}\rho_2}{\sim} \left[\begin{array}{rrr|c} 1 & 1 & 1 & 1 \\ 0 & -3 & 0 & -1 \\ 0 & 0 & 0 & \alpha-\frac{4}{3} \end{array}\right]$$

It is clear that the ranks are the same (and hence a solution exists) if and only if $\alpha = \frac{4}{3}$. In this case the rank is 2 and the solution is 1-parametric.

3.4 Invertible matrices

We showed within *the algebra of matrices* section that every matrix A has an additive inverse, denoted by $-A$, which is the unique $m \times n$ matrix X that satisfies the equation $A + X = 0$. We also gave an example showing that the multiplicative inverse does not need to exist. In this section we will explore when the multiplicative inverse exists.

Definition 3.75. *Let A be an $m \times n$ matrix. Then an $n \times m$ matrix X is said to be a* ***left inverse*** *of A if it satisfies the equation $XA = I_n$; and a* ***right inverse*** *of A if it satisfies the equation $AX = I_m$.*

Example 3.76. *Consider the matrices*

$$A = \begin{bmatrix} 1 & 1 \\ 4 & 3 \\ 3 & 4 \\ 0 & 0 \end{bmatrix}, \quad X = \begin{bmatrix} -3 & 1 & 0 & a \\ -3 & 0 & 1 & b \end{bmatrix}.$$

A simple computation shows that $XA = I_2$, and so A has infinitely many left inverses (one for every choice of a, b). In contrast, A has no right inverse. To see this, it suffices to observe that if Y were a 2×4 matrix such that $AY = I_4$ then we would require $[AY]_{4,4} = 1$ which is not possible since all the entries in the fourth row of A are 0.

Example 3.77. *The matrix*

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

has a common unique left inverse and a unique right inverse, namely

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

Theorem 3.78. *Let A be an $m \times n$ matrix. Then*

1. *A has a right inverse if and only if $r(A) = m$;*
2. *A has a left inverse if and only if $r(A) = n$.*

Proof. 1. Suppose that the $n \times m$ matrix X is a right inverse of A, so that we have $AX = I_m$. If x_i denotes the i-th column of X then this equation can be expanded to the m equations

$$Ax_i = \Delta_i,$$

for $i = 1, \ldots, m$, where Δ_i denotes the i-th column of I_m.

Now each of the matrix equations $Ax_i = \Delta_i$ represents a consistent system of m equations in n unknowns and so, for each i we have $r(A) = r(A \mid \Delta_i)$ by Theorem 3.68.

Since $\Delta_1, \ldots, \Delta_m$ are linearly independent, it follows by considering column ranks that

$$\begin{aligned} r(A) &= r(A \mid \Delta_1) \\ &= r(A \mid \Delta_1 \mid \Delta_2) \\ &= \ldots \\ &= r(A \mid \Delta_1 \mid \Delta_2 \mid \cdots \mid \Delta_m) = r(A \mid I_m) = m. \end{aligned}$$

For the last equality, as $\Delta_1, \ldots, \Delta_m$ are linearly independent, we have $r(A \mid I_m) \geq m$ as $r(I_m) = m$. Also, every column of A is a linear combination of the columns of I_m, so $r(A \mid I_m) = m$.

Conversely, suppose that the rank of A is m. Then necessarily we have that $n \geq m$. Consider the Hermite form of A^T. Since $H(A^T)$ is an $n \times m$ matrix and

$$r(H(A^T)) = r(A^T) = r(A) = m,$$

we see that $H(A^T)$ is of the form

$$H(A^T) = \begin{bmatrix} I_m \\ 0_{n-m,m} \end{bmatrix}.$$

As this is row-equivalent to A^T, there exists an $n \times n$ matrix Y, which is a product of an appropriate number of elementary matrices which correspond to the elementary row operations used to find the reduced echelon form of A_T, such that

$$YA^T = \begin{bmatrix} I_m \\ 0_{n-m,m} \end{bmatrix}.$$

Taking transposes, we obtain

$$AY^T = \begin{bmatrix} I_m & 0_{m,n-m} \end{bmatrix}.$$

Now let Z be the $n \times m$ matrix consisting of the first m columns of Y^T. Then from the form of the immediately preceding equation we see that $AZ = I_m$, whence Z is a right inverse of A.

2. Since it is clear that A has a left inverse if and only if its transpose has a right inverse, the result follows by applying (1) to the transpose of A.

□

Theorem 3.79. *If a matrix A has both a left inverse X and a right inverse Y then necessarily*

1. *A is square;*

2. *$X = Y$.*

Proof. 1. Suppose that A is of size $m \times n$. Then, by Theorem 3.78, the existence of a right inverse forces $r(A) = m$, and the existence of a left inverse forces $r(A) = n$. Hence $m = n$ and so A is square.

2. If A is of size $p \times p$ then $XA = I_p = AY$ gives, by the associativity of matrix multiplication,

$$X = XI_p = X(AY) = (XA)Y = I_pY = Y.$$

□

For square matrices we have the following stronger result.

Theorem 3.80. *If A is an $n \times n$ matrix then the following statements are equivalent:*

1. *A has a left inverse;*
2. *A has a right inverse;*
3. *$r(A) = n$;*
4. *the Hermite form of A is I_n;*
5. *A is a product of elementary matrices.*

Proof. We first establish the equivalence of $(1), (2), (3), (4)$. That $(1), (2), (3)$ are equivalent is immediate from Theorem 3.78.

$(3) \Rightarrow (4)$: If A is of rank n then the Hermite form of A must have n non-zero rows, hence n corner entries 1. The only possibility is I_n.

$(4) \Rightarrow (3)$: This is clear from the fact that $r(I_n) = n$.

We complete the proof by showing that $(3) \Leftrightarrow (5)$.

$(3) \Rightarrow (5)$: If A is of rank n then, since $(3) \Rightarrow (1)$, A has a left inverse X. Since $XA = I_n$ we see that X has a right inverse A so, since $(2) \Rightarrow (4)$, there is a finite string of elementary matrices $F_1, F_2, \ldots, F_q$ such that

$$F_q \ldots F_2F_1X = I_n.$$

Consequently we have

$$A = I_nA = (F_q \ldots F_2F_1X)A = F_q \ldots F_2F_1(XA) = F_q \ldots F_2F_1$$

and so A is a product of elementary matrices.

$(5) \Rightarrow (3)$: Suppose now that $A = E_1E_2 \ldots E_p$ where each E_i is an elementary matrix. Observe that E_p is of rank n since it is obtained from I_n by a single elementary operation which has no effect on rank. Also, pre-multiplication by an elementary matrix is equivalent to an elementary row operation, which has no effect on rank. It follows that the rank of the product $E_1E_2 \ldots E_p$ is the same as the rank of E_p, which is n. Thus the rank of A is n. □

It is immediate from the above important result that if a square matrix A has a one-sided inverse then this is a two-sided inverse (i.e. both a left inverse and a right inverse). In what follows we will always use the word *inverse* to mean two-sided inverse. By a Theorem 3.22, inverses that exist are unique. When it exists, we denote the unique inverse of the square matrix A by A^{-1}.

Definition 3.81. *If A has an inverse then we say that A is* ***invertible***.

If A is an invertible $n \times n$ matrix then so is A^{-1}. In fact, since $AA^{-1} = I_n = A^{-1}A$ and inverses are unique, we have that A is an inverse of A^{-1} and so $(A^{-1})^{-1} = A$.

We note at this point that since, by Theorem 3.80, every product of elementary matrices is invertible, we can assert that B is row-equivalent to A if and only if there exists an invertible matrix E such that $B = EA$.

Another useful feature of Theorem 3.80 is that it provides a relatively simple method of determining whether or not A has an inverse, and of computing A^{-1} when it does exist. The method consists of reducing A to Hermite form: if this turns out to be I_n then A is invertible; and if the Hermite form is not I_n then A has no inverse.

In practice there is no need to compute the elementary matrices required at each stage. We simply begin with the array $A \mid I_n$ and apply the elementary row operations to the entire array. In this way the process can be described by

$$A \mid I_n \sim E_1A \mid E_1 \sim E_2E_1A \mid E_2E_1 \sim \ldots.$$

At each stage we have an array of the form

$$S \mid Q \equiv E_i \ldots E_2E_1A \mid E_i \ldots E_2E_1$$

in which $QA = S$. If A has an inverse then the Hermite form of A will be I_n and the final configuration will be

$$I_n \mid E_p \ldots E_2E_1$$

so that $E_p \ldots E_2E_1A = I_n$ and consequently $A^{-1} = E_p \ldots E_2E_1$.

Example 3.82. *Find the inverse of*

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

Solution. Applying the procedure we just described, we obtain

$$\left[\begin{array}{ccc|ccc} 1 & 2 & 3 & 1 & 0 & 0 \\ 1 & 3 & 4 & 0 & 1 & 0 \\ 1 & 4 & 4 & 0 & 0 & 1 \end{array}\right] \begin{array}{c} \sim \\ \rho_2 - \rho_1 \\ \rho_3 - \rho_1 \end{array} \left[\begin{array}{ccc|ccc} 1 & 2 & 3 & 1 & 0 & 0 \\ 0 & 1 & 1 & -1 & 1 & 0 \\ 0 & 2 & 1 & -1 & 0 & 1 \end{array}\right] \begin{array}{c} \\ \sim \\ \rho_3 - 2\rho_2 \end{array} \left[\begin{array}{ccc|ccc} 1 & 2 & 3 & 1 & 0 & 0 \\ 0 & 1 & 1 & -1 & 1 & 0 \\ 0 & 0 & -1 & 1 & -2 & 1 \end{array}\right]$$

$$\begin{array}{c} \sim \\ \rho_1 - 2\rho_2 \\ \cdot(-1)\rho_3 \end{array} \left[\begin{array}{ccc|ccc} 1 & 0 & 1 & 3 & -2 & 0 \\ 0 & 1 & 1 & -1 & 1 & 0 \\ 0 & 0 & 1 & -1 & 2 & -1 \end{array}\right] \begin{array}{c} \rho_1 - \rho_3 \\ \rho_2 - \rho_3 \\ \sim \end{array} \left[\begin{array}{ccc|ccc} 1 & 0 & 0 & 4 & -4 & 1 \\ 0 & 1 & 0 & 0 & -1 & 1 \\ 0 & 0 & 1 & -1 & 2 & -1 \end{array}\right]$$

so A has an inverse and

$$A^{-1} = \begin{bmatrix} 4 & -4 & 1 \\ 0 & -1 & 1 \\ -1 & 2 & -1 \end{bmatrix}$$

We will now consider some further results concerning inverses.

First note that if A and B are invertible $n \times n$ matrices then in general $A + B$ is not invertible. This is easily illustrated by taking $A = I_n$ and $B = -I_n$ and observing that the zero $n \times n$ matrix is not invertible. However, as the following result shows, products of invertible matrices are invertible.

Theorem 3.83. *Let A and B be $n \times n$ matrices. If A and B are invertible then so is AB, moreover,*

$$(AB)^{-1} = B^{-1}A^{-1}.$$

Proof. It suffices to observe that

$$ABB^{-1}A^{-1} = AI_nA^{-1} = AA^{-1} = I_n$$

and hence $B^{-1}A^{-1}$ is the right inverse of AB. By Theorem 3.80, AB therefore has an inverse, and $(AB)^{-1} = B^{-1}A^{-1}$. $\square$

Corollary 3.84. *If A is invertible then so is A^m for every positive integer m, moreover,*

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

Proof. The proof is by induction. The result is trivial for $m = 1$. As for the inductive step, suppose that it holds for m. Then by Theorem 3.83 we have

$$(A^{-1})^{m+1} = A^{-1}(A^{-1})^m = A^{-1}(A^m)^{-1} = (A^mA)^{-1} = (A^{m+1})^{-1},$$

whence it holds for $m + 1$. $\square$

Theorem 3.85. *If A is invertible then so is its transpose A^T, moreover,*

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

Proof. By Theorem 3.24 we have

$$I_n = I_n^T = (AA^{-1})^T = (A^{-1})^TA^T$$

and so $(A^{-1})^T$ is a left inverse, hence the inverse, of A^T. $\square$

Definition 3.86. *An $n \times n$ matrix A is said to be* ***orthogonal*** *if $AA^T = I_n = A^TA$.*

By Theorem 3.80, we see that in fact only one of this equalities is necessary. An orthogonal matrix is therefore an invertible matrix whose inverse is its transpose.

Example 3.87. *A 2×2 matrix is orthogonal if and only if it is of one of the forms*

$$\begin{bmatrix} a & b \\ -b & a \end{bmatrix}, \quad \begin{bmatrix} a & b \\ b & -a \end{bmatrix}$$

in which $a^2 + b^2 = 1$.

In fact, suppose that

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

is orthogonal. Then from the equation

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} \cdot \begin{bmatrix} a & c \\ b & d \end{bmatrix} = AA^T = I_2$$

we obtain

$$a^2 + b^2 = 1, \quad ac + bd = 0, \quad c^2 + d^2 = 1.$$

From the equation

$$\begin{bmatrix} a & c \\ b & d \end{bmatrix} \cdot \begin{bmatrix} a & b \\ c & d \end{bmatrix} = A^T A = I_2$$

we obtain

$$a^2 + c^2 = 1, \qquad b^2 + d^2 = 1, \qquad ab + cd = 0.$$

As $ab + cd = 0$, $ab = -cd$ and so $a^2b^2 = c^2d^2$. As $a^2 = 1 - c^2$ and $d^2 = 1 - b^2$, substituting we obtain $(1 - c^2)b^2 = c^2(1 - b^2)$. This implies $b^2 = c^2$ and so $b = \pm c$. An analogous argument gives $a = \pm d$. Finally, if $a = d$ then as $ab + cd = 0$, we see $db + cd = 0$ or $b = -d$. Similarly, if $a = -d$, then $b = d$. Hence the matrix is one of the given form.

3.5 The determinant of a square matrix

Associated with every square matrix is a scalar called the **determinant**. Perhaps the most striking property of the determinant of a matrix is the fact that it tells us if the matrix is invertible. On the other hand, there is obviously a limit to the amount of information about a matrix which can be carried by a single scalar, and this is probably why determinants are considered less important today than, say, a hundred years ago. Nevertheless, there are other objects that are associated with an arbitrary square matrix (eg. the characteristic polynomial, that you will still learn about), that carry a vast amount of information about the matrix.

3.5.1 Permutations and the definition of a determinant

Let $A = (a_{ij})$ be an $n \times n$ matrix. Our first task is to show how to define the determinant of A, which will be written either

$$\det(A)$$

or else in the extended form

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

For $n = 1$ and $n = 2$ the definition is simple enough:

$$|a_{11}| = a_{11} \text{ and } \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = a_{11}a_{22} - a_{12}a_{21}.$$

Example 3.88.

$$\det(-6) = |-6| = -6$$

$$\begin{vmatrix} 2 & -3 \\ 4 & 1 \end{vmatrix} = 2 \cdot 1 - (-3) \cdot 4 = 14$$

Where does the expression $a_{11}a_{22} - a_{12}a_{21}$ come from? The motivation is provided by linear systems. Suppose that we want to solve the linear system

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

for unknowns x_1 and x_2. Eliminate x_2 by subtracting a_{12} times equation 2 from a_{22} times equation 1; in this way we obtain

$$(a_{11}a_{22} - a_{12}a_{21})x_1 = b_1a_{22} - a_{12}b_2.$$

This equation expresses x_1 as the quotient of a pair of 2×2 determinants:

$$x_1 = \frac{\begin{vmatrix} b_1 & a_{12} \\ b_2 & a_{22} \end{vmatrix}}{\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix}},$$

provided, of course, that the denominator does not vanish. There is a similar expression for x_2.

The preceding calculation indicates that 2×2 determinants are likely to be of significance for linear systems. And this is confirmed if we try the same computation for a linear system of three equations in three unknowns. While the resulting solutions are complicated, they do suggest the following definition for $\det(A)$ where $A = (a_{ij})_{3\times3}$:

$$a_{11}a_{22}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32}$$

$$-a_{12}a_{21}a_{33} - a_{13}a_{22}a_{31} - a_{11}a_{23}a_{32}$$

What informations can we obtain from this expression? In the first place it contains six terms, each of which is a product of three entries of A. The second subscripts in each term correspond to the six ways of ordering the integers $1, 2, 3$, namely

$$1, 2, 3 \quad 2, 3, 1 \quad 3, 1, 2 \quad 2, 1, 3 \quad 3, 2, 1 \quad 1, 3, 2.$$

Also each term is a product of three entries of A, while three of the terms have positive signs and three have negative signs. There is something of a pattern here, but how can one tell which terms are to get a plus sign and which are get a minus sign? The answer is given by **permutations**.

Permutations

Definition. Let n be a fixed positive integer. By a **permutation** of the integers $1, 2, \ldots, n$ we will mean an arrangement of these integers in some definite order.

For example, as has been observed before, there are six permutations of the integers $1, 2, 3$. In general, a permutation of $1, 2, \ldots, n$ can be written in the form

$$i_1, i_2, \ldots, i_n$$

where $i_1, i_2, \ldots, i_n$ are the integers $1, 2, \ldots, n$ in some order.

Thus to construct a permutation we have only to choose distinct integers $i_1, i_2, \ldots, i_n$ from the set $\{1, 2, \ldots, n\}$. Clearly there are n choices for i_1; once i_1 has been chosen, it can not be chosen again, so there are just $n-1$ choices for i_2; since i_1 and i_2 can not be chosen again, there are $n-2$ choices for i_3, and so on. There will be only one possible choice for i_n since $n-1$ integers have already been selected. The number of ways of constructing a permutation is therefore equal to the product of these numbers

$$n(n-1)(n-2)\cdots 2\cdot 1,$$

which is written as

$$n!$$

and referred to as "n factorial". Thus we can state the following basic result.

Theorem 3.89. *The number of permutations of the integers* $1, 2, \ldots, n$ *equals*

$$n! = n(n-1)\cdots 2\cdot 1.$$

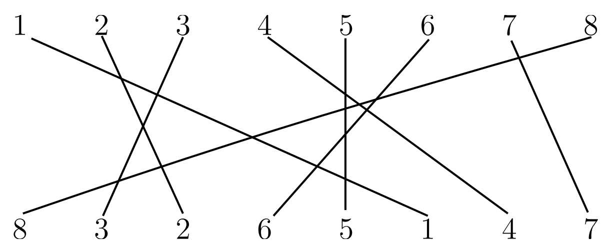
Even and odd permutations

Definition. A permutation of the integers $1, 2, \ldots, n$ is called **even** or **odd** according to whether the number of inversions of the natural order $1, 2, \ldots, n$ that are present in the permutation is even or odd respectively.

For example, the permutation $1, 3, 2$ involves a single inversion, for 3 comes before 2; so this is an odd permutation. For permutations of longer sequences of integers it is advantageous to count inversions by means of what is called a crossover diagram. This is best explained by an example.

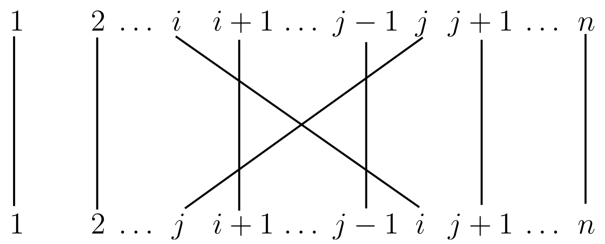
Example 3.90. *Is the permutation* $8, 3, 2, 6, 5, 1, 4, 7$ *even or odd?*

Solution. The procedure is to write the integers 1 through 8 in the natural order in a horizontal line, and then to write down the entries of the permutation in the line below. Join each integer i in the top line to the same integer i where it appears in the bottom line, taking care to avoid multiple intersections. The number of intersections or **crossovers** will be the number of inversions present in the permutation:



Since there are 15 crossovers in the diagram, this permutation is odd.

Definition. A **transposition** is a permutation that is obtained from $1, 2, \ldots, n$ by interchanging just two integers. Thus $2, 1, 3, 4, \ldots, n$ is an example of a transposition. An important fact about transpositions is that they are always odd.



Theorem 3.91. *Transpositions are odd permutations.*

Proof. Consider the transposition which interchanges i and j, with $i < j$ say. The crossover diagram for this transposition is shown below. Each of the $j - i - 1$ integers $i + 1, i + 2, \ldots, j - 1$ gives 2 crossovers, while i and j add one more. Hence the total number of crossovers in the diagram equals $2(j - i - 1) + 1$, which is odd. $\square$

It is important to determine the numbers of even and odd permutations.

Theorem 3.92. *If $n > 1$, there are $\frac{1}{2}(n!)$ even permutations of $1, 2, \ldots, n$ and the same number of odd permutations.*

Proof. If the first two integers are interchanged in a permutation, it is clear from the crossover diagram that an inversion is either added or removed. Thus the operation changes an even permutation to an odd permutation and an odd permutation to an even one. This makes it clear that the numbers of even and odd permutations must be equal. Since the total number of permutations is $n!$, the result follows. $\square$

Example 3.93. *The even permutations of $1, 2, 3$ are*

$$1, 2, 3 \qquad 2, 3, 1 \qquad 3, 1, 2;$$

while the odd permutations are

$$2, 1, 3 \qquad 3, 2, 1 \qquad 1, 3, 2.$$

Definition 3.94. *The **sign** of a permutation $i_1, i_2, \ldots, i_n$, denoted by $sign(i_1, i_2, \ldots, i_n)$ is $+1$ if the permutation is even end -1 if the permutation is odd.*

Example 3.95. $sign(3, 2, 1) = -1$

Permutation matrices

Before proceeding to the formal definition of a determinant, we pause to show how permutations can be represented by matrices.

Definition An $n \times n$ matrix is called a **permutation matrix** if it can be obtained from the identity matrix I_n by rearranging the rows or columns.

For example, the permutation matrix

$$\begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}$$

is obtained from I_3 by cyclically permuting the columns, $C_1 \to C_2 \to C_3 \to C_1$. Permutation matrices are easy to recognize since each row and each column contains a single 1, while all other entries are zero.

Consider a permutation $i_1, i_2, \ldots, i_n$ of $1, 2, \ldots, n$ and let P be the permutation matrix which has (j, i_j) entry equal to 1 for $j = 1, 2, \ldots, n$, and all other entries zero. This means that P is obtained from I_n by rearranging the columns in the manner specified by the permutation $i_1, i_2, \ldots, i_n$, that is, $C_j \to C_{i_j}$. Then, as matrix multiplication shows,

$$P \cdot \begin{bmatrix} 1 \\ 2 \\ \vdots \\ n \end{bmatrix} = \begin{bmatrix} i_1 \\ i_2 \\ \vdots \\ i_n \end{bmatrix}.$$

Thus the effect of a permutation on the order $1, 2, ..., n$ is reproduced by pre-multiplication by the corresponding permutation matrix.

Example 3.96. *The permutation matrix which corresponds to the permutation* $4, 2, 1, 3$ *is obtained from* I_4 *by the column replacements* $C_1 \to C_4$, $C_2 \to C_2$, $C_3 \to C_1$, $C_4 \to C_3$. *It is*

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

and indeed

$$P \cdot \begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \end{bmatrix} = \begin{bmatrix} 4 \\ 2 \\ 1 \\ 3 \end{bmatrix}.$$

Remark 3.97. *The easiest way to compute the permutation matrix (and the standard way to write permutations) is to write the permutation as*

$$\begin{pmatrix} 1 & 2 & 3 & 4 \\ 4 & 2 & 1 & 3 \end{pmatrix} \equiv (143)(2)$$

Definition of a determinant

We are now in a position to define the general $n \times n$ determinant.

Definition. Let $A = (a_{ij})_{n \times n}$ be an $n \times n$ matrix over some field of scalars. Then the determinant of A is the scalar defined by the equation

$$\det(A) = \sum_{i_1, i_2, \ldots, i_n} \operatorname{sign}(i_1, i_2, \ldots, i_n) a_{1i_1} a_{2i_2} \cdots a_{ni_n}$$

where the sum is taken over all permutations $i_1, i_2, \ldots, i_n$ of $1, 2, \ldots, n$.

Thus $\det(A)$ is a sum of $n!$ terms each of which involves a product of n elements of A, one from each row and one from each column. A term has a positive or negative sign according to whether the corresponding permutation is even or odd respectively. One determinant which can be immediately evaluated from the definition is that of I_n:

$$\det(I_n) = 1.$$

This is because only the permutation $1, 2, \dots, n$ contributes a non-zero term to the sum that defines $\det(I_n)$.

If we specialise the above definition to the cases $n = 1, 2, 3$, we obtain the expressions for $\det(A)$ given at the beginning of the section. For example, let $n = 3$; the even and odd permutations were listed in Example 3.93. If we write down the terms of the determinant in the same order, we obtain

$$a_{11}a_{22}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32}$$

$$-a_{12}a_{21}a_{33} - a_{13}a_{22}a_{31} - a_{11}a_{23}a_{32}$$

We could in a similar fashion write down the general 4×4 determinant as a sum of $4! = 24$ terms, 12 with a positive sign and 12 with a negative sign. Of course, it is clear that the definition does not provide a convenient means of computing determinants with large numbers of rows and columns; we will shortly see that much more efficient procedures are available.

Example 3.98. *What term in the expansion of the* 8×8 *determinant* $\det(a_{ij})$ *corresponds to the permutation* $8, 3, 2, 6, 5, 1, 4, 7$*?*

Solution. We saw before that this permutation is odd, so its sign is -1; hence the term is

$$-a_{18}a_{23}a_{32}a_{46}a_{55}a_{61}a_{74}a_{87}.$$

Minors and cofactors

In the theory of determinants certain *subdeterminants* called **minors** prove to be a useful tool. Let $A = (a_{ij})$ be an $n \times n$ matrix. The (i, j) *minor* M_{ij} of A is defined to be the determinant of the submatrix of A that remains when row i and column j of A are deleted. The (i, j) **cofactor** A_{ij} of A is simply the minor with an appropriate sign:

$$A_{ij} = (-1)^{i+j} M_{ij}.$$

Example 3.99. *Let*

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

Then

$$M_{23} = \begin{vmatrix} a_{11} & a_{12} \\ a_{31} & a_{32} \end{vmatrix} = a_{11}a_{32} - a_{12}a_{31}$$

and

$$A_{23} = (-1)^{2+3} M_{23} = a_{12}a_{31} - a_{11}a_{32}.$$

One reason for the introduction of cofactors is that they provide us with methods of calculating determinants called **row expansion and column expansion**. These are a great improvement on the defining sum as a means of computing determinants. The next result tells us how they operate.

Theorem 3.100. *Let $A = (a_{ij})$ be an $n \times n$ matrix. Then*

1. $\det(A) = \Sigma_{k=1}^{n} a_{ik}A_{ik}$, *(expansion by row i);*

2. $\det(A) = \Sigma_{k=1}^{n} a_{kj}A_{kj}$, *(expansion by column j).*

Proof. Is omitted. □

The theorem provides a practical method of computing determinants although for determinants of larger size there are even more efficient methods, as we will see.

Example 3.101. *Compute the determinant*

$$\begin{vmatrix} 1 & 2 & 0 \\ 4 & 2 & -1 \\ 6 & 2 & 2 \end{vmatrix}.$$

Solution. We may, for example, expand by row 1, obtaining

$$1 \cdot (-1)^2 \begin{vmatrix} 2 & -1 \\ 2 & 2 \end{vmatrix} + 2 \cdot (-1)^3 \begin{vmatrix} 4 & -1 \\ 6 & 2 \end{vmatrix} + 0 \cdot (-1)^4 \begin{vmatrix} 4 & 2 \\ 6 & 2 \end{vmatrix}$$

$$= 6 - 28 + 0 = -22.$$

Alternatively, we could expand by column 2:

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

$$= -28 + 4 + 2 = -22.$$

However there is an obvious advantage in expanding by a row or column which contains as many zeros as possible.

The determinant of a triangular matrix can be written down at once, an observation which is used frequently in calculating determinants.

Theorem 3.102. *The determinant of an upper or lower triangular matrix equals the product of the entries on the principal diagonal of the matrix.*

Proof. Suppose that $A = (a_{ij})_{n\times n}$ is, say, upper triangular, and expand $\det(A)$ by column 1. The result is the product of a_{11} and an $(n-1)\times(n-1)$ determinant which is also upper triangular. Repeat the operation until a 1×1 determinant is obtained (or use mathematical induction). □

3.5.2 Basic properties of determinants

We now proceed to develop the theory of determinants, establishing a number of properties which will allow us to compute determinants more efficiently.

Theorem 3.103. *If A is an $n \times n$ matrix, then*

$$\det(A^T) = \det(A).$$

Proof. The proof is by mathematical induction. The statement is certainly true if $n = 1$ since then $A^T = A$. Let $n > 1$ and assume that the theorem is true for all matrices with $n-1$ rows and columns. Expansion by row 1 gives

$$\det(A) = \sum_{j=1}^{n} a_{1j} A_{1j}.$$

Let B denote the matrix A^T. Then $a_{ij} = b_{ji}$. By induction on n, the determinant A_{1j} equals its transpose. But this is just the $(j, 1)$ cofactor B_{j1} of B. Hence $A_{1j} = B_{j1}$ and the above equation becomes

$$\det(A) = \sum_{j=1}^{n} b_{j1} B_{j1}.$$

However the right hand side of this equation is simply the expansion of $\det(B)$ by column 1; thus $\det(A) = \det(B)$. □

A useful feature of this result is that it sometimes enables us to deduce that a property known to hold for the rows of a determinant also holds for the columns.

Theorem 3.104. *A determinant with two equal rows (or two equal columns) is zero.*

Proof. Suppose that the $n \times n$ matrix A has its j-th and k-th rows equal. We have to show that $\det(A) = 0$. Let $i_1, i_2, \ldots, i_n$ be a permutation of $1, 2, \ldots, n$; the corresponding term in the expansion of $\det(A)$ is $\operatorname{sign}(i_1, i_2, \ldots, i_n) a_{1i_1} a_{2i_2} \cdots a_{ni_n}$. Now if we switch i_j and i_k in this product, the sign of the permutation is changed, but the product of the a's remains the same since $a_{ji_k} = a_{ki_k}$ and $a_{ki_j} = a_{ji_j}$. This means that the term under consideration occurs a second time in the defining sum for $\det(A)$, but with the opposite sign. Therefore all terms in the sum cancel and $\det(A)$ equals zero. □

Remark 3.105. *A determinant with a row (or column) where all entries are* 0 *is also zero. The same is true if rows (or columns) are not linearly independent.*

The next result describes the effect on a determinant when applying a row or column operation to the associated matrix.

Theorem 3.106. *Let A be an $n \times n$ matrix.*

1. *If a single row (or column) of a matrix A is multiplied by a scalar c, the resulting matrix has determinant equal to* $c(\det(A))$*.*

2. *If two rows (or columns) of a matrix A are interchanged, the effect is to change the sign of the determinant.*

3. *The determinant of a matrix A is not changed if a multiple of one row (or column) is added to another row (or column).*

Proof.
1. The effect of the operation is to multiply every term in the sum defining $\det(A)$ by c. Therefore the determinant is multiplied by c.

2. Here the effect of the operation is to switch two entries in each permutation of $1, 2, \ldots, n$; we have already seen that this changes the sign of a permutation, so it multiplies the determinant by -1.

3. Suppose that we add c times row j to row k of the matrix: here we assume that $j < k$. If C is the resulting matrix, then $\det(C)$ equals

$$\Sigma \mathrm{sign}(i_1, i_2, \ldots, i_n) a_{1i_1} \cdots a_{ji_j} \cdots (a_{ki_k} + c a_{ji_k}) \cdots a_{ni_n}$$

which is turn equals the sum of

$$\Sigma \mathrm{sign}(i_1, i_2, \ldots, i_n) a_{1i_1} \cdots a_{ji_j} \cdots a_{ki_k} \cdots a_{ni_n}$$

and

$$c \cdot \Sigma \mathrm{sign}(i_1, i_2, \ldots, i_n) a_{1i_1} \cdots a_{ji_j} \cdots a_{ji_k} \cdots a_{ni_n}.$$

Now the first of these sums is simply $\det(A)$, while the second sum is the determinant of a matrix in which rows j and k are identical, so it is zero by one of the previous theorems. Hence $\det(C) = \det(A)$.

□

Now let us see how use of these properties can lighten the task of evaluating a determinant. Let A be an $n \times n$ matrix whose determinant is to be computed. Then elementary row operations can be used as in Gaussian elimination to reduce A to row echelon form B. But B is an upper triangular matrix, say

$$B = \begin{bmatrix} b_{11} & b_{12} & \cdots & b_{1n} \\ 0 & b_{22} & \cdots & b_{2n} \\ \cdot & \cdot & \cdots & \cdot \\ 0 & 0 & \cdots & b_{nn} \end{bmatrix},$$

so by a previous theorem we obtain $\det(B) = b_{11} b_{22} \cdots b_{nn}$. Thus all that has to be done is to keep track of the changes in $\det(A)$ produced by the row operations.

Example 3.107. *Compute the determinant of*

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

Solution. We will calculate the determinant of D in two ways, first using row expansion and then by transforming the matrix in row echelon form.

1. Expansion by the first row (because it has most zero entries):

$$\begin{aligned} \det(A) = 0 - 1 \cdot \begin{vmatrix} 1 & 1 & 1 \\ -2 & 3 & 3 \\ 1 & -2 & -3 \end{vmatrix} + 2 \cdot \begin{vmatrix} 1 & 1 & 1 \\ -2 & -2 & 3 \\ 1 & -2 & -3 \end{vmatrix} - 3 \cdot \begin{vmatrix} 1 & 1 & 1 \\ -2 & -2 & 3 \\ 1 & -2 & -2 \end{vmatrix} \\ = -1 \cdot (-9 + 3 + 4 - 3 - 6 + 6) + 2 \cdot (6 + 3 + 4 + 2 - 6 + 6) \\ - 3 \cdot (4 + 3 + 4 + 2 - 4 + 6) \\ = -1 \cdot (-5) + 2 \cdot 15 - 3 \cdot 15 = 5 + 30 - 45 = -10 \end{aligned}$$

2. Transforming the matrix in row echelon form:

$$\begin{bmatrix} 0 & 1 & 2 & 3 \\ 1 & 1 & 1 & 1 \\ -2 & -2 & 3 & 3 \\ 1 & -2 & -2 & -3 \end{bmatrix} \begin{matrix} \rho_1 \leftrightarrow \rho_2 \\ \\ \sim \\ \\ \end{matrix} \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 3 \\ -2 & -2 & 3 & 3 \\ 1 & -2 & -2 & -3 \end{bmatrix} \begin{matrix} \\ \sim \\ \rho_3 + 2\rho_2 \\ \rho_4 - \rho_1 \end{matrix} \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 3 \\ 0 & 0 & 5 & 5 \\ 0 & -3 & -3 & -4 \end{bmatrix}$$

$$\begin{matrix} \sim \\ \frac{1}{5}\rho_3 \\ \rho_4 + 3\rho_2 \end{matrix} \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 3 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 3 & 5 \end{bmatrix} \begin{matrix} \sim \\ \\ \rho_4 - 3\rho_3 \end{matrix} \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 3 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 2 \end{bmatrix}$$

We had one switch of two rows and one row multiplication by $\frac{1}{5}$. Therefore $\det(A) = -5 \cdot (1 \cdot 1 \cdot 1 \cdot 2) = -10$

3.5.3 Application of determinants

3.5.3.1 Invertible matrices

An important property of the determinant of a square matrix is that it tells us whether the matrix is invertible.

Theorem 3.108. *An $n \times n$ matrix A is invertible if and only if* $\det(A) \neq 0$.

Proof. We know by Theorem 3.80 that A has a left and right inverse (which are necessarily equal by Theorem 3.79) if and only if the Hermite form of A is I_n. The Hermite for of A is I_n if and only if $\det(A)$ is a non-zero multiple of 1 by Theorem 3.106. Thus A is invertible if and only if $\det(A) \neq 0$. □

Corollary 3.109. *A linear system $AX = 0$ with n equations in n unknowns has a non-trivial solution if and only if* $\det(A) = 0$.

Theorem 3.110. *If A and B are any $n \times n$ matrices, then*

$$\det(AB) = \det(A) \cdot \det(B).$$

Corollary 3.111. *Let A and B be $n \times n$ matrices. If $AB = I_n$, then $BA = I_n$, and thus $B = A^{-1}$.*

Proof. For $1 = \det(AB) = \det(A) \cdot \det(B)$, so $\det(A) \neq 0$ and therefore A is invertible. Therefore $BA = A^{-1}(AB)A = A^{-1}I_nA = I_n$. □

Corollary 3.112. *If A is an invertible matrix, then* $\det(A^{-1}) = \frac{1}{\det(A)}$.

Proof. Clearly $1 = \det(I_n) = \det(AA^{-1}) = \det(A) \cdot \det(A^{-1})$ and from here the statement follows. □

We will now see how the determinant can be of help when computing the inverse of a matrix. We start with a definition.

Definition 3.113. *Let $A = (a_{ij})$ be an $n \times n$ matrix. Then the* ***adjoint matrix*** *adj(A) of A is the $n \times n$ matrix whose (i, j) element is the (j, i) cofactor A_{ji} of A.*

Example 3.114. *Find adj(A) if*

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

Solution. We begin by computing all the 9 cofactors:

$$A_{11} = (-1)^2 \begin{vmatrix} -1 & 3 \\ -3 & 4 \end{vmatrix} = 5 \qquad A_{21} = (-1)^3 \begin{vmatrix} 2 & 1 \\ -3 & 4 \end{vmatrix} = -11 \qquad A_{31} = (-1)^4 \begin{vmatrix} 2 & 1 \\ -1 & 3 \end{vmatrix} = 7$$

$$A_{12} = (-1)^3 \begin{vmatrix} 6 & 3 \\ 2 & 4 \end{vmatrix} = -18 \qquad A_{22} = (-1)^4 \begin{vmatrix} 1 & 1 \\ 2 & 4 \end{vmatrix} = 2 \qquad A_{32} = (-1)^5 \begin{vmatrix} 1 & 1 \\ 6 & 3 \end{vmatrix} = 3$$

$$A_{13} = (-1)^4 \begin{vmatrix} 6 & -1 \\ 2 & -3 \end{vmatrix} = -16 \qquad A_{23} = (-1)^5 \begin{vmatrix} 1 & 2 \\ 2 & -3 \end{vmatrix} = 7 \qquad A_{33} = (-1)^6 \begin{vmatrix} 1 & 2 \\ 6 & -1 \end{vmatrix} = -13$$

Therefore

$$\mathrm{adj}(A) = \begin{bmatrix} 5 & -11 & 7 \\ -18 & 2 & 3 \\ -16 & 7 & -13 \end{bmatrix}$$

The significance of the adjoint matrix is made clear by the next two results.

Theorem 3.115. *If A is an $n \times n$ matrix, then*

$$A\mathit{adj}(A) = \det(A) \cdot I_n = \mathit{adj}(A)A.$$

Proof. The (i, j) entry of the matrix product $A\mathrm{adj}(A)$ is

$$\sum_{k=1}^{n} a_{ik}(\mathrm{adj}(A))_{kj} = \sum_{k=1}^{n} a_{ik}A_{jk}.$$

If $i = j$, this is just the expansion of $\det(A)$ by row i; on the other hand, if $i \neq j$, the sum is also a row expansion of a determinant, but one in which rows i and j are identical. We know that if a matrix has two identical rows, the sum will vanish. This means that the off-diagonal entries of the matrix product $A\mathrm{adj}(A)$ are zero, while the entries on the diagonal all equal $\det(A)$. Therefore $A\mathrm{adj}(A)$ is the scalar matrix $(\det(A))I_n$, as claimed. The second statement can be proved in a similar fashion. □

The last Theorem leads to a formula for the inverse of an invertible matrix.

Theorem 3.116. *If A is an invertible matrix, then $A^{-1} = \frac{1}{\det(A)}\mathit{adj}(A)$.*

Proof. In the first place, remember that A^{-1} exists if and only if $\det(A) \neq 0$. From $A\mathrm{adj}(A) = \det(A) \cdot I_n$ we obtain

$$A\left(\frac{1}{\det(A)}\mathrm{adj}(A)\right) = \frac{1}{\det(A)}A\mathrm{adj}(A) = I_n$$

and the result follows. □

Example 3.117. *Find the inverse of*

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

in two different ways.

Solution. First we calculate $\det(A) = 8 + 0 + 0 - 0 - 2 - 2 = 4$, so A is invertible.

- With cofactors:

$$A_{11} = (-1)^2 \begin{vmatrix} 2 & -1 \\ -1 & 2 \end{vmatrix} = 3 \quad A_{21} = (-1)^3 \begin{vmatrix} -1 & 0 \\ -1 & 2 \end{vmatrix} = 2 \quad A_{31} = (-1)^4 \begin{vmatrix} -1 & 0 \\ 2 & -1 \end{vmatrix} = 1$$

$$A_{12} = (-1)^3 \begin{vmatrix} -1 & -1 \\ 0 & 2 \end{vmatrix} = 2 \quad A_{22} = (-1)^4 \begin{vmatrix} 2 & 0 \\ 0 & 2 \end{vmatrix} = 4 \quad A_{32} = (-1)^5 \begin{vmatrix} 2 & 0 \\ -1 & -1 \end{vmatrix} = 2$$

$$A_{13} = (-1)^4 \begin{vmatrix} -1 & 2 \\ 0 & -1 \end{vmatrix} = 1 \quad A_{23} = (-1)^5 \begin{vmatrix} 2 & -1 \\ 0 & -1 \end{vmatrix} = 2 \quad A_{33} = (-1)^6 \begin{vmatrix} 2 & -1 \\ -1 & 2 \end{vmatrix} = 3$$

Therefore $A^{-1} = \frac{1}{4} \begin{bmatrix} 3 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 3 \end{bmatrix} = \begin{bmatrix} \frac{3}{4} & \frac{1}{2} & \frac{1}{4} \\ \frac{1}{2} & 1 & \frac{1}{2} \\ \frac{1}{4} & \frac{1}{2} & \frac{3}{4} \end{bmatrix}$

- Using basic row operations on the augmented matrix $A \mid I_n$

$$\left[\begin{array}{rrr|rrr} 2 & -1 & 0 & 1 & 0 & 0 \\ -1 & 2 & -1 & 0 & 1 & 0 \\ 0 & -1 & 2 & 0 & 0 & 1 \end{array}\right] \underset{\sim}{\rho_2 + \tfrac{1}{2}\rho_1} \left[\begin{array}{rrr|rrr} 2 & -1 & 0 & 1 & 0 & 0 \\ 0 & \frac{3}{2} & -1 & \frac{1}{2} & 1 & 0 \\ 0 & -1 & 2 & 0 & 0 & 1 \end{array}\right] \begin{array}{c} \frac{1}{2}\rho_1 \\ \frac{2}{3}\rho_2 \\ \sim \end{array} \left[\begin{array}{rrr|rrr} 1 & -\frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 \\ 0 & 1 & -\frac{2}{3} & \frac{1}{3} & \frac{2}{3} & 0 \\ 0 & -1 & 2 & 0 & 0 & 1 \end{array}\right]$$

$$\underset{\rho_3 + \rho_2}{\sim} \left[\begin{array}{rrr|rrr} 1 & -\frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 \\ 0 & 1 & -\frac{2}{3} & \frac{1}{3} & \frac{2}{3} & 0 \\ 0 & 0 & \frac{4}{3} & \frac{1}{3} & \frac{2}{3} & 1 \end{array}\right] \begin{array}{c} \rho_1 + \frac{1}{2}\rho_2 \\ \sim \\ \frac{3}{4}\rho_3 \end{array} \left[\begin{array}{rrr|rrr} 1 & 0 & -\frac{1}{3} & \frac{2}{3} & \frac{1}{3} & 0 \\ 0 & 1 & -\frac{2}{3} & \frac{1}{3} & \frac{2}{3} & 0 \\ 0 & 0 & 1 & \frac{1}{4} & \frac{1}{2} & \frac{3}{4} \end{array}\right] \begin{array}{c} \rho_1 + \frac{1}{3}\rho_3 \\ \rho_2 + \frac{2}{3}\rho_3 \\ \sim \end{array}$$

$$\left[\begin{array}{rrr|rrr} 1 & 0 & 0 & \frac{3}{4} & \frac{1}{2} & \frac{1}{4} \\ 0 & 1 & 0 & \frac{1}{2} & 1 & \frac{1}{2} \\ 0 & 0 & 1 & \frac{1}{4} & \frac{1}{2} & \frac{3}{4} \end{array}\right]$$

The given formula also gives us a fast way to compute inverses of 2×2 matrices, namely if $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$, then

$$A^{-1} = \frac{1}{\det(A)} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}.$$

Despite the neat formula that we gave, for matrices with four or more rows it is usually faster to use elementary row operations to compute the inverse; for to find the adjoint of an $n \times n$ matrix one must compute n determinants each with $n-1$ rows and columns.

We already had seen an application of determinants in geometry, remember the formula for a plane given by 3 points. Next we will show how determinants can be used to study systems of linear equations.

3.5.3.2 Systems of linear equations - Cramer's rule

Theorem 3.118 (Cramer's rule)**.** *If $Ax = b$ is a linear system of n equations in n unknowns and $\det(A)$ is not zero, then the unique solution of the linear system can be written in the form*

$$x_i = \frac{\det(M_i)}{\det(A)}$$

for $i = 1, \ldots, n$, where M_i is the matrix obtained from A when column i is replaced by b.

Proof. Consider a linear system of n equations in n unknowns $x_1, x_2, \ldots, x_n$

$$Ax = b,$$

where the coefficient matrix A has non-zero determinant. The system has a unique solution, namely $x = A^{-1}b$. Using the formula for the inverse of an invertible matrix we obtain

$$x = A^{-1}b = \frac{1}{\det(A)}(\mathrm{adj}(A)b).$$

From the matrix product $\mathrm{adj}(A)b$ we can read off the i-th unknown as

$$x_i = \sum_{j=1}^{n}((\mathrm{adj}(A))_{ij}b_j)/\det(A) = (\sum_{j=1}^{n} b_j A_{ji})/\det(A).$$

Now the second sum is a determinant; in fact it is $\det(M_i)$ where M_i is the matrix obtained from A when column i is replaced by B. Hence the solution of the linear system can be expressed in the form $x_i = \det(M_i)/\det(A)$ for $i = 1, 2, \ldots, n$. □

Remark 3.119. *Note that Cramer's rule can only be used when A is a square matrix with* $\det(A) \neq 0$.

Example 3.120. *Solve the following system using Cramer's rule, if possible.*

$$\begin{array}{rcrcrcr} 2x_1 & + & 3x_2 & + & 5x_3 & = & 10 \\ 3x_1 & + & 7x_2 & + & 4x_3 & = & 3 \\ x_1 & + & 2x_2 & + & 2x_3 & = & 3 \end{array}$$

Solution. First we calculate the determinant of the matrix A

$$\det(A) = \begin{vmatrix} 2 & 3 & 5 \\ 3 & 7 & 4 \\ 1 & 2 & 2 \end{vmatrix} = 28 + 12 + 30 - 35 - 18 - 16 = 1$$

Since $\det(A) \neq 0$ the system can be solved using Cramer's rule. So we proceed by calculating M_1, M_2 and M_3.

$$\det(M_1) = \begin{vmatrix} 10 & 3 & 5 \\ 3 & 7 & 4 \\ 3 & 2 & 2 \end{vmatrix} = 140 + 36 + 30 - 105 - 18 - 80 = 3$$

Therefore $x_1 = \frac{\det(M_1)}{\det(A)} = \frac{3}{1} = 3$

$$\det(M_2) = \begin{vmatrix} 2 & 10 & 5 \\ 3 & 3 & 4 \\ 1 & 3 & 2 \end{vmatrix} = 12 + 40 + 45 - 15 - 60 - 24 = -2$$

Therefore $x_2 = \frac{\det(M_2)}{\det(A)} = \frac{-2}{1} = -2$

$$\det(M_3) = \begin{vmatrix} 2 & 3 & 10 \\ 3 & 7 & 3 \\ 1 & 2 & 3 \end{vmatrix} = 42 + 9 + 60 - 70 - 27 - 12 = 2$$

Therefore $x_3 = \frac{\det(M_3)}{\det(A)} = \frac{2}{1} = 2$

3.5.3.3 Rank of a matrix

At last we will see how determinants can help to determine the rank of a matrix.

Claim 3.121. *The rank of a matrix A is equal to the size of the largest square submatrix A' of A with* $\det(A') \neq 0$.

Example 3.122. *Determine the rank of* $A = \begin{bmatrix} 3 & 1 & 2 & 5 \\ 1 & 2 & -1 & 2 \\ 4 & 3 & 1 & 7 \end{bmatrix}$ *using determinants.*

Solution. Since A has 3 rows $r(A) \leq 3$. Hence we first need to verify if there exists a 3×3 submatrix of A with non-zero determinant.

$$\begin{vmatrix} 1 & 2 & 5 \\ 2 & -1 & 2 \\ 3 & 1 & 7 \end{vmatrix} = -7 + 12 + 10 + 15 - 28 - 2 = 37 - 37 = 0$$

$$\begin{vmatrix} 3 & 2 & 5 \\ 1 & -1 & 2 \\ 4 & 1 & 7 \end{vmatrix} = -21 + 16 + 5 + 20 - 14 - 6 = 41 - 41 = 0$$

$$\begin{vmatrix} 3 & 1 & 5 \\ 1 & 2 & 2 \\ 4 & 3 & 7 \end{vmatrix} = 42 + 8 + 15 - 40 - 7 - 18 = 65 - 65 = 0$$

$$\begin{vmatrix} 3 & 1 & 2 \\ 1 & 2 & -1 \\ 4 & 3 & 1 \end{vmatrix} = 6 - 4 + 6 - 16 - 1 + 9 = 21 - 21 = 0$$

Therefore $r(A) \leq 2$ and now we follow the same procedure for 2×2 submatrices.

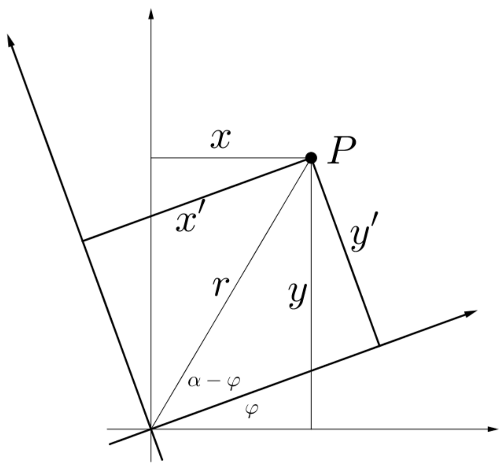
$$\begin{vmatrix} 3 & 1 \\ 1 & 2 \end{vmatrix} = 6 - 1 = 5$$

Since we found a 2×2 submatrix wit non-zero determinant we can conclude that $r(A) = 2$.

3.6 Application of matrices

3.6.1 In analytic geometry - an example

In analytic geometry, various transformations of the coordinate axes may be described using matrices. By way of example, suppose that in the two-dimensional cartesian plane we rotate the coordinate axes in an anti-clockwise direction through an angle φ, as illustrated in the figure below.



Let us compute the new coordinates (x', y') of the point P, whose old coordinates were (x, y).

From the diagram we have $x = r\cos\alpha$ and $y = r\sin\alpha$ so

$$\begin{aligned} x' = r\cos(\alpha-\varphi) &= r\cos\alpha\cos\varphi + r\sin\alpha\sin\varphi \\ &= x\cos\varphi + y\sin\varphi; \end{aligned}$$

$$\begin{aligned} y' = r\sin(\alpha-\varphi) &= r\sin\alpha\cos\varphi + r\cos\alpha\sin\varphi \\ &= y\cos\varphi - x\sin\varphi. \end{aligned}$$

These equations give x' and y' in terms of x, y and φ. They can be expressed in the matrix form

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} \cos\varphi & \sin\varphi \\ -\sin\varphi & \cos\varphi \end{bmatrix} \cdot \begin{bmatrix} x \\ y \end{bmatrix}.$$

The 2×2 matrix

$$R_\varphi = \begin{bmatrix} \cos\varphi & \sin\varphi \\ -\sin\varphi & \cos\varphi \end{bmatrix}$$

is called the **rotation matrix** associated with φ and it has the following property:

$$R_\varphi R_\varphi^T = I_2 = R_\varphi^T R_\varphi.$$

In fact, we have

$$
\begin{aligned}
R_\varphi R_\varphi^T &= \begin{bmatrix} \cos\varphi & \sin\varphi \\ -\sin\varphi & \cos\varphi \end{bmatrix} \cdot \begin{bmatrix} \cos\varphi & -\sin\varphi \\ \sin\varphi & \cos\varphi \end{bmatrix} \\
&= \begin{bmatrix} \cos^2\varphi + \sin^2\varphi & -\cos\varphi\sin\varphi + \sin\varphi\cos\varphi \\ -\sin\varphi\cos\varphi + \cos\varphi\sin\varphi & \sin^2\varphi + \cos^2\varphi \end{bmatrix} \\
&= \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix},
\end{aligned}
$$

and similarly, one can verify, $R_\varphi^T R_\varphi = I_2$.