Linear Algebra 1

Department of Mathematics and Computer Science
Technische Universiteit Eindhoven

2022-2023

Lecture notes for Linear Algebra 1 (2WF20)

Contents

Preface

These are the lecture notes for the course Linear Algebra 1 (2WF20). Together with 2WF30 Linear Algebra 2 this course provides a solid background in linear algebra for mathematics students at TU/e. It focuses on computational aspects, such as solution methods for systems of linear equations, and matrix arithmetic, but also on the theoretical structure of linear algebra, providing theorems and proofs in the setting of vector spaces.

Linear algebra is at the basis of most fields in mathematics, ranging from algebra, analysis to optimization and probability theory. So its language and results will reoccur in many courses in the bachelor Applied Mathematics. In many applications of mathematics, linear algebra often also occurs in an explicit or implicit way.

The course starts with a chapter on complex numbers, a bit of an outlier. This topic is directly related to vector geometry in dimension 2, but is also used in Linear Algebra 2 when eigenvalues and eigenspaces are discussed. Its algebraic aspects will also be further discussed in 2WF50 Algebra and Discrete Mathematics. The analysis aspects of complex numbers (differentiation and integration) are treated in the course 2WA80 Complex Analysis.

If you notice any mistakes in these lecture notes, please let me know.

Hans Sterk (editor of the Lecture notes, various people have contributed to parts of earlier versions)
Summer 2022

Chapter 1

Complex numbers

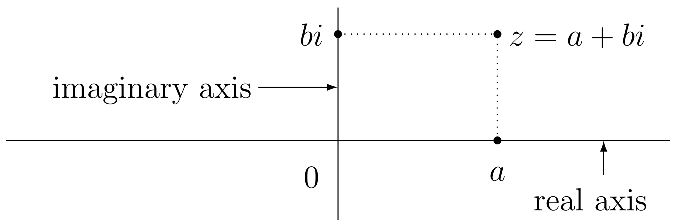
1.1 Arithmetic with complex numbers

1.1.1 We can view the real numbers as points on the 'real line'. In a similar way, we can view the extended number system to be discussed, the complex numbers, as points in the plane. We discuss the operations addition and multiplication and the properties of these new numbers, which justify the name 'numbers'. In particular, we discuss in this section

- the notion of a complex number;
- the addition/subtraction and multiplication/division of complex numbers;
- the description of a complex number using its absolute value and argument;
- the complex conjugate of a complex number;
- various arithmetical rules;
- the geometric interpretation of complex numbers and their usage in plane geometry.

1.1.2 Start with the usual coordinate system in the plane. Now call the horizontal axis the *real axis* and the vertical axis the *imaginary axis*. Every point in the plane is determined by its coordinates, say a and b, which are real numbers. We will call the point (a, b) a *complex number*, and usually denote it by $a+bi$.

The points $(a, 0)$ will simply be denoted by a, and the points $(0, b)$ on the second axis by bi. In particular, i denotes the point $(0, 1)$. So $(1, 2)$ becomes $1 + 2i$, and $(0, 3)$ becomes $3i$. We often denote a complex number by z or w. The set of complex numbers is denoted by $\mathbb{C}$.



1.1.3 Addition: the sum of two complex numbers

The *sum* of two complex numbers $z_1 = a_1 + b_1\, i$ and $z_2 = a_2 + b_2\, i$ is defined as follows:

$$z_1 + z_2 := (a_1 + a_2) + (b_1 + b_2)i.$$

For example, $(1 + i) + (-2 + 4i) = -1 + 5i$. The addition of complex numbers therefore corresponds to coordinatewise addition of points in the plane. Geometrically, it corresponds to vector addition.

The addition of complex numbers satisfies various properties, which we know from the real numbers. For instance, the addition is *commutative*: for all complex numbers z and w the equality $z + w = w + z$ holds. To prove this, write $z = a + bi$ and $w = c + di$ (with a, b, c, d real) and apply the definition first to $z + w$ to get $(a + c) + (b + d)i$ and then to $w + z$ to get $(c + a) + (d + b)i$. Since $a + c = c + a$ and $b + d = d + b$ (property of the real numbers!) we conclude that $z + w = w + z$.

In a similar way we can show that the addition is *associative*: for all complex numbers z_1, z_2, z_3 we have $(z_1 + z_2) + z_3 = z_1 + (z_2 + z_3)$. This is useful in the sense that if we encounter a sum $z_1 + z_2 + z_3$ of three complex numbers we can deal with this sum by adding $z_1 + z_2$ and z_3, or by adding z_1 and $z_2 + z_3$.

1.1.4 The product of two complex numbers

The multiplication of complex numbers is defined in a surprising way, but the definition is justified by its incredible usefulness, as we will see later.

We define the *product* of two complex numbers $z_1 = a_1 + b_1\, i$ and $z_2 = a_2 + b_2 i$ as follows:

$$z_1 z_2 := (a_1 a_2 - b_1 b_2) + (a_1 b_2 + a_2 b_1)\, i.$$

In particular, $i^2 = -1$. This definition of the multiplication looks pretty complicated, but is in fact easy to memorize and use in the following way. Just expand $(a_1 + b_1\, i)\ (a_2 + b_2\, i)$, using the rules you are used to from the real numbers, and then use one additional property, namely that $i^2 = -1$. For instance,

$$(1+i)\ (-2+4i) = -2+4i+(-2)i+i(4i) = -2+2i-4 = -6+2i.$$

Since we have defined the multiplication by using the usual rules for dealing with expressions containing symbols, it is not that surprising that the complex numbers share some arithmetical properties with the real numbers. The verification is fairly straightforward (cf. the properties of addition). Here are the most often used properties: *commutativity*, i.e., $zw = wz$ for all complex numbers z and w; *associativity*, i.e., $(z_1z_2)z_3 = z_1(z_2z_3)$ for all complex numbers z_1, z_2, z_3; *distributivity*: $z_1(z_2+z_3) = z_1z_2 + z_1z_3$ for all complex numbers z_1, z_2, z_3. Also: $z \cdot 0 = 0 (= 0 + 0i)$ for all complex numbers z.

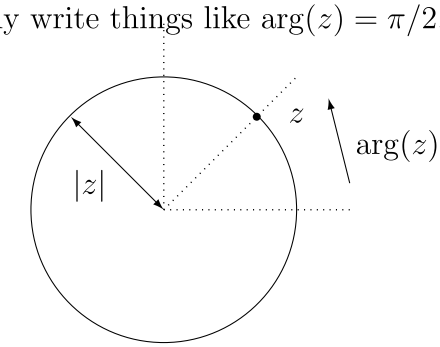
By $-z$ we mean $(-1)\cdot z$. If $z = a + bi$ (a, b real), then $-z = -a - bi$ as is easy to see. Note that for all complex numbers z we have $z + -z = 0$.

1.1.5 Absolute value and argument

Multiplication of complex numbers has a geometric interpretation in terms of scaling and rotations. To see this, we introduce the notions of absolute value and argument of a complex number, which are closely related to the *polar coordinates* (r, φ) of a point (x, y) in the plane. See also 1.1.9 and 1.1.12. Here are the definitions:

- the *absolute value* is the distance to the origin (the r of the pair (r, φ));
- the *argument* is the angle between the positive real axis and the directed segment from the origin to the complex number. The argument is only defined for nonzero complex numbers and determined up to multiples of 2π (the φ of the pair (r, φ)).

The absolute value of the complex number $z = a + bi$ (with a and b real!) is denoted by $|z|$, and equals $\sqrt{a^2+b^2}$. For instance, $|1+2i| = \sqrt{5}$. The argument is determined up to a multiple of 2π. If we choose the argument of z in the interval $(-\pi, \pi]$, then we call this value the *principal value* of z. The argument of $z \neq 0$ is denoted by $\arg(z)$. So $\arg(1+i) = \dfrac{\pi}{4} + 2k\pi$, k an integer. We also write $\arg(1+i) = \frac{\pi}{4} \pmod{2\pi}$ to indicate that the argument is determined up to multiples of 2π ('modulo 2π'). (Sometimes we are a bit sloppy and simply write things like $\arg(z) = \pi/2$, etc.)



If a complex number z has absolute value $|z|$ and argument φ, then the cartesian coordinates of the corresponding point in the plane are $|z|\cos\varphi$ and $|z|\sin\varphi$, respectively, so that

$$z = |z|\cos\varphi + i|z|\sin\varphi, \quad \text{or}$$
$$z = |z|(\cos\varphi + i\sin\varphi).$$

1.1.6 Example. A few examples:

- The complex number i has absolute value $\sqrt{0^2 + 1^2} = 1$ and argument $\pi/2$ (up to multiples of 2π).
- The complex number $\cos t + i \sin t$ (with t real) corresponds to the point $(\cos t, \sin t)$ on the unit circle. It has absolute value $\sqrt{\cos^2 t + \sin^2 t} = 1$ and argument t (up to multiples of 2π). Likewise, the complex number $r(\cos t + i \sin t)$ (with r and t real and $r > 0$) has absolute value r and argument t (up to multiples of 2π).
- The complex number $-1+i$ has absolute value $\sqrt{(-1)^2 + 1^2} = \sqrt{2}$; the principal value of the argument is $3\pi/4$. The complex number $-1 + i$ can also be written in the following way: $\sqrt{2}\,(\cos(3\pi/4) + i\sin(3\pi/4))$. The advantage of this representation is that we see immediately what the absolute value and the argument of $-1 + i$ are.

1.1.7 Real and imaginary part of a complex number

If $z = a + bi$ with $a, b \in \mathbb{R}$, then a is called the *real part* of z, denoted by $\mathrm{Re}(z)$, and b is called the *imaginary part* of z, denoted by $\mathrm{Im}(z)$. For instance, $\mathrm{Im}(2 + 3i) = 3$.

Re en Im can be viewed as maps from $\mathbb{C}$ to $\mathbb{R}$: you put a complex number in, and a real number comes out. Note that the imaginary part of a complex number is real! The following properties can be easily verified using the definitions. For alle z_1, z_2 we have:

$$\begin{aligned}
&\mathrm{Re}(z_1+z_2) = \mathrm{Re}(z_1) + \mathrm{Re}(z_2),\\
&\mathrm{Im}(z_1+z_2) = \mathrm{Im}(z_1) + \mathrm{Im}(z_2),\\
&\mathrm{Re}(z_1 z_2) = \mathrm{Re}(z_1)\mathrm{Re}(z_2) - \mathrm{Im}(z_1)\mathrm{Im}(z_2),\\
&\mathrm{Im}(z_1 z_2) = \mathrm{Re}(z_1)\mathrm{Im}(z_2) + \mathrm{Im}(z_1)\mathrm{Re}(z_2).
\end{aligned}$$

It is useful to know the first two by heart.

From the absolute value and the argument of an arbitrary complex number $z \neq 0$ it is straightforward to find its real and imaginary parts:

$$\begin{aligned}
\mathrm{Re}(z) &= |z|\cos(\arg(z)),\\
\mathrm{Im}(z) &= |z|\sin(\arg(z)).
\end{aligned}$$

Conversely, the absolute value can be deduced from the real and imaginary parts:

$$|z| = \sqrt{\mathrm{Re}(z)^2 + \mathrm{Im}(z)^2}\ .$$

The (principal value of the) argument of the complex number z is often thought to equal

$$\arg(z) = \arctan\left(\frac{\mathrm{Im}(z)}{\mathrm{Re}(z)}\right)\ ,$$

but this is not true in general (check for $z = -1 - i$). It is true if $\mathrm{Re}(z) > 0$. There is a (for this course not important) formula valid for all $z \neq 0$ which are not on the negative real axis,

$$\arg(z) = 2\arctan\frac{\mathrm{Im}(z)}{|z| + \mathrm{Re}(z)}\ .$$

1.1.8 Triangle inequality

The absolute value has the following property. For all z and w:

$$|z+w| \ \leq\ |z| + |w|.$$

This property is called the *triangle inequality*. See exercise 6 for a proof.

1.1.9 Absolute value and argument: computational rules

The absolute value and argument satisfy the following important properties. For all z, w (nonzero in the case of the argument):

$$\begin{aligned}&|zw| = |z| \cdot |w|\\ &\arg(zw) = \arg(z) + \arg(w) \pmod{2\pi}\end{aligned}$$

(We recall that (mod 2π) stands for 'modulo/up to multiples of 2π'.) These properties are based on properties of the sine and cosine. To prove them (it suffices to consider nonzero complex numbers), take two arbitrary complex numbers $z_1 \neq 0$ and $z_2 \neq 0$ with polar coordinates $(|z_1|, \varphi_1)$ and $(|z_2|, \varphi_2)$, respectively, so that

$$\begin{aligned}z_1 &= |z_1| \cos\varphi_1 + i\,|z_1| \sin\varphi_1,\\ z_2 &= |z_2| \cos\varphi_2 + i\,|z_2| \sin\varphi_2.\end{aligned}$$

Then

$$\begin{aligned}z_1 z_2 &= |z_1|\,|z_2|\Big((\cos\varphi_1\cos\varphi_2 - \sin\varphi_1\sin\varphi_2) + \\ &\qquad\qquad i(\cos\varphi_1\sin\varphi_2 + \cos\varphi_2\sin\varphi_1)\Big)\\ &= |z_1|\,|z_2|\Big(\cos(\varphi_1+\varphi_2) + i\sin(\varphi_1+\varphi_2)\Big).\end{aligned}$$

Since the absolute value of the complex number $r(\cos t + i\sin t)$ (with r, t real and $r > 0$) is r and its argument is t (up to multiples of 2π) we find:

$$\begin{aligned}|z_1\, z_2| &= |z_1|\ |z_2|,\\ \arg(z_1\, z_2) &= \arg(z_1) + \arg(z_2) \pmod{2\pi}.\end{aligned} \tag{1.1}$$

By repeated application of these rules we find for n complex numbers z_1, $z_2, \ldots, z_n$ (all nonzero in the case of the argument):

$$\begin{aligned}|z_1\, z_2 \cdots z_n| &= |z_1|\ |z_2| \cdots |z_n|,\\ \arg(z_1\, z_2 \cdots z_n) &= \arg(z_1) + \arg(z_2) + \cdots + \arg(z_n) \pmod{2\pi}.\end{aligned} \tag{1.2}$$

If all $z_1, z_2, \ldots, z_n$ are equal to z, then we find:

$$\begin{aligned}|z^n| &= |z|^n,\\ \arg(z^n) &= n\arg(z) \pmod{2\pi}.\end{aligned} \tag{1.3}$$

1.1.10 The quotient of two complex numbers

For every nonzero complex number z there exists an 'inverse', '$1/z$'. To describe it, note that $1/z$ should satisfy

$$z\,(1/z) = 1,$$

and so, using formula (1.1), it should satisfy

$$\begin{aligned} |z|\;|1/z| \;&=\; |1| = 1\;, \\ \arg(z) + \arg(1/z) &= \arg(1) = 0. \end{aligned}$$

Consequently, for nonzero z we can define $1/z$ in terms of polar coordinates by

$$\begin{aligned} \left|\tfrac{1}{z}\right| \; &= \tfrac{1}{|z|}, \\ \arg(1/z) \; &= -\arg(z). \end{aligned}$$

The *quotiënt* z/w of the complex numbers z and w (with $w \neq 0$) is defined as the product

$$z\,(1/w).$$

In terms of polar coordinates (again using formula (1.1)) we obtain for the quotient:

$$\begin{aligned} |z/w| \; &= |z|/|w|, \\ \arg(z/w) \; &= \arg(z) - \arg(w) \quad (\text{mod } 2\pi). \end{aligned} \tag{1.4}$$

The quotient satisfies the usual rules, like

$$\frac{z_1}{z_2}\,\frac{w_1}{w_2} = \frac{z_1 w_1}{z_2 w_2}.$$

The real and imaginary part of $1/z$ can be obtained as follows. Suppose $z = a + bi$ with $a, b \in \mathbb{R}$ and not both equal to 0. Then

$$\frac{1}{z} = \frac{1}{a+bi} = \frac{a-bi}{(a+bi)(a-bi)} = \frac{a-bi}{a^2+b^2}.$$

Here, we use that $(a+bi)(a-bi) = a^2 + b^2$ (verify!). For example,

$$\frac{1}{3+4i} = \frac{1}{3+4i}\cdot\frac{3-4i}{3-4i} = \frac{3-4i}{3^2+4^2} = \frac{3-4i}{25},$$

and

$$\frac{1+i}{2+3i} = \frac{1+i}{2+3i}\cdot\frac{2-3i}{2-3i} = \frac{5-i}{13}.$$

1.1.11 The complex conjugate

If $z = a+bi$ with $a, b \in \mathbb{R}$ is a complex number, then $a-bi$ is called the *complex conjugate* of z, denoted by $\bar{z}$. Geometrically, conjugation is a reflection in the real axis. The following properties are easy to verify, geometrically, or by using the definition.

$$\begin{array}{rcl} \mathrm{Re}(z) & = & \mathrm{Re}(\bar{z}), \\ \mathrm{Im}(z) & = & -\mathrm{Im}(\bar{z}), \\ \mathrm{Re}(z) & = & \frac{1}{2}(z+\bar{z}), \\ \mathrm{Im}(z) & = & \frac{1}{2i}(z-\bar{z}), \\ |z| & = & |\bar{z}|, \\ \arg(z) & = & -\arg(\bar{z}), \\ \overline{z_1+z_2} & = & \overline{z_1}+\overline{z_2}, \\ \overline{z_1 z_2} & = & \overline{z_1}\cdot\overline{z_2}, \\ z+\bar{z} & = & 2\,\mathrm{Re}(z), \\ z\,\bar{z} & = & |z|^2. \end{array}$$

Note that the last formula implies

$$\frac{1}{z} = \frac{\bar{z}}{|z|^2}\ ,$$

in accordance with 1.1.10. So $1/z$ is obtained by first reflecting in the real axis and then scaling by a factor $|z|^2$.

By way of example, we provide the details of the proof of the identity $\overline{z_1+z_2} = \overline{z_1}+\overline{z_2}$. Write $z_1 = a_1+b_1 i$ and $z_2 = a_2+b_2 i$ (with a_i and b_i real for $i = 1, 2$). Then $z_1+z_2 = a_1+b_1 i+a_2+b_2 i = (a_1+a_2)+(b_1+b_2)i$ so that

$$\overline{z_1+z_2} = \overline{(a_1+a_2)+(b_1+b_2)i} = (a_1+a_2)-(b_1+b_2)i.$$

Now $(a_1+a_2)-(b_1+b_2)i = (a_1-b_1 i)+(a_2-b_2 i) = \overline{z_1}+\overline{z_2}$ and we are done.

1.1.12 Geometric interpretations

Let $z = a+bi$ and $w = c+di$ be two complex numbers. The absolute value $|z-w|$ is the distance between z and w, since $|z-w| = \sqrt{(c-a)^2+(d-b)^2}$ and using the Pythagorean theorem applied to (a, b) and (c, d). The following examples show the usefulness of this interpretation.

- The solutions to the equation $|z-i| = 5$ are precisely the complex numbers at distance 5 from i. So they form a circle with center i and radius 5. Of course, we can rewrite the equation in terms of coordinates to get the well-know equation of a circle: if $z = x + iy$, then

$$|z - i| = 5 \Leftrightarrow |z - i|^2 = 25 \Leftrightarrow |x + (y - 1)i|^2 = 25 \Leftrightarrow x^2 + (y - 1)^2 = 25.$$

- Given two complex numbers z_1 and z_2, the equation $|z - z_1| = |z - z_2|$ describes the points with equal distance to z_1 and z_2. This is precisely the perpendicular bisector of the segment with endpoint z_1 and z_2.

Complex multiplication with a fixed complex number also has a geometric interpretation, which follows from 1.1.9. Let's illustrate this for multiplication by i. If $z \neq 0$, then $|zi| = |z| \cdot |i| = |z|$, and $\arg(zi) = \arg(z) + \arg(i) = \arg(z) + \pi/2$. In other words, zi can be obtained from z by rotating z (around the origin) through an angle $\pi/2$. Similarly, multiplication by $1 + i$ comes down to rotating z through an angle $\pi/4$ and scaling by a factor $\sqrt{2}$. In general, if $z = r(\cos t + i \sin t)$, with $r > 0$, then multiplication by z produces a scaling by r and a rotation through an angle t.

1.1.13 In computations with complex numbers we can use the representation in terms of their real and imaginary parts ('$a + bi$', cartesian coordinates) or in terms of their absolute values and arguments ('$|z|$' and '$\arg(z)$', or $r(\cos t + i \sin t)$). Here is a rough indication of when to use which representation.

- If the computation mainly involves additions, then try using cartesian coordinates first.
- If the computation mainly involves multiplications (and powers), and no additions, then try absolute values and arguments first.

Also note the following: two complex numbers are equal if and only if their real parts and their imaginary parts are equal. Two nonzero complex numbers are equal if and only if their absoluate values are equal and their arguments are equal up to multiples of 2π.

1.1.14 **Example.** We solve the equation $z^2 = 2i$ in two different ways.

- (First solution) Here we write $z = x + iy$ (with x and y real) and substitute
$$x^2 + 2ixy - y^2 = 2i.$$

Looking at the real and imaginary parts produces the two (real) equations $x^2 - y^2 = 0$ and $2xy = 2$. From the first equation we deduce for the real x and y that $x = y$ or $x = -y$. Substituting in the second equation then yields $x^2 = 1$ (in case $x = y$), and $x^2 = -1$ (in case $x = -y$). The equation $x^2 = -1$ has no real solutions, and the equation $x^2 = 1$ leads to $x = 1$ or $x = -1$. So we find the two solutions $1 + i$ en $-1 - i$ of $z^2 = -1$.

- (Second solution) First observe that any solution has to be nonzero. Using the absolute value and argument we then get:

$$\begin{aligned} z^2 = 2i \quad &\Leftrightarrow |z^2| = 2 \quad \text{and} \quad \arg(z^2) = \pi/2 + 2k\pi \quad (k \text{ integral}) \\ &\Leftrightarrow |z|^2 = 2 \quad \text{and} \quad 2\arg(z) = \pi/2 + 2k\pi \quad (k \text{ integral}) \\ &\Leftrightarrow |z| = \sqrt{2} \quad \text{and} \quad \arg(z) = \pi/4 + k\pi \quad (k \text{ integral}). \end{aligned}$$

This leads to the following two solutions:

$$\sqrt{2}\,(\cos(\pi/4) + i\sin(\pi/4)) \quad \text{and} \quad \sqrt{2}\,(\cos(5\pi/4) + i\sin(5\pi/4))$$

(we only use the values $k = 0$ and $k = 1$, since for $k = 2$ we find the same solution as for $k = 0$, for $k = 3$ we find the same solution as for $k = 1$, etc.). Verify that these two numbers are $1 + i$ and $-1 - i$, respectively.

Note that the second approach is to be preferred if the exponent is bigger, for instance, $z^6 = -1$ (try the first approach and you'll quickly see why).

1.1.15 Example. We prove that $zw = 0$ implies $z = 0$ or $w = 0$ in two different ways. (In fact: $zw = 0 \Leftrightarrow z = 0$ or $w = 0$.)

- First proof. If $z = 0$ we are done. Suppose that $z \neq 0$, then we need to show that $w = 0$. From $zw = 0$ we infer $|zw| = 0$ so that $|z| \cdot |w| = 0$. So now we are in the situation where the product of the two real numbers $|z|$ and $|w|$ is 0. Since $|z| \neq 0$ we conclude that $|w| = 0$ (using the properties of real numbers). But then $w = 0$.

- Second proof. If $z = 0$ we are done. Suppose that $z \neq 0$. Now multiply $zw = 0$ on both sides with $1/z$:

$$\frac{1}{z} \cdot (zw) = \frac{1}{z} \cdot 0 (= 0).$$

Rewrite the left-hand side using associativity: $\frac{1}{z} \cdot (zw) = (\frac{1}{z} \cdot z)w = w$. So $w = 0$.

A third approach would be to write $z = x + iy$ and $w = u + iv$ (with x, y, u, v real) and analyse $(x + iy)(u + iv) = 0$. This approach is computationally more involved. The previous two proofs show the usefulness of the arithmetic rules for complex numbers.

1.2 The exponential function, sine and cosine

So far we have discussed addition, substraction, multiplication and division of complex numbers. In this section we turn to the complex exponential function, the complex sine and cosine function.

1.2.1 Definition. (Complex exponential function) For very complex number z we define the complex number e^z by giving its absolute value and argument:

$$\begin{aligned} |e^z| &= e^{\mathrm{Re}(z)} , \\ \arg(e^z) &= \mathrm{Im}(z) . \end{aligned}$$

1.2.2 Note the use of the real exponential function in this definition. The definition of the complex exponential function agrees with the real exponential function for real numbers z, since for a *real* number $z = x + i \cdot 0$, our new definition 1.2.1 yields $|e^z| = e^x$ and $\arg(e^z) = \mathrm{Im}(z) = 0$. So e^z equals the real exponential e^z.

Note furthermore that $e^z \neq 0$ for all complex z because $|e^z| = e^{\mathrm{Re}(z)} \neq 0$ (the real exponential function has no zeros).

1.2.3 Example. The complex number $e^{\pi i}$ has absolute value $e^{\mathrm{Re}(\pi i)} = e^0 = 1$ and argument $\mathrm{Im}(\pi i) = \pi$, so that $e^{\pi i} = -1$. The number $e^{1+\pi i/2}$ has absolute value $e^{\mathrm{Re}(1+\pi i/2)} = e^1$ and argument $\mathrm{Im}(1+\pi i/2) = \pi/2$, so that $e^{1+\pi i/2} = e\, i$.

1.2.4 Example. To solve the equation $e^z = 1 + i$ we compare the absolute values and arguments of the left-hand and right-hand sides. Writing $z = x + iy$ this gives us the following two equations:

$$e^x = \sqrt{2} \quad \text{and} \quad y = \frac{\pi}{4} + 2k\pi \quad \text{with } k \text{ integral.}$$

The (infinite) solution set is therefore:

$$\{\frac{1}{2}\log(2) + i(\frac{\pi}{4} + 2k\pi) \mid k \in \mathbb{Z}\},$$

where log denotes the natural logarithm.

1.2.5 Theorem. $e^{z_1}\, e^{z_2} = e^{z_1+z_2}$ *for all complex numbers* z_1 *and* z_2*.*

Proof. We prove this equality by showing that both sides have the same absolute value and the same argument. Here is the computation for the absolute value (note that we use arithmetical rules for the absolute value, and for the real exponential function):

$$\begin{aligned}|e^{z_1}e^{z_2}| &= |e^{z_1}|\,|e^{z_2}| = e^{\mathrm{Re}(z_1)}e^{\mathrm{Re}(z_2)} = e^{\mathrm{Re}(z_1)+\mathrm{Re}(z_2)}\ ,\\ |e^{z_1+z_2}| &= e^{\mathrm{Re}(z_1+z_2)} = e^{\mathrm{Re}(z_1)+\mathrm{Re}(z_2)}\ .\end{aligned}$$

So $e^{z_1}\ e^{z_2}$ and $e^{z_1+z_2}$ have the same absolute value. Similarly,

$$\begin{aligned}\arg(e^{z_1}e^{z_2}) &= \arg(e^{z_1}) + \arg(e^{z_2}) = \mathrm{Im}(z_1) + \mathrm{Im}(z_2)\ ,\\ \arg(e^{z_1+z_2}) &= \mathrm{Im}(z_1+z_2) = \mathrm{Im}(z_1) + \mathrm{Im}(z_2)\ .\end{aligned}$$

So $e^{z_1}\ e^{z_2}$ and $e^{z_1+z_2}$ have the same argument. This concludes the proof. □

1.2.6 Theorem. $(e^z)^n$ *exists for every integral exponent* n*. It satisfies (for every complex number* z *and integral* n*) the following property:*

$$(e^z)^n = e^{nz}\ .$$

Proof. The existence statement follows from the fact that $e^z \neq 0$. So we turn to the proof of the equality $(e^z)^n = e^{nz}$. For positive integral n this property is a consequence of Theorem 1.2.5 (and for the mathematicians: use mathematical induction). For $n = 0$ both sides equal 1.

Next we turn to the case $n = -1$. Because of 1.1.10 the absolute value of $1/e^z$ equals

$$|(e^z)^{-1}| = |e^z|^{-1} = \left(e^{\mathrm{Re}(z)}\right)^{-1} = e^{-\mathrm{Re}(z)} = e^{\mathrm{Re}(-z)}\,.$$

Analogously, we find for the argument:

$$\arg\left((e^z)^{-1}\right) = -\arg(e^z) = -\mathrm{Im}(z) = \mathrm{Im}(-z)\ .$$

So $(e^z)^{-1} = e^{-z}$. Applying Theorem 1.2.5 the equality $(e^z)^n = e^{nz}$ follows for all integral negative n. □

1.2.7 Corollary. *$e^{2\pi in} = 1$ and $e^{\pi in} = (-1)^n$ for every integral n.*

Both properties follow directly from the definition of e^z and Theorem 1.2.6.

1.2.8 Theorem. *The function e^z is periodic and has period $2\pi i$.*

Proof. $e^{z+2\pi i} = e^z e^{2\pi i} = e^z \cdot 1 = e^z$. □

1.2.9 The formulas in Corollary 1.2.7 are a special case of a general property. Let φ be a real number. Then $e^{i\varphi}$ is a complex number with absolute value 1 and argument φ, so:

Property. For every real number φ we have

$$e^{i\varphi} = \cos\varphi + i\sin\varphi \ .$$

This relation connects the complex exponential function and the (real) sine and cosine. Moreover, it provides another short way of representing a complex number:

$$z = |z| \ e^{i\arg z}.$$

For instance, $1 + i = \sqrt{2}\, e^{\pi i/4}$.

1.2.10 From 1.2.9 the following relation for real φ follows:

$$e^{-i\varphi} = \cos\ \varphi - i\sin\ \varphi \ .$$

From $e^{i\varphi} = \cos\varphi + i\sin\varphi$ and $e^{-i\varphi} = \cos\ \varphi - i\sin\ \varphi$ we deduce that for every real φ the following relations hold:

$$\cos\ \varphi = \tfrac{1}{2}\Big(e^{i\varphi} + e^{-i\varphi}\Big) \ ,$$

$$\sin\ \varphi = \tfrac{1}{2i}\Big(e^{i\varphi} - e^{-i\varphi}\Big) \ .$$

Based on this we *define* for an arbitary complex number z the (complex) *sine* and *cosine* as follows.

1.2.11 Definition. (Complex sine and cosine)

$$\cos\ z = \tfrac{1}{2}\Big(e^{iz} + e^{-iz}\Big) \ ,$$

$$\sin\ z = \tfrac{1}{2i}\Big(e^{iz} - e^{-iz}\Big) \ .$$

it follows from 1.2.9 that these definitions agree with the real sine and cosine, i.e., if you take z to be a real number in the new definition, then $\sin z$ and $\cos z$ are simply the usual real sine and cosine of z, respectively.

1.2.12 Theorem. *Sine and cosine are periodic and have period 2π. Also,*

$$\sin^2(z) + \cos^2(z) = 1 \quad \forall z \in \mathbb{C}.$$

Proof.

$$\begin{aligned} \sin(z + 2\pi) &= \tfrac{1}{2i}\left(e^{i(z+2\pi)} - e^{-i(z+2\pi)}\right) = \tfrac{1}{2i}\left(e^{iz}e^{2\pi i} - e^{-iz}e^{-2\pi i}\right) \\ &= \tfrac{1}{2i}\left(e^{iz} - e^{-iz}\right) = \sin z, \end{aligned}$$

since $e^{\pm 2\pi i}$ has absolute value 1 and argument $\pm 2\pi$, and therefore equals 1. The proof that the cosine is periodic with period 2π is similar.

The relation $\sin^2(z) + \cos^2(z) = 1$ is proved by substituting the defining expressions for $\sin(z)$ and $\cos(z)$:

$$\left(\frac{1}{2}(e^{iz}+e^{-iz})\right)^2+\left(\frac{1}{2i}(e^{iz}-e^{-iz})\right)^2 = \frac{1}{4}(e^{2iz}+2+e^{-2iz})-\frac{1}{4}(e^{2iz}-2+e^{-2iz}) = 1.$$

□

1.2.13 Example. Solve the equation

$$\cos(z) = 2\ .$$

It is clear that there are no real solutions. But there turn out to be complex solutions. First rewrite the equation in terms of the exponential function:

$$\frac{1}{2}\left(e^{iz} + e^{-iz}\right) = 2\ .$$

Now set $w = e^{iz}$, then $w \neq 0$ because of 1.2.2, and we find:

$$\begin{aligned} &w + \tfrac{1}{w} = 4\ , \\ &w^2 - 4w + 1 = 0\ , \\ &(w - 2)^2 = 3\ , \\ &w = 2 \pm \sqrt{3}\ . \end{aligned}$$

So we arrive at:

$$|e^{iz}| = e^{-\mathrm{Im}(z)} = |w| = 2 \pm \sqrt{3}\ ,\ \text{dus } \mathrm{Im}(z) = -\ \log(2 \pm \sqrt{3})\ .$$

$$\arg(e^{iz}) = \mathrm{Re}(z) = \arg w = 0 \quad (\mathrm{mod}\ 2\pi),\ \text{dus } \mathrm{Re}(z) = k\ 2\pi\ ,\ k \in \mathbb{Z}\ .$$

Therefore all solutions are

$$\begin{aligned} z &= -\,i \log(2 + \sqrt{3}) + k \cdot 2\pi\ ,\ k \in \mathbb{Z}, \\ z &= -\,i \log(2 - \sqrt{3}) + k \cdot 2\pi\ ,\ k \in \mathbb{Z}. \end{aligned}$$

Note that the absolute value of the complex cosine is not bounded by 1 like for the real cosine.

1.2.14 Theorem.

$$\begin{aligned} \overline{e^z} &= e^{\bar{z}}, \\ \overline{\sin(z)} &= \sin(\bar{z}), \\ \overline{\cos(z)} &= \cos(\bar{z}). \end{aligned}$$

Proof.

$$\begin{aligned} |\overline{e^z}| &= |e^z| = e^{\mathrm{Re}(z)} = e^{\mathrm{Re}(\bar{z})}, \\ \arg\left(\overline{e^z}\right) &= -\ \arg(e^z) = -\,\mathrm{Im}(z) = \mathrm{Im}(\bar{z}), \end{aligned}$$

so $\overline{e^z} = e^{\bar{z}}$.

$$\begin{aligned} \overline{\sin(z)} &= \overline{\frac{1}{2i}\left(e^{iz} - e^{-iz}\right)} = \left(\overline{\frac{1}{2i}}\right)\left(\overline{e^{iz}} - \overline{e^{-iz}}\right) = -\frac{1}{2i}\left(e^{\bar{i}\bar{z}} - e^{-\bar{i}\bar{z}}\right) \\ &= -\frac{1}{2i}\ \left(e^{-i\bar{z}} - e^{i\bar{z}}\right) = \frac{1}{2i}\left(e^{i\bar{z}} - e^{-i\bar{z}}\right) \\ &= \sin(\bar{z}). \end{aligned}$$

The formula for $\overline{\cos(z)}$ can be proved in a similar way. $\square$

1.2.15 The formula $e^z e^w = e^{z+w}$ for the complex exponential function is useful in deriving trigonometric formulas. For instance, start with $e^{2it} = e^{it}e^{it}$ (for real t) and rewrite this relation as follows:

$$\cos(2t) + i\sin(2t) = (\cos t + i\sin t)(\cos t + i\sin t)$$

Since $(\cos t + i \sin t)(\cos t + i \sin t) = \cos^2 t - \sin^2 t + 2i \sin t \cos t$ we find upon comparing real and imaginary parts:

$$\cos(2t) = \cos^2 t - \sin^2 t \text{ en } \sin(2t) = 2 \sin t \cos t.$$

This formula (and many similar ones) also turn out to hold for complex values of t; this is easily verified by applying the definition of the complex sine and cosine.

In a similar way the formula $e^{ia}e^{ib} = e^{i(a+b)}$ leads to

$$\begin{aligned} \cos(a+b) &= \cos a \cos b - \sin a \sin b; \\ \sin(a+b) &= \sin a \cos b + \cos a \sin b. \end{aligned}$$

1.3 Complex polynomials

1.3.1 Solving equations is very important in mathematics. Complex numbers enable us to solve more equations than with just the real numbers. They also enable us to see the connections between various types of equations. In this section we will discuss various types of polynomial equations.

1.3.2 Complex polynomials

An expression of the form

$$a_n z^n + a_{n-1} z^{n-1} + \cdots + a_1 z + a_0$$

in which $a_0, \ldots, a_n$ are complex numbers, is called a *complex polynomial* in z. If $a_n \neq 0$ then n is the *degree* of the polynomial. The numbers $a_0, \ldots, a_n$ are called the *coefficients* of the polynomial. If they are all real, then the polynomial is called *real*.

Let $p(z)$ be a polynomial. If $p(\alpha) = 0$, then α is called a *zero* or *root* of the polynomial. It is also called a *solution* of the polynomial equation $p(z) = 0$.

The following formulas (see (1.2) and (1.3)) will play an important role in this section:

$$\begin{aligned} |z_1 z_2 \cdots z_n| &= |z_1|\ |z_2| \cdots |z_n|, \\ \arg(z_1 z_2 \cdots z_n) &= \arg(z_1) + \arg(z_2) + \cdots + \arg(z_n) \pmod{2\pi} \end{aligned}$$

and, in particular, if $z_1 = z_2 = ... = z_n = z$:

$$\begin{aligned} |z^n| &= |z|^n, \\ \arg(z^n) &= n \arg(z) \pmod{2\pi}. \end{aligned}$$

1.3.3 Example. Here is how we use these formulas to solve the equation

$$z^3 = i \ .$$

(If you rewrite it like $z^3 - i = 0$ you see that it comes from a polynomial equation.) We solve this equation by comparing the absolute values and arguments of both sides of the equation. First note that any solution is nonzero. Now turn to the absolute values:

$$|z^3| = |z|^3 = |i| = 1 \, , \text{ so } |z| = 1 \ .$$

And for the arguments we find (here we use that $z \neq 0$)

$$\arg(z^3) = 3 \arg(z) = \arg(i) = \frac{\pi}{2} + k \cdot 2\pi \, , \; k \in \mathbb{Z} \ ,$$

so that

$$\arg(z) = \frac{\pi}{6} + k \cdot \frac{2\pi}{3} \, , \; k \in \mathbb{Z} \ .$$

Up to multiples of 2π we find three distinct values for $\arg(z)$, for $k = 0$, $k = 1$, $k = 2$. There are therefore precisely three solutions of the equation:

$$z = \cos\frac{\pi}{6} + i \sin\frac{\pi}{6}, \quad z = \cos\frac{5\pi}{6} + i \sin\frac{5\pi}{6}, \quad z = \cos\frac{9\pi}{6} + i \sin\frac{9\pi}{6}.$$

Depending on what the solutions are needed for, other forms may be more practical, like $e^{i\pi/6}$, $e^{5i\pi/6}$, $e^{3i\pi/2}$, or: $\frac{1}{2}\sqrt{3} + \frac{1}{2}i$, $-\frac{1}{2}\sqrt{3} + \frac{1}{2}i$, $-i$.

1.3.4 The equation of the previous example is a special case of the equation

$$z^n = a,$$

in which n is a nonnegative integer and a is a complex number with $a \neq 0$. This type of equation lends itself very well for the approach using absolute values and arguments. From $z^n = a$ we obtain

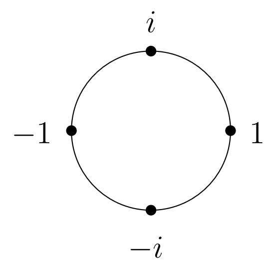
$$|z^n| = |z|^n = |a| \, , \text{ dus } |z| = \sqrt[n]{|a|} \ ,$$

$$\arg(z^n) = n \arg(z) = \arg(a) + k \cdot 2\pi \, , \; k \in \mathbb{Z} \ ,$$

$$\arg(z) = \frac{1}{n}\arg(a) + k \cdot \frac{2\pi}{n} \, , \; k \in \mathbb{Z} \ .$$

Up to multiples of 2π we find, for $k = 0, ..., n-1$, exactly n distinct values for $\arg z$.

So the equation $z^n = a$ has n disctinct solutions which are all located ar regular angular intervals on the circle with center 0 and radius $\sqrt[n]{|a|}$, i.e. the circle with equation $|z| = \sqrt[n]{|a|}$. In the figure, the four solutions of $z^4 = 1$ are drawn.



The equation $z^n = 0$ has 'n coinciding' solutions, or a solution $z = 0$ with *multiplicity* n. See 1.3.6.

The equation
$$z^2 = a\,,\;\; a \neq 0$$
is also a special case. It has two solutions, both with absolute value $\sqrt{|a|}$ and with arguments $\frac{1}{2}\arg(a)$ and $\frac{1}{2}\arg(a) + \pi$, respectively.

Next, we investigate the zeros of a complex polynomial $p(z)$ of degree n.

1.3.5 Theorem. *If α is a root of the complex polynomial $p(z)$ of degree $n > 0$, i.e., $p(\alpha) = 0$, then $z - \alpha$ is a factor of $p(z)$, i.e., there exists a complex polynomial $q(z)$ of degree $n-1$ such that*
$$p(z) = (z - \alpha)\, q(z)\ .$$

Proof. For every complex number α (regardless of it being a zero of $p(z)$ or not). we can divide $p(z)$ by $z - \alpha$ (e.g., by using the technique of *long division*). We then get a quotient $q(z)$ and a remainder r (a constant):
$$p(z) = (z - \alpha)q(z) + r\ .$$
If α is a zero, so that $p(\alpha) = 0$, then we obtain
$$0 = p(\alpha) = (\alpha - \alpha)q(\alpha) + r$$
and so $r = 0$. $\square$

1.3.6 If $p(\alpha) = 0$, then $p(z)$ can be written as $(z-\alpha)q(z)$ for some polynomial $q(z)$. If $q(\alpha) = 0$, then $q(z)$ also contains a factor $z-\alpha$, and so we have (for some polynomial $s(z)$)

$$p(z) = (z-\alpha)^2 s(z)\,, \quad \text{enz.}$$

Definition. α is called a zero of *multiplicity* m of the complex polynomial $p(z)$ if there exists a polynomial $t(z)$ with $t(\alpha) \neq 0$ such that

$$p(z) = (z-\alpha)^m t(z)\ .$$

The multiplicity of a zero α is the number of factors $z-\alpha$ of $p(z)$.

1.3.7 If $p(z)$ is a polynomial of degree n, then the preceeding discussion implies that thet total number of zeros, counted with multiplicities, is at most n. In fact, we have the following even stronger property.

1.3.8 Theorem. (Fundamental theorem of algebra) *Every complex polynomial of degree n has exactly n zeros if every zero is counted with its multiplicity.*

The proof is beyond the scope of this course (and requires advanced techniques). Note that we did prove the theorem for the special class of polynomials of the form $z^n - a$. Also note that the theorem doesn't hold over the real numbers: x^2+1 is a real polynomial of degree 2 without any zeros in $\mathbb{R}$.

1.3.9 Polynomials of degree 1

For polynomials of degree 1 the zero is easily found as follows

$$az + b = 0\,,\ a \neq 0$$

implies $z = -\,b/a$.

1.3.10 Polynomials of degree 2

Next we turn to polynomials of degree 2:

$$az^2 + bz + c\,,\ a \neq 0\ .$$

Rewrite the polynomial as follows:

$$az^2+bz+c = a(z^2+\frac{b}{a}z)+c = a(z+\frac{b}{2a})^2+\frac{4ac-b^2}{4a}\ .$$

Now let $w = z + \frac{b}{2a}$, then we find the (quadratic) equation

$$w^2 = \frac{b^2 - 4ac}{4a^2} \ ,$$

which is of the type we discussed before. It has two solutions for w (unless $b^2 - 4ac = 0$), see 1.3.4, from which the two solutions for z follow immediately.

The technique we used in rewriting the quadratic equation is called *completing the square*. Note that we do not use the *abc* formula, since we haven't defined square roots.

1.3.11 Example. Consider the equation

$$z^2 + (2 + 4i)z + i = 0 \ .$$

Completing the square yields

$$(z + (1 + 2i))^2 = -3 + 3i \ ,$$

so that (see 1.3.4.):

$$\begin{aligned} z + 1 + 2i \ &= \sqrt[4]{18}(\cos(\frac{3\pi}{8}) + i\sin(\frac{3\pi}{8})) \quad \text{or} \\ z + 1 + 2i \ &= \sqrt[4]{18}(\cos(\frac{11\pi}{8}) + i\sin(\frac{11\pi}{8})). \end{aligned}$$

Finally,

$$\begin{aligned} z \ &= -1 + \sqrt[4]{18}\cos(\frac{3\pi}{8}) + i(-2 + \sqrt[4]{18}\sin\frac{3\pi}{8}) \quad \text{or} \\ z \ &= -1 - \sqrt[4]{18}\cos(\frac{3\pi}{8}) + i(-2 - \sqrt[4]{18}\sin\frac{3\pi}{8}). \end{aligned}$$

1.3.12 Degree 3 and higher

For polynomials of degree 3 and 4 there exist (complicated) ways to produce the solutions in an algorithmic manner (there are formulas like the *abc* formula). For polynomials of degree 5 and higher such algorithms and formulas do not exist. In those cases we have to use numerical methods to approximate the zeros. Of course, in some specific cases one may be able to find (some of) the exact solutions, for instance for equations of the form $z^n - a = 0$.

The following theorem deals with polynomials whose coefficients are all real. In that case any non-real solution automatically produces a 'twin' solution.

1.3.13 Theorem. *Let*

$$p(z) = a_n z^n + a_{n-1} z^{n-1} + \cdots + a_1 z + a_0 \ ,$$

where $a_n, a_{n-1}, ..., a_1, a_0$ *are all real. If the complex number* α *satisfies* $p(\alpha) = 0$*, then* $p(\bar{\alpha}) = 0$.

Proof. Since α is a zero, we have

$$a_n \alpha^n + a_{n-1} \alpha^{n-1} + \cdots + a_1 \alpha + a_0 = 0 \ .$$

Taking the complex conjugate yields

$$\overline{a_n \alpha^n + a_{n-1} \alpha^{n-1} + \cdots + a_1 \alpha + a_0} = \bar{0} = 0 \ .$$

Applying the rules for complex conjugation 1.1.11 (and the fact that $\overline{a_k} = a_k$ for $k = 1, \ldots, n$) then produces

$$a_n \bar{\alpha}^n + a_{n-1} \bar{\alpha}^{n-1} + \cdots + a_1 \bar{\alpha} + a_0 = 0,$$

so $p(\bar{\alpha}) = 0$. □

1.3.14 Corollary. *Every nonzero polynomial with real coefficients can be factored as a product of real polynomials of degree 1 and 2.*

Proof. Let $p(z)$ be a polynomial with real coefficients. If α is a real zero of $p(z)$, then $p(z)$ can be written as

$$p(z) = (z - \alpha) q(z) \ ,$$

where $q(z)$ is also a polynomial with real coefficients. If α is a non-real zero of $p(z)$, then $\bar{\alpha}$ is also a zero and

$$\begin{aligned} p(z) &= (z - \alpha)(z - \bar{\alpha}) \ r(z) \\ &= (z^2 - (\alpha + \bar{\alpha}) z + \alpha\bar{\alpha}) \ r(z) \\ &= (z^2 - 2\mathrm{Re}(\alpha) z + |\alpha|^2) \ r(z) \ . \end{aligned}$$

The first factor has real coefficients, so $r(z)$ has real coefficients. Since the degrees of $q(z)$ and $r(z)$ are less than the degree of $p(z)$, we can repeat this construction until we get to the point where the degrees of the quotients are 0. □

1.3.15 Example. Suppose we wish to factor the polynomial

$$p(z) = z^5 - 6z^4 + 25z^3 - z^2 + 6z - 25.$$

into real factors of degrees 1 and 2. Suppose also that we know that $3-4i$ is a zero of the polynomial. Because the polynomial has real coefficients, $3+4i$ is also a zero, and so the polynomial has a factor

$$(z - 3 + 4i)(z - 3 - 4i) = z^2 - 6z + 25.$$

Long division then gives

$$p(z) = (z^2 - 6z + 25)(z^3 - 1)\ .$$

The last factor has a zero $z = 1$ and therefore contains a factor $z - 1$:

$$p(z) = (z^2 - 6z + 25)(z - 1)(z^2 + z + 1)\ .$$

The third factor, $z^2 + z + 1$, has no real factors of degree 1. (Since both zeros of this polynomial are non-real.)

1.4 Geometry with complex numbers

1.4.1 Complex numbers have proven their usefulness in many branches of mathematics (and other disciplines as well). In this section we explore the relation with plane geometry, and show how complex numbers add to the techniques for addressing geometric problems. In the next chapter we will see a similar role for vectors.

In the sequel we identify the plane from classical geometry with the complex plane. Points are then described by complex numbers. Of course, we have the freedom to put for instance the origin at a suitable place in a given problem. We will assume some basic facts from classical geometry at certain points in our discussion. The book [4] elaborates more extensively on the use of complex numbers in planar geometry.

1.4.2 Lines and segments in the (complex) plane
If $z \neq 0$ is a complex number, then the numbers tz with t real describe the line through 0 and z. The segment with endpoints 0 and z is decribed by taking t in the interval $[0, 1]$. We sometimes denote this segment by $[0, z]$. The midpoint of the segment $[0, z]$ is the complex number $\dfrac{1}{2}\, z$.

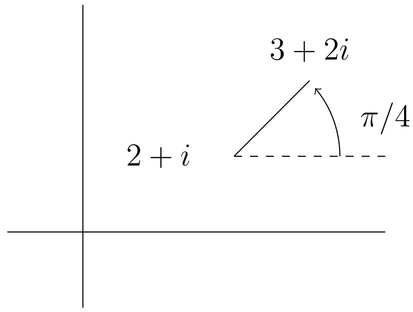


Figure 1.1: *The length of the segment* $[2+i, 3+2i]$ *equals* $\sqrt{2}$ *and its angle with the positive real axis is* $\pi/4$.

For two distinct complex numbers z and w the complex numbers of the form $z + t(w - z)$ with t real run through the points of the line through z and w. The segment with endpoints z and w (whose points correspond to parameter values t in the interval $[0, 1]$) is denoted by $[z, w]$. For $t = \dfrac{1}{2}$ we find the midpoint of the segment $[z, w]$:

$$w + \frac{1}{2}(z - w) \text{ or } \frac{1}{2}(z + w).$$

The length of the segment $[z, w]$ is equal to the distance between z and w, i.e., $|w - z|$. The complex number $w - z$ not only encodes the information on the length of the segment $[z, w]$, but also on the segment's direction via its argument. For example, the segment $[2 + i, 3 + 2i]$ has length $|1 + i| = \sqrt{2}$, and it makes an angle of $\pi/4$ radians with the positive real axis since $\arg(1 + i) = \pi/4$.

The lines through z_1 and z_2 (with $z_1 \neq z_2$) and through w_1 and w_2 (with $w_1 \neq w_2$), respectively, are parallel if and only if $w_2 - w_1$ is a real multiple of $z_2 - z_1$, or, equivalently, $\dfrac{w_2 - w_1}{z_2 - z_1}$ is real (or: has argument $k\pi$, k an integer).

1.4.3 Example. This example illustrates the use of complex numbers in handling segments in a triangle. Suppose $\triangle ABC$ is a triangle in the plane. Let D be the midpoint of AC and let E be the midpoint of BC. We will show, using complex numbers, that DE is parallel with AB and that the length of segment AB is twice the length of segment DE.

To show this, let A, B, C correspond to the complex numbers z_1, z_2, z_3, respectively (it turns out to be irrelevant where the origin is). Then D and

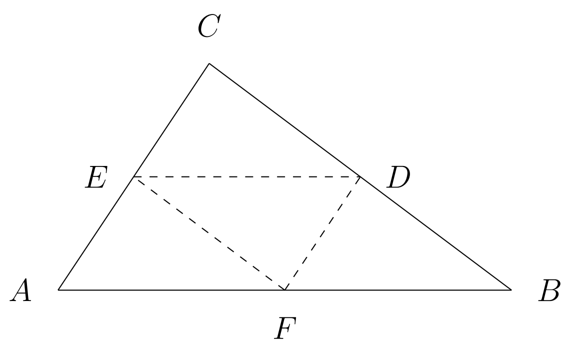


Figure 1.2: $\triangle ABC$ *and the triangle* DEF *of midpoints.*

E correspond to

$$\frac{1}{2}(z_1 + z_3) \text{ and } \frac{1}{2}(z_2 + z_3)\,,$$

respectively. Since

$$\frac{1}{2}(z_2 + z_3) - \frac{1}{2}(z_1 + z_3) = \frac{1}{2}(z_2 - z_1)$$

we conclude that DE and AB are parallel ($\frac{1}{2}(z_2 - z_1)$ is a real multiple of $z_2 - z_1$) and that the length $|\frac{1}{2}(z_2 - z_1)|$ of segment DE is half that of segment AB.

1.4.4 Translations

Let u be a complex number. The map $T : \mathbb{C} \to \mathbb{C}$ given by $T(z) = z + u$ is a *translation over* u. Translations 'preserve shapes', so, for example, they transform straight lines into straight lines.

1.4.5 Rotations and circles

If w is a complex number with absolute value 1 and argument α radians, the map $R : \mathbb{C} \to \mathbb{C}$ defined by $R(z) = zw$ is a rotation through α radians. To see this, note that

$$|R(z)| = |zw| = |z| \cdot |w| = |z|,$$

so that $R(z)$ and z are at the same distance from the origin (they are both on the circle with center 0 and radius $|z|$), and

$$\arg(R(z)) = \arg(zw) = \arg(z) + \arg(w) = \arg(z) + \alpha \pmod{2\pi},$$

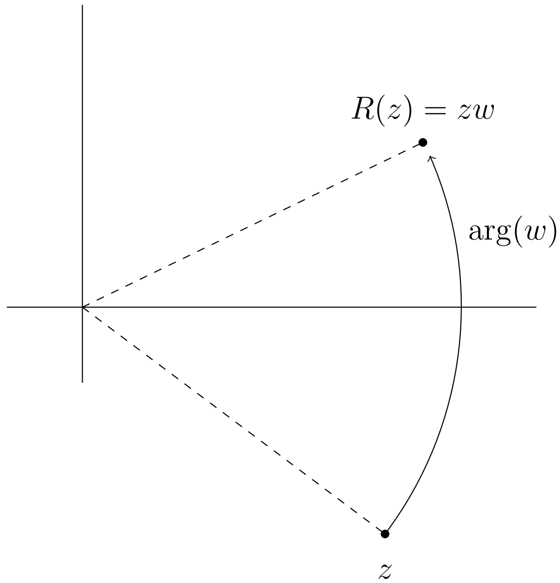


Figure 1.3: *Rotating* z *around the origin through* $\arg(w)$ *radians, where* w *has absolute value* 1.

so that the argument of $R(z)$ is α radians more than that of z. Another way of saying this is: if z is on the circle C with equation $|z| = r$, then $R(z)$ is also on C.

Similarly, if the absolute value of $w \neq 0$ differs from 1, then multiplication by w defines a transformation of the plane in which each complex number is rotated through $\arg(w)$ radians and is scaled by a factor $|w|$.

A circle with center z_0 and radius r consists of all complex numbers z satisfying

$$|z - z_0| = r.$$

Depending on the situation, alternative descriptions may be useful. Here are a few equvalent descriptions.

- $|z - z_0|^2 = r^2$ or $(z - z_0)(\bar{z} - \bar{z}_0) = r^2$, where in the last equation we have used the fact that $|w|^2 = w\bar{w}$ for every complex number w.

- If we write $z_0 = x_0 + iy_0$ and $z = x + iy$ (with x_0, y_0, x, y real), then $|z - z_0|^2 = r^2$ expands into the well-known 'real' equation of a circle: $(x - x_0)^2 + (y - y_0)^2 = r^2$.

- An explicit way of describing all points on the circle with equation $|z - z_0| = r$ is as follows: $z - z_0$ has absolute value r and is therefore of the form $r\,e^{it}$ (or $r(\cos t + i \sin t)$), where t is the argument (modulo 2π) of $z - z_0$. So we find that z can be described as $z_0 + re^{it}$. This is an example of a *parametric equation* of the circle $|z - z_0| = r$. (Of course, there are many, $z_0 + re^{2it}$, $z_0 - re^{it}$ are two more examples.)

For example, suppose you are asked to show that $\overline{w} = \dfrac{1}{w}$ for every complex number w on the circle $C : |z| = 1$, then you could proceed as follows. Let w be an arbitrary complex number on C, then w can be written as e^{it} for some real t. Then $\overline{w} = \overline{e^{it}} = e^{-it}$ by Theorem 1.2.6, and $\dfrac{1}{w} = \dfrac{1}{e^{it}} = e^{-it}$ by Theorem 1.2.14, and so we are done. (An alternative approach is to write w in the form $\cos t + i \sin t$, etc.)

1.4.6 Example. If the complex numbers z and w have the same absolute value ($\neq 0$) and if the angle between (the segments connecting 0 with) z and w is equal to α, then the argument of the quotient z/w is α or $-\alpha$ so that $z/w = e^{i\alpha}$ of $z/w = e^{-i\alpha}$. Another way of phrasing this is to say that $z = we^{i\alpha}$ or $z = we^{-i\alpha}$.

If, for instance, in $\triangle\alpha\beta\gamma$ (so a triangle with vertices α, β, γ) $\gamma - \alpha = e^{\pm\pi i/3}(\beta - \alpha)$, then this can be read as: the segment $[\alpha, \gamma]$ is obtained from the segment $[\alpha, \beta]$ by a rotation through $\pm\pi/3$ radians. In particular, these two segments have the same length:

$$|\gamma - \alpha| = |e^{\pm\pi i/3}(\beta - \alpha)| = |e^{\pm\pi i/3}| \cdot |(\beta - \alpha)| = |\beta - \alpha|.$$

And of course, the triangle is then equilateral by the congruence criterion SAS (side-angle-side). Here is a verification that $|\gamma - \beta| = |\beta - \alpha|(= |\gamma - \alpha|)$ using complex numbers. First rewrite $\gamma - \beta$ as follows:

$$\gamma - \beta = \gamma - \alpha + \alpha - \beta = e^{\pm\pi i/3}(\beta - \alpha) + \alpha - \beta = (e^{\pm\pi i/3} - 1)(\beta - \alpha).$$

Now note that $e^{\pm\pi i/3} - 1 = \dfrac{1}{2} \pm \dfrac{1}{2} i\sqrt{3} - 1 = -\dfrac{1}{2} \pm \dfrac{1}{2} i\sqrt{3}$ whose absolute value is 1. So

$$|\gamma - \alpha| = |(e^{\pm\pi i/3} - 1)(\beta - \alpha)| = |e^{\pm\pi i/3} - 1| \cdot |\beta - \alpha| = |\beta - \alpha|.$$

Note that this example really comes down to the fact that the complex numbers 0, $e^{\pi i/3}$ and $e^{\pi i/3} - 1$ are the vertices of an equilateral triangle.

1.4.7 Example. If $\triangle z_1z_2z_3$ is a triangle, then for every complex number $w \neq 0$ the triangles $\triangle z_1z_2z_3$ and $\triangle(wz_1)(wz_2)(wz_3)$ (multiply each vertex z_i by w) are similar. There are various ways to see this. One way is to compare lengths of corresponding sides (using the rules for absolute values):

$$\frac{|wz_2 - wz_1|}{|z_2 - z_1|} = \frac{|w| \cdot |z_2 - z_1|}{|z_2 - z_1|} = |w|,$$

and

$$\frac{|wz_3 - wz_1|}{|z_3 - z_1|} = \frac{|w| \cdot |z_3 - z_1|}{|z_3 - z_1|} = |w|,$$

and, similarly, $|wz_3 - wz_2| = |w| \cdot |z_3 - z_2|$. So the triangles are similar by the *sss* criterion (side-side-side).

Of course, you can also compare the angles of triangle $\triangle(wz_1)(wz_2)(wz_3)$ with those of triangle $\triangle z_1z_2z_3$, e.g., $\arg(\frac{wz_1}{wz_2}) = \arg(\frac{z_1}{z_2})$. So both triangles have the same angles and are therefore similar.

1.4.8 Example. Let $\triangle ABC$ be a triangle. Let $BCDE$ and $ACFG$ be two squares erected externally on the sides BC and AC, respectively, as in the illustration. Let H be the midpoint of DF. Prove that HC and AB are perpendicular. The idea in the following proof is to connect perpendicularity with

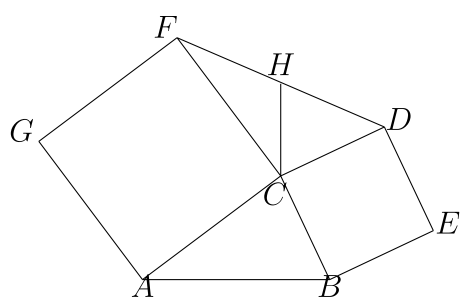


Figure 1.4: *Triangle $\triangle ABC$ with two squares.*

multiplication by i.

Put the origin in C (the point C seems central to the configuration, so looks like a reasonable choice to make the computations easier) and denote vertex A by the complex number z and vertex B by w. Then vertex D corresponds to iw (rotate B around C through $90°$) vertex F corresponds to

$-iz$ (rotate vertex A through $-90°$). The midpoint of segment DF is then $\frac{1}{2}(iw - iz)$. Since

$$\frac{1}{2}(iw - iz) = \frac{1}{2} \cdot i \cdot (w - z)$$

and since $w - z$ corresponds to segment AB, we find that HC is indeed perpendicular to AB (and has half its length).

1.4.9 The nine-point circle I

Let $\triangle ABC$ or, in complex terms, $\triangle\alpha\beta\gamma$ be a triangle, where the origin is chosen in the center of the circumcircle. Suppose that $|\alpha| = |\beta| = |\gamma| = 1$. The points $\frac{1}{2}(\beta + \gamma)$, $\frac{1}{2}(\alpha + \gamma)$ and $\frac{1}{2}(\alpha + \beta)$ are the midpoints D, E, F of the three sides BC, AC and AB, respectively. It follows from classical geometry that the segments connecting the origin with each of these midpoints are perpendicular to the corresponding sides of the triangle. The point $\frac{1}{3}(\alpha + \beta + \gamma)$ is the *centroid* Z of triangle $\triangle\alpha\beta\gamma$.

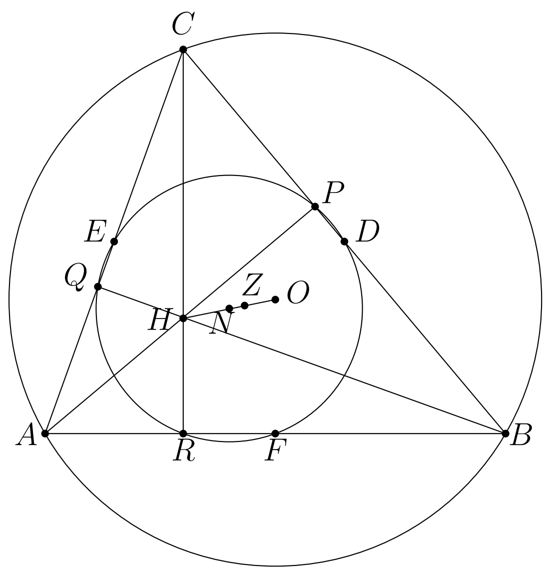


Figure 1.5: *The circumcircle of $\triangle ABC$ with center O, and the circumcircle of $\triangle DEF$ with center N. The centroid Z, the orthocenter H and the altitudes are also shown.*

The point $h = \alpha + \beta + \gamma$ is also special: $h - \gamma = \alpha + \beta = 2 \cdot \frac{1}{2}(\alpha + \beta)$, so

$h-\gamma$ and AB are perpendicular and $h-\gamma$ is twice as long as OF. So the point h is on the altitude from C. Similarly, h is on the altitudes through B and A, respectively. So the three altitudes are concurrent. Their common point H (or h in terms of complex numbers) is called the *orthocenter* of triangle $\triangle ABC$. (By the way: in every triangle the three altitudes are concurrent; the assumptions we have made are not restrictive; do you see why?)

The point $\frac{1}{2}(\alpha+\beta+\gamma)$ or $h/2$ is also special. To see this, consider the distances between this point and the three midpoints of the sides of $\triangle ABC$:

$$|\frac{1}{2}(\alpha+\beta+\gamma)-\frac{1}{2}(\beta+\gamma)| = |\frac{1}{2}\,\alpha| = \frac{1}{2}.$$

The distances to the midpoints E and F are equal to $\frac{1}{2}$ and so the point $h/2$ is the midpoint N of the circumcircle of triangle $\triangle DEF$ (see figure 2.15).

1.4.10 The nine-point circle II

The circle through D, E, and F turns out to pass through the midpoints of

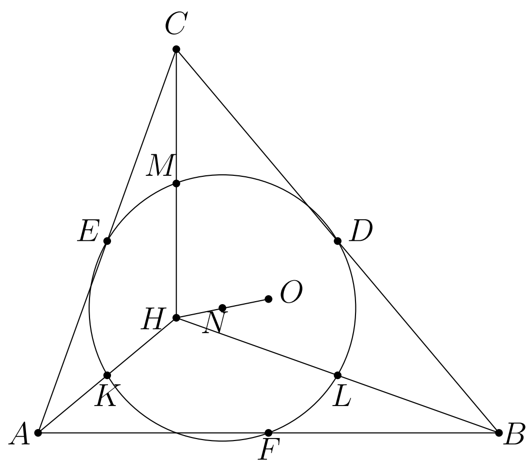


Figure 1.6: *The circumcircle of $\triangle DEF$ with center N also passes through the midpoints of segments AH, BH en CH.*

the three segments connecting h and the three vertices A, B and C, respectively. The distances of h to these midpoints is equal to $\frac{1}{2}$:

$$|\frac{1}{2}h - \frac{1}{2}(h+\alpha)| = \frac{1}{2}\,|-\alpha| = \frac{1}{2},$$

etc. By now we have: *the midpoints of the sides of* $\triangle ABC$ *and the midpoints of the segments* HA*,* HB *en* HC *lie on the same circle.*

This circle turns out to also pass through the three feet of the altitudes of $\triangle ABC$ (as figure 2.15 suggests). For this reason the circle is called the *nine-point circle* of $\triangle ABC$. The proof is discussed in exercise 24.

1.4.11 Other transformations

Translations, rotations and reflections (and their compositions) have the special property that they are bijections of the plane, i.e., they have inverses. There also exist transformations that are not bijections (and still useful), like the transformation sending z to z^2. We will not discuss them in this course.

1.5 Notes

More worked examples can be found in [6] (see the bibliography at the end of the lecture notes). The role of complex numbers in geometry is extensively discussed in [4].

Complex numbers have a rich history starting already in the 16th century. The search for formulas to solve polynomial equations of degrees 3 and higher really took off with people like Cardano, Tartaglia and Ferrari. Solutions to equations of degrees 3 and 4 can already be found in Cardano's *Ars magna* published in 1545. *Math History*

The construction of complex numbers is an example of the construction of an arithmetical system. Another example is the system $\mathbb{Z}/n\mathbb{Z}$ of integers modulo n. Such constructions are discussed in the various algebra courses. *Algebra*

A rigorous definition in terms of pairs of real numbers was given by Sir William Hamilton (1805–1865), see [1], p. 524. He defined the addition on such pairs by $(a, b)+(c, d) = (a+c, b+d)$, and the multiplication by $(a, b)\cdot(c, d) = (ac-bd, ad+bc)$. By agreeing to write a instead of (a, b) (for real a) and i for $(0, 1)$, we arrive at the usual notation $a + bi$. Hamilton's approach to define complex numbers in terms of the familiar real numbers contributed to the demystification of complex numbers. Hamilton generalized his construction to an arithmetical system with elements of the form $a + bi + cj + dk$ (with $a, b, c, d \in \mathbb{R}$), where $i^2 = -1$, $j^2 = -1$, $k^2 = -1$, $ij = k = -ji$, $jk = i = -kj$, $ki = j = -ik$. This is the famous arithmetical system of the *quaternions*.

Complex numbers are useful for linear algebra since they enable us to solve polynomial equations related to linear transformations, as will be discussed in Linear Algebra 2. Polynomials are discussed in more detail in the algebra courses. They play an important role in many branches of mathematics, ranging from numerical mathematics to cryptology. *Linear Algebra*

The Fundamental Theorem of Algebra has a long history in itself. It took many decades in the 18th and 19th century and the efforts of mathematicians like d'Alembert, Argand, Gauss to produce a rigorous proof (many candidate proofs contained a subtle gap which could only be filled after the development of rigorous analysis and topology), see [1]. A proof that uses complex integration is discussed in Complex Analysis. The fact that there do not exist explicit formulas for solving polynomial equations of degrees 5 and higher requires a substantial amount of algebra. *Complex Analysis*

The analysis of functions $f : \mathbb{C} \to \mathbb{C}$, i.e., limits, continuity, differentiation, integration, is also discussed in the course on complex analysis. Complex analysis is extensively used not only in mathematics, but also in electrical engineering and in mathematical physics.

1.6 Exercises

§1

1 Write each of the following complex numbers in the form $a+bi$ with a en b real:

a. $(2+3i)(1-i)$,

b. $(-\frac{1}{2}+\frac{1}{2}i\sqrt{3})(-\frac{1}{2}-\frac{1}{2}i\sqrt{3})$,

c. $\dfrac{1}{4-3i}$,

d. $\dfrac{7+i}{1+2i}$,

e. $\dfrac{9-3i}{1+3i}$,

f. $\dfrac{z}{(z+1)^2}$, met $z=\frac{1}{2}\sqrt{2}+\frac{1}{2}\sqrt{2}i$.

2 Write each of the following complex numbers in the form $r(\cos\varphi + i\sin\varphi)$, with $r>0$ en $-\pi \le \varphi \le \pi$, and draw these numbers in the complex plane:

a. -3,

b. $2i$,

c. $1+i$,

d. $\sqrt{3}+i$,

e. $5+12i$,

f. $4-4i$.

3 In the complex plane, draw the real and complex axes and an arbitrary complex number z (not on the real axis).

a. Sketch the following complex numbers and briefly describe how they are related geometrically to z:

$$z+2, \quad -2z, \quad \frac{1}{z}, \quad z-2i, \quad iz, \quad \overline{z}, \quad -i\overline{z}.$$

b. Similar question for: $z(\cos(\pi/2)-i\sin(\pi/2))$, $3z(\cos(7\pi/6)+i\sin(7\pi/6))$ and $z(\cos(2\pi/3)+i\sin(2\pi/3))$.

4 In the complex plane, draw the complex numbers $z \in \mathbb{C}$ that satisfy both

$$|z+1-i|^2 \le 2 \quad \text{and} \quad \frac{\pi}{2} \le \arg z \le \frac{3\pi}{4}\ .$$

5 Determine all complex numbers z that satisfy

a. $|z-i|=|z+3i|$,
b. $|z-3i|=|4+2i-z|$,
c. $\mathrm{Re}(z^2)=\mathrm{Im}(z^2)$,
d. $\mathrm{Re}(z^2+1)=0$ and $|z|=\sqrt{2}$,
e. $\arg(z/\overline{z})=\frac{2\pi}{3}$.

6 Prove the triangle inequality $|z_1+z_2|\leq|z_1|+|z_2|$ in the following steps.

a. Prove the inequality in the case $z_1=1$. To do this, write z_2 in the form $z_2=r(\cos t+i\sin t)$ and analyse $|1+r\cos t+ir\sin t|^2$.

b. Prove the inequality $|z_1+z_2|\leq|z_1|+|z_2|$ in the case $z_1\neq 0$ by dividing both sides by $|z_1|$.

Show furthermore that for all complex numbers $z_1,z_2,\ldots,z_n$ the following inequality holds:

$$|z_1+z_2+\cdots+z_n|\leq|z_1|+|z_2|+\cdots+|z_n|.$$

§2

7 Draw each of the following complex numbers in the plane and write them in the form $a+bi$ (with a,b real):
a. $2e^{\pi i/2}$,
b. $3e^{2\pi i/3}$,
c. $\sqrt{2}e^{\pi i/4}$,
d. $e^{5\pi i/3}$,
e. $e^{(-\pi i/3)+3}$,
f. $e^{-5\pi i/6+2k\pi i}$, $k\in\mathbf{Z}$.

8 Solve each of the following equations:
a. $e^z=1+i$,
b. $e^z=1+\sqrt{3}i$,
c. $e^{\mathrm{Re}(z)}=5$,
d. $e^{|z|}=1$,
e. $e^{-z^2}=-i$,
f. $e^{2iz}=\frac{1+i}{1-i}$.

9 Use the definitions of the complex cosine and sine to show each of the following statements.

a. $\sin 2z=2\sin z\cos z$ for all $z\in\mathbb{C}$.

b. $\cos 2z = \cos^2 z - \sin^2 z$ for all $z \in \mathbb{C}$.

10 Solve each of the following equations:

a. $\frac{1}{2}(e^{iz} + e^{-iz}) = 0$,

b. $\sin(2z) = 4$.

§3

11 Solve each of the following equations and draw the solutions in the complex plane.

a. $z^6 = 1$,

b. $z^3 = 8$,

c. $z^4 = 16i$,

d. $(z+i)^4 = -1$,

e. $(z+2-i)^6 = 27i$,

f. $z^2 = \overline{z}$,

g. $z^3 = -\overline{z}$.

12 Solve each of the following equations and draw the solutions in the complex plane.

a. $z^2 + z + 1 = 0$,

b. $z^2 - 2iz + 8 = 0$,

c. $z^2 - (4+2i)z + 3 + 4i = 0$,

d. $z^2(i + z^2) = -6$.

13 a. The equation $z^3 + (2-3i)z^2 + (-2-6i)z - 4 = 0$ has a solution $z = i$. Determine the other solutions.

b. The equation $z^4 + 4z^3 + 3z^2 - 14z + 26 = 0$ has a solution $z = 1 + i$. Determine the other solutions.

c. Suppose 5 and $1+2i$ are zeros of degree 3 polynomial with real coefficients. Determine such a polynomial.

d. Suppose i and $2 - 3i$ are zeros of a degree 4 polynomial with real coefficients. Determine such a polynomial.

14 Factor in real factors of lowest possible degrees:

a. $z^4 - 3z^2 - 4$,

b. $z^3 + 3z^2 + 4z + 2$,

c. $z^4 + z^3 + 2z^2 + z + 1$.

15 a. Compute $(1+i)^{11}$.

b. Suppose the complex number z satisfies

$$z^4 = 8\sqrt{3} + 8i \quad \text{and} \quad \frac{\pi}{2} \le \arg(z) \le \pi.$$

Determine the exact values of $|z^{23}|$ and $\arg(z^{23})$ (the argument taken in the interval $[0, 2\pi]$).

16 Prove that for all positive integers n *De Moivre's formula* holds for real φ:

$$(\cos\varphi + i\sin\varphi)^n = \cos n\varphi + i\sin n\varphi.$$

(After the French mathematician A. de Moivre (1667–1754)). Use it to express $\cos 3\varphi$ and $\sin 4\varphi$ in terms of $\cos\varphi$ and $\sin\varphi$.

17 Let $p(z) = az^2 + bz + c$ be a complex polynomial with $a \neq 0$. Prove the following statement using the steps outlined below: The polynomial $p(z)$ has a zero with multiplicity 2 if and only if $b^2 - 4ac = 0$.

- First part: *If p has a zero with multiplicity* 2*, then* $b^2 - 4ac = 0$.
  So suppose that λ is such a zero of $p(z)$. Now show:

  1) $p(z) = az^2 + bz + c = a(z - \lambda)^2$.
  2) Express b and c in terms of λ and verify that $b^2 - 4ac = 0$.

- Second part: *If* $b^2 - 4ac = 0$*, then $p(z)$ has a zero with multiplicity* 2.
  Complete the square in $p(z) = az^2 + bz + c$ and use $b^2 - 4ac = 0$.

§4

18 Prove each of the following statements.

a. The complex number z is real if and only if $\overline{z} = z$.

b. The complex number z is purely imaginary (i.e., real part equal to 0) if and only if $z + \overline{z} = 0$.

c. The (segments connecting 0 with the) complex numbers z and w (both $\neq 0$) are parallel (their arguments differ by an integral multiple of π) if and only if $\overline{z}w = z\overline{w}$. [Hint: analyse the quotient z/w.]

d. The (segments connecting 0 with the) complex numbers z and w (both $\neq 0$) are perpendicular (their arguments differ by $\pi/2$ up to multiples of π) if and only if $\overline{z}w + z\overline{w} = 0$. [Hint: what can you say of the quotient z/w if z are w perpendicular?]

19 (**Reflections**) In terms of complex numbers complex conjugation describes a reflection in the real axis: $z = x + iy$ is mapped to $x - iy$ or $\overline{z}$.

a. Show that a reflection in the imaginary axis maps z into $-\overline{z}$.

b. A reflection of z in a line through the origin making an angle of α radians with the positive real axis can be described as follows: first rotate z around the origin through $-\alpha$ radians, then reflect the result in the real axis, and, finally, rotate through α radians. Describe the resulting complex number in terms of z.

c. The angle between the lines ℓ and m through the origin is α radians. We first reflect z in ℓ and then the result in m. Show that the composition of these two reflections is a rotation through 2α radians around the origin. [Hint: assume the angle between ℓ and the positive real axis is β radians, and the angle between m and the real axis is $\beta + \alpha$ radians.]

20 Let $\triangle ABC$ be an equilateral triangle, whose vertices A, B, C correspond to the complex numbers α, β, γ, respectively.

a. In this item we put the origin in A. Show that the vertices can be represented in the following way: 0, z, $\exp(\pi i/3)\, z$.

b. Let $\rho = e^{\pi i/3}$ and let $\omega = \rho^2$. Verify that $\rho^3 = -1$, and $1 - \rho + \rho^2 = 0$, and $1 + \omega + \omega^2 = 0$.

c. From this item onwards, the origin is not necessarily located in one of the vertices. Prove that $\gamma - \alpha = \rho(\beta - \alpha)$ or $\gamma - \alpha = \bar{\rho}(\beta - \alpha)$.

d. Prove that $\alpha + \omega\beta + \omega^2\gamma = 0$ or $\alpha + \omega^2\beta + \omega\gamma = 0$ if $\triangle ABC$ is equilateral.

e. Prove that $\triangle ABC$ is equilateral if $\alpha + \omega\beta + \omega^2\gamma = 0$ or $\alpha + \omega^2\beta + \omega\gamma = 0$.

21 Let ℓ be the line through the two distinct complex numbers v and w. Then ℓ consists of all complex numbers of the form $v + t(w - v)$ with t real.

a. Prove: if, for a complex number z with $z \neq v$ and $z \neq w$, the quotient $\dfrac{z-w}{z-v}$ is a real number, say t, then z is on the line ℓ.

b. Prove: if z, distinct from v and w, is on ℓ, then the quotient $\dfrac{z-w}{z-v}$ is a real number.

c. Prove that v, w, z are collinear (lie on one line) if and only if

$$(z - w)(\bar{z} - \bar{v}) = (\bar{z} - \bar{w})(z - v).$$

22 Suppose $ABCD$ and $AB'C'D'$ are two squares in the plane that a) have vertex A in common, b) have the same orientation of the vertices, and c) lie outside one another. Let P be the intersection of the diagonals AC and BD; let Q be the intersection of the diagonals AC' and $B'D'$; let R be the midpoint of the segmen BD', and let S be the midpoint of the diagonal $B'D$. Prove that $PQRS$ is a square by first showing that segment PS transforms into PR by a rotation through 90°. (Do not denote complex numbers corresponding to P, etc., by P, etc.; use for instance corresponding small letters.)

23 Reflecting in the line through u and v
Let ℓ be the line in the complex plane through the points u and v. In this exercise we determine the mirror image of z when we reflect z in ℓ.

a. Suppose $u = 0$. Show that we can write z in the form $z = re^{it} \cdot v$ for some real r and t. What is the mirror image of z in this case?

b. Back to the general case: show that z can be written as $u + re^{it} \cdot (v - u)$. Use this to show that the mirror image of z is equal to

$$u + (\bar{z} - \bar{u})\frac{v - u}{\bar{v} - \bar{u}}\,.$$

c. Show that this expression simplifies to

$$u + v - uv\bar{z}$$

if $|u| = |v| = 1$. [Hint: use that $1/u = \overline{u}$.]

24 The nine-point circle

The nine-point circle, (see 1.4.9 and 1.4.10) also passes through the feet of the three altitudes. In this exercise we discuss a proof of this fact.

a. The altitude from A intersects the circumcircle of $\triangle ABC$ in A'. Show that the corresponding complex number α'

$$\frac{\alpha - \alpha'}{\beta - \gamma} + \frac{\bar{\alpha} - \bar{\alpha}'}{\bar{\beta} - \bar{\gamma}} = 0.$$

[Hint: since $\alpha - \alpha'$ and $\beta - \gamma$ are perpendicular, the quotient $\dfrac{\alpha - \alpha'}{\beta - \gamma}$ is purely imaginary.]

b. Now use $\bar{\alpha} = 1/\alpha$, $\bar{\beta} = 1/\beta$, etc., to deduce that

$$\alpha' = -\frac{\beta\,\gamma}{\alpha}.$$

[Note: an alternative approach would be to compute the mirror image of $h = \alpha + \beta + \gamma$ in the line AB with the formula from exercise 23c) and to verify that this mirror image in on the circumcircle of $\triangle ABC$.]

c. Show that the segments BH and BA' have the same length.

d. Conclude that the foot P of the altitude from A is

$$\frac{1}{2}(h - \frac{\beta\,\gamma}{\alpha}).$$

e. Show that the distance between $h/2$ and P equals $\dfrac{1}{2}$. Conclude that the nine-point circle passes through the three feet P, Q, R of the altitudes.

1.6.1 Exercises from old exams

25 Determine all complex numbers z that satisfy

$$\left|\frac{\bar{z} \cdot z}{(1-z)^2}\right| = 1.$$

26 Solve the following equation in $\mathbb{C}$:

$$e^{2iz} = \frac{1+i}{1-i}.$$

27 a. Sketch the set of $z \in \mathbb{C}$ that satisfy

$$|\arg(z)| = \frac{\pi}{4},$$

and the set of $z \in \mathbb{C}$ that satisfy

$$|z + 2i| = |z - 3|.$$

b. Determine all $z \in \mathbb{C}$ satisfying

$$|\arg(z)| = \frac{\pi}{4} \quad \text{and} \quad |z + 2i| = |z - 3|.$$

28 The complex number $1+2i$ is a zero of the polynomial $z^4-2z^3+9z^2-8z+20$. Find the factorization in factors of lowest possible degrees and determine the remaining zeros.

29 Let $p(z)$ be a complex polynomial. Prove that $p(z)$ is a real polynomial (i.e., all its coefficients are real) if and only if $p(z) = \overline{p(\overline{z})}$ for all $z \in \mathbb{C}$.

30 Solve in $\mathbb{C}$:

$$z^3 = i\,\overline{z}.$$

31 Suppose the squares $ABCD$ and $A'B'C'D'$ have the same orientation (so going from A to B to C to D and going from A' to B' to C' to D' is both clockwise or counterclockwise). Prove that the midpoints of the segments AA', BB', CC', and DD' are the vertices of a square.

Chapter 2

Vectors in two and three dimensions

2.1 Vectors in dimensions two and three

2.1.1 The mathematical concept of a vector was originally introduced to represent quantities that have both magnitude and direction. Velocity and force are two examples of such quantities in physics. In this chapter we will consider vectors in the plane and in space. The notions we encounter here will be generalized to abstract vector spaces in later chapters.

The crucial point with vectors is that they can be added and multiplied by numbers. This opens the way to applying algebra in the plane and in space. In this chapter this arithmetic of vectors is discussed, the use of vectors in describing lines and planes and their relative positions (including distances, angles, and the cross product). Also, the use of vectors in geometric problems is briefly discussed. Like complex numbers, vectors provide a tool for dealing with geometry.

2.1.2 Vectors

A vector corresponds with an arrow in the plane or in space, and is determined by its direction and its magnitude (length). Therefore, an arrow, translated parallel to itself to any point in space but with the same direction and magnitude, is considered to represent the same vector. Such translated arrows are called equivalent, i.e. represent the same vector[1].

[1]Don't confuse this with a force vector (or any other vector valued quantity) applied

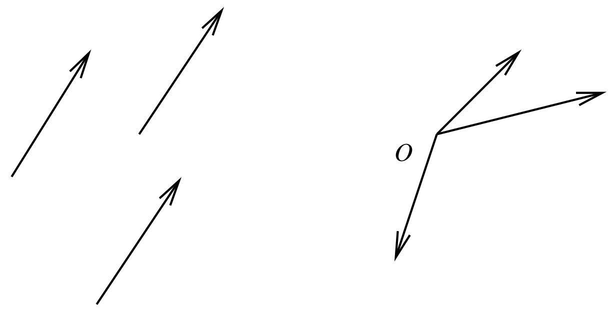


Figure 2.1: *On the left, representations of the same vector are drawn: direction and length of each arrow are the same, but the heads and tails differ. On the right, different vectors are drawn with the same starting point, namely a chosen origin O in the plane.*

In the sequel, we will usually choose an origin in the plane or space, and assume our vectors start there. Sometimes we will deviate from this convention. This will hardly lead to any confusion.

In these notes, vectors will be denoted by an underlined letter as follows:

$$\underline{v}.$$

In the literature other notations are used, such as $\mathbf{v}$, $\vec{v}$ or $\bar{v}$.

2.1.3 The zero vector

There is one special vector which has no direction, and whose length equals 0. This vector is called the *zero vector* and is denoted by $\underline{0}$.

2.1.4 Scalar multiplication

Let $\underline{v}$ be a vector, and λ a real number. Then $\lambda\underline{v}$ denotes the *scalar product* of λ and $\underline{v}$, and is defined as the vector that points in the same (if $\lambda > 0$) or opposite (if $\lambda < 0$) direction as $\underline{v}$, and whose length equals $|\lambda|$ times the length of $\underline{v}$. For $\lambda = 0$ we define $\lambda\underline{v} = \underline{0}$, the zero vector. We also call $\lambda\underline{v}$ a *scalar multiple* or *multiple* of $\underline{v}$. The real number λ is called a *scalar*. Scalars are often denoted by Greek letters, but this is only a tradition and not necessary. Sometimes a multiplication symbol is used for clarity as in $3 \cdot \underline{v}$.

to a physical point in space! Although the force has, as a vector, many mathematical representations, there is physically only one force with one point of application, and we cannot translate the force to another point without changing the physics.

Furthermore, we write $\underline{v}$ for $1\,\underline{v}$, $-\underline{v}$ for $(-1)\underline{v}$, $-3\underline{v}$ for $(-3)\underline{v}$, etc. The vector $-\underline{v}$ is called the *opposite* of $\underline{v}$.

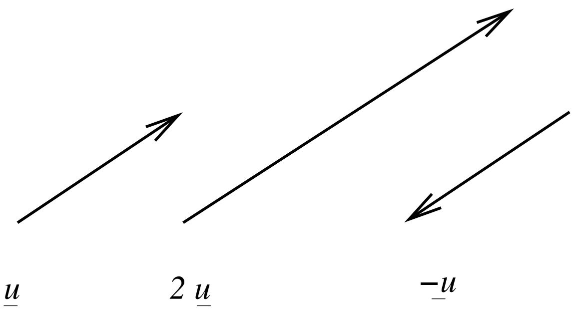


Figure 2.2: *Scalar multiplication.*

For any vector $\underline{v}$ and scalars λ and μ, we have:

- $0 \cdot \underline{v} = \underline{0}$;
- $\lambda(\mu\underline{v}) = (\lambda\mu)\underline{v}$.

  In words: if the vector $\underline{v}$ is first multiplied by μ and the resulting vector is multiplied by λ, then the result equals the scalar product of the scalar $\lambda\mu$ and the vector $\underline{v}$.

2.1.5 The addition of vectors

If $\underline{u}$ and $\underline{v}$ are two vectors starting from the same point (by translating this can always be arranged), then their sum $\underline{u} + \underline{v}$ is by definition the vector which starts from that same point and ends at the point equal to the fourth point of the parallelogram spanned by $\underline{u}$ and $\underline{v}$. The sum $\underline{u}$ en $\underline{v}$ can also be obtained by joining the starting point (or tail) of $\underline{v}$ to the endpoint (or head) of $\underline{u}$, or vice versa. If the vectors have the same or opposite directions, then only this second construction works. Note furthermore that $\underline{u} + \underline{0} = \underline{u}$ for every vector $\underline{u}$.

Here are the arithmetic rules and some remarks regarding this addition (no proofs):

- **Associativity** of the addition:

$$(\underline{u} + \underline{v}) + \underline{w} = \underline{u} + (\underline{v} + \underline{w})$$

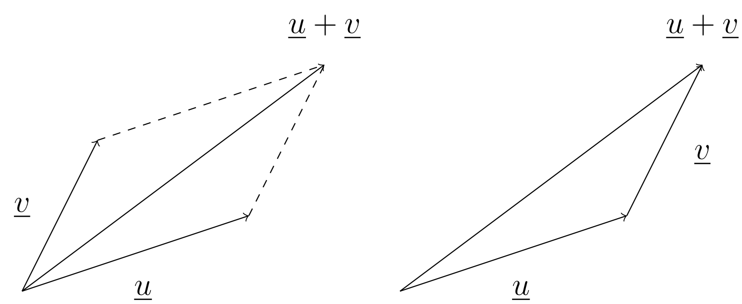


Figure 2.3: *Addition of vectors: on the left the construction using a parallelogram, on the right the head-to-tail construction, joining the tail of the second vector to the head of the first one.*

for all vectors $\underline{u}$, $\underline{v}$, $\underline{w}$. Note that the addition is only defined for *two* vectors, and not for three or more. So if you want to add three vectors, you will have to split the problem in various additions of two vectors. For instance, you could add the first and second vector, and then add the result to the third vector, so this corresponds to $(\underline{u}+\underline{v})+\underline{w}$. Associativity tells you that it doesn't matter in which way you split the problem, the answer is always the same. That's the (justified) reason we often leave out the brackets (we sometimes put in brackets to clarify calculations for the reader). So we often simply write $\underline{v}_1 + \underline{v}_2 + \underline{v}_3 + \underline{v}_4$ for the addition of four vectors instead of, say,

$$\underline{v}_1 + ((\underline{v}_2 + \underline{v}_3) + \underline{v}_4).$$

- **Commutativity** of the addition:

$$\underline{v} + \underline{w} = \underline{w} + \underline{v}$$

for all vectors $\underline{v}$ and $\underline{w}$. This is obvious from the parallellogram construction. It implies that you can change whenever needed the order of the vectors in additions. For instance, $\underline{u} + \underline{v} + \underline{w} = \underline{w} + \underline{u} + \underline{v}$. Here is how this specific equality follows from commutativity:

$$\underline{u} + \underline{v} + \underline{w} = \underline{u} + \underline{w} + \underline{v} = \underline{w} + \underline{u} + \underline{v}.$$

From now on, you don't have to supply such proofs any time you use commutativity, unless a proof is explicitly asked for.

- Instead of $\underline{v} + -\underline{w}$ we usually write $\underline{v} - \underline{w}$ (*subtraction* of vectors).

Here are the arithmetic rules that involve both addition and scalar multiplication.

- **Distributivity** of the scalar multiplication over addition:
$$\lambda(\underline{v} + \underline{w}) = \lambda\underline{v} + \lambda\underline{w}$$
for all vectors $\underline{v}$, $\underline{w}$ and for all scalars λ.

- **Distributivity** of the scalar addition over the scalar multiplication:
$$(\lambda + \mu)\underline{v} = \lambda\underline{v} + \mu\underline{v}$$
for all scalars λ, μ and all vectors $\underline{v}$.

The sum of any vector $\underline{v}$ and its opposite $-\underline{v}$ always yields the zero vector:
$$\underline{v} - \underline{v} = \underline{0}.$$

2.1.6 Linear combinations

If $\underline{v}_1, \underline{v}_2, \ldots, \underline{v}_n$ are n vectors and $\lambda_1, \lambda_2, \ldots, \lambda_n$ are n real numbers (scalars), then the vector
$$\lambda_1\underline{v}_1 + \lambda_2\underline{v}_2 + \cdots + \lambda_n\underline{v}_n$$
is called a *linear combination* of the vectors $\underline{v}_1, \underline{v}_2, \ldots, \underline{v}_n$. Linear combinations are the vectors we can build out of a given set of vectors using addition and scalar multiplication.

So $2\underline{u} - 3\underline{v} + 2\underline{w}$ is a linear combination of $\underline{u}, \underline{v}, \underline{w}$.

2.1.7 Examples. The following examples show that computations with vectors involving addition and scalar multiplication only are fairly easy. Note that we cannot multiply two (or more) vectors (but see §2.5).

a) $3\underline{v} - \underline{w} + 2\underline{v} + 3\underline{w} = 5\underline{v} + 2\underline{w}$. Here are the detailed steps, using the various arithmetic rules. By commutativity
$$3\underline{v} - \underline{w} + 2\underline{v} + 3\underline{w} = 3\underline{v} + 2\underline{v} - \underline{w} + 3\underline{w}.$$
Next, distributivity and the fact that $-\underline{w} = (-1)\underline{w}$ imply
$$3\underline{v} + 2\underline{v} - \underline{w} + 3\underline{w} = (3 + 2)\underline{v} + (-1 + 3)\underline{w} = 5\underline{v} + 2\underline{w}.$$
Note that because of associativity we didn't place brackets. Otherwise, the first step of the computation would have looked like:
$$(3\underline{v}+(-\underline{w}+2\underline{v}))+3\underline{w} = (3\underline{v}+(2\underline{v}-\underline{w}))+3\underline{w} = ((3\underline{v}+2\underline{v})-\underline{w})+3\underline{w} = \ldots$$

b) The opposite of $\lambda \underline{v}$ is $-\lambda\underline{v}$. Here, λ is an arbitrary scalar. A proof could run as follows: $\lambda \underline{v} + -\lambda\underline{v} = (\lambda - \lambda)\underline{v} = 0 \cdot \underline{v} = \underline{0}$.

c) Do you see why $\underline{v} + \frac{1}{2}(\underline{w} - \underline{v}) = \frac{1}{2}(\underline{v} + \underline{w})$? This equality provides two ways to consider the midpoint of a segment. Do you see which ones?

d) By using the various arithmetic rules, you find:

$$(-\underline{u} + 2\underline{v} + 3\underline{w}) + (2\underline{u} - \underline{v} + \underline{w}) = \underline{u} + \underline{v} + 4\underline{w}.$$

2.2 Vector descriptions of lines and planes

2.2.1 By choosing an origin O, every point P corresponds uniquely with the vector p that starts in O and ends in P. On the other hand, every vector $\underline{p}$ determines uniquely the point P given by its head, if we let $\underline{p}$ start in O. In this way we have a correspondence between points and vectors. Sometimes, when there is no risk of misinterpretation, we will make no explicit distinction between a point and the corresponding vector. From here on we will assume that an origin O is defined. This point corresponds with the zero vector $\underline{0}$.

2.2.2 Lines

The scalar multiples $\underline{x} = \lambda\underline{v}$ of a vector $\underline{v}$, with $\underline{v} \neq \underline{0}$, run through all points (vectors) of a straight line ℓ through the origin.

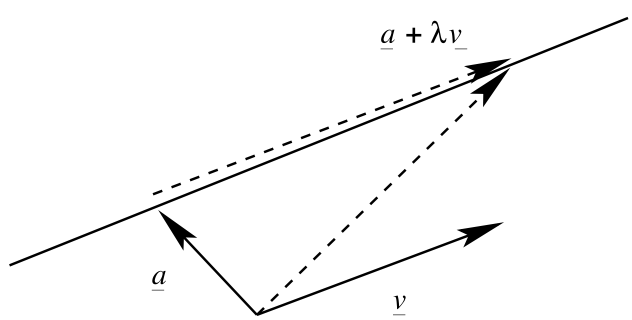


Figure 2.4: *A parametric representation of a straight line with supporting vector $\underline{a}$ nd direction vector $\underline{v}$. Each vector on the line can be obtained by adding an appropriate scalar multiple of $\underline{v}$ to $\underline{a}$.*

If $\underline{a}$ is a second vector, then for any λ,

$$\underline{x} = \underline{a} + \lambda\underline{v}$$

is a point on a straight line m through $\underline{a}$ and parallel with ℓ. We call

$$\ell : \underline{x} = \lambda\underline{v} \text{ and } m : \underline{x} = \underline{a} + \lambda\underline{v}$$

a *parametric representations* or *vector representations* of the lines ℓ and m, respectively. The vector $\underline{v}$ is a so-called *direction vector*. The vector $\underline{a}$ is a so-called *supporting vector* (or position vector) of the line m (we may call $\underline{0}$ a supporting vector of the line ℓ). The scalar λ is called a *parameter*. Of course, instead of λ any other letter can be used.

Summarizing: for a parameter or vector representation of a straight line we need a vector (with its endpoint) on the line and a direction vector. Note that both vectors are not unique. Every non-zero scalar multiple $\mu\underline{v}$ of the direction vector $\underline{v}$ is also a possible direction vector. Any vector $\underline{b}$ (with endpoint) on the line m can be used as supporting vector.

2.2.3 Example. (Supporting and direction vectors of a line are not unique) The vector $\underline{p} + \underline{v}$ is on the line ℓ with parametric description $\underline{x} = \underline{p} + \lambda\underline{v}$: just take $\lambda = 1$. This vector $\underline{p} + \underline{v}$ may serve as a supporting vector of ℓ, since the vectors $\underline{p} + \underline{v} + \mu\underline{v}$ run through the same vectors for varying μ as the vectors $\underline{p} + \lambda\underline{v}$ (for varying λ). This follows easily from the equalities $\underline{p}+\underline{v}+\mu\underline{v} = \underline{p}+(1+\mu)\underline{v}$ and $\underline{p}+\lambda\underline{v} = \underline{p}+\underline{v}+(\lambda-1)\underline{v}$. In fact, every vector on ℓ may serve as a supporting vector.

Similarly, $2\underline{v}$, $-3\underline{v}$, $\pi\,\underline{v}$ are direction vectors of ℓ. For instance, the vectors $\underline{p} + \mu(2\underline{v})$ run through the vectors of ℓ for varying μ.

2.2.4 Planes

Planes in space can also be represented in terms of a vector or parametric representation. For this we need one vector whose endpoint is in the plane (a supporting vector) and two direction vectors which are not scalar multiples of each other. Since we use two direction vectors, we also need two parameters.

The plane U through the origin and with direction vectors $\underline{u}$ and $\underline{v}$ has the following parametric description:

$$U : \underline{x} = \lambda\underline{u} + \mu\underline{v}.$$

The plane V with supporting vector $\underline{a}$ and direction vectors $\underline{u}$ and $\underline{v}$ has the following parametric representation:

$$V : \underline{x} = \underline{a} + \lambda\underline{u} + \mu\underline{v}.$$

Just as with lines, neither the supporting vectors nor the direction vectors are uniquely determined.

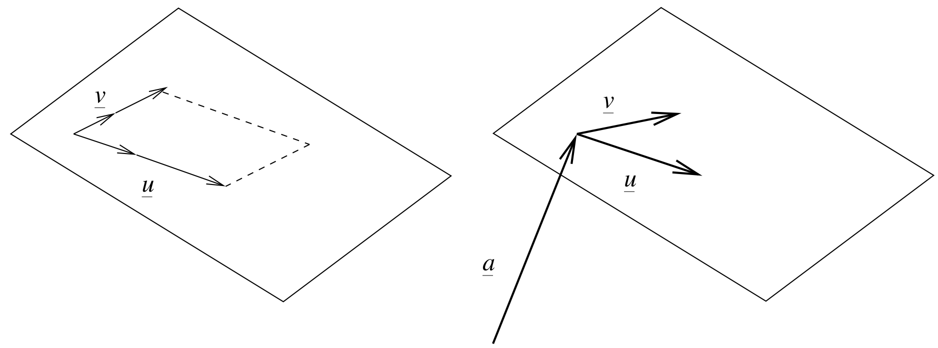


Figure 2.5: *On the left a plane through the origin. On the right a plane through $\underline{a}$ with direction vectors $\underline{u}$ and $\underline{v}$.*

2.2.5 Example. (Supporting and direction vectors of planes are not unique)
The plane V with parametric equation $V : \underline{x} = \underline{a} + \lambda\underline{u} + \mu\underline{v}$ can, for instance, also be described in the following way:

$$\underline{x} = \underline{a} + \rho(\underline{u} + \underline{v}) + \sigma(\underline{u} - \underline{v}).$$

To see this we have to verify that every vector of the form $\underline{a} + \lambda\underline{u} + \mu\underline{v}$ can also be written in the form $\underline{a} + \rho(\underline{u} + \underline{v}) + \sigma(\underline{u} - \underline{v})$, and vice versa. The following two equalities show this:

$$\begin{aligned} &\underline{a} + \lambda\underline{u} + \mu\underline{v} = \underline{a} + \tfrac{1}{2}(\lambda + \mu)(\underline{u} + \underline{v}) + \tfrac{1}{2})(\lambda - \mu)(\underline{u} - \underline{v}) \\ &\underline{a} + \rho(\underline{u} + \underline{v}) + \sigma(\underline{u} - \underline{v}) = \underline{a} = (\rho + \sigma)\underline{u} + (\rho - \sigma)\underline{v}. \end{aligned}$$

In fact, one can prove in a similar way that any two linear combinations of $\underline{u}$ and $\underline{v}$ that are not multiples of one another, may serve as direction vectors. For example, the pair $2\underline{u} + 3\underline{v}$, $2\underline{u} - 5\underline{v}$ is such a couple.

As with lines, any vector on V can serve as supporting vector of V. For example, $\underline{a} + 3\underline{u} + 5\underline{v}$ is such a vector.

2.3 Bases, coordinates, and equations

2.3.1 To be able to do concrete computations, it is useful to describe vectors using numbers (in their role as coordinates). For this purpose we need the notions *basis* and *coordinates*.

2.3.2 Basis

- **The plane**
  In the plane we need two vectors which are not multiples of each other,

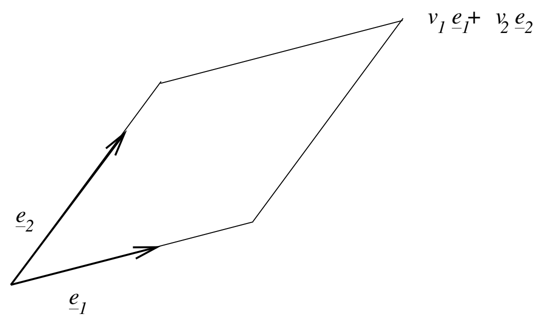


Figure 2.6: *Using the basis $\underline{e}_1$, $\underline{e}_2$ any vector in the plane can be described with the use of two coordinates.*

  for instance $\underline{e}_1$ and $\underline{e}_2$. Every vector $\underline{v}$ in the plane can be expressed in a unique way as a linear combination of the vectors $\underline{e}_1$, $\underline{e}_2$,

$$\underline{v} = v_1\,\underline{e}_1 + v_2\,\underline{e}_2,$$

  for some scalars v_1 and v_2, unique determined by $\underline{v}$.

- **3-dimensional space**
  In space we choose three vectors $\underline{e}_1$, $\underline{e}_2$, $\underline{e}_3$ that are not coplanar (i.e., whose endpoints do not lie in one plane with the origin). Then any vector $\underline{x}$ can be written as a linear combination of these three vectors:

$$\underline{v} = v_1\,\underline{e}_1 + v_2\,\underline{e}_2 + v_3\,\underline{e}_3,$$

  for unique scalars v_1, v_2 en v_3 ($v_1\underline{e}_1$ is a kind of 'projection' of $\underline{v}$ onto the line $\underline{x} = \lambda\underline{e}_1$, so that v_1 is determined, etc.).

  The vectors $\underline{e}_1$, $\underline{e}_2$, $\underline{e}_3$ are said to form a basis *basis* of space and the numbers v_1, v_2, v_3 are called the *coordinates* of the vector $\underline{v}$ with respect to this basis. If the vectors $\underline{e}_1$, $\underline{e}_2$, $\underline{e}_3$ are mutually perpendicular and have length 1, then the basis is called an *orthonormal basis*.

2.3.3 The vector spaces $\mathbb{R}^2$ and $\mathbb{R}^3$

Via coordinates every vector $\underline{v}$ in space corresponds to a unique triple of coordinates, v_1, v_2, v_3, say. The triple is usually denoted as an element of

$\mathbb{R}^3$. Such a triple is called a *coordinate vector*. The addition and scalar multiplication translate as follows into coordinates:

$$\begin{array}{lcl} \underline{v}+\underline{w} & \leftrightarrow & (v_1+w_1, v_2+w_2, v_3+w_3) \\ \lambda\,\underline{v} & \leftrightarrow & (\lambda v_1, \lambda v_2, \lambda v_3) \end{array}$$

(where $\underline{w}$ corresponds to (w_1, w_2, w_3)). With the choice of a basis, we have *coordinatized* space with the set $\mathbb{R}^3$. Note that the coordinates depend on the specific basis (and position of the origin). In this *coordinate space* we can add *coordinate vectors* and multiply them by scalars. In Chapter 4 we will see that space and its coordinate space are special cases of a *vector space.* In a similar way, $\mathbb{R}^2$ provides coordinates for vectors in the plane. For us, main role of the coordinate plane $\mathbb{R}^2$ and the coordinate space $\mathbb{R}^3$ is to able to translate vector problems into problems with numbers. The zero vector in the plane and in space, respectively, correspond to $(0,0)$ and $(0,0,0)$, respectively.

In practice we often 'identify' the coordinate space (plane) with the space (plane) of vectors itself. We then speak of the line in $\mathbb{R}^2$ or $\mathbb{R}^3$, vectors in $\mathbb{R}^2$ or $\mathbb{R}^3$, the plane $\mathbb{R}^2$, the space $\mathbb{R}^3$, a parametric equation of a line in $\mathbb{R}^2$, etc. We will write $\underline{a} = (a_1, a_2, a_3)$ if (a_1, a_2, a_3) is the coordinate vector of $\underline{a}$, even though this is strictly speaking not correct.

2.3.4 Describing lines in the plane with coordinates

Suppose $\ell : \underline{x} = \underline{a} + \lambda\underline{v}$ is a (vector parametric equation of a) line in the plane, and suppose $\underline{e}_1$, $\underline{e}_2$ is a basis of the plane. Let (a_1, a_2) and (v_1, v_2) be the coordinate vectors of the supporting vector $\underline{a}$ and the direction vector $\underline{v}$, then, in terms of coordinates, we find the following parametric equation for ℓ:

$$\ell : (x_1, x_2) = (a_1, a_2) + \lambda(v_1, v_2).$$

Sometimes it is useful to use column notation:

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} + \lambda \begin{pmatrix} v_1 \\ v_2 \end{pmatrix}.$$

Still another way is to write $x_1 = a_1 + \lambda v_1$ and $x_2 = a_2 + \lambda v_2$.

By eliminating λ from these two relations, we obtain an equation for ℓ. Multiply both sides of $x_1 = a_1 + \lambda v_1$ by v_2 and both sides of $x_2 = a_2 + \lambda v_2$ by v_1 and subtract:

$$v_2 x_1 - v_1 x_2 = v_2 a_1 - v_1 a_2.$$

This yields a linear equation in the variables x_1 and x_2. Lines do not have unique equations. For instance, the equations $x_1+2x_2 = 3$ and $2x_1+4x_2 = 6$ obviously describe the same line.In fact, multiplying both sides of an equation by the same nonzero scalar doesn't change the solution set.

Note that a vector parametric equation $\ell : \underline{x} = \underline{a} + \lambda\underline{v}$ of a line gives an explicit description of the vectors on the line: every value of λ produces a vector (or coordinate vector) on the line.

An equation describes the vectors on the line implicitly: you only know which relation the coordinates of a vector need to satisfy in order to be the coordinates of a vector on the line.

2.3.5 Describing lines in space with coordinates

A parametric equation $\ell : \underline{x} = \underline{a}+\lambda\underline{v}$ with $\underline{a} = (a_1, a_2, a_3)$ and $\underline{v} = (v_1, v_2, v_3)$ is

$$\ell : (x_1, x_2, x_3) = (a_1, a_2, a_3) + \lambda(v_1, v_2, v_3)$$

or, in column notation:

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix} + \lambda \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix}.$$

A line in space can also be described using two linear equations, because a line can be seen as the intersection of two planes and every plane can be described by a linear equation (extensive details on this follow in Chapter 4). For instance, the system $x_1 + x_2 + x_3 = 1$, $2x_1 - x_3 = 0$ describes the line $\underline{x} = (0, 1, 0) + \lambda(1, -3, 2)$ (by substitution you can verify that every vector satisfies both equations). A way to find this parametric equation from the two linear equations is to choose x_1 as parameter, call it λ, and then deduce that $x_3 = 2\lambda$ and $x_2 = 1 - x_1 - x_3 = 1 - 3\lambda$. The computational techniques behind this will be discussed in Chapter 3.

2.3.6 Describing planes in space in terms of coordinates

A parametric description $V : \underline{x} = \underline{a} + \lambda\underline{u} + \mu\underline{v}$ of a plane can be written out in coordinates in various ways.

- Parametric description in 'row notation':

$$(x_1, x_2, x_3) = (a_1, a_2, a_3) + \lambda(u_1, u_2, u_3) + \mu(v_1, v_2, v_3).$$

- Parametric description in 'column notation':

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix} + \lambda \begin{pmatrix} u_1 \\ u_2 \\ u_3 \end{pmatrix} + \mu \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix} .$$

- Or simply each coordinate separately:

$$\begin{aligned} x_1 &= a_1 + \lambda u_1 + \mu v_1 \\ x_2 &= a_2 + \lambda u_2 + \mu v_2 \\ x_3 &= a_3 + \lambda u_3 + \mu v_3 . \end{aligned}$$

Upon eliminating the parameters λ and μ an equation of the plane appears, a linear equation in x_1, x_2, x_3,

$$d_1 x_1 + d_2 x_2 + d_3 x_3 = d_4,$$

for certain d_1, d_2, d_3, d_4. At least one of the coefficients d_1, d_2, d_3 should be nonzero.

2.3.7 Examples. a) $\underline{x} = (1,2) + \lambda(3,-1)$ and $\underline{x} = (1,2) + \mu(-6,2)$ describe the same line. Why?

b) To determine an equation of the line $\ell : \underline{x} = (1,2) + \lambda(3,-1)$, we start with $x_1 = 1 + 3\lambda$ and $x_2 = 2 - \lambda$. Now multiply the 2nd equation by 3 and add the result to the 1st equation:

$$x_1 + 3x_2 = 1 + 3\lambda + 3(2-\lambda) = 7.$$

So an equation of the line ℓ is $x_1 + 3x_2 = 7$.

c) Suppose $2x_1 - x_2 + 3x_3 = 4$ is the equation of the plane V. To determine a parametric description we proceed as follows. If you assign any value, say λ, to x_2 and any value, say μ, to x_3, then x_1 is determined: $x_1 = 2 + \lambda/2 - 3\mu/2$. So

$$x_1 = 2 + \lambda/2 - 3\mu/2,\ x_2 = \lambda,\ x_3 = \mu.$$

In vector notation:

$$(x_1, x_2, x_3) = (2+\lambda/2-3\mu/2, \lambda, \mu) = (2,0,0) + \lambda(1/2,1,0) + \mu(-3/2,0,1).$$

Then a vector parametric description is

$$V : \underline{x} = (2, 0, 0) + \lambda(\frac{1}{2}, 1, 0) + \mu(-\frac{3}{2}, 0, 1).$$

To avoid fractions, you could also take

$$V : \underline{x} = (2, 0, 0) + \rho(1, 2, 0) + \sigma(-3, 0, 2).$$

Do you see why?

d) To find an equation of the plane V with vector parametric equation $\underline{x} = (2, 0, 0) + \lambda(1, 1, 0) + \mu(0, 2, 1)$, we eliminate λ and μ from the three expressions $x_1 = 2 + \lambda$, $x_2 = \lambda + 2\mu$ and $x_3 = \mu$, for instance as follows (more systematic methods will be discussed in later chapters):

- Since $x_3 = \mu$ we can replace μ in $x_2 = \lambda + 2\mu$ by x_3: $x_2 = \lambda + 2x_3$.
- Now subtract $x_2 = \lambda + 2x_3$ from $x_1 = 2 + \lambda$: $x_1 - x_2 = 2 - 2x_3$. So an equation is

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

2.4 Distances, Angles and the Inner Product

2.4.1 Computations involving the length of a vector, the distance and angle between two vectors (the definitions are given below) become easier with the notion of *inner product*. This section is devoted to a discussion of these notions.

We start with the plane or with the 3-dimensional space and introduce an origin. The *length* of a vector $\underline{x}$ is the distance between head and tail of (any representative arrow of) the vector. The length of $\underline{x}$ is denoted by $\| \underline{x} \|$. The *distance* between the two vectors $\underline{u}$ and $\underline{v}$ is by definition the length of the difference $\underline{u} - \underline{v}$ (or $\underline{v} - \underline{u}$), so $\| \underline{u} - \underline{v} \|$.

2.4.2 The inner product
The *inner product* of two vectors $\underline{u}$ and $\underline{v}$, both $\neq \underline{0}$, is defined as

$$\| \underline{u} \| \cdot \| \underline{v} \| \cdot \cos\varphi,$$

where φ is the angle between the vectors $\underline{u}$ and $\underline{v}$ (note the role of the cosine: the sign of the angle doesn't matter). If one (or both) of the vectors is the zero vector, then the inner product is, by definition, 0. We denote the inner product by

$$(\underline{u}, \underline{v}).$$

Here is an example. Suppose the vectors $\underline{u}$, $\underline{v}$ have length 4 and the angle

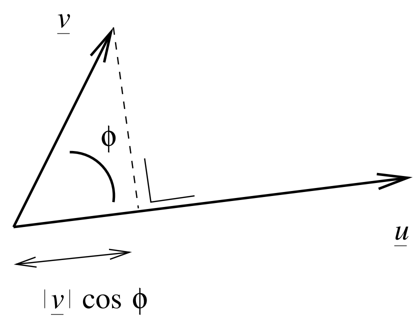


Figure 2.7: *If the angle between the vectors $\underline{u}$ and $\underline{v}$ is at most $\pi/2$, then their inner product equals the product of the length of $\underline{u}$ and the length of the projection of $\underline{v}$ on the line $\underline{x} = \lambda\underline{u}$.*

between them is 60° (or $\pi/3$ radians), then

$$(\underline{u}, \underline{v}) = 4 \cdot 4 \cos 60^\circ = 4 \cdot 4 \cdot \frac{1}{2} = 8.$$

If the angle is 120°, then the inner product changes into -8. In particular, the inner product can be negative. In the literature other notations occur, like $\underline{u} \bullet \underline{v}$ (the inner product is sometimes called the dot product). Here are some remarks and properties (we do not always treat the case of zero vectors separately).

- If $\underline{v} = \underline{u}$, both $\neq \underline{0}$, then the angle is 0, so the cosine is 1 and we obtain $(\underline{u}, \underline{u}) = \parallel \underline{u} \parallel^2$. Or:

  $$\parallel \underline{u} \parallel = \sqrt{(\underline{u}, \underline{u})}.$$

  This relation also holds in case $\underline{u} = \underline{0}$. Note that $(\underline{u}, \underline{u}) \geq 0$ for every vector $\underline{u}$, and that $(\underline{u}, \underline{u}) = 0$ occurs precisely if $\underline{u} = \underline{0}$.

- If the inner product of two nonzero vectors is known as well as their lengths, then (the cosine of) the angle between the vectors can be computed:

  $$\cos\varphi = \frac{(\underline{u}, \underline{v})}{\parallel \underline{u} \parallel \cdot \parallel \underline{v} \parallel}.$$

Once we work with coordinates this is often useful.

- **Symmetry of the inner product**:
  $(\underline{u}, \underline{v}) = (\underline{v}, \underline{u})$ for all vectors $\underline{u}$ and $\underline{v}$. This follows immediately from the definition, since

  $$\| \underline{u} \| \cdot \| \underline{v} \| \cdot \cos\varphi = \| \underline{v} \| \cdot \| \underline{u} \| \cdot \cos\varphi.$$

  (The angle between the vectors is the same in both cases.)

- Behaviour with respect to **scalar multiplication**:
  $\lambda(\underline{u}, \underline{v}) = (\lambda\underline{u}, \underline{v}) = (\underline{u}, \lambda\underline{v})$ for all vectors $\underline{u}$, $\underline{v}$ and for every scalar λ. Fill in the details yourself (distinguish the cases $\lambda > 0$, $\lambda < 0$ and $\lambda = 0$).

- Behaviour with respect to **vector addition**:

  $$\begin{aligned}(\underline{u} + \underline{v}, \underline{w}) &= (\underline{u}, \underline{w}) + (\underline{v}, \underline{w}),\\ (\underline{u}, \underline{v} + \underline{w}) &= (\underline{u}, \underline{v}) + (\underline{u}, \underline{w})\end{aligned}$$

  for all vectors $\underline{u}$, $\underline{v}$ en $\underline{w}$.

- **Orthogonality**:
  If two non-zero vectors have inner product 0, then they are perpendicular (the angle between them is $\pm 90°$ or $\pm\pi/2$) since the cosine of the angle between them is 0. Conversely, if two non-zero vectors are perpendicular, then their inner product is 0. Now the zero vector has inner product 0 with any vector, and we agree to say that the zero vector is perpendicular to any vector. This is a convenient convention since then we have: The inner product of two vectors is 0 if and only if the two vectors are perpendicular.

2.4.3 Examples. Although lengths and angles are maybe what we are really interested in, the inner product is so useful because of the arithmetic rules it satisfies. For instance, $\| \underline{u} + \underline{v} \|$ usually differs from $\| \underline{u} \| + \| \underline{v} \|$, but $(\underline{u} + \underline{v}, \underline{u} + \underline{v})$ is easy to expand using the rules. Often it is therefore useful to translate problems involving lengths and angles into problems with inner products. Here are some examples demonstrating the use of the inner product's properties.

a) Suppose that $(\underline{u}, \underline{v}) = 2$. Using the arithmetic rules the inner product $(3\,\underline{u}, -4\,\underline{v})$ is computed as follows:

$$(3\,\underline{u}, -4\,\underline{v}) = 3(\underline{u}, -4\,\underline{v}) = 3 \cdot -4\,(\underline{u}, \underline{v}) = -12 \cdot 2 = -24.$$

b) If $\|\ \underline{u}\ \|= 2$, $\|\ \underline{v}\ \|= 3$ en $(\underline{u}, \underline{v}) = 1$, then, using the arithmetic rules again, you can determine, for instance, $(\underline{u}+\underline{v}, \underline{u}-2\underline{v})$. Here is the first step of the computation:

$$(\underline{u}+\underline{v}, \underline{u}-2\underline{v}) = (\underline{u}, \underline{u}-2\underline{v}) + (\underline{v}, \underline{u}-2\underline{v}).$$

Next, we turn to the first term on the right-hand side, $(\underline{u}, \underline{u}-2\underline{v})$:

$$(\underline{u}, \underline{u}-2\underline{v}) = (\underline{u}, \underline{u}) + (\underline{u}, -2\underline{v}) = (\underline{u}, \underline{u}) - 2(\underline{u}, \underline{v}).$$

Now $(\underline{u}, \underline{u}) =\|\ \underline{u}\ \|^2= 4$ and $(\underline{u}, \underline{v}) = 1$, so we find $(\underline{u}, \underline{u}-2\underline{v}) = 4-2 = 2$. In a similar way we deal with the term $(\underline{v}, \underline{u}-2\underline{v})$:

$$(\underline{v}, \underline{u}-2\underline{v}) = (\underline{v}, \underline{u}) + (\underline{v}, -2\underline{v}) = (\underline{v}, \underline{u}) - 2(\underline{v}, \underline{v}) = 1 - 2 \cdot 9 = -17.$$

So we find $(\underline{u}+\underline{v}, \underline{u}-2\underline{v}) = 2 - 17 = -15$.

2.4.4 The inner product and coordinates

Let $\underline{e}_1$, $\underline{e}_2$ be a basis of the plane consisting of two perpendicular vectors of length 1 (an *orthonormal basis*). In terms of coordinates we find an easy to memorize expression (in the case the basis is not orthonormal the expressions become complicated, we will not discuss that situation). Now suppose $\underline{v} = v_1\underline{e}_1 + v_2\underline{e}_2$ and $\underline{w} = w_1\underline{e}_1 + w_2\underline{e}_2$ are two vectors in the plane, then we find, using the properties of the inner product and using that $(\underline{e}_1, \underline{e}_1) = 1$, $(\underline{e}_1, \underline{e}_2) = 0$, $(\underline{e}_2, \underline{e}_1) = 0$, $(\underline{e}_2, \underline{e}_2) = 1$:

$$\begin{aligned}(\underline{v}, \underline{w}) &= (v_1\underline{e}_1 + v_2\underline{e}_2, w_1\underline{e}_1 + w_2\underline{e}_2)\\ &= v_1w_1(\underline{e}_1, \underline{e}_1) + v_1w_2(\underline{e}_1, \underline{e}_2) + v_2w_1(\underline{e}_2, \underline{e}_1) + v_2w_2(\underline{e}_2, \underline{e}_2)\\ &= v_1w_1 + v_2w_2.\end{aligned}$$

So:

$$(\underline{v}, \underline{w}) = v_1w_1 + v_2w_2.$$

In particular, we obtain an easy (and well-known) expression for the length of a vector $\underline{v} = v_1\underline{e}_1 + v_2\underline{e}_2$:

$$\|\ \underline{v}\ \|= \sqrt{v_1^2 + v_2^2}.$$

In terms of coordinates, the distance between $\underline{u} = (u_1, u_2)$ and $\underline{v} = (v_1, v_2)$ equals

$$\| \underline{u} - \underline{v} \| = \sqrt{(u_1 - v_1)^2 + (u_2 - v_2)^2}.$$

The cosine of the angle φ between the vectors (both $\neq \underline{0}$) $\underline{v} = v_1\underline{e}_1 + v_2\underline{e}_2$ and $\underline{w} = w_1\underline{e}_1 + w_2\underline{e}_2$ is equal to

$$\cos\varphi = \frac{(\underline{v}, \underline{w})}{\| \underline{v} \| \cdot \| \underline{w} \|} = \frac{v_1w_1 + v_2w_2}{\sqrt{v_1^2 + v_2^2} \cdot \sqrt{w_1^2 + w_2^2}}.$$

In a similar way, using an orthonormal basis $\underline{e}_1$, $\underline{e}_2$, $\underline{e}_3$ in space (so vectors of length 1 and mutually perpendicular), we have the following expression in coordinates for the inner product of the vectors $\underline{v} = v_1\underline{e}_1 + v_2\underline{e}_2 + v_3\underline{e}_3$ and $\underline{w} = w_1\underline{e}_1 + w_2\underline{e}_2 + w_3\underline{e}_3$:

$$v_1w_1 + v_2w_2 + v_3w_3.$$

The coordinate expression for the length of the vector $\underline{v}$ is

$$\| \underline{v} \| = \sqrt{v_1^2 + v_2^2 + v_3^2}.$$

The distance between $\underline{u}$ and $\underline{v}$ equals

$$\| \underline{u} - \underline{v} \| = \sqrt{(u_1 - v_1)^2 + (u_2 - v_2)^2 + (u_3 - v_3)^2}.$$

Finally, the cosine of the angle between the vectors $\underline{v}$ and $\underline{w}$ (both $\neq \underline{0}$) equals

$$\cos\varphi = \frac{(\underline{v}, \underline{w})}{\| \underline{v} \| \cdot \| \underline{w} \|} = \frac{v_1w_1 + v_2w_2 + v_3w_3}{\sqrt{v_1^2 + v_2^2 + v_3^2} \cdot \sqrt{w_1^2 + w_2^2 + w_3^2}}.$$

2.4.5 $\mathbb{R}^2$, $\mathbb{R}^3$ and the standard inner product

Motivated by the previous discussion, we introduce th so-called *standard inner product* in $\mathbb{R}^2$ and $\mathbb{R}^3$, viewed as vector spaces themselves (more on this in later chapters). A vector in $\mathbb{R}^2$ is a pair of real numbers like (a_1, a_2). The *standard inner product* of two vectors $\underline{a} = (a_1, a_2)$ and $\underline{b} = (b_1, b_2)$ in $\mathbb{R}^2$ is defined as

$$(\underline{a}, \underline{b}) := a_1b_1 + a_2b_2.$$

Similarly, he standard inner product of two vectors $\underline{a} = (a_1, a_2, a_3)$ and $\underline{b} = (b_1, b_2, b_3)$ in $\mathbb{R}^3$ is defined as

$$(\underline{a}, \underline{b}) := a_1b_1 + a_2b_2 + a_3b_3.$$

2.4.6 Example. The angle φ between the vectors $\underline{u} = (1,0)$ and $\underline{v} = (1,1)$ in $\mathbb{R}^2$ can be determined as follows.

$$\cos\varphi = \frac{(\underline{u},\underline{v})}{\|\,\underline{u}\,\| \cdot \|\,\underline{v}\,\|} = \frac{1\cdot 1 + 0\cdot 1}{\sqrt{1^2+0^2}\cdot\sqrt{1^2+1^2}} = \frac{1}{\sqrt{2}} = \frac{1}{2}\sqrt{2}.$$

So the angle is $\pi/4$ (or $45°$).

2.4.7 Normal vectors

If $\underline{u} = (u_1,u_2,u_3)$ and $\underline{v} = (v_1,v_2,v_3)$ are two vectors in the plane V with equation $2x_1 - x_2 + 3x_3 = 6$, then $2u_1 - u_2 + 3u_3 = 6$ en $2v_1 - v_2 + 3v_3 = 6$. Subtracting yields

$$2(u_1 - v_1) - (u_2 - v_2) + 3(u_3 - v_3) = 0.$$

We can rephrase this equality as an inner product:

$$((2,-1,3),(u_1 - v_1, u_2 - v_2, u_3 - v_3)) = 0.$$

This means that the difference $\underline{u} - \underline{v}$ is perpendicular to $(2,-1,3)$. In particular, $(2,-1,3)$ is a vector which is perpendicular to all direction vectors of the plane. We call $(2,-1,3)$ a *normal vector* of the plane.

In general, if $a_1x_1 + a_2x_2 + a_3x_3 = d$ is an equation of the plane V, then we can rewrite it in the form of an inner product:

$$(\underline{a},\underline{x}) = d,$$

where $\underline{a} = (a_1,a_2,a_3)$ and $\underline{x} = (x_1,x_2,x_3)$. If $\underline{u}$ and $\underline{v}$ are two vectors in the plane, then $(\underline{a},\underline{u}) = d$ and $(\underline{a},\underline{v}) = d$. Subtracting yields $(\underline{a},\underline{u}) - (\underline{a},\underline{v}) = 0$, so that, using the properties of the inner product:

$$(\underline{a},\underline{u} - \underline{v}) = 0.$$

In other words, $\underline{u} - \underline{v}$ is perpendicular to $\underline{a}$. In particular, direction vectors of the plane V are all perpendicular to $\underline{a}$. The vector $\underline{a}$ is called a *normal vector* of the plane.

The situation for lines in the plane is similar. If $a_1x_1 + a_2x_2 = d$ is an equation of a line, then (a_1,a_2) is a normal vector of the line. This vector is perpendicular to every direction vector of the line.

2.4.8 Pythagoras
If $\underline{u}$ and $\underline{v}$ are perpendicular vectors, then we find for the square of the length of the sum vector $\underline{u}+\underline{v}$:

$$\begin{aligned}\| \underline{u}+\underline{v} \|^2 &= (\underline{u}+\underline{v}, \underline{u}+\underline{v}) \\ &= (\underline{u}, \underline{u}) + 2(\underline{u}, \underline{v}) + (\underline{v}, \underline{v}) \\ &= (\underline{u}, \underline{u}) + (\underline{v}, \underline{v}) \\ &= \| \underline{u} \|^2 + \| \underline{v} \|^2 .\end{aligned}$$

Similarly, $\| \underline{u}-\underline{v} \|^2 = \| \underline{u} \|^2 + \| \underline{v} \|^2$. This is a vector form of the

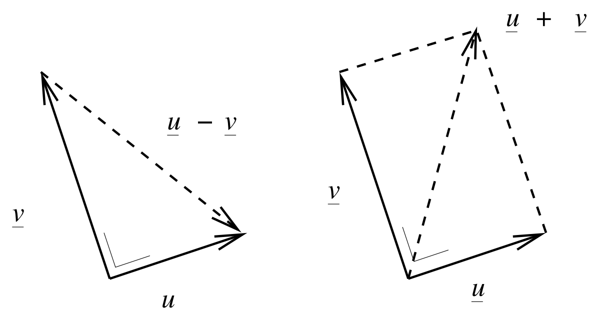


Figure 2.8: *If $\underline{u}$ and $\underline{v}$ are perpendicular, then $\| \underline{u}+\underline{v} \|^2 = \| \underline{u} \|^2 + \| \underline{v} \|^2$ and $\| \underline{u}-\underline{v} \|^2 = \| \underline{u} \|^2 + \| \underline{v} \|^2$. The figure illustrates the relation with the Pythagorean theorem.*

Pythagorean theorem: the triangle with vertices the endpoints of $\underline{0}$, $\underline{u}$, $\underline{v}$ is a right triangle whose sides opposite the hypothenuse have lengths $\| \underline{u} \|$ and $\| \underline{v} \|$, respectively. The hypothenuse has length $\| \underline{u}-\underline{v} \|$.

Give a similar geometric interpretation of the equality $\| \underline{u}+\underline{v} \|^2 = \| \underline{u} \|^2 + \| \underline{v} \|^2$.

2.4.9 Example. We determine the distance between (the endpoint of) $\underline{p} = (1, 2)$ and the line $\ell : \underline{x} = (8, 1) + \lambda(3, -4)$. To this end we first determine a vector $\underline{q}$ on ℓ such that $\underline{p}-\underline{q}$ is perpendicular to ℓ, i.e., perpendicular to $\underline{a}$. To find $\underline{q}$, we solve:

$$((1, 2) - (8, 1) - \lambda(3, -4), (3, -4)) = 0 \text{ or } (-7) \cdot 3 - 9\lambda + 1 \cdot (-4) - 16\lambda = 0.$$

This leads to $\lambda = -1$. So $\underline{q} = (5, 5)$. The distance between $\underline{p}$ and $\underline{q}$ is $\sqrt{(5-1)^2 + (5-2)^2} = 5$. This is also the distance between $\underline{p}$ and the line ℓ:

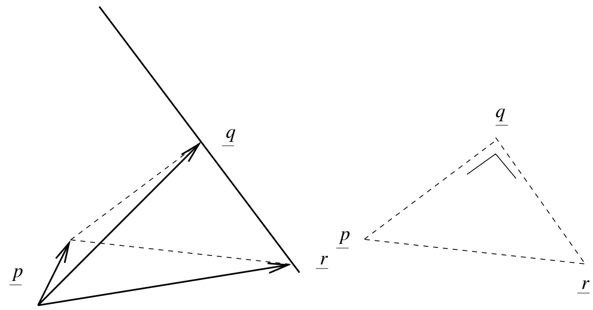


Figure 2.9: *To compute the distance between* $\underline{p}$ *and the line* ℓ*, we determine a vector* $\underline{q}$ *on* ℓ *such that* $\underline{p} - \underline{q}$ *is perpendicular to* ℓ*. If* $\underline{r}$ *is an arbitrary vector on* ℓ*, then the right-hand figure illustrates that the distance between* $\underline{p}$ *and* $\underline{r}$ *is greater than (or equal to) the distance between* $\underline{p}$ *and* $\underline{q}$ *because of the Pythagorean theorem.*

for every vector on ℓ, its distance to $\underline{p}$ turns out to be at least as big. Here is why. If $\underline{r}$ is a vector on ℓ, then we should compare $\| \underline{p} - \underline{r} \|$ and $\| \underline{p} - \underline{q} \|$. Since $\underline{p} - \underline{q}$ is perpendicular to $\underline{q} - \underline{r}$ (why?), we can apply the Pythagorean theorem to the triangle with vertices $\underline{p}$, $\underline{q}$, $\underline{r}$. In vector language: we apply Pythagoras to the vectors $\underline{u} = \underline{p} - \underline{q}$, $\underline{v} = \underline{q} - \underline{r}$ and their sum $\underline{u} + \underline{v} = \underline{p} - \underline{r}$. We obtain:

$$\| \underline{p} - \underline{r} \|^2 = \| \underline{p} - \underline{q} \|^2 + \| \underline{q} - \underline{r} \|^2 \ .$$

Evidently, $\| \underline{p} - \underline{r} \| \geq \| \underline{p} - \underline{q} \|$ (with equality if and only if $\underline{q} = \underline{r}$). So $\| \underline{p} - \underline{q} \|$ is the distance between $\underline{p}$ and ℓ.

2.5 The cross product

2.5.1 The definition of the cross product

The inner product of two vectors is a real number. There is also a construction that assigns to two vectors in space a new vector with special and useful properties. We discuss this construction on the level of coordinates, so in $\mathbb{R}^3$.

The *cross product* $\underline{v} \times \underline{w}$ of the vectors $\underline{v} = (v_1, v_2, v_3)$ and $\underline{w} = (w_1, w_2, w_3)$ is by definition the vector

$$(v_2w_3 - v_3w_2, v_3w_1 - v_1w_3, v_1w_2 - v_2w_1).$$

This looks pretty complicated, but turns out to always produce a vector perpendicular to both $\underline{v} = (v_1, v_2, v_3)$ and $\underline{w} = (w_1, w_2, w_3)$. In a sense it provides a 'universal' answer to the question: what is a vector perpendicular to two given vectors in $\mathbb{R}^3$? In $\mathbb{R}^2$, the analogue would be the much easier question: what is a vector perpendicular to a given vector (a, b)? In this case a useful answer is easy to see: $(-b, a)$.

The cross product has more properties, like the following: its length equals $\| \underline{v} \| \cdot \| \underline{w} \| \cdot \sin\varphi$, where φ is the angle $(0 \leq \varphi \leq \pi)$ between $\underline{v}$ and $\underline{w}$. This length equals the surface area of the parallelogram spanned by $\underline{v}$ and $\underline{w}$.

Here is a list with properties of the cross product. They can all be verified by expanding the relevant expressions, although d) requires some thought, see below. There are more properties but they are beyond the scope of these lecture notes. For all $\underline{v}$, $\underline{w}$, etc., we have:

a) $\underline{v} \times \underline{v} = \underline{0}$.

b) The cross product of $\underline{v}$ and $\underline{w}$ is perpendicular to both $\underline{v}$ and $\underline{w}$, i.e., the corresponding inner products are 0:

$$(\underline{v} \times \underline{w}, \underline{v}) = 0 \text{ and } (\underline{v} \times \underline{w}, \underline{w}) = 0.$$

This property can be used to determine, for instance, a normal vector to a plane given its direction vectors.

c) The cross product is antisymmetric:

$$\underline{v} \times \underline{w} = -(\underline{w} \times \underline{v}).$$

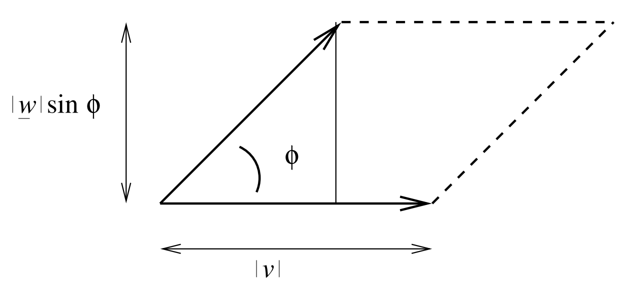


Figure 2.10: *The length of the cross product of $\underline{v}$ and $\underline{w}$ is the area of the parallelogram spanned by $\underline{v}$ and $\underline{w}$.*

d) The length of the cross product equals

$$\| \underline{v} \times \underline{w} \| = \| \underline{v} \| \cdot \| \underline{w} \| \cdot \sin\varphi,$$

where φ is the angle between $\underline{v}$ en $\underline{w}$. This is precisely the surface area of the parallelogram spanned by $\underline{v}$ and $\underline{w}$ ('base times heigth', where the base has length $\| \underline{v} \|$ and the height equals $\| \underline{w} \| \cdot \| \sin\varphi \|$).

e) The connection with the vector addition is as follows:

$$\underline{u} \times (\underline{v} + \underline{w}) = \underline{u} \times \underline{v} + \underline{u} \times \underline{w} \quad \text{en} \quad (\underline{v} + \underline{w}) \times \underline{u} = \underline{v} \times \underline{u} + \underline{w} \times \underline{u}.$$

f) The connection with scalar multiplication is as follows:

$$\lambda(\underline{v} \times \underline{w}) = (\lambda\underline{v}) \times \underline{w} = \underline{v} \times (\lambda\underline{w}).$$

The properties b) and d) almost determine the cross product, but not quite: the cross product could still point in two different directions perpendicular to the plane spanned by $\underline{v}$ and $\underline{w}$. Which direction to choose is based on the *right hand rule* if you put your right hand along $\underline{v}$ in such a way that your fingers curl from $\underline{v}$ to $\underline{w}$ (so either your litte finger or your index finger touches $\underline{v}$), then your thumb points in the direction of $\underline{v} \times \underline{w}$.

2.5.2 On the proof of property d)

Property d) on the length of the cross product is best approached by some subtle manoeuvring to deal with the factor $\sin\varphi$. Here are the various steps. Firstly, instead of proving that $\| \underline{v} \times \underline{w} \| = \| \underline{v} \| \cdot \| \underline{w} \| \cdot \sin\varphi$, we show that $\| \underline{v} \times \underline{w} \|^2$ equals

$$\| \underline{v} \|^2 \cdot \| \underline{w} \|^2 \cdot \sin^2\varphi.$$

Secondly, replace $\sin^2\varphi$ by $1 - \cos^2\varphi$ and use that $(\underline{v}, \underline{w}) = \| \underline{v} \| \cdot \| \underline{w} \| \cdot \cos\varphi$ to replace $\| \underline{v} \|^2 \cdot \| \underline{w} \|^2 \cdot \sin^2\varphi$ by

$$\| \underline{v} \|^2 \cdot \| \underline{w} \|^2 \cdot (1 - \cos^2\varphi) = \| \underline{v} \|^2 \cdot \| \underline{w} \|^2 - (\underline{v}, \underline{w})^2.$$

So now we are reduced to showing that $\| \underline{v} \times \underline{w} \|^2$ equals $\| \underline{v} \|^2 \cdot \| \underline{w} \|^2 - (\underline{v}, \underline{w})^2$, i.e., $(v_2w_3 - v_3w_2)^2 + (v_3w_1 - v_1w_3)^2 + (v_1w_2 - v_2w_1)^2$ equals

$$(v_1^2 + v_2^2 + v_3^2)(w_1^2 + w_2^2 + w_3^2) - (v_1w_1 + v_2w_2 + v_3w_3)^2.$$

This verification is straightforward and left to the reader.

2.5.3 The volume of a parallelepiped

The volume of the parallelepiped P 'spanned' by the vectors $\underline{a}$, $\underline{b}$, $\underline{c}$ can be

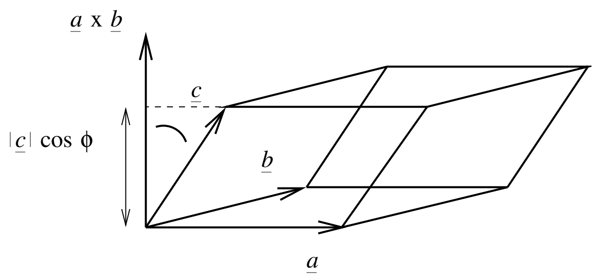


Figure 2.11: *The volume of the parallelepiped equals the absolute value of* $(\underline{a} \times \underline{b}, \underline{c})$.

expressed using the inner product and cross product. To obtain this expression, note that the volume is the product of the area of the parallelogram spanned by $\underline{a}$ and $\underline{b}$ and the height. The area of the parallelogram is $\| \underline{a} \times \underline{b} \|$ as we saw before. Since $\underline{a} \times \underline{b}$ is perpendicular to the parallelogram, the height equals the (length of the) projection of $\underline{c}$ on $\underline{a} \times \underline{b}$, i.e., the absolute value of $\| \underline{c} \| \cdot \cos\varphi$, where φ is the angle between $\underline{c}$ and $\underline{a} \times \underline{b}$. So, the volume of the parallelepiped is

$$\| \underline{a} \times \underline{b} \| \cdot \| \underline{c} \| \cdot |\cos\varphi| = |(\underline{a} \times \underline{b}, \underline{c})|.$$

In conclusion, the volume of the parallelepiped spanned by $\underline{a}$, $\underline{b}$, $\underline{c}$ is

$$|(\underline{a} \times \underline{b}, \underline{c})|.$$

2.5.4 Examples. a) A normal vector to the plane V with parametric description $\underline{x} = (1,2,3) + \lambda(1,2,1) + \mu(3,1,0)$ is

$$(1,2,1) \times (3,1,0) = (-1,3,-5).$$

An equation of the plane is therefore $-x_1 + 3x_2 - 5x_3 = d$ for some d. Now substitute $(1,2,3)$ and we find that $d = -10$. An equation is $-x_1 + 3x_2 - 5x_3 = -10$.

b) The surface area of the triangle with vertices $(0,0,0)$, $(1,2,1)$, $(2,-1,3)$ equals

$$\frac{1}{2} \| (1,2,1) \times (2,-1,3) \| = \frac{1}{2} \| (7,-1,-5) \| = \frac{1}{2}\sqrt{65}\,.$$

2.6 Vectors and geometry

2.6.1 Vectors can be a useful tool in addressing geometric problems. Below we describe how to translate a geometric situation into the language of vectors and how to solve (some) geometric problems using our vector techniques. Not all geometric problems lend themselves for such an approach, but vector techniques, just like complex numbers, do provide additional tools for dealing with geometric problems.

2.6.2 The medians in a triangle

A well-known theorem in geometry states that the three medians in a triangle are concurrent, i.e., have a point in common, the *centroid*. The *medians* (segments or lines as the situation requires) join the vertices of a triangle to the midpoints of the opposite sides.

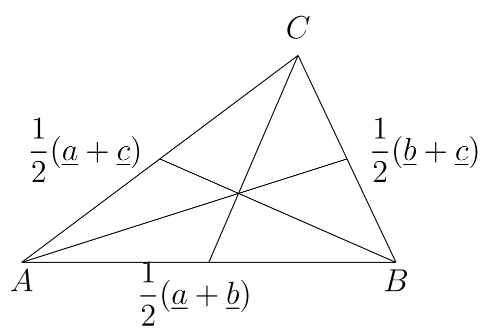


Figure 2.12: *The three medians in $\triangle ABC$ are concurrent. The vector description of the midpoints of the sides is given.*

For triangle $\triangle ABC$ we denote the vectors corresponding to the vertices as follows: $\underline{a}$, $\underline{b}$, $\underline{c}$. The midpoint of side BC corresponds to the vector $\frac{1}{2}(\underline{b}+\underline{c})$. A parametric description of the median through A is then

$$\underline{x} = \underline{a} + \lambda\left(\frac{1}{2}(\underline{b}+\underline{c}) - \underline{a}\right).$$

The other two medians have the following descriptions:

$$\underline{x} = \underline{b} + \mu \left(\frac{1}{2}(\underline{a} + \underline{c}) - \underline{b} \right)$$
$$\underline{x} = \underline{c} + \rho \left(\frac{1}{2}(\underline{a} + \underline{b}) - \underline{c} \right).$$

The question whether the three medians have a point in common comes down to the question whether the parameters λ, μ and ρ can be chosen in such a way that the three parametric descriptions describe the same vector. The answer is 'yes.' Indeed, for λ, μ, ρ all equal to $2/3$ we obtain the common vector $\frac{1}{3}(\underline{a}+\underline{b}+\underline{c})$. This gives a vector description of the centroid of a triangle. Note that it looks like an 'average' of the three vectors corresponding to the vertices.

Since we need the parameter value $2/3$, the vector approach also shows that the medians, now viewed as segments, divide one another in the ratio $2 : 1$.

Note that the common value for λ, μ, ρ can also be computed. Try finding that value by rewriting the vector equation

$$\underline{a} + \lambda \left(\frac{1}{2}(\underline{b} + \underline{c}) - \underline{a} \right) = \underline{b} + \mu \left(\frac{1}{2}(\underline{a} + \underline{c}) - \underline{b} \right)$$

in the form $(2-2\lambda-\mu)\underline{a}+(\lambda-2+2\mu)\underline{b}+(\lambda-\mu)\underline{c} = \underline{0}$. Note that this equation does not imply that the coefficients of $\underline{a}$, $\underline{b}$, $\underline{c}$ are all 0 (why?). Fortunately, we don't need that, we just need *a* solution. This is a subtle point. If you don't trust it, put the origin in A, say, and the computations (and subtleties) simplify.

2.6.3 A parallelogram in a quadrangle

A second example concerns an arbitrary quadrangle $ABCD$ in the plane in which no two points coincide. Let E, F, G, H, respectively, be the midpoints of (the segments) AB, BC, CD, AD, respectively. The figure suggests the theorem: quadrangle $EFGH$ is a parallelogram (regardless of the position of the points A, B, C, D)!

Again, we use vectors. We have to show that EF and HG are parallel and equal in length. In vector language this comes down to showing that $\underline{e} - \underline{f} = \pm(\underline{h} - \underline{g})$, where we indicate vectors corresponding to the points

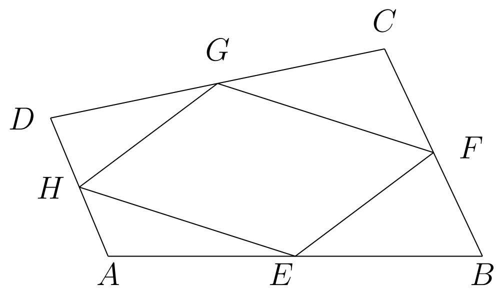


Figure 2.13: *The midpoints of the sides of quadrangle $ABCD$ form a parallelogram.*

in the obvious way. First we express the vectors $\underline{e}$, $\underline{f}$, $\underline{g}$, $\underline{h}$ in terms of the vectors $\underline{a}$, $\underline{b}$, $\underline{c}$, $\underline{d}$:

$$\underline{e} = \frac{1}{2}(\underline{a}+\underline{b}),\ \underline{f} = \frac{1}{2}(\underline{b}+\underline{c}),\ \underline{g} = \frac{1}{2}(\underline{c}+\underline{d}),\ \underline{h} = \frac{1}{2}(\underline{a}+\underline{d}).$$

Then we analyze $\underline{e}-\underline{f}$ and $\underline{h}-\underline{g}$:

$$\underline{e}-\underline{f} = \frac{1}{2}(\underline{a}+\underline{b}) - \frac{1}{2}(\underline{b}+\underline{c}) = \frac{1}{2}(\underline{a}-\underline{c})$$

and

$$\underline{h}-\underline{g} = \frac{1}{2}(\underline{a}+\underline{d}) - \frac{1}{2}(\underline{c}+\underline{d}) = \frac{1}{2}(\underline{a}-\underline{c}).$$

This finishes the proof.

2.6.4 The altitudes in a triangle are concurrent

The *altitudes* of a triangle are the segments (or lines, when convenient) connecting a vertex of the triangle with the (unique) point on the opposite side (extended if necessary) so that the segment (or line) is perpendicular to this opposite side.

We use vectors and the inner product to show that the three altitudes in a triangle are concurrent, i.e., pass through a common point, the so-called *orthocenter* of the triangle.

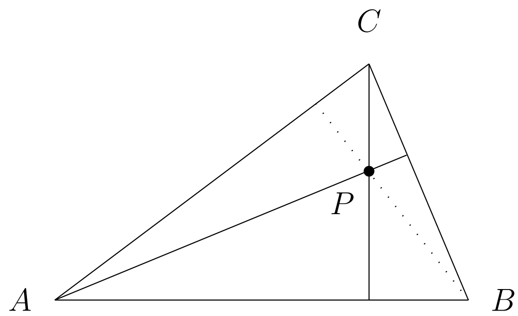


Figure 2.14: *The altitudes in $\triangle ABC$ are concurrent. The altitude from B is dashed.*

So, let $\triangle ABC$ be a triangle. Suppose the altitudes from A and C meet in P. The vector corresponding to P is denoted by $\underline{p}$. The fact that AP is perpendicular to BC and CP is perpendicular to AB translates as follows:

$$\begin{array}{c} \underline{p}-\underline{a} \perp \underline{b}-\underline{c} \text{ or } (\underline{p}-\underline{a}, \underline{b}-\underline{c}) = 0, \\ \underline{p}-\underline{c} \perp \underline{a}-\underline{b} \text{ or } (\underline{p}-\underline{c}, \underline{a}-\underline{b}) = 0. \end{array} \tag{2.1}$$

In order to prove that P is on the altitude from B, we will show that $\underline{p}-\underline{b}$ and $\underline{a}-\underline{c}$ are perpendicular. First we use the bilinearity of the inner product to rewrite the expressions in (2.1):

$$\begin{array}{c} (\underline{p}, \underline{b}) + (\underline{a}, \underline{c}) = (\underline{p}, \underline{c}) + (\underline{a}, \underline{b}) \\ (\underline{p}, \underline{a}) + (\underline{c}, \underline{b}) = (\underline{p}, \underline{b}) + (\underline{c}, \underline{a}) \end{array}$$

Adding (the left-hand sides and right-hand sides, respectively, of) these equations yields

$$(\underline{p}, \underline{a}) + (\underline{c}, \underline{b}) = (\underline{p}, \underline{c}) + (\underline{a}, \underline{b}),$$

which can be rewritten as

$$(\underline{p}-\underline{b}, \underline{a}-\underline{c}) = 0.$$

So we are done.

Note that we haven't used the freedom to choose an origin. Choosing a convenient origin might simplify the computations. In our case, a clever choice would be to put the origin in P. Please check yourself in what way the computation then simplifies.

2.6.5 The nine-point circle revisited

Vectors can also be used to analyse the nine-point circle of a triangle ABC, i.e., the circumcircle of the midpoints of the sides of the triangle, discussed in the previous chapter using complex numbers.

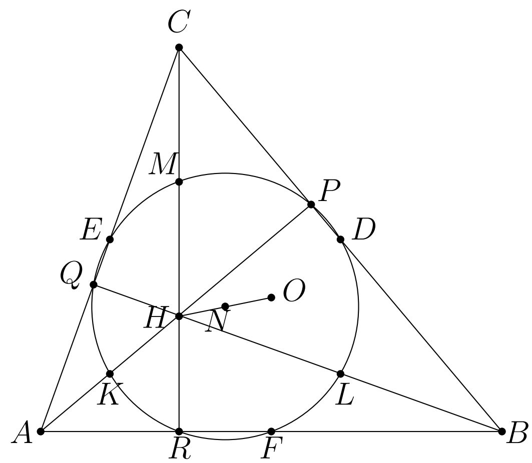


Figure 2.15: *The nine points D, E, F, K, K, L, M, P, Q, R on the circumcircle of triangle DEF with center N. The orthocenter H and the center O of the circumcircle of $\triangle ABC$ are also drawn.*

Choose the origin O in the circumcenter of $\triangle ABC$. Let $\underline{a}$, $\underline{b}$, etc., denote the vectors corresponding to the vertices A, B, etc. Then $\| \underline{a} \| = \| \underline{b} \| = \| \underline{c} \|$. Denote by D, E, F the midpoints of the sides BC, AC, AB, respectively. Then $\underline{d} = \frac{1}{2}(\underline{b} + \underline{c})$, $\underline{e} = \frac{1}{2}(\underline{a} + \underline{c})$, and $\underline{f} = \frac{1}{2}(\underline{a} + \underline{b})$.

Let N be the point corresponding to the vector $\underline{n} = \frac{1}{2}(\underline{a} + \underline{b} + \underline{c})$. This is the center of the circle through D, E and F, since the distance between N and each of these points is $\frac{1}{2} \| \underline{a} \|$:

$$\begin{aligned}
\| \underline{d} - \underline{n} \| &= \| \tfrac{1}{2}(\underline{b} + \underline{c}) - \tfrac{1}{2}(\underline{a} + \underline{b} + \underline{c}) \| = \tfrac{1}{2} \| \underline{a} \|, \\
\| \underline{e} - \underline{n} \| &= \| \tfrac{1}{2}(\underline{a} + \underline{c}) - \tfrac{1}{2}(\underline{a} + \underline{b} + \underline{c}) \| = \tfrac{1}{2} \| \underline{b} \|, \\
\| \underline{f} - \underline{n} \| &= \| \tfrac{1}{2}(\underline{a} + \underline{b}) - \tfrac{1}{2}(\underline{a} + \underline{b} + \underline{c}) \| = \tfrac{1}{2} \| \underline{c} \| .
\end{aligned}$$

The orthocenter H of $\triangle ABC$ turns out to correspond to the vector $\underline{h} = \underline{a} + \underline{b} + \underline{c}$. To prove this, we show that $\underline{h}$ is on the three altitudes. For instance, $\underline{h} - \underline{a} \perp \underline{c} - \underline{b}$ follows from

$$(\underline{h} - \underline{a}, \underline{c} - \underline{b}) = (\underline{b} + \underline{c}, \underline{c} - \underline{b}) = 0.$$

And similarly for the two other altitudes.

Let K be the midpoint of segment AH. Then $\underline{k} = \frac{1}{2}(\underline{a} + \underline{a} + \underline{b} + \underline{c})$. The distance between K and N equals

$$\| \underline{k} - \underline{n} \| = \| \frac{1}{2}(\underline{a} + \underline{a} + \underline{b} + \underline{c}) - \frac{1}{2}(\underline{a} + \underline{b} + \underline{c}) \| = \frac{1}{2} \| \underline{a} \| .$$

So K is also on the circumcircle of triangle DEF. Similar computations show that the midpoints L of BH and M of CH are on this circle.

That the circle also passes through the feet of the three altitudes is left as an exercise.

2.6.6 In later chapters we will handle rotations and reflections using vectors.

2.7 Notes

This chapter serves as a quick and slightly informal introduction to 'concrete' vectors in the plane and in space. In Chapter 4 the general notion of a vector space will be discussed. The notions and techniques discussed in this chapter (and their extensions presented in the following chapters) are of direct use in many branches of mathematics (algebra, analysis, statistics, optimization) and other disciplines like physics.

2.8 Exercises

§1

1 Given arbitrary vectors $\underline{u}$ and $\underline{v}$, draw the vectors

a. $2\underline{u} + 3\underline{v}$,

b. $\underline{u} - \underline{v}$.

2 Use the computational rules for vectors (and indicate which one you use in each step) to verify that

$$\underline{v}_1 + ((\underline{v}_2 + \underline{v}_3) + \underline{v}_4) = (\underline{v}_2 + \underline{v}_1) + (\underline{v}_4 + \underline{v}_3).$$

§2

3 Let $\underline{u}$ and $\underline{v}$ be two (distinct) vectors.

a. Why is
$$\underline{x} = \underline{u} + \lambda(\underline{v} - \underline{u})$$
a vector parametric equation of the line through (the endpoints of) $\underline{u}$ and $\underline{v}$?

b. Which of the following vector parametric equations describes the same line?
$$\underline{x} = (1-\lambda)\underline{u} + \lambda\underline{v},\ \ \underline{x} = \underline{v} + \mu(\underline{u} - \underline{v}),\ \ \underline{x} = 2\underline{v} - \underline{u} + \rho(\underline{u} - \underline{v}).$$

c. Is $-2\underline{u} + 3\underline{v}$ on the line?

4 Suppose $\underline{u}$, $\underline{v}$, $\underline{w}$ are distinct vectors in space.

a. Show that
$$\underline{x} = \underline{u} + \lambda(\underline{v} - \underline{u}) + \mu(\underline{w} - \underline{u})$$
is a vector parametric equation of the plane through $\underline{u}$, $\underline{v}$ and $\underline{w}$ (where we assume that none of the three vectors is on the line through the remaining two).

b. Which of the following is also a parametric equation of this plane?
$$\begin{aligned}\underline{x} &= (1-\lambda-\mu)\underline{u} + \lambda\underline{v} + \mu\underline{w},\\ \underline{x} &= \underline{v} + \lambda(\underline{v} - \underline{u}) + \mu(\underline{w} - \underline{u}),\\ \underline{x} &= \underline{u} + \lambda(\underline{w} - \underline{v}) + \mu(\underline{w} - \underline{u}).\end{aligned}$$

5 The line ℓ has parametric equation $\underline{x} = \underline{u} + \lambda(\underline{v} - \underline{u})$.

a. For which values of λ is $\underline{x}$ between $\underline{u}$ and $\underline{v}$?

b. For which value of λ is $\underline{x}$ the midpoint of the segment connecting the endpoints of $\underline{u}$ and $\underline{v}$?

c. If $\underline{x}$ divides the segment connecting the endpoints of $\underline{u}$ and $\underline{v}$ in the ratio $2:1$, then what is the value of λ?

§3

6 Determine a parametric equation for each of the lines in a) and b) and for each of the planes in c) and d).

a. The line passing through $(2,1,5)$ and $(5,-1,4)$.

b. The line passing through $(1,2)$ and $(2,4)$.

c. The plane passing through $(1,2,2)$, $(0,1,1)$ and $(1,3,2)$.

d. The plane containing the line $\underline{x} = (-2,1,3) + \lambda(1,2,-1)$ and the point $(4,0,3)$.

7 Determine whether $(3,4,0)$ is on the line with parametric description $\underline{x} = (1,2,1) + \lambda(2,2,-1)$. Are $\underline{x} = (3,4,0) + \lambda(2,2,-1)$ and $\underline{x} = (1,2,1) + \mu(-2,-2,1)$ vector parametric equations of the same line?

8 Determine an equation for each of the following lines.

a. $\underline{x} = (1,3) + \lambda(2,-1)$.

b. $\underline{x} = (2,2) + \lambda(1,-1)$.

c. $\underline{x} = (3,4) + \lambda(0,2)$.

9 Determine a parametric equation for each of the following lines in $\mathbb{R}^2$.

a. $2x_1 + 3x_2 = 3$.

b. $3x_1 - 4x_2 + 7 = 0$.

c. $2x_2 = 5$.

10 Determine an equation for each of the following planes.

a. $\underline{x} = (2,0,1) + \lambda(1,0,2) + \mu(1,-1,0)$.

b. $\underline{x} = (1,1,1) + \lambda(1,1,0) + \mu(0,1,1)$.

c. $\underline{x} = \lambda(4,1,1) + \mu(0,1,-1)$.

11 Determine a parametric equation for each of the following planes.

a. $x_1 + x_2 - 3x_3 = 5$.

b. $2x_1 + 3x_2 + 5x_3 = 0$.

c. $x_2 = 5$.

§4

12 Draw a vector $\underline{u}$ of length 2 in the plane. Draw all vectors in the plane having inner product 1 with $\underline{u}$.

13 Use the properties of the inner product to prove the following:

a. $(\lambda\underline{u}, \mu\underline{v}) = \lambda\mu(\underline{u}, \underline{v})$ for all vectors $\underline{u}$, $\underline{v}$ and all scalars λ and μ.

b. $(\underline{u} + \underline{v}, \underline{u} - \underline{v}) = \parallel \underline{u} \parallel^2 - \parallel \underline{v} \parallel^2$ for all vectors $\underline{u}$ and $\underline{v}$.

14 a. Compute the length of the vector $(-2, 2, 1)$.

b. Compute the distance between the vectors $(1, -1, 1)$ and $(1, -4, 5)$.

c. Compute the angle between the vectors $(1, 1, 2)$ and $(1, 1, -1)$.

d. Determine the number a so that the vector $(1, -2, a)$ is perpendicular to the vector $(3, 1, -1)$.

15 In each of the following cases determine an equation of the line passing through the given point and which is perpendicular to the given line. Determine in each case the distance between the given point and line.

a. $P = (3, 2)$ and $\ell : \underline{x} = (2, 1) + \lambda(1, -1)$.

b. $P = (1, 2)$ and $\ell : 3x_1 - 4x_2 = 20$.

16 The plane V has the following equation: $2x - y + 2z = 18$.

a. Determine the distance between $(0, 0, 0)$ and V.

b. The plane $W : 2x - y + 2z = 24$ is parallel to V. Compute the distance between V and W.

§5

17 Use the cross product to determine a normal vector and an equation of each of the following planes.

a. $\underline{x} = (1, 2, 2) + \lambda(1, -1, 0) + \mu(0, 1, 1)$.

b. $\underline{x} = (2, 1, 0) + \lambda(1, 2, 0) + \mu(0, 2, 3)$.

18 Compute using the cross product:

a. The surface area of the triangle with vertices $(1, 1, 0)$, $(2, 1, 1)$, $(1, 3, 3)$.

b. The surface area of the triangle with vertices $(2, 0)$, $(5, 1)$, $(1, 4)$.

c. The volume of the parallellepiped spanned by $(1, 1, 1)$, $(2, 2, 3)$, $(1, 0, 1)$.

§6

19 In 2.6.2 parametric vector descriptions were given of the three medians in a triangle. The values of the parameters corresponding to the centroid were not computed there, but simply given. In this exercise we take a look at the way these values can be computed.

a) Show that intersecting the medians through A and B, respectively, leads to the equation

$$(2 - 2\lambda - \mu)\underline{a} + (\lambda - 2 + 2\mu)\underline{b} + (\lambda - \mu)\underline{c} = \underline{0}.$$

b) Equation a) is satisfied if all coefficients are equal to 0. What does this mean for λ en μ? Show that ρ can be obtained in a similar way.

c) Does equation a) imply that all coefficients must be 0?

d) Take a look at c) under the assumption that the origin is *not* in the plane of the triangle.

20 Three non-collinear points determine a triangle. Similarly, four non-coplanar points in space determine a (not necessarily regular) tetrahedron.

a) Define, analogously to the notion of a median in a triangle, the notion of a median in a tetrahedron $ABCD$.

b) Show that the four medians in a tetrahedron are concurrent, i.e., pass through one point (the centroid), and describe the centroid in terms of vectors.

21 The *perpendicular bisector* of a segment is the line through the midpoint of the segment which is perpendicular to the segment. In this exercise we are going to prove that the three perpendicular bisectors of a triangle are concurrent, i.e., pass through a common point.

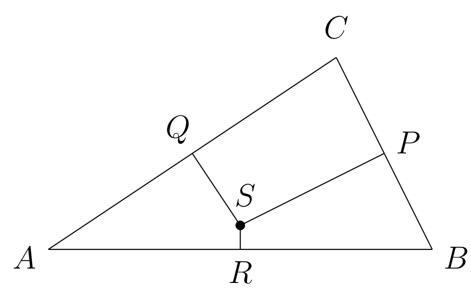


Figure 2.16: *The perpendicular bisectors of a triangle pass through a common point.*

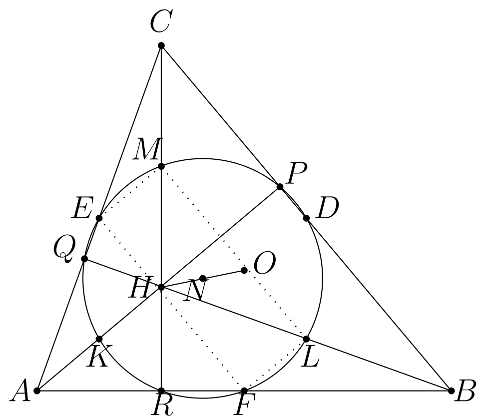
Let P, Q, R be the midpoints of the sides BC, AC, AB of triangle $\triangle ABC$.

a) Let S be the intersection point of the perpendicular bisectors of AC and BC. Which inner products with the vectors $\underline{s} - \underline{p}$ and $\underline{s} - \underline{q}$ must be 0?

b) Use the properties of the inner product to prove that $\underline{s} - \underline{r}$ is perpendicular to $\underline{a} - \underline{b}$. Conclusion?

c) Another property of a perpendicular bisector of a segment is that every point on it has equal distances to the endpoints of the segment. Use vector parametric equations of the perpendicular bisector of AC to show that each point on this bisector has equal distances to A and C.

22 (**The nine-point circle**) In this exercise we show that the feet of the altitudes of $\triangle ABC$ are also on the nine-point circle.



a) Use vectors to show that $EFLM$ is a rectangle (show that LF and EM are parallel and have equal length, that EF and ML are parallel and have equal length, and that $LF \perp EF$). Conclude that MF is a diameter of the nine-point circle.

b) Why is $MR \perp RF$? Conclude that R is on the nine-point circle (use Thales or show that $\| \underline{r} - \underline{n} \| = \frac{1}{2} \| \underline{a} \|$ using $(\underline{r} - \underline{f}, \underline{m} - \underline{r}) = 0$).

2.8.1 Exercises from old exams

23 Let ℓ: $\underline{x} = (3, 1, 2) + \lambda(1, 3, -2)$, $\underline{p} = (6, 10, -4)$, and let V be the plane with equation $x + y + z = 0$.

a) Show that $\underline{p}$ is on ℓ and determine the intersection of ℓ and V.

b) The perpendicular projection of the line ℓ on V, i.e., the collection of perpendicular projections of the vectors on ℓ) is a line. Determine a vector parametric equation of this line.

24 In $\triangle ABC$ (the points A, B, C are non-collinear) P is the midpoint of the segment BC and R is the point on the line AB such that A is the midpoint of the segment BR. Use vectors to determine the point of intersection Q of the lines PR and AC, and show that $AQ : QC = 1 : 2$.

25 Suppose ℓ is the line in $\mathbb{R}^3$ given by ℓ: $\underline{x} = (3,0,3) + \lambda(1,-2,2)$, and V is the plane given by the equation $2x + 2y + z = 0$.

a) Show that ℓ and V do not intersect.

b) Determine the distance between ℓ and V.

c) Suppose the plane W contains ℓ and suppose the vector $(1,0,7)$ is a direction vector of W. Determine a vector parametric equation of the line of intersection of V and W.

26 Let $\ell : \underline{x} = \lambda \underline{a}$ and $m : \underline{x} = \mu \underline{b}$ be two distinct lines in the plane. Suppose the vectors $\underline{a}$ and $\underline{b}$ both have length 1. Let $\underline{p} \neq \underline{0}$ be a vector such that its angle with $\underline{a}$ equals its angle with $\underline{b}$. Determine, for each real scalar α, the perpendicular projections of $\alpha\,\underline{p}$ on ℓ and m, respectively. Use this to prove that the distance between $\alpha\,\underline{p}$ and ℓ equals the distance between $\alpha \cdot \underline{p}$ and m.

27 Let V be the plane with equation $2x + y + 3z = 0$ and let ℓ be the line with parametric description $\underline{x} = (4,0,2) + \lambda(1,1,-1)$.

a) Show that ℓ and V do not intersect.

b) Determine the perpendicular projection of ℓ onto V.

28 Let ABC be a triangle in the plane (so A, B, C are not collinear). Suppose P is a point on the line AB such that A is the midpoint of the segment PB. Let R be the point on the segment BC such that $BR : RC = 2 : 1$. Choose a convenient origin and denote the vectors corresponding to points in the usual way: $\underline{c}$ corresponds to C, etc. Use vectors to determine the point of intersection Q of the lines PR and AC. Also determine the ratio $AQ : QC$.

Chapter 3

Matrices and systems of linear equations

3.1 Matrices

3.1.1 Matrices are rectangular arrays of numbers (or, more generally, elements from some arithmetical structure, like polynomials) which turn out to be useful in many places. In this chapter we discuss the arithmetic of matrices and the role of matrices in solving systems of linear equations.

This first section deals with

- the notion of a matrix,
- the addition and (scalar) multiplication of matrices,
- special matrices such as the zero matrix, the identity matrix, the transpose of a matrix,
- the inverse of a matrix.

3.1.2 What is a matrix?
A *matrix* is a rectangular array of numbers or elements from some arithmetical structure, like

$$A = \begin{pmatrix} 1 & 0 & 4 & -2 \\ 0 & 2 & 0 & 1 \end{pmatrix} .$$

In this example the matrix consists of 2 *rows* and 4 *columns*. If the matrix consists of n rows and m columns, then we say that the matrix is a n by m

or $n \times m$ matrix. So our example is a 2×4 matrix. We often denote matrices by capitals.

The numbers in a matrix are called the *entries*, *elements* or *coefficients* of the matrix. The elements of a matrix A are usually indicated by subindices in the following way: $A = (a_{ij})$ or $A = (A_{ij})$, where a_{ij} or A_{ij} denotes the element in the i^{th} row and the j^{th} column. If we denote the elements in the example by a_{ij}, then

$$a_{12} = a_{21} = a_{23} = 0\,, \quad a_{13} = 4\,.$$

Two matrices $A = (a_{ij})$ and $B = (b_{ij})$ are equal, $A = B$, if they have the same dimensions, i.e., A and B have the same number of rows and the same number of columns, and if all corresponding elements are equal, i.e., $a_{ij} = b_{ij}$ for all admissible i and j.

The set of all $n \times m$-matrices is denoted by $M_{n,m}$. In case it is relevant to know which set the coefficients of a matrix belong to, then this is denoted as follows: $M_{n,m}(\mathbb{R})$, $M_{n,m}(\mathbb{C})$, etc.

Note that, apart from numbers, variables and polynomials occasionally occur as entries in our matrices.

3.1.3 Matrix arithmetic: the addition of matrices

Matrices of the same dimension can be added in an obvious way (multiplication is discussed below).

Let $A = (a_{ij})$ and $B = (b_{ij})$ be two $n \times m$-matrices. Define, for all $1 \leq i \leq n$ and $1 \leq j \leq m$,

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

The $n \times m$-matrix C with entries c_{ij} is called the *sum* of the matrices A and B. We write $C = A + B$. If the dimensions of two matrices A and B are not the same, the sum is not defined.

Here are two examples:

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

The sum of $\left(\begin{smallmatrix} 1 & -1 & 3 \\ 2 & 0 & 4 \end{smallmatrix}\right)$ and $\left(\begin{smallmatrix} -1 & -1 \\ 4 & 2 \end{smallmatrix}\right)$ doesn't exist because the dimensions don't match.

3.1.4 Property. Let $A = (a_{ij})$, $B = (b_{ij})$, $C = (c_{ij})$ be three $n \times m$-matrices. It is straightforward to check the following two properties:

$$\begin{array}{ll} A + B = B + A & (\textit{commutativity}), \\ (A + B) + C = A + (B + C) & (\textit{associativity}). \end{array}$$

For instance, the second property can be proved as follows. First note that $A+B$, $B+C$, $(A+B)+C$, $A+(B+C)$ are all $n \times m$ matrices. Next note that the element in position ij of the matrix $(A+B)+C$ is $((A+B)+C)_{ij} = (A+B)_{ij} + c_{ij} = (a_{ij} + b_{ij}) + c_{ij}$, and that the element in position ij of the matrix $A + (B + C)$ equals $a_{ij} + (b_{ij} + c_{ij})$ (similar computation); of course, these numbers are equal (here we use the associativity of the real or complex numbers).

Due to the associativity we can just speak of $A+B+C$ without specifying which addition is carried out first, etc., since it doesn't matter for the result. Likewise we don't necessarily need brackets in expressions like $A+B+C+D$, since all ways of obtaining this sum, for instance as $(A+B)+(C+D)$ or as $A+(B+(C+D))$, lead to the same result.

3.1.5 Matrix arithmetic: scalar multiplication

There is also a useful way of multiplying matrices by numbers (scalars), called *scalar multiplication*. Let λ be a number and let A be a $n \times m$-matrix. The matrix λA is obtained by multiplying *every* element of A by λ. For example,

$$2 \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix} = \begin{pmatrix} 2 & 4 & 6 \\ 8 & 10 & 12 \end{pmatrix}.$$

In general terms, by definition the element $(\lambda A)_{ij}$ in position ij of λA equals $\lambda \cdot a_{ij}$, where $A = (a_{ij})$.

We have the following properties for scalar multiplication (for all scalars λ, μ and all matrices A and B of the same dimension):

$$\begin{aligned} 1\,A &= A, \\ (\lambda + \mu)A &= \lambda A + \mu A, \\ \lambda(A + B) &= \lambda A + \lambda B, \\ \lambda(\mu A) &= (\lambda\mu)A. \end{aligned}$$

The verifications are easy exercises. For instance, the last property is proved by comparing the elements in position ij of both sides (for all i, j):

$$(\lambda(\mu A))_{ij} = \lambda \cdot (\mu A)_{ij} = \lambda(\mu \cdot a_{ij}) = (\lambda\mu) \cdot a_{ij} = ((\lambda\mu)A)_{ij}.$$

3.1.6 Matrix arithmetic: multiplication

Multiplying matrices is more complicated and, at least in the beginning, not so intuitive.

To begin with, we only define the product AB of two matrices A and B if the *rows* of A have the same length as the *columns* of B. So if A is a $m \times n$-matrix, then B has to be a $n \times p$-matrix for some p. The *product matrix* $C = AB$ is defined by

$$c_{ij} = a_{i1}b_{1j} + a_{i2}b_{2j} + \ldots + a_{in}b_{nj} \quad i = 1, \ldots, m;\ j = 1, \ldots, p.$$

So the matrix C is a $m \times p$-matrix.

3.1.7 Examples. Here are some examples that can be verified using the definition of the product of matrices.

- $$\begin{pmatrix} 1 & 2 & -1 \\ 0 & 1 & 1 \end{pmatrix} \begin{pmatrix} 1 & -1 \\ -2 & 0 \\ 3 & 0 \end{pmatrix} = \begin{pmatrix} -6 & -1 \\ 1 & 0 \end{pmatrix},$$
- $$\begin{pmatrix} 1 & -1 \\ -2 & 0 \\ 3 & 0 \end{pmatrix} \begin{pmatrix} 1 & 2 & -1 \\ 0 & 1 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 1 & -2 \\ -2 & -4 & 2 \\ 3 & 6 & -3 \end{pmatrix}.$$

In particular, we observe that AB and BA are not necessarily the same. We say that matrix multiplication is not commutative. In the example AB and BA even have distinct sizes.

- Even if A and B are *square* matrices of the same size, i.e., both are $n \times n$ matrices for some n, then AB and BA may still differ:

$$\begin{pmatrix} 1 & -3 \\ 3 & 4 \end{pmatrix} \begin{pmatrix} -1 & 2 \\ 1 & 5 \end{pmatrix} = \begin{pmatrix} -4 & -13 \\ 1 & 26 \end{pmatrix},$$

$$\begin{pmatrix} -1 & 2 \\ 1 & 5 \end{pmatrix} \begin{pmatrix} 1 & -3 \\ 3 & 4 \end{pmatrix} = \begin{pmatrix} 5 & 11 \\ 16 & 17 \end{pmatrix}.$$

- If AB exists, then BA need not exist:

$$\begin{pmatrix} 1 & 2 \\ 3 & 1 \end{pmatrix} \begin{pmatrix} -1 \\ 2 \end{pmatrix} = \begin{pmatrix} 3 \\ -1 \end{pmatrix},$$

$$\begin{pmatrix} -1 \\ 2 \end{pmatrix} \begin{pmatrix} 1 & 2 \\ 3 & 1 \end{pmatrix} \quad \text{does not exist!}$$

3.1.8 Property. For matrices of the correct dimensions, various arithmetic rules, similar to those for ordinary real or complex numbers, hold. Here are the most important ones.

$$\begin{gathered} A(B+C) = AB + AC \text{ and } (E+F)G = EG + FG\ , \\ (\lambda A)B = \lambda(AB)\ , \\ \lambda(\mu A) = (\lambda\mu)A, \\ (AB)C = A(BC)\ . \end{gathered}$$

These rules follow from the definitions, but especially the third one requires some effort. When we deal with linear transformations, we will discuss an easy proof.

As a consequence of these rules we can, for example, simply write λAB instead of $(\lambda A)B$ or $\lambda(AB)$. Similarly, we can write ABC instead of $(AB)C$ or $A(BC)$. Of course, putting in brackets may sometimes be useful to clarify a computation.

3.1.9 The zero matrix

The $n \times m$ matrix all of whose entries are 0 is called the *$n \times m$-zero matrix*. We usually denote this matrix by O. For every $n \times m$ matrix A we have $A + O = O + A = A$. Note: for every size $n \times m$ there is exactly one zero matrix with that size. To avoid confusion we sometimes write $O_{n,m}$ for the $n \times m$ zero matrix.

3.1.10 The opposite matrix

The matrix $B = (-1)A$ satisfies $A + B = O$ and is called the (additive) *opposite* of A. Instead of $(-1)A$ we usually write $-A$.

3.1.11 The identity matrix

The $n \times n$-matrix

$$I = \begin{pmatrix} 1 & 0 & \ldots & 0 \\ 0 & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \ldots & 0 & 1 \end{pmatrix}$$

satisfies the following property: $IA = AI = A$ for every $n \times n$-matrix A. This is easily verified using the definition of matrix multiplication (and it is a good exercise to first try the case $n = 2$). The matrix I is called the *$n \times n$-identity matrix*. It plays a similar role in matrix arithmetic as the number 1 in multiplications of numbers. If necessary, we emphasize the dimensions by writing I_n.

3.1.12 The (multiplicative) inverse of a matrix

If, given a $n \times n$ matrix A, there exists a matrix $n \times n$ B with $AB = BA = I$, then B is called the *inverse* of A. We usually denote this inverse by A^{-1}. This is justified by the fact that such a B is unique: for if B' also satisfies $AB' = B'A = I$, then $B = BI = B(AB') = (BA)B' = IB' = B'$. Also note that is B is the inverse of A, then A is of course the inverse of B.

We really have to impose two conditions, $AB = I$ and $BA = I$, since matrix multiplication is not commutative, and so it is not obvious that once a $n \times n$ matrix B satisfies $AB = I$, it also satisfies $BA = I$.

Fortunately, using linear transformations (Linear Algebra 2), it is fairly easy to show for a $n \times n$ matrix B that $AB = I$ implies $BA = I$ (and conversely, if $BA = I$ then $AB = I$). So, if $AB = I$ (or $BA = I$), then B is an inverse of A. Without using linear transformations, the proof is more complicated, which is why we skip it here.

The square matrix 0 has no inverse since $OB = BO = O$ for every matrix B (of the same size). But there also exist non-zero matrices without an inverse.

3.1.13 Example. The following two matrices are each other's inverse:

$$\begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}, \quad \begin{pmatrix} 1 & -1 \\ -1 & 2 \end{pmatrix}.$$

Next consider

$$A = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}.$$

If

$$B = \begin{pmatrix} x & u \\ y & v \end{pmatrix}$$

is the inverse of A, then

$$\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} x & u \\ y & v \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix},$$

so

$$x + y = 1,\ x + y = 0,\ u + v = 0,\ u + v = 1.$$

It is clear that there are no solutions for x, y, u, v.

3.1.14 Let A and B be $n \times n$ matrices and suppose that A^{-1} and B^{-1} exist. Then

$$(A\,B)^{-1} = B^{-1}A^{-1}\ ,$$

since

$$(AB)(B^{-1}A^{-1}) = A(B\,B^{-1})A^{-1} = A(I\,A^{-1}) = A\,A^{-1} = I$$

and, similarly, $(B^{-1}A^{-1})(AB) = I$.

3.1.15 The transpose of a matrix
If $A = (a_{ij})$ is a $n \times m$ matrix, then its *transpose* A^T is the $m \times n$ matrix whose i-th row equals the i-th column of A (for $i = 1, \ldots, m$), so the i, j-th entry of A^T equals a_{ji}. The j-th column is then automatically equal to the j-th row of A. In other words, you get the matrix A^T from the matrix A by interchanging the roles of rows and columns.

3.1.16 Examples.

$$A = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix}, \quad A^T = \begin{pmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{pmatrix}\ .$$

$$A = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}, \quad A^T = \begin{pmatrix} 1\,2\,3 \end{pmatrix}\ .$$

In short, transposing is 'taking the mirror image with respect to the so-called main diagonal' (the main diagonal consists of the elements with indices $11, 22, \ldots$).

3.1.17 Property. It follows directly from the definition that the following rules hold (supposing in each case that the operations can be carried out):

$$\begin{aligned} (A+B)^T &= A^T + B^T, \\ (\lambda A)^T &= \lambda A^T, \\ (A\,B)^T &= B^T A^T, \\ (A^T)^T &= A. \end{aligned}$$

3.2 Row reduction

Row reduction refers to a useful type of algorithm on matrices, that enables us, as one of its main applications, to solve systems of linear equations in a systematic and efficient way. In this section we discuss the details about row reduction, and in the next section we discuss its application to solving systems of linear equations. More applications will follow in later chapters.

3.2.1 Row reduction

The main ingredient of row reduction consists of the following three so-called *elementary row operations* that can be applied to (the rows of) a given matrix:

- Interchange the order of the rows (in particular, interchange two rows).
- Multiply every entry in a row by a nonzero constant.
- Replace a row by the sum of this row and a scalar multiple of another row.

These row operations are inspired by the process of solving systems of linear equations, in which interchanging equations, multiplying equations by a scalar, and adding a multiple of an equation to another, are used to simplify and solve the equations. The relation between row operations and solving systems of linear equations is discussed in the next section.

First we discuss an example of how to use these elementary operations to change the given matrix into a special form with 'many zeros'.

3.2.2 Example. Consider the matrix

$$A = \begin{pmatrix} 1 & 2 & -4 & 8 \\ -1 & -2 & 6 & -4 \\ 1 & 4 & -2 & 0 \end{pmatrix} .$$

We use the first row to get as many zeros as possible in the first column. Therefore we add the first row to the second, and subtract it from the third row (we work from top to bottom). We find:

$$\begin{pmatrix} 1 & 2 & -4 & 8 \\ 0 & 0 & 2 & 4 \\ 0 & 2 & 2 & -8 \end{pmatrix} .$$

Next, we try to achieve the same in the second column without ruining the first column. So we don't use the first row, but instead interchange the second and the third row:

$$\begin{pmatrix} 1 & 2 & -4 & 8 \\ 0 & 2 & 2 & -8 \\ 0 & 0 & 2 & 4 \end{pmatrix}.$$

Then we divide the second row by 2:

$$\begin{pmatrix} 1 & 2 & -4 & 8 \\ 0 & 1 & 1 & -4 \\ 0 & 0 & 2 & 4 \end{pmatrix}.$$

Now we can use the second row to produce zeros in the second column. So we subtract the second row twice from the first (note that this doesn't affect the first column!):

$$\begin{pmatrix} 1 & 0 & -6 & 16 \\ 0 & 1 & 1 & -4 \\ 0 & 0 & 2 & 4 \end{pmatrix}.$$

In the next step, we use the third row. We first divide it by 2,

$$\begin{pmatrix} 1 & 0 & -6 & 16 \\ 0 & 1 & 1 & -4 \\ 0 & 0 & 1 & 2 \end{pmatrix}$$

and then add the new third row 6 times to the first row, and subtract it from the second row. Note that this doesn't alter the first two columns.

$$\begin{pmatrix} 1 & 0 & 0 & 28 \\ 0 & 1 & 0 & -6 \\ 0 & 0 & 1 & 2 \end{pmatrix}.$$

We can't go any further since that would affect the first three columns. The matrix obtained is called the (*row*) *reduced echelon form* of A.

3.2.3 Row reduction: description of the steps

We now turn to the general description of the row reduction procedure in the form of an 'algorithm.' Do compare this description with the example just given.

The *first step* consists of the following.

- Let n_1 be the index of the first column (from the left) that contains a non-zero element.
- If necessary interchange two rows so that the first element of the n_1-th column is non-zero.
- Divide each element of the first row by the first element of the n_1-th column so that we obtain a situation with $a_{1n_1} = 1$.
- Use the first row to produce zeros in all other entries of the n_1-th column.

Now suppose we have carried out m steps of this kind. In the resulting matrix the first m rows have been used and the last column we have dealt with is the n_m-th. Then we do the following:

- Let n_{m+1} be the index of the first column that contains a non-zero element in one of the spots with index at least $m+1$.
- If necessary interchange the $m+1$-th row with one of the next rows so that the $m+1$-th element of the n_{m+1}-th column is non-zero.
- Divide the $m+1$-th row by this element so that $a_{m+1,n_{m+1}} = 1$.
- Use the $m+1$-th row to produce zeros in the other entries of the n_{m+1}-th column.

This process stops if all rows have been used or if we are left with rows consisting of zeros only.

The result of these row reduction steps is a matrix in so-called *row reduced echelon form* or simply *reduced echelon form*. It looks as follows in the first case:

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

and as follows in the second case:

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

The entries with a $*$ can be any numbers.

3.2.4 The shape of a row reduced matrix

A matrix in row reduced form has the following properties:

- Every row starts with (possibly zero) zeros. Its first nonzero entry (if there is any) is 1 (its *leading entry*). The column containing this 1 has zeros in all other entries.
- Every non-zero row starts with more zeros than the row directly above it In particular, if there are any 'zero rows' (rows consisting of zeros only), they are all below the non-zero rows.

The matrix

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

is not in row reduced form, because the second row doesn't satisfy the first condition: there is a -1 above the 1 in the second column.

3.3 Systems of linear equations

3.3.1 In this section we explain the connection between row reducing matrices and solving systems of linear equations. The resulting procedure for solving such systems is also called *Gaussian elimination* after C.F. Gauss (1777-1855).

3.3.2 Systems of linear equations

The equation

$$3x_1 - 4x_2 + 5x_3 = 7$$

is called a *linear equation* in the variables x_1, x_2, x_3. The numbers 3, -4, 5, 7 are called the *coefficients* of this equation; the number 7 is also called the right-hand side of the equation. If we are dealing with several such equations, then we speak of a system of linear equations.

In such a *system of linear equations*

$$\begin{array}{llllllll} a_{11} & x_1 + & a_{12} & x_2 & + \ldots + & a_{1m} & x_m & = b_1 \\ a_{21} & x_1 + & a_{22} & x_2 & + \ldots + & a_{2m} & x_m & = b_2 \\ \vdots & & & & & \vdots & & \vdots \\ a_{n1} & x_1 + & a_{n2} & x_2 & + \ldots + & a_{nm} & x_m & = b_n \end{array} \tag{3.1}$$

the matrix

$$A = \begin{pmatrix} a_{11} & \ldots & a_{1m} \\ \vdots & & \vdots \\ a_{n1} & \ldots & a_{nm} \end{pmatrix}$$

is called the *coefficient matrix* and the row $\underline{b} = (b_1, \ldots, b_n)$ (or column $(b_1, \ldots, b_n)^\top$) is called the *right-hand side*. If $b_i = 0$ for all i then the system is called *homogeneous*, and otherwis it is called *inhomogeneous*. Note that we can write the system (3.1) as follows in matrix form:

$$A \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix}.$$

A sequence of numbers $(p_1, \ldots, p_n)$ is called a *solution* of (3.1) if

$$A \begin{pmatrix} p_1 \\ \vdots \\ p_n \end{pmatrix} = \begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix}.$$

We can represent the system (3.1) by the matrix

$$\begin{pmatrix} a_{11} & \ldots & a_{1m} & b_1 \\ \vdots & & \vdots & \vdots \\ a_{n1} & \ldots & a_{nm} & b_n \end{pmatrix}.$$

containing all the a_{ij} and b_k. This matrix is often denoted as $(A|\underline{b})$ and is called the *extended coefficient matrix* of the system. The vertical bar is sometimes used to distinguish between the the two types of coefficients.

3.3.3 Examples. If the matrix $(A|\underline{b})$ is in row reduced form, then it is easy to describe the solutions of the system. Here are some examples.

- $\left(\begin{array}{cc|c} 1 & 0 & 2 \\ 0 & 1 & 3 \end{array}\right)$ so $\begin{array}{l} x_1 = 2, \\ x_2 = 3. \end{array}$
- $\left(\begin{array}{ccc|c} 1 & 0 & 5 & 0 \\ 0 & 1 & -2 & 0 \\ 0 & 0 & 0 & 1 \end{array}\right)$. The last equation has the form $0x_1 + 0x_2 + 0x_3 = 1$ or $0 = 1$. This equation has no solutions. We call the system *inconsistent*.
- $\left(\begin{array}{ccc|c} 1 & 0 & 5 & 2 \\ 0 & 1 & -2 & 3 \\ 0 & 0 & 0 & 0 \end{array}\right)$. Every triple (p_1, p_2, p_3) satisfies the last equation, so that we can just as well leave out this equation. What remains is:
$$\begin{array}{cccc} x_1 & & +5x_3 & = 2 \ , \\ & x_2 & -2x_3 & = 3 \ . \end{array}$$
We now assign x_3 (a variable from a column we can't do anything with) an arbitrary value, say λ. Then we find
$$\begin{array}{l} x_1 = 2 - 5\lambda, \\ x_2 = 3 + 2\lambda, \end{array}$$
so that
$$(x_1, x_2, x_3) = (2, 3, 0) + \lambda(-5, 2, 1).$$

3.3.4 Solutions of systems in row reduced form
Here is the general procedure for writing down the solutions of a system. Suppose that $(A|\underline{b})$ is in row reduced form.

$$\left[\begin{array}{cccccccccccccccccc}
0 & \dots & 0 & 1 & * & \dots & * & 0 & * & \dots & * & 0 & * & \dots & 0 & * & \dots & * \\
0 & \dots & 0 & 0 & 0 & \dots & 0 & 1 & * & \dots & * & 0 & * & \dots & 0 & * & \dots & * \\
0 & \dots & 0 & 0 & 0 & \dots & 0 & 0 & 0 & \dots & 0 & 1 & * & \dots & 0 & * & \dots & * \\
\dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\
\dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & 0 & * & \dots & * \\
0 & \dots & 0 & 0 & 0 & \dots & 0 & 0 & 0 & \dots & 0 & 0 & 0 & \dots & 1 & * & \dots & * \\
0 & \dots & 0 & 0 & 0 & \dots & 0 & 0 & 0 & \dots & 0 & 0 & 0 & \dots & 0 & 0 & \dots & 0 \\
\vdots & & \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots & \vdots & & \vdots \\
0 & \dots & 0 & 0 & 0 & \dots & 0 & 0 & 0 & \dots & 0 & 0 & 0 & \dots & 0 & 0 & \dots & 0
\end{array}\right]$$

- If a row $(0, \dots, 0, 1)$ occurs then the corresponding equation and therefore the system of equations no solutions; the system is inconsistent.
- Assign parameters to every variable corresponding to a column with an $*$ (when the system is in row reduced echelon form). Then solve the remaining variables (corresponding to columns with a 1 in the row reduced echelon form). We then find:
  - Exactly one solution if in the row reduced echelon form no column with an *ast* occurs,
  - ∞ many solutions otherwise.

We return to this below.

3.3.5 Row operations in terms of the system of equations

Next we turn to the situation of a system of linear equations (3.1) where $(A|\underline{b})$ is not in row reduced echelon form. The result we will deduce below is that the solutions of a system do not change when we apply row operations to the corresponding matrix: interchanging rows (equations), multiplying a row (equation) by a non-zero scalar, adding a multiple of a row (equation) to one of the other rows (equations).

Consider the equations

$$v : a_1x_1 + a_2x_2 + \cdots + a_mx_m = b \quad \text{and} \quad w : c_1x_1 + c_2x_2 + \cdots + c_mx_m = d.$$

We define the sum of these equations by

$$v + w : (a_1 + c_1)x_1 + \cdots + (a_m + c_m)x_m = b + d$$

and, for an arbitrary (real or complex) scalar α, the scalar product by

$$\alpha v : \alpha a_1x_1 + \cdots + \alpha a_mx_m = \alpha b\ .$$

Note that if we represent the equations v and w by the rows

$$\begin{aligned} v &: (a_1, a_2, \dots, a_m, b), \\ w &: (c_1, c_2, \dots, c_m, d), \end{aligned}$$

the equation $v + w$ is represented by the sum of the rows

$$v + w : (a_1 + c_1, a_2 + c_2, \dots a_m + c_m, b + d),$$

and the equation αv is represented by the scalar product

$$\alpha v : (\alpha a_1, \alpha a_2, \ldots, \alpha a_m, \alpha b) \ .$$

3.3.6 Applying row operations doesn't change the solutions of the system
For the technique of applying row operations to work, it is essential that in each step the solution set remains the same. We show this by proving that each type of row operation doesn't change the solution set.

- **Changing the order of the equations.**
  Since we are not changing the individual equations, it is immediately clear that the solutions do not change when we change the order of the equations.

- **Multiplication of an equation by a non-zero factor.**
  Let $(p_1, \ldots, p_m)$ be a solution of the equation

  $$v : a_1x_1 + \cdots + a_mx_m = b \ .$$

  This means that

  $$a_1p_1 + \cdots + a_mp_m = b \ .$$

  Now let α be number different from 0. Multiplying left-hand side and righ-hand side by α produces the equality

  $$\alpha a_1p_1 + \cdots + \alpha a_mp_m = \alpha b \ ,$$

  so that $(p_1, \ldots, p_m)$ is a solution of the eqution αv. So every solution of the equation v is also a solution of the equation αv. For the same reason, every solution of the equation αv is a solution of $\frac{1}{\alpha}\alpha v = v$, because $\alpha \neq 0$. So the equations v and αv have the same solutions. In conclusion, if we multiply one of the equations by a non-zero scalar, the solutions of the equation do not change. And then the solution set of the system does not change.

- **Replacing the equations v and w by v and $v + \beta w$.**
  Let $(p_1, \ldots, p_m)$ be a solution of the equations

  $$v : a_1x_1 + \cdots + a_mx_m = b \quad \text{and} \quad w : c_1x_1 + \cdots + c_mx_m = d.$$

  This means that

  $$a_1p_1 + \cdots + a_mp_m = b \quad \text{and} \quad c_1p_1 + \cdots + c_mp_m = d.$$

From these two equalities we infer that for any scalar β

$$(c_1 + \beta a_1)p_1 + \cdots + (c_m + \beta a_m)p_m = d + \beta b \ .$$

So every solution of the equations v and w is also a solution of the equations v and $w+\beta v$. For the same reason we have that every solution of the equations v and $w + \beta v$ is also a solution of the equations v en $(w + \beta v) - \beta v$, i.e., a solution of the equations v and w.

The system of equations v and w therefore has the same solutions as the equations v and $w + \beta v$. So the solution set of a system doesn't change if we add a scalar multiple of one of the equations to one of the other equations.

Since the process of row reducing consists of applying (at the level of matrices) such operations consecutively, we conclude that applying row operations doesn't change the solution set of a system of equations.

3.3.7 The procedure for solving a system of linear equations

Form the above discussion we arrive at the following procedure for solving a system of linear equations:

- Start with the extended coefficient matrix $(A|\underline{b})$ corresponding to the system,
- Use row operations to determine the row reduced echelon form,
- Then determine the solutions.

3.3.8 Example. To solve the system

$$\begin{array}{rcrcrl} x_1 & + & 2x_2 & + & 3x_3 & = 2 \ , \\ & & x_2 & + & 2x_3 & = 1 \ , \\ 3x_1 & + & x_2 & + & x_3 & = 3 \end{array}$$

we first represent it in matrix form:

$$\begin{pmatrix} 1 & 2 & 3 & 2 \\ 0 & 1 & 2 & 1 \\ 3 & 1 & 1 & 3 \end{pmatrix} .$$

Applying row operations yields

$$\begin{pmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & 1 \end{pmatrix} .$$

So the system has precisely one solution:

$$(x_1, x_2, x_3) = (1, -1, 1) .$$

3.3.9 Example. The system

$$\begin{aligned} x_1 + 2x_2 + 3x_3 - x_4 &= 1 , \\ 2x_1 + 3x_2 - 2x_3 + 3x_4 &= 1 , \\ 4x_1 + 7x_2 + 4x_3 + x_4 &= 3 \end{aligned}$$

has the following matrix representation:

$$\begin{pmatrix} 1 & 2 & 3 & -1 & 1 \\ 2 & 3 & -2 & 3 & 1 \\ 4 & 7 & 4 & 1 & 3 \end{pmatrix} .$$

Applying row operations produces the following row reduced echelon form

$$\begin{pmatrix} 1 & 0 & -13 & 9 & -1 \\ 0 & 1 & 8 & -5 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix} .$$

We now assign the values λ and μ to the variables x_3 and x_4, respectively: $x_3 = \lambda$ and $x_4 = \mu$. Then

$$x_1 = -1 - 9\mu + 13\lambda ,$$

$$x_2 = 1 + 5\mu - 8\lambda ,$$

so that

$$(x_1, x_2, x_3, x_4) = (-1, 1, 0, 0) + \lambda(13, -8, 1, 0) + \mu(-9, 5, 0, 1) .$$

The advantage of this latter way of describing the solutions is that it shows that the solution set is a 'plane in 4-dimensional space.' We'll return to this in the chapter on vectors spaces.

3.3.10 Remark. a) One can prove that the row reduced echelon form of a matrix is unique: in whatever way you apply the row operations, you'll always end up with the same row reduced echelon form. A proof can be found in Thomas Yuster, *The reduced row echelon form of a matrix is unique: A simple proof*, Mathematics Magazine, vol. 57, No. 2 (1984).

b) Of course, in practical situations it is not always necessary to find the row reduced echelon form of a system of equations in order to find the solutions. Often, the solutions can be read off some steps before you reach this form.

3.4 Notes

James Joseph Sylvester (1814–1897) introduced the term *matrix* for a rectangular array of numbers. In the Philosophical Magazine (1851) he wrote: "I have in previous papers defined a "Matrix" as a rectangular array of terms, out of which different systems of determinants may be engendered, as from the womb of a common parent". Determinants will be discussed in Chapter 5.

Matrices turn out to be a useful way of storing and handling data. In this chapter, we have used them to store and manipulate the coefficients of systems of linear equations. We will come across various other usages of matrices in the following chapters (by the way, they are also used in many other mathematics courses). The importance of matrices is in the arithmetic operations like addition and multiplication that allow for efficient handling of data.

In a way linear equations are the simplest kind of equations in mathematics. They belong to the few classes of equations for which, in theory, a complete solution procedure exists, Gaussian elimination, the details of which have been discussed in this chapter. This procedure is not the end of the story regarding linear equations, since, for instance, equations whose coefficients vary very much in size, or systems with a very large number of equations cause different problems. Such systems often occur in practice and the mathematics to deal with them is discussed in the course *Numerical Linear Algebra*.

Numerical Linear Algebra

Linear equations are a relatively simple form of polynomial equations. The latter type is much more difficult to handle. Algorithms to determine exact solutions of systems of polynomial equations are discussed in the course *Algoritms in Algebra and Number Theory*.

Algebra

Many problems in linear algebra (and also outside linear algebra) are closely related to solving systems of linear equations. Knowing the techniques to solve them is essential for a fruitful study of the remaining chapters, but is also useful for other courses.

3.5 Exercises

§1

1 Determine AB, BA, $A(B-2C)$, AD, $CC^\top$, $C^\top C$, $DD^\top$, $D^\top D$, where

$$A = \begin{pmatrix} 1 & -1 & 2 \\ 0 & 3 & 4 \end{pmatrix},\ B = \begin{pmatrix} 4 & -1 \\ 0 & -2 \\ -3 & 3 \end{pmatrix},\ C = \begin{pmatrix} 2 & 2 \\ 1 & -1 \\ 1 & -3 \end{pmatrix},\ D = \begin{pmatrix} 2 \\ -1 \\ 3 \end{pmatrix}.$$

2 Determine $A+B$, $(A-B)C$, $A^\top B$, $AA^\top$, $A^\top C^\top$, where

$$A = \begin{pmatrix} 1+i & i & -i \\ 1 & -1 & 1 \end{pmatrix},\ B = \begin{pmatrix} 1-i & -i & i \\ 1 & -1 & 1 \end{pmatrix},\ C = \begin{pmatrix} 1-i & 1 \\ i & 1 \\ i & 1 \end{pmatrix}.$$

3 The 2×3 matrices $A = (a_{kl})$ and $B = (b_{kl})$ are given by $a_{kl} = k+l$, $b_{kl} = k-l$. Determine $A+B$, $A-2B$, $A^\top B$, $AB^\top$.

4 The 3×2 matrices $A = (a_{kl})$ and $B = (b_{kl})$ are given by $a_{kl} = k+li$, $b_{kl} = k-li$. Determine $A+B$, $A-B$, $A^\top B$, $AB^\top$.

5 a. Compute the inverse of the matrix

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

b. Suppose A is invertible with inverse A^{-1}. Determine the inverse of each of the following matrices: λA ($\lambda \neq 0$), A^2, $A^\top$, A^{-1}.

6 Let V be the set of matrices of the form

$$\begin{pmatrix} a & -b \\ b & a \end{pmatrix} \text{ met } a, b \text{ reëel.}$$

a. Show that $A+B \in V$ and $AB \in V$ for alle $A \in V$ and $B \in V$.

b. Compare, for real numbers a, b, c, d, the sum and the product of the complex numbers $a+bi$, $c+di$ with the sum and the product of the matrices

$$\begin{pmatrix} a & -b \\ b & a \end{pmatrix},\quad \begin{pmatrix} c & -d \\ d & c \end{pmatrix}.$$

Conclude that the matrices in V ‘behave’ similarly with respect to addition and multiplication as the complex numbers.

c. Prove that

$$\begin{pmatrix} \cos\varphi & -\sin\varphi \\ \sin\varphi & \cos\varphi \end{pmatrix}^n = \begin{pmatrix} \cos n\varphi & -\sin n\varphi \\ \sin n\varphi & \cos n\varphi \end{pmatrix}$$

for all positive integers n.

§2

7 Use row reduction to transform the following matrices into row reduced echelon form.

a.

$$\begin{pmatrix} 1 & 2 & -3 & -11 \\ 2 & 5 & -5 & -11 \\ -1 & -1 & 7 & 43 \end{pmatrix},$$

b.

$$\begin{pmatrix} 0 & 0 & 1 & 1 & 3 \\ 1 & 2 & 2 & 2 & 8 \\ 1 & 2 & 3 & 3 & 11 \end{pmatrix}.$$

8 The operations used in row reduction can also be brought about by multiplying with suitable matrices. This connection is discussed in this exercise.

a. In the 3×3 identity matrix interchange the 2nd and 3rd row. Let E be the resulting matrix. Now compute the product

$$E\begin{pmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \end{pmatrix}.$$

Find by analogy the matrix you need to multiply with (from the left or from the right?) to accomplish swapping the i-th and j-th rows of an $m \times n$ matrix.

b. In the 3×3 identity matrix multiply the 2nd row by 7. Let F be the resulting matrix. Now compute the product

$$F\begin{pmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \end{pmatrix}.$$

Find by analogy the matrix you need to multiply with (from the left or from the right?) to accomplish multiplication of the i-th row of an $m \times n$ matrix by λ.

c. In the 3×3 identity matrix add 5 times the 3-rd row to the first row and call the resulting matrix G. Compute the product

$$G\begin{pmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \end{pmatrix}.$$

Find by analogy the matrix you need to multiply with (from the left or from the right?) so that in a $m \times n$ matrix λ times the i-th row is added to the j-th row.

§3

9 Solve each of the following systems of linear equations.

a.

$$\begin{array}{rrrrl} x_1 & +2x_2 & +3x_3 & -x_4 & = 0, \\ 2x_1 & +3x_2 & -x_3 & +3x_4 & = 0, \\ 4x_1 & +6x_2 & +x_3 & +2x_4 & = 0; \end{array}$$

b.

$$\begin{array}{rrrrl} 3x_1 & +x_2 & +2x_3 & -x_4 & = 0, \\ 2x_1 & -x_2 & +x_3 & +x_4 & = 0, \\ 5x_1 & +5x_2 & +4x_3 & -5x_4 & = 0, \\ 2x_1 & +9x_2 & +3x_3 & -9x_4 & = 0; \end{array}$$

c.

$$\begin{array}{rrrrl} x_1 & -x_2 & +x_3 & +2x_4 & = 2, \\ 2x_1 & -3x_2 & +4x_3 & -x_4 & = 3, \\ x_1 & & -x_3 & +7x_4 & = 3. \end{array}$$

10 Solve each of the following systems of linear equations.

a.

$$\begin{array}{rrrl} & x_2 & +2x_3 & = 1, \\ x_1 & +2x_2 & +3x_3 & = 2, \\ 3x_1 & +x_2 & +x_3 & = 3; \end{array}$$

b.

$$\begin{array}{rrrrrl} x_1 & +x_2 & +2x_3 & +3x_4 & -2x_5 & = 1, \\ 2x_1 & +4x_2 & & & -8x_5 & = 3, \\ & -2x_2 & +4x_3 & +6x_4 & +4x_5 & = 0; \end{array}$$

c.

$$\begin{array}{rrrl} x_1 & +2x_2 & & = 0, \\ x_1 & +4x_2 & -2x_3 & = 4, \\ 2x_1 & & +4x_3 & = -8, \\ 3x_1 & & +6x_3 & = -12, \\ -2x_1 & -8x_2 & +4x_3 & = -8; \end{array}$$

d.

$$\begin{array}{rrrl} x_1 & +x_2 & -2x_3 & = 0, \\ 2x_1 & +x_2 & -3x_3 & = 0, \\ 4x_1 & -2x_2 & -2x_3 & = 0, \\ 6x_1 & -x_2 & -5x_3 & = 0, \\ 7x_1 & -3x_2 & -4x_3 & = 1. \end{array}$$

11 Solve each of the following systems of linear equations.

a.

$$\begin{array}{rrrl} z_1 & +iz_2 & +z_3 & = 1, \\ & z_2+ & (i+1)z_3 & = 0, \\ -iz_1 & +z_2 & & = 0; \end{array}$$

b.

$$\begin{array}{rrl} (1-\lambda)z_1 & -2z_2 & = 0, \\ 5z_1 & +(3-\lambda)z_2 & = 0, \end{array}$$

for $\lambda = 2 + 3i$, and for $\lambda = 2 - 3i$;

c.

$$\begin{array}{rrrl} \lambda z_1 & -z_2 & & = 0, \\ & \lambda z_2 & +z_3 & = 0, \\ z_1 & & +\lambda z_3 & = 0, \end{array}$$

for $\lambda = 1$, $\lambda = e^{2\pi i/3}$, and for $\lambda = e^{-2\pi i/3}$.

12 Let $a = -\frac{1}{2} + \frac{1}{2}i\sqrt{3}$. Show that $a^2 + a + 1 = 0$ and solve the following system of linear equations.

$$\begin{array}{rrrl} z_1 & -z_2 & +z_3 & = 0, \\ z_1 & +az_2 & +a^2 z_3 & = 1, \\ -z_1 & -a^2 z_2 & -az_3 & = 1. \end{array}$$

13 Determine for each value of λ the solution(s) of the following system of linear equations.

$$\begin{array}{rrrl} \lambda x_1 & +x_2 & +x_3 & = 2, \\ x_1 & +\lambda x_2 & +x_3 & = 3. \end{array}$$

3.5.1 Exercises from old exams

14 Solve for each value of λ the following system of linear equations.

$$\begin{array}{rrrcl} x_1 & & -2x_3 & = & \lambda+4, \\ -2x_1 & +\lambda x_2 & +7x_3 & = & -14, \\ -x_1 & +\lambda x_2 & +6x_3 & = & \lambda-12. \end{array}$$

15 Solve for each value of λ the following system of linear equations.

$$\begin{array}{rrrrcl} x_1 & +2x_2 & +3x_3 & -6x_4 & = & -1, \\ 2x_1 & +x_2 & & +9x_4 & = & 2+\lambda, \\ -3x_1 & -3x_2 & -3x_3 & -3x_4 & = & 1+\lambda. \end{array}$$

Chapter 4

Vector spaces

4.1 Vector spaces and linear subspaces

4.1.1 When mathematicians notice mathematical objects with similar properties in various places of mathematics, they try to develop a common generalization. The concept of a vector space is such a notion. Vector spaces play an important role in many branches of mathematics (and other sciences). In this section we introduce

- the arithmetical rules ('axiom's') of a vector space,
- linear subspaces,
- (parametric representations of) lines and planes in vector spaces.

4.1.2 The notion of a vector space
The notion of a vector is usually associated with an arrow in the plane or space (starting in the origin) as we discussed in Chapter 2. In that setting we noticed that addition and scalar multiplication satisfy various properties. Here are eight important ones. For all vectors $\underline{p}$, $\underline{q}$, $\underline{r}$ and all scalars (numbers) λ, μ we have:

1. $\underline{p}+\underline{q}=\underline{q}+\underline{p}$,
2. $(\underline{p}+\underline{q})+\underline{r}=\underline{p}+(\underline{q}+\underline{r})$,
3. there is a *zero vector* $\underline{0}$ with the property $\underline{p}+\underline{0}=\underline{p}$ for every $\underline{p}$
4. every vector $\underline{p}$ has an *opposite* $-\underline{p}$ such that $\underline{p}+-\underline{p}=\underline{0}$ (we also write $\underline{p}-\underline{p}=\underline{0}$)
5. $1\,\underline{p}=\underline{p}$,

6. $(\lambda\mu)\underline{p} = \lambda(\mu\underline{p})$,
7. $(\lambda + \mu)\underline{p} = \lambda\underline{p} + \mu\underline{p}$,
8. $\lambda(\underline{p} + \underline{q}) = \lambda\underline{p} + \lambda\underline{q}$.

Now matrix addition and scalar multiplication of, say, $m \times n$ matrices satisfy similar properties. The similarities observed in the setting of vectors in the plane, of matrices, and of other examples, have led to the idea of introducing an abstract notion of which vectors in the plane or space, and matrices are examples. This is the notion of a *vector space* in which the starting point is any set together with two operations on the elements of this set, called 'addition' and 'scalar multiplication', in which the above eight 'axioms' hold. The elements of the set are then called *vectors*. A vector space is also sometimes called a *linear space*. In these lecture notes we denote vectors by underlined symbols[1], like $\underline{v}$. The scalars can be real or complex numbers. In the first case we are dealing with a *real vector space*, in the second case with a *complex vector space*. There do exist vector spaces over other sets of scalars but they are beyond the scope of this course.

From the eight rules described above we can derive some more (obvious looking) arithmetical rules that hold for vectors in an abstract setting (note that in the abstract setting we only know so far that our set satisfies the eight axioms; any other rule, even if it looks trivial, requires a proof). For instance, for every scalar λ the equality $\lambda\,\underline{0} = \underline{0}$ holds, and for every vector $\underline{a}$ we have $0\,\underline{a} = \underline{0}$ (see exercise 27).

Some more rules (that we will not discuss and proof in detail here; but see exercise 27) and remarks:

- The zero vector $\underline{0}$ is unique (in a given vector space), the opposite of a vector is unique.

- Stricly speaking, a sum of, say, three vectors $\underline{v}_1$, $\underline{v}_2$, $\underline{v}_3$ (or more) is not defined; only the sum of two vectors is. To deal with three vectors, just take $(\underline{v}_1 + \underline{v}_2) + \underline{v}_3$ (why is this sum defined?). Another option is to define the sum as $\underline{v}_1 + (\underline{v}_2 + \underline{v}_3)$, and the associativity garantuees that the two given options give the same answer. This is the reason that we usually just write $\underline{v}_1 + \underline{v}_2 + \underline{v}_3$ and only care about brackets if they are of help in a computation or proof. For more than three vectors something similar can be shown, so that a sum of n vectors $\underline{v}_1 + \cdots + \underline{v}_n$ is meaningful. For instance, a way of defining the sum of four vectors $\underline{v}_1, \ldots, \underline{v}_4$ is as follows: $(\underline{v}_1 + \underline{v}_2) + (\underline{v}_3 + \underline{v}_4)$. But, $((\underline{v}_1 + \underline{v}_2) + \underline{v}_3) + \underline{v}_4$ could also be the definition, and, again by an

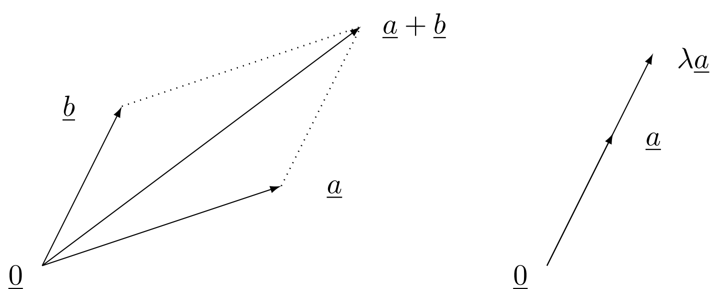
[1] In the literature you'll come across various other notations: $\vec{v}$, $\bar{v}$, $\mathbf{v}$

associativity argument (do you see how?), the two 'definitions' produce the same vector.

Finally, even though vectors in the plane or in space are just two examples of vector spaces, they are important in shaping our intuition. These examples are often a good guide, even when working in a totally different vector space.

4.1.3 Example. The first example is the 'space of arrows' in the plane or in space. We fix a point O, the *origin*. For every point P let $\underline{p}$ be the arrow from O to P. Our vector space to be consists of all such arrows; we denote it by E^2 (the plane) or E^3 (space).

The operations 'addition' and 'scalar multiplication' are defined as suggested in the figure. Using geometry the eight axioms of a vector space can be checked, but we will not discuss the details of this verification. The vector spaces E^2 and E^3 are examples of *real vector spaces*.



4.1.4 Example. Let $n \geq 1$ be an integer and let $\mathbb{R}^n = \{(a_1, \ldots, a_n) \mid a_1, \ldots, a_n \in \mathbb{R}\}$. For any two n-tuples of real numbers $\underline{a} = (a_1, \ldots, a_n)$ and $\underline{b} = (b_1, \ldots, b_n)$ from $\mathbb{R}^n$, and any scalar α we define the sum and the scalar product as follows:

$$\begin{aligned} \underline{a} + \underline{b} &= (a_1 + b_1, \ldots, a_n + b_n), \\ \alpha\, \underline{a} &= (\alpha a_1, \ldots, \alpha a_n) \quad (\alpha \text{ real}). \end{aligned}$$

One can easily verify that $\mathbb{R}^n$ with these two operations the axioms of a (real) vector space are satisfied. By way of example, We'll check the first one. The first axiom requires that $\underline{a} + \underline{b} = \underline{b} + \underline{a}$ for all $\underline{a}$ and $\underline{b}$. Now

$$\underline{a} + \underline{b} = (a_1 + b_1, \ldots, a_n + b_n)$$

and

$$\underline{b} + \underline{a} = (b_1 + a_1, \ldots, b_n + a_n),$$

where $\underline{a} = (a_1, \ldots, a_n)$ and $\underline{b} = (b_1, \ldots, b_n)$. Since $a_i + b_i = b_i + a_i$, $i = 1, \ldots, n$ (this is a property of the real numbers), we conclude that indeed $\underline{a} + \underline{b} = \underline{b} + \underline{a}$.

Note that the zero vector is $(0, 0, \ldots, 0)$, and the opposite of the vector $(a_1, \ldots, a_n)$ is $(-a_1, \ldots, -a_n)$.

We remark that, strictly speaking, E^2 is not the same space as $\mathbb{R}^2$ and E^3 is *not* $\mathbb{R}^3$, since an arrow is not an array of numbers. There is a close connection, and we will come back to that.

In a similar way we can turn the set $\mathbb{C}^n$ of n-tuples of complex numbers into a complex vector space.

4.1.5 Example. The set $M_{n,m}$ of $n \times m$-matrices with matrix addition and the usual scalar multiplication is a vector space with zero vector the $n \times m$ zero matrix. The opposite of a matrix A is the matrix $-A$. Depending on which numbers we use in the matrix and as scalars, we obtain a real or complex vector space. Sometimes the notations $M_{n,m}(\mathbb{R})$ and $M_{n,m}(\mathbb{C})$ are used to denote these two types.

4.1.6 Example. Let p and q be two polynomials of degree at most n:

$$\begin{aligned} p &= a_n x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0, \\ q &= b_n x^n + b_{n-1}x^{n-1} + \cdots + b_1 x + b_0. \end{aligned}$$

Now define the sum and the scalar product as follows:

$$\begin{aligned} p + q &= (a_n + b_n)x^n + (a_{n-1} + b_{n-1})x^{n-1} + \cdots + (a_1 + b_1)x + (a_0 + b_0), \\ \lambda p &= \lambda a_n x^n + \lambda a_{n-1}x^{n-1} + \cdots + \lambda a_1 x + \lambda a_0. \end{aligned}$$

This addition and scalar multiplication satisfy the eight axioms. The zero vector is the zero polynomial (all coefficients equal to 0), and the opposite of $a_n x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0$ is of course the polynomial $-a_n x^n - a_{n-1}x^{n-1} - \cdots - a_1 x - a_0$. If we only allow polynomials with real coefficients and if we use real scalars in the scalar multiplication, then the vector space is real. If we admit complex coefficients and complex scalars, the vector space is complex.

4.1.7 Example. Consider the set of all functions from a non-empty set X to the real numbers. Addition and scalar multiplication can be defined as follows:

$$\begin{aligned} (f + g)(x) &= f(x) + g(x), \\ (\alpha f)(x) &= \alpha\, f(x). \end{aligned}$$

Then X becomes a real vector space (the zero vector is the 'zero function' which sends every $x \in X$ to 0; the opposite $-f$ of a function f is the function $(-1)f$).

Of course, in a similar way a complex vector space can be constructed.

4.1.8 Next, we discuss subsets of a vector space V that are themselves vector spaces (with the two operations 'inherited' from V). A typical example is the subset $\{(x, y) \in \mathbb{R}^2 \mid y = 0\}$ of the vector space $\mathbb{R}^2$. It is easy to verify that this subset with the addition $(u, 0)+(x, 0) = (u+x, 0)$ and the scalar multiplication $\lambda(u, 0) = (\lambda u, 0)$ (simply add and multiply them as vectors in $\mathbb{R}^2$) is itself a vector space (with zero vector $(0, 0)$ and opposite $(-u, 0)$ of $(u, 0)$).

Suppose W is a non-empty subset of the vector space V. For any two vectors in V that actually lie in W, there is a sum vector in V because we know how to add vectors in V. But there is no garantuee that this sum vector is itself in W. A similar remark holds for scalar multiples of vectors from W: such multiples lie in V but not necessarily in W. If such sums and scalar multiples always lie in W, then W turns out to be a vector space itself. We call such a subset a *linear subspace of* V.

4.1.9 Definition. (Linear subspace) A *non-empty* subset W of a vector space V is called a *linear subspace* of V if for all $\underline{p}, \underline{q} \in W$ and for all scalars λ we have

$$\begin{aligned} \underline{p} + \underline{q} &\in W, \\ \lambda \underline{p} &\in W. \end{aligned}$$

Equivalently: for all $\underline{p}, \underline{q} \in W$ and for all scalars λ, μ

$$\lambda \underline{p} + \mu \underline{q} \in W.$$

To verify that such a W is indeed a vector space itself, we need to check the eight axioms. This turns out to be easy. For instance, to check the first axiom, we need to verify that $\underline{v} + \underline{w} = \underline{w} + \underline{v}$ for every $\underline{v}$, $\underline{w} \in W$. But we already know that $\underline{v}+\underline{w} = \underline{w}+\underline{v}$ for every $\underline{v}$, $\underline{w} \in V$, and so the equality certainly holds if $\underline{v}$, $\underline{w}$ belong to a subset of V! Most axioms hold for similar reasons.

As for the zero vector: V's zero vector turns out to lie in W. To see this, take any $\underline{p}$ in W (here we use the fact that W is non-empty!) and take the scalar 0. Then $0 \cdot \underline{p} = \underline{0}$ is in W by the above requirements for a linear subspace.

By using the equality $-\underline{w} = (-1)\underline{w}$ one easily shows in a similar way that the opposite of $\underline{w} \in W$ is itself in W.

So linear subspaces are vector spaces themselves. Conversely, if a subset of a vector space V is a vector space itself (with the addition and scalar multiplication from V), then the subset obviously satisfies the above conditions for a linear subspace.

Caution: note that subspaces are required to be non-empty.

4.1.10 Here is a useful observation that sometimes helps in deciding that a subset is *not* a linear subspace.

If W is a linear subspace of the vector space V, then W contains the zero vector $\underline{0}$ of V as we just saw. An equivalent formulation is: if the subset W of the vector space V does *not* contain V's zero vector, then W is not a linear subspace.

For instance, the subset W of $\mathbb{R}^2$ defined by $W = \{(x, y) \in \mathbb{R}^2 \mid y = 1\}$ is not a linear subspace since it doesn't contain $(0, 0)$.

4.1.11 Example. The subset $V : 3x - 2y + z = 6$ in $\mathbb{R}^3$ is not a linear subspace of $\mathbb{R}^3$, since the zero vector of $\mathbb{R}^3$ is not in V.

Here is a different way to arrive at the same conclusion. Take two cleverly chosen vectors in V, say $\underline{a} = (2, 0, 0)$ and $\underline{b} = (0, -3, 0)$. Then $\underline{a} \in V$ and $\underline{b} \in V$ but $\underline{a} + \underline{b} = (2, -3, 0) \notin V$, as is easily verified by substituting the coordinates in the equation. So V (with the addition and scalar multiplication from $\mathbb{R}^3$) is not closed with respect to the addition.

The set (plane) U with equation $3x - 2y + z = 0$ is a linear subspace of $\mathbb{R}^3$. To verify this we first note that U is nonempty because $(0, 0, 0) \in U$. Next we turn to the condition that the sum of any two vectors from U should be in U. Sp let $\underline{a} = (a_1, a_2, a_3) \in U$ and $\underline{b} = (b_1, b_2, b_3) \in U$. Then

$$\begin{aligned} 3a_1 - 2a_2 + a_3 &= 0, \\ 3b_1 - 2b_2 + b_3 &= 0. \end{aligned}$$

Adding yields:

$$3(a_1 + b_1) - 2(a_2 + b_2) + (a_3 + b_3) = 0,$$

so $\underline{a} + \underline{b} = (a_1 + b_1, a_2 + b_2, a_3 + b_3) \in U$.

In a similar way one can verify that $\alpha\, \underline{a} \in U$ for every $\alpha \in \mathbb{R}$.

Note that in order to prove that a subset is a linear subspace it is not enough to show that $\underline{0}$ belongs to that subset. For instance, the subset $W = \{(x, y) | y = x^2\}$ of $\mathbb{R}^2$ contains $(0, 0)$, but W is not a linear subspace because $(1, 1)$ is in W but $2 \cdot (1, 1)$ is not.

4.1.12 Example. In the vector space V of all real polynomials of degree at most 3, the subset $W = \{\, p(x) \in V \mid p(1) = 0 \,\}$, i.e., the set of polynomials having a zero at 1, is a linear subspace. This subset contains for example the polynomial $p(x) = x^2 - 1$. Here is the proof that W is indeed a linear subspace.

- $W \neq \emptyset$ because the zero polynomial is in W;
- if p, $q \in W$ and $\lambda, \mu \in \mathbb{R}$, then $\lambda p + \mu q \in W$ because

$$(\lambda p + \mu q)(1) = \lambda \cdot p(1) + \mu \cdot q(1) = \lambda \cdot 0 + \mu \cdot 0 = 0.$$

4.1.13 Example. The solutions of a homogeneous system of linear equations

$$A\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix},$$

where A is an $m \times n$ matrix, form a linear subspace of the vector space $M_{n,1}$ of $n \times 1$ matrices (over $\mathbb{R}$ or $\mathbb{C}$): first of all, the solution set is non-empty because the $n \times 1$ zero matrix O is a solution; secondly, if X and Y are solutions and λ and μ are scalars, then (using the rules for matrix multiplication) $A(\lambda X + \mu Y) = \lambda AX + \mu AY = O + O = O$, so that $\lambda X + \mu Y$ is also a solution.

The solution set of an inhomogeneous system $AX = B$, where B is a nonzero $m \times 1$ matrix, is never a linear subspace because the $n \times 1$ zero matrix is not a solution.

In the next section we discuss so-called (linear) *spans*, an important class of linear subspaces. Here is another, more abstract example involving a chain of linear subspaces.

4.1.14 Example. Consider the vector space V of all functions on $\mathbb{R}$, with sum $f+g$ and scalar product αf defined by

$$\begin{aligned} (f+g)(x) &= f(x) + g(x) \quad \text{for all } x \in \mathbb{R}, \\ (\alpha f)(x) &= \alpha\, f(x) \quad \text{for all } \ x \in \mathbb{R}. \end{aligned}$$

Now polynomials (more precisely, polynomial functions) form a nonempty subset of V. The sum of two such functions and the scalar product of such a function are again polynomial functions. So the set P of all polynomials forms a linear subspace of V.

Here is a further refinement of this statement. The sum of two polynomials of degree at most n and the scalar product of a polynomial of degree at most n are again polynomials of degree at most n. So for every nonnegative integer n the set P_n of all polynomials of degree at most n is a linear subspace of V. So we have the following chain of linear subspaces:

$$P_0 \subset P_1 \subset P_2 \subset \cdots \subset P_n \subset \cdots \subset P \subset V.$$

Note that no two of these subspaces are equal.

Next, we define the notions *line* and *plane* in the general setting of vector spaces.

4.1.15 If $\underline{p}$ and $\underline{v} \neq \underline{0}$ are two vectors in E^3 (or E^2), then geometrically it is clear that the endpoints of the vectors

$$\underline{x} = \underline{p} + \lambda \underline{v}, \quad \lambda \in \mathbb{R} \tag{4.1}$$

are on the line through the endpoint of $\underline{p}$ and parallel with $\underline{v}$. The formula (4.1) is calles a *parametric equation* of this line. Since the expression $\underline{p} + \lambda\underline{v}$ is built from a scalar product of a vector and a sum of vectors, we can, by analogy, state the following definition in any vector space.

4.1.16 Definition. (Line) Let $\underline{p}$ and $\underline{v}$ be two vectors in a vector space en suppose $\underline{v} \neq \underline{0}$. Then the set of vectors of the form

$$\underline{x} = \underline{p} + \lambda\underline{v}, \quad \lambda \in \mathbb{R} \text{ or } \mathbb{C},$$

is called a *line* in the vector space. The vector $\underline{p}$ is called a *position vector* of the line and the vector $\underline{v}$ a *direction vector*. We call the description $\underline{x} = \underline{p} + \lambda\underline{v}$ a *parametric equation* or *parametric representation* of the line.

4.1.17 Example. All solutions of the differential equation

$$y' + 2y = 2x$$

are

$$y = (x - \frac{1}{2}) + ce^{-2x}.$$

The solution of this differential equation is therefore a line in the space of all functions on $\mathbb{R}$. Its position vector is the function $x - \frac{1}{2}$ and its direction vector is the function e^{-2x}.

4.1.18 Similarly, if $\underline{p}$, $\underline{v}$, $\underline{w}$ are three vectors in E^3 such that $\underline{v} \neq \underline{0}$, $\underline{w} \neq \underline{0}$, and such that $\underline{v}$ and $\underline{w}$ are not multiples of one another. (In the next section, we will formulate this as: $\underline{v}$ and $\underline{w}$ are linearly independent. Geometrically it is clear that the endpoints of the vectors

$$\underline{x} = \underline{p} + \lambda\underline{v} + \mu\underline{w}, \quad \lambda, \mu \in \mathbb{R}$$

describe a plane passing through the endpoint of $\underline{p}$ and parallel to $\underline{v}$ and $\underline{w}$. This motivates the following generalization.

4.1.19 Definition. (Plane) Let $\underline{p}, \underline{v}, \underline{w}$ be three vectors in a vector space and suppose $\underline{v} \neq \underline{0}$, $\underline{w} \neq \underline{0}$, and $\underline{v}$ and $\underline{w}$ are not multiples of one another. The set of vectors

$$\underline{x} = \underline{p} + \lambda\underline{v} + \mu\underline{w}, \quad \lambda, \mu \in \mathbb{R} \quad (\text{of } \mathbb{C}) \tag{4.2}$$

is calles a *plane* in the vector space with *position vector* $\underline{p}$ and *direction vectors* $\underline{v}$ and $\underline{w}$. The description (4.2) is called a *parametric equation* (or parametric representation) of the plane.

4.1.20 **Example.** The real solutions of the differential equation

$$y'' + y = x$$

are

$$y = x + c_1 \cos x + c_2 \sin x \quad \text{with } c_1,\, c_2 \in \mathbb{R}.$$

So the solution set is a plane in the vector space of all functions on $\mathbb{R}$, with position vector the function x, and with direction vectors the functions $\sin x$ and $\cos x$.

4.1.21 In $\mathbb{R}^3$, consider the set

$$V = \{(x, y, z) \mid 2x + 3y - z = 4\}.$$

Take $x = \lambda$, $y = \mu$, then $z = 2\lambda + 3\mu - 4$ so that

$$V:\ \underline{x} = (0, 0, -4) + \lambda(1, 0, 2) + \mu(0, 1, 3).$$

This is a parametric description of the plane V in $\mathbb{R}^3$ where we have used the vector $(0, 0 - 4)$ as position vector and the vectors $(1, 0, 2)$ and $(0, 1, 3)$ as direction vectors. (If, for example, you take $x = \lambda$ and $z = \mu$ you will find a different parametric description, but which is just as good.)

The equation $2x + 3y - z = 4$ is an *equation* of the plane ($4x + 6y = 8 + 2z$ is another one). By solving the equation we have produced a *vector parametric equation* (or parametric equation(s)) of the plane.

Conversely, starting from a vector parametric equation of, say, a plane in $\mathbb{R}^3$, we can derive an equation of the plane. We use an example to demonstrate one of the ways to find such an equation. Consider the plane

$$W : \underline{x} = (1, 0, 1) + \lambda(1, 1, -1) + \mu(2, -1, -1),$$

i.e.,

$$\begin{array}{rcrrr} x & = & 1 & +\lambda & +2\mu, \\ y & = & & \lambda & -\mu, \\ z & = & 1 & -\lambda & -\mu. \end{array}$$

From the last two equations we solve for λ and μ and find

$$\lambda = \frac{1}{2}y - \frac{1}{2}z + \frac{1}{2} \quad \text{and} \quad \mu = -\frac{1}{2}y - \frac{1}{2}z + \frac{1}{2}.$$

Substituting in the first of the three equations yields

$$2x + y + 3z = 5,$$

an equation of the plane W.

A systematic way to find the answer is to rephrase the parametric equations as a system of linear equations in λ and μ with the following extended coefficient matrix:

$$\left(\begin{array}{cc|l} 1 & 2 & x-1 \\ 1 & -1 & y \\ -1 & -1 & z-1 \end{array}\right).$$

Row reducing produces:

$$\left(\begin{array}{cc|l} 1 & 0 & -x-2z+3 \\ 0 & 1 & x+z-2 \\ 0 & 0 & 2x+y+3z-5 \end{array}\right).$$

In the last row we see an equation appearing: $2x+y+3z-5=0$.

4.1.22 A linear equation in $\mathbb{R}^3$ describes a plane in $\mathbb{R}^3$. A linear equation in $\mathbb{R}^4$ is not a plane; if you turn the equation into a parametric description, three parameters appear instead of the expected two in the case of a plane. A plane in $\mathbb{R}^4$ can be described in terms of *two* linear equations. We illustrate this with an example. (The notion that is really behind all this is that of *dimension*, to which we will turn our attention later on in this chapter. Then all details should become fully clear.)

In any vector space, the plane through $\underline{p}, \underline{q}$ and $\underline{r}$ (assumed not to be on a single line) can be described with the following vector parametric equation:

$$\underline{x} = \underline{p} + \lambda(\underline{q} - \underline{p}) + \mu(\underline{r} - \underline{p}).$$

We find $\underline{p}$ for $\lambda = \mu = 0$, $\underline{q}$ for $\lambda = 1$, $\mu = 0$, and $\underline{r}$ for $\lambda = 0$, $\mu = 1$.

Next we turn to the plane W in $\mathbb{R}^4$ through $(0,1,0,1)$, $(-1,0,2,1)$, $(1,1,0,0)$. A vector parametric description of the plane is

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

or

$$\begin{array}{ccccl} x & = & & -\lambda & +\mu, \\ y & = & 1 & -\lambda & , \\ z & = & & 2\lambda & , \\ u & = & 1 & & -\mu. \end{array}$$

Using the last two equations we express λ and μ in terms of z and u and use the results in the first two equations. We find

$$\begin{array}{ccccccc} 2x & & +z & +2u & = & 2, \\ & 2y & +z & & = & 2. \end{array}$$

Every point of W is a solution of this system and conversely (for the converse, solve the system of two linear equations).

4.1.23 Every vector on the line

$$\ell : \underline{x} = \underline{p} + \lambda \underline{v}$$

can be used as position vector of the line. To show this, let's look at the line

$$m : \underline{x} = (\underline{p} + \alpha \underline{v}) + \lambda \underline{v},$$

where α is an arbitrary scalar (which we assume fixed in the discussion!)), and show that ℓ and m are actually the same line.

First, by rewriting $(\underline{p} + \alpha\underline{v}) + \lambda\underline{v}$ as $\underline{p} + (\alpha + \lambda)\underline{v}$ we see that every vector on m is also on ℓ. By rewriting $\underline{p} + \lambda\underline{v}$ as $(\underline{p} + \alpha\underline{v}) + (\lambda - \alpha)\underline{v}$ we conclude that every vector on ℓ is also on m. (Note how placing brackets and the arithmetic rules for vectors play a role in the proof.)

In terms of sets we have shown that

$$\{\underline{p} + \lambda\underline{v} \mid \lambda \in \mathbb{R}\} = \{(\underline{p} + \alpha\underline{v}) + \lambda\underline{v} \mid \lambda \in \mathbb{R}\}\,.$$

(For complex vector spaces we should of course replace $\mathbb{R}$ by $\mathbb{C}$.)

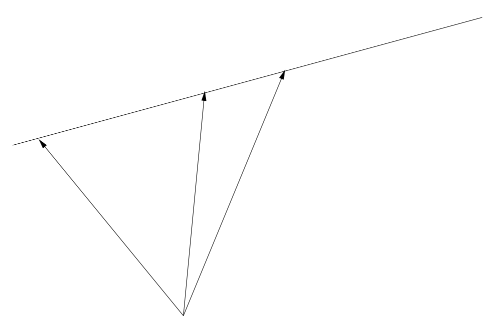


Figure 4.1: Every vector of a line may serve as position vector.

This remark implies that the line ℓ is a linear subspace if and only if $\underline{0} \in \ell$. Here are the details for the 'if' part. If $\underline{0} \in \ell$, then we can use $\underline{0}$ as a position vector of the line and describe the line by the scalar multiples $\lambda\,\underline{v}$ of $\underline{v}$. Since the sum $\lambda\underline{v} + \mu\underline{v}$ can be written as $(\lambda + \mu)\underline{v}$, this sum is again a multiple of $\underline{v}$ and therefore on ℓ. Of course, since $\mu(\lambda\underline{v}) = (\mu\lambda)\underline{v}$, we see that scalar multiples of vectors on l are themselves on ℓ.

Similar remarks hold for planes: every vector on a plane can serve as position vector of the plane, and a plane is a linear subspace if and only if the zero vector is on the plane.

In a similar way as above one can show that any nonzero multiple of $\underline{v}$ can serve as direction vector of the line $\ell : \underline{x} = \underline{p} + \lambda \underline{v}$. Planes can also have many pairs of direction vectors (no details here).

4.2 Spans, linearly (in)dependent systems

4.2.1 The concept of a vector space generalizes many particular situations, such as vectors in the plane, polynomials, matrices. What we gain by having this abstract notion is that whatever we prove there automatically holds in every concrete example of a vector space (the price we pay is that we have to get used to working with the abstract concept).

In this section we will illustrate this by developing the notion of *dimension* of a vector space ('the size of a vector space'). Along the way we will need various other notions and results. To develop these notions we usually let concrete examples be the source of inspiration.

In this section we concentrate on

- linear combinations of vectors,
- systems and spans of vectors,
- linearly dependent sets and linearly independent sets of vectors,
- bases and dimension.

4.2.2 **Definition. (Linear combination)** Let $\underline{a}_1, \ldots, \underline{a}_n$ be vectors in a vector space V. The vector $\underline{x}$ is called a *linear combination* of $\underline{a}_1, \ldots, \underline{a}_n$ if there exist scalars $\lambda_1, \ldots, \lambda_n$ such that

$$\underline{x} = \lambda_1 \underline{a}_1 + \lambda_2 \underline{a}_2 + \cdots + \lambda_n \underline{a}_n.$$

In this situation we also say that the vector $\underline{x}$ *depends* (or is linearly dependent) on the vectors $\underline{a}_1, \ldots, \underline{a}_n$.

4.2.3 **Example.** A linear combination of the vectors $(1, 1, -1)$ and $(2, 0, 1)$ in $\mathbb{R}^3$ is, for example, the vector $(-1, 3, -5) = 3\,(1, 1, -1) - 2\,(2, 0, 1)$.

4.2.4 **Definition. (Span of vectors)** Let $\underline{a}_1, \ldots, \underline{a}_n$ be vectors in a vector space. The set of all linear combinations of $\underline{a}_1, \ldots, \underline{a}_n$ is called the *span* of the vectors $\underline{a}_1, \ldots, \underline{a}_n$ and is denoted as $< \underline{a}_1, \ldots, \underline{a}_n >$.

4.2.5 **Theorem.** *Spans are linear subspaces, i.e., if $\underline{a}_1, \ldots, \underline{a}_n$ are vectors in the vector space V, then $< \underline{a}_1, \ldots, \underline{a}_n >$ is a linear subspace of V.*

Proof. Of course, the span is non-empty (it contains the zero vector).

Now let $\underline{p}$ and $\underline{q}$ be vectors in $< \underline{a}_1, \dots, \underline{a}_n >$ and suppose

$$\underline{p} = p_1\underline{a}_1 + \cdots + p_n\underline{a}_n \quad \text{and} \quad \underline{q} = q_1\underline{a}_1 + \cdots + q_n\underline{a}_n.$$

Then

$$\underline{p} + \underline{q} = (p_1 + q_1)\underline{a}_1 + \cdots + (p_n + q_n)\underline{a}_n \in < \underline{a}_1, \dots, \underline{a}_n > .$$

Also, for every scalar λ:

$$\lambda\underline{p} = \lambda p_1\underline{a}_1 + \cdots + \lambda p_n\underline{a}_n \in < \underline{a}_1, \dots, \underline{a}_n > .$$

So sums and scalar multiples of vectors from the span belong to the span, which finishes the proof. □

4.2.6 Example. The span $< (2,1,0), (1,0,1) >$ is precisely the plane with equation $x - 2y - z = 0$: with $y = \lambda$ and $z = \mu$ we get $(x, y, z) = (2\lambda + \mu, \lambda, \mu) = \lambda(2,1,0) + \mu(1,0,1)$, and, by definition, these vectors run through the span $< (2,1,0), (1,0,1) >$.

4.2.7 Example. In a vector space V consider a line passing through the origin:

$$l : \underline{x} = \lambda\underline{v}.$$

This line equals the span $< \underline{v} >$, so it is a linear subspace as we saw before in 4.1.23.

Similarly, the plane

$$V : \underline{x} = \lambda\underline{v} + \mu\underline{w}$$

passing through the origin equals the span $< \underline{v}, \underline{w} >$.

4.2.8 Example. In $\mathbb{R}^3$ consider the vectors $\underline{a} = (1, 1, -2)$, $\underline{b} = (-1, 1, 0)$, $\underline{c} = (0, 1, -1)$ and let $V = < \underline{a}, \underline{b}, \underline{c} >$. We see immediately that $2\underline{c} - \underline{a} = \underline{b}$. Now take an arbitrary $\underline{x} \in V$. Then $\underline{x}$ can be written as

$$\underline{x} = x_1\underline{a} + x_2\underline{b} + x_3\underline{c},$$

for some scalars x_1, x_2, x_3, so that

$$\begin{aligned} \underline{x} &= x_1\underline{a} + x_2(2\underline{c} - \underline{a}) + x_3\,\underline{c} \\ &= (x_1 - x_2)\underline{a} + (2x_2 + x_3)\underline{c}. \end{aligned}$$

So every vector from V is a linear combination of $\underline{a}$ and $\underline{c}$, and therefore V is contained in the span $< \underline{a}, \underline{c} >$.

Conversely, every linear combination of $\underline{a}$ and $\underline{c}$ is of course a linear combination of $\underline{a}, \underline{b}$, $\underline{c}$ (add $0\,\underline{b}$ to such a linear combination of $\underline{a}$ and $\underline{c}$), so that $< \underline{a}, \underline{c} >$ is contained in V. We conclude: $V =< \underline{a}, \underline{c} >$.

Verify yourself that $\underline{a} = 2\underline{c} - \underline{b}$ and hence $V =< \underline{b}, \underline{c} >$; also $\underline{c} = \frac{1}{2}\underline{a} + \frac{1}{2}\underline{b}$ from which we conclude that $V =< \underline{a}, \underline{b} >$.

Also check that V is the plane $x + y + z = 0$.

4.2.9 Manipulating spans

In the previous example we considered a span of a set of vectors that could also be spanned by a smaller set of vectors. Such a smaller set is likely to be of more practical value than the bigger one (as we will see later on). Our next goal is therefore to discuss operations with which we can reduce the number of vectors spanning a given linear subspace in a systematic way, and even find 'minimal spanning sets.' You will notice that these operations are very similar to row reduction operations.

4.2.10 Theorem. *Let $\underline{a}_1, \ldots, \underline{a}_n$ be vectors in a vector space V. The span $< \underline{a}_1, \ldots, \underline{a}_n >$ doesn't change if we*

1. *change the order of the vectors,*
2. *multiply one of the vectors by a scalar $\neq 0$, i.e., replace, say, $\underline{a}_i$ by $\lambda \underline{a}_i$ with $\lambda \neq 0$,*
3. *add a scalar multiple of one of the vectors to one of the other vectors, i.e., replace, say, $\underline{a}_i$ by $\underline{a}_i + \alpha \underline{a}_j$ with $j \neq i$.*

The span also doesn't change if we

4. *insert the zero vector, for instance, $< \underline{a}_1, \ldots, \underline{a}_n >=< \underline{a}_1, \ldots, \underline{a}_n, \underline{0} >$, or leave out the zero vector (if of course the zero vector was one of the $\underline{a}_i$),*
5. *insert a linear combination $\lambda_1 \underline{a}_1 + \cdots + \lambda_n \underline{a}_n$ of $\underline{a}_1, \ldots, \underline{a}_n$,*
6. *leave out $\underline{a}_i$ if this vector is a linear combination of the other $\underline{a}_j$.*

Proof. The proof that changing the order (1), and inserting or leaving out the zero vector (4) doesn't affect the span is almost trivial, so we leave that to the reader. To prove 2) we first observe that the equality

$$\lambda_1 \underline{a}_1 + \lambda_2 \underline{a}_2 + \cdots + \lambda_k (\alpha \underline{a}_k) + \cdots + \lambda_n \underline{a}_n = \lambda_1 \underline{a}_1 + \lambda_2 \underline{a}_2 + \cdots + (\lambda_k \alpha) \underline{a}_k + \cdots + \lambda_n \underline{a}_n$$

shows that every linear combination of $\underline{a}_1, \ldots, \alpha \underline{a}_k, \ldots, \underline{a}_n$ (only $\underline{a}_k$ is multiplied by the scalar α) is a linear combination of $\underline{a}_1, \ldots, \underline{a}_k, \ldots, \underline{a}_n$. Likewise,

$$\lambda_1 \underline{a}_1 + \lambda_2 \underline{a}_2 + \cdots + \lambda_k \underline{a}_k + \cdots + \lambda_n \underline{a}_n = \lambda_1 \underline{a}_1 + \lambda_2 \underline{a}_2 + \cdots + \frac{\lambda_k}{\alpha} (\alpha \underline{a}_k) + \cdots + \lambda_n \underline{a}_n$$

shows that, for $\alpha \neq 0$, every linear combination of $\underline{a}_1, \ldots, \underline{a}_k, \ldots, \underline{a}_n$ is a linear combination of $\underline{a}_1, \underline{a}_2, \ldots, \alpha\underline{a}_k, \ldots, \underline{a}_n$. This finishes the proof of 2).

We now show that the span doesn't change if we add a multiple of one of the vectors to one of the other vectors. By possibly changing the order of the vectors we may consider the situation where we add $\alpha\underline{a}_2$ to $\underline{a}_1$. Since every linear combination of $\underline{a}_1 + \alpha\underline{a}_2, \underline{a}_2, \ldots, \underline{a}_n$ is a linear combination of $\underline{a}_1, \ldots, \underline{a}_n$ we have

$$< \underline{a}_1 + \alpha\underline{a}_2, \underline{a}_2, \ldots, \underline{a}_n > \subset < \underline{a}_1, \ldots, \underline{a}_n > .$$

But then

$$< (\underline{a}_1 + \alpha\underline{a}_2) - \alpha\underline{a}_2, \underline{a}_2, \ldots, \underline{a}_n > \subset < \underline{a}_1 + \alpha\underline{a}_2, \underline{a}_2, \ldots, \underline{a}_n > .$$

Part 5) follows from the previous ones as follows: if $\underline{b} = \lambda_1\underline{a}_1 + \cdots + \lambda_n\underline{a}_n$, then

$$\begin{aligned}
< \underline{a}_1, \ldots, \underline{a}_n > \quad &= < \underline{a}_1, \ldots, \underline{a}_n, \underline{0} > \\
&= < \underline{a}_1, \ldots, \underline{a}_n, \underline{0} + \lambda_1\underline{a}_1 > \\
&\vdots \\
&= < \underline{a}_1, \ldots, \underline{a}_n, \lambda_1\underline{a}_1 + \cdots + \lambda_n\underline{a}_n > \\
&= < \underline{a}_1, \ldots, \underline{a}_n, \underline{b} > .
\end{aligned}$$

Here we have used properties 4) and 3).

We leave the proof of the last item of the theorem to the reader. □

4.2.11 Example. By repeatedly applying the above rules, we see that (regardless of the vector space we are working in)

$$\begin{aligned}
< \underline{a} + 2\underline{b}, \underline{a} - \underline{b}, \underline{a} + \underline{b} > \quad &= < \underline{a} + 2\underline{b}, \underline{a} - \underline{b}, \underline{a} + \underline{b} + (\underline{a} - \underline{b}) > \\
&= < \underline{a} + 2\underline{b}, \underline{a} - \underline{b}, 2\underline{a} > \\
&= < \underline{a} + 2\underline{b} - (1/2) \cdot 2\underline{a}, \underline{a} - \underline{b} - (1/2) \cdot 2\underline{a}, 2\underline{a} > \\
&= < 2\underline{b}, -\underline{b}, 2\underline{a} > = < \underline{b}, -\underline{b}, \underline{a} > \\
&= < \underline{b}, -\underline{b} + \underline{b}, \underline{a} > = < \underline{b}, \underline{0}, \underline{a} > \\
&= < \underline{a}, \underline{b} > .
\end{aligned}$$

In particular we observe that we only need two vectors. Matrices can be used to perform the above manipulations in a systematic way. Here are the details. As in the case of systems of linear equations we restrict to writing down the coefficients of $\underline{a}$ and $\underline{b}$ in the various vectors in the various stages of the rewriting process. Collect the coefficients of the three vectors we start with in rows and apply row reduction:

$$\text{row reduce} \quad \begin{pmatrix} 1 & 2 \\ 1 & -1 \\ 1 & 1 \end{pmatrix} \quad \text{to normal form} \quad \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix}.$$

We interpret this computation as $<\underline{a}+2\underline{b}, \underline{a}-\underline{b}, \underline{a}+\underline{b}>=<\underline{a}, \underline{b}>$. This matrix approach provides both a shorthand notation and a systematic way of finding the answer.

Here is yet another way of handling spans.

4.2.12 Theorem. (Exchange theorem) *If* $V=<\underline{a}_1,\ldots,\underline{a}_n>$ *and* $\underline{b}=\lambda_1\underline{a}_1+\ldots+\lambda_i\underline{a}_i+\ldots+\lambda_n\underline{a}_n$ *with* $\lambda_i\neq 0$, *then*

$$V=<\underline{a}_1,\ldots,\underline{a}_n>=<\underline{a}_1,\ldots\underline{a}_{i-1},\underline{b},\underline{a}_{i+1},\ldots,\underline{a}_n>.$$

4.2.13 This theorem states that, if a vector $\underline{b}\in<\underline{a}_1,\ldots,\underline{a}_n>$ can be written as a linear combination of $\underline{a}_1,\ldots,\underline{a}_n$, where the coefficient of $\underline{a}_i$ is nonzero, then we can replace the vector $\underline{a}_i$ by $\underline{b}$ without altering the span of the vectors.

Proof of 4.2.12. Consider $V=<\underline{a}_1,\ldots,\underline{a}_i,\ldots,\underline{a}_n>$. Now first multiply $\underline{a}_i$ by λ_i ($\neq 0$) and then add to i^{e} the vector $\lambda_1\underline{a}_1,\ldots,\lambda_{i-1}\underline{a}_{i-1}$, $\lambda_{i+1}\underline{a}_{i+1},\ldots,\lambda_n\underline{a}_n$. These steps leave the span the samen, so that $V=<\underline{a}_1,\ldots,\underline{a}_{i-1},\underline{b},\underline{a}_{i+1},\ldots,\underline{a}_n>$. $\square$

4.2.14 In 4.2.11 we have seen an example of a space spanned by three vectors, but which can also be spanned by two vectors. We now discuss how to find such 'minimal' systems of vectors spanning a given space. Apart from the theorems 4.2.10 and 4.2.12, the notion of a linear (in)dependent system of vectors plays a central role.

4.2.15 Definition. (Linearly (in)dependent set of vectors) A set or system of vectors $\underline{a}_1,\ldots,\underline{a}_n$ is called *linearly dependent* if at least one of the vectors is a linear combination of the others. The vectors are called *linearly independent* if none of the vectors is a linear combination of the others. We often say in such situations that the vectors $\underline{a}_1,\ldots,\underline{a}_n$ are linearly (in)dependent, or that the set $\{\underline{a}_1,\ldots,\underline{a}_n\}$ is linearly independent.

4.2.16 A more practical way to decide if a set of vectors is linearly (in)dependent is based on the following equivalent formulation.

4.2.17 Definition. (Linearly (in)dependent vectors: practical version) The system of vectors $\underline{a}_1,\ldots,\underline{a}_n$ is linearly independent if and only if the only solution of the equation

$$\lambda_1\underline{a}_1+\lambda_2\underline{a}_2+\cdots+\lambda_n\underline{a}_n=\underline{0} \tag{4.3}$$

in $\lambda_1,\lambda_2,\ldots,\lambda_n$ is: $\lambda_1=0,\lambda_2=0,\ldots,\lambda_n=0$.

A *non–trivial relation* between the vectors $\underline{a}_1,\ldots,\underline{a}_n$ is an equality of the form

$$\lambda_1\underline{a}_1+\lambda_2\underline{a}_2+\cdots+\lambda_n\underline{a}_n=\underline{0},$$

where at least one of the coefficients $\lambda_1, \lambda_2, \ldots, \lambda_n$ is non-zero.

An equivalent formulation for a linearly dependent system of vectors is: the vectors $\underline{a}_1, \ldots, \underline{a}_n$ are linearly dependent if and only if there exists a non-trivial relation between the vectors $\underline{a}_1, \ldots, \underline{a}_n$.

Proof. We restrict ourselves to the proof of the first equivalence, and leave the second one to the reader.

First we deal with the implication $\Rightarrow$). If the equation (4.3) has a solution with, say, $\lambda_i \neq 0$, then

$$\underline{a}_i = \frac{-\lambda_1}{\lambda_i}\underline{a}_1 + \frac{-\lambda_{i-1}}{\lambda_i}\underline{a}_{i-1} + \frac{-\lambda_{i+1}}{\lambda_i}\underline{a}_{i+1} + \frac{-\lambda_n}{\lambda_i}\underline{a}_n,$$

so that $\underline{a}_i$ is a linear combination of the other vectors, contradicting the assumption. Conversely, suppose the equation (4.3) only has the zero solution. If, for example, $\underline{a}_i$ is a linear combination of the vectors $\underline{a}_1, \ldots, \underline{a}_{i-1}, \underline{a}_{i+1}, \ldots, \underline{a}_n$, say,

$$\underline{a}_i = \alpha_1\underline{a}_1 + \cdots + \alpha_{i-1}\underline{a}_{i-1} + \alpha_{i+1}\underline{a}_{i+1} + \cdots + \alpha_n\underline{a}_n,$$

then we obtain the following non-trivial relation

$$\alpha_1\underline{a}_1 + \cdots + \alpha_{i-1}\underline{a}_{i-1} - 1 \cdot \underline{a}_i + \alpha_{i+1}\underline{a}_{i+1} + \cdots + \alpha_n\underline{a}_n = \underline{0},$$

contradicting the assumption. □

4.2.18 Examples. The following examples illustrate definition 4.2.17.

- The vectors
$$\begin{aligned} \underline{e}_1 &= (1, 0, 0, \ldots, 0), \\ \underline{e}_2 &= (0, 1, 0, \ldots, 0), \\ &\vdots \\ \underline{e}_n &= (0, 0, 0, \ldots, 1), \end{aligned}$$
in $\mathbb{R}^n$ are linearly independent, since the equation
$$\lambda_1\underline{e}_1 + \cdots + \lambda_n\underline{e}_n = (0, \ldots, 0),$$
i.e., $(\lambda_1, \ldots, \lambda_n) = (0, \ldots, 0)$, has $\lambda_1 = \cdots = \lambda_n = 0$ as its only solution.

- The vectors $(1, 2, 2)$ and $(0, 1, -1)$ in $\mathbb{R}^3$ are linearly independent; here is the proof. If $a(1, 2, 2) + b(0, 1, -1) = (0, 0, 0)$, then we rewrite this as $(a, 2a + b, 2a - b) = (0, 0, 0)$, and easily conclude $a = b = 0$.

- The vectors $(1,0,1)$, $(0,1,1)$, $(1,1,0)$, $(2,2,2)$ in $\mathbb{R}^3$ are not linearly dependent. To see this, consider the equation

$$a(1,0,1)+b(0,1,1)+c(1,1,0)+d(2,2,2)=(0,0,0)$$

  in a,b,c,d. This is a system of linear equations in a,b,c,d, whose solutions are $(a,b,c,d)=\lambda(1,1,1,-1)$. So the vectors satisfy a non-trivial relation, e.g., for $\lambda=1$:

$$(1,0,1)+(0,1,1)+(1,1,0)-(2,2,2)=(0,0,0).$$

  So the vectors are not linearly independent.

- The functions sin and cos in the space of real functions $\mathbb{R}\to\mathbb{R}$ are linearly independent. Suppose

$$a\ \sin + b\ \cos = 0 \quad (= \text{de nulfunctie}),$$

  then, *since this is an equality of functions*, we find dat that for every real number t the relation $a\sin(t)+b\cos(t)=0$ holds. Now we choose a few 'smart' values for t to deduce that a and b are 0: for $t=0$ we get $b\cos(0)=0$ so that $b=0$, and for $t=\pi/2$ we get $a\sin(\pi/2)=0$ so that $a=0$.

  Of course, in general there may exist dependences between functions. For instance, the formula $\sin(2t)=2\sin(t)\cos(t)$ tells us that the functions $t\mapsto\sin(2t)$ en $t\mapsto\sin(t)\cos(t)$ are not linearly independent.

- In example 4.2.8 a non-trivial relation between the vectors $\underline{a}$, $\underline{b}$ and $\underline{c}$ exists, namely $\underline{a}+\underline{b}-2\underline{c}=\underline{0}$. In particular, these vectors are not linearly independent.

4.2.19 Finding an independent set from a spanning set of vectors

Every set $\{\underline{a}_1,\ldots,\underline{a}_n\}$ in a vector space V which does not consist of zero vectors only can be 'reduced' to a linearly independent set with the same span in the following way. Let $U=<\underline{a}_1,\ldots,\underline{a}_n>$. If $\underline{a}_1,\ldots\underline{a}_n$ is linearly independent, then we are done. If not, then at least one of the vectors is a linear combination of the others. This vector can be left out without changing the span. We are then left with a set of $n-1$ vectors that still spans U. If this set is linearly independent, we are ready. If not, we repeat the procedure we just described. After at most $n-1$ steps we find a linearly independent set spanning U.

Next we turn to a special property of linearly independent sets spanning a given vector space: if two (finite) linearly independent sets span the same space, they contain the same number of vectors. That number is what we call the *dimension* of the space. Here are the preparations for this result.

4.2.20 Theorem. *Suppose* $V = < \underline{a}_1, \ldots, \underline{a}_n >$ *and suppose* $\underline{b}_1, \ldots, \underline{b}_m$ *is a linearly independent set of vectors in* V*. Then* $m \leq n$*.*

Proof. The vector $\underline{b}_1$ is a linear combination of $\underline{a}_1, \ldots, \underline{a}_n$. Since $\underline{b}_1$ is not the zero vector, at least of the coefficients of the $\underline{a}_i$ is not zero. So, Theorem 4.2.12 enables us to exchange the vectors $\underline{b}_1$ and one of the $\underline{a}_i$. Without loss of generality we may assume that we exchange $\underline{b}_1$ and $\underline{a}_1$ (if necessary relabel the $\underline{a}_i$). So

$$V = < \underline{b}_1, \underline{a}_2, \ldots, \underline{a}_n > .$$

Now the vector $\underline{b}_2$ is a linear combination of the vectors on the right-hand side. Again, at least of the coefficients of the $\underline{a}_2, \ldots, \underline{a}_n$ must be $\neq 0$ (otherwise, $\underline{b}_1$ would be a multiple of $\underline{b}_1$). So we can exchange $\underline{b}_2$ and one the vectors $\underline{a}_2, \ldots, \underline{a}_n$, again by Theorem 4.2.12. Possibly after relabeling, we may assume that we exchange $\underline{b}_2$ and $\underline{a}_2$. So:

$$V = < \underline{b}_1, \underline{b}_2, \underline{a}_3, \ldots, \underline{a}_n > .$$

Continue in the same way. By Theorem 4.2.12 *every* $\underline{b}_i$ can be exchanged, so that $m \leq n$. □

4.2.21 Theorem. *If the vector space* V *is the span of each of the systems of independent vectors* $\underline{a}_1, \ldots, \underline{a}_n$ *and* $\underline{b}_1, \ldots, \underline{b}_m$*, then* $m = n$*.*

Proof. Apply the previous theorem to the system of independent vectors $\underline{b}_1, \ldots, \underline{b}_m$ in $< \underline{a}_1, \ldots, \underline{a}_n >$ to conclude that $m \leq n$. Then apply the theorem to the system of independent vectors $\underline{a}_1, \ldots, \underline{a}_n$ in $< \underline{b}_1, \ldots, \underline{b}_m >$ to conclude $n \leq m$. So $m = n$. □

4.2.22 Definition. (Basis and dimension) A linearly independent set spanning a vector space V is called a *basis* of V. The number of elements in the basis is called the *dimension* of V is denoted as $\dim(V)$.

4.2.23 If there isn't a finite basis of V (and V does not consist of $\underline{0}$ only), then we say $\dim(V) = \infty$.

The case $V = \{\underline{0}\}$ is a bit special. The space V contains only one vector, $\underline{0}$, but this vector is not linearly independent since $3\,\underline{0} = \underline{0}$ (do you see why?). We usually say that the emptyset $\emptyset$ is a basis and that the dimension of V is 0.

4.2.24 Examples. Here are some vector spaces and their dimensions.

- Geometrically it is clear that $\dim(E^2) = 2$ and $\dim(E^3) = 3$.

- In $\mathbb{R}^n$ the set containing the vectors

$$\begin{aligned} \underline{e}_1 &= (1,0,0,\ldots,0),\\ \underline{e}_2 &= (0,1,0,\ldots,0),\\ &\vdots\\ \underline{e}_n &= (0,0,0,\ldots,1), \end{aligned}$$

  is a linearly independent set spanning $\mathbb{R}^n$: we already established that these vectors are linearly independent, so we only need to check that every vector $(x_1, x_2, \ldots, x_n)$ can be written as a linear combination of these vectors, but that is clear from the expression $x_1\underline{e}_1 + \cdots + x_n\underline{e}_n$. So $\dim(\mathbb{R}^n) = n$. The basis $\underline{e}_1, \ldots, \underline{e}_n$ is usually called the *standard basis* of $\mathbb{R}^n$. Similarly, $\dim(\mathbb{C}^n) = n$. We also use the term standard basis in this case.

- Let P_n be the set of all (real or complex) polynomials in x of degree at most n. Then $P_n = < 1, x, x^2, \ldots, x^n >$. The polynomials $1, x, x^2, \ldots, x^n$ are linearly independent as we now prove. Suppose $\alpha_0 + \alpha_1 x + \alpha_2 x^2 + \cdots + \alpha_n x^n = 0$ for all x and suppose not all coefficients α_j are equal to 0. Then the polynomial on the left-hand side would have at most n zeros, while the zero polynomial on the right-hand side is identically 0. This contradiction shows that all α_j have to be 0. So $\{1, x, \ldots, x^n\}$ is a basis of P_n and $\dim(V) = n + 1$.

4.2.25 Here are some consequences of the definitions. If V is a vector space with $\dim(V) = n < \infty$, then *every* basis of V consists of exactly n vectors. We use this to prove the following statements about the m vectors $\underline{b}_1, \ldots, \underline{b}_m$ in V.

1. If $m < n$, then $< \underline{b}_1, \ldots, \underline{b}_m > \subset\subset V$ (the notation $\subset\subset$ indicates that the left-hand side is contained in but not equal to the right-hand side). Here is why. If $< \underline{b}_1, \ldots, \underline{b}_m >$ were equal to V, then using our technique to reduce the set $\{\underline{b}_1, \ldots, \underline{b}_m\}$ to a basis would produce a basis of V with at most m elements, contradicting the assumption on the dimension being n.

2. If $m > n$, then the set $\{\underline{b}_1, \ldots, \underline{b}_m\}$ is not linearly independent because of Theorem 4.2.20.

3. If $m = n$, then the vectors $\underline{b}_1, \ldots, \underline{b}_m$ are a basis of V if and only if the vectors are linearly independent. To prove this, first assume the set is not linearly independent. Then $< \underline{b}_1, \ldots, \underline{b}_m >$ can be spanned by less than m vectors and then item 1) implies $< \underline{b}_1, \ldots, \underline{b}_m > \subset\subset V$. So our set can't be a basis. Conversely, if $\underline{b}_1, \ldots, \underline{b}_n$ is linearly independent and its span is not equal to V, then there is a vector $\underline{a} \in V$ with $\underline{a} \notin < \underline{b}_1, \ldots, \underline{b}_n >$, so that $\{\underline{b}_1, \ldots, \underline{b}_n, \underline{a}\}$ is linearly independent (see Theorem 4.2.29 below) and

contains more than n vectors, contradicting the fact that $\dim(V) = n$. So $< \underline{b}_1, \ldots, \underline{b}_n >= V$ and $\underline{b}_1, \ldots, \underline{b}_n$ is a basis of V.

4.2.26 If V is a vector space with $\dim(V) = \infty$, then there is an infinite sequence of vectors $\underline{a}_1, \underline{a}_2, \ldots$ with

$$\underline{a}_{n+1} \notin < \underline{a}_1, \ldots, \underline{a}_n >$$

for every n. To see this, choose $\underline{a}_1 \neq \underline{0}$ in V. If $V = < \underline{a}_1 >$, then $\dim(V) = 1$. So there must be a $\underline{a}_2 \in V$ with $\underline{a}_2 \notin < \underline{a}_1 >$. If $V = < \underline{a}_1, \underline{a}_2 >$, then $\dim(V) = 2$, so $< \underline{a}_1, \underline{a}_2 > \subset\subset V$. Now choose $\underline{a}_3 \in V$, $\underline{a}_3 \notin < \underline{a}_1, \underline{a}_2 >$, etc.

The infinite sequence $\underline{a}_1, \underline{a}_2, \ldots$ that we find in this way has the desired property. Moreover, for every n the set $\{\underline{a}_1, \ldots, \underline{a}_n\}$ is linearly independent. This follows from Theorem 4.2.29 below. Here we see an important distinction between finite dimensional and infinite dimensional vector spaces: in an infinite dimensional vector space there exist arbitrarily large linearly independent sets, whereas in finite dimensional vector spaces the number of vectors in a linearly independent set is at most the dimension of the vector space.

4.2.27 Finding bases

Back to the problem of finding 'economical' sets spanning a vector space. The previous discussions show that this comes down to finding bases of vector spaces. Here are two obviuous ways of finding bases (sometimes ad hoc methods are quicker).

- As we saw before, if a vector space is given as a span of vectors, then by eliminating 'dependencies' we find a basis. In the case of vectors in $\mathbb{R}^n$ or $\mathbb{C}^n$, our row reduction operations from Chapter 3 are useful. We come back to this in Theorem 4.3.9 below. By using coordinates these techniques can also be used in other cases (see Theorem 4.3.7).

- If we do not have a set of vectors spanning a given vector space V, we make one ourselves in the following way. Start with a vector $\underline{a}_1 \neq \underline{0}$ in V (if possible, otherwise, we are already done). If $\langle \underline{a}_1 \rangle \neq V$, then choose a vector $\underline{a}_2 \notin < \underline{a}_1 >$. The vectors $\underline{a}_1, \underline{a}_2$ are linearly independent by Theorem 4.2.29 below. If $\langle \underline{a}_1, \underline{a}_2 \rangle \neq V$, then choose $\underline{a}_3 \notin < \underline{a}_1, \underline{a}_2 >$. The vectors $\underline{a}_1, \underline{a}_2, \underline{a}_3$ are linearly independent by Theorem 4.2.29. Etc. We illustrate this technique in the following example.

4.2.28 Example. We complete $\underline{a}_1 = (1, -1, 0)$ to a basis of $\mathbb{R}^3$. Choose a second vector outside $< (1, -1, 0) >$, e.g., $\underline{a}_2 = (0, 0, 1)$. Since $\dim(\mathbb{R}^3) = 3$ we look for a third vector outside the span $< (1, -1, 0), (0, 0, 1) >$, e.g., $\underline{a}_3 = (1, 1, 0)$ (of course we have to check that $\underline{a}_3$ is a valid choice). By Theorem 4.2.29 below we can now conclude that $\{\underline{a}_1, \underline{a}_2, \underline{a}_3\}$ is a basis (you may also prove it directly from the definition if you prefer).

4.2.29 Theorem. *If the set of vectors $\{\underline{a}_1, \ldots, \underline{a}_n\}$ in the vector space V satisfies*

$$\underline{a}_1 \neq \underline{0},\ \underline{a}_2 \notin < \underline{a}_1 >,\ \underline{a}_3 \notin < \underline{a}_1, \underline{a}_2 >,\ \ldots, \underline{a}_n \notin < \underline{a}_1, \ldots, \underline{a}_{n-1} >,$$

then the vectors are linearly independent.

Proof. To show linear independence, we consider the equation

$$\lambda_1 \underline{a}_1 + \lambda_2 \underline{a}_2 + \cdots + \lambda_n \underline{a}_n = \underline{0}$$

in $\lambda_1, \ldots, \lambda_n$. If $\lambda_n \neq 0$, then

$$\underline{a}_n = \frac{-\lambda_1}{\lambda_n} \underline{a}_1 + \cdots + \frac{-\lambda_{n-1}}{\lambda_n} \underline{a}_{n-1},$$

contradicting the fact that $\underline{a}_n \notin < \underline{a}_1, \ldots, \underline{a}_{n-1} >$. So $\lambda_n = 0$. In a similar way we derive that $\lambda_{n-1} = 0, \ldots, \lambda_2 = 0$. Finally, from $\lambda_1 \underline{a}_1 = \underline{0}$ and $\underline{a}_1 \neq \underline{0}$ we conclude that $\lambda_1 = 0$ (see exercise 27).

4.3 Coordinates

4.3.1 Coordinates
Bases are 'minimal' systems of vectors spanning a vector space. They have another special property which will enable us to use coordinates. If $\underline{a}_1, \ldots, \underline{a}_n$ span V, then every $\underline{x} \in V$ can be written in the form

$$\underline{x} = x_1 \underline{a}_1 + \cdots + x_n \underline{a}_n. \tag{4.4}$$

The coefficients need not be unique. For example, consider the space V from example 4.2.8; for the vector $\underline{b}$ we have

$$\begin{aligned} \underline{b} &= 0\underline{a} + 1\underline{b} + 0\underline{c} \\ &= -1\underline{a} + 0\underline{b} + 2\underline{c}. \end{aligned}$$

However, if $\underline{a}_1, \ldots, \underline{a}_n$ is a basis, then the coefficients in (4.4) *are* unique. Here is why. If $\underline{x} = y_1 \underline{a}_1 + \cdots + y_n \underline{a}_n$, then by subtracting we deduce

$$\underline{0} = (x_1 - y_1)\underline{a}_1 + \cdots + (x_n - y_n)\underline{a}_n,$$

so that $x_1 = y_1, \ldots, x_n = y_n$ because the system is linearly independent. We can therefore represent *every* vector $\underline{x}$ in V in a *unique* way with such coefficients. We define coordinates and coordinate vectors as follows.

4.3.2 Definition. (Coordinates) Let $\underline{a}_1, \ldots, \underline{a}_n$ be a basis of the vector space V. If

$$\underline{x} = x_1\underline{a}_1 + \cdots + x_n\underline{a}_n$$

then the coefficients $x_1, \ldots, x_n$ are called the *coordinates* of the vector $\underline{x}$ *with respect to this basis*. The vector $(x_1, \ldots, x_n)$ is called the *coordinate vector* of $\underline{x}$ and is itself a vector in $\mathbb{R}^n$ of $\mathbb{C}^n$.

Note: coordinates depend on the basis used!

4.3.3 Example. Let V be the vector space of polynomials of degree at most 2. Consider the polynomials

$$\begin{array}{ll} p_0: & p_0(x) = 1, \\ p_1: & p_1(x) = x, \\ p_2: & p_2(x) = x^2. \end{array}$$

Let p be an arbitrary polynomial in this space, say, $ax^2 + bx + c$. Then $p = ap_2 + bp_1 + cp_0$, so that

$$V =< p_0, p_1, p_2 > .$$

The polynomials p_0, p_1, p_2 are linearly independent: suppose $\alpha_0 p_0 + \alpha_1 p_1 + \alpha_2 p_2 = 0$ (the zero polynomial). This means

$$\alpha_0 + \alpha_1 x + \alpha_2 x^2 = 0 \quad \text{for } \textit{all } x.$$

If $(\alpha_0, \alpha_1, \alpha_2) \neq (0, 0, 0)$, then the left-hand side polynomial would have at most two zeros, which is not the case (since the polynomial is also equal to the zero polynomial). So $(\alpha_0, \alpha_1, \alpha_2) = (0, 0, 0)$ and p_0, p_1, p_2 are linearly independent. Therefore the polynomials p_0, p_1, p_2 form a basis of V and (c, b, a) is the coordinate vector of p with respect to this basis.

4.3.4 Example. The vectors $(1, 1)$ and $(1, -1)$ form a basis of $\mathbb{R}^2$. The linear independency is easily derived from the fact that $a(1, 1) + b(1, -1) = (0, 0)$ implies $a = b = 0$. From 4.2.25 we then derive that these two independent vectors are a basis of the space.

The coordinate vector of $(5, 3)$ with respect to this basis can be found by looking for a c and d such that $c(1, 1) + d(1, -1) = (5, 3)$. Solving leads to $c = 4$ and $d = 1$. The coordinate vector of $(5, 3)$ w.r.t. the new basis is then $(4, 1)$.

4.3.5 Coordinates of sums and scalar multiples

Let $\underline{a}_1, \ldots, \underline{a}_n$ be a basis of the vector space V. If $\underline{x} \in V$ has coordinate vector $(x_1, \ldots, x_n)$ w.r.t. this basis and $\underline{y} \in V$ has coordinate vector $(y_1, \ldots, y_n)$, then we easily verify that the coordinate vectors of $\underline{x} + \underline{y}$ and $\alpha\underline{x}$ are $(x_1 + y_1, \ldots, x_n + y_n)$

and $(\alpha x_1, \ldots, \alpha x_n)$, respectively. E.g., for the scalar multiple $\alpha \underline{x}$ this follows from the fact that $\alpha(x_1\underline{a}_1 + \cdots + x_n\underline{a}_n) = (\alpha x_1)\underline{a}_1 + \cdots (\alpha x_n)\underline{a}_n$.

So addition and scalar multiplication in V correspond nicely to the usual addition and scalar multiplication in $\mathbb{R}^n$ (or $\mathbb{C}^n$).

4.3.6 By coordinatising questions on vector spaces can be reduced to questions on $\mathbb{R}^n$ or $\mathbb{C}^n$, where we have our machinery of techniques at our disposal. (Of course, finally we translate the answers found in the coordinate world back to the original setting.)

The vector spaces $\mathbb{R}^n$ and $\mathbb{C}^n$ model every real and complex vector space of dimension n. Mathematicians say that every real n-dimensional vector space is *isomorphic* to $\mathbb{R}^n$, and every complex n-dimensional vector space is *isomorphic* to $\mathbb{C}^n$.

Often we say that $\mathbb{R}^2$ is the plane, but that is, strictly speaking, wrong. Upon choosing a basis $(\underline{a}_1, \underline{a}_2)$ in E^2 every vector $\underline{x}$ can be written in the form $\underline{x} = x_1\underline{a}_1 + x_2\underline{a}_2$ and computations with vectors can be 'translated' into computations with elements in $\mathbb{R}^2$. The choice of basis is of importance here. Since we are so used to the use of coordinate vectors in the plane, it is tempting to say that these two space are the same.

The following theorem shows that linear independency can be checked at the level of coordinates.

4.3.7 Theorem. *Let α be a basis of the n–dimensional vector space V and let $\{\underline{a}_1, \ldots, \underline{a}_m\}$ be a set of vectors in V. Then:*

- *$\{\underline{a}_1, \ldots, \underline{a}_m\}$ is linearly indenpendent if and only if the corresponding set of coordinate vectors (in $\mathbb{R}^n$ or $\mathbb{C}^n$) w.r.t. α is linearly independent.*
- *$\{\underline{a}_1, \ldots, \underline{a}_m\}$ is a basis if and only if the corresponding set of coordinate vectors (in $\mathbb{R}^n$ or $\mathbb{C}^n$) w.r.t. α is a basis.*

Proof. We only prove the first item, since the second item is a direct consequence of it. Suppose $\underline{b}_1, \ldots, \underline{b}_m$ are the coordinate vectors of $\underline{a}_1, \ldots, \underline{a}_m$. The coordinate vector of $\lambda_1\underline{a}_1 + \cdots + \lambda_m\underline{a}_m$ is equal to $\lambda_1\underline{b}_1 + \cdots + \lambda_m\underline{b}_m$. So if $\lambda_1\underline{a}_1 + \cdots + \lambda_m\underline{a}_m = \underline{0}$, then we also have $\lambda_1\underline{b}_1 + \cdots + \lambda_m\underline{b}_m = (0, \ldots, 0)$ and conversely, since the coordinate vector of the zero vector is $(0, \ldots, 0)$. A non-trivial relation between $\underline{a}_1, \ldots, \underline{a}_m$ translates into a non-trivial relation between the coordinate vectors $\underline{b}_1, \ldots, \underline{b}_m$. In other words: $\underline{a}_1, \ldots, \underline{a}_m$ is linearly dependent if and only if $\underline{b}_1, \ldots, \underline{b}_m$ is linearly dependent.

4.3.8 Relation with systems of linear equations

The above considerations also lead to a new view on systems of linear equations.

Let

$$\begin{array}{rcl} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1m}x_m &=& b_1 \\ \vdots && \vdots \\ a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nm}x_m &=& b_n \end{array}$$

be such a system. Let $\underline{k}_1, \ldots, \underline{k}_m$ be the columns of the coefficient matrix and let $\underline{b} = (b_1, \ldots, b_n)^T$. Then we can write the system as

$$x_1\underline{k}_1 + x_2\underline{k}_2 + \cdots + x_m\underline{k}_m = \underline{b}.$$

So we try to write $\underline{b}$ as a linear combination of the columns of the matrix. The system has at least one solution if and only if $\underline{b}$ is contained in the span of the columns, and precisely one solution if and only if the columns $\underline{k}_1, \ldots, \underline{k}_m$ are moreover linearly independent. In that case, the solution produces the coordinates of the vector $\underline{b}$ w.r.t. the basis $\{\underline{k}_1, \ldots, \underline{k}_m\}$.

4.3.9 Theorem. *The nonzero rows of a matrix in row reduced echelon form are linearly independent.*

Proof. A matrix in row reduced form has the following shape (see 3.2.3):

$$\begin{bmatrix}
0 & \ldots & 0 & 1 & * & \ldots & * & 0 & * & \ldots & * & 0 & * & \ldots & 0 & * & \ldots & * \\
0 & \ldots & 0 & 0 & 0 & \ldots & 0 & 1 & * & \ldots & * & 0 & * & \ldots & 0 & * & \ldots & * \\
0 & \ldots & 0 & 0 & 0 & \ldots & 0 & 0 & 0 & \ldots & 0 & 1 & * & \ldots & 0 & * & \ldots & * \\
\ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\
\ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & 0 & * & \ldots & * \\
0 & \ldots & 0 & 0 & 0 & \ldots & 0 & 0 & 0 & \ldots & 0 & 0 & 0 & \ldots & 1 & * & \ldots & * \\
0 & \ldots & 0 & 0 & 0 & \ldots & 0 & 0 & 0 & \ldots & 0 & 0 & 0 & \ldots & 0 & 0 & \ldots & 0 \\
\vdots & & \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots & \vdots & & \vdots \\
0 & \ldots & 0 & 0 & 0 & \ldots & 0 & 0 & 0 & \ldots & 0 & 0 & 0 & \ldots & 0 & 0 & \ldots & 0
\end{bmatrix}$$

Suppose the first p rows $\underline{r}_1, \ldots, \underline{r}_p$ are $\neq \underline{0}$ and suppose

$$\alpha_1\underline{r}_1 + \alpha_2\underline{r}_2 + \cdots + \alpha_p\underline{r}_p = \underline{0}.$$

Now consider the columns containing only zeros except for one 1. Then we find, respectively,

$$\alpha_1 = 0,\ \alpha_2 = 0,\ \ldots, \alpha_p = 0.$$

□

4.3.10 The considerations above fully explain the techniques announced in 4.2.27. From Theorems 4.2.10 and 4.3.9 we find how to construct a basis for a given span in $\mathbb{R}^n$

or $\mathbb{C}^n$: Consider the spanning vectors as rows of a matrix, take the row reduced echelon form of this matrix and take the nonzero rows. In an 'abstract' vector space we can use these techniques if use coordinates.

4.3.11 Example. Consider the following vectors in $\mathbb{R}^4$:

$$\underline{a} = (1,0,2,0), \quad \underline{b} = (1,1,2,1), \quad \underline{c} = (2,-1,4,-1) \quad \text{en} \quad \underline{d} = (1,3,2,3).$$

Suppose $V =< \underline{a}, \underline{b}, \underline{c}, \underline{d} >$. Now form the matrix

$$\begin{pmatrix} 1 & 0 & 2 & 0 \\ 1 & 1 & 2 & 1 \\ 2 & -1 & 4 & -1 \\ 1 & 3 & 2 & 3 \end{pmatrix}.$$

Row reducing doesn't change the span of the rows. The row reduced form is:

$$\begin{pmatrix} 1 & 0 & 2 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

The nonzero rows $(1,0,2,0)$ en $(0,1,0,1)$ produce a basis of V.

4.3.12 Example. Consider the following polynomials in the vector space of polynomials of degree at most 2

$$\begin{array}{lll} p_1: & p_1(x) & = x^2 + 2x - 3, \\ p_2: & p_2(x) & = x^2 - 2x + 1, \\ p_3: & p_3(x) & = x^2 - 5x + 4. \end{array}$$

and consider their span $V = < p_1, p_2, p_3 >$. Choose as basis: $(1, x, x^2)$. With respect to this basis, the coordinate vectors of the three polynomials are

$$\begin{array}{ll} p_1: & (-3,2,1), \\ p_2: & (1,-2,1), \\ p_3: & (4,-5,1). \end{array}$$

Next we use Theorems 4.3.7 and 4.3.9 to find a basis for V. Collect the coordinate vectors as rows in a matrix and row reduce:

$$\begin{pmatrix} -3 & 2 & 1 \\ 1 & -2 & 1 \\ 4 & -5 & 1 \end{pmatrix} \sim \begin{pmatrix} 1 & 0 & -1 \\ 0 & 1 & -1 \\ 0 & 0 & 0 \end{pmatrix}.$$

So the first two rows yield the basis $(1,0,-1)$, $(0,1,-1)$ of the span of the coordinate vectors of p_1, p_2, p_3. So the polynomials $1 - x^2$ and $x - x^2$ are a basis of V.

4.4 Notes

The notion of a vector space (or linear space) is the central concept in linear algebra; it is the (or a) formalized version of our intuition of space. Its strength lies in the fact that vector spaces can be used in many different situations. One of the first to describe vector spaces using axioms was Giuseppe Peano (1858–1932). In this course we only touch upon the precise role of the axioms. Probably you do not even notice that we use rules such as $\underline{0} + \underline{0} = \underline{0}$ and $0 \cdot \underline{a} = \underline{0}$, which we didn't prove (but see the exercises).

Vector spaces are used in many different situations, also outside mathematics. For instance, to describe notions like speed, acceleration force and impulse in mechanics, and fields in electromagnetism. In signal theory (to handle visual or audio signals) and in quantum mechanics vector spaces of functions are important. They tend to be infinite dimensional. *Mechanics*

As for geometry, although vector spaces are used to model 'flat' objects like lines and planes, vector spaces are also of help in describing tangent spaces to curved objects.

Finally, instead of using real or complex numbers, also other systems of scalars are possible, and most results still hold! In coding theory and cryptology, such number systems, like the integers modulo 2 are used (i.e., the numbers 0 and 1 with the rules $0+0=0$, $0+1=1+0=1$, $1+1=0$, $1 \cdot 0 = 0 \cdot 1 = 0$ and $1 \cdot 1 = 1$). *Coding theory* *Cryptology*

4.5 Exercises

§1

1 In each of the following cases decide if the subsets of $\mathbb{R}^3$ are linear subspaces of $\mathbb{R}^3$:

$$\begin{aligned} W_1 &= \{(x_1, x_2, x_3) \in \mathbb{R}^3 \mid x_2 \in \mathbb{Q}\}, \\ W_2 &= \{(x_1, x_2, x_3) \in \mathbb{R}^3 \mid x_1 + 2x_2 + 3x_3 = 0\}, \\ W_3 &= \{(x_1, x_2, x_3) \in \mathbb{R}^3 \mid x_1 + x_2 = 1\}, \\ W_4 &= \{(x_1, x_2, x_3) \in \mathbb{R}^3 \mid x_1 \geq 0\}. \end{aligned}$$

2 In each of the following cases decide if the subsets of $\mathbb{C}^3$ are linear subspaces of $\mathbb{C}^3$:

$$\begin{aligned} W_1 &= \{(z_1, z_2, z_3) \in \mathbb{C}^3 \mid z_1 + iz_2 + (1+i)z_3 = 0\}, \\ W_2 &= \{(z_1, z_2, z_3) \in \mathbb{C}^3 \mid z_1 + iz_2 + (1+i)\overline{z}_3 = 0\}, \\ W_3 &= \{(z_1, z_2, z_3) \in \mathbb{C}^3 \mid \mathrm{Re}(z_1) + \mathrm{Im}(z_2) = 0\}, \\ W_4 &= \{(z_1, z_2, z_3) \in \mathbb{C}^3 \mid \overline{z}_1 + i\overline{z}_2 = 0\}. \end{aligned}$$

3 Check whether the following subsets of the vector space of 2×3-matrices with real entries are linear subspaces:

a. the matrices of the form
$$\begin{pmatrix} a_{11} & 0 & 0 \\ a_{21} & a_{22} & 0 \end{pmatrix},$$

b. the matrices
$$\begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{pmatrix}, \quad \text{where } a_{11} + a_{22} = 0.$$

4 Consider the vector space V of all functions defined on $\mathbb{R}$. Check whether the following subsets of V are linear subspaces:

a. the functions f with $f(1) = 0$,

b. the functions f with $f(1) = 2$,

c. the functions f with $f(0) = f(1)$,

d. the functions f with $\lim_{x\to\infty} f(x) = 0$,

e. the functions f with $f(0) = 1 + f(1)$,

f. the functions f with $f(x) = f(-x)$ for all x,

g. the functions f with $f(x) \geq 0$ for all $x \in \mathbb{R}$,

h. the functions f with f a real polynomial of degree ≥ 3.

5 Determine a parametric representation of each of the following planes in $\mathbb{R}^3$:

a. $x + 4y - 5z = 7$,

b. $2x - 4y + z = 0$,

c. $2x + 4y + 4z = 7$.

6 Determine an equation of each of the following planes in $\mathbb{R}^3$:

a. $V: \underline{x} = (3, 0, 1) + \lambda(1, 1, 2) + \mu(0, 0, 1)$,

b. $V: \underline{x} = (1, 1, 1) + \lambda(2, 3, 1) + \mu(1, 0, 1)$,

c. $V: \underline{x} = (1, 1, 3) + \lambda(0, 1, 1) + \mu(-1, 0, 1)$.

7 Determine a parametric representation of each of the following planes in $\mathbb{R}^4$:

a. $V = \{(x, y, z, u) \mid x + 2y - 3z - u = 2,\ 2x + y + 6z + u = 7\}$,

b. $V = \{(x, y, z, u) \mid y + 2z - 2u = 1,\ 3y + 6z = 9\}$.

8 Consider the following planes in $\mathbb{R}^4$ which are given in parametric representation. Describe each of them by a system of two linear equations:

a. $U: \underline{x} = (1, 0, 0, 0) + \lambda(1, 1, 1, 0) + \mu(1, 2, 2, 2)$,

b. $V: \underline{x} = (-1, 1, 3, 5) + \lambda(2, -1, 1, 1) + \mu(-1, -1, 1, -2)$.

9 Let l be the line in $\mathbb{R}^3$ through $(3, 2, 1)$ and $(-3, 5, 4)$ and let V be the plane with equation $3x - y + 2z = 4$. Determine the intersection of l and V.

10 Let U and W be linear subspaces of the vector space V.

a. Prove that the intersection $U \cap W$ is also a linear subspace of V.

b. Prove that the subset $\{\, u + w \mid u \in U,\ w \in W \,\}$ is a linear subspace of V. (This subspace is often denoted by $U + W$.)

c. Show by example that the union $U \cup W$ is not necessarily a linear subspace of V.

§2, 3

11 Show that:

a. $(1,3)$ and $(3,9)$ form a linearly dependent set of vectors in $\mathbb{R}^2$,

b. $(1,2)$ and $(2,-1)$ form a linearly independent set of vectors in $\mathbb{R}^2$,

c. $(0,1,2)$, $(1,2,3)$, $(1,1,1)$ form a linearly dependent set of vectors in $\mathbb{R}^3$;

d. $(-1,5,5,3)$, $(-1,2,1,1)$, $(1,1,3,1)$ form a linearly dependent set of vectors in $\mathbb{R}^4$,

e. in $\mathbb{R}^3$, the vector $(1,2,1)$ is not a linear combination of $(1,3,2)$ and $(1,1,1)$,

f. in $\mathbb{R}^3$, the vector $(1,1,1)$ is a linear combination of $(3,-1,4)$, $(1,-3,2)$ and $(2,6,1)$,

g. in $\mathbb{R}^3$, the vector $(1,0,0)$ is not a linear combination of $(3,-1,4)$, $(1,-3,2)$ and $(2,6,1)$.

12 Which of the following systems are linearly independent:

a. e^{2t}, $\sin t$, $\cos t$ in the space of all functions on $\mathbb{R}$,

b. e^{it}, $\sin t$, $\cos t$ in the space of all functions on $\mathbb{R}$,

c. the polynomials z^2+1, z^3+z and $z+i$ in the vector space of polynomials.

d. 2, t, $\sin t$, $\cos t$ in the space of all functions on $\mathbb{R}$,

e. 2, t, $\sin^2 t$, $\cos^2 t$ in the space of all functions on $\mathbb{R}$,

13 Determine a basis of each of the following spans:

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

b. $<(1,1,1),(1,2,1),(1,0,1)>$ in $\mathbb{R}^3$,

c. $<(1,0,0,1),(2,1,1,3),(0,0,1,0)>$ in $\mathbb{R}^4$,

d. $<(3,-1,4,7),(1,-3,2,5),(2,6,1,-2)>$ in $\mathbb{R}^4$,

e. $<(3,-1,4,7),(1,-3,2,5),(5,3,2,-1)>$ in $\mathbb{R}^4$,

f. $<(3,-1,4,7),(1,-3,2,5),(2,6,1,-2),(0,4,-1,4)>$ in $\mathbb{R}^4$.

14 Check if the vectors $(2,-2,7,5)$, $(i,1+i,i,1)$, $(2+3i,-3+2i,1,-2+2i)$ belong to the span $<(3,-2,3,1),(2,1,-2,-1),(1,1,2,3)>$ in $\mathbb{C}^4$.

15 In the vector space of polynomials the following polynomials are given: $f(x) = x+1$, $g(x) = (x+1)^2$. Check if the polynomials x^2+3x+1, x^2-1, $3x^2-4x-7$ belong to the span $< f, g >$.

16 Let $\underline{a}$ and $\underline{b}$ be two vectors in a vector space. Prove that $< \underline{a}, \underline{b} >=< \underline{a}-\underline{b}, \underline{a}+\underline{b} >$.

17 Suppose the vectors $\underline{a}, \underline{b}, \underline{c}$ are linearly independent. Determine whether the following systems are linearly dependent:

a. $\underline{a}+\underline{b}$, $\underline{a}+\underline{b}-\underline{c}$, $2\underline{a}+\underline{b}+\underline{c}$;

b. $\underline{a}+\underline{b}+\underline{c}$, $\underline{a}+2\underline{b}$, $\underline{c}-\underline{b}$;

c. $\underline{a}+2\underline{b}$, $\underline{a}+\underline{c}$, $\underline{c}$.

18 Let $\underline{a_1}, \ldots, \underline{a_k}$ be a basis of the linear subspace U of the vector space V. Let $\underline{a}_{k+1}, \underline{a}_{k+2}, \ldots, \underline{a}_n$ be vectors in V. Show that $\underline{a}_1, \underline{a}_2, \ldots, \underline{a}_n$ is a basis of V if and only if the following holds:

$$\underline{a}_{k+1} \notin U, \quad \underline{a}_{k+2} \notin < \underline{a}_1, \underline{a}_2, \ldots, \underline{a}_{k+1} >, \ldots, \underline{a}_n \notin < \underline{a}_1, \underline{a}_2, \ldots, \underline{a}_{n-1} > .$$

19 Determine a parametric representation of the plane containing the point $(0,1,1)$ and the line $\underline{x} = (1,-1,0) + \lambda(2,1,1)$. Also determine an equation of this plane.

20 In $\mathbb{R}^3$ the subset V and the line l are described as follows

$$V : \underline{x} = (1,1,1) + \lambda(1,a,a^2) + \mu(1,a,4), \quad l : \underline{x} = \sigma(1,3,3).$$

a. For which values of a is V a plane?

b. For which values of a do l and V not intersect (which means that l and V are parallel)?

21 a. Determine all vectors in $\mathbb{R}^3$ which belong to both of these spans:

$$U_1 =< (-4,1,3), (-2,3,1) > \quad \text{and} \quad U_2 =< (-1,5,4), (3,-1,2) >,$$

i.e., determine the intersection $U_1 \cap U_2$.

b. Determine all vectors in $\mathbb{R}^4$ which belong to both of the following spans:

$$V_1 =< (4,3,2,1), (1,0,0,0) > \quad \text{and} \quad V_2 =< (2,1,0,0), (3,2,1,0) >,$$

i.e., determine $V_1 \cap V_2$.

22 Determine a basis and the dimension of the following subspaces of $\mathbb{C}^3$:

a. $< (3,-1,4), (1,-3,2), (5,3,2) >$,

b. $< (i, 1+2i, 1+i), (i, 1+3i, 2+2i), (2+2i, 5+i, 2) >$,

c. $< (2i, 3i, i), (1+i, 2+i, i), (-1+2i, -1+3i, -1+i) >$.

23 Determine a basis and the dimension of each of the following spans (in the space of functions from $\mathbb{R}$ to $\mathbb{R}$):

a. $< e^{2t}, t^2, t >$,

b. $< 2, t, \sin^2 t, \cos^2 t >$,

c. $< e^{2t}, e^{-t}, \cosh t, \sinh t >$,

d. $< 2x^3 + x^2 - x + 5, x^3 + 2x^2 + 10, -2x^2 + x, 2x^3 - 8x^2 + x - 10 >$.

24 Determine the dimension of each of the following subspaces:

a. the subspace of P_5 of the polynomials which are zero in 2,

b. the space $M_{3,3}$ of 3×3–matrices,

c. the subspace of $M_{3,3}$ consisting of the matrices A with $A = A^\top$,

d. the subspace of $M_{3,3}$ consisting of the matrices A for which $A + A^\top = 0$,

e. the subspace of $M_{2,3}$ consisting of the matrices A, for which

$$A \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} = 0,$$

f. the subspace of $M_{2,2}$ consisting of the matrices A, for which

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

25 In the vector space V the linearly independent set $\underline{a}, \underline{b}, \underline{c}$ is given. Determine the dimension of each of the following spans:

a. $< \underline{a} - \underline{b}, \underline{a} + \underline{b} + \underline{c}, -2\underline{a} - \underline{c} >$,

b. $< \underline{a} - \underline{b}, \underline{a} + \underline{b}, \underline{a} + \underline{b} + \underline{c} >$,

26 Determine the coordinates of each of the following vectors with respect to the given bases:

a. $(2,3)$ with respect to $(1,0),(1,1)$,

b. $(1,2,3)$ with respect to $(1,1,1),(1,0,1),(0,0,1)$,

c. x^2 with respect to $1-x, 1-x^2, x+x^2$,

d. $\cos 2t$ with respect to $4, \sin^2 t$.

27 In this problem we will prove some further properties of vectors. We need the eight axioms mentioned in 4.1.2.

a. If $\underline{a}+\underline{0}=\underline{a}+\underline{b}$, then $\underline{b}=\underline{0}$. Prove this by adding the opposite $-\underline{a}$ of $\underline{a}$ to both sides.

b. For all scalars $\lambda\,\underline{0}=\underline{0}$. Indicate which axioms are used in the following derivation. We have: $\lambda\,\underline{0}=\lambda(\underline{0}+\underline{0})=\lambda\,\underline{0}+\lambda\underline{0}$. But we also have: $\lambda\,\underline{0}=\lambda\,\underline{0}+\underline{0}$. So $\lambda\,\underline{0}+\lambda\underline{0}=\lambda\,\underline{0}+\underline{0}$, so that part a. implies $\lambda\,\underline{0}=\underline{0}$.

c. This item focuses on the equality $0\,\underline{a}=\underline{0}$ (for every $\underline{a}$). Finish the following proof: $0\,\underline{a}=(0+0)\underline{a}=0\,\underline{a}+0\,\underline{a}$.

d. The zero vector $\underline{0}$ is unique, i.e., if $\underline{0}'$ also satisfies $\underline{a}+\underline{0}'=\underline{a}$ for all $\underline{a}$, then $\underline{0}=\underline{0}'$. Prove this by considering the expression $\underline{0}+\underline{0}'$.

e. The opposite $-\underline{a}$ of a given vector $\underline{a}$ is unique. Suppose $\underline{b}$ is also an opposite of $\underline{a}$. Then finish the following chain of equalities to provide the proof: $-\underline{a}=-\underline{a}+\underline{0}=-\underline{a}+(\underline{a}+\underline{b})$.

28 Prove: a system of vectors $\underline{a}_1,\ldots,\underline{a}_n$ in a vector space V is a basis of V if and only if every vector from V can be written as a linear combination of the vectors $\underline{a}_1,\ldots,\underline{a}_n$ in exactly one way.

29 V is a finite dimensional vector space; U and W are subspaces of V.

a. Assume that each vector in V can be written as a sum of a vector from U and a vector from W. (Then we write $V=U+W$.) Show that $\dim(V)\le\dim(U)+\dim(W)$.

b. If $V=U+W$ and $U\cap W=\{\underline{0}\}$, then every vector from V can be written in exactly one way as a sum of a vector from U and a vectr from W (i.e. if $\underline{x}=\underline{u}+\underline{w}=\underline{u}'+\underline{w}'$ with $\underline{u},\underline{u}'\in U$ and $\underline{w},\underline{w}'\in W$, then $\underline{u}=\underline{u}'$ and $\underline{w}=\underline{w}'$). Derive this from the relation $\underline{u}-\underline{u}'=\underline{w}'-\underline{w}$. Also show that in this situation $\dim(V)=\dim(U)+\dim(W)$.

4.5.1 Exercises from old exams

30 a. If $\underline{a}, \underline{b}, \underline{c}$ is a linearly independent system of vectors in a vector space V, then prove that $\underline{a}+\underline{b}$, $\underline{a}-\underline{b}$, $\underline{a}-2\underline{b}+\underline{c}$ is a linearly independent system.

b. In the real vector space $\mathbb{P}_2$ of polynomials of degree at most 2, the following polynomials are given:

$$p_1(x) = x^2 + 2x - 3, \quad p_2(x) = x^2 + cx + 1, \quad p_3(x) = x^2 - 5x + 4.$$

For which value(s) of c is $p_1(x)$, $p_2(x)$, $p_3(x)$ a basis of $\mathbb{P}_2$?

31 $C^\infty(\mathbb{R})$ is the vector space of infinitely differentiable functions $f : \mathbb{R} \to \mathbb{R}$.

a. Determine a basis of the subspace

$$U = \langle 1 + 2x, 4x - e^{3x}, 2 + e^{3x} \rangle$$

of $C^\infty(\mathbb{R})$. What is the dimension of U?

b. Determine the intersection $U \cap W$, where $W = \langle 1, e^x, e^{3x} \rangle$.

32 a. Let $U = \{\, (x, y, z, u) \mid x+y+z = 0,\ y+z+u = 0 \,\}$ be a plane in $\mathbb{R}^4$. Determine a vector parametric representation of U.

b. Determine the intersection of the plane U from part a) and the plane

$$V : \underline{x} = (2, 2, 2, 1) + \lambda(1, 3, 2, 0) + \mu(1, 2, 3, 0).$$

Chapter 5

Rank and inverse of a matrix, determinants

5.1 Rank and inverse of a matrix

5.1.1 In this section we first concentrate on the linear subspaces spanned by the rows and columns of a matrix, respectively. Then we study the relation with systems of linear equations and with the inverse of a matrix.

5.1.2 Definition. (Row and column space) Let A be a matrix with n rows and m columns. Then every row has m entries so that these rows can be seen as vectors in in $\mathbb{R}^m$ or $\mathbb{C}^m$; the subspace spanned by the rows is called the *row space* of the matrix. Similarly, every column is an element of $\mathbb{R}^n$ or $\mathbb{C}^n$; the space spanned by the columns is called the *column space* of the matrix.

5.1.3 We agree to consider length n sequences of numbers, but written in column form, as elements of $\mathbb{R}^n$ or $\mathbb{C}^n$, respectively, and that, when convenient, we consider elements of $\mathbb{R}^n$ or $\mathbb{C}^n$, as columns. Of course, we try to avoid any confusion in doing so. For instance, we write: the system $A\underline{x} = \underline{b}$ with $\underline{x} \in \mathbb{R}^n$, where $\underline{x}$ is then seen as a column vector.

5.1.4 Example. Let

$$A = \begin{pmatrix} 1 & -1 & 3 & 7 \\ 2 & 1 & 1 & 5 \end{pmatrix}.$$

Then the row space is

$$< (1, -1, 3, 7), (2, 1, 1, 5) > \subseteq \mathbb{R}^4$$

and the column space is

$$< (1,2), (-1,1), (3,1), (7,5) > \subseteq \mathbb{R}^2.$$

5.1.5 The row and column spaces of a matrix seem to be quite unrelated, since they are usually subspaces of different vector spaces. Yet their dimensions are the same! To show this we first connect the matrix product to linear combinations of the columns (or rows). The following example shows how a matrix product can be rewritten as a linear combination of the columns of the 3×2–matrix.

$$\begin{aligned}\begin{pmatrix} 1 & 3 & -1 \\ 2 & -2 & 5 \end{pmatrix} \begin{pmatrix} 3 \\ 2 \\ 6 \end{pmatrix} &= \begin{pmatrix} 3 \cdot 1 + 2 \cdot 3 + 6 \cdot (-1) \\ 3 \cdot 2 + 2 \cdot (-2) + 6 \cdot 5 \end{pmatrix} \\ &= 3 \begin{pmatrix} 1 \\ 2 \end{pmatrix} + 2 \begin{pmatrix} 3 \\ -2 \end{pmatrix} + 6 \begin{pmatrix} -1 \\ 5 \end{pmatrix}.\end{aligned}$$

In general, the matrix product

$$\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nm} \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_m \end{pmatrix}$$

can be rewritten as

$$x_1 \begin{pmatrix} a_{11} \\ \vdots \\ a_{n1} \end{pmatrix} + x_2 \begin{pmatrix} a_{12} \\ \vdots \\ a_{n2} \end{pmatrix} + \cdots + x_m \begin{pmatrix} a_{1m} \\ \vdots \\ a_{nm} \end{pmatrix},$$

i.e., as a linear combination of the columns $\underline{k}_1, \ldots, \underline{k}_m$ of the matrix A. In a similar way, the product

$$\begin{pmatrix} y_1 & \cdots & y_n \end{pmatrix} \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nm} \end{pmatrix}$$

can be viewed as a linear combination $y_1\underline{r}_1 + \cdots + y_n\underline{r}_n$ of the rows $\underline{r}_1, \ldots, \underline{r}_n$ of the matrix. A first conclusion from these considerations is the following.

5.1.6 Theorem. *The system of linear equations (in matrix form) $A\underline{x} = \underline{b}$ has a solution if and only if $\underline{b}$ belongs to the column space of A.*

5.1.7 Two linear combinations of the columns can be described by using a $m \times 2$ matrix instead of a $m \times 1$ matrix. Likewise for more than two linear combinations. For example, in the following matrix product the two columns of the 2×2–matrix from the right-hand side are linear combinations of the three columns of the 2×3-matrix from the left-hand side:

$$\begin{pmatrix} 1 & 3 & -1 \\ 2 & -2 & 5 \end{pmatrix} \begin{pmatrix} 3 & 1 \\ 2 & -4 \\ 6 & 2 \end{pmatrix} = \begin{pmatrix} 3 & -13 \\ 32 & 20 \end{pmatrix}.$$

In general, the columns in the product AB are linear combinations of the columns of A (to prove this, verify that the ℓ-th column of AB is $B_{1\ell}$ times the first column of A plus $B_{2\ell}$ times the second column of A etc.); the rows of AB are linear combinations of the rows of B.

Now suppose that the columns of A span a subspace of dimension k. Let $\underline{c}_1, \ldots, \underline{c}_k$ be a basis of this space. Arrange these vectors as columns in a $n \times k$–matrix C. Each of the m columns of A is a linear combination of these columns of C; these linear combinations are assembled in the matrix product

$$CX = A,$$

in which X is a $k \times m$–matrix.

Now concentrate on the rows in this equality: every row of A is a linear combination of the rows of X. Since the number of rows of X equals k, the row space has dimension at most k. So

$$\dim(\text{rowspace}) \leq \dim(\text{columnspace}).$$

Note that this equality applies to *any* matrix. In particular, if we apply it to A^T and note that the dimension of the row space (column space) of A^T equals the dimension of the column space (row space) of A, we find:

$$\dim(\text{columnspace}) \leq \dim(\text{rowspace}).$$

Summarizing:

5.1.8 Theorem. *For every matrix A the dimension of the row space equals the dimension of the column space.*

5.1.9 Definition. (Rank) The *rank* of a matrix is by definition the dimension of its row or column space. Notation: $\operatorname{rank}(A)$.

5.1.10 Determining the rank of a matrix is straightforward: row reduce till you reach the row reduced echelon form and count the number of nonzero rows.

In the remainder of this section we concentrate on $n \times n$–matrices.

5.1.11 Theorem. *For a $n \times n$–matrix A the following statements are equivalent:*

1. *the rank of A is n;*
2. *the rows of A are linearly independent;*
3. *the columns of A are linearly independent;*
4. *the row reduced echelon form of A is the identity matrix.*

Proof. We first show that 1), 2) and 3) are equivalent, and then that 1) and 4) are equivalent.

If the rank of A is n, then the n rows and the n columns span an n-dimensional space, hence must be linearly independent. So 1) implies 2) and 3). Conversely, if the n rows (or columns) are linearly independent, then the rows (columns) span a n-dimensional space and so the rank must be n. So 2) implies 1), and 3) implies 1). Hence 1), 2), 3) are equivalent.

Since A and its row reduced echelon form have the same rank, we see that 4) implies 1). Now suppose A has rank n and consider the last row of the row reduced echelon form. It can't be the zero row, because then the row space is spanned by $n-1$ rows and its dimension would be at most $n-1$. So there must be a 1 in the last row. Since the last row starts with more zeros than the $n-1$-th row, and this $n-1$-th row starts with more zeros than the $n-2$-th row, etc, this 1 must be in position n, n. In a similar way (or use induction), you show that, for $k = 1, \ldots, n-1$, the k-th row is the k-th row of the identity matrix. So 1) implies 4) and we are done. $\square$

5.1.12 Corollary. *A system $A\underline{x} = \underline{b}$ of n linear equations in n variables has exactly one solution if and only if the rank of the coefficient matrix equals n.*

Proof. If the rank of the coefficient matrix is n, then the columns are a basis of $\mathbb{R}^n$ (or $\mathbb{C}^n$). Every vector in $\mathbb{R}^n$ (or $\mathbb{C}^n$) can then be written in a unique way as a linear combination of the columns, i.e., the system $A\underline{x} = \underline{b}$ has exactly one solution.

If the rank of A is less than n, then the columns of A are not linearly independent. This means that the system has infinitely many solutions, or no solutions at all (this last case means that $\underline{b}$ is not in the column space of A). So if the system has exactly one solution, the rank of A must be n. $\square$

5.1.13 Theorem. *If the $m \times n$–matrix A has rank k, then the solution space of the homogeneous system of linear equations $A\underline{x} = \underline{0}$ in n variables has dimension $n-k$. In other words: the dimension of the solution space equals the number of variables minus the number of 'independent' conditions.*

Proof. Since the rank of A nor the solution space of $A\underline{x} = \underline{0}$ changes if we replace A by its row reduced echelon form, we can restrict our attention to the case that A is already in row reduced echelon form. The rank of A is then equal to the number of nonzero rows (see Theorem 4.3.9). The first place (from the left) in row j containing a 1 corresponds to variable x_{i_j}, say. So the k variables $x_{i_1}, \ldots, x_{i_k}$ cannot be assigned a value arbitrarily, while the remaining $n-k$ variables, say $x_{j_1}, \ldots, x_{j_{n-k}}$ in spots $j_1, \ldots, j_{n-k}$ can be assigned arbitrary values, $\lambda_{j_1}, \ldots, \lambda_{j_{n-k}}$, say. The solutions can then be written as $\lambda_{j_1}\underline{a}_{j_1} + \cdots + \lambda_{j_{n-k}}\underline{a}_{j_{n-k}}$. The vector $\underline{a}_{j_l}$ has a 1 in position j_l, while the remaining vectors have a zero in that position. This implies that these $n-k$ vectors are linearly independent (why?). So the dimension of the solution space is $n-k$. □

5.1.14 Inverse of a matrix

Finally, we use the previous results to discuss a procedure to find the inverse of an $n \times n$–matrix, if this inverse exists. The main result ifs the following theorem; the proof provides a way to compute inverses.

5.1.15 Theorem. *The inverse of an $n \times n$–matrix A exists if and only if* $\mathrm{rank}(A) = n$.

Proof. Let A be an $n \times n$–matrix of rank n. Then the columns $\underline{k}_1, \ldots, \underline{k}_n$ of A are linearly independent and are a basis of $\mathbb{R}^n$ (or $\mathbb{C}^n$). We are looking for an $n \times n$–matrix X with elements x_{ij} so that

$$A \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ \vdots & \vdots & & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nn} \end{pmatrix} = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & \cdots & 0 & 1 \end{pmatrix}.$$

The elements of the i-th column $(x_{1i}, \ldots, x_{ni})$ of X should satisfy $x_{1i}\underline{k}_1 + \cdots + x_{ni}\underline{k}_n = \underline{e}_i$. Since the columns of A are a basis of $\mathbb{R}^n$ (or $\mathbb{C}^n$), such a column exists (and is in fact unique). So there is a unique inverse.

Conversely, if an inverse matrix X exists, then we conclude from

$$x_{1i}\underline{k}_1 + \cdots + x_{ni}\underline{k}_n = \underline{e}_i \quad \text{for all}\, i$$

that $\underline{e}_1, \ldots, \underline{e}_n$ are in the column space of A. The n columns of A therefore span the n-dimensional space $\mathbb{R}^n$ (or $\mathbb{C}^n$), hence must be linearly independent (otherwise this n-dimensional space could be spanned with less columns contradicting the fact that the dimension is n). So the rank of A is n by Theorem 5.1.11. □

5.1.16 Computing inverses
The proof just given provides a way to compute inverses. To find the i-th column of the inverse, you need to solve the system of linear equations with extended matrix $(A|\underline{e}_i)$. Since the resulting n systems all have the same coefficient matrix A, these systems can be solved *simultaneously*! To actually do this, consider the matrix $(A|\underline{e}_1,\ldots,\underline{e}_n) = (A|I)$. Row reduce and read off the solutions on the right of the vertical bar: $(A|I)$ is being reduced to $(I|A^{-1})$ if the inverse exists. Whether this inverse exists can be concluded during the process: if the rank of A turns out to be less than n, then row reducing of $(A|I)$ produces a matrix in which the last row is

$$(0\ldots 0|*\ldots *),$$

where the last n $*$'s cannot all be 0 (the rows of I are linearly independent so must remain nonzero in the row reducing process). So the system has no solutions.

In *Linear Algebra 2* we will be able to prove in a simple way that if a matrix B satisfies $A\,B = I$, then automatically $B\,A = I$. The matrix computed according to the above procedure is therefore indeed the inverse of A.

5.1.17 Example. To determine the inverse of the matrix

$$A = \begin{pmatrix} 4 & 0 & 1 \\ 2 & 2 & 0 \\ 3 & 1 & 1 \end{pmatrix}$$

we consider $(A|I)$, i.e.,

$$\left(\begin{array}{ccc|ccc} 4 & 0 & 1 & 1 & 0 & 0 \\ 2 & 2 & 0 & 0 & 1 & 0 \\ 3 & 1 & 1 & 0 & 0 & 1 \end{array}\right),$$

and row reduce:

$$\left(\begin{array}{ccc|ccc} 1 & 0 & 0 & \frac{1}{2} & \frac{1}{4} & -\frac{1}{2} \\ 0 & 1 & 0 & -\frac{1}{2} & \frac{1}{4} & \frac{1}{2} \\ 0 & 0 & 1 & -1 & -1 & 2 \end{array}\right).$$

The first column of the inverse matrix is then the first column on the right of the vertical bar, etc. So we find:

$$A^{-1} = \begin{pmatrix} \frac{1}{2} & \frac{1}{4} & -\frac{1}{2} \\ -\frac{1}{2} & \frac{1}{4} & \frac{1}{2} \\ -1 & -1 & 2 \end{pmatrix} = \frac{1}{4}\begin{pmatrix} 2 & 1 & -2 \\ -2 & 1 & 2 \\ -4 & -4 & 8 \end{pmatrix}.$$

5.2 Determinants

5.2.1 In the previous section we have seen that an $n \times n$-matrix A of rank n has some pleasant properties: its inverse exists and the system of linear equations $(A|\underline{b})$ has exactly one solution. To determine the rank of a matrix can be done using row reduction and a simple count of nonzero rows. A second technique which we will discuss, is to determine the so-called determinant of an $n \times n$-matrix. This is a number which is nonzero if and only if the rank of the matrix equals n. Computing determinants is fairly easy (at least if n is relatively small), but the theory behind is more complicated.

The plan for this section is as follows.

- First we discuss the case of 2×2–determinants in detail, because almost all aspects of determinants can be illustrated in this case.
- Then we turn to the definition of a general $n \times n$–determinant.
- Next we concentrate on various ways to compute determinants, like expanding with respect to a row or column and the role of row reducing.
- Finally, we discuss the connection with systems of linear equations (Cramer's rule) and with inverse matrices.

5.2.2 Definition. (2×2–determinant) The 2×2–*determinant* of a 2×2–matrix

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

is the number $ad - bc$. Notations: $\det(A)$ and

$$\begin{vmatrix} a & b \\ c & d \end{vmatrix}.$$

5.2.3 It's easy to show that $\det(A) \neq 0 \Leftrightarrow \operatorname{rang}(A) = 2$. If, for instance, the second row is a multiple of the first, say $(c, d) = \lambda(a, b)$, then it easily follows that $\det(A) = a(\lambda b) - b(\lambda a) = 0$. Similarly, if the first row is a multiple of the second row. Conversely, if $\det(A) = 0$ and $a \neq 0$, then from $ad = bc$ we get $d = (bc/a)$ so that $(c, d) = (c/a)(a, b)$. Finish the proof yourself by considering the case $a = 0$.

The number $\det(A)$ also plays a role in solving the system of equations $A\underline{x} = \underline{p}$ in two variables. If A_1 is the matrix obtained from A by replacing the first column by $\underline{p}$, and A_2 the matrix obtained from A by replacing the second column of A by $\underline{p}$, then the unique solution of the system

$$\begin{aligned} ax_1 + bx_2 &= p_1 \\ cx_1 + dx_2 &= p_2 \end{aligned}$$

is

$$x_1 = \frac{\det(A_1)}{\det(A)} = \frac{p_1 d - b p_2}{ad - bc}, \quad x_2 = \frac{\det(A_2)}{\det(A)} = \frac{a p_2 - p_1 c}{ad - bc},$$

provided $\det(A) \neq 0$. You can check that this is really the solution, but we will come across a nice proof in 5.2.22. There is a similar formula for the inverse of a 2×2–matrix of rank 2.

The determinant $\det(A)$ also plays a role in surface area computations. (We haven't defined surface area's exactly, so this discussion is only by way of illustration.) Figure 5.1 provides a 'proof by pictures' of the fact that the area of a parallellogram spanned by the vectors (a, b) and (c, d) is equal to $ad - bc = \det(A)$. This number is called the 'oriented area' since it can be negative. The 'true' area is obtained by taking the absolute value. The expression $ad - bc$ is not linear, but

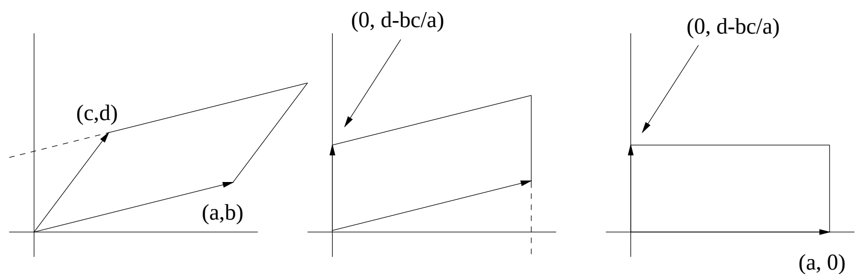


Figure 5.1: De determinant en de oppervlakte van een parallellogram.

is what is called *bilinear* (as will be discussed later). Most properties of the 2×2–determinant recur when we discuss $n \times n$–determinants. Consider the determinant as a function of the two rows (or columns, but for the moment we focus on rows) of the matrix: $\det(\underline{a}_1, \underline{a}_2)$. Then the properties that we mean are mainly (for all choices of vectors, scalars):

1. bilinearity:

$$\begin{aligned} \det(\lambda_1 \underline{b}_1 + \lambda_2 \underline{b}_2, \underline{a}_2) &= \lambda_1 \det(\underline{b}_1, \underline{a}_2) + \lambda_2 \det(\underline{b}_2, \underline{a}_2), \\ \det(\underline{a}_1, \lambda_1 \underline{b}_1 + \lambda_2 \underline{b}_2) &= \lambda_1 \det(\underline{a}_1, \underline{b}_1) + \lambda_2 \det(\underline{a}_1, \underline{b}_2), \end{aligned}$$

   (sometimes we say: linear in both entries);

2. antisymmetry: $\det(\underline{a}_1, \underline{a}_2) = -\det(\underline{a}_2, \underline{a}_1)$ (interchanging two vectors introduces a minus sign); if the two vectors are equal, then we conclude: $\det(\underline{a}, \underline{a}) = 0$ since this number is equal to its negative;

3. normalization: $\det(\underline{e}_1, \underline{e}_2) = 1$.

These properties are easy to verify. For instance, the second property follows from:

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

It seems a bit pompous to desribe an easy expression like $ad-bc$ with these abstract looking properties, but in higher dimensions the description with the properties is extremely useful as opposed to explicit formulas for $n \times n$ determinants.

The determinant is unique in the sense that any function D of pairs of vectors having the three properties mentioned above must be the determinant function. To show this, we first note that $D(\underline{a}, \underline{a}) = 0$ (use antisymmetry). Next, using bilinearity we find:

$$\begin{aligned} D((a, b), (c, d)) &= D(a\underline{e}_1 + b\underline{e}_2, c\underline{e}_1 + d\underline{e}_2) \\ &= aD(\underline{e}_1, c\underline{e}_1 + d\underline{e}_2) + bD(\underline{e}_2, c\underline{e}_1 + d\underline{e}_2) \\ &= acD(\underline{e}_1, \underline{e}_1) + adD(\underline{e}_1, \underline{e}_2) + bcD(\underline{e}_2, \underline{e}_1) + bdD(\underline{e}_2, \underline{e}_2) \\ &= ad\, D(\underline{e}_1, \underline{e}_2) + bc\, D(\underline{e}_2, \underline{e}_1) \\ &= (ad - bc)D(\underline{e}_1, \underline{e}_2) = ad - bc. \end{aligned}$$

In higher dimensions a similar computation (which we will skip) shows that the $n \times n$ determinant is unique.

Now we turn to the $n \times n$ determinant and start with a definition of a determinant function of n vectors in $\mathbb{R}^n$ or $\mathbb{C}^n$.

5.2.4 Definition. (Determinant function) A *determinant function* on $\mathbb{R}^n$ or $\mathbb{C}^n$ is a function D that assigns to every n-tuple of vectors $\underline{a}_1, \ldots, \underline{a}_n$ a real (or complex) number and has the properties:

1. **Multilinearity**:

$$\begin{aligned} &D(\underline{a}_1, \ldots, \underline{a}_{i-1}, \sum_{k=1}^{m} \beta_k \underline{b}_k, \underline{a}_{i+1}, \ldots, \underline{a}_n) \\ &\quad = \sum_{k=1}^{m} \beta_k D(\underline{a}_1, \ldots, \underline{a}_{i-1}, \underline{b}_k, \underline{a}_{i+1}, \ldots, \underline{a}_n), \end{aligned}$$

   and similar expressions for $D(\sum_{k=1}^{m} \beta_k \underline{b}_k, \underline{a}_2, \ldots, \underline{a}_n)$ up to $D(\underline{a}_1, \ldots, \underline{a}_{n-1}, \sum_{k=1}^{m} \beta_k \underline{b}_k)$. We sometimes say that D is linear in every entry.

2. **Antisymmetry**: by interchanging two vectors the determinant function changes sign.

3. **Normalization**: $D(\underline{e}_1, \underline{e}_2, \ldots, \underline{e}_n) = 1$.

5.2.5 Note that the definition doesn't garantee that determinant functions exist for all n (for $n = 2$ we have seen that there is one). We will show that there is, for every n, precisely one determinant function, and we will discuss ways to compute such determinants. We will usually call the unique determinant function simply the *determinant*.

5.2.6 Suppose D is a determinant function, then D satisfies the following properties.

- $D(\ldots, \underline{a}, \ldots, \underline{a}, \ldots) = 0$, i.e., of two of the vectors $\underline{a}_1, \ldots, \underline{a}_n$ are equal then the determinant is 0: on the one hand, by interchanging two vectors the determinant changes sign because of antisymmetry, on the other hand the determinant stays the same since the two vectors are the same. But if a number equals minus that number, it has to be 0.

- $D(\underline{a}_1, \ldots, \underline{a}_n) = 0$ if the vectors $\underline{a}_1, \ldots, \underline{a}_n$ are not linearly independent. Suppose for instance that

  $$\underline{a}_1 = \alpha_2 \underline{a}_2 + \cdots + \alpha_n \underline{a}_n.$$

  Then

  $$\begin{aligned} D(\underline{a}_1, \underline{a}_2, \ldots, \underline{a}_n) &= D(\sum_{k=2}^{n} \alpha_k \underline{a}_k, \underline{a}_2, \ldots, \underline{a}_n) \\ &= \sum_{k=2}^{n} \alpha_k D(\underline{a}_k, \underline{a}_2, \ldots, \underline{a}_n) = 0 \end{aligned}$$

  because of the previous item.

5.2.7 Using these conditions we can write out what the $n \times n$ determinant should be: just like for 2×2 matrices, write every row as a linear combination of the standard basis vectors and use the multilinearity to expand the determinant as a sum of many determinants with standard basis vectors as rows (in some order). Here is how to do that. Consider

$$A = \begin{pmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2n} \\ \vdots & \ldots & \ldots & \vdots \\ a_{n1} & \ldots & \ldots & a_{nn} \end{pmatrix}$$

with rows $\underline{a}_1, \ldots, \underline{a}_n$. Then write for every i

$$\underline{a}_i = \sum_{j=1}^{n} a_{ij} \underline{e}_j$$

and so

$$\begin{aligned} D\ (\underline{a}_1, \ldots, \underline{a}_n) &= D\Big(\sum_{j_1=1}^{n} a_{1j_1} \underline{e}_{j_1}, \ldots, \sum_{j_n=1}^{n} a_{nj_n} \underline{e}_{j_n} \Big) \\ &= \sum_{j_1=1}^{n} \ldots \sum_{j_n=1}^{n} a_{1j_1} \ldots a_{nj_n} D(\underline{e}_{j_1}, \ldots, \underline{e}_{j_n})\ . \end{aligned}$$

This is a sum with many terms if n gets big: there are n summation indices each one of which assumes n values, so the number of terms is n^n. For $n = 8$ this already amounts to 16.777.216 terms.

In reality there are less terms, since if two of the indices are equal, then we are dealing with a determinant with two equal rows and such a determinant is 0. So, if D exists, then

$$D(\underline{a}_1, \ldots, \underline{a}_n) = \sum_{\text{alle } j_i \text{ distinct}} a_{1j_1} \ldots a_{nj_n} D(\underline{e}_{j_1}, \ldots, \underline{e}_{j_n})\ .$$

Now which indices $(j_1, \ldots, j_n)$ occur in this sum? From the fact that all numbers $j_1, \ldots, j_n$ should be distinct and lies in between 1 and n we conclude that in $(j_1, \ldots, j_n)$ every number between 1 and n occurs precisely once. Such a sequence is called a *permutation* of the numbers $1, \ldots, n$. For example, the permutations of $1, 2, 3$ are

$$(1,2,3),\ (1,3,2),\ (2,1,3),\ (2,3,1),\ (3,1,2),\ (3,2,1).$$

These permutations can be listed as follows: choose an element from $\{1, 2, 3\}$. This can be done in three ways. Then there are two left to choose from in the next step. After that, there is only one choice left for the third element. So there are $3 \times 2 \times 1 = 3! = 6$ permutations of the numbers 1, 2, 3.

In general, there are $n! = n \cdot (n-1) \cdots 2 \cdot 1$ ('n factorial') permutations for the numbers $1, 2, \ldots, n$.

The number of terms is therefore $n!$, which is substantially less than n^n, but still large. For instance, $8! = 40.320$. Unfortunately, a further a priori reduction is not possible (though we will see that there are ways to compute determinants avoiding writing down all these terms). Since any sequence $j_1, \ldots, j_n$ consists of all numbers $1, \ldots, n$, by repeatedly interchanging two elements in the sequence in a clever way (so-called transpositions), we can attain the sequence $1, \ldots, n$. Every step in which we interchange two elements introduces a factor -1 , so that,

$$D(\underline{a}_1, \ldots, \underline{a}_n) = \sum_{j_1, \ldots, j_n} \pm a_{1j_1} \ldots a_{nj_n}, \tag{5.1}$$

where the sum runs through all $n!$ permutations of $(1,\ldots,n)$ and where a term is preceded by $+1$ if the corresponding permutation can be changed into $(1,\ldots,n)$ by an even number of 'transpositions', and by -1 otherwise.

The formula shows that for any n there is at most one determinant function (defined by the formula (5.1) just given). Conversely, one can prove that (5.1) satisfies the requirements from Definition 5.2.4 (the proof comes down to proving that the parity of the number of transpositions, i.e., whether you need an even or odd number, involved in changing a permutation into $1,\ldots,n$ does not depend on the choice of transpositions used). We will skip this proof, since it belongs to the domain of algebra. Given that determinants exist, we now concentrate on the cases $n=2$ and $n=3$. The discussions for $n>3$ are similar.

5.2.8 2×2 **determinant**

Take $n=2$ and consider the matrix with rows $\underline{a}_1=(a,b)$ en $\underline{a}_2=(c,d)$.

There are only two permutations in this case: $(1,2)$ and $(2,1)$. The first one comes with a plus sign in (5.1) and the second is one step of interchanging away from $(1,2)$ and so comes with a minus sign. the rows. So:

$$\begin{vmatrix} a & b \\ c & d \end{vmatrix} = a\,d - b\,c\ ,$$

in agreement with definition 5.2.2.

5.2.9 3×3 **determinant**

Take $n=3$ and consider

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

There are 6 permutations of $(1,2,3)$; the ones that come with a plus sign are $(1,2,3)$, $(2,3,1)$, $(3,1,2)$, and the ones with a minus sign are $(1,3,2)$, $(2,1,3)$, $(3,2,1)$. We find:

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

This expression is also known as *Sarrus' rule*. This rule is easy to remember. Just put copies of the first two columns on the right-hand side of the matrix,

$$\begin{matrix} a_{11} & a_{12} & a_{13} & a_{11} & a_{12} \\ a_{21} & a_{22} & a_{23} & a_{21} & a_{22} \\ a_{31} & a_{32} & a_{33} & a_{31} & a_{32} \end{matrix}$$

and then take the sum of the products of the elements on each of the diagonals (from upper left to lower right) and subtract the products of the 'anti diagonals' (from upper right to lower left).

By inspection of the terms we see that each of a_{11}, a_{12}, a_{13} occurs linearly in every term reflecting the fact that the determinant is really linear in the first vector. It takes some more work to verify from the formula that interchanging $\underline{a}_1$ and $\underline{a}_2$, or $\underline{a}_2$ and $\underline{a}_3$, or $\underline{a}_3$ and $\underline{a}_1$, only changes the sign of the determinant (and we will not write out the details). Finally, if $a_{11} = a_{22} = a_{33} = 1$ and $a_{ij} = 0$ for $i \neq j$, then $D(\underline{e}_1, \underline{e}_2, \underline{e}_3) = 1$. These considerations show that the determinant exists for $n = 3$.

5.2.10 In general one can prove that expression (5.1) indeed defines a determinant function for every n. The proof is quite involved and we will not provide details in this course. The subtle point is the sign: one needs that the parity of the number of steps you need in rewriting a given permutation doesn't depend on the way you actually carry out these steps.

In practice, the expression (5.1) is almost never useful. To actually compute determinants there are much better ways than using this formula as we will see below. Here are a number of results on determinants and some words on the proofs.

5.2.11 Theorem. *For every $n \times n$-matrix A:*

$$\det(A) = \det(A^T)\ .$$

Sketch of proof. Note that (5.1) implies that every term in $\det(A)$ is a product of n elements of the matrix in such a way that every row and every column contribute to this product. The same holds for all terms of $\det(A^T)$ so that $\det(A)$ and $\det(A^T)$ are sums of the same terms except maybe for the signs. That the signs are the same is less trivial and is part of the theory of permutations. □

5.2.12 Theorem. *All $n \times n$ matrices A and B satisfy*

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

Sketch of proof. We relate this property to the construction of a special multilinear function D that turns out to be the determinant up to a factor. To define $D(\underline{x}_1, \ldots, \underline{x}_n)$ we collect the vectors as rows in a matrix X and define

$$D(\underline{x}_1, \ldots, \underline{x}_n) = \det(XB).$$

Then it is not difficult to show (but it takes some writing) that D is multilinear and anti-symmetric. Since $D(\underline{e}_1, \ldots, \underline{e}_n) = \det(B)$ (and not necessarily 1), we conclude that D must be $\det(B)$ times the determinant. In other words, $D = \det(B)\det$. In particular, if we use the rows $\underline{a}_1, \ldots, \underline{a}_n$ of our matrix A, we get

$$\det(AB) = D(\underline{a}_1, \ldots, \underline{a}_n) = \det(B)\det(A).$$

$\square$

5.2.13 Practical rules for computing determinants

To actually compute a determinant more practical approaches are available. We will concentrate on two of them: one involves row (or column) operations, the other one concerns the so-called expansion of a determinant with respect to a row or column in order to reduce the computation of an $n \times n$ determinant to the computation of n determinants of size $(n-1)\times(n-1)$. In practice, the combination of these two techniques is quite effient.

5.2.14 Expansion with respect to a row or column

For a matrix A we denote by A_{ij} the matrix obtained from A by deleting the i-th row and the j-th column. We call A_{ij} a *submatrix* of A. For example, if

$$A = \begin{pmatrix} -1 & 0 & 2 \\ 2 & 1 & 3 \\ 1 & 3 & -1 \end{pmatrix},$$

then

$$A_{12} = \begin{pmatrix} 2 & 3 \\ 1 & -1 \end{pmatrix} \text{ en } A_{33} = \begin{pmatrix} -1 & 0 \\ 2 & 1 \end{pmatrix}.$$

Consider the $n \times n$-matrix

$$A = \begin{pmatrix} a_{11} & \ldots & a_{1n} \\ a_{21} & & \vdots \\ \vdots & & \vdots \\ a_{n1} & \ldots & a_{nn} \end{pmatrix}.$$

Every term of $\det(A)$ contains factors from the first column. First consider those terms containing the factor a_{11}. Then (5.1) shows that all other factors of such a term do not come from the first row or column. These other factors therefore come from the submatrix A_{11}. Next we consider the terms containing the factor a_{21}. The other factors in such a term must come from A_{21} for similar reasons, etc. A detailed inspection yields:

$$\det(A) = a_{11}\det(A_{11}) - a_{21}\det(A_{21}) + \ldots + (-1)^{n+1}a_{n1}\det(A_{n1}).$$

This formula is known as *expansion across (or along) the first column.*

There are similar formulas for the expansion across any row or column. Such expansions reduce the computation of a 'big' determinant to the computation of smaller determinants.

5.2.15 **Theorem. (Expansion across a row or column)** *A determinant can be computed by expansion across a row or column:*

- *Expansion across the i-th row:*

$$\det(A) = \sum_{j=1}^{n} (-1)^{i+j} a_{ij} \det(A_{ij}).$$

- *Expansion across the j-th column:*

$$\det(A) = \sum_{i=1}^{n} (-1)^{i+j} a_{ij} \det(A_{ij}).$$

5.2.16 **Example.** Such an expansion is especially useful if a row or column contains many zeros. The expansion across such a row or column then reduces the computation to only a few smaller determinants. Here is an example concerning a so-called (upper of lower) *triangular matrix*: a matrix whose entries above or below the main diagonal are 0's. Repeatedly expanding across the first column yields:

$$\begin{aligned}
\det(A) &= \begin{vmatrix} a_{11} & * & \dots & \dots & * \\ 0 & a_{22} & & & \vdots \\ \vdots & \ddots & \ddots & & \vdots \\ \vdots & \ddots & \ddots & a_{n-1,n-1} & * \\ 0 & \dots & \dots & 0 & a_{nn} \end{vmatrix} \\
&= a_{11} \begin{vmatrix} a_{22} & * & \dots & \dots & * \\ 0 & \ddots & & & \vdots \\ \vdots & \ddots & \ddots & & \vdots \\ \vdots & \ddots & \ddots & a_{n-1,n-1} & * \\ 0 & \dots & \dots & 0 & a_{nn} \end{vmatrix} \\
&= a_{11}a_{22} \begin{vmatrix} a_{33} & \dots & * \\ & \ddots & \vdots \\ 0 & & a_{nn} \end{vmatrix} = a_{11}a_{22}\cdots \begin{vmatrix} a_{n-1,n-1} & * \\ 0 & a_{nn} \end{vmatrix} \\
&= a_{11}a_{22}\dots a_{n-1,n-1}a_{nn} \, .
\end{aligned}$$

Of course, this result can also be obtained directly from (5.1); suppose all entries below the main diagonal are 0, then if a term contains a factor from above the diagonal, then there must also be a factor from below the diagonal which is 0. Such a term is therefore 0. The only term that remains is the product of the entries on the main diagonal.

5.2.17 Row operations and determinants

Row and column operations can be used to transform a matrix into one with 'many' 0's so that expansion across a suitable row or column is efficient. Row and column operations do influence the value of a determinant, so some bookkeeping is required. Let A be a square matrix $n \times n$. We first deal with row operations.

- The antisymmetry of the determinant function implies: interchanging two rows changes the determinant by a minus sign. So if A changes into B by interchanging two rows of A, then $\det(B) = -\det(A)$.

- The multilinearity implies: if one row of A is multiplied by λ to produce matrix B, then $\det(B) = \lambda \det(A)$, or:

$$\det(\underline{a}_1, \dots, \underline{a}_{i-1}, \lambda\underline{a}_i, \underline{a}_{i+1}, \dots, \underline{a}_n) = \lambda \det(\underline{a}_1, \dots, \underline{a}_i, \dots, \underline{a}_n).$$

- The multilinearity also implies: if a multiple of one row of A is added to another row of A, then the resulting matrix B has the same determinant as A, i.e., $\det(B) = \det(A)$.

The proofs are straightforward. By way of example, we prove the third property. Suppose we add $\lambda \underline{a}_j$ to $\underline{a}_i, i \neq j$. Then linearity implies:

$$\det(\ldots, \underline{a}_i + \lambda \underline{a}_j, \ldots, \underline{a}_j, \ldots)$$

$$= \det(\ldots, \underline{a}_i, \ldots, \underline{a}_j, \ldots) + \lambda \det(\ldots, \underline{a}_j, \ldots, \underline{a}_j, \ldots).$$

The first determinant on the right-hand side is $\det(A)$, while the second one is 0 since two of the rows are equal, see 5.2.6.

5.2.18 Column operations and determinants As $\det(A) = \det(A^T)$ similar properties hold if we replace rows by columns in 5.2.17. In particular, if we are only interested in the value of the determinant of A, then we can apply row operations and column operations in any order we like. However, if we use both types of operations, we do loose information on the row and column space of the matrix (but for the value of the determinant this is not relevant).

5.2.19 Example.

$$\det(A) = \begin{vmatrix} 0 & 1 & 2 & -1 \\ 2 & 5 & -7 & 3 \\ 0 & 3 & 6 & 2 \\ -2 & -5 & 4 & -2 \end{vmatrix} = \begin{vmatrix} 0 & 1 & 2 & -1 \\ 2 & 5 & -7 & 3 \\ 0 & 3 & 6 & 2 \\ 0 & 0 & -3 & 1 \end{vmatrix} =$$

$$-2 \begin{vmatrix} 1 & 2 & -1 \\ 3 & 6 & 2 \\ 0 & -3 & 1 \end{vmatrix} = -2 \begin{vmatrix} 1 & 2 & -1 \\ 0 & 0 & 5 \\ 0 & -3 & 1 \end{vmatrix} = 10 \begin{vmatrix} 1 & 2 \\ 0 & -3 \end{vmatrix} = -30.$$

Here is a description of the steps taken. First we add the 2-nd row to the 4-th. Then we expanded across the first column. Then we subtract the 1-st row 3 times from the 2-nd. Then we expand the 3×3 determinant across the 2-nd row. Finally, the 2×2 determinant is computed using 5.2.8.

5.2.20 Example.

$$\det(A) = \begin{vmatrix} 2 & 0 & 0 & 8 \\ 1 & -7 & -5 & 0 \\ 3 & 8 & 6 & 0 \\ 0 & 7 & 5 & 4 \end{vmatrix} = 2\begin{vmatrix} 1 & 0 & 0 & 4 \\ 1 & -7 & -5 & 0 \\ 3 & 8 & 6 & 0 \\ 0 & 7 & 5 & 4 \end{vmatrix}$$

$$= 8\begin{vmatrix} 1 & 0 & 0 & 1 \\ 1 & -7 & -5 & 0 \\ 3 & 8 & 6 & 0 \\ 0 & 7 & 5 & 1 \end{vmatrix} = 8\begin{vmatrix} 1 & 0 & 0 & 1 \\ 1 & -7 & -5 & 0 \\ 3 & 8 & 6 & 0 \\ -1 & 7 & 5 & 0 \end{vmatrix}$$

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

In the first two steps we 'extract' a factor 2 from the 1-st row and then a factor 4 from the last column. Then we expand across the last column. The result is a 3×3 determinant in which the 1-st and 3-rd row are multiples of one another, so that this determinant is 0.

The following theorem is related to Theorems 5.1.11 and 5.1.15, and to Corollary 5.1.12:

5.2.21 Theorem. *Let A be an $n \times n$-matrix and let $\underline{b}$ be a vector in $\mathbb{R}^n$ or $\mathbb{C}^n$.*

1. *rank* $(A) = n \Leftrightarrow \det(A) \neq 0$.
2. *A is invertible if and only if* $\det(A) \neq 0$.
3. *The system of linear equations $A\underline{x} = \underline{b}$ has exactly one solution if and only if* $\det(A) \neq 0$.

Proof.

1. From 5.2.17 we first conclude that row and column operations may change the value of the determinant, but not the property of 'being 0': if $\det(A) \neq 0$, then any row or column operation produces a matrix whose determinant is $\neq 0$; and if $\det(A) = 0$ then any row or column operation produces a matrix whose determinant is 0. If rank $(A) = n$, then A can be transformed into the identity matrix I whose determinant is 1, so that, by the previous considerations, $\det(A) \neq 0$; if rank $(A) < n$, then one of the rows must be a linear combination of the other rows and therefore $\det(A) = 0$ by 5.2.6.

2. Follows from (1) and Theorem 5.1.15.

3. Follows from (1) and Corollary 5.1.12.

□

5.2.22 Cramer's rule

There is a remarkable role for determinants in solving systems of n linear equations in n variables. Let A be the coefficient matrix of such a system $A\underline{x} = \underline{b}$, or explicitly:

$$\begin{array}{ll} a_{11}x_1 + a_{12}x_2 & + \ldots + a_{1n}x_n = b_1 \\ a_{21}x_1 + a_{22}x_2 & + \ldots + a_{2n}x_n = b_2 \\ & \vdots \\ a_{n1}x_1 + a_{n2}x_2 & + \ldots + a_{nn}x_n = b_n \ . \end{array} \tag{5.2}$$

This system of linear equations has exactly one solution if rank $(A) = n$, so if $\det(A) \neq 0$. The remarkable thing is that using determinants we can write down an explicit expression for the solution. Here is how such an expression if derived.

Let $\underline{k}_1, \ldots, \underline{k}_n$ be the columns of A, and let $x_1, \ldots, x_n$ be the (coordinates of the) solution vector. Then

$$\underline{b} = x_1\underline{k}_1 + x_2\underline{k}_2 + \ldots + x_n\underline{k}_n \ .$$

Now replace the j-th column of A by $\underline{b}$ an denote the resulting matrix by $A_j(\underline{b})$. Then

$$\begin{aligned} \det(A_j(\underline{b})) &= \det(\underline{k}_1, \ldots, \underline{k}_{j-1}, \sum_{i=1}^{n} x_i\, \underline{k}_i, \underline{k}_{j+1}, \ldots, \underline{k}_n) \\ &= \sum_{i=1}^{n} x_i \det(\underline{k}_1, \ldots, \underline{k}_{j-1}, \underline{k}_i, \underline{k}_{j+1}, \ldots, \underline{k}_n) \\ &= x_j \det(\underline{k}_1, \ldots, \underline{k}_n) = x_j \det(A), \end{aligned}$$

because in the sum all determinants with $i \neq j$ are 0 since they contain two equal vectors. The solution of the system is therfore:

$$x_j = \frac{\det(A_j(\underline{b}))}{\det(A)} \ , \ j = 1, \ldots, n.$$

This is called *Cramer's rule*. This rule is of importance in theoretical considerations on determinants, but has limited practical value. Already for $n \geq 3$, the rule is quite impractical for actually solving a system.

For $n = 2$, Cramer's rule applied to the system

$$\begin{aligned} ax + by &= c \\ dx + ey &= f \end{aligned}$$

leads to

$$x = \frac{\begin{vmatrix} c & b \\ f & e \end{vmatrix}}{\begin{vmatrix} a & b \\ d & e \end{vmatrix}} = \frac{ce - bf}{ae - bd}, \quad y = \frac{\begin{vmatrix} a & c \\ d & f \end{vmatrix}}{\begin{vmatrix} a & b \\ d & e \end{vmatrix}} = \frac{af - cd}{ae - bd},$$

whenever $ae - bd \neq 0$.

5.2.23 Cramer's rule can also be used to derive an explicit formula for the inverse of an invertible square matrix. Again, for $n \geq 3$ this formula is mainly of theoretical importance. Using row operations to find the inverse is much more efficient in practice.

To derive this formula in the case of a 2×2–matrix, we have to solve two systems of linear equations:

$$\begin{aligned} ax_{11} + bx_{21} &= 1 \\ cx_{11} + dx_{21} &= 0 \end{aligned} \quad \text{en} \quad \begin{aligned} ax_{12} + bx_{22} &= 0 \\ cx_{12} + dx_{22} &= 1. \end{aligned}$$

Cramer's rule then produces the following result:

$$A^{-1} = \frac{1}{ad - bc} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}.$$

5.2.24 The 2×2 determinant can be interpreted as a surface area as we saw in the beginning of the chapter. Similarly, $n \times n$ determinants have geometric interpretations: they 'measure' volumes of parallelepipeds spanned by the rows (or columns) of the matrix in in $\mathbb{R}^n$.

5.3 Notes

Determinants were popular among 19th century mathematicians. The name *determinant* was coined by the French mathematician Augustin-Louis Cauchy (1789–1846). All sorts of determinantal identities were derived. Cramer's rule goes back to Gabriel Cramer (1704–1752), even though Cramer himself did not give a proof of the rule. Nowadays, the importance of determinants in various branches of mathematics (and fields where mathematics is applied) is very clear. In analysis *Analysis 2, 3* determinants show up in the substitution rule for multiple integrals. For instance, when introducing new variables in a double integral a 2×2 determinant appears in the transformed integral. When using polar coordinates $x = r\cos\phi$, $y = r\sin\phi$ this looks as follows:

$$\begin{aligned} \int\int f(x,y)\,dx\,dy &= \int\int f(r\cos\phi, r\sin\phi) \begin{vmatrix} \cos\phi & \sin\phi \\ -r\sin\phi & r\cos\phi \end{vmatrix} dr\,d\phi \\ &= \int\int f(r\cos\phi, r\sin\phi) r\,dr\,d\phi. \end{aligned}$$

The proof that determinant functions exist in all dimensions requires knowledge of permutations beyond the scope of these lecture notes. An elegant construction of a determinant function is the following: let $\phi_1, \ldots, \phi_n$ be the dual basis of a basis of $\mathbb{R}^n$ (dual bases occur in Linear Algebra 2) and define

$$\det(\underline{a}_1, \ldots, \underline{a}_n) = \sum_{\sigma \in S_n} \operatorname{sgn}(\sigma) \phi_{\sigma(1)}(\underline{a}_1) \cdots \phi_{\sigma(n)}(\underline{a}_n),$$

where S_n is the set of all $n!$ permutations of $\{1, \ldots, n\}$. Permutations are discussed more extensively in the courses on algebra. They are also useful in describing symmetries like the 48 symmetries of the cube. The proof that there are no general *Algebra* formulas for finding the roots of polynomials of degrees ≥ 5 also uses permutations.

Determinant functions are special cases of multilinear functions, which play an important role in the theory of line, surface and volume integrals (theorems of Gauss, Green, Stokes) and in differential geometry.

5.4 Exercises

§1

1 Determine a basis for the row space and a basis for the column space for each of the following matrices:

a. $\begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 2 & 3 & 4 \end{pmatrix}$,

b. $\begin{pmatrix} 1 & 1 & 0 & 1 \\ -1 & 2 & 1 & 1 \\ -1 & 8 & 3 & 5 \end{pmatrix}$,

c. $\begin{pmatrix} 1 & i & 1+i \\ 1+i & 1 & 2+i \\ 2+i & -1 & 1+i \end{pmatrix}$.

2 Determine the rank of each of the following matrices:

a. $\begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}$,

b. $\begin{pmatrix} 1 & -2 & 1 \\ -1 & 3 & -2 \\ -1 & 1 & 0 \end{pmatrix}$,

c. $\begin{pmatrix} -4 & 1 & 0 & 1 \\ -2 & 0 & 2 & 1 \\ 0 & 2 & -3 & 0 \\ -7 & 2 & 0 & 2 \end{pmatrix}$,

d. $\begin{pmatrix} 1 & 2 & 3 \\ 2 & 4 & 5 \\ 3 & 5 & 6 \end{pmatrix}$.

3 Consider the matrices in the previous exercise. Deetermine the inverse of each of the matrices whenever the inverse exists. Check your answers by using $AA^{-1} = A^{-1}A = I$.

4 Determine the inverse of each of the following matrices:

a. $\begin{pmatrix} i & 1 \\ 1 & i \end{pmatrix}$,

b. $\begin{pmatrix} 1 & 0 & 1 \\ 1 & 0 & -1 \\ 0 & i & 0 \end{pmatrix}$.

5 a. Let A be an $n \times m$–matrix and let B be an $n \times 1$–matrix. Prove that there exists an $m \times 1$–matrix X with $AX = B$ if and only if B belongs to the column space of A.

b. (Notation as in a.) What is the relation between $\text{rank}(A)$ and $\text{rank}(A|B)$ if the system $AX = B$ has a solution?

c. Let A and B be two matrices such that the product AB exists. Show that the column space of AB is contained in the column space of A, and that the row space of AB is contained in the row space of B.

6 a. For each $\lambda \in \mathbb{C}$ determine the rank of the matrix

$$\begin{pmatrix} \lambda & 1 \\ 1 & \lambda^3 \end{pmatrix}.$$

b. For each $\lambda \in \mathbb{R}$ determine the rank of the matrix

$$\begin{pmatrix} 2 & 3 & 2+\lambda \\ 1+\lambda & 4-\lambda & 3 \end{pmatrix}.$$

§2

7 Determine the following determinants:

a. $\begin{vmatrix} i & 1 \\ -1 & i \end{vmatrix}$,

b. $\begin{vmatrix} 0 & 0 & \pi \\ 0 & e & 1 \\ i & \sqrt{2} & \frac{1}{2} \end{vmatrix}$,

c. $\begin{vmatrix} 7 & -3 & 8 \\ 9 & 11 & 5 \\ 8 & 8 & 3 \end{vmatrix}$,

d. $\begin{vmatrix} 4 & 4 & 8 \\ 0 & 1 & 2 \\ 0 & 0 & 2 \end{vmatrix}$,

e. $\begin{vmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \end{vmatrix}$,

f. $\begin{vmatrix} 3 & -4 & 1 \\ -2 & 3 & -1 \\ 1 & 1 & 2 \end{vmatrix}$,

g. $\begin{vmatrix} 7 & 9 & 4 \\ 6 & 5 & 4 \\ 2 & 5 & 3 \end{vmatrix}$,

h. $\begin{vmatrix} 0 & 0 & 2 \\ 0 & 2 & 6 \\ 1 & 2 & 2 \end{vmatrix}$.

8 Determine the following determinants:

a. $\begin{vmatrix} 4 & 6 & 2 & 5 \\ 6 & 9 & 3 & 7 \\ 5 & 8 & 6 & 2 \\ 7 & 10 & 8 & 3 \end{vmatrix}$,

b. $\begin{vmatrix} 11 & 3 & -4 & 5 \\ -13 & -4 & 5 & -6 \\ 18 & 5 & -5 & 7 \\ 8 & 2 & -2 & 3 \end{vmatrix}$,

c. $\begin{vmatrix} 1 & 0 & 0 & 1 \\ 0 & \pi & 0 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 \end{vmatrix}$,

d. $\begin{vmatrix} 1 & 1 & 0 & 0 \\ 1 & 2 & 1 & 0 \\ 1 & 3 & 3 & 1 \\ 1 & 4 & 6 & 4 \end{vmatrix}$,

e. $\begin{vmatrix} 1 & 3 & -1 & 2 \\ -1 & 1 & 0 & -1 \\ 1 & 2 & 1 & -2 \\ 0 & 3 & 4 & 1 \end{vmatrix}$,

f. $\begin{vmatrix} 1 & 2 & 3 & -3 & 6 \\ 3 & 6 & 9 & 2 & 7 \\ 5 & 3 & 2 & 3 & -1 \\ 7 & 2 & 5 & 4 & 1 \\ 2 & -1 & 3 & 1 & 1 \end{vmatrix}$.

9 Let $A = \begin{pmatrix} 1 \\ 4 \\ 2 \end{pmatrix}$ and $B = \begin{pmatrix} 2 & -1 & 2 \end{pmatrix}$. Calculate $\det(AB)$ and $\det(BA)$.

10 Consider the $n \times n$-matrix A given by $a_{kl} = k + l$. Calculate $\det(A)$.

11 Check if the matrix

$$A = \begin{pmatrix} 1 & 1 & i \\ 1 & i & 1 \\ i & 1 & 1 \end{pmatrix}$$

has an inverse.

12 a. Show that the system of linear equations

$$\begin{array}{rrrcr} 3x_1 & -4x_2 & +x_3 & = & 4, \\ -2x_1 & +3x_2 & -x_3 & = & -3, \\ x_1 & +x_2 & +2x_3 & = & 3. \end{array}$$

has exactly one solution.

b. Show that the system of equations

$$\begin{array}{rrrcr} x_1 & +x_2 & +x_3 & = & 2, \\ 2x_1 & -x_2 & -3x_3 & = & -1, \\ & 3x_2 & +5x_3 & = & 5. \end{array}$$

has more than one solution.

c. Show that thes system of linear equations

$$\begin{array}{ccccc} x_1 & +ix_2 & +2x_3 & = & i, \\ (1+i)x_1 & +x_2 & +x_3 & = & 1, \\ (-1+2i)x_1 & +(-1+i)x_2 & +3ix_3 & = & 0. \end{array}$$

has no solutions.

13 a. Use Cramer's rule to solve for y in the system of linear equations

$$\begin{array}{rrrcr} 2x & +3y & -2z & = & -1, \\ 3x & +y & +5z & = & 11, \\ x & +4y & -3z & = & -2. \end{array}$$

b. Use Cramer's rule to solve for x in the system of linear equations

$$\begin{array}{rrrcl} 3x & +4y & +2z & = & 6, \\ 4x & +6y & +3z & = & 6, \\ 2x & +3y & +z & = & 1. \end{array}$$

14 For which α, β, γ are the vectors (α, β, γ), $(\alpha, 2\beta, 2\gamma)$, $(2\alpha, 2\beta, \gamma)$ in $\mathbb{R}^3$ linearly dependent?

15 a. Show that any $n \times n$-matrix A and any scalar α the relation $\det(\alpha A) = \alpha^n \det(A)$ holds.

b. Suppose the square matrix A satisfied $A^{-1} = A^\top$. What conclusion can you draw on $\det(A)$?

c. Of the $n \times n$-matrix A it is given that $A^\top = -A$. What follows from this for $\det(A)$?

d. If A is an $n \times n$-matrix and S is an $n \times n$-matrix for which S^{-1} exists, then $\det(S^{-1}AS) = \det(A)$. Prove this.

16 a. Give an example of a square matrix A, different from the zero matrix, for which A^2 equals the zero matrix.

b. Suppose A is a square matrix with $A^2 = 0$ (zero matrix). Determine $\det(A)$?

c. Give an example of the square matrix A, different from the zero matrix, for which $A^2 = A$.

d. What are the possibilities for $\det(A)$ if $A^2 = A$?

17 a. Show that

$$\begin{vmatrix} 1 & 1 & 1 \\ a & b & c \\ a^2 & b^2 & c^2 \end{vmatrix} = (b-a)(c-a)(c-b).$$

b. Can you generalize this to $n \times n$–determinants? (These so-called Vandermonde-determinants are named after Alexandre–Theophile Vandermonde (1735–1796).)

5.4.1 Exercises from old exams

18 For which value(s) of a does the matrix

$$A = \begin{pmatrix} 1 & 1 & a \\ 0 & a & 1 \\ 1 & 1 & 3 \end{pmatrix}$$

have an inverse? Determine the inverse of A for these value(s).

19 Determine all the values of a for which the vectors $(a^2+1, a, 1)$, $(2, 1, a)$, $(1, 0, 1)$ in $\mathbb{R}^3$ are linearly dependent.

20 a. For each $\lambda \in \mathbb{R}$ the following matrix is given:

$$A_\lambda = \begin{pmatrix} 1 & \lambda & 0 \\ \lambda & 1 & \lambda^2 - 1 \\ 0 & 2 & -1 \end{pmatrix}.$$

Determine the inverse of A_λ for each value of λ for which A_λ is invertible.

b. For which value(s) of λ does the system

$$\begin{array}{rrrl} x_1 + & \lambda x_2 & & = \lambda^2 \\ \lambda x_1 + & x_2 & +(\lambda^2-1)x_3 & = \lambda^2 \\ & 2x_2 & -x_3 & = 0 \end{array}$$

have no solutions?

Chapter 6

Inner product spaces

6.1 Inner product, length and angle

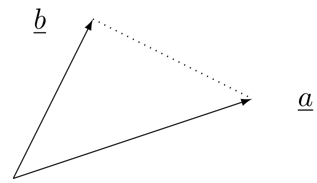
6.1.1 In Chapter 4 we defined the notion of a vector space. The elements of such a vector space are called vectors and 'behave' as far as addition and scalar multiplication is concerned as the vectors in the plane or in space.

Now vectors in the plane and in space have two more characteristics: such a vector has a length and two such (nonzero) vectors determine an angle. In this chapter we will generalize these notions to general vector spaces. It may not come as a surprise that actually the notion of inner product will be central in our discussion.

In this section we

- introduce inner products and inner product spaces;
- define the notions length, distance and angle (in particular being perpendicular) in real vector spaces;
- discuss the Cauchy-Schwarz inequality (needed for instance to introduce the notion of angle)
- discuss the Pythagorean theorem.

6.1.2 To give as to where the notion of an inner product comes from, we take a look at the plane.



In a triangle in which two sides correspond to the vectors $\underline{a}$ and $\underline{b}$, the third side has length $\| \underline{a} - \underline{b} \|$. If φ is the angle between the vectors $\underline{a}$ and $\underline{b}$, then the cosine rule tells us that

$$\| \underline{a} - \underline{b} \|^2 = \| \underline{a} \|^2 + \| \underline{b} \|^2 - 2 \; \| \underline{a} \| \cdot \| \underline{b} \| \cos \varphi.$$

In this equality the product $\| \underline{a} \| \cdot \| \underline{b} \| \cos \varphi$ on the right-hand side is of importance to us; it is the inner product of the vectors $\underline{a}$ and $\underline{b}$ in the plane. By choosing $\underline{a} = \underline{b}$ we recover the square of the length of $\underline{a}$. The angle between two (nonzero) vectors is also present in this expression. Now in abstract vector spaces we can't start with such an explicit formula containing undefined factors. The way out is to generalize properties of the inner product: symmetry and (bi)linearity as discussed in Chapter 2.

6.1.3 Definition. (Inner product) Let V be a real vector space. An *inner product*[1] on V is a function which assigns to any two vectors $\underline{a}, \underline{b}$ from V a real number denoted by $(\underline{a}, \underline{b})$ in such a way that

1. $(\underline{a}, \underline{b})$ is linear in both entries:

$$\begin{aligned} (\lambda \underline{v} + \mu \underline{w}, \underline{b}) &= \lambda(\underline{v}, \underline{b}) + \mu(\underline{w}, \underline{b}) \\ (\underline{a}, \lambda \underline{v} + \mu \underline{w}) &= \lambda(\underline{a}, \underline{v}) + \mu(\underline{a}, \underline{w}) \end{aligned}$$

   for all scalars and all vectors.

2. $(\underline{a}, \underline{b}) = (\underline{b}, \underline{a})$ for all $\underline{a}, \underline{b} \in V$. (So we need only impose linearity in the first entry in the previous item, because symmetry will then automatically imply linearity in the second entry.)

3. $(\underline{a}, \underline{a}) \geq 0$ for all $\underline{a} \in V$, and $(\underline{a}, \underline{a}) = 0$ implies $\underline{a} = \underline{0}$.

A real vector space with an inner product is often called a real *inner product space*.

[1] An inner product is sometimes called a 'dot product' with notation $\underline{a} \cdot \underline{b}$.

6.1.4 **Definition.** Let V be a complex vector space. An *inner product* on V is a function that assigns to every two vectors $\underline{a}, \underline{b}$ in V a complex number, denoted by $(\underline{a}, \underline{b})$, in such a way that

1. $(\underline{a}, \underline{b})$ is linear in $\underline{a}$,
2. $(\underline{a}, \underline{b}) = \overline{(\underline{b}, \underline{a})}$ for all $\underline{a}, \underline{b} \in V$,
3. $(\underline{a}, \underline{a}) \geq 0$ for all $\underline{a} \in V$, and $(\underline{a}, \underline{a}) = 0$ implies $\underline{a} = \underline{0}$.

A complex vector space with an inner product is often called a (complex) *inner product space*.

6.1.5 An inner product on a real vector space takes real values, an inner product on a complex vector space assumes complex values. Although there are many similarities between real and complex inner products one significant difference concerns the second entry: a complex inner product is not linear in the second entry.

$$\begin{aligned}\left(\underline{a}, \sum_{i=1}^{n} \beta_i \underline{b}_i\right) &= \overline{\left(\sum_{i=1}^{n} \beta_i \underline{b}_i, \underline{a}\right)} = \overline{\sum_{i=1}^{n} \beta_i \ (\underline{b}_i, \underline{a})} = \sum_{i=1}^{n} \overline{\beta}_i \overline{(\underline{b}_i, \underline{a})} \\ &= \sum_{i=1}^{n} \overline{\beta_i} \ (\underline{a}, \underline{b}_i) \ .\end{aligned}$$

6.1.6 **Example. (Standard inner product)** There are many ways to define an inner product on $\mathbb{R}^n$ or $\mathbb{C}^n$, but there is one which deserves a special name, the so-called *standard inner product*: if $\underline{a} = (a_1, a_2, \ldots, a_n)$ en $\underline{b} = (b_1, b_2, \ldots, b_n)$ then

$$(\underline{a}, \underline{b}) = a_1\overline{b_1} + a_2\overline{b_2} + \cdots + a_n\overline{b_n};$$

in $\mathbb{R}^n$ this definition reduces to

$$(\underline{a}, \underline{b}) = a_1 b_1 + a_2 b_2 + \cdots + a_n b_n.$$

Verify that these are really inner products using 6.1.3 and 6.1.4. The standard inner product generalizes the usual inner product in $\mathbb{R}^2$, but it turns out to play a special role among inner products as we will see later.

6.1.7 **Example.** Let V be the set of continuous *real* valued functions on the interval $[a, b]$ (with $b > a$). Then pointwise addition and scalar multiplication turn V into a vector space. For $f, g \in V$ we define

$$(f, g) = \int_a^b f(x)\, g(x) dx.$$

It's easy to verify that this defines an inner product on V, except maybe the second part of property 3 (the remaining verifications are left to the reader). To prove the second part of property 3 we proceed as follows. Take any $f \in V$ and suppose $f(\alpha) \neq 0$ for some $\alpha \in [a, b]$. Then continuity of f implies that there is an interval around α so that $|f(x)| > \frac{1}{2}|f(\alpha)| > 0$ for all x in that interval (if you are familiar with the ε-δ definition of continuity: use $\varepsilon = \frac{1}{2}|f(\alpha)|$). Let δ be the length of the interval. Then:

$$(f, f) \ = \ \int_a^b |f(x)|^2 dx \geq \frac{1}{4}\delta|f(\alpha)|^2 > 0.$$

So if $(f, f) \ = \ 0$, then $f(x)$ must equal 0 for all $x \in [a, b]$.

6.1.8 Consider a real or complex inner product space V. Since $(\underline{a}, \underline{a})$ is a nonnegative real number for every vector $\underline{a}$, the square root $\sqrt{(\underline{a}, \underline{a})}$ is a real number. We use this to define the notion of length.

6.1.9 Definition. (Length and distance) In an inner product space the *length* or *norm* $\| \underline{a} \|$ of a vector $\underline{a}$ is defined as

$$\| \underline{a} \| = \sqrt{(\underline{a}, \underline{a})}.$$

The *distance* between the vectors $\underline{a}$ and $\underline{b}$ is by definition the length of the vector $\underline{a} - \underline{b}$, i.e., $\| \underline{a} - \underline{b} \|$.

6.1.10 Examples. Using the standard inner product, the length of $(1, 1, 1, 1) \in \mathbb{R}^4$ equals

$$\sqrt{1^2 + 1^2 + 1^2 + 1^2} = 2.$$

The distance between $(2, 1, 3, 4)$ and $(5, 1, 7, 4)$ is

$$\sqrt{(2-5)^2 + (1-1)^2 + (3-7)^2 + (4-4)^2} = \sqrt{9 + 16} = 5.$$

In the inner product space V from example 6.1.7 (where we take $a = 0$ and $b = 1$) the length of the function/vector $x \mapsto x^2$ is

$$\sqrt{\int_0^1 x^2 \cdot x^2 \, dx} = \frac{1}{5}\sqrt{5}.$$

6.1.11 The introduction of the angle between vectors requires some thought plus a famous theorem, the Cauchy-Schwarz inequality. This inequality is also of importance in other branches of mathematics.

6.1.12 **Theorem. (The Cauchy-Schwarz inequality)** *For all $\underline{a}, \underline{b}$ in the inner product space V the following inequality holds:*

$$|(\underline{a}, \underline{b})| \ \leq \| \ \underline{a} \ \| \ \ \| \ \underline{b} \ \| \ .$$

Proof. First we deal with the case that $\underline{a} \ = \ \underline{0}$. If $\underline{a} \ = \ \underline{0}$, then the linearity of the inner product implies: $(\underline{a}, \underline{b}) \ = \ (0\underline{a}, \underline{b}) \ = \ 0(\underline{a}, \underline{b}) \ = \ 0$ for all $\underline{b}$, so in particular $(\underline{a}, \underline{a}) \ = \ 0$ and $\| \ \underline{a} \ \| = \ 0$. In this the inequality is even an equality.
From now on we assume $\underline{a} \neq \underline{0}$. Choose $\underline{b} \in V$ and let φ be the argument of $(\underline{a}, \underline{b})$. Define $\underline{a}^* \ = \ e^{-i\varphi}\underline{a}$. Then

$$(\underline{a}^*, \underline{b}) \ = \ e^{-i\varphi}(\underline{a}, \underline{b}) \in \mathbb{R} \ .$$

Take $\lambda \in \mathbb{R}$ and consider

$$f(\lambda) \ = \ (\lambda\underline{a}^* + \underline{b}, \lambda\underline{a}^* + \underline{b}) \ .$$

Then $f(\lambda) \geq 0$ for every $\lambda \in \mathbb{R}$ because of definitions 6.1.3 and 6.1.4. Since $f(\lambda)$ is a quadratic function in λ,

$$\begin{aligned} f(\lambda) \ &= \ (\lambda\underline{a}^* + \underline{b}, \lambda\underline{a}^* + \underline{b}) \\ &= \ \lambda^2(\underline{a}^*, \underline{a}^*) + \lambda(\underline{a}^*, \underline{b}) + \lambda(\underline{b}, \underline{a}^*) + (\underline{b}, \underline{b}) \\ &= \| \ \underline{a}^* \ \|^2 \ \lambda^2 + 2\lambda(\underline{a}^*, \underline{b}) + \| \ \underline{b} \ \|^2 \ \ (\text{want } (\underline{a}^*, \underline{b}) \in \mathbb{R}), \end{aligned}$$

its discriminant must be less than or equal to 0, i.e.,

$$\begin{aligned} 4 \ \ &(\underline{a}^*, \underline{b})^2 - 4 \| \ \underline{a}^* \ \|^2 \| \ \underline{b} \ \|^2 \leq \ 0 \ , \\ &(\underline{a}^*, \underline{b})^2 \ \leq \| \ \underline{a}^* \ \|^2 \| \ \underline{b} \ \|^2 \ , \\ &|(\underline{a}^*, \underline{b})| \ \leq \| \ \underline{a}^* \ \| \ \ \| \ \underline{b} \ \| \ \ . \end{aligned}$$

This implies

$$|e^{i\varphi}(\underline{a}, \underline{b})| \ = \ |(\underline{a}, \underline{b})| \ \leq \| \ e^{i\varphi}\underline{a} \ \| \ \ \| \ \underline{b} \ \| = \| \ \underline{a} \ \| \ \ \| \ \underline{b} \ \| \ \ .$$

□

6.1.13 The Cauchy–Schwarz inequality is also known as the Cauchy–Schwarz–Bunyakovsky inequality. As for the proof, it is a good exercise to redo the proof in the real case (it simplifies considerably).

6.1.14 **Examples.** Consider the inner product spaces from example 6.1.6. The Cauchy–Schwarz inequality then implies the following formula in $\mathbb{C}^n$:

$$\left| \sum_{i=1}^{n} \ a_i b_i \right|^2 \ \leq \ \left(\sum_{i=1}^{n} |a_i|^2 \right) \ \left(\sum_{i=1}^{n} |b_i|^2 \right)$$

for every pair $(a_1, \ldots, a_n)$, $(b_1, \ldots, b_n)$ in $\mathbb{C}^n$. For example, for $\underline{a} = (1, 1, 2)$ and for all (x, y, z) in $\mathbb{R}^3$ we get:

$$(x + y + 2z)^2 \leq (1^2 + 1^2 + 2^2)(x^2 + y^2 + z^2) = 6(x^2 + y^2 + z^2).$$

In the vector space V from example 6.1.7 the Cauchy–Schwarz inequality leads to the following inequality for every pair of continuous functions f and g on $[a, b]$:

$$\left| \int_a^b f(x)\overline{g(x)} dx \right|^2 \;\leq\; \int_a^b |f(x)|^2 \, dx \int_a^b |g(x)|^2 \, dx \;.$$

Inequalities like this one are used to give estimates on, e.g., integrals. For example, for $x \mapsto e^{x^2}$ and $x \mapsto e^{-x^2}$, considered on the interval $[0, 1]$, we have:

$$1 = \left| \int_0^1 e^{x^2} \, e^{-x^2} \, dx \right|^2 \leq \int_0^1 e^{2x^2} \, dx \int_0^1 e^{-2x^2} \, dx.$$

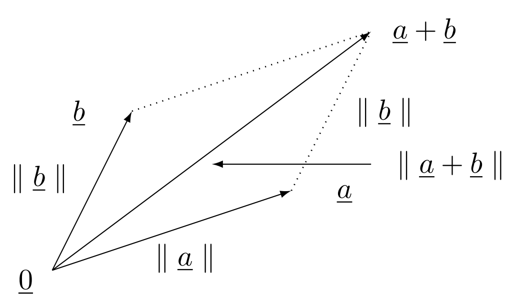
6.1.15 Theorem. *Let V be an inner product space. Then for all vectors $\underline{a}$ and $\underline{b}$ in V and all scalars λ we have:*

1. $\| \underline{a} \| \geq 0$, *and* $\| \underline{a} \| = 0$ *if and only if* $\underline{a} = \underline{0}$.

2. **The triangle inequality***:*

$$\| \underline{a} + \underline{b} \| \leq \| \underline{a} \| + \| \underline{b} \|,$$

a generalization of the famous inequality in the plane or in space:



3. $\| \lambda \underline{a} \| = |\lambda| \; \| \underline{a} \|$.

Proof. The first part follows directly from the third condition on inner products. The third part follows from the linearity of the inner product and $(\underline{a}, \underline{b}) = \overline{(\underline{b}, \underline{a})}$:

$$(\lambda \underline{a}, \lambda \underline{a}) \;=\; \lambda(\underline{a}, \lambda \underline{a}) \;=\; \lambda\bar{\lambda}(\underline{a}, \underline{a}) \;=\; |\lambda|^2 \; (\underline{a}, \underline{a}),$$

so

$$\| \lambda \underline{a} \| = |\lambda| \ \| \underline{a} \| .$$

For property 2, the triangle inequality, we need the Cauchy–Schwarz inequality:

$$\begin{aligned} \| \underline{a} + \underline{b} \|^2 &= (\underline{a} + \underline{b}, \underline{a} + \underline{b}) = (\underline{a}, \underline{a}) + (\underline{b}, \underline{b}) + (\underline{a}, \underline{b}) + (\underline{b}, \underline{a}) \\ &\leq \| \underline{a} \|^2 + \| \underline{b} \|^2 + 2|(a, b)| \\ &\leq \| \underline{a} \|^2 + \| \underline{b} \|^2 + 2 \| \underline{a} \| \ \| \underline{b} \| = (\| \underline{a} \| + \| \underline{b} \|)^2 . \end{aligned}$$

Upon taking square roots on both sides, we find the desired inequality. □

6.1.16 The triangle inequality provides an upper bound on the length of the sum of two vectors. A lower bound can also be extracted from this inequality as follows. Replace $\underline{a}$ by $\underline{a} - \underline{b}$ in the triangle inequality (we can apply the inequality to any two vectors!). We then find (for all $\underline{a}$ and $\underline{b}$):

$$\| (\underline{a} - \underline{b}) + \underline{b} \| \leq \| \underline{a} - \underline{b} \| + \| \underline{b} \| ,$$

$$\| \underline{a} - \underline{b} \| \geq \| \underline{a} \| - \| \underline{b} \| .$$

This inequality is also valid if we now replace $\underline{b}$ by $-\underline{b}$:

$$\| \underline{a} + \underline{b} \| \geq \| \underline{a} \| - \| -\underline{b} \| = \| \underline{a} \| - \| \underline{b} \|$$

(using the third property of an inner product). Interchanging $\underline{a}$ and $\underline{b}$ lproduces

$$\| \underline{b} + \underline{a} \| \geq \| \underline{b} \| - \| \underline{a} \| ,$$

so that finally (for all $\underline{a}$ and $\underline{b}$):

$$\| \underline{a} + \underline{b} \| \geq | \ \| \underline{a} \| - \| \underline{b} \| \ | .$$

6.1.17 Normed vector space

A vector space with the notion of length satisfying the properties in the theorem is called a *normed vector space*. They are of importance in functional analysis where spaces of functions paly a special role. They are beyond the scope of this course.

6.1.18 Angle

The Cauchy–Schwarz inequality enables us to *define* the notion of *angle* between two (nonzero) vectors in a real vector space. Here are the details. Let $\underline{a} \neq \underline{0}, \underline{b} \neq \underline{0}$. Then the Cauchy–Schwarz inequality implies

$$-1 \leq \frac{(\underline{a}, \underline{b})}{\| \underline{a} \| \ \| \underline{b} \|} \leq 1 .$$

So there is a real number φ satisfying

$$\frac{(a,b)}{\| \underline{a} \| \ \| \underline{b} \|} = \cos\varphi$$

or

$$(\underline{a},\underline{b}) = \| \underline{a} \| \ \| \underline{b} \| \cos\varphi. \tag{6.1}$$

We call this number (usually taken in the interval $(-\pi,\pi]$) the *angle* between the two vectors. In practice, you usually first compute the cosine of the angle via (6.1).

The notion of angle in a complex vector space is more complicated. It is beyond the scope of this course. We do define perpendicularity.

6.1.19 Definition. (Perpendicular vectors) Let V be an inner product space. The vectors $\underline{a} \in V$ and $\underline{b} \in V$ are *perpendicular*, denoted by $\underline{a} \perp \underline{b}$, if $(\underline{a},\underline{b}) = 0$. (In this definition $\underline{a}$ and/or $\underline{b}$ are allowed to be the zero vector; also note that $(\underline{a},\underline{b}) = 0 \Leftrightarrow (\underline{b},\underline{a}) = 0$.)

6.1.20 Example. In $\mathbb{R}^n$ or $\mathbb{C}^n$ (with the standard inner product) the vectors $\underline{e}_i$ and $\underline{e}_j$ are perpendicular for every i and j with $i \neq j$. Moreover, every $\underline{e}_i$ has length 1.

6.1.21 Example. Let V be the inner product space of continuous functions on $[0,2\pi]$ (see example 6.1.7) and consider for all $n \in \mathbb{Z}$ the function

$$e_n = e^{inx} .$$

If $n \neq m$, then $e_n \perp e_m$, because

$$\begin{aligned}(e_n,e_m) &= \int_0^{2\pi} e^{inx}\overline{e^{imx}}dx = \int_0^{2\pi} e^{i(n-m)x}dx \\ &= \tfrac{1}{i(n-m)} \left(e^{i(n-m)2\pi} - e^0\right) = 0 .\end{aligned}$$

Also,

$$\begin{aligned}\| e_n \|^2 &= (e_n,e_n) = \int_0^{2\pi} e^{inx}\overline{e^{inx}}dx \\ &= \int_0^{2\pi} \left|e^{inx}\right|^2 dx = \int_0^{2\pi} dx = 2\pi ,\end{aligned}$$

and therefore

$$\| e_n \| = \sqrt{2\pi} \text{ voor alle } n \in \mathbb{Z} .$$

This inner product space of functions plays an important role in Fourier analysis, a field with applications in for instance signal analysis.

6.1.22 Theorem. (Pythagorean theorem) *Let V be an inner product space.*

1. *If the vectors $\underline{a}$ and $\underline{b}$ are perpendicular, then*
$$\| \underline{a} + \underline{b} \|^2 = \| \underline{a} \|^2 + \| \underline{b} \|^2 .$$

2. *If the vectors $\underline{a}_1 \ldots, \underline{a}_k$ are mutually perpendicular, i.e., $(\underline{a}_i, \underline{a}_j) = 0$ if $i \neq j$, then*
$$\| \underline{a}_1 + \cdots + \underline{a}_k \|^2 = \| \underline{a}_1 \|^2 + \cdots + \| \underline{a}_k \|^2 .$$

Proof. To prove the first part, we use the properties of the inner product to expand $\| \underline{a} + \underline{b} \|^2$ (and we use $(\underline{a}, \underline{b}) = (\underline{b}, \underline{a}) = 0$):

$$\begin{aligned} \| \underline{a} + \underline{b} \|^2 &= (\underline{a} + \underline{b}, \underline{a} + \underline{b}) = (\underline{a}, \underline{a}) + (\underline{a}, \underline{b}) + (\underline{b}, \underline{a}) + (\underline{b}, \underline{b}) \\ &= (\underline{a}, \underline{a}) + (\underline{b}, \underline{b}) = \| \underline{a} \|^2 + \| \underline{b} \|^2 . \end{aligned}$$

The second part follows from the first one plus an induction argument. We leave this to the reader. □

6.2 Orthogonal complements and orthonormal bases

6.2.1 In terms of the standard inner product on $\mathbb{R}^3$ the equation $2x_1 + 3x_2 - x_3 = 0$ can be rewritten as
$$((2, 3, -1), (x_1, x_2, x_3)) = 0.$$

So the solutions of the equation are precisely the vectors $\underline{x} \in \mathbb{R}^3$ which are perpendicular to $(2, 3, -1)$. There is a similar interpretation of the solutions of a homogeneous system of linear equations $A\underline{x} = \underline{0}$. If $A = (a_{ij})$ is a $m \times n$–matrix, then the i-th element of the matrix product $A\underline{x}$, i.e., $a_{i1}x_1 + \cdots + a_{in}x_n$, equals the standard inner product $(\underline{a}_i, \underline{x})$ of the i-th row of A and the vector $\underline{x}$. The solutions of the system are therefore the vectors which are perpendicular to all rows of A.

There are more situations where perpendicularity is useful as we will see. In this section we discuss

- The set of vectors perpendicular to a given subspace (orthogonal complement),
- sets of vectors of length 1 which are mutually perpendicular (orthonormal sets),
- the role of orthonormal sets of vectors in computing orthogonal projections and in working with coordinates,

- the Gram-Schmidt procedure to transform a given set of vectors into an orthonormal set.

6.2.2 Orthogonal complement
Let V be a real or complex inner product space and let W be a linear subspace of V. The set of vectors which are perpendicular to all vectors of W is denoted by $W^\perp$. More precisely:

6.2.3 Definition. Let W be a linear subspace of an inner product space V. The *orthogonal complement* of W is the set

$$W^\perp = \{\underline{x} \in V \mid (\underline{x}, \underline{w}) = 0 \text{ voor alle } \underline{w} \in W\} \, .$$

6.2.4 Here are some properties of the orthogonal complement $W^\perp$:

- $W^\perp$ is a linear subspace of V. The orthogonal complement $W^\perp$ is non-empty since $\underline{0} \in W^\perp$. If $\underline{x} \in W^\perp$ and $\underline{y} \in W^\perp$, then for all $\underline{w} \in W$ $(\underline{x}+\underline{y}, \underline{w}) = (\underline{x}, \underline{w}) + (\underline{y}, \underline{w}) = 0$, so that $\underline{x}+\underline{y} \in W^\perp$, and $(\alpha\underline{x}, \underline{w}) = \alpha(\underline{x}, \underline{w}) = 0$, so that $\alpha\underline{x} \in W^\perp$.

- $W \cap W^\perp = \{\underline{0}\}$, i.e., a linear subspace W and its orthogonal complement $W^\perp$ only have the zero vector in common. Suppose $\underline{x} \in W \cap W^\perp$. Since $\underline{x} \in W^\perp$ we have $(\underline{x}, \underline{w}) = 0$ for all $\underline{w} \in W$. Now apply this to $\underline{w} = \underline{x} \in W$. Then $(\underline{x}, \underline{x}) = 0$ and we get $\underline{x} = \underline{0}$.

To compute the orthogonal complement of a span $< \underline{a}_1, \ldots, \underline{a}_n >$ it suffices to find the vectors perpendicular to each of the vectors $\underline{a}_1, \ldots, \underline{a}_n$. We formulate this result as a theorem.

6.2.5 Theorem. *If* $W =< \underline{a}_1, \ldots, \underline{a}_n >$, *then*

$$W^\perp = \{\underline{x} \in V \mid (\underline{a}_i, \underline{x}) = 0 \text{ for } i = 1, \ldots, n\} \, .$$

Proof. If $\underline{x} \in W^\perp$, then $\underline{x}$ is perpendicular to all vectors in W, in particular, $\underline{x}$ is perpendicular to $\underline{a}_1, \ldots, \underline{a}_n$. This means that $(\underline{a}_i, \underline{x}) = 0$ for $i = 1, \ldots, n$. Conversely, let $\underline{x}$ be a vector with the property that $(\underline{a}_i, \underline{x}) = 0$ for $i = 1, \ldots, n$. An arbitrary vector $\underline{w} \in W$ can be written as a linear combination $\underline{w} = \sum_{i=1}^n \alpha_i \underline{a}_i$. Then $(\underline{w}, \underline{x}) = (\sum_{i=1}^n \alpha_i \underline{a}_i, \underline{x}) = \sum_{i=1}^n \alpha_i (\underline{a}_i, \underline{x}) = 0$ by linearity of the inner product. So $\underline{x}$ is perpendicular to every vector in W. Consequently, $\underline{x} \in W^\perp$. $\square$

6.2.6 Example. In $\mathbb{R}^3$ we determine all vectors which are perpendicular to $(1,2,-1)$, i.e., the orthogonal complement of the subspace $l =< (1,2,-1) >$. A vector $\underline{x} = (x,y,z)$ is in this complement if and only if $((1,2,-1),\underline{x}) = 0$, or $x+2y-z=0$. So $l^\perp$ is the plane $V : x+2y-z=0$.

Next we determine the orthogonal complement of the plane V. We first determine a parametric representation of V. Let $z=\lambda$ and $y=\mu$, then $x=\lambda-2\mu$ and

$$V \ = < (1,0,1), (-2,1,0) > \ .$$

$V^\perp$ consists precisely of all vectors (x,y,z) satisfying

$$\begin{array}{rcl} ((1,0,1),(x,y,z)) & = & x \qquad\quad +z = 0, \\ ((-2,1,0),(x,y,z)) & = & -2x \quad +y \qquad = 0. \end{array}$$

Neem $x=\lambda$, dan is $y=2\lambda$ en $z=-\lambda$, dus

$$V^\perp =< (1,2,-1) >= l\ ,$$

which is maybe not so surprising (it is actually an example of a general statement on orthogonal complements).

In a similar way we find that the orthogonal complement of the line $l =< (a,b,c) >$ in $\mathbb{R}^3$ is the plane $V : ax+by+cz=0$, and that, conversely, $V^\perp =< (a,b,c) >$.

6.2.7 Orthonormal sets of vectors

Sets of length 1 vectors which are mutually orthogonal are of special importance. Such sets of vectors are called *orthonormal sets*. An example is the standard basis $\underline{e}_1,\ldots,\underline{e}_n$ of $\mathbb{R}^n$ or $\mathbb{C}^n$. Orthonormal sets are useful in computing orthogonal projections, distances, and coordinates.

6.2.8 Definition. Let V be an inner product space. The vectors $\underline{e}_1,\ldots,\underline{e}_n$ in V form an *orthonormal set* if for $1 \le i,j \le n$

$$(\underline{e}_i,\underline{e}_j) = \begin{cases} 0 \text{ als } i \neq j\ , \\ 1 \text{ als } i = j\ . \end{cases}$$

If moreover $\underline{e}_1,\ldots,\underline{e}_n$ is a basis of V, then the set is called an *orthonormal basis* of V.

6.2.9 Example. The set of vectors

$$\frac{1}{\sqrt{2}}(1,-1,0),\ \frac{1}{\sqrt{2}}(1,1,0),\ (0,0,1)$$

in $\mathbb{R}^3$ is an orthonormal set. It is not difficult to see that this set is linearly independent (see also the following theorem) and therefore an orthonormal basis of $\mathbb{R}^3$.

6.2.10 Theorem. *Let V be an inner product space.*

1. *Orthonormal sets of vectors in V are linearly independent.*
2. *If $\underline{a}_1, \ldots, \underline{a}_n$ is an orthonormal basis of V, then the coordinates of $\underline{x}$ with respect to this basis are $(\underline{x}, \underline{a}_1), \ldots, (\underline{x}, \underline{a}_n)$, respectively. In the case of a real vector space we have*

$$\| \underline{x} \|^2 = (\underline{x}, \underline{a}_1)^2 + \cdots + (\underline{x}, \underline{a}_n)^2,$$

i.e., the length of $\underline{x}$ equals the length of the coordinate vector of $\underline{x}$ with respect to the standard inner product. (There is a similar equality in the complex setting, we won't go into that.)

Proof. 1) To prove linear independence, we study the equation

$$\lambda_1 \underline{a}_1 + \cdots + \lambda_n \underline{a}_n = \underline{0} \tag{6.2}$$

in $\lambda_1, \ldots, \lambda_n$. Now take the inner product on both sides of (6.2) with the vector $\underline{a}_j$:

$$(\lambda_1 \underline{a}_1 + \cdots + \lambda_n \underline{a}_n, \underline{a}_j) = (\underline{0}, \underline{a}_j) = 0.$$

Since $\underline{a}_1, \ldots, \underline{a}_n$ is an orthonormaal set, the left-hand side simplifies to λ_j. Consequently, $\lambda_j = 0$.

2) To prove the second statement, we first note that since $\underline{a}_1, \ldots, \underline{a}_n$ is a basis of V there are scalars $\mu_1, \ldots, \mu_n$ with $\underline{x} = \mu_1 \underline{a}_1 + \cdots + \mu_n \underline{a}_n$. To determine the μ_j, we again take the inner product of both sides with $\underline{a}_j$:

$$(\underline{x}, \underline{a}_j) = (\mu_1 \underline{a}_1 + \cdots + \mu_n \underline{a}_n, \underline{a}_j).$$

Expanding the right-hand side yields $(\underline{x}, \underline{a}_j) = \mu_j$. The equality involving the length is a direct consequence of the Pythagorean theorem. 6.1.22. □

6.2.11 Example. The set 6.2.9 is an orthonormal set and therefore linearly independent by Theorem 6.2.10. So the three vectors are a basis. The coordinates of $\underline{a} = (2, 2, 2)$ with respect to this basis are

$$(\underline{a}, \frac{1}{\sqrt{2}}(1, -1, 0)) = 0,\ (\underline{a}, \frac{1}{\sqrt{2}}(1, 1, 0)) = 2\sqrt{2},\ (\underline{a}, (0, 0, 1)) = 2.$$

So:

$$(2, 2, 2) = 2\sqrt{2}\frac{1}{\sqrt{2}}\,(1, 1, 0) + 2\,(0, 0, 1).$$

6.2.12 Orthogonal projection

Orthonormal sets are useful in determining orthogonal projections on subspaces (and such projections are useful since they 'solve' shortest distance problems). Let W be a linear subspace of the inner product space V. The *orthogonal projection* of $\underline{x} \in V$ on the subspace W is the vector $\underline{y}$ in W such that $\underline{x} - \underline{y}$ is perpendicular to the subspace W, i.e., belongs to $W^\perp$. (The vector $\underline{y}$ turns out to be unique as we will see below.)

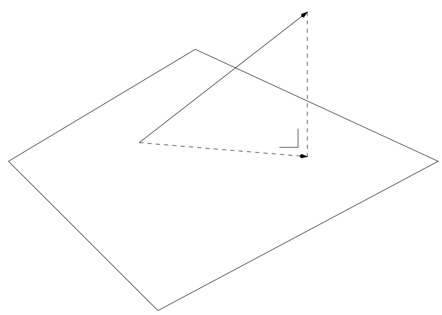


Figure 6.1: Loodrechte projectie.

Suppose that $\underline{a}_1, \ldots, \underline{a}_k$ is an orthonormal basis of the subspace W. Then the vector $\underline{y}$ can be written as $\lambda_1\underline{a}_1 + \cdots + \lambda_k\underline{a}_k$ for some scalars $\lambda_1, \ldots, \lambda_k$. The condition that $\underline{x} - \underline{y}$ is orthogonal to W is equivalent to

$$\begin{aligned}
(\underline{x} - (\lambda_1\underline{a}_1 + \cdots + \lambda_k\underline{a}_k), \underline{a}_1) &= 0\\
&\vdots\\
(\underline{x} - (\lambda_1\underline{a}_1 + \cdots + \lambda_k\underline{a}_k), \underline{a}_k) &= 0.
\end{aligned}$$

Expanding these inner products yields

$$\lambda_1 = (\underline{x}, \underline{a}_1), \ldots, \lambda_k = (\underline{x}, \underline{a}_k).$$

In particular, there is precisely one such vector $\underline{y}$. We denote the projection of $\underline{x}$ by $P_W(\underline{x})$. We state this result in the following theorem. We also show that $P_W(\underline{x})$ is the vector in W with minimal distance to $\underline{x}$.

6.2.13 Theorem. *Let $\underline{a}_1, \ldots, \underline{a}_k$ be an orthonormal basis of the linear subspace W of the inner product space V. If $P_W(\underline{x})$ is the orthogonal prrojection of $\underline{x}$ on W, then:*

1. *$\underline{x} - P_W(\underline{x})$ is perpendicular to every vector from W.*
2. *The orthogonal projection $P_W(\underline{x})$ of $\underline{x} \in V$ on $W =< \underline{a}_1, \ldots, \underline{a}_k >$ equals $P_W(\underline{x}) = (\underline{x}, \underline{a}_1)\underline{a}_1 + \cdots + (\underline{x}, \underline{a}_k)\underline{a}_k$.*

3. $\| \underline{x} - P_W(\underline{x}) \| = \min_{\underline{z} \in W} \| \underline{x} - \underline{z} \|$, *i.e., the orthogonal projection is the unique vector in W with minimal distance to $\underline{x}$.*

4. $\| P_W(\underline{x}) \| \leq \| \underline{x} \|$ *with equality occurring if and only if $\underline{x} = P_W(\underline{x})$.*

Proof. The first item is just the definition; the second part has been proved above. So we turn to the third item. Take any vector $\underline{z}$ in W. We will compare

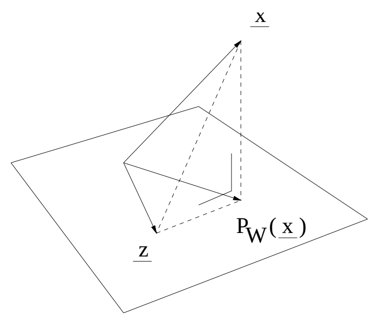


Figure 6.2: Orthogonal projection and shortest distance.

$\| \underline{x} - P_W(\underline{x}) \|$ and $\| \underline{x} - \underline{z} \|$. To this end, write $\underline{x} - \underline{z} = (\underline{x} - P_W(\underline{x})) + (P_W(\underline{x}) - \underline{z})$. The component $P_W(\underline{x}) - \underline{z}$ is in W and is therefore orthogonal to $\underline{x} - P_W(\underline{x})$. The Pythagorean theorem 6.1.22 now implies

$$\| \underline{x} - \underline{z} \|^2 = \| \underline{x} - P_W(\underline{x}) \|^2 + \| P_W(\underline{x}) - \underline{z} \|^2 .$$

Since lengths are non-negative we find

$$\| \underline{x} - P_W(\underline{x}) \| \leq \| \underline{x} - \underline{z} \|$$

with equality if and only if $\| P_W(\underline{x}) - \underline{z} \| = 0$, i.e., if and only if $P_W(\underline{x}) = \underline{z}$.

The last statement in the theorem follows in a similar way from the Pythagorean theorem 6.1.22: since $\underline{x} - P_W(\underline{x})$ and $P_W(\underline{x})$ are perpendicular (note that the second vector is in W), we have

$$\| \underline{x} \|^2 = \| \underline{x} - P_W(\underline{x}) \|^2 + \| P_W(\underline{x}) \|^2 .$$

So $\| P_W(\underline{x}) \| \leq \| \underline{x} \|$ with equality precisely if $\underline{x} = P_W(\underline{x})$. □

6.2.14 Example. The orthogonal projection of the vector $(1, 0, 1) \in \mathbb{R}^3$ onto the the line $l = < (1, 2, 1) >$ can be computed as follows. First we divide $(1, 2, 1)$ by its length $\sqrt{6}$ to get a vector on the line with length 1:

$$l = < \frac{1}{\sqrt{6}} (1, 2, 1) > .$$

Next we apply Theorem 6.2.13 to find

$$((1,0,1), \frac{1}{\sqrt{6}}(1,2,1)) \cdot \frac{1}{\sqrt{6}}(1,2,1) = \frac{1}{3}(1,2,1).$$

In general, the orthogonal projection of $\underline{x}$ onto the line $< \underline{a} >$, where $\underline{a}$ has length 1, equals $(\underline{x}, \underline{a})\underline{a}$.

6.2.15 Constructing orthonormal bases

Coordinates with respect to orthonormal bases can be easily found by taking appropriate inner products. Our next task is to construct orthonormal sets and bases.

6.2.16 Theorem. *Let $\underline{a}_1, \ldots, \underline{a}_n$ be linearly independent vectors in an inner product space V. There exists a constructive process to transform this set of vectors into an orthonormal set $\underline{e}_1, \ldots, \underline{e}_n$ in such a way that*

$$< \underline{a}_1, \ldots, \underline{a}_i > = < \underline{e}_1, \ldots, \underline{e}_i > \text{ voor } i = 1, \ldots, n\ .$$

In particular, every finite dimensional inner product space has an orthonormal basis.

6.2.17 Proof: the Gram-Schmidt process

We prove this theorem by providing an algorithm that actually produces the required orthonormal basis $\underline{e}_1, \ldots, \underline{e}_n$. This algorithm can be carried out by a computer or, in simple cases, by hand. It is called the *Gram-Schmidt process*.

- **Step 1** First we replace $\underline{a}_1$ by a vector of length 1 in $\langle \underline{a}_1 \rangle$:

$$\underline{e}_1 = \frac{\underline{a}_1}{\| \underline{a}_1 \|}.$$

  Then $\underline{e}_1$ is an orthonormal set and we have $< \underline{a}_1 > = < \underline{e}_1 >$.

- **Step** i: Suppose we have found the orthonormal set $\underline{e}_1, \ldots, \underline{e}_i$ already. In particular:

$$< \underline{a}_1, \ldots, \underline{a}_i > = < \underline{e}_1, \ldots, \underline{e}_i > \ .$$

  Then we construct $\underline{e}_{i+1}$ as follows. We first compute the orthogonal projection of $\underline{a}_{i+1}$ onto $< \underline{e}_1, \ldots, \underline{e}_i >$. This projection is $(\underline{a}_{i+1}, \underline{e}_1)\underline{e}_1 + \cdots + (\underline{a}_{i+1}, \underline{e}_i)\underline{e}_i$ by Theorem 6.2.13. By definition of orthogonal projection, the difference

$$\underline{a}^*_{i+1} = \underline{a}_{i+1} - [(\underline{a}_{i+1}, \underline{e}_1)\underline{e}_1 + \cdots + (\underline{a}_{i+1}, \underline{e}_i)\underline{e}_i]$$

is perpendicular to $<\underline{e}_1,\ldots,\underline{e}_i>$, and, in particular, perpendicular to each of the vectors $\underline{e}_1,\ldots,\underline{e}_i$. Now $\underline{a}^*_{i+1}$ is not the zero-vector because otherwise $\underline{a}_{i+1}$ would be a linear combination of $\underline{e}_1,\ldots\underline{e}_i$, and therefore of $\underline{a}_1,\ldots,\underline{a}_i$, contradicting the fact that the $\underline{a}_j$ are linearly independent. Construct a vector of length 1 spanning the same line as $\underline{a}^*_{i+1}$ by dividing by its length $\parallel \underline{a}^*_{i+1} \parallel$:

$$\underline{e}_{i+1} := \frac{\underline{a}^*_{i+1}}{\parallel \underline{a}^*_{i+1} \parallel} = \frac{\underline{a}_{i+1} - \sum_{k=1}^{i}(\underline{a}_{i+1},\underline{e}_k)\underline{e}_k}{\parallel \underline{a}_{i+1} - \sum_{k=1}^{i}(\underline{a}_{i+1},\underline{e}_k)\underline{e}_k \parallel} \ .$$

Since $\underline{e}_{i+1} \in <\underline{e}_1,\ldots,\underline{e}_i>^{\perp}$, the vectors $\underline{e}_1,\ldots,\underline{e}_i,\underline{e}_{i+1}$ form an orthonormal set. Moreover,

$$<\underline{a}_1,\ldots,\underline{a}_i,\underline{a}_{i+1}> = <\underline{e}_1,\ldots,\underline{e}_i,\underline{a}_{i+1}> = <\underline{e}_1,\ldots,\underline{e}_i,\underline{e}_{i+1}>,$$

where the first equality follows from $<\underline{e}_1,\ldots,\underline{e}_i> = \langle \underline{a}_1,\ldots,\underline{a}_i\rangle$, and the second equality follows from the way we have constructed $\underline{e}_{i+1}$: we have added a linear combination of $\underline{e}_1,\ldots,eu_i$ to $\underline{a}_{i+1}$ and multiplied by a non-zero scalar. Both operations do not change the span.

Step 2 is then used to construct $\underline{e}_2$ from $\underline{e}_1$, then $\underline{e}_3$ from $\underline{e}_1$ and $\underline{e}_2$, etc.

6.2.18 Example. Consider the plane

$$V : x + y - 2z = 0 \ .$$

in $\mathbb{R}^3$. Take $z=\lambda$, $y=\mu$, then $x = 2\lambda-\mu$ and it easily follows that $V = <(2,0,1),(-1,1,0)>$. We construct an orthonormal basis of V as follows. First we let

$$\underline{e}_1 = \frac{(2,0,1)}{\parallel (2,0,1) \parallel} = \frac{1}{\sqrt{5}}(2,0,1) \ .$$

Then we let

$$\begin{aligned} \underline{a}^*_2 &= \underline{a}_2 - P_{<\underline{e}_1>}\underline{a}_2 = (-1,1,0) - (\,(-1,1,0),(\tfrac{2}{\sqrt{5}},0,\tfrac{1}{\sqrt{5}})\,)\,(\tfrac{2}{\sqrt{5}},0,\tfrac{1}{\sqrt{5}}) \\ &= \tfrac{1}{5}(-1,5,2) \end{aligned}$$

so that

$$\underline{e}_2 = \frac{\underline{a}^*_2}{\parallel \underline{a}^*_2 \parallel} = \frac{1}{\sqrt{30}}(-1,5,2) \ .$$

An orthonormal basis is therefore $\{\frac{1}{\sqrt{5}}(2,0,1),\ \frac{1}{\sqrt{30}}(-1,5,2)\}$. Note that if you start with other vectors spanning V, or with the same vectors but in a different order, you usually obtain a different orthonormal basis. For example, if you start with the vectors $(-1,1,0)$, $(2,0,1)$, i.e., just the order has changed, then you find the orthonormal basis $\{\frac{1}{\sqrt{2}}(-1,1,0), \frac{1}{\sqrt{3}}(1,1,1)\}$ of V.

6.2.19 **Example.** We determine an orthonormal basis for the span of the vectors $\underline{a} = (1,1,1,1)$, $\underline{b} = (1,-1,2,0)$, $\underline{c} = (5,0,1,-4)$ in $\mathbb{R}^4$.
In the first step we obtain:

$$\underline{e}_1 = \frac{\underline{a}}{\| \underline{a} \|} = \frac{1}{2}(1,1,1,1) .$$

Next, $P_{<\underline{e}_1>}(\underline{b}) = (\underline{b}, \underline{e}_1)\underline{e}_1 = \frac{1}{2}(1,1,1,1)$, so $\underline{b} - P_{<\underline{e}_1>}(\underline{b}) = \frac{1}{2}(1,-3,3,-1)$. Therefore

$$\underline{e}_2 = \frac{(1,-3,3,-1)}{\| (1,-3,3,-1) \|} = \frac{1}{2\sqrt{5}}(1,-3,3,-1) .$$

In the next step, let $P_{<\underline{e}_1,\underline{e}_2>}(\underline{c}) = (\underline{c}, \underline{e}_1)\underline{e}_1 + (\underline{c}, \underline{e}_2)\underline{e}_2 = \frac{1}{2}(1,1,1,1) + \frac{3}{5}(1,-3,3,-1)$ $= \frac{1}{10}(11,-13,23,-1)$. Then $\underline{c} - P_{<\underline{e}_1,\underline{e}_2>}(\underline{c}) = \frac{13}{10}(3,1,-1,-3)$ and

$$\underline{e}_3 = \frac{(3,1,-1,-3)}{\| (3,1,-1,-3) \|} = \frac{1}{2\sqrt{5}}(3,1,-1,-3).$$

The set $\{\underline{e}_1, \underline{e}_2, \underline{e}_3\}$ is the required orthonormal basis.

6.2.20 **Orthonormal sets and coordinates**
Now that we know that every finite-dimensional inner product space V has an orthonormal basis and that we can find such bases from given ones using the Gram-Schmidt process, we turn to coordinates with respect to orthonormal bases.

Let $\{\underline{e}_1, \ldots, \underline{e}_n\}$ be an orthonormal basis of V and let

$$\underline{x} = \sum_{i=1}^{n} x_i \underline{e}_i \in V , \quad \underline{y} = \sum_{j=1}^{n} y_j \underline{e}_j \in V$$

be two arbitrary vectors in V. Then

$$(\underline{x}, \underline{y}) = \left(\sum_{i=1}^{n} x_i \underline{e}_i, \sum_{j=1}^{n} y_j \underline{e}_j \right) = \sum_{i=1}^{n} \sum_{j=1}^{n} x_i \overline{y_j} (\underline{e}_i, \underline{e}_j) = \sum_{i=1}^{n} x_i \overline{y_i}.$$

This means that if we use coordinates with respect to the orthonormal basis, the inner product of $\underline{x}$ and $\underline{y}$ equals the 'ordinary' inner product (in $\mathbb{R}^n$ or $\mathbb{C}^n$ depending on whether we work in a real or complex vector space) of their coordinate vectors.

So, if we work with coordinates with respect to an orthonormal basis, computations in an n-dimensional real or complex inner product space can be 'translated' into computations in $\mathbb{R}^n$ or $\mathbb{C}^n$ with the standard inner product.

We next turn to an interesting relation between a linear subspace and its orthogonal complement. Let W be a k–dimensional lineare subspace of an n–dimensional inner product space V. Choose any basis $\underline{a}_1, \ldots, \underline{a}_k$ of W and supplement it with vectors $\underline{a}_{k+1}, \ldots, \underline{a}_n$ to a basis $\underline{a}_1, \ldots, \underline{a}_n$ of V. Apply the Gram-Schmidt process to $\underline{a}_1, \ldots, \underline{a}_n$ to find an orthonormal basis $\underline{e}_1, \ldots, \underline{e}_n$ of V such that $\underline{e}_1, \ldots, \underline{e}_k$ is an orthonormal basis of W. By Theorem 6.2.10 every vector $\underline{x} \in V$ can now be written as

$$\underline{x} = \sum_{i=1}^{k} (\underline{x}_i, \underline{e}_i)\ \underline{e}_i + \sum_{i=k+1}^{n} (\underline{x}, \underline{e}_i)\ \underline{e}_i \ = \ P_w(\underline{x}) + \sum_{i=k+1}^{n} (\underline{x}, \underline{e}_i)\ \underline{e}_i.$$

Then Theorem 6.2.13 implies that

$$W^\perp = < \underline{e}_{k+1}, \ldots, \underline{e}_n >$$

(or: if $\underline{x} \in W^\perp$ then $P_w(\underline{x}) = \underline{0}$ and $\underline{x} \in < \underline{e}_{k+1}, \ldots, \underline{e}_n >$; if $\underline{x} \in < \underline{e}_{k+1}, \ldots, \underline{e}_n >$, then, by orthonormality, $(\underline{x}, \underline{e}_i) = 0$ for $i = 1, \ldots, k$, hence $\underline{x} \in W^\perp$).

So the orthogonal complement $W^\perp$ of W has dimension $n - k$ and the set $\{\underline{e}_{k+1}, \ldots, \underline{e}_n\}$ is an orthonormal basis of $W^\perp$. In conclusion:

6.2.21 Theorem. *Let W be a k-dimensional linear subspace of an n-dimensional inner product space V. Then there exists an orthonormal basis $\{\underline{e}_1, \ldots, \underline{e}_n\}$ of V such that $W = < \underline{e}_1, \ldots, \underline{e}_k >$ and $W^\perp = < \underline{e}_{k+1}, \ldots, \underline{e}_n >$. In particular,*

$$\dim V = \dim W + \dim W^\perp. \tag{6.3}$$

6.2.22 Homogeneous systems and orthogonal complements

We have come across the relation $\dim V = \dim W + \dim W^\perp$ before, but in a different guise. Let $\underline{a}_1, \ldots, \underline{a}_m$ be the rows of an $m \times n$–matrix A. The matrix product $A\underline{x}$ of A with a (column) vector $\underline{x} \in \mathbb{R}^n$ can be viewed as a column of inner products of each of the vectors $\underline{a}_i$ with $\underline{x}$:

$$\begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} (\underline{a}_1, \underline{x}) \\ \vdots \\ (\underline{a}_m, \underline{x}) \end{pmatrix}.$$

So the set of solutions of the system $A\underline{x} = \underline{0}$ is exactly the orthogonal complement of $< \underline{a}_1, \ldots, \underline{a}_m >$. If the rank of A is k, then we find, using Theorem 6.2.21, that the dimension of the solution space is $n - k$, in agreement with Theorem 5.1.13.

6.2.23 Example. In $\mathbb{R}^4$ consider the linear subspace V given by

$$\begin{aligned} x \quad\quad +2z - u &= 0, \\ x + y \quad\quad\quad\quad &= 0. \end{aligned}$$

Solving these equations show that

$$V \;\; = < (1,-1,0,1), (-2,2,1,0) > .$$

The vectors $(1,0,2,-1)$ and $(1,1,0,0)$ (produced from the coefficients of the equations) are perpendicular to V. Since $\dim(V) = 2$ we have $\dim(V^\perp) = 2 (= 4-2)$; as the two vectors $(1,0,2,-1)$, $(1,1,0,0)$ are linearly independent we conclude

$$V^\perp = < (1,0,2,-1),\ (1,1,0,0) > .$$

Now apply Gram-Schmidt to the two spanning vectors of V and the two spanning vectors of $V^\perp$. Then we find

$$\frac{1}{\sqrt{3}}(1,-1,0,1), \frac{1}{\sqrt{33}}(-2,2,3,4), \frac{1}{\sqrt{6}}(1,0,2,-1), \frac{1}{\sqrt{66}}(5,6,-2,1)$$

which form an orthonormal basis of $\mathbb{R}^4$ in such a way that the first two vectors span V and the last two vectors span $V^\perp$.

6.2.24 Projection onto $W^\perp$
If W is a linear subspace of the finite-dimensional inner product space V and $P_W(\underline{x})$ is the orthogonal projection of $\underline{x}$ onto W, then it seems plausible that the vector $\underline{x} - P_W(\underline{x})$, which is perpendicular to W, is the orthogonal projection of $\underline{x}$ onto $W^\perp$. To show that this is really the case we must check that $\underline{x} - (\underline{x} - P_W(\underline{x})) = P_W(\underline{x})$ is perpendicular to all vectors from $W^\perp$, i.e., that it belongs to $(W^\perp)^\perp$. As $P_W(\underline{x}) \in W$ it suffices to show that $W = (W^\perp)^\perp$. Now it's evident that every vector from W is perpendicular to every vector from $W^\perp$ (from the definition of $W^\perp$!), so that $W \subseteq (W^\perp)^\perp$. The dimension formula (6.3) implies that $\dim(W^\perp)^\perp = \dim V - \dim W^\perp = \dim W$. Hence $W = (W^\perp)^\perp$.

This observation on projection onto $W^\perp$ is useful in computations, since one of the projections may be easier to compute (directly) than the other.

6.2.25 Example. We determine the orthogonal projection of $(1,2,1) \in \mathbb{R}^3$ onto the plane $W : x + y + z = 0$. The orthogonal complement $W^\perp$ is the line $< (1,1,1) >$ and the projection of $(1,2,1)$ onto this line is

$$((1,2,1), \frac{1}{\sqrt{3}}(1,1,1)) \frac{1}{\sqrt{3}}(1,1,1) = \frac{4}{3}(1,1,1).$$

The orthogonal projection of $(1,2,1)$ onto W is therefore

$$(1,2,1) - \frac{4}{3}(1,1,1) = \frac{1}{3}(-1,2,-1).$$

6.3 The QR-decomposition

6.3.1 The relation between a linearly independent set of vectors in $\mathbb{R}^m$ and the resulting orthonormal set, derived using the Gram-Schmidt process, can also be expressed in terms of matrices. This leads to a way to express an $m \times n$–matrix whose columns are linearly independent as a product of two matrices with special properties.

6.3.2 The Gram-Schmidt process in terms of matrices
Let $\underline{a}_1, \ldots, \underline{a}_n \in \mathbb{R}^m$ be n linearly independent vectors. Collect them as columns in an $m \times n$-matrix A. Applying Gram-Schmidt to these vectors produces an orthononormal set $\underline{u}_1, \ldots, \underline{u}_n \in \mathbb{R}^m$. Now every vector $\underline{a}_k$ is a linear combination of the first k vectors $\underline{u}_1, \ldots, \underline{u}_k$ from the orthonormal set:

$$\underline{a}_k = r_{1k}\underline{u}_1 + \cdots + r_{kk}\underline{u}_k + 0 \cdot \underline{u}_{k+1} + \cdots 0 \cdot \underline{u}_n$$

with $r_{kk} \neq 0$ (why?). By replacing $\underline{u}_k$ by $-\underline{u}_k$ if necessary, we can arrange that every $r_{kk} > 0$. Now collect the vectors $\underline{u}_1, \ldots, \underline{u}_n$ as columns in the $m \times n$-matrix Q and the numbers r_{ij} in the $n \times n$-matrix R (column k of R contains the coefficients that we used for $\underline{a}_k$). Then we get the following equality of matrices:

$$A = QR,$$

in which R is an upper triangular matrix with positive entries on the diagonal. This decomposition is usually called the *QR-decomposition* of the matrix A.

The matrix R can be found in various ways in concrete problems. One of these ways is the following, which uses the orthonormality of the set $\{\underline{u}_1, \ldots, \underline{u}_n\}$. The equalities

$$\underline{u}_k^\top \cdot \underline{u}_\ell = (\underline{u}_k, \underline{u}_\ell) = \delta_{k\ell}$$

can be rewritten as

$$Q^\top Q = I_n.$$

But then

$$Q^\top A = Q^\top QR = I_n R = R.$$

Another way is via careful bookkeeping when carrying out the Gram-Schmidt process.

6.3.3 Example. Applying the Gram-Schmidt process to the linearly independent vectors $(2, 0, 1), (-1, 1, 0)$ yields the orthonormal vectors

$$\frac{1}{\sqrt{5}}(2, 0, 1),\ \frac{1}{\sqrt{30}}(-1, 5, 2).$$

The corresponding QR-decomposition is therefore

$$\begin{pmatrix} 2 & -1 \\ 0 & 1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} \frac{2}{\sqrt{5}} & \frac{-1}{\sqrt{30}} \\ 0 & \frac{5}{\sqrt{30}} \\ \frac{1}{\sqrt{5}} & \frac{2}{\sqrt{30}} \end{pmatrix} \cdot \begin{pmatrix} \sqrt{5} & -\frac{2}{\sqrt{5}} \\ 0 & \frac{6}{\sqrt{30}} \end{pmatrix}.$$

Here, R has been found, for instance, via $R = Q^\top A$.

6.4 Notes

The Cauchy–Schwarz inequality is named after A.–L. Cauchy (1789–1857) and H.A. Schwarz (1843–1921). Cauchy described the inequality in terms of sequences of numbers, whereas Schwarz worked in function spaces. The inequality is sometimes also named after V.Y. Bunyakovsky (1804–1889), who came up with it independently, also in the setting of function spaces.

Analysis

Inner product spaces of functions will be further discussed in more advanced analysis courses. Applications of such inner product spaces can be found in, for instance, signal analysis and in quantum mechanics.

Relativity theory

A variation of the inner product, in which non-zero vectors need not have a positive length, occurs in relativity theory. Implicitly, a bit of this can be seen in the classification of quadratic forms in Linear Algebra 2.

The Pythagorean theorem has its roots, of course, in geometry, but has surprising and useful consequences in cleverly chosen function spaces. An example is the inequality

$$\sum_{n=1}^{\infty} \frac{1}{n^2} \leq \frac{\pi^2}{6},$$

which can be derived using an inner product space as in example 6.1.7. (By the way, the inequality turns out to be an equality, a famous result due to Euler.)

The Gram-Schmidt process is named after the Dane J.P. Gram (1850–1916) and the German E. Schmidt (1876–1959) and was introduced in the setting of function spaces.

Linear Algebra

Orthogonal projections can be used to derive the method of least squares. This method is an essential tool in, e.g., dealing with measurements. The notions length, angle, orthogonality will reoccur in Linear Algebra 2 in the study of orthogonal maps, like reflections and rotations.

The data of an inner product can be neatly stored in a so-called *Gram–matrix*. If $\{\underline{a}_1, \ldots, \underline{a}_n\}$ is a basis of the inner product space V, then the entry in position i, j of this $n \times n$–matrix is the inner product $(\underline{a}_i, \underline{a}_j)$.

6.5 Exercises

§1

1 Which of the following expressions define an inner product in $\mathbb{R}^2$ (where $\underline{a} = (a_1, a_2)$, $\underline{b} = (b_1, b_2)$)?

a. $(\underline{a}, \underline{b}) = a_1 b_1 + 2a_2 b_2$,

b. $(\underline{a}, \underline{b}) = a_1 b_1 - a_2 b_2$,

c. $(\underline{a}, \underline{b}) = \begin{pmatrix} a_1 & a_2 \end{pmatrix} \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix} \begin{pmatrix} b_1 \\ b_2 \end{pmatrix}$,

d. $(\underline{a}, \underline{b}) = \begin{pmatrix} a_1 & a_2 \end{pmatrix} \begin{pmatrix} 4 & 2 \\ 2 & 1 \end{pmatrix} \begin{pmatrix} b_1 \\ b_2 \end{pmatrix}$.

2 Use the Cauchy-Schwarz inequality to prove that for all real numbers a, b, c the following inequality holds:

$$|a + 2b + 2c| \leq 3\sqrt{a^2 + b^2 + c^2}.$$

3 Compute the distance between the given points in $\mathbb{R}^3$:

a. $(2, 1, 3)$ en $(-1, 2, 4)$;

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

4 Determine the angle between the given pairs of vectors:

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

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

c. $(2, 3, 0, 2, 1)$ en $(0, 2, 2, 2, 2)$ in $\mathbb{R}^5$;

d. $(1, 4, 4, 1, 1, 1)$ en $(0, -1, 3, 2, 1, 1)$ in $\mathbb{R}^6$.

5 In the real inner product space V the (mutually distinct) vectors $\underline{a}, \underline{b}, \underline{c}, \underline{d}$ satisfy $\underline{a} - \underline{b} \perp \underline{c} - \underline{d}$ and $\underline{a} - \underline{c} \perp \underline{b} - \underline{d}$. Prove that $\underline{a} - \underline{d} \perp \underline{b} - \underline{c}$.

6 a. Prove the following equality for every pair of vectors $\underline{a}$ and $\underline{b}$ in $\mathbb{R}^n$:

$$\|\underline{a} + \underline{b}\|^2 + \|\underline{a} - \underline{b}\|^2 = 2\|\underline{a}\|^2 + 2\|\underline{b}\|^2.$$

Also give a geometrical interpretation.

b. Prove the following inequality for every pair of vectors $\underline{a}$ and $\underline{b}$ in $\mathbb{R}^n$:

$$4|(\underline{a},\underline{b})| \;\leq\; (\underline{a}+\underline{b},\underline{a}+\underline{b})+(\underline{a}-\underline{b},\underline{a}-\underline{b}).$$

7 In the real inner product space V the vectors $\underline{a}$, $\underline{b}$ satisfy $\|\underline{a}\| = \|\underline{b}\|$. Prove that $\underline{a}+\underline{b}$ and $\underline{a}-\underline{b}$ are orthogonal. Give a geometrical interpretation.

8 Determine all vectors $\underline{u}$ in the inner product space V such that $(\underline{u},\underline{x}) = 0$ for all $\underline{x} \in V$. [Hint: take $\underline{x} = \underline{u}$.]

§2

9 Determine the orthogonal complement of each of the following subspaces of $\mathbb{R}^3$:

a. $<(1,1,2),(0,1,0)>$.

b. $<(1,1,2),(0,1,0),(4,5,8)>$.

c. $<(3,1,0),(2,1,-4),(5,1,4)>$.

d. $\{(x_1,x_2,x_3) \mid 2x_1 - x_2 = 5x_3\}$.

10 Consider the following vectors in $\mathbb{R}^4$

$$\underline{a} = (4,7,-6,1),\ \underline{b} = (2,6,2,-2),\ \underline{c} = (-2,-1,8,-3),\ \underline{d} = (-2,-6,-2,2).$$

Determine $<\underline{a},\underline{b},\underline{c},\underline{d}>^{\perp}$.

11 a. Let $W =<(1,0,1,2),(1,1,0,1)>$ be a subspace of $\mathbb{R}^4$.

i) Determine a basis of $W^{\perp}$.

ii) Determine $\underline{v} \in W$ and $\underline{w} \in W^{\perp}$ such that $\underline{v}+\underline{w} = (2,2,3,2)$.

b. Let $W =<(1,1,-1,0),(2,1,-1,-1),(1,-2,0,3)>$ be a subspace of $\mathbb{R}^4$.

i) Determine a basis of $W^{\perp}$.

ii) Determine $\underline{v} \in W$ and $\underline{w} \in W^{\perp}$ such that $\underline{v}+\underline{w} = (3,6,-1,3)$.

12 In $\mathbb{R}^4$, let $\underline{a} = (1,2,-2,0)$, $\underline{b} = (2,1,0,4)$, $\underline{c} = (5,7,3,2)$. Determine the orthogonal projection of $\underline{c}$ on $<\underline{a},\underline{b}>^{\perp}$.

13 In $\mathbb{R}^5$ let $U =<(1,0,1,0,1),(1,1,1,1,1)>$.

a. Give a basis of $U^{\perp}$.

b. Decompose the vector $(3,2,2,0,1)$ in a component in U and a component in $U^\perp$, i.e., find $\underline{u} \in U$ and $\underline{v} \in U^\perp$ such that $(3,2,2,0,1) = \underline{u} + \underline{v}$.

14 a. In $\mathbb{R}^3$ let $V = \{(x,y,z) \mid x+y+z=0\}$. Determine orthonormal bases of V and $V^\perp$. Determine the orthogonal projection of $(1,2,1)$ onto V.

b. Let $l = \{(x,y,z) \mid x+y=0,\ x+y+2z=0\}$ be a line in $\mathbb{R}^3$. Determine orthonormal bases of l and of $l^\perp$.

c. Let

$$W =< (1,0,0,-1), (1,1,-1,0), (0,1,0,-2) >$$

be a subspace of $\mathbb{R}^4$. Determine an orthonormal basis of W and one of $W^\perp$.

15 Let $W =< (1,2,2,4), (3,2,2,1), (1,-2,-8,-4) >$ be a subspace of $\mathbb{R}^4$.

a. Determine an orthonormal basis of W.

b. Extend this orthonormal basis to one of $\mathbb{R}^4$. What are the coordinate vectors of $(1,1,1,1)$ and $(4,4,4,5)$, respectively, with respect to this basis?

16 Let $\underline{a}_1 = (1,1,1,1)$, $\underline{a}_2 = (3,3,-1,-1)$, $\underline{a}_3 = (7,9,3,5)$ be vectors in $\mathbb{R}^4$. Consider the subspaces $W_1 =< \underline{a}_1 >$, $W_2 =< \underline{a}_1, \underline{a}_2 >$, $W_3 =< \underline{a}_1, \underline{a}_2, \underline{a}_3 >$.

a. Determine orthonormal bases of W_1, W_2, W_3, respectively.

b. Determine the orthogonal projection of $(1,2,1,2)$ onto W_2.

c. Extend the basis of W_3 to an orthonormal basis of $\mathbb{R}^4$. What are the coordinates of the vector $(1,2,1,2)$ with respect to this basis?

17 Let

$$A_1 =< (1,1,1,2,1) >^\perp \quad \text{and} \quad A_2 =< (2,2,3,6,2) >^\perp$$

be subspaces in $\mathbb{R}^5$. Determine an orthonormal basis of $A_1 \cap A_2$.

§3

18 Determine the QR-decomposition of the matrix with columns $(1,1,1,1)$, $(3,3,-1,-1)$, $(7,9,3,5)$ (see exercise 16).

6.5.1 Exercises from old exams

19 In $\mathbb{R}^4$ with the standard inner product let $U =< (1,1,1,1), (1,0,2,-1) >$.

a. Determine an orthonormal basis of U.

b. Determine the orthogonal projection of $(-1,4,4,-1)$ onto U.

c. What is the distance of $\underline{a} = (-1,4,4,-1)$ to U, i.e. the minimum of the distances of $\underline{a}$ to vectors from U?

20 In $\mathbb{R}^4$ let $A_1 : (\underline{a}_1, \underline{x}) = 0$ and $A_2 : (\underline{a}_2, \underline{x}) = 0$, where $\underline{a}_1 = (1,1,1,1)$ and $\underline{a}_2 = (1,-2,2,4)$.

a. Determine the angle between the vectors $\underline{a}_1$ and $\underline{a}_2$.

b. Determine a parametric representation for $A_1 \cap A_2$, i.e. the collection of vectors that belong to both A_1 and A_2.

21 Let $\underline{a}$, $\underline{b}$, $\underline{c}$ be vectors in the real inner product space V such that

$$(\underline{a},\underline{a}) = (\underline{b},\underline{b}) = (\underline{c},\underline{c}) = 1,\ (\underline{a},\underline{b}) = \frac{1}{2},\ (\underline{a},\underline{c}) = 0,\ (\underline{b},\underline{c}) = 0.$$

a. Show that $\{\underline{a},\underline{b},\underline{c}\}$ is a linearly independent set of vectors.

b. Determine the orthogonal projection of $\underline{a}+\underline{b}+\underline{c}$ onto $< \underline{a} >$.

c. Determine the orthogonal projection of $\underline{a}+\underline{b}+\underline{c}$ onto $< \underline{a}, \underline{b} >$.

Appendix A

Prerequisites and some words on proving

A.1 Sets

Sets consist of elements. The way to denote that a is an element of the set A (or belongs to A) is as follows:

$$a \in A.$$

By definition, two sets A and B are the same, $A = B$, if they contain exactly the same elements. We usually describe sets in one of the following ways:

- **Enumeration of a set's elements between curly brackets**. For example:

  $$\{1, 2, 3, 5\}, \quad \{1, 2, 3, \ldots\}, \quad \{1, 2, 3, 5, 3\}, \quad \{2, \sqrt{3}, x^2 - 1\}.$$

  The dots in the second example indicate that the reader is expected to recognize the pattern and knows that 4, 5, etc., also belong to the set. The first and third sets are equal: the order in which the elements are listed and repetitions of elements are unimportant.

  In mathematics $(a, 2, \pi)$ denotes an *ordered list* in which the order of the elements and repetitions do matter. So $(1, 2, 3)$, $(1, 2, 2, 3)$, $(1, 3, 2, 2)$ are all distinct lists. We use such lists in this course mainly in the setting of coordinates.

- **Description of a set using defining properties**. Examples:

  $$\{x \mid x \text{ is an even integer}\}, \quad \{y \mid y \text{ is real and } y < 0\}.$$

We also write

$$\{x \in \mathbf{Z} \mid x \text{ even}\}, \quad \{y \in \mathbf{R} \mid y < 0\},$$

so that it is immediately clear in which set we are working.

Here is a list with often used notations regarding sets.

$\emptyset$	the empty set
$a \notin A$	a is *not* an element of A
$A \subset B$ (ook: $B \supset A$) (or $A \subseteq B$)	A is a subset of B (or: A is contained in B) i.e., for every $a \in A$ we have $a \in B$
$A \not\subset B$	A is not a subset of B
$A \cap B := \{x \mid x \in A \text{ and } x \in B\}$	the *intersection* of A and B
$A \cup B := \{x \mid x \in A \text{ or } x \in B\}$	the *union* of A and B
$A - B := \{x \mid x \in A \text{ and } x \notin B\}$ (or: $A \setminus B$)	the (set)*difference* of A and B
$A \times B := \{(a, b) \mid a \in A,\, b \in B\}$	the (cartesian) *product* of A and B
$A_1 \times A_2 \times \cdots \times A_n := \{(a_1, a_2, \ldots, a_n) \mid a_1 \in A_1, a_2 \in A_2, \ldots, a_n \in A_n\}$	the *product* of $A_1, A_2, \ldots, A_n$
$A^n := \{(a_1, a_2, \ldots, a_n) \mid a_1, \ldots, a_n \in A\}$	special case

To prove that two sets A and B are equal an often used strategy is to prove the following statements separately: $A \subset B$ and $B \subset A$.

A.2 Maps

If A and B are sets, then a *map* f from A to B is a rule that assigns to every element a of A an element $f(a)$ of B, called the image of a under f. Notation: $f : A \to B$. The set A is called the *domain* of the map, the set B is called the *codomain*. If B is a set of numbers, the term *function* is often used instead of the term map. In the setting of vector spaces the term *transformation* is often used. Two maps are the same if they have the same domain, codomain, and if they assign to every element of the domain the same image. The set of all images $f(a)$ is called the *range* of the map f. In set notation: $\{f(a) \mid a \in A\}$ or $\{b \in B \mid \exists a \in A[b = f(a)]\}$.

Some more notions and notations regarding maps:

$f : A \to B$	map with domain A and codomain B (other letters are also allowed!)
$f(a)$	the image of a
$f : a \mapsto b$	f assigns b to a or: a is mapped to b, or: f maps a to b
$f(D) := \{f(d) \mid d \in D\}$	the *image* of D, where D is a subset of A
$f(A)$	special case: the *image* of f
$f^{-1}(E) := \{a \in A \mid f(a) \in E\}$ (or: $f^{\leftarrow}(E)$)	the *preimage* of E (E a subset of B)
$f^{-1}(b)$ instead of $f^{-1}(\{b\})$	keeps the notation simple
$f : A \to B$ *injective* (f is an *injective map*)	for every $a, a' \in A$: $f(a) = f(a') \Rightarrow a = a'$; or: for every $a, a' \in A$: if $a \neq a'$ then $f(a) \neq f(a')$
$f : A \to B$ *surjective* (f is a *surjective map*)	for every b there is an $a \in A$ with $f(a) = b$, i.e., $f(A) = B$
$f : A \to B$ *bijective* (f is a *bijection*)	f is injective and surjective (so: for every $b \in B$ there is exactly one $a \in A$ with $f(a) = b$)

If $f : A \to B$ is a bijection, then for every $b \in B$ there is exactly one $a \in A$ with $f(a) = b$. Therefore, we can define a map from B to A by: $b \mapsto a$ if $f(a) = b$. This map (which only exists if f is a bijection) is called the *inverse* of f and is denoted by f^{-1}. (Be careful: this symbol is used for different notions.)

A.3 Some trigonometric relations

Here are some trigonometric relations (x and y are arbitrary real numbers):

- $\cos^2(x) + \sin^2(x) = 1$;
- $\sin(x + 2\pi) = \sin(x)$ and $\cos(x + 2\pi) = \cos(x)$;
- $\sin(\pi - x) = \sin(x)$ and $\cos(\pi - x) = -\cos(x)$;
- $\sin(\pi + x) = -\sin(x)$ and $\cos(\pi + x) = -\cos(x)$;
- $\sin(\pi/2 - x) = \cos(x)$ and $\cos(\pi/2 - x) = \sin(x)$;
- $\sin(2x) = 2\sin(x)\cos(x)$ and $\cos(2x) = \cos^2(x) - \sin^2(x)$;
- $\sin(x + y) = \sin(x)\cos(y) + \cos(x)\sin(y)$ and $\cos(x + y) = \cos(x)\cos(y) - \sin(x)\sin(y)$.

A.4 The Greek alphabet

In mathematics we often use Greek symbols. Below is the Greek alphabet, where we have added a * to symbols that occur frequently in Linear algebra A and B.

name	**lower case**	**upper case**
alpha	α *	A
beta	β *	B
gamma	γ *	Γ
delta	δ *	Δ
epsilon	ε	E
zeta	ζ	Z
eta	η	H
theta	θ of ϑ	Θ
iota	ι	I
kappa	κ	K
lambda	λ *	Λ
mu	μ *	M
nu	ν	N
xi	ξ	Ξ
omikron	o	O
pi	π	Π
rho	ρ *	R
sigma	σ *	Σ
tau	τ *	T
upsilon	υ	Υ
phi	ϕ of φ *	Φ
chi	χ	X
psi	ψ *	Ψ
omega	ω *	Ω

A.5 Proving

A typical aspect of mathematics is its aim to structure (many of) its results in the form of mathematical statements (usually, so-called theorems, propositions, lemma's, and corollaries, see below) that are being accompanied by mathematical proofs, i.e., precise argumentations why the statements are true. These argumentations involve logic, mathematical concepts, previous results, and language (the latter to make them readable for humans; one can also construct more formal proofs that are better suited for computers).

In the various courses you will be taking in the mathematics program, mathematical proofs play an important role. Below we present some advice related to proving. The course 2WF40 *Set theory and algebra aims* at discussing this in much more detail.

- **Terminology**. The most common ways in which you will encounter mathematical results is in one of the following forms.
  - A *theorem*: A mathematical statement that is known to be true since a proof exists. For instance, the Pythagorean theorem in plane geometry. It usually has the form: *If such and such conditions hold, then such and such is the case.* For example: *If in triangle ABC the angle $\angle ABC$ is a right angle, then $AB^2 + BC^2 = AC^2$.*
  - A *proposition*: Similar as a theorem, but regarded as somewhat less significant than a theorem. For instance, *the product of any two even integers is an integer* could be an example of such a mathematical statement. Minor results are sometimes not even given a name like proposition, but are, for instance, just mentioned (and proved) in the running text.
  - A *lemma*: a mathematical statement that mainly serves as an aid in proving a theorem or proposition.
  - A *corollary*: a mathematical result that is an immediate consequence of a proposition or theorem.

  In mathematics, we also use *definitions* to make the precise meaning of the various concepts used clear. For instance, a definition specifies the meaning of even and odd integers (given that we know what integers and multiplication mean): *an even integer is an integer that can be written in the form $2m$, where m itself is an integer; an odd integer is an integer that can be written in the form $2m + 1$, where m is an integer.* Here is a definition of prime numbers: *a prime number is an integer which is greater then 1 and whose only positive divisors are 1 and the integer itself.* Of course, a previous definition should explain the notion of a *divisor*, etc.

- **Quantifiers**. We often come across statements like *there exists...* or *for all...*. Instead of words, we sometimes also use symbols for them. The most common ones are:
  - $\exists$: this is the symbol for *there exist(s)*. For instance, $\exists a \in \mathbb{R}, a > 0$.
  - $\forall$: this is the symbol for *for all*. For instance, $\forall a \in \mathbb{R}, a^2 \geq 0$.

They are mostly used if they occur together, like in: *for all* $z \in \mathbb{C}$ *there exists a* $w \in \mathbb{C}$ *such that* $z = w^2$. With symbols:

$$\forall z \in \mathbb{C}, \exists w \in \mathbb{C}, z = w^2.$$

Note that in the mathematical literature, slightly different notations also occur, like

$$\forall z \in \mathbb{C}, \exists w \in \mathbb{C}\,[\, z = w^2\,].$$

- **Types of proof**. Although there is no clear list of types of proofs, some (parts of) proofs are clearly of a recognizable type. We review a few of them. Of course, it takes experience to find out how to build a proof in any given situation.

  - **Proof by computation**. The proof (or part of it) consists mainly of a computation. For instance, the proof of the statement *for all real numbers* a *and* b *we have* $a^2+2ab+2b^2 \geq 0$ could run as follows: Let a and b be arbitrary real numbers. Now rewrite $a^2+2ab+2b^2$ as follows:

    $$a^2 + 2ab + 2b^2 = a^2 + 2ab + b^2 + b^2 = (a + b)^2 + b^2.$$

    Since squares of real numbers are non-negative and sums of non-negative numbers are again non-negative, we conclude $(a+b)^2 + b^2 \geq 0$. Hence $a^2+2ab+2b^2 \geq 0$. (Of course, the proof uses various properties of real numbers that need proofs themselves.)

  - **Proof by mathematical induction**. It usually concerns a family of statements $P(n)$ for all positive integers or all non-negative integers. For instance, suppose $P(n)$ is the statement: for every positive integer n the sum $1 + 2 + \cdots + n$ equals $\frac{1}{2}n(n + 1)$. A proof by induction first handles the statement $P(1)$, the so-called base case. The next step, the so-called induction step, is to prove that, if $P(n)$ is true for a fixed but arbitrary n (the so-called induction hypothesis), then $P(n + 1)$ is also true. These two steps together suffice to prove the original statement, the idea being that $P(1)$ is true (first step), but then also $P(2)$ (special case of the second step), and then also $P(3)$, etc.

    In the case at hand, the base case $P(1)$ is simple to verify: we need to verify that 1 equals $\frac{1}{2} \cdot 1 \cdot 2$, which is clearly true.

    For the induction step: let n be an arbitrary positive integer and assume that $P(n)$ is true, so that $1 + 2 + \cdots + n = \frac{1}{2}n(n + 1)$. Now consider $1 + 2 + \cdots + n + (n + 1)$. We need to prove that this sum equals $\frac{1}{2}(n+1)(n+2)$. To do so, rewrite the sum as follows, where the first equality sign is based on the induction hypothesis:

$$\begin{aligned}(1+2+\cdots+n)+(n+1) &= \tfrac{1}{2}n(n+1)+(n+1)\\ &= \tfrac{1}{2}(n^2+n+2n+2)\\ &= \tfrac{1}{2}(n+1)(n+2).\end{aligned}$$

So $P(n+1)$ is also true and we are done by induction.

– **Proof by contradiction**. In this case you assume the result to be proven false, and then derive a contradiction. You then conclude that assuming the result false is wrong, so the result is actually true.

  A famous proof by contradiction concerns the irrationality of $\sqrt{2}$ (this means: $\sqrt{2}$ cannot be written as the quotient of two integers). In the following proof we assume some elementary properties of the integers, like squares of even (odd) integers are even (odd). So here is the proof. We argue by contradiction. Assume the statement is false, so that there do exist two integers a and b such that $\sqrt{2} = a/b$. We may assume that both integers are positive (if they are both negative, then change the signs of both of them), and we may assume they are not both even (if they were both even, then divide a and b by 2 or if needed a sufficiently higher power of 2). Next, we aim for a contradiction. To do so, first rewrite $\sqrt{2} = a/b$ as $2b^2 = a^2$. Since the expression on the left-hand side is even, the expression on the right-hand side is even and hence a must be even, say $a = 2a'$ for some integer a', and so b must be odd. Using $a = 2a'$ we obtain $2b^2 = 4(a')^2$ so that $b^2 = 2(a')^2$. We now conclude that b is even. So b is both even and odd, clearly a contradiction. Hence our original assumption that $\sqrt{2}$ can be expressed as a fraction is false.

– **Proof by distinguishing cases**. Here you split the proof in a number of cases (that together cover everything). A simple example is the following. Consider the statement: Let x, y be real numbers. If $xy = 0$ then $x = 0$ or $y = 0$.

  To prove this statement we distinguish the case that $x = 0$ and the case $x \neq 0$. Here is the proof: Let x, y be real numbers that satisfy $xy = 0$. **Case 1**: $x = 0$. Then of course we are done. **Case 2**: $x \neq 0$. In this case we multiply $xy = 0$ on both sides by $1/x$ to obtain $\frac{1}{x}\cdot(xy) = \frac{1}{x}\cdot 0 = 0$. Rewriting $\frac{1}{x}\cdot(xy)$ as $(\frac{1}{x}\cdot x)\cdot y = y$ we get $y = 0$ and we are done in this second case. Since both cases cover all possibilities, this finishes the proof.

Some advice on finding proofs.

- Make sure you know what is meant by the statement to be proven.
- To get a good idea of the result, first try to consider some special cases if possible.
- Look for results, computational techniques, etc., that may have something to do with the statement.
- Maybe a sketch of the situation or a scheme of related results and techniques can help understand the situation.
- Consider which type of proof might apply.

Some advice on writing proofs.

- Write down what is given and what you need to prove.
- Use sentences, mixed with mathematical expressions.
- Indicate the strategy of the proof. For instance: We will prove this statement by mathematical induction. Or: We will argue by contradiction. Keep in mind that you help the reader follow your argument.
- If you want to prove a statement of the form *for all* $a \in A$ ..., start the proof for instance by (possibly after some introductory remarks): *Let* $a \in A$. Or: *Let* a *be an arbitrary element of* A.
- If you want to prove that there exists an element in a set with a certain property, then indicate the element and proceed to show that it has the desired property. For instance: *Choose* a *such that...*
- A proof of a statement of the form *if* A *then* B (in symbols: $A \Rightarrow B$), starts for instance as follows: *Assume* A. (Then continue to show B.)
  * For instance, to prove the statement *If* a *is an even integer, then* a^2 *is an even integer*, you could proceed as follows: Assume a is an even integer. Since a is even there exists an integer m such that $a = 2m$ (definition of even integers). Square both sides to obtain $a^2 = 4m^2$. Rewrite $4m^2$ as $2 \cdot 2m^2$. Since $2m^2$ is an integer, we have shown that a^2 is an integer.
- If you want to prove a statement of the form $A \Leftrightarrow B$, i.e., A implies B *and* B implies A, make sure both implications are taken care of. A place where this can easily go wrong is in dealing with equations. If you solve an equation by rewriting it in several steps, then make sure the rewriting can be 'undone' (or build in a check of the answers at the end). The following example (in the real numbers) shows what can go

wrong.

$$\begin{aligned}
&-\sqrt{x-4} = x-4\\
&x-4 = (x-4)^2\\
&x^2-9x+20=0\\
&(x-5)(x-4)=0\\
&x=4 \text{ or } x=5
\end{aligned}$$

You cannot go back from the second step to the first step, so these two statements are not equivalent. If x satisfies $\sqrt{x-4} = x-4$, then it satisfies $x-4=(x-4)^2$, but not conversely.

- Sometimes people end proofs by writing q.e.d. (from the Latin quod erat demonstrandum, what was to be proved) or the symbol $\Box$.

Counterexamples

A counterexample to a mathematical statement is a concrete example that shows that the mathematical statement does not hold in this particular case *and hence is not true in general.* For instance, the statement

every integer of the form p^2-1, where p is a prime number, is not prime

is false since for $p=2$, the integer $2^2-1=3$ is prime. (However, the statement *every integer of the form p^2-1, where p is a prime greater than 2, is not prime*, is true. Can you prove it?

If you cannot find a counterexample, that doesn't necessarily mean of course that the statement is false (you may simply have failed to find a counterexample). Counterexamples are often used to sharpen the conditions in a statement so that hopefully a sensible true statement remains. But it may also serve to show that no reasonable mathematical result is to be expected.

Appendix B

Answers to most of the exercises

Chapter 1: Complex numbers

1. a) 5+i d) $\frac{9}{5} - \frac{13}{5}i$
   b) 1 e) $-3i$
   c) $\frac{4}{25} + \frac{3}{25}i$ f) $1 - \frac{\sqrt{2}}{2}$

2. a) $r = 3, \quad \varphi = \pi$ d) $r = 2, \quad \varphi = \frac{\pi}{6}$
   b) $r = 2, \quad \varphi = \frac{\pi}{2}$ e) $r = 13, \quad \varphi = \arctan\frac{12}{5} = 2\arctan\frac{2}{3}$
   c) $r = \sqrt{2}, \quad \varphi = \frac{\pi}{4}$ f) $r = 4\sqrt{2}, \quad \varphi = -\frac{\pi}{4}$

5. a) $\mathrm{Im}(z) = -1$, the perpendicular bisector of i and $-3i$
   b) perpendicular bisector of $3i$ and $4 + 2i$
   c) all z satisfying $\arg(z) = \frac{\pi}{8} + k\pi$ or $\arg(z) = \frac{5\pi}{8} + k\pi, \ k = 0, -1$
   d) $z_1 = +\frac{1}{2}\sqrt{2} + \frac{1}{2}i\sqrt{6}$, $z_2 = -\frac{1}{2}\sqrt{2} + \frac{1}{2}i\sqrt{6}$,
   $z_3 = +\frac{1}{2}\sqrt{2} - \frac{1}{2}i\sqrt{6}$, $z_4 = -\frac{1}{2}\sqrt{2} - \frac{1}{2}i\sqrt{6}$,
   e) alle z met $\arg(z) = \frac{1}{3}\pi, \ |z| \neq 0$ en $\arg(z) = \frac{4}{3}\pi, \ |z| \neq 0$.

7. a) $2i$ c) $1 + i$ e) $\mathrm{e}^3(\frac{1}{2} - \frac{1}{2}i\sqrt{3})$
   b) $-\frac{3}{2} + \frac{3}{2}i\sqrt{3}$ d) $\frac{1}{2} - \frac{1}{2}i\sqrt{3}$ f) $-\frac{1}{2}\sqrt{3} - \frac{1}{2}i$

8. a) $z = \frac{1}{2}\ln 2 + (\frac{\pi}{4} + 2k\pi)i,\ k \in \mathbb{Z}$
   b) $z = \ln 2 + (\frac{\pi}{3} + 2k\pi)i,\ k \in \mathbb{Z}$
   c) $\mathrm{Re}(z) = \ln 5,\ \mathrm{Im}(z)$ arbitrary
   d) $z = 0$
   e) $z = \pm\frac{1}{2}\sqrt{4k+1}\,\sqrt{\pi}(1+i),\ k = 0, 1, \cdots$
   $z = \pm\frac{1}{2}\sqrt{4k-1}\,\sqrt{\pi}(1-i),\ k = 1, 2, \cdots$
   f) $z = \frac{\pi}{4} + k\pi,\ k \in \mathbb{Z}$

10. a) $z = \frac{\pi}{2} + k\pi,\ k \in \mathbb{Z}$
    b) $z = x + iy$ met $x = \frac{\pi}{4} + k\pi;\ y = -\frac{1}{2}\ln(4 \pm \sqrt{15}),$
    $k \in \mathbb{Z}$

11. a) $z_1 = 1 \quad z_2 = \frac{1}{2}(1 + i\sqrt{3}) \quad z_3 = \frac{1}{2}(-1 + i\sqrt{3})$
    $z_4 = -1 \quad z_5 = \frac{1}{2}(-1 - i\sqrt{3}) \quad z_6 = \frac{1}{2}(1 - i\sqrt{3})$
    b) $z_1 = 2:\ z_{2,3} = -(1 \pm i\sqrt{3})$
    c) $z_1 = 2(\cos\frac{\pi}{8} + i\sin\frac{\pi}{8}) \quad z_2 = 2(-\sin\frac{\pi}{8} + i\cos\frac{\pi}{8})$
    $z_3 = 2(-\cos\frac{\pi}{8} - i\sin\frac{\pi}{8}) \quad z_4 = 2(\sin\frac{\pi}{8} - i\cos\frac{\pi}{8})$
    d) $z_1 = \frac{1}{2}\sqrt{2} + (\frac{1}{2}\sqrt{2} - 1)i \quad z_2 = -\frac{1}{2}\sqrt{2} + (\frac{1}{2}\sqrt{2} - 1)i$
    $z_3 = \frac{1}{2}\sqrt{2} - (\frac{1}{2}\sqrt{2} + 1)i \quad z_4 = -\frac{1}{2}\sqrt{2} - (\frac{1}{2}\sqrt{2} + 1)i$
    e) $z_k = i - 2 + \sqrt{3}(\cos(\frac{\pi}{12} + k\frac{\pi}{3}) + i\sin(\frac{\pi}{12} + k\frac{\pi}{3})),$
    $k = 0, 1, \cdots, 5.$
    f) $z_1 = 0;\ z_2 = 1;\ z_{3,4} = -\frac{1}{2}(1 \pm \sqrt{3}i)$
    g) $z = 0$ and $z = \mathrm{e}^{i\varphi}$ met $\varphi = (\frac{1}{4} + \frac{1}{2}k)\pi,\ k \in \mathbb{Z}$ (d.w.z. $\pm\frac{\sqrt{2}}{2} \pm i\frac{\sqrt{2}}{2}$)

12. a) $z = -\frac{1}{2} \pm \frac{1}{2}i\sqrt{3}$
    b) $z_1 = -2i$ and $z_2 = 4i$
    c) $z_{1,2} = 2 + i$
    d) $z_{1,2} = \pm\frac{1}{2}\sqrt{6}(1 - i);\ z_{3,4} = \pm(1 + i)$

13. a) $z_2 = 2i;\ z_3 = -2$
    b) $z_2 = 1 - i;\ z_{3,4} = -3 \pm 2i$
    c) $z^3 - 7z^2 + 15z - 25$
    d) $z^4 - 4z^3 + 14z^2 - 4z + 13$

14. a) $(z-2)(z+2)(z^2+1)$
    b) $(z+1)(z^2+2z+2)$
    c) $(z^2+1)(z^2+z+1)$

15. a) $-32 + 32i$
    b) $|z^{23}| = 2^{23},\ \arg(z^{23}) = \frac{11}{24}\pi$

16. $\cos^3\varphi - 3\sin^2\varphi\cos\varphi$
    $4\cos^3\varphi\sin\varphi - 4\cos\varphi\sin^3\varphi$

18. a. If $z = x + iy$, then $\overline{z} = x - iy$. If z is real, then both $z = x$ and $\overline{z} = x$. Conversely, if $x + iy = x - iy$, then $y = 0$, so z is real.

    b. Proof is similar to that of a)

    c. z and w are parallel if and only if z/w is real. Then use a).

19. a. $z = x + iy$ is mapped to $-x + iy$ which is $-\overline{z}$.

    b. $z \mapsto e^{2i\alpha} \cdot \overline{z}$

20. a. Without loss of generality assume the vertices are 0, z and ρz with $|\rho| = 1$. Then $|z| = |z - \rho z| = |1 - \rho| \cdot |z|$ so that $|1 - \rho| = 1$. From $|\rho| = |1 - \rho| = 1$ you obtain via $\rho = a + bi$ that $\rho = \dfrac{1}{2} \pm i\dfrac{\sqrt{3}}{2}$.

21. a. From $\dfrac{z - w}{z - v} = t$ with t real (and $\neq 1$) you get $z - w = t(z - v)$ so that $(1 - t)z = w - tv$. Then $z = \dfrac{1}{1 - t}\,w - \dfrac{t}{1 - t}\,v$, so $z = uw + (1 - u)v = v + u(w - v)$ with $u = 1/(1 - t)$ real.

23. a. $re^{-it} \cdot v$

25. $(1/2) + bi$, $b \in \mathbb{R}$

26. $(\pi/4) + k\pi$ with $k \in \mathbb{Z}$

27. b) $(5/2) - i(5/2)$, $(1/2) + i(1/2)$.

28. Factorization: $(z^2 - 2z + 5)(z^2 + 4)$; zeros: $1 \pm 2i$, $\pm 2i$

30. Solutions are:

$$0,\ \cos(\pi/8) + i\sin(\pi/8),\ \cos(5\pi/8) + i\sin(5\pi/8),$$
$$\cos(9\pi/8) + i\sin(9\pi/8),\ \cos(13\pi/8) + i\sin(13\pi/8).$$

    (Solutions can also be given in exponential notation, of course.)

31. Put vertex A of square $ABCD$ in the origin, then the vertices of the square can be described as follows: 0, w, $(1 + i)w$, iw (check!). Now $A'B'C'D'$ is also of this form apart from a translation, so: $u, z + u, z(1 + i) + u, zi + u$. The midpoints (times 2 for computational convenience) are u, $w + z + u$, $(1 + i)(w + z) + u, i(w + z) + u$. Apart from a translation over u we get 0, $w + z$, $(1 + i)(w + z)$, $i(w + z)$.

Chapter 2: Vector geometry in dimensions 2 and 3

3. b) All
   c) The vector is on the line: take $\lambda = 3$.

4. b) all

5. a) $0 \leq \lambda \leq 1$
   b) $\lambda = 1/2$
   c) $\lambda = 2/3$

6. a) $\underline{x} = (2, 1, 5) + \lambda(3, -2, -1)$
   b) $\underline{x} = \lambda(1, 2)$
   c) $\underline{x} = (1, 2, 2) + \lambda(1, 1, 1) + \mu(0, 1, 0)$
   d) $\underline{x} = (-2, 1, 3) + \lambda(1, 2, -1) + \mu(6, -1, 0)$

7. Yes; yes.

8. a) $x + 2y = 7$
   b) $x + y = 4$
   c) $x = 3$

9. a) $\underline{x} = (0, 1) + \lambda(3, -2)$
   b) $\underline{x} = (-1, 1) + \lambda(4, 3)$
   c) $\underline{x} = (0, 5/2) + \lambda(1, 0)$

10. a) $2x + 2y - z = 3$
    b) $x - y + z = 1$
    c) $x - 2y - 2z = 0$

11. a) $\underline{x} = (5, 0, 0) + \lambda(1, -1, 0) + \mu(3, 0, 1)$
    b) $\underline{x} = \lambda(3, -2, 0) + \mu(0, 5, -3)$
    c) $\underline{x} = (0, 5, 0) + \lambda(1, 0, 0) + \mu(0, 0, 1)$

14. a) 3
    b) 5
    c) $\pi/2$ radians
    d) $a = 1$

15. a) $x_1 - x_2 = 1$; $\sqrt{2}$
    b) $4x_1 + 3x_2 = 10$; 5

16. a) 6
    b) 2

17. a) normal vector $(-1,-1,1)$; $-x_1 - x_2 + x_3 = -1$
    b) normal vector $(6,-3,2)$; $6x_1 - 3x_2 + 2x_3 = 9$

18. a) $\sqrt{17}/2$
    b) $13/2$
    c) 2

20. b) A median is the line through a vertex of $ABCD$ and the midpoint of the opposite site.

23. a) intersection point $(0,-8,8)$
    b) $\underline{x} = (0,-8,8) + \lambda(1,7,-8)$

24. $\underline{q} = \frac{1}{3} \cdot \underline{c}$ where the origin is in A.

25. b) 3
    c) $\underline{x} = (2,0,4) + \sigma(1,-2,2)$

26. b) $\underline{x} = (2,-1,-1) + \rho(1,1,-1)$

Chapter 3: Linear equations

1. $AB = \begin{pmatrix} -2 & 7 \\ -12 & 6 \end{pmatrix}$; $\quad BA = \begin{pmatrix} 4 & -7 & 4 \\ 0 & -6 & -8 \\ -3 & 12 & 6 \end{pmatrix}$;

   $A(B-2C) = \begin{pmatrix} -8 & 13 \\ -26 & 36 \end{pmatrix}$; $\quad AD = \begin{pmatrix} 9 \\ 9 \end{pmatrix}$;

   $CC^T = \begin{pmatrix} 8 & 0 & -4 \\ 0 & 2 & 4 \\ -4 & 4 & 10 \end{pmatrix}$; $\quad C^TC = \begin{pmatrix} 6 & 0 \\ 0 & 14 \end{pmatrix}$;

   $DD^T = \begin{pmatrix} 4 & -2 & 6 \\ -2 & 1 & -3 \\ 6 & -3 & 9 \end{pmatrix}$; $\quad D^TD = 14$

2. $A + B = \begin{pmatrix} 2 & 0 & 0 \\ 2 & -2 & -2 \end{pmatrix}$; $\quad (A-B)C = \begin{pmatrix} 2+2i & 2i \\ 0 & 0 \end{pmatrix}$;

   $A^TB = \begin{pmatrix} 3 & -i & i \\ i & 2 & -2 \\ -i & -2 & 2 \end{pmatrix}$; $\quad AA^T = \begin{pmatrix} -2+2i & 1-i \\ 1-i & 3 \end{pmatrix}$;

   $A^TC = \begin{pmatrix} 3 & i & i \\ i & -2 & -2 \\ -i & 2 & 2 \end{pmatrix}$

3. $A+B = \begin{pmatrix} 2 & 2 & 2 \\ 4 & 4 & 4 \end{pmatrix}$; $\qquad A-2B = \begin{pmatrix} 2 & 5 & 8 \\ 1 & 4 & 7 \end{pmatrix}$;

   $A^TB = \begin{pmatrix} 3 & -2 & -7 \\ 4 & -3 & -10 \\ 5 & -4 & -13 \end{pmatrix}$; $\quad AB^T = \begin{pmatrix} -11 & -2 \\ -14 & -2 \end{pmatrix}$

4. $A+B = \begin{pmatrix} 2 & 2 \\ 4 & 4 \\ 6 & 6 \end{pmatrix}$; $\qquad A-B = \begin{pmatrix} 2i & 4i \\ 2i & 4i \\ 2i & 4i \end{pmatrix}$;

   $A^TB = \begin{pmatrix} 17 & 20-6i \\ 20+6i & 26 \end{pmatrix}$; $\quad AB^T = \begin{pmatrix} 7 & 9+3i & 11+6i \\ 9-3i & 13 & 17+3i \\ 11-6i & 17-3i & 23 \end{pmatrix}$

5. a) $\frac{1}{2}\begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix}$

   b) $\frac{1}{\lambda}A^{-1}$; $(A^{-1})^2$; $(A^{-1})^T$; A

7. a) $\begin{pmatrix} 1 & 0 & 0 & 2 \\ 0 & 1 & 0 & 4 \\ 0 & 0 & 1 & 7 \end{pmatrix}$ $\quad$ b) $\begin{pmatrix} 1 & 2 & 0 & 0 & 2 \\ 0 & 0 & 1 & 1 & 3 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$

8. multiply all from the left by:

   a) $\begin{array}{cc} & \begin{array}{cccccc} & & i & j & & m \end{array} \\ \begin{array}{c} \\ \\ i \\ j \\ \\ m \end{array} & \begin{pmatrix} 1 & & & & & \\ & \ddots & & & & \\ & & 0 & 1 & & \\ & & 1 & 0 & & \\ & & & & \ddots & \\ & & & & & 1 \end{pmatrix} \end{array}$ $\quad$ b) $\begin{array}{cc} & \begin{array}{ccccc} & & & & m \end{array} \\ \begin{array}{c} \\ \\ i \\ \\ m \end{array} & \begin{pmatrix} 1 & & & & \\ & \ddots & & & \\ & & \lambda & & \\ & & & \ddots & \\ & & & & 1 \end{pmatrix} \end{array}$ $\quad$ c) $\begin{array}{cc} & \begin{array}{cccccc} & & & j & & m \end{array} \\ \begin{array}{c} \\ \\ i \\ \\ \\ m \end{array} & \begin{pmatrix} 1 & & & & & \\ & \ddots & & & & \\ & & \lambda & 1 & & \\ & & & \ddots & & \\ & & & & \ddots & \\ & & & & & 1 \end{pmatrix} \end{array}$

9. a) $\underline{x} = \lambda(17, -13, 4, 3)$
   b) $\underline{x} = \lambda(3, 1, -5, 0) + \mu(0, 1, 0, 1)$
   c) $\underline{x} = (3, 1, 0, 0) + \lambda(1, 2, 1, 0) + \mu(7, 5, 0, -1)$

10. a) $\underline{x} = (1, -1, 1)$
    b) no solutions
    c) $\underline{x} = (0, 0, -2) + \lambda(-2, 1, 1)$
    d) no solutions

11. a) $\underline{z} = (-1+i, -1-i, 1)$
    b) $\underline{z} = \mu(2, -1-3i)$ and $\underline{z} = \mu(2, -1+3i)$
    c) $\underline{z}_1 = \mu(1, 1, -1)$; $\underline{z}_2 = \mu(1, \mathrm{e}^{\frac{2}{3}\pi i}, \mathrm{e}^{\frac{1}{3}\pi i})$; $\underline{z}_3 = \mu(1, \mathrm{e}^{-\frac{2}{3}\pi i}\mathrm{e}^{-\frac{1}{3}\pi i})$;

12. $\underline{z} = \frac{1}{2a+1}(2, 3, 1) = -\frac{1}{3}i\sqrt{3}(2, 3, 1)$

13. $\lambda = 1$: inconsistent; $\lambda \neq 1$: $(\dfrac{1}{1-\lambda}, 0, 3 - \dfrac{1}{1-\lambda}) + \mu(1, 1, -\lambda - 1)$

14. For $\lambda \neq 0$: $(x_1, x_2, x_3) = (\lambda, 2, -2)$; for $\lambda = 0$: $x_1 = 0$, $x_3 = -2$, $x_2 = \mu$ (arbitrary)

15. $\lambda \neq -1$: inconsistent system; for $\lambda = -1$: $(1, -1, 0, 0) + \alpha(1, -2, 1, 0) + \beta(-8, 7, 0, 1)$

Chapter 4: Vector spaces

1. No, yes, no, no

2. Yes, no, no, yes

3. Yes, yes

4. a), c), d), f): yes; b), e), g), h): no

5. a) $\underline{x} = (7, 0, 0) + \lambda(-4, 1, 0) + \mu(5, 0, 1)$
   b) $\underline{x} = \lambda(1, 0, -2) + \mu(0, 1, 4)$
   c) $\underline{x} = (\frac{7}{2}, 0, 0) + \lambda(-2, 1, 0) + \mu(-2, 0, 1)$

6. a) $x - y = 3$
   b) $3x - y - 3z = -1$
   c) $x - y + z = 3$

7. a) $\underline{x} = (0, 0, 3, -11) + \lambda(1, 0, -1, 4) + \mu(0, 1, -1, 5)$
   b) $\underline{x} = (0, 3, 0, 1) + \lambda(1, 0, 0, 0) + \mu(0, 2, -1, 0)$

8. a) $2x - 2y + u = 2$; $y = z$
   b) $y + z = 4$; $-x - y + u = 5$

9. $\underline{x} = (1, 3, 2)$

12. a), c), d): independent; b), e): dependent

13. a) $\{(1, 0), (0, 1)\}$
    b) $\{(0, 1, 0), (1, 0, 1)\}$
    c) $\{(1, 0, 0, 1), (0, 1, 0, 1), (0, 0, 1, 0)\}$
    d) $\{(0, 4, -1, -4), (1, 5, 0, -3)\}$
    e) $\{(3, -1, 4, 7), (1, -3, 2, 5), (5, 3, 2, -1)\}$
    f) $\{(1, 5, 0, 0), (0, 4, -1, 0), (0, 0, 0, 1)\}$

14. Yes, no, yes

15. No, yes, yes

17. b): dependent; a), c): independent

19. $\underline{x} = (1, -1, 0) + \lambda(2, 1, 1) + \mu(-1, 2, 1)$; $x + 3y - 5z = -2$

20. a) $a \neq \pm 2$
    b) $a = 3$

21. a) $< (-2, 3, 1) >$
    b) $< \underline{0} >$

22. a) basis $\{(1, 0, 0), (0, 1, 0), (0, 0, 1)\}$ $\dim W_1 = 3$
    b) basis $\{(i, 1 + 1, 0), (0, i, 1 + i)\}$ $\dim W_2 = 2$
    c) basis $\{(0, 1, -1), (1, 2, 0)\}$ $\dim W_3 = 2$

23. a) basis $\{\mathrm{e}^{2t}, t^2, t\}$ $\dim = 3$
    b) basis $\{t, \sin^2 t, \cos^2 t\}$ $\dim = 3$
    c) basis $\{\mathrm{e}^{2t}, \mathrm{e}^{-t}, \mathrm{e}^{t}\}$ $\dim = 3$
    d) basis $\{2x^3 + x^2 - x + 5, x^3 + 2x^2 + 10, -2x^2 + x\}$ $\dim = 3$

24. a) $\dim = 5$
    b) $\dim = 9$
    c) $\dim = 6$
    d) $\dim = 3$
    e) $\dim = 4$
    f) $\dim = 2$

25. a) $\dim = 2$
    b) $\dim = 3$

26. a) $(-1, 3)$
    b) $(2, -1, 2)$
    c) $(\frac{1}{2}, -\frac{1}{2}, \frac{1}{2})$
    d) $(\frac{1}{4}, -2)$

30. b) $c \neq -2$

31. a) 2; b) $\langle 2 + e^{3x} \rangle$

32. a) $\lambda(0, -1, 1, 0) + \mu(1, -1, 0, 1)$; b) $(1, -1, 0, 1) + \mu(0, -1, 1, 0)$

Chapter 5: Rank and Inverse, Determinants

1. a) row space: $< (1,1,1,1),(0,1,2,3) >$
   column space: $< (1,1),(0,1) >$

   b) row space: $< (1,1,0,1),(0,3,1,2) >$
   column space: $< (1,-1,-1),(0,1,3) >$

   c) row space: $< (0,1,1),(1,0,1) >$
   column space: $< (1,1+i,2+i),(0,2-i,-2i) >$

2. a) rank $= 2$; b) rank $= 2$; c) rank $= 4$; d) rank $= 3$

3. $$\begin{pmatrix} 2 & -1 \\ -1 & 1 \end{pmatrix} ; \begin{pmatrix} -2 & 0 & 0 & 1 \\ 9 & 3 & 2 & -6 \\ 6 & 2 & 1 & -4 \\ -16 & -3 & -2 & 10 \end{pmatrix} ; \begin{pmatrix} -2 & 3 & -1 \\ 0 & -3 & 2 \\ 1 & 1 & -1 \end{pmatrix} ; \begin{pmatrix} 1 & -3 & 2 \\ -3 & 3 & -1 \\ 2 & -1 & 0 \end{pmatrix}$$

4. a) $\frac{1}{2}\begin{pmatrix} -i & 1 \\ 1 & -i \end{pmatrix}$

   b) $\frac{1}{2}\begin{pmatrix} 1 & 1 & 0 \\ 0 & 0 & -2i \\ 1 & -1 & 0 \end{pmatrix}$

5. b) $\text{rank}(A) = \text{rank}(A|B)$

6. a) $\lambda^4 = 1 (\lambda = \pm 1 \vee \lambda = \pm i) : \text{rank} = 1, \ \text{anders} : \text{rank} = 2$

   b) $\lambda = 1 : \ \text{rank} = 1, \ \lambda \neq 1 : \ \text{rank} = 2$

7. a) 0; b) $-\pi e i$; c) -216; d) 8; e) 1; f) 4; g) -45; h) -4

8. a) -8; b) 0; c) $\pi - 1$; d) 1; e) 73; f) 484

9. 0; 2

10. $$|A| = \begin{cases} 2, & n = 1 \\ -1, & n = 2 \\ 0, & n \geq 3 \end{cases}$$

11. yes

12. a) $\det(A) \neq 0$

    b) $\det(A) = 0$ and $(2,-1,5) \in$ column space

    c) $\det(A) = 0$ and $(i,1,0) \notin$ column space

13. a) $y = 1$; b) $x = -6$

14. $\alpha = 0 \vee \beta = 0 \vee \gamma = 0$

15. b) $\det(A) = \pm 1$
    c) $\det(A) = 0$ for n odd

16. a) $\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$
    b) 0
    c) I
    d) $\det(A) = 0$ or $\det(A) = 1$

18. Voor $a \neq 0$ and $a \neq 3$:

$$\frac{1}{a^2 - 3a} \begin{pmatrix} -3a+1 & -a+3 & a^2-1 \\ -1 & a-3 & 1 \\ a & 0 & -a \end{pmatrix}$$

19. $a = 0$ and $a = 1$

20. a) for $\lambda \neq 1, -1$:

$$\frac{1}{1-\lambda^2} \begin{pmatrix} 1-2\lambda^2 & \lambda & \lambda^3-\lambda \\ \lambda & -1 & 1-\lambda^2 \\ 2\lambda & -2 & 1-\lambda^2 \end{pmatrix}$$

b) $\lambda = -1$

Chapter 6: Inner product spaces

1. a) yes
   b) no
   c) yes
   d) no

3. a) $\sqrt{11}$
   b) $\sqrt{29}$

4. a) $\frac{2}{3}\pi$
   b) $\frac{1}{3}\pi$
   c) $\frac{1}{4}\pi$
   d) $\frac{1}{3}\pi$

8. $\underline{u} = \underline{0}$

9. a) $< (2, 0, -1) >$
   b) $< (2, 0, -1) >$
   c) $< (0, 0, 0) >$
   d) $< (2, -1, -5) >$

10. $< (5,-2,1,0),(-2,1,0,1) >$

11. a) i) $< (1,-1,-1,0),(0,-1,-2,1) >$
    ii) $\underline{v} = (2,1,1,3)$, $\underline{w} = (0,1,2,-1)$
    b) i) $< (1,2,3,1) >$
    ii) $\underline{v} = (2,4,-4,2)$, $\underline{w} = (1,2,3,1)$

12. $(2,4,5,-2)$

13. a) $< (1,0,0,0,-1),(0,1,0,-1,0),(0,0,1,0,-1) >$
    b) $(2,1,2,1,2)$ and $(1,1,0,-1,-1)$

14. a) $V^{\perp} =< \frac{1}{3}\sqrt{3}(1,1,1) >$; $V =< \frac{1}{2}\sqrt{2}(1,0,-1), \frac{1}{6}\sqrt{6}(1,-2,1) >$
    b) $l^{\perp} =< \frac{1}{2}\sqrt{2}(1,1,0),(0,0,1) >$; $l =< \frac{1}{2}\sqrt{2}(1,-1,0) >$
    c) $W =< \frac{1}{2}\sqrt{2}(1,0,0,-1), \frac{1}{10}\sqrt{10}(1,2,-2,1), \frac{1}{3}\sqrt{3}(1,-1,0,1) >$,
    $W^{\perp} =< \frac{1}{15}\sqrt{15}(1,2,3,1) >$

15. a) $W =< \frac{1}{5}(1,2,2,4), \frac{1}{15}(12,4,4,-7), \frac{1}{15}(6,2,-13,4) >$
    b) complete with $\frac{1}{15}(6,-13,2,4)$
    $(4,4,4,5) \to (8,3,0,0)$

16. a) $W_1 =< \underline{b}_1 >$, $\underline{b}_1 = \frac{1}{2}(1,1,1,1)$
    $W_2 =< \underline{b}_1, \underline{b}_2 >$, $\underline{b}_2 = \frac{1}{2}(1,1,-1,-1)$
    $W_3 =< \underline{b}_1, \underline{b}_2, \underline{b}_3 >$, $\underline{b}_3 = \frac{1}{2}(1,-1,1,-1)$
    b) $\frac{3}{2}\underline{b}_1$

17. $< \frac{1}{2}\sqrt{2}(1,0,0,0,-1), (\frac{1}{5}\sqrt{5}(0,0,2,-1,0), \frac{1}{6}\sqrt{6}(1,-2,0,0,1) >$

18.
$$\begin{pmatrix} 1 & 3 & 7 \\ 1 & 3 & 9 \\ 1 & -1 & 3 \\ 1 & -1 & 5 \end{pmatrix} = \frac{1}{2}\begin{pmatrix} 1 & 1 & -1 \\ 1 & 1 & 1 \\ 1 & -1 & -1 \\ 1 & -1 & 1 \end{pmatrix} \cdot \begin{pmatrix} 2 & 2 & 12 \\ 0 & 4 & 4 \\ 0 & 0 & 2 \end{pmatrix}$$

19. a) $(1/2)(1,1,1,1)$, $(1/\sqrt{20})(1,-1,3,-3)$; b) $(2,1,3,0)$; c) $2\sqrt{5}$

20. a) $\pi/3$; b) $\underline{x} = \lambda(-4,1,3,0) + \mu(2,0,-3,1)$

21. b) $(3/2)\underline{a}$; c) $\underline{a} + \underline{b}$

Bibliography

[1] Carl B. Boyer. *A history of mathematics.* John Wiley & Sons, Inc., New York, (1989, 2nd edition with Uta C. Merzbach)

[2] Jan van de Craats. *Vectoren en matrices.* Epsilon Uitgaven, Utrecht (2000).

This book (in Dutch) overlaps quite a bit with the course material. But of course, the author uses his own approach.

[3] David C. Lay. *Linear algebra and its applications.* Pearson/Addison Wesley, Boston, etc. (2006, 3rd ed. update)

A pleasant book to read. Contains many exercises. Regularly, newer editions are being published.

[4] Liang-shin Hahn. *Complex numbers & Geometry.* The Mathematical Association of America (1994)

This nice little book discusses the role of complex numbers in plane geometry more extensively then we do in the lecture notes.

[5] Paul Halmos. *Finite–dimensional vector spaces.* Van Nostrand (1958)

A classic with emphasis on the abstract theory of vector spaces.

[6] Murray R. Spiegel. *Theory and problems of complex variables.* Schaum's Outline Series, McGraw-Hill (1974)

The books in the Schaum's Outline Series contain many problems with complete solutions.

[7] Gilbert Strang. *Linear algebra and its applications*. Harcourt Brace etc., San Diego (1988, 3rd edition)

Clearly written text on linear algebra in which matrices are central. In contrast to Halmos's book, no emphasis on abstract theory. Strang has written more books on linear algebra.

Index