Linear Algebra

Analysis Series

M. Winklmeier

Chigüiro Collection

Work in progress. Use at your own risk.

Contents

Chapter 1

Systems of Linear Equations

Bla bla bla

1.1 Examples of systems of linear equations

Let us start with a few examples of linear systems of linear equations.

Example 1.1. Assume that a zoo has birds and cats. All of their animals combined, they have 60 heads and 200 legs. How many birds and cats are in the zoo?

Solution. First, we give names to the quantities we want to calculate. So let B = number of birds, C = number of cats in the zoo. If we write the information given in the exercise in formulas, we obtain

$$\begin{array}{cr@{\;}l@{\qquad}l} \textcircled{1} & b + c = & 60, & \text{(total number of heads)} \\ \textcircled{2} & 2b + 4c = & 200, & \text{(total number of legs)} \end{array}$$

since each bird has 1 head and 2 legs and each cat has 1 head and legs. Equation ① tells us that $B = 60 - C$. If we insert this into equation ②, we find

$$200 = 2(60 - C) + 4C = 120 - 2C + 4C = 120 + 2C \quad\Longrightarrow\quad 2c = 80 \quad\Longrightarrow\quad c = 40.$$

This implies that $B = 60 - C = 60 = 40 = 20$. Note that in our calculations and arguments, all the arrow all go "from left to right", so we found that the only possible solution is $B = 40, C = 20$. Inserting this in the original equation shows that this is indeed a solution. So there are 40 birds and 20 cats. ◇

Let us put one more equation into the zoo.

Example 1.2. Assume that a zoo has birds and cats. All of their animals combined, they have 60 heads and 200 legs. Moreover, there are 140 cage and in every cage there are either 2 birds or 3 cats. How many birds and cats are in the zoo?

Solution. Again, let B = number of birds, C = number of cats in the zoo. The information of the exercise gives the following equations:

$$\begin{array}{cr l}
\text{①} & B+C = 60, & \text{(total number of heads)}\\
\text{②} & 2B+4C = 200, & \text{(total number of legs)}\\
\text{③} & 2B+3C = 140. & \text{(total number of cages)}
\end{array}$$

As in the previous exercise, we obtain from that $B = 40, C = 20$. Clearly, this also satisfies equation ③. ◇

Example 1.3. Assume that a zoo has birds and cats. All of their animals combined, they have 60 heads and 200 legs. Moreover, there are 100 cage and in every cage there are either 2 birds or 3 cats. How many birds and cats are in the zoo?

Solution. Again, let B = number of birds, C = number of cats in the zoo. The information of the exercise gives the following equations:

$$\begin{array}{cr l}
\text{①} & B+C = 60, & \text{(total number of heads)}\\
\text{②} & 2B+4C = 200, & \text{(total number of legs)}\\
\text{③} & 2B+3C = 100. & \text{(total number of cages)}
\end{array}$$

As in the previous exercise, we obtain from that $B = 40, C = 20$. However, this does not satisfy equation ③; so there is no way to choose B and C such that all three equations are satisfied simultaneously. Therefore, a zoo as in this example does not exist. ◇

We give a few more examples.

Example 1.4. Find a polynomial P of degree at most 3 with

$$P(0) = 1, \quad P(1) = 7, \quad P'(0) = 3, \quad P'(2) = 23. \tag{1.1}$$

Solution. A polynomial of degree at most 3 is known, if we know its 4 coefficients. In this exercise, the unknowns are the coefficients of the polynomial P. We can write $P(x) = \alpha x^3 + \beta x^2 + \gamma x + \delta$ and we have to find $\alpha, \beta, \gamma, \delta$ such that (1.1) is satisfied. Note that $P'(x) = 3\alpha x^2 + 2\beta x + \gamma$. Hence (1.1) is equivalent to the following system of equations:

$$\left.\begin{array}{r}
P(0) = 1,\\
P(1) = 7,\\
P'(0) = 3,\\
P'(2) = 23.
\end{array}\right\}
\quad\iff\quad
\begin{cases}
\text{①} & \delta = 1,\\
\text{②} & \alpha + \beta + \gamma + \delta = 7,\\
\text{③} & \gamma = 3,\\
\text{④} & 24\alpha + 8\beta + 2\gamma + \delta = 23.
\end{cases}$$

Clearly, $\delta = 1$ and $\gamma = 3$. If we insert this in the remaining equations, we obtain a system of two equations for the two unknowns α, β:

$$\begin{array}{cr}
\textcircled{\scriptsize 2'} & \alpha + \beta = 3,\\
\textcircled{\scriptsize 4'} & 24\alpha + 8\beta = 16.
\end{array}$$

From ②' we obtain $\beta = 4-\alpha$. If we insert this into ④', we get that $16 = 24\alpha + 8(4-\alpha) = 16\alpha + 32$, that is, $\alpha = (32-16)/16 = 1$. So the only possible solution is

$$\alpha = 1, \quad \beta = 2, \quad \gamma = 3, \quad \delta = 1.$$

It is easy to verify that the polynomial $P(x) = x^3 + 2x^2 + 3x + 1$ has all the desired properties. ◇

Example 1.5. A pole is 5 metres long and shall be coated with varnish. There are two types of varnish available: The green one adds 3 g per 50 cm to the pole, the red one adds 6 g per meter to the pole. Is it possible to coat the pole in a combination of the varnishes so that the total weight added is

(a) 35 g? (b) 30 g?

Solution. (a) We call g the length of the pole which will be covered in green and r the length of the pole which will be covered in red. Then we obtain the system of equations

$$\begin{array}{cl} ① & g + r = 5 \qquad \text{(total length)} \\ ② & 6g + 6r = 35 \qquad \text{(total weight)} \end{array}$$

The first equation gives $r = 5 - g$. Inserting into the second equation yields $35 = 6g + 6(5-g) = 30$ which is a contradiction. This shows that there is no solution.

(b) As in (a), we obtain the system of equations

$$\begin{array}{cl} ① & g + r = 5 \qquad \text{(total length)} \\ ② & 6g + 6r = 30 \qquad \text{(total weight)} \end{array}$$

Again, the first equation gives $r = 5-g$. Inserting into the second equation yields $30 = 6g+6(5-g) = 30$ which is always true, independently of how we choose g and r as long as ① is satisfied. This means that in order to solve the system of equations, it is sufficient to solve only the first equation since then the second one is automatically satisfied. So we have infinitely many solutions. Any pair g, r such that $g + r = 5$ gives a solution. So for any g that we choose, we only have to set $r = 5 - g$ and we have a solution of the problem. Of course, we could also fix r and then choose $g = 5 - r$ to obtain a solution.

For example, we could choose $g = 1$, then $r = 4$, or $g = 0.00001$, then $r = 4.99999$, or $r = -2$ then $g = 7$. Clearly, the last example does not make sense for the problem at hand, but it still does satisfy our system of equations. ◇

All the examples were so-called linear systems of linear equations. Let us define what we mean by this,

Definition 1.6. A $m \times n$ *system of linear equations* is a system of m linear equations for n unknowns of the form

$$\begin{array}{llll} a_{11}x_1 & + a_{12}x_2 & + \cdots + a_{1n}x_n & = b_1 \\ a_{21}x_1 & + a_{22}x_2 & + \cdots + a_{2n}x_n & = b_2 \\ \vdots & & \vdots & \vdots \\ a_{m1}x_1 & + a_{m2}x_2 & + \cdots + a_{mn}x_n & = b_m. \end{array}$$

The unknowns are $x_1, \dots, x_n$. The numbers a_{ij} and b_i ($i = 1, \dots, m$, $j = 1, \dots, n$) are given. The numbers a_{ij} are called the *coefficients of the linear system* and numbers $b_1, \dots, b_n$ are called the *right side of the linear system.*

In the special case when all b_i are equal to 0, the system is called a *homogeneous*; otherwise it is called *inhomogeneous.*

The *coefficient matrix* A of the system is the collection of all coefficients a_{ij}

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

The coefficient matrix is nothing else than the collection of the coefficients a_{ij} ordered in some sort of table or rectangle such that the place of the coefficient a_{ij} is in the ith row of the jth column.

Let us come back to our examples.

Example 1.1: This is a 2×2 system with coefficients $a_{11} = 1$, $a_{11} = 1$, $a_{21} = 2$, $a_{22} = 4$ and right hand side $b_1 = 60, b_2 = 200$. The system has a unique solution. The coefficient matrix is

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

Example 1.2: This is a 3×2 system with coefficients $a_{11} = 1$, $a_{11} = 1$, $a_{21} = 2$, $a_{22} = 4$, $a_{31} = 2$, $a_{32} = 3$, and right hand side $b_1 = 60, b_2 = 200$, $b_3 = 140$. The system has a unique solution. The coefficient matrix is

$$A = \begin{pmatrix} 1 & 1 \\ 2 & 4 \\ 2 & 3 \end{pmatrix}.$$

Example 1.3: This is a 3×2 system with coefficients $a_{11} = 1$, $a_{11} = 1$, $a_{21} = 2$, $a_{22} = 4$, $a_{31} = 2$, $a_{32} = 3$, and right hand side $b_1 = 60, b_2 = 200$, $b_3 = 100$. The system has no solution. The coefficient matrix is the same as in Example 1.2.

Example 1.4: This is a 4×4 system with coefficients $a_{11} = 0$, $a_{12} = 0$, $a_{13} = 0$, $a_{14} = 1$, $a_{21} = 1$, $a_{22} = 1$, $a_{23} = 1$, $a_{24} = 1$, $a_{31} = 0$, $a_{32} = 0$, $a_{33} = 1$, $a_{34} = 0$, $a_{41} = 24$, $a_{42} = 8$, $a_{43} = 2$, $a_{44} = 1$, and right hand side $b_1 = 1, b_2 = 7$, $b_3 = 3$, $b_4 = 23$. The system has a unique solution. The coefficient matrix is

$$A = \begin{pmatrix} 0 & 0 & 0 & 1 \\ 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 0 \\ 24 & 8 & 2 & 1 \end{pmatrix}.$$

Example 1.5: This is a 2×2 system with coefficients $a_{11} = 1$, $a_{11} = 6$, $a_{21} = 1$, $a_{22} = 6$. In case (a) the right hand side is $b_1 = 5, b_2 = 35$ and the system has no solution.
In case (b) the right hand side is $b_1 = 5, b_2 = 30$ and the system has infinite solutions.
In both cases, the coefficient matrix is

$$A = \begin{pmatrix} 1 & 6 \\ 1 & 6 \end{pmatrix}.$$

Given an $m \times n$ system of linear equations, two important solutions arise:

- *Existence*: Does the system have a solution?
- *Uniqueness*: If the system has a solution, is it unique?

As we saw, in Examples 1.1, 1.2, 1.4, 1.5 (b) solutions do exist. In Example 1.5 (b) the solution is not unique (on the contrary: it has infinite solutions!). Examples 1.3 and 1.5(a) do not admit solutions.

More generally, we would like to be able so say something about the structure of solutions of linear systems. For example, is it possible that there is only one solution? That there are exactly two solutions? That there are infinite solutions? That there is is no solution? Can we give criteria for existence and/or uniqueness of solutions? Can we give criteria for existence of infinite solutions?

(Spoiler alert: *A system of linear equations has either no or exactly one or infinite solutions. It is not possible that it has, e.g., exactly* 7 *solutions.*)

Before answering these questions for general $m \times n$ systems, we will have a closer look at 2×2 systems in the next section.

1.2 Linear 2×2 systems of equations

Let us come back to the equation from Example 1.1. For convenience, we write now x instead of B and y instead of C. Recall that the system of equations that we are interested in solving is

$$\begin{aligned} ① \qquad x + \quad y &= 60, \\ ② \qquad 2x + 4y &= 200. \end{aligned} \tag{1.2}$$

We want to give a geometric meaning to this system of equations. To this end we think of pairs x, y as points (x, y) in the plane. Let's forget about equation ② for a moment and concentrate only on ①. Clearly, there are infinitely many solutions. If we choose an arbitrary x, we can always find y such that ① satisfied (just take $y = 60 - x$). Similarly, if we choose any y, then we only have to take $x = 60 - y$ and we obtain a solution of ①.
Now, where in the xy-plane lie *all* solutions of ①? Clearly, ① is equivalent to $y = 60 - x$ which we easily identify of the equation of the line L_1 in the xy-plane which passes through $(0, 60)$ and has slope -1. In summary, a pair (x, y) is a solution of ① if and only if it lies on the line L_1.

If we apply the same reasoning to ②, we find that a pair (x, y) satisfies ② if and only if (x, y) lies on the line L_2 in the xy-plane given by $y = \frac{1}{4}(200 - 2x)$ (this is the line in the xy-plane passing through $(9, 50)$ with slope $-\frac{1}{2}$).

Now it is clear that a pair (x, y) satisfies both ① and ② if and only if it lies both on L_1 and L_2. So finding the solution of our system (1.2) is the same as finding the intersection of the two lines L_1 and L_2. From elementary geometry we know that there are exactly three possibilities:

(i) L_1 and L_2 are not parallel. Then they intersect in exactly one point.

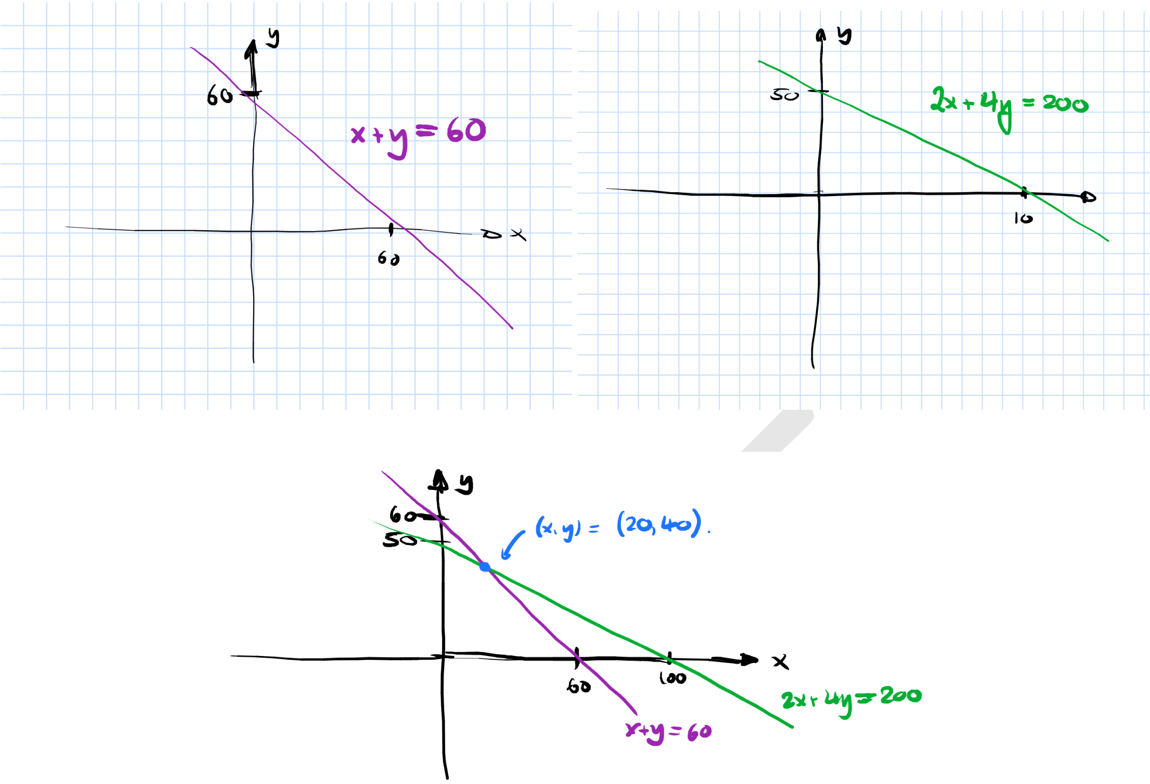

Figure 1.1: Example 1.1. Graphs of L_1, L_2 and their intersection.

(ii) L_1 and L_2 are parallel and not equal. Then they do not intersect.

(iii) L_1 and L_2 are parallel and equal. Then $L_1 = L_2$ and they intersect in infinite points (they intersect in every point of $L_1 = L_2$).

In our example we know that the slope of L_1 is -1 and that the slope of L_2 is $-\frac{1}{2}$, so they are not parallel and therefore intersect in exactly one point. Consequently, the system (1.2) has exactly one solution, see Figure 1.1

If we look again at Example 1.5, we see that in Case (a) we look for the intersection of the lines

$$L_1 : y = 5 - x, \qquad L_2 : y = \frac{35}{6} - x.$$

Both lines have slope -1 so they are parallel. Since the constant terms in both lines are not equal, they never intersect, showing that the system of equations has no solution, see Figure 1.2.
In Case (b), the two lines that we have to intersect are

$$G_1 : y = 5 - x, \qquad G_2 : y = 5 - x.$$

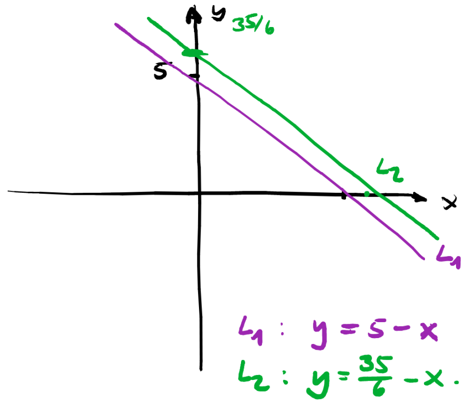

FIGURE 1.2: Example 1.5. Graphs of G_1, G_2.

We see that $G_1 = G_2$, so every point on G_1 (or G_2) is solution of the system and therefore we have infinite solutions.

Now let us consider the general case.

One linear equation with two unknowns

The general form of one linear equation with two unknowns is

$$\alpha x + \beta y = \gamma. \tag{1.3}$$

For the set of solutions, there are three possibilities:

(i) *The set of solutions forms a line.* This happens if at least one of the coefficients α or β is different from 0. If $\beta \neq 0$, then set of all solutions is equal to the line $L : y = -\frac{\alpha}{\beta} x + \frac{\gamma}{\beta}$ which is a line with slope $-\frac{\alpha}{\gamma}$. If $\beta = 0$ and $\alpha \neq 0$, then the set of solutions of (1.3) is a line parallel to the y-axis passing through $(\frac{\gamma}{\alpha})$.

(ii) *The set of solutions is all of the plane.* This happens if $\alpha = \beta = \gamma = 0$. In this case, clearly every pair (x, y) is a solution of (1.3).

(iii) *The set of solutions is empty.* This happens if $\alpha = \beta = 0$ and $\gamma \neq 0$. In this case, no pair (x, y) can be a solution of (1.3) since the left hand side is always 0.

Two linear equations with two unknowns

The general form of one linear equation with two unknowns is

$$\begin{aligned} ① \quad Ax + By &= U \\ ② \quad Cx + Dy &= V. \end{aligned} \tag{1.4}$$

We are using the letters A, B, C, D instead of $a_{11}, a_{12}, a_{21}, a_{22}$ in order to make the calculations more readable. If we interprete the system of equations as intersection of two geometrical objects, we already know how the possible solutions will be:

- *A point* if ① and ② describe two non-parallel lines.
- *A line* if ① and ② describe the same line; or if one of the equations is a plane and the other one is a line.
- *A plane* if both equations describe a plane.
- *The empty set* if the two equations describe parallel but different lines; or if one of the equations has no solution.

In summary, we have:

Remark 1.7. The system (1.4) has either exactly 1 solution or infinite solutions or no solution.

It is not possible to have for instance exactly 7 solutions.

Exercise. How is the situation if we had a system of 3 linear equations for 2 unknowns?

Proof of Remark 1.7. Now we want proof the Remark 1.7 algebraically and we want to find a criteria on a, b, c, d which allows us to decide easily how many solutions there are. Let's look at the different cases.

Case 1. $B \neq 0$. In this case we can solve ① for y and obtain $y = \frac{1}{B}(U - Ax)$. In ② this gives $Cx + \frac{D}{B}(U - Ax) = V$. If we put all terms with x on one side and all other terms on the other side, we obtain

②' $(AD - BC)x = DU - BV$

(i) If $AD - BC \neq 0$ then there is at most one solution, namely $x = \frac{DU-BV}{AD-BC}$ and consequently $y = \frac{1}{B}(U - Ax) = \frac{AV-CU}{AD-BC}$. Inserting these expressions for x and y in our system of equations, we see that they indeed solve the system (1.4), so that we have exactly one solution.

(ii) If $AD - BC = 0$, then equation ②' reduces to $0 = DU - BV$. This equation has either no solution (if $DU - BV \neq 0$) or infinite solutions (if $DU - BV = 0$). Since ① has infinite solutions, it follows that the system (1.4) has either no solution or infinite solutions.

Case 2. $D \neq 0$. In this case we can solve ② for y and obtain $y = \frac{1}{D}(V - Cx)$. In ② this gives $Ax + \frac{B}{D}(V - Cx) = U$. If we put all terms with x on one side and all other terms on the other side, we obtain

②' $(AD - BC)x = DU - BV$

We have the same subcases as before:

(i) If $AD - BC \neq 0$ then there is exactly one solution, namely $x = \frac{DU-BV}{AD-BC}$ and consequently $y = \frac{1}{B}(U - Ax) = \frac{AV-CU}{AD-BC}$.

(ii) If $AD - BC = 0$, then equation ②' reduces to $0 = DU - BV$. This equation has either no solution (if $DU - BV \neq 0$) or infinite solutions (if $DU - BV = 0$). Since ② has infinite solutions, it follows that the system (1.4) has either no solution or infinite solutions.

Case 3. $B = 0$ and $D = 0$. Observe that in this case $AD - BC = 0$. In this case the system (1.4) reduces to

$$Ax = U, \quad Cx = V. \tag{1.5}$$

We see that the system no longer depends on y. So, if the system (1.5) has at least one solution, then we automatically have infinite solutions since we can choose y freely. If the system (1.5) has no solution, then the original system (1.4) cannot have a solution either.

Note that there are no other cases for the coefficients than these three cases. ☐

Summing up, we find the following theorem:

Theorem 1.8. *The system of linear equations*

$$\begin{aligned} \textcircled{1} \quad & Ax + By = U \\ \textcircled{2} \quad & Cx + Dy = V. \end{aligned} \tag{1.6}$$

has

(i) *exactly one solution if and only if $AD - BC \neq 0$. In this case, the solution is*

$$x = \frac{DU - BV}{AD - BC}, \quad y = \frac{AV - CU}{AD - BC}. \tag{1.7}$$

(ii) *no solution or infinite solutions if $AB - BC = 0$.*

Definition 1.9. The number $d := AD - BC$ is called the *determinant* of the system (1.6).

Later we will generalise this concept to systems with more equations and more variables.

Remark 1.10. Let us see how this connects to our geometric interpretation of the system of equations. Assume that $B \neq 0$ and $D \neq 0$. Then we can solve ① and ② for y obtain equations for lines

$$L_1: \quad y = -\frac{A}{B}x + \frac{1}{B}U, \qquad L_2: \quad y = -\frac{C}{D}x + \frac{1}{D}V.$$

The two lines intersect in exactly one point if and only if they have different slopes, i.e., if $-\frac{A}{B} \neq -\frac{C}{D}$. After multiplication by $-BD$ we see that this is the same as $AD \neq BC$, or $AD - BC \neq 0$.
On the other hand, the lines are parallel (and hence have either no intersection or are equal) if $-\frac{A}{B} \neq -\frac{C}{D}$. This is the case if and only if $AD = BC$, or in other word, if $AD - BC = 0$.

Exercise. Consider the cases when $B = 0$ or $D = 0$ and make the connection between Theorem 1.8 and the geometric interpretation of the system of equations.

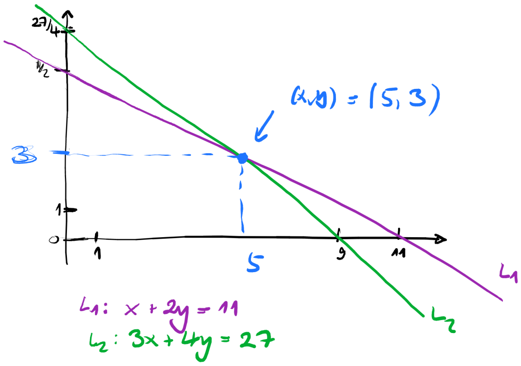

FIGURE 1.3: Example 1.11(a). Graphs of L_1, L_2 and their intersection $(5, 3)$.

Let us consider same examples.

Examples 1.11. (a)

$$\begin{aligned} &① \quad x + 2y &&= 11 \\ &② \quad 3x + 4y &&= 27. \end{aligned}$$

Clearly, the determinant is $d = 4 - 6 = -2 \neq 0$. So we expect *exactly one solution*.

We can check this easily: The first equation gives $x = 11 - 2y$. Inserting this into the second equations leads to

$$3(11 - 2y) + 4y = 27 \quad \Longrightarrow \quad -2y = -6 \quad \Longrightarrow \quad y = 3 \quad \Longrightarrow \quad x = 11 - 2 \cdot 3 = 5.$$

So the solution is $x = 5, y = 3$. (If we did not have Theorem 1.8, we would have to check that this is not only a candidate for a solution, but indeed is one.)

Exercise. Check that the formula (1.7) is satisfied.

(b)

$$\begin{aligned} &① \quad x + 2y &&= 1 \\ &② \quad 2x + 4y &&= 5. \end{aligned}$$

Here, the determinant is $d = 4 - 4 = 0$, so we expect *either no solution or infinite solutions*. The first equations gives $x = 1 - 2y$. Inserting into the second equations gives $2(1 - 2y) + 4y = 5$. We see that the terms with y cancel and we obtain $2 = 5$ which is a contradiction. Therefore, the system of equations has *no solution*.

(c)

$$\begin{aligned} &① \quad x + 2y &&= 1 \\ &② \quad 3x + 6y &&= 3. \end{aligned}$$

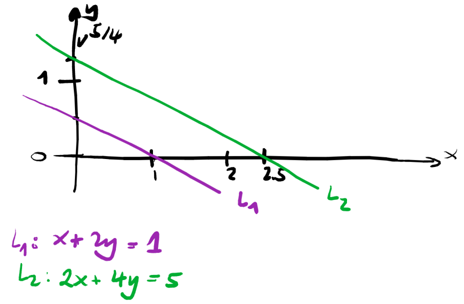

Figure 1.4: Example 1.11(b). The lines L_1, L_2 are parallel and do not intersect.

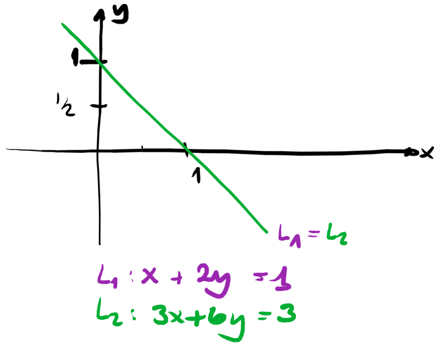

Figure 1.5: Example 1.11(c). The lines L_1, L_2 are equal.

The determinant is $d = 6 - 6 = 0$, so again we expect *either no solution or infinite solutions*. The first equations gives $x = 1-2y$. Inserting into the second equations gives $3(1-2y)+6y = 3$. We see that the terms with y cancel and we obtain $3 = 3$ which is true. Therefore, the system of equations has *infinite solutions* given by $x = 1 - 2y$.

Remark. This was somewhat clear since we can obtain the second equation from the first one by multiplying both sides by 3 which shows that both equations carry the same information and we loose nothing if we simply forget about one of them.

Example 1.12. Find all $k \in \mathbb{R}$ such that the system

$$\begin{array}{rrl} ① & kx+(15/2-k)y= & 1 \\ ② & 4x+ \qquad\qquad 2ky= & 3 \end{array}$$

has exactly one solution.

Solution. We only need to calculate the determinant and find all k such that it is different from zero. So let's start by calculating

$$d = k \cdot 2k - (15/2 - k) \cdot 4 = 2k^2 + 4k - 30 = 2(k^2 + 2k - 15) = 2[(k+1)^2 - 16].$$

So we see that there are exactly two values for k where $d = 0$, namely $k = -1 \pm 4$, that is $k_1 = 3$, $k_2 = -5$. For all other k, we have that $d \neq 0$.
So the answer is: The system has exactly one solution if and only if $k \in \mathbb{R} \setminus \{-5, 3\}$. ◇

Remark 1.13. 1. Note that the answer does not depend on the right hand side of the system of the equation. Only the coefficients on the left hand side determine if there is exactly one solution or not.

2. If we wanted, we could also calculate the solution x, y in the case $k \in \mathbb{R} \setminus \{-3, 1\}$. We could do it by hand or use (1.7). Either way, we find

   $$x = \frac{1}{d}[2k - 3(15/2 - k)] = \frac{5k - 45/2}{2k^2 + 4k - 30}, \qquad y = \frac{1}{d}[6k - 4] = \frac{6k - 4}{2k^2 + 4k - 30}.$$

   Note that the denominators would become 0 if $k = -5$ or $k = 3$.

3. What happens if $k = -3$ or $k = 1$? In both cases, $d = 0$, so we will either have no solution or infinite solutions.

   If $k = -3$, then the system becomes

   $$3x + 9/2y = 1, \qquad 4x + 6y = 3.$$

   Multiplying the first equation by $4/3$, we obtain

   $$4x - 6y = \frac{4}{9}, \qquad 4x - 6y = 3$$

   which clearly cannot be satisfied simultaneously.

   If $k = 5$, then the system becomes

   $$5x + 5/2y = 1, \qquad 4x + 10y = 3.$$

   Multiplying the first equation by $4/5$, we obtain

   $$4x - +10 = \frac{4}{5}, \qquad 4x + 10y = 3$$

   which clearly cannot be satisfied simultaneously.

1.3 Summary

1.4 Exercises

Chapter 2

$\mathbb{R}^2$ and $\mathbb{R}^3$

2.1 Vectors in $\mathbb{R}^2$

Recall that the xy-plane is the set of all pairs (x, y) with $x, y \in \mathbb{R}$. We will denote it by $\mathbb{R}^2$.

Maybe you already encountered vectors in a physics lecture. For instance velocities and forces are described by vectors. The velocity of a particle says how fast and in which direction the particle moves. Usually, a velocity are represented by an arrow which points in the direction in which the particle moves and whose length is proportional to the magnitude of the velocity.

A force has strength and a direction so it is represented by an arrow which point in the direction in which it acts and with length proportional to its strength.

Observe that it is not important where in the space $\mathbb{R}^2$ or $\mathbb{R}^3$ we put the arrow. As long it points in the same direction and has the same length, it is considered the same vector. We call two arrows *equivalent* if they have the same direction and the same length. A *vector* is the set of all arrows which are equivalent to a given arrow. Each specific arrow in this set is called a *representation* of the vector. A special representation is the arrow that starts in the origin $(0, 0)$.

Given two points P, Q in the xy-plane, we write $\overrightarrow{PQ}$ for the vector which is represented by the arrow that starts in P and ends in Q.

Example 2.1.

Let $P(1,1)$ and $Q(3,4)$ be points in the xy-plane. The arrow from P to Q is $\overrightarrow{PQ} = \begin{pmatrix} 2 \\ 3 \end{pmatrix}$.

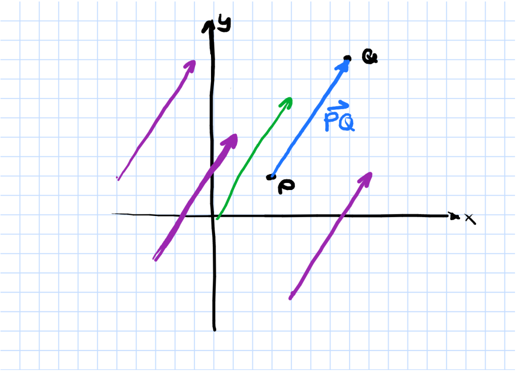

FIGURE 2.1: The vector $\overrightarrow{PQ}$ and several of its representations. The green arrow is the special representation whose initial point in is in the origin.

We can identify a point $P(p_1, p_2)$ in the xy-plane with the vector starting in $(0,0)$ and ending in P. We denote this vector by $\overrightarrow{0P}$ or $\begin{pmatrix} p_1 \\ p_2 \end{pmatrix}$ or sometimes by $(p_1,\, p_2)^t$ in order to save space (the subscript ^{t} stands for “transposed”). p_1 is called the *x-coordinate* or the *x-component* of $\vec{v}$ and p_2 is called the *y-coordinate* or the *y-component* of $\vec{v}$.

On the other hand, given a vector (a,b), then it describes a unique point in the xy-plane, namely the tip of the arrow which represents the given vector and starts in the origin.

So we can identify the set of all vectors in $\mathbb{R}^2$ with $\mathbb{R}^2$ itself.

Observe that the slope of the arrow $\vec{v} = (a,b)$ is $\frac{b}{a}$ if $a \neq 0$. If $a = 0$, then we obtain a vector which is parallel to the y-axis. Vectors are usually denoted by a small letter with an arrow on top.

If a vector is given, e.g., as $\vec{v} = (2,5)^t$, then this is an arrow whose tip would be at the point $(2,5)$ if its initial point is in the origin. If it is anywhere else, then we find the tip if we move 2 units to the right parallel to the x-axis and 5 units up parallel to the y-axis.

A very special vector is the zero vector $(0,0)^t$. Is is usually denoted by $\vec{0}$.

In order to distinguish numbers in $\mathbb{R}$ from vectors, we call them *scalars*.

Now we want to do algebra with vectors. If we think of a force and we double its strength then the corresponding vector should be twice as long. If we multiply the force by 5, then the length of the corresponding vector should be 5 times as long, that is, if for instance a force $\vec{F} = (3,4)$ is given, then $5\vec{F}$ should be $(5 \cdot 3,\, 5 \cdot 4) = (15, 20)$. In general, if a vector $\vec{v} = (a,b)$ is given, then $c\vec{v} = (ca, cb)$. Note that the resulting vector is always parallel to the original one. If $c > 0$, then the resulting vector points in the same direction as the original one, if $c < 0$, then it points in the opposite direction, see Figure 2.2

How should we sum two vectors? Again, let us think of forces. Assume we have two forces $\vec{F}_1$ and $\vec{F}_2$ both acting on the same particle. Then we get the resulting force by drawing the arrow representing $\vec{F}_1$ and at its tip put the initial point of the arrow representing $\vec{F}_2$. The total force is then represented by the arrow starting in the initial point of $\vec{F}_1$ and ending in the tip of $\vec{F}_2$.

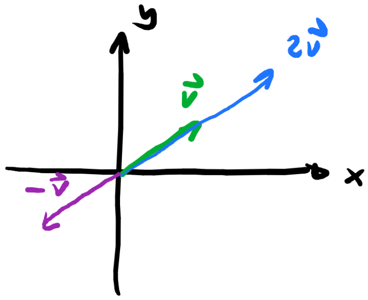

FIGURE 2.2: Multiplication of a vector by a scalar.

Exercise. Convince yourself that we obtain the same result if we start with $\vec{F}_2$ and put the initial point of $\vec{F}_1$ at the tip of $\vec{F}_2$.

We could also think of the sum of velocities. For example, if the have a train with velocity $\vec{v}_t$ and on the train a passenger is moving with relative velocity $\vec{v}_p$, then the total velocity is the vector sum of the two.
Now assume that $\vec{F}_1 = (a, b)^t$ and $\vec{F}_2 = (p, q)^t$. Algebraically, we obtain the components of their sum by summing the components: $\vec{F}_1 + \vec{F}_2 = (a + p, b + q)$, see Figure 2.3. When you do vector sums, you should always think in triangles (or polygons if you sum more than two vectors).

Exercise. Given two points $P(p_1, p_2)$, $Q(q_1, q_2)$ in the xy-plane. Convince yourself that $\overrightarrow{0P} + \overrightarrow{PQ} = \overrightarrow{0Q}$ and consequently $\overrightarrow{PQ} = \overrightarrow{0Q} - \overrightarrow{0P}$.
How could you write $\overrightarrow{QP}$ in terms of $\overrightarrow{0P}$ and $\overrightarrow{0Q}$? What is its relation with $\overrightarrow{PQ}$?

We sum up:

Definition 2.2. Let $\vec{v} = \begin{pmatrix} a \\ b \end{pmatrix}$, $\vec{w} = \begin{pmatrix} p \\ q \end{pmatrix}$, $c \in \mathbb{R}$. Then:

Vector sum: $$\vec{v} + \vec{w} = \begin{pmatrix} a \\ b \end{pmatrix} + \begin{pmatrix} p \\ q \end{pmatrix} = \begin{pmatrix} a + p \\ b + q \end{pmatrix},$$

Product with a scalar: $$c\vec{v} = c \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} ca \\ cb \end{pmatrix},$$

With this definition, it is easy to see that for arbitrary vectors $\vec{u}, \vec{v}, \vec{w} \in \mathbb{R}^2$ and scalars $\alpha, \beta \in \mathbb{R}$ the so-called *vector space axioms* hold:

Vector Space Axioms.

(a) **Associativity:** $\vec{(}u + \vec{v}) + \vec{w} = \vec{u} + (\vec{v} + \vec{w})$

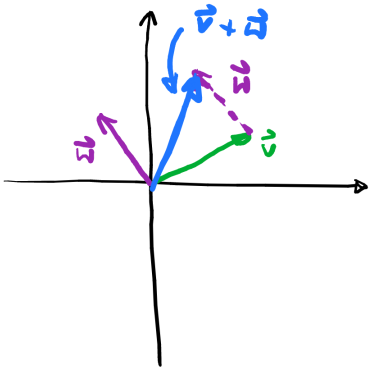

Figure 2.3: Sum of two vectors.

(b) **Commutativity:** $\vec{v} + \vec{w} = \vec{w} + \vec{v}$.

(c) **Identity element of addition:** For every $\vec{v} \in \mathbb{R}^2$, we have $\vec{0} + \vec{v} = \vec{v} + \vec{0} = \vec{v}$.

(d) **Inverse element:** For every $\vec{v} \in \mathbb{R}^2$, we have an inverse element $\vec{v'}$ such that $\vec{v} + \vec{v'} = \vec{0}$, namely $\vec{v'} = -\vec{v}$.

(e) **Identity element of multiplication by scalar:** For every $\vec{v} \in \mathbb{R}^2$, we have that $1\vec{v} = \vec{v}$.

(f) **Compatibility:** For every $\vec{v} \in \mathbb{R}^2$ and $a, b \in \mathbb{R}$, we have that $(ab)\vec{v} = a(b\vec{v})$.

(g) **Distributivity laws:** For all $\vec{v}, \vec{w} \in \mathbb{R}^2$ and $a, b \in \mathbb{R}$, we have

$$(a+b)\vec{v} = a\vec{v} + b\vec{v} \quad \text{and} \quad a(\vec{v} + \vec{w}) = a\vec{v} + a\vec{w}.$$

These axioms are fundamental for linear algebra. We will come back to them later when we deal with abstract vector spaces in Chapter 4.

Let us look at some more geometric properties of vectors. Clearly a vector is known if we know its length and its angle with the x-axis.
From the Pythagoras theorem it is clear that the length of a vector $\vec{v} = (a, b)^t$ is $\sqrt{a^2 + b^2}$.

Definition 2.3 (Norm of a vector in $\mathbb{R}^2$). The *length* $\vec{v} = \begin{pmatrix} a \\ b \end{pmatrix} \in \mathbb{R}^2$ is denoted by $\|\vec{}\|$. It is given by

$$\|\vec{v}\| = \sqrt{a^2 + b^2}\ .$$

Other names for the length of $\vec{v}$ are *magnitude of* $\vec{v}$ or *norm of* $\vec{v}$.

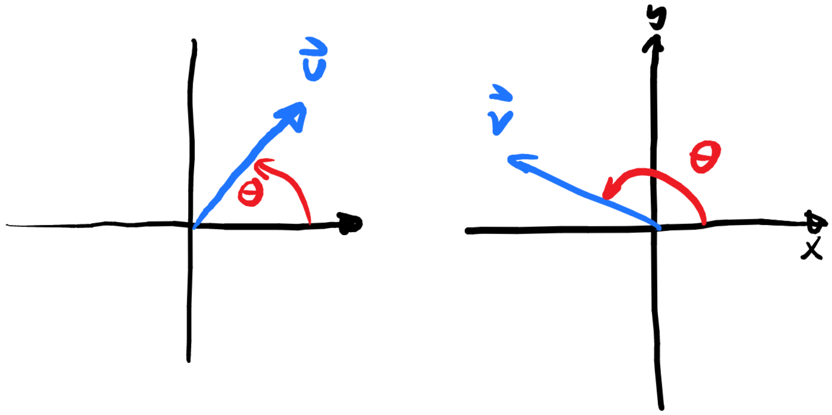

FIGURE 2.4: Length and angle of a vector.

As already mentioned earlier, the slope of vector $\vec{v}$ is $\frac{b}{a}$ if $a \neq 0$. If φ is the angle of the vector $\vec{v}$ with the x-axis then $\tan\varphi = \frac{b}{a}$ if $a \neq 0$. If $a = 0$, then $\varphi = 0$ or $\varphi = \pi$. Recall that the range of arctan is $(-\pi/2,\ \pi/2)$, so we cannot simply take arctan of the fraction $\frac{a}{b}$ in order to obtain φ. Observe that $\arctan\frac{b}{a} = \arctan -b{-}a$, however the angles of the vectors $(a,b)^t$ and $(-a,-b)^t$ are parallel but point in opposite directions, so they do not have the same angle with the x-axis. From geometry, we find

$$\varphi = \begin{cases} \arctan\frac{b}{a} & \text{if } a > 0, \\ \pi - \arctan\frac{b}{a} & \text{if } a < 0, \\ \pi/2 & \text{if } a = 0,\ b > 0, \\ -\pi/2 & \text{if } a = 0,\ b < 0. \end{cases}$$

Note that this formula gives angles with values $[-\pi/2,\ 3\pi/2)$.

Proposition 2.4 (Properties of the norm). *Let $\lambda \in \mathbb{R}$ and $\vec{v},\ \vec{w} \in \mathbb{R}^2$. Then the following is true:*

(i) $\|\lambda\vec{v}\| = |\lambda|\|\vec{v}\|$,

(ii) $\|\vec{v} + \vec{w}\| \leq \|\vec{v}\| + \|\vec{w}\|$,

(iii) $\|\vec{v}\| = 0$ *if and only if* $\vec{v} = \vec{0}$.

Proof. Let $\vec{v} = (a,b)^t$, $\vec{w} = (c,d)^t \in \mathbb{R}^2$ and $\lambda \in \mathbb{R}$.

(i)
$$\begin{aligned}\|\lambda\vec{v}\| &= \|\lambda(a,b)^t\| = \|(\lambda a,\ \lambda b)^t\| = \sqrt{(\lambda a)^2 + (\lambda b)^2} = \sqrt{\lambda^2(a^2+b^2)} = |\lambda|\sqrt{a^2+b^2} \\ &= |\lambda|\|\vec{v}\|.\end{aligned}$$

(ii) This will be shown later in XXX.

(iii) Since $\|\vec{v}\| = \sqrt{a^2 + b^2}$ it follows that $\|\vec{}\| = 0$ if and only if $a = 0$ and $b = 0$. This is the case if and only if $\vec{v} = \vec{0}$.

$\square$

Definition 2.5. A vector $\vec{v} \in \mathbb{R}^2$ is called a *unit vector* if $\|\vec{v}\| = 1$.

Note that every vector $\vec{v} \neq \vec{0}$ defines a unit vector pointing in the same direction as itself by $\|\vec{v}\|^{-1}\vec{v}$.

Remark 2.6. (i) The tip of every unit vector lies on the unit circle, and every vector whose initial point is the origin and whose tip lies on the unit circle is a unit vector.

(ii) Every unit vector is of the from $\begin{pmatrix} \cos\varphi \\ \sin\varphi \end{pmatrix}$ where φ is its angle with the positive x-axis.

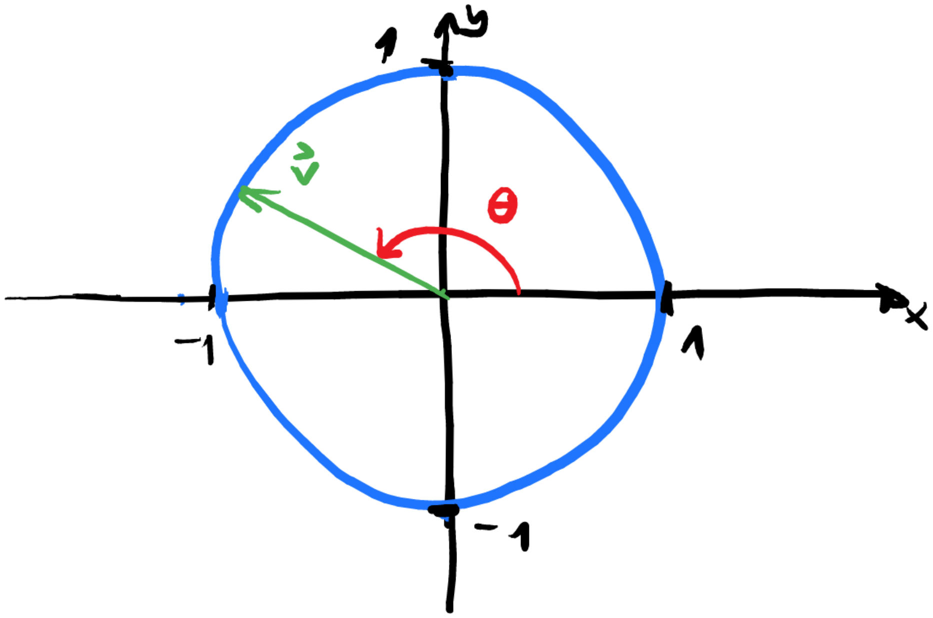

FIGURE 2.5: Unit vectors.

Finally, we define two very special unit vectors:

$$\vec{e}_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \qquad \vec{e}_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}.$$

Clearly, $\vec{e_1}$ is parallel to the x-axis, $\vec{e_2}$ is parallel to the y-axis and $\|\vec{e}_1\| = \|\vec{e}_2\| = 1$.

Remark 2.7. Every vector $\vec{v} = \begin{pmatrix} a \\ b \end{pmatrix}$ can be written as

$$\vec{v} = \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} a \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ b \end{pmatrix} = a\vec{e}_1 + b\vec{e}_2.$$

Remark 2.8. Another notation for $\vec{e}_1$ and $\vec{e}_2$ is $\hat{\imath}$ and $\hat{\jmath}$.

2.2 Inner product and orthogonal projections

Let us start with a definition.

Definition 2.9. Sean $\vec{v} = \begin{pmatrix} v_1 \\ v_2 \end{pmatrix}, \vec{w} = \begin{pmatrix} w_1 \\ w_2 \end{pmatrix}$ vectors in $\mathbb{R}^2$. The *inner product* of $\vec{v}$ and $\vec{w}$ is

$$\langle \vec{v}\,, \vec{w} \rangle := v_1 w_1 + v_2 w_2.$$

The inner product is also called *scalar product* or *dot product* and it can also be denoted by $\vec{v} \cdot \vec{w}$.

We usually prefer the notation $\langle \vec{v}\,, \vec{w} \rangle$ since this notation is used frequently in physics and extends naturally to abstract vector spaces with an inner product. Moreover, the the notation with the dot seems to suggest that the dot product behaves like a usual product, but it does not, see Remark 2.12.

Before we give properties of the inner product, we want to calculate a few examples.

Examples 2.10.

(i) $\left\langle \begin{pmatrix} 2 \\ 3 \end{pmatrix}, \begin{pmatrix} -1 \\ 5 \end{pmatrix} \right\rangle = 2 \cdot (-1) + 3 \cdot 5 = -2 + 15 = 13.$

(ii) $\left\langle \begin{pmatrix} 2 \\ 3 \end{pmatrix}, \begin{pmatrix} 2 \\ 3 \end{pmatrix} \right\rangle = 2^2 + 3^2 = 4 + 9 = 13.$ Note that this is equal to $\left\| \begin{pmatrix} 2 \\ 3 \end{pmatrix} \right\|^2$.

(iii) $\left\langle \begin{pmatrix} 2 \\ 3 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \end{pmatrix} \right\rangle = 2, \quad \left\langle \begin{pmatrix} 2 \\ 3 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix} \right\rangle = 3,$

(iv) $\left\langle \begin{pmatrix} 2 \\ 3 \end{pmatrix}, \begin{pmatrix} -3 \\ 2 \end{pmatrix} \right\rangle = 0.$

Proposition 2.11 (Properties of the inner product). *Let $\vec{u}$, $vecv, \vec{w} \in \mathbb{R}^2$ and $\lambda \in \mathbb{R}$. Then the following holds.*

(i) $\langle \vec{v}\,, \vec{v}\rangle = \|\vec{v}\|^2$. *In dot notation:* $\vec{v}\cdot\vec{v} = \|\vec{v}\|^2$.

(ii) $\langle \vec{u}\,, \vec{v}\rangle = \langle \vec{v}\,, \vec{u}\rangle$. *In dot notation:* $\vec{u}\cdot\vec{v} = \vec{v}\cdot\vec{u}$.

(iii) $\langle \vec{u}\,, \vec{v}+\vec{w}\rangle = \langle \vec{u}\,, \vec{v}\rangle + \langle \vec{u}\,, \vec{w}\rangle$. *In dot notation:* $\vec{u}\cdot(\vec{v}+\vec{w}) = \vec{u}\cdot\vec{v} + \vec{u}\cdots\vec{w})$.

(iv) $\langle \lambda\vec{u}\,, \vec{v}\rangle = \lambda\langle \vec{u}\,, \vec{v}\rangle$. *In dot notation:* $(\lambda\vec{u})\cdot\vec{v} = \lambda(\vec{u}\cdot\vec{v})$.

Proof. Let $\vec{u} = \begin{pmatrix} u_1 \\ u_2 \end{pmatrix}$,$\vec{v} = \begin{pmatrix} v_1 \\ v_2 \end{pmatrix}$ and $\vec{w} = \begin{pmatrix} w_1 \\ w_2 \end{pmatrix}$.

(i) $\langle \vec{v}\,, \vec{v}\rangle = v_1^1 + v_2^2 = \|\vec{v}\|^2$.

(ii) $\langle \vec{u}\,, \vec{v}\rangle = u_1v_1 + u_2v_2 = v_1u_1 + v_2u_2 = \langle \vec{v}\,, \vec{u}\rangle$.

(iii)

$$\begin{aligned}\langle \vec{u}\,, \vec{v}+\vec{w}\rangle &= \left\langle \begin{pmatrix} u_1 \\ u_2 \end{pmatrix}, \begin{pmatrix} v_1+w_1 \\ v_2+w_2 \end{pmatrix}\right\rangle = u_1(v_1+w_1) + u_2(v_2+w_2) = u_1v_1 + u_2v_2 + u_1w_1 + u_2w_2 \\ &= \left\langle \begin{pmatrix} u_1 \\ u_2 \end{pmatrix}, \begin{pmatrix} v_1 \\ v_2 \end{pmatrix}\right\rangle + \left\langle \begin{pmatrix} u_1 \\ u_2 \end{pmatrix}, \begin{pmatrix} w_1 \\ w_2 \end{pmatrix}\right\rangle = \langle \vec{u}\,, \vec{v}\rangle + \langle \vec{u}\,, \vec{w}\rangle.\end{aligned}$$

(iv) $\langle \lambda\vec{u}\,, \vec{v}\rangle = \langle \begin{pmatrix} \lambda u_1 \\ \lambda u_2 \end{pmatrix}, \begin{pmatrix} v_1 \\ v_2 \end{pmatrix}\rangle = \lambda u_1v_1 + \lambda u_2v_2 = \lambda(u_1v_1 + u_2v_2) = \lambda\langle \vec{u}\,, \vec{v}\rangle$. □

Remark 2.12. Observe that the proposition says that the inner product is commutative and distributive, so has some properties of "usual multiplication" that we are used to from the product in $\mathbb{R}$ or $\mathbb{C}$, but there are some properties that show that the inner product is NOT a product:

(a) The inner products takes to vectors and gives back a number, so it gives back an object which is not of the same type as the two things we put in.

(b) In Example 2.10(iv) we saw that it may happen that $\vec{v} \neq \vec{0}$ and $\vec{w} \neq \vec{0}$ but still $\langle \vec{v}\,, \vec{w}\rangle = 0$, something that is impossible for a "decent" product.

(c) Given a vector $\vec{v} \neq 0$ and a number $c \in \mathbb{R}$, there are many solutions of the equation $\langle \vec{v}\,, \vec{x}\rangle = c$ for the vector $\vec{x}$, in stark contrast to the usual product in $\mathbb{R}$ or $\mathbb{C}$. As an example, look at Example 2.10(i) and (ii). Therefore it makes NO sense to write something like $\vec{v}^{-1}$.

(d) There is no such thing as a neutral element for scalar multiplication.

Now let us see what the inner product is good for. We will see that inner product between two vectors is connected to the angle between them and it will help us to define orthogonal projections of one vector onto another.
Let us start with a definition.

Definition 2.13. Let $\vec{v}, \vec{w}$ be vectors in $\mathbb{R}^2$. The *angle between $\vec{v}$ and $\vec{w}$* is the smallest nonnegative angle between them, see Figure 2.6. It is denoted by $\sphericalangle(\vec{v}, \vec{w})$.

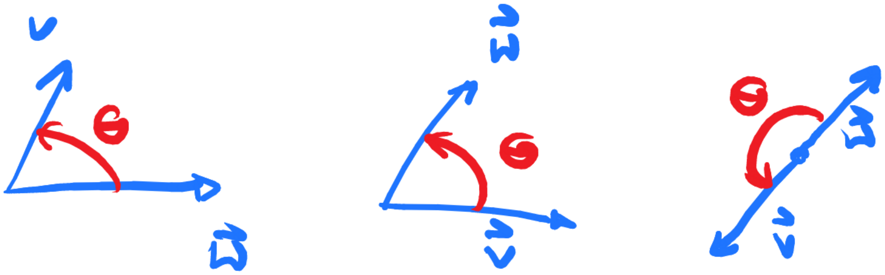

FIGURE 2.6: Angle between two vectors. XXXXXX Faltan π y 0.

The following properties of the angle are easy to see.

Proposition 2.14. (i) *Note that by definition,* $\sphericalangle(\vec{v}, \vec{w}) \in [0, \pi]$.

(ii) $\sphericalangle(\vec{v}, \vec{w}) = \sphericalangle(\vec{w}, \vec{v})$.

(iii) *If* $\lambda > 0$, *then* $\sphericalangle(\lambda\vec{v}, \vec{w}) = \sphericalangle(\vec{v}, \vec{w})$.

(iv) *If* $\lambda < 0$, *then* $\sphericalangle(\lambda\vec{v}, \vec{w}) = \pi - \sphericalangle(\vec{v}, \vec{w})$.

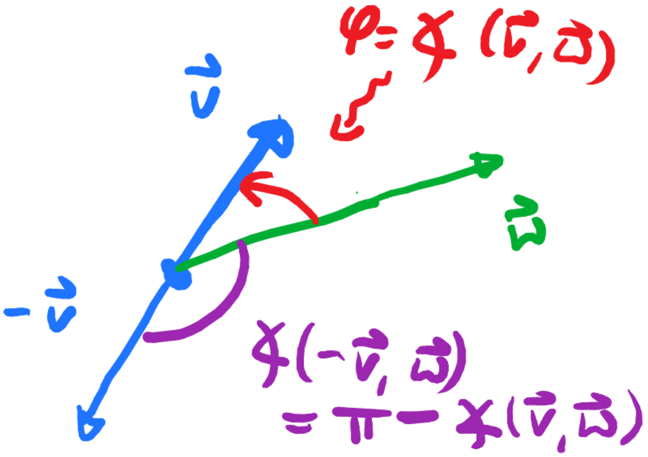

FIGURE 2.7: Angle between vectors $\vec{v}$ and $\vec{w}$.

Definition 2.15. (a) Two vectors $\vec{v}$ and $\vec{w}$ are called *parallel* if $\sphericalangle(\vec{v}, \vec{w}) = 0$ or π. In this case we use the notation $\vec{v} \parallel \vec{w}$.

(b) Two vectors $\vec{v}$ and $\vec{w}$ are called *orthogonal* or *perpendicular* if $\sphericalangle(\vec{v}, \vec{w}) = \pi/2$. In this case we use the notation $\vec{v} \perp \vec{w}$.

The following properties should be known from geometry. We will proof them after Theorem 2.19.

Proposition 2.16. *Let $\vec{v}, \vec{w}$ be vectors in $\mathbb{R}^2$. Then:*

(i) *$\vec{v} \parallel \vec{w}$ and $\vec{v} \neq \vec{0}$, then there exists $\lambda \in \mathbb{R}$ such that $\vec{w} = \lambda \vec{v}$.*

(ii) *If $\vec{v} \parallel \vec{w}$ and $\lambda, \mu \in \mathbb{R}$, then also $\lambda \vec{v} \parallel \mu \vec{w}$.*

(iii) *If $\vec{v} \perp \vec{w}$ and $\lambda, \mu \in \mathbb{R}$, then also $\lambda \vec{v} \perp \mu \vec{w}$.*

Remark 2.17. Observe that (i) is wrong if we do not assume that $\vec{v} \neq \vec{0}$ because if $\vec{v} = \vec{0}$, then it is parallel to every vector $\vec{w}$ in $\mathbb{R}^2$, but there is no $\lambda \in \mathbb{R}$ such that $\lambda \vec{v}$ could ever become different from $\vec{0}$.

Further observe that the reverse direction in (ii) is true only if $\lambda \neq 0$ and $\mu \neq 0$.

Without proof, we state the following theorem which should be known.

Theorem 2.18 (Cosine Theorem). *Let a, b, c be the sides or a triangle and let φ be the angle between the sides a and b. Then*

$$c^2 = a^2 + b^2 - 2ab \cos \varphi. \tag{2.1}$$

Theorem 2.19. *Let $\vec{v}, \vec{w} \in \mathbb{R}^2$ and let $\varphi = \sphericalangle(\vec{v}, \vec{w})$. Then*

$$\boxed{\langle \vec{v}\,, \vec{w} \rangle = \|\vec{v}\| \|\vec{w}\| \cos \varphi.}$$

Proof. The vectors $\vec{v}$ and $\vec{w}$ define a triangle in $\mathbb{R}^2$, see Figure 2.8

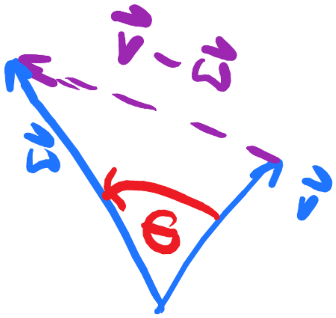

Figure 2.8: Triangle given by $\vec{v}$ and $\vec{w}$.

Now we apply the cosine theorem with $a = \|\vec{v}\|$, $b = \|\vec{w}\|$, $c = \|\vec{v} - w\|$. We obtain

$$\|\vec{v} - \vec{w}\|^2 = \|\vec{v}\|^2 + \|\vec{w}\|^2 - 2\|\vec{v}\|\|\vec{w}\| \cos\varphi. \tag{2.2}$$

On the other hand,

$$\begin{aligned} \|\vec{v} - \vec{w}\|^2 &= \langle \vec{v} - \vec{w}\,, \vec{v} - \vec{w} \rangle = \langle \vec{v}\,, \vec{v} \rangle - \langle \vec{v}\,, \vec{w} \rangle - \langle \vec{w}\,, \vec{v} \rangle + \langle \vec{w}\,, \vec{w} \rangle = \langle \vec{v}\,, \vec{v} \rangle - 2\langle \vec{v}\,, \vec{w} \rangle + \langle \vec{w}\,, \vec{w} \rangle \\ &= \|\vec{v}\|^2 - 2\langle \vec{v}\,, \vec{w} \rangle + \|\vec{w}\|^2. \end{aligned} \tag{2.3}$$

Comparison of (2.2) and (2.3) show that

$$\|\vec{v}\|^2 + \|\vec{w}\|^2 - 2\|\vec{v}\|\|\vec{w}\| \cos\varphi = \|\vec{v}\|^2 - 2\langle \vec{v}\,, \vec{w} \rangle + \|\vec{w}\|^2,$$

which gives the claimed formula. ☐

A very important consequence of this theorem is that we can now determine if two vectors ara parallel or perpendicular to each other by simply calculating their inner product as can be seen from the following corollary.

Corollary 2.20. *Let* $\vec{v}, \vec{w} \in \mathbb{R}^2$ *and* $\varphi = \sphericalangle(\vec{v}, \vec{w})$. *Then:*

(i) $|\langle \vec{v}\,, \vec{w} \rangle| \leq \|\vec{v}\|\,\|\vec{w}\|$.

(ii) $\vec{v} \parallel \vec{w} \qquad \iff \qquad \|\vec{v}\|\,\|\vec{w}\| = |\langle \vec{v}\,, \vec{w} \rangle|$.

(iii) $\vec{v} \perp \vec{w} \qquad \iff \qquad \langle \vec{v}\,, \vec{w} \rangle = 0$.

Proof. (i) From Theorem 2.19 we have that $|\langle \vec{v}\,, \vec{w} \rangle| = \|\vec{v}\|\,\|\vec{w}\| \cos\varphi \leq \|\vec{v}\|\,\|\vec{w}\|$ since $0 \leq \cos\varphi \leq 1$.

The claims in (ii) and (iii) are clear if one of the vectors is equal to $\vec{0}$ since the zero vector is parallel and orthogonal to every vector in $'R^2$. So let us assume now that $\vec{v} \neq \vec{0}$ and $\vec{w} \neq \vec{0}$.

(ii) From Theorem **??** we have that $|\langle \vec{v}\,, \vec{w}\rangle| = \|\vec{v}\|\,\|\vec{w}\|$ if and only if $\cos\varphi = 1$. This is the case if and only if $\varphi = 0$ or π, that is, if and only if $\vec{v}$ and $\vec{w}$ are parallel.

(iii) From Theorem **??** we have that $|\langle \vec{v}\,, \vec{w}\rangle| = 0$ if and only if $\cos\varphi = 0$. This is the case if and only if $\varphi = \pi/2$, that is, if and only if $\vec{v}$ and $\vec{w}$ are perpendicular. ☐

With this corollary, the proof of Proposition 2.16(ii) and (iii) is now easy and left to the reader.

Example 2.21. Theorem **??** lets us calculate the angle of a given vector with the x-axis easily (see Figure 2.9):

$$\cos\varphi_x = \frac{\langle \vec{v}\,, \vec{\mathrm{e}}_1\rangle}{\|\vec{v}\|\|\vec{e}_1\|}, \qquad \cos\varphi_y = \frac{\langle \vec{v}\,, \vec{\mathrm{e}}_2\rangle}{\|\vec{v}\|\|\vec{e}_2\|}.$$

If we now use that $\|\vec{e}_1\| = \|\vec{e}_2\| = 1$ and that $\langle \vec{v}\,, \vec{\mathrm{e}}_1\rangle = v_1$ and $\langle \vec{v}\,, \vec{\mathrm{e}}_2\rangle = v_2$, then

$$\cos\varphi_x = \frac{v_1}{\|\vec{v}\|}, \qquad \cos\varphi_y = \frac{v_2}{\|\vec{v}\|}.$$

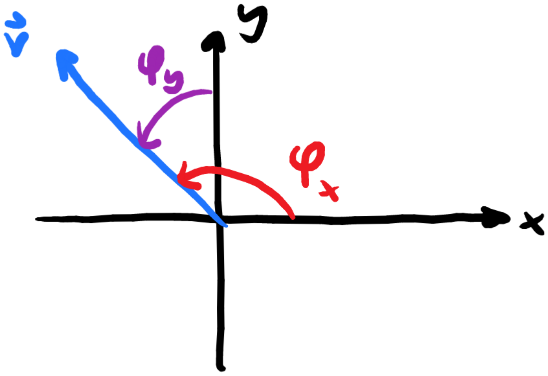

FIGURE 2.9: Angle of $\vec{v}$ with the axes.

Orthogonal Projections in $\mathbb{R}^2$.

Let $\vec{v}$ and $\vec{w}$ be vectors in $\mathbb{R}^2$ and $\vec{w} \neq \vec{0}$. We want to find the orthogonal projection of $\vec{v}$ onto $\vec{w}$. Geometrically, we find it as follows: We move $\vec{v}$ such that its initial point coincides with that of $\vec{w}$. Then we extend $\vec{w}$ to a line and construct a line that passes through the tip of $\vec{v}$. The vector from the initial point to the intersection of the two lines is the see Figure 2.10

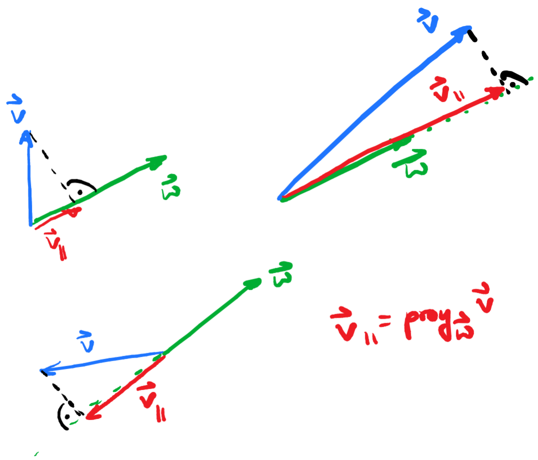

FIGURE 2.10: Orthogonal projections in $\mathbb{R}^2$.

We denote the orthogonal projection of $\vec{v}$ onto $\vec{w}$ by $\text{proj}_{\vec{w}} \vec{v}$, or sometimes by $\vec{v}_{\|}$ it is clear on which vector we are projecting. By construction of $\text{proj}_{\vec{w}} \vec{v}$ it is clear that

- $\text{proj}_{\vec{w}} \vec{v}$ is parallel to $\vec{w}$,
- $\vec{v} - \text{proj}_{\vec{w}} \vec{v}$ is orthogonal to $\vec{w}$. Therefore, we sometimes write $\vec{v}_{\perp} = \vec{v} - \text{proj}_{\vec{w}} \vec{v}$.

This procedure allows us to write $\vec{v}$ as sum of a vector parallel to $\vec{w}$ and one orthogonal to $\vec{w}$. How we can calculate these two vectors, is the content of the next theorem.

Theorem 2.22. *Let $\vec{v}$ and $\vec{w}$ be vectors in $\mathbb{R}^2$ and $\vec{w} \neq \vec{0}$. Then*

$$\text{proj}_{\vec{w}} \vec{v} = \frac{\langle \vec{v}\,, \vec{w}\rangle}{\|\vec{w}\|^2}\, \vec{w}. \tag{2.4}$$

Before we prove the formula, note that it seems to make sense. The right hand side is a multiple of $\vec{w}$, so it is parallel to $\vec{w}$ as it should be. Moreover, it does not depend on $\|w\|$ as it should be because it should not matter if we project on $\vec{w}$ or on $5\vec{w}$ or on $-0.4\vec{w}$; only the direction of $\vec{w}$ matters, not its length.

Proof. Let $\vec{v}_{\|} = \text{proj}_{\vec{w}} \vec{v}$ and $\vec{v}_{\perp} = \vec{v} - \vec{v}_{\|}$. Then $\vec{v} = \vec{v}_{\|} + \vec{v}_{\perp}$. Since $\vec{v}_{\|} \parallel \vec{w}$, there exists a $\lambda \in \mathbb{R}$ such that $\vec{v}_{\|} = \lambda \vec{w}$, so we only need to determine λ. For this, we write

$$\begin{aligned}
& \vec{v} = \lambda \vec{w} + \vec{v}_{\perp} \\
\implies \quad & \langle \vec{v}\,, \vec{w}\rangle = \langle \lambda \vec{w} + \vec{v}_{\perp}\,, \vec{w}\rangle = \langle \lambda \vec{w}\,, \vec{w}\rangle + \underbrace{\langle \vec{v}_{\perp}\,, \vec{w}\rangle}_{=0 \text{ since } \vec{v}_{\perp} \perp \vec{w}} = \langle \lambda \vec{w}\,, \vec{w}\rangle = \lambda \langle \vec{w}\,, \vec{w}\rangle = \lambda \|\vec{w}\|^2 \\
\implies \quad & \lambda = \frac{\langle \vec{v}\,, \vec{w}\rangle}{\|\vec{w}\|^2}
\end{aligned}$$

So it follows that

$$\operatorname{proj}_{\vec{w}} \vec{v} = \vec{v}_{\|} = \lambda \vec{w} = \frac{\langle \vec{v}\,, \vec{w}\rangle}{\|\vec{w}\|^2}\vec{w}. \qquad \square$$

Remark 2.23. (i) $\operatorname{proj}_{\vec{w}} \vec{v}$ depends only of the direction of $\vec{w}$. It does not depend on its length.

Proof. By our geometric intuition, this should be clear. But we can see this also from the formula. Suppose we want to project on $c\vec{w}$ for some $c \in \mathbb{R} \setminus \{0\}$. Then

$$\operatorname{proj}_{c\vec{w}} \vec{v} = \frac{\langle \vec{v}\,, c\vec{w}\rangle}{\|c\vec{w}\|^2}(c\vec{w}) = \frac{c\langle \vec{v}\,, \vec{w}\rangle}{c^2\|\vec{w}\|^2}(c\vec{w}) = \frac{\langle \vec{v}\,, \vec{w}\rangle}{\|\vec{w}\|^2}\vec{w} = \operatorname{proj}_{\vec{w}} \vec{v}. \qquad \square$$

(ii) For every $c \in \mathbb{R}$, we have that $\operatorname{proj}_{\vec{w}}(c\vec{v}) = c \operatorname{proj}_{\vec{w}} \vec{v}$.

Proof. Again, by geometric considerations, this should be clear. The corresponding calculus is

$$\operatorname{proj}_{\vec{w}}(c\vec{v}) = \frac{\langle c\vec{v}\,, \vec{w}\rangle}{\|\vec{w}\|^2}\vec{w} = \frac{c\langle \vec{v}\,, \vec{w}\rangle}{\|\vec{w}\|^2}\vec{w} = c \operatorname{proj}_{\vec{w}} \vec{v}. \qquad \square$$

(iii) As special cases of the above, we find $\operatorname{proj}_{\vec{w}}(-\vec{v}) = \operatorname{proj}_{\vec{w}} \vec{v}$ and $\operatorname{proj}_{-\vec{w}} \vec{v} = -\operatorname{proj}_{\vec{w}} \vec{v}$.

(iv) $\vec{v} \parallel \vec{w} \quad\Longrightarrow\quad \operatorname{proj}_{\vec{w}} \vec{v} = \vec{v}$.

(v) $\vec{v} \perp \vec{w} \quad\Longrightarrow\quad \operatorname{proj}_{\vec{w}} \vec{v} = \vec{0}$.

(vi) $\operatorname{proj}_{\vec{w}} \vec{v}$ is the unique vector in $\mathbb{R}^2$ such that

$$\vec{v} - \operatorname{proj}_{\vec{w}} \vec{v} \perp \vec{v} \quad \text{and} \quad \operatorname{proj}_{\vec{w}} \vec{v} \parallel \vec{w}.$$

We end this section with some examples.

Example 2.24. Let $\vec{u} = 2\vec{\mathrm{e}}_1 + 3\vec{\mathrm{e}}_2$, $\vec{v} = 4\vec{\mathrm{e}}_1 - \vec{\mathrm{e}}_2$.

(i) $\operatorname{proj}_{\vec{\mathrm{e}}_1} \vec{u} = \frac{\langle \vec{u}\,, \vec{\mathrm{e}}_1\rangle}{\|\vec{\mathrm{e}}_1\|^2}\vec{\mathrm{e}}_1 = \frac{2}{1^2}\vec{\mathrm{e}}_1 = 2\vec{\mathrm{e}}_1$.

(ii) $\operatorname{proj}_{\vec{\mathrm{e}}_2} \vec{u} = \frac{\langle \vec{u}\,, \vec{\mathrm{e}}_2\rangle}{\|\vec{\mathrm{e}}_2\|^2}\vec{\mathrm{e}}_2 = \frac{3}{1^2}\vec{\mathrm{e}}_2 = 3\vec{\mathrm{e}}_2$.

(iii) Similarly, we can calculate $\operatorname{proj}_{\vec{\mathrm{e}}_1} \vec{v} = 4\vec{e}_1$, $\operatorname{proj}_{\vec{\mathrm{e}}_2} \vec{v} = -\vec{e}_2$.

(iv) $\operatorname{proj}_{\vec{u}} \vec{v} = \frac{\langle \vec{u}\,, \vec{v}\rangle}{\|\vec{u}\|^2}\vec{u} = \dfrac{\left\langle \begin{pmatrix} 2 \\ 3 \end{pmatrix}, \begin{pmatrix} 5 \\ -1 \end{pmatrix} \right\rangle}{\|\vec{u}\|^2}\vec{u} = \frac{8-3}{2^2+3^2}\vec{u} = \frac{5}{13}\vec{u} = \frac{5}{13}\begin{pmatrix} 2 \\ 3 \end{pmatrix}$.

(v) $\operatorname{proj}_{\vec{v}} \vec{u} = \frac{\langle \vec{v}\,, \vec{u}\rangle}{\|\vec{v}\|^2}\vec{v} = \dfrac{\left\langle \begin{pmatrix} 4 \\ -1 \end{pmatrix}, \begin{pmatrix} 2 \\ 3 \end{pmatrix} \right\rangle}{\|\vec{u}\|^2}\vec{u} = \frac{8-3}{4^2+(-1)^2}\vec{v} = \frac{5}{17}\vec{v} = \frac{5}{17}\begin{pmatrix} 4 \\ -1 \end{pmatrix}$.

Example 2.25 (Angle with coordinate axes). Let $\vec{v} = \begin{pmatrix} a \\ b \end{pmatrix} \in \mathbb{R}^2 \setminus \{\vec{0}\}$.

Then $\cos\sphericalangle(\vec{v}, \vec{e}_1) = \frac{a}{\|\vec{v}\|}$, $\cos\sphericalangle(\vec{v}, \vec{e}_2) = \frac{b}{\|\vec{v}\|}$, hence

$$\vec{v} = \begin{pmatrix} a \\ b \end{pmatrix} = \|\vec{v}\| \begin{pmatrix} \cos\sphericalangle(\vec{v}, \vec{e}_1) \\ \cos\sphericalangle(\vec{v}, \vec{e}_2) \end{pmatrix}.$$

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

In this section we extend our calculations from $\mathbb{R}^2$ to $\mathbb{R}^3$. Recall that $\mathbb{R}^3$ is the space of all points $P(a, b, c)$ with $a, b, c \in \mathbb{R}$. This is a model for our usual physical everyday space. Recall that the distance between two points $P(p_1, p_2, p_3)$ and $Q(q_1, q_2, q_3)$ is $\overline{PQ} = \sqrt{(q_1 - p_1)^2 + (q_2 - p_2)^2 + (q_3 - p_3)^2}$.

As in $\mathbb{R}^2$, we can identify every point in $\mathbb{R}^3$ with the arrow that starts in the origin of coordinate system and ends in the given point. The set of all arrows with the same length and the same direction is called a vector in $\mathbb{R}^3$. Again, we denote a vector in $\mathbb{R}^3$ as a column

$$\vec{v} = \begin{pmatrix} a \\ b \\ c \end{pmatrix}.$$

In order to save space, we will also use the notation $(a, b, c)^t$, where, as in $\mathbb{R}^2$, the superscript t stands for *transposed.*

Definition 2.26. Let $\vec{v} = \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix}, \vec{w} = \begin{pmatrix} w_1 \\ w_2 \\ w_3 \end{pmatrix} \in \mathbb{R}^3$ and $c \in \mathbb{R}$. We define the sum of $\vec{v}$ and $\vec{w}$ and the product of the scalar c with the vector $\vec{v}$ as follows:

$$\vec{v} + \vec{w} = \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix} + \begin{pmatrix} w_1 \\ w_2 \\ w_3 \end{pmatrix} = \begin{pmatrix} v_1 + w_1 \\ v_2 + w_2 \\ v_3 + w_3 \end{pmatrix}, \qquad c\vec{v} = \begin{pmatrix} cv_1 \\ cv_2 \\ cv_3 \end{pmatrix}.$$

It is easy to see that $\mathbb{R}^3$ with this sum and product satisfies the vector space axioms on page 19.

As in $\mathbb{R}^2$, we define an *inner product*

$$\langle \vec{v}\,, \vec{w} \rangle = \left\langle \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix}, \begin{pmatrix} w_1 \\ w_2 \\ w_3 \end{pmatrix} \right\rangle = v_1 w_1 + v_2 w_2 + v_3 + w_3$$

and a *norm*

$$\|\vec{v}\| = \left\| \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix} \right\| := \sqrt{v_1^2 + v_2^2 + v_3^2}.$$

We also use the words *magnitude* or *length* of $\vec{w}$. .

Two vectors in $\mathbb{R}^3$ which are not parallel generate a plane. Then we can measure the angle between the two vectors in this plane as if it was $\mathbb{R}^2$ and we call it the *angle between the two vectors.*
As in $\mathbb{R}^2$, we have the following properties:

(i) *Symmetry of the inner product:* For all vectors $\vec{v}, \vec{w} \in \mathbb{R}^3$, we have that $\langle \vec{v}\,, \vec{w} \rangle = \langle \vec{w}\,, \vec{v} \rangle$.

(ii) *Bilinearity of the inner product:* For all vectors $\vec{u}, \vec{v}, \vec{w} \in \mathbb{R}^3$ and all $c \in \mathbb{R}$, we have that $\langle \vec{u}\,, \vec{v} + \vec{w} \rangle = \langle \vec{u}\,, \vec{v} \rangle + c\langle \vec{u}\,, \vec{w} \rangle$.

(iii) *Relation of the inner product with the angle between vectors:* Let $\vec{v}, \vec{w} \in \mathbb{R}^3$ and let $\varphi = \sphericalangle(\vec{v}, \vec{w})$. Then

$$\langle \vec{v}\,, \vec{w} \rangle = \|\vec{v}\|\,\|\vec{w}\|\,\cos\varphi.$$

Remark 2.27. Actually, the inner product usually is used to *define* the angle between two vectors by the formula above.

In particular, we have (cf. Proposition 2.16):

$$\begin{array}{llllll}
\text{(a)} & \vec{v} \parallel \vec{w} & \iff & \sphericalangle(\vec{v}, \vec{w}) \in \{0, \pi\} & \iff & |\langle \vec{v}\,, \vec{w} \rangle| = \|\vec{v}\|\,\|\vec{w}\| \\
\text{(b)} & \vec{v} \perp \vec{w} & \iff & \sphericalangle(\vec{v}, \vec{w}) = \pi/2 & \iff & \langle \vec{v}\,, \vec{w} \rangle = 0.
\end{array}$$

(iv) *Relation of norm and inner product:* For all vectors $\vec{v} \in \mathbb{R}^3$, we have that $\|\vec{v}\|^2 = \langle \vec{v}\,, \vec{v} \rangle$.

(v) *Properties of the norm:* For all vectors $\vec{v}, \vec{w} \in \mathbb{R}^3$ and scalars $c \in \mathbb{R}$, we have that $\|c\vec{v}\| = |c|\|\vec{v}\|$ and $\|\vec{v} + \vec{w}\| \leq \|\vec{v}\| + \|\vec{w}\|$.

(vi) *Orthogonal projections of one vector onto another:* For all vectors $\vec{v}, \vec{w} \in \mathbb{R}^3$ the *orthogonal projection of $\vec{v}$ onto $\vec{w}$* is

$$\operatorname{proj}_{\vec{w}} \vec{v} = \frac{\langle \vec{v}\,, \vec{w} \rangle}{\|\vec{w}\|^2}\,\vec{w}.$$

As in $\mathbb{R}^3$, we have three sort of special vectors which are parallel to the coordinate system:

$$\vec{\mathrm{e}}_1 := \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \quad \vec{\mathrm{e}}_2 := \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \quad \vec{\mathrm{e}}_3 := \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}.$$

Another notation for them is $\hat{\imath}, \hat{\jmath}, \hat{k}$.
For a given vector $\vec{v} \neq \vec{0}$, we can now easily determine its angle with the coordinate axes:

$$\begin{aligned}
\varphi_x = \sphericalangle(\vec{v}, \vec{\mathrm{e}}_1) \quad &\implies \quad \cos\varphi_x = \frac{\langle \vec{v}\,, \vec{e}_1 \rangle}{\|\vec{v}\|\,\|\vec{\mathrm{e}}_1\|} = \frac{v_1}{\|\vec{v}\|}, \\
\varphi_y = \sphericalangle(\vec{v}, \vec{\mathrm{e}}_2) \quad &\implies \quad \cos\varphi_x = \frac{\langle \vec{v}\,, \vec{e}_2 \rangle}{\|\vec{v}\|\,\|\vec{\mathrm{e}}_2\|} = \frac{v_2}{\|\vec{v}\|}, \\
\varphi_z = \sphericalangle(\vec{v}, \vec{\mathrm{e}}_3) \quad &\implies \quad \cos\varphi_x = \frac{\langle \vec{v}\,, \vec{e}_3 \rangle}{\|\vec{v}\|\,\|\vec{\mathrm{e}}_3\|} = \frac{v_3}{\|\vec{v}\|}.
\end{aligned}$$

Esto nos dice que

$$\vec{v} = \|\vec{v}\| \begin{pmatrix} \cos\varphi_x \\ \cos\varphi_y \\ \cos\varphi_z \end{pmatrix}.$$

If we take the norm both sides of the equation, we find

$$(\cos\varphi_x)^2 + (\cos\varphi_y)^2 + (\cos\varphi_z)^2 = 1.$$

2.4 Cross product

In this section we define the so-called *cross product*. Another name for it its *vector product*. It takes two vectors and gives back two vectors. It does have several properties which makes it look like a product, however we will see that it is NOT a product. Here is the definition.

Definition 2.28 (Cross product). Let $\vec{v} = \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix}, \vec{w} = \begin{pmatrix} w_1 \\ w_2 \\ w_3 \end{pmatrix} \in \mathbb{R}^3$. Their *cross product* or *vector product* is

$$\vec{v} \times \vec{w} = \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix} \times \begin{pmatrix} w_1 \\ w_2 \\ w_3 \end{pmatrix} := \begin{pmatrix} v_2w_3 - v_3w_2 \\ v_3w_1 - v_1w_3 \\ v_1w_2 - v_2w_1 \end{pmatrix}.$$

Remark 2.29. The cross product exists only in $\mathbb{R}^3$!

Before we collect some easy properties of the cross product, let us calculate a few examples.

Examples 2.30. Let $\vec{u} = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}, \vec{v} = \begin{pmatrix} 5 \\ 6 \\ 7 \end{pmatrix}$.

- $\vec{u} \times \vec{v} = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} \times \begin{pmatrix} 5 \\ 6 \\ 7 \end{pmatrix} = \begin{pmatrix} 2\cdot7 - 3\cdot6 \\ 3\cdot5 - 1\cdot7 \\ 1\cdot6 - 2\cdot5 \end{pmatrix} = \begin{pmatrix} 14-18 \\ 15-7 \\ 6-10 \end{pmatrix} = \begin{pmatrix} -4 \\ 8 \\ -4 \end{pmatrix}.$
- $\vec{v} \times \vec{u} = \begin{pmatrix} 5 \\ 6 \\ 7 \end{pmatrix} \times \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} = \begin{pmatrix} 6\cdot3 - 7\cdot2 \\ 7\cdot1 - 3\cdot5 \\ 5\cdot2 - 6\cdot1 \end{pmatrix} = \begin{pmatrix} 18-14 \\ 7-15 \\ 10-6 \end{pmatrix} = \begin{pmatrix} 4 \\ -8 \\ 4 \end{pmatrix}.$
- $\vec{v} \times \vec{e}_1 = \begin{pmatrix} 5 \\ 6 \\ 7 \end{pmatrix} \times \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 6\cdot0 - 7\cdot0 \\ 7\cdot0 - 7\cdot1 \\ 5\cdot0 - 6\cdot1 \end{pmatrix} = \begin{pmatrix} 0 \\ -7 \\ -6 \end{pmatrix}.$

Proposition 2.31 (Properties of the cross product). *Let $\vec{u}, \vec{v}, \vec{w} \in \mathbb{R}^3$ and let $c \in \mathbb{R}$. Then:*

(i) $\vec{u} \times \vec{0} = \vec{0} \times \vec{u} = \vec{0}$.

(ii) $\vec{u} \times \vec{v} = -\vec{v} \times \vec{u}$.

(iii) *$\vec{u}\times(\vec{v}+\vec{w})=(\vec{u}\times\vec{v})+(\vec{u}\times\vec{w})$.*

(iv) *$(c\vec{u})\times\vec{v}=c(\vec{u}\times\vec{v})$.*

(v) *$\vec{u}\parallel\vec{v} \quad\Longrightarrow\quad \vec{u}\times\vec{v}=\vec{0}$. In particular, $\vec{v}\times\vec{v}=\vec{0}$.*

(vi) *$\langle\vec{u}\,,\vec{v}\times\vec{w}\rangle=\langle\vec{u}\times\vec{v}\,,\vec{w}\rangle$.*

(vii) *$\langle\vec{u}\,,\vec{u}\times\vec{v}\rangle=0$ and $\langle\vec{v}\,,\vec{u}\times\vec{v}\rangle=0$, in particular*

$$\boxed{\vec{v}\perp\vec{v}\times\vec{u}, \quad \vec{u}\perp\vec{v}\times\vec{u}}$$

that means that the vector $\vec{v}\times\vec{w}$ is orthogonal to both $\vec{v}$ and $\vec{w}$.

Proof. The proofs of the formulas (i) to (v) are easy calculations (you should do them!).

(vi) The proof is a long but straightforward calculation:

$$\begin{aligned}\langle\vec{u}\,,\vec{v}\times\vec{w}\rangle &= \left\langle \begin{pmatrix}u_1\\u_2\\u_3\end{pmatrix}, \begin{pmatrix}v_1\\v_2\\v_3\end{pmatrix}\times\begin{pmatrix}w_1\\w_2\\w_3\end{pmatrix}\right\rangle = \left\langle \begin{pmatrix}u_1\\u_2\\u_3\end{pmatrix}, \begin{pmatrix}v_2w_3-v_3w_2\\v_3w_1-w_3v_1\\v_1w_2-v_2w_1\end{pmatrix}\right\rangle\\ &= u_1(v_2w_3-v_3w_2)+u_2(v_3w_1-v_1w_3)+u_3(v_1w_2-v_2w_1)\\ &= u_1v_2w_3-u_1v_3w_2+u_2v_3w_1-u_2v_1w_3+u_3v_1w_2-u_3v_2w_1\\ &= u_2v_3w_1-u_3v_2w_1+u_3v_1w_2-u_1v_3w_2+u_1v_2w_3-u_2v_1w_3\\ &= (u_2v_3-u_3v_2)w_1+(u_3v_1-u_1v_3)w_2+(u_1v_2-u_2v_1)w_3\\ &= \langle\vec{u}\times\vec{v}\,,\vec{w}\rangle.\end{aligned}$$

(vii) It follows from (vi) and (v) that

$$\langle\vec{u}\,,\vec{u}\times\vec{v}\rangle=\langle\vec{u}\times\vec{u}\,,\vec{v}\rangle=\langle\vec{0}\,,\vec{v}\rangle=0.$$ ☐

Note that the cross product is distributive but it is not commutative nor associative.

Recall that for the inner product we proved the formula $\langle\vec{v}\,,\vec{w}\rangle=\|\vec{v}\|\,\|\vec{w}\|\cos\varphi$ where φ is the angle between the two vectors, see Theorem 2.19. In the next theorem we will prove a similar relation for the cross product.

Theorem 2.32. *Let $\vec{v},\vec{w}$ be vectors in $\mathbb{R}^3$ and let φ be the angle between them. Then*

$$\boxed{\|\vec{v}\times\vec{w}\|=\|\vec{v}\|\,\|\vec{w}\|\sin\varphi}$$

Proof. A long, but straightforward calculations shows that $\|\vec{v}\times\vec{w}\|^2=\|\vec{u}\|^2\|\vec{w}\|^2-\langle\vec{v}\,,\vec{w}\rangle^2$. Now it follows from Theorem 2.19 that

$$\begin{aligned}\|\vec{v}\times\vec{w}\|^2 &= \|\vec{u}\|^2\|\vec{w}\|^2-\langle\vec{v}\,,\vec{w}\rangle^2=\|\vec{u}\|^2\|\vec{w}\|^2-\|\vec{v}\|^2\|\vec{w}\|^2(\cos\varphi)^2\\ &= \|\vec{u}\|^2\|\vec{w}\|^2(1-(\cos\varphi)^2)=\|\vec{u}\|^2\|\vec{w}\|^2(\sin\varphi)^2.\end{aligned}$$

Observe that $\sin\varphi\geq 0$ because $\varphi\in[0,\pi]$. So if we take the square root we we do not need to take the absolute value and we arrive at the claimed formula. ☐

Application: Area of a parallelogram and volume of a parelellepiped

Area of a parallelogram

Let $\vec{v}$ and $\vec{w}$ be two vectors in $\mathbb{R}^3$. Then they define a parallelogram (if the vectors are parallel or one of them is equal to $\vec{0}$, it is a *degenerate parallelogram*).

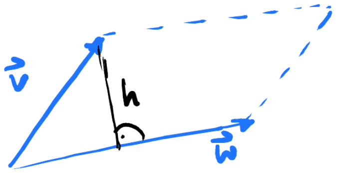

FIGURE 2.11: Parallelogram spanned by $\vec{v}$ and $\vec{w}$.

Proposition 2.33 (Area of a parallelogram). *The area of the parallelogram spanned by the vectors $\vec{v}$ and $\vec{w}$ is*

$$A = \|\vec{v} \times \vec{w}\|. \tag{2.5}$$

Proof. The area of a parallelogram is the product of the length of its base with the height. We can take $\vec{w}$ as base. Let φ be the angle between $\vec{w}$ and $\vec{v}$. Then we obtain that $h = \|\vec{v}\| \sin\varphi$ and therefore, with the help of Theorem 2.32

$$A = \|\vec{w}\|h = \|\vec{w}\|\|\vec{v}\| \sin\varphi = \|\vec{v} \times \vec{w}\|. \qquad \square$$

Note that in the case when $\vec{v}$ and $\vec{w}$ are parallel, this gives the right answer $A = 0$.

Any three vectors in $\mathbb{R}^3$ define a parallelepiped.

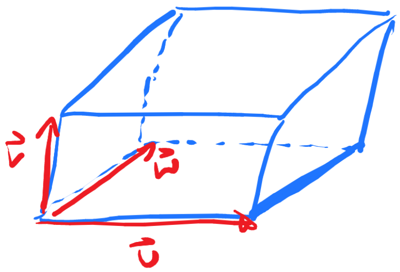

FIGURE 2.12: Parallelepiped spanned by $\vec{u}, \vec{v}, \vec{w}$.

Proposition 2.34 (Volume of a parallelepiped). *The volume of the parallelepiped spanned by the vectors $\vec{u}$, $\vec{v}$ and $\vec{w}$ is*

$$V = \|\vec{u}(\vec{v} \times \vec{w})\|. \tag{2.6}$$

Proof. The volume of a parallelepiped is the product of the area of its base with the height. Let us take the parallelogram spanned by $\vec{v}, \vec{w}$ as base. If $\vec{v}$ and $\vec{w}$ are parallel or one or them is equal to $\vec{0}$, then (2.6) is true because $V = 0$ and $\vec{v} \times \vec{w} = \vec{0}$ in this case.
Now let us assume that they are not parallel. By Proposition 2.33 we already know that its base has area $A = \|\vec{v} \times \vec{w}\|$. The height is the length of the orthogonal projection of $\vec{u}$ onto the normal vector of the plane spanned by $\vec{v}$ and $\vec{w}$. We already know that $\vec{v} \times \vec{w}$ is such a normal vector. Hence we obtain that

$$h = \|\operatorname{proj}_{\vec{v}\times\vec{w}} \vec{u}\| = \left\| \frac{\langle \vec{u}\,, \vec{v} \times \vec{w} \rangle}{\|\vec{v} \times \vec{w}\|^2} \vec{v} \times \vec{w} \right\| = \frac{|\langle \vec{u}\,, \vec{v} \times \vec{w} \rangle|}{\|\vec{v} \times \vec{w}\|^2} \|\vec{v} \times \vec{w}\| = \frac{|\langle \vec{u}\,, \vec{v} \times \vec{w} \rangle|}{\|\vec{v} \times \vec{w}\|}.$$

We can take $\vec{w}$ as base. Let φ be the angle between $\vec{w}$ and $\vec{v}$. Then we obtain that $h = \|\vec{v}\| \sin\varphi$ and therefore, with the help of Theorem 2.32

$$A = \|\vec{w}\| h = \|\vec{w}\| \|\vec{v}\| \sin\varphi = \|\vec{v} \times \vec{w}\|.$$

Therefore, the volume of the parallelepiped is

$$V = Ah = \|\vec{v} \times \vec{w}\| \frac{|\langle \vec{u}\,, \vec{v} \times \vec{w} \rangle|}{\|\vec{v} \times \vec{w}\|} = |\langle \vec{u}\,, \vec{v} \times \vec{w} \rangle|. \qquad \square$$

Corollary 2.35. *Let $\vec{u}, \vec{v}, \vec{w} \in \mathbb{R}^3$. Then*

$$|\langle \vec{u}\,, \vec{v} \times \vec{w} \rangle| = |\langle \vec{v}\,, \vec{w} \times \vec{u} \rangle| = |\langle \vec{w}\,, \vec{u} \times \vec{v} \rangle|.$$

Proof. The formula holds because each of the expressions describes the volume of the parallelepiped spanned by the three given vectors since we can take any of the sides of the parallelogram as its base. ☐

2.5 Lines and planes in $\mathbb{R}^3$

Lines

In order to know a line in $\mathbb{R}^3$ completely, it is not necessary to know all its points. It is sufficient to know either

(a) two different points P, Q on the line

or

(b) one point P on the line and the direction of the line.

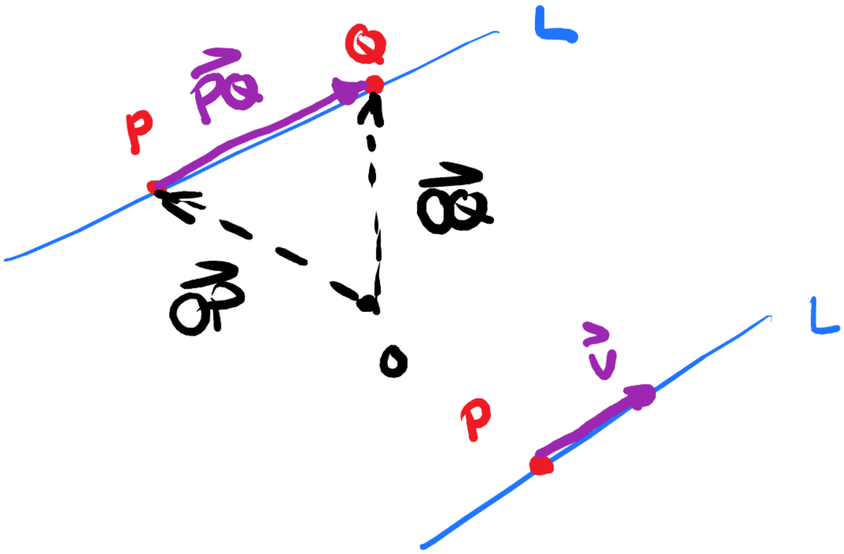

FIGURE 2.13: Line L given (a) by two points P, Q on L, (b) by a point P on L and the direction of L.

Clearly, both descriptions are equivalent. If we have two different points P, Q on the line L, then its direction is given by the vector $\overrightarrow{PQ}$. If on the other hand we are given a point P on L and a vector $\vec{v}$ which is parallel to L, then we easily get another point Q on L by $\overrightarrow{OQ} = \overrightarrow{0P} + \vec{v}$.

Now we want to give formulas for the line.

Vector equation

Given two points $P(p_1, p_2, p_3)$ and $Q(q_1, q_2, q_3)$ with $P \neq Q$, there is exactly one line L which passes through both points. In formulas, this line is described as

$$L = \left\{ \overrightarrow{0P} + t\overrightarrow{PQ} : t \in \mathbb{R} \right\} = \left\{ \begin{pmatrix} p_1 + (q_1 - p_1)t \\ p_2 + (q_2 - p_2)t \\ p_3 + (q_3 - p_3)t \end{pmatrix} : t \in \mathbb{R} \right\} \tag{2.7}$$

If we are given a point $P(p_1, p_2, p_3)$ on L and a vector $\vec{v} = \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix} \neq \vec{0}$ parallel to L, then

$$L = \left\{ \overrightarrow{0P} + t\vec{v} : t \in \mathbb{R} \right\} = \left\{ \begin{pmatrix} p_1 + v_1 t \\ p_2 + v_2 t \\ p_3 + v_3 t \end{pmatrix} : t \in \mathbb{R} \right\} \tag{2.8}$$

The formulas (2.7) and (2.8) are called *vector equation* for the line L. Note that they are the same if we set $v_1 = q_1 - p_1$, $v_2 = q_2 - p_2$, $v_3 = q_3 - p_3$. We will mostly use the notation with the v's since it is shorter. The vector $\vec{v}$ is called *directional vector* of the line L. Observe that if $\vec{v}$ is a directional vector for L, then $c\vec{v}$ is so too for every $c \in \mathbb{R} \setminus \{0\}$.

Parametric equation

From the formula (2.8) it is clear that a point (x, y, z) belongs to L if and only if there exists $t \in \mathbb{R}$ such that

$$\begin{aligned} x &= p_1 + tv_1, \\ y &= p_2 + tv_2, \\ z &= p_3 + tv_3. \end{aligned} \tag{2.9}$$

If we had started with (2.7), then had obtained

$$\begin{aligned} x &= p_1 + t(q_1 - p_1), \\ y &= p_2 + t(q_2 - p_2), \\ z &= p_3 + t(q_3 - p_3) \end{aligned} \tag{2.10}$$

The system of equations (2.9) or (2.10) are called the *parametric equations* of L. Here, t is the parameter.

Symmetric equation

Observe that for $(x, y, z) \in L$, the three equations in (2.9) must hold for the same t. So if we assume that $v_1, v_2, v_3 \neq 0$, then we can solve for t and we obtain that

$$\frac{x - p_1}{v_1} = \frac{y - p_2}{v_2} = \frac{z - p_3}{v_3} \tag{2.11}$$

If we use (2.10) then we obtain

$$\frac{x-p_1}{q_1-p_1} = \frac{y-p_2}{q_2-p_2} = \frac{z-p_3}{q_3-p_3}. \tag{2.12}$$

The system of equations (2.11) or (2.12) is called the *symmetric equation* of L.
If for instance, $v_1 = 0$ and $v_2, v_3 \neq 0$, then the symmetric equation would be

$$x = p_1, \quad \frac{y-p_2}{v_2} = \frac{z-p_3}{v_3}.$$

This is a line which is parallel to the yz-plane.
If $v_1 = v_2 = 0$ and $v_3 \neq 0$, then the symmetric equation would be

$$x = p_1, \quad y = p_2, z \in \mathbb{R}.$$

This is a line which is parallel to the z-axis.

Remark 2.36. It is important to observe that a given line has many different parametrizations. For example, the vector equation that we write down depends on the points we choose on L. Clearly, we have infinitely many possibilities to do so.

Example 2.37. The following equations describe the same line:

$$L = \left\{ \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} + t \begin{pmatrix} 4 \\ 5 \\ 6 \end{pmatrix} : t \in \mathbb{R} \right\} = \left\{ \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} + t \begin{pmatrix} 8 \\ 10 \\ 12 \end{pmatrix} : t \in \mathbb{R} \right\} = \left\{ \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} + t \begin{pmatrix} -4 \\ -5 \\ -6 \end{pmatrix} : t \in \mathbb{R} \right\}$$

$$= \left\{ \begin{pmatrix} 5 \\ 7 \\ 9 \end{pmatrix} + t \begin{pmatrix} 4 \\ 5 \\ 6 \end{pmatrix} : t \in \mathbb{R} \right\}$$

Two lines G and L in $'R^3$ are parallel if and only if their directional vectors are parallel.

Planes

In order to know a plane in $\mathbb{R}^3$ completely, it is sufficient to

(a) three points P, Q on the plane that do not lie on a line,

or

(b) one point P on the plane and two non-parallel vectors $\vec{v}, \vec{w}$ which are both parallel the plane,

or

(c) one point P on the plane and a vector $\vec{n}$ which is perpendicular to the plane,

FIGURE 2.14: Plane π given (a) by three points P, Q, R on π, (b) by a point P on L and two vectors $\vec{v}, \vec{w}$ parallel to π. (c) by a point P on L and a vector $\vec{n}$ perpendicular to π.

First, let us see how we can pass from one description to another. Clearly, the descriptions ((a)) and ((b)) are equivalent because given three points P, Q, R on π which do not lie on a line, we can form the vectors $\overrightarrow{PQ}$ and $\overrightarrow{PR}$. Theses vectors are then parallel to the plane π but are not parallel with each other. (Of course, we also could have taken $\overrightarrow{QR}$ and $\overrightarrow{QP}$ or $\overrightarrow{RP}$ and $\overrightarrow{RQ}$.) If, on the other hand, we have one point P on π and two vectors $\vec{v}$ and $\vec{w}$, parallel to π and $\vec{v} \nparallel \vec{w}$, then we can easily get two other points on π, for instance by $\overrightarrow{0Q} = \overrightarrow{0P} + \vec{v}$ and $\overrightarrow{0R} = \overrightarrow{0P} + \vec{w}$. Then the three points P, Q, R lie on π and do not lie on a plane.
In formulas, we can now describe our plane π as

$$\pi = \left\{ (x, y, z) : \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \overrightarrow{0P} + s\vec{v} + t\vec{w} \quad \text{for some } s, t \in \mathbb{R} \right\}$$

Now we want to use the normal vector of the plane to describe it. Assume that we are given a point P on π and a normal vector $\vec{n}$ perpendicular to the plane. This means that every vector which is parallel to the plane π must be perpendicular to $\vec{n}$. If we take an arbitrary point $Q(x, y, z) \in \mathbb{R}^3$, then $Q \in \pi$ if and only if $\overrightarrow{PQ}$ is parallel to π, that means that $\overrightarrow{PQ}$ is orthogonal to $\vec{n}$. Recall that two vectors are perpendicular if and only if their inner product is 0, so $Q \in \pi$ if and only if

$$\begin{aligned} 0 = \langle n \,, \overrightarrow{PQ} \rangle &= \left\langle \begin{pmatrix} n_1 \\ n_2 \\ n_3 \end{pmatrix} , \begin{pmatrix} x - p_1 \\ y - p_2 \\ z - p_3 \end{pmatrix} \right\rangle = n_1(x - p_1) + n_2(y - p_2) + n_3(z - p_3) \\ &= n_1 x + n_2 y + n_3 z - (n_1 p_1 + n_2 p_2 + n_3 - p_3) \end{aligned}$$

If we set $d = n_1 p_1 + n_2 p_2 + n_3 - p_3$, then it follows that a point $Q(x, y, z)$ belongs to π if and only if its coordinates satisfy

$$n_1 x + n_2 y + n_3 z = d. \tag{2.13}$$

Equation (2.13) is called the *normal equation* for the plane π.

Remark 2.38. As before, note that the normal equation for a plane is not unique. For instance,

$$x + 2y + 3z = 5 \qquad \text{and} \qquad 2x + 4y + 6z = 10$$

describe the same plane. The reason is that "the" normal vector of a plane is not unique. Given one normal vector $\vec{n}$, than every $c\vec{n}$ with $c \in \mathbb{R} \setminus \{0\}$ is also a normal vector to the plane.

Definition 2.39. The *angle between two planes* is the angle between their normal vectors.

Note that this definition is consistent with the fact that two planes are parallel if and only if their normal vectors are parallel.

Remark 2.40.

- Assume a plane is given as in ((b)) (that is, we know a point P on π and two vectors $\vec{v}$ and $\vec{w}$ parallel to π but with $\vec{v} \nparallel \vec{w}$). In order to have description as in ((c)) (that is one point on $\imath$ and a normal vector), we only have to find a vector $\vec{n}$ that is perpendicular to both $\vec{v}$ and $\vec{w}$. Proposition 2.31(vii) tells us how to do this: we only need to calculate $\vec{v} \times \vec{w}$.

- Assume a plane is given as in ((c)) (that is, we know a point P on π and its normal vector). In order to find vectors $\vec{v}$ and $\vec{w}$ as in ((b)), we can guess either find two solutions of $\vec{x} \times \vec{n} = 0$ which are not parallel. Or we find only one solution $\vec{v}$ which usually is easy to guess and then calculate $\vec{w} = \vec{v} \times \vec{n}$. This vector is perpendicular to $\vec{n}$ and therefore it is parallel to the plane. It is also perpendicular to $\vec{v}$ and therefore it is not parallel to $\vec{v}$. In total, this vector $\vec{w}$ does what we need.

2.6 Intersections of lines and planes in $\mathbb{R}^3$

Intersection of lines

Given two lines G and L in $\mathbb{R}^3$, there are three possibilities:

(a) The lines intersect in exactly one point. In this case, they cannot be parallel.

(b) The lines intersect in infinitely many points. In this case, the lines have to be equal. In particular the have to be parallel.

(c) The lines do not intersect. Not that in contrast to the case in $\mathbb{R}^2$, the lines do not have to be parallel for this to happen. For example, the line $L : x = y = 1$ is a line parallel to the z-axis passing through $(1, 1, 0)$, and $G : x = z = 0$ is a line parallel to the y-axis passing through $(0, 0, 0)$, The lines do not intersect and they are not parallel.

Example 2.41. We consider four lines $L_j = \{\vec{p}_j + t\vec{v}_j : t \in \mathbb{R}\}$ with

$$\text{(i)}\quad \vec{v}_1 = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}, \quad \vec{p}_1 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}, \qquad \text{(ii)}\quad \vec{v}_2 = \begin{pmatrix} 2 \\ 4 \\ 6 \end{pmatrix}, \vec{p}_2 = \begin{pmatrix} 2 \\ 4 \\ 7 \end{pmatrix},$$

$$\text{(iii)}\quad \vec{v}_3 = \begin{pmatrix} 1 \\ 1 \\ 2 \end{pmatrix}, \vec{p}_3 = \begin{pmatrix} -1 \\ 0 \\ 0 \end{pmatrix}, \qquad \text{(iv)}\quad \vec{v}_4 = \begin{pmatrix} 1 \\ 1 \\ 2 \end{pmatrix}, \vec{p}_4 = \begin{pmatrix} 3 \\ 0 \\ 5 \end{pmatrix}.$$

We will calculate their mutual intersections.

$\boxed{L_1 \cap L_2 = L_1}$

Proof. A point $Q(x, y, z)$ belongs to $L_1 \cap L_2$ if and only if it belongs both to L_1 and L_2. This means that there must exist an $s \in \mathbb{R}$ such that $\overrightarrow{0Q} = \vec{p}_1 + s\vec{v}_1$ and there must exist a $t \in \mathbb{R}$ such that $\overrightarrow{0Q} = \vec{p}_2 + t\vec{v}_2$. Note the s and t are different parameters. So we are looking for s and t such that

$$\vec{p}_1 + s\vec{v}_1 = \vec{p}_2 + t\vec{v}_2, \qquad \text{that is} \qquad \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} + s \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} = \begin{pmatrix} 2 \\ 4 \\ 7 \end{pmatrix} + t \begin{pmatrix} 2 \\ 4 \\ 6 \end{pmatrix} \tag{2.14}$$

Once we have solved this for s and t, we insert the into the equation for L_1 and L_2 respectively, and obtain Q. Note that (2.14) in reality is a system of three equations: one equation for each component of the vector equation. Writing it out, and solving each equation for s, we obtain

$$\begin{array}{rcl} 0 + s = 2 + 2t & & s = 2 + 2t \\ 0 + 2s = 4 + 4t & \iff & s = 2 + 2t \\ 1 + 3s = 7 + 6t & & s = 2 + 2t. \end{array}$$

This means that we have infinitely many solutions: Given any point R on L_1, there is a corresponding $s \in \mathbb{R}$ such that $\overrightarrow{0R} = \vec{p}_1 + s\vec{v}_1$. Now if we choose $t = (s-2)/2$, then $\overrightarrow{0R} = \vec{p}_2 + t\vec{v}_2$ holds, hence $R \in L_2$ too. If on the other hand we have a point $R' \in L_2$, then there is a corresponding $t \in \mathbb{R}$ such that $\overrightarrow{0R'} = \vec{p}_2 + t\vec{v}_2$. Now if we choose $s = 2 + 2t$, then $\overrightarrow{0R'} = \vec{p}_1 + t\vec{v}_1$ holds, hence $R' \in L_2$ too. In summary, we showed that $L_1 = L_2$. ☐

Remark 2.42. We could also have seen that the directional vectors of L_1 and L_2 are parallel. In fact, $\vec{v}_2 = 2\vec{v}_1$. It then suffices to show that L_1 and L_2 have at least one point in common in order to conclude that the lines are equal.

$\boxed{L_1 \cap L_3 = \{(1, 2, 4)\}}$

Proof. As before, we need to find $s, t \in \mathbb{R}$ such that

$$\vec{p}_1 + s\vec{v}_1 = \vec{p}_3 + t\vec{v}_3, \qquad \text{that is} \qquad \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} + s \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} = \begin{pmatrix} -1 \\ 0 \\ 0 \end{pmatrix} + t \begin{pmatrix} 1 \\ 1 \\ 2 \end{pmatrix}. \tag{2.15}$$

We write this as a system of equations, we get

$$\begin{array}{lrcl} \text{①} & 0 + s = -1 + t & & \text{①} \quad s - t = -1 \\ \text{②} & 0 + 2s = 0 + t & \iff & \text{②} \quad 2s - t = 0 \\ \text{③} & 1 + 3s = 0 + 2t & & \text{③} \quad 3s - 2t = -1 \end{array}$$

From ① it follows that $s = t - 1$. Inserting in ② gives $0 = 2(t-1) - t = t - 2$, hence $t = 2$. From ① we then obtain that $s = 2 - 1 = 1$. Observe that so far we used only equations ① and ②. In order to see if we really found a solution, we must check if it is consistent with ③. Inserting our candidates for s and t, we find that $3 \cdot 1 - 2 \cdot 2 = -1$ which is consistent with ③.
So we have exactly one point of intersection. In order to find it, we put $s = 1$ in the equation for L_1:

$$\overrightarrow{0Q} = \vec{p}_1 + 1 \cdot \vec{v}_1 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} + \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \\ 4 \end{pmatrix},$$

hence the intersection point is $Q(1, 2, 4)$.
In order to check if this result is correct, we can put $t = 2$ in the equation for L_3. The result must be the same. The corresponding calculation is:

$$\overrightarrow{0Q} = \vec{p}_3 + 2 \cdot \vec{v}_3 = \begin{pmatrix} -1 \\ 0 \\ 0 \end{pmatrix} + \begin{pmatrix} 2 \\ 2 \\ 4 \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \\ 4 \end{pmatrix},$$

which confirms that the intersection point is $Q(1, 2, 4)$. ☐

$\boxed{L_1 \cap L_4 = \varnothing}$

Proof. As before, we need to find $s, t \in \mathbb{R}$ such that

$$\vec{p}_1 + s\vec{v}_1 = \vec{p}_4 + t\vec{v}_4, \qquad \text{that is} \qquad \begin{pmatrix}0\\0\\1\end{pmatrix} + s\begin{pmatrix}1\\2\\3\end{pmatrix} = \begin{pmatrix}3\\0\\5\end{pmatrix} + t\begin{pmatrix}1\\1\\2\end{pmatrix}. \tag{2.16}$$

We write this as a system of equations, we get

$$\begin{array}{rrcl} ① & s &=& 3 + \ t \\ ② & 2s &=& \quad\ \ t \\ ③ & 1 + 3s &=& 5 + 2t \end{array} \qquad \Longleftrightarrow \qquad \begin{array}{rrcl} ① & s - \ t &=& 3 \\ ② & 2s - \ t &=& 0 \\ ③ & 3s - 2t &=& 5 \end{array}$$

From ① it follows that $s = t + 3$. Inserting in ② gives $0 = 2(t + 3) - t = t + 6$, hence $t = -6$. From ① we then obtain that $s = -6 + 3 = -3$. Observe that so far we used only equations ① and ②. In order to see if we really found a solution, we must check if it is consistent with ③. Inserting our candidates for s and t, we find that $3 \cdot (-3) - 2 \cdot (-6) = 3$ which is inconsistent with ③. Therefore we conclude that there is no pair of real numbers s, t which satisfies all three equations ①–③ simultaneously, so the two lines do not intersect. ☐

Exercise. Show that $L_3 \cap L_4 = \varnothing$.

Intersection of planes

Given two planes π_1 and π_2 in $\mathbb{R}^3$, there are two possibilities:

(a) The planes intersect. In this case, they necessarily intersect in infinitely many points. The intersection is either a line. In this case π_1 and π_2 are not parallel. Or the intersection is a plane. In this case $\pi_1 = \pi_2$.

(b) The planes do not intersect. In this case, the planes must be parallel and not equal.

Example 2.43. We consider the following four planes:

$$\pi_1 : x + y + 2z = 3, \quad \pi_2 : 2x + 2y + 4z = 3, \quad \pi_3 : 2x + 2y + 4z = 6, \quad \pi_4 : x + y - 2z = 5.$$

We will calculate their mutual intersections.

$\boxed{\pi_1 \cap \pi_2 = \varnothing}$

Proof. The set of all points $Q(x, y, z)$ which belong both to π_1 and π_2 is the set of all x, y, z which simultaneously satisfy

$$\begin{array}{rl} ① & x + \ y + 2z = 3, \\ ② & 2x + 2y + 4z = 3. \end{array}$$

Now clearly, if x, y, z satisfies ①, then it cannot satisfy ② (the right side would be 6). We can see this more formally if we solve ①, e.g., for x and then insert into ②. We obtain from ①: $x = 3 - y - 2z$. Inserting into ② leads to

$$3 = 2(3 - y - 2z) + 2y + 4z = 6,$$

which is absurd.

Geometrically, this was to be expected. The normal vectors of the planes are $\vec{n}_1 = \begin{pmatrix} 1 \\ 1 \\ 2 \end{pmatrix}$ and $\vec{n}_2 = \begin{pmatrix} 2 \\ 2 \\ 4 \end{pmatrix}$ respectively. Since they are parallel, the planes are parallel and therefore they either are equal or they have empty intersection. Now we see that for instance $(3,0,0) \in \pi_1$ but $(3,0,0) \notin \pi_2$, so the planes cannot be equal. Therefore they have empty intersection. ☐

$\boxed{\pi_1 \cap \pi_3 = \pi_1}$

Proof. The set of all points $Q(x,y,z)$ which belong both to π_1 and π_3 is the set of all x,y,z which simultaneously satisfy

$$\begin{array}{rl} ① & x + y + 2z = 3, \\ ② & 2x + 2y + 4z = 6. \end{array}$$

Clearly, both equations are equivalent: if x,y,z satisfies ①, then it also satisfies ② and vice versa. Therefore, $\pi_1 = \pi_3$. ☐

$$\boxed{\pi_1 \cap \pi_4 = \left\{ \begin{pmatrix} 4 \\ 0 \\ -\frac{1}{2} \end{pmatrix} + t \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix} : t \in \mathbb{R} \right\}.}$$

Proof. First, we notice that the normal vectors $\vec{n}_1 = \begin{pmatrix} 1 \\ 1 \\ 2 \end{pmatrix}$ and $\vec{n}_4 = \begin{pmatrix} 1 \\ 1 \\ -2 \end{pmatrix}$ are not parallel, so we expect that the solution is a line in $\mathbb{R}^3$.
The set of all points $Q(x,y,z)$ which belong both to π_1 and π_4 is the set of all x,y,z which simultaneously satisfy

$$\begin{array}{rl} ① & x + y + 2z = 3, \\ ② & x + y - 2z = 5. \end{array}$$

Equation ① shows that $x = 3 - y - 2z$. Inserting into ② leads to $5 = 3 - y - 2z + y - 2z = 3 - 4z$, hence $z = -\frac{1}{2}$. Putting this into ①, we find that $x + y = 3 - 2z = 4$. So in summary, the intersection consists of all points (x,y,z) which satisfy

$$z = -\frac{1}{2}, \quad x = 4 - y \quad \text{with} \quad y \in \mathbb{R} \quad \text{arbitrary},$$

in other words,

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 4 - y \\ y \\ -\frac{1}{2} \end{pmatrix} = \begin{pmatrix} 4 \\ 0 \\ -\frac{1}{2} \end{pmatrix} + \begin{pmatrix} -y \\ y \\ 0 \end{pmatrix} = \begin{pmatrix} 4 \\ 0 \\ -\frac{1}{2} \end{pmatrix} + y \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix} \quad \text{with} \quad y \in \mathbb{R} \quad \text{arbitrary}.$$ ☐

Intersection of several lines and planes

If we wanted to intersect for instance, 5 planes in $\mathbb{R}^3$, then we would have to solve a system of 5 equations for 3 unknowns. Or if we wanted to intersect 7 lines in $\mathbb{R}^3$, then we had to solve a system of 3 equations for 7 unknowns. If we do it like here, this could become quite messy. So the next chapter is devoted to find a systematic way how to solve a system of m linear equations for n unknowns.

2.7 Summary

$$\begin{aligned} x - 2y - 4z &= 1 \\ 3x - y - z &= -1 \\ x - 11y + 22z &= 110 \end{aligned}$$

Faltan Figures 11, 12.

Chapter 3

Linear Systems and Matrices

Chapter 4

Vector spaces and linear maps

In the following, $\mathbb{K}$ always denotes a field. In this chapter, you may always think of $\mathbb{K} = \mathbb{R}$, though almost everything is true also for other fields, like $\mathbb{C}$, $\mathbb{Q}$ or $\mathbb{F}_p$ where p is a prime number. Later, in Chapter **??** it will be more useful to work with $\mathbb{K} = \mathbb{C}$.

In this rather we will first work with abstract vector spaces. We will first discuss their basic properties. Then, in Section 4.2 we will talk about subspaces. These are subsets of vector space which are themselves vector spaces. In Section 4.3 we will introduce basis and dimension of a vector space. These concepts are fundamental in linear algebra since they allow to classify all finite dimensional vector spaces. In a certain sense, all n dimensional vector spaces over the same field $\mathbb{K}$ are the same. In Chapter **??** we will study linear maps between vector spaces.

4.1 Definitions and basic properties

First we recall the definition of an abstract vector space from Chapter 2.

Definition 4.1. Let V be a set together with two operations

$$\begin{aligned} \text{vector sum} \quad & + : V \times V \to V, \quad (v, w) \mapsto v + w, \\ \text{product of a scalar and a vector} \quad & \cdot : \mathbb{K} \times V \to V, (\lambda, v) \mapsto \lambda \cdot v. \end{aligned}$$

Note that we will usually write λv instead of $\lambda \cdot v$. Then V (or more precisely, $(V, +, \cdot)$) is called a *vector space* if for all $u, v, w \in V$ and all $\lambda, \mu \in \mathbb{K}$ the following holds:

(a) **Associativity:** $(u + v) + w = u + (v + w)$.

(b) **Commutativity:** $v + w = w + v$.

(c) **Identity element of addition:** There exists an element $\mathbb{0} \in V$, called the *additive identity* such that for every $v \in \mathbb{R}^2$, we have $\mathbb{0} + v = v + \mathbb{0} = v$.

(d) **Inverse element:** For all $v \in V$, we have an inverse element v' such that $v + v' = \mathbb{0}$.

(e) **Identity element of multiplication by scalar:** For every $v \in V$, we have that $1v = v$.

(f) **Compatibility:** For every $v \in V$ and $\lambda, \mu \in \mathbb{R}$, we have that $(\lambda\mu)v = \lambda(\mu v)$.

(g) **Distributivity laws:** For all $v, w \in V$ and $\lambda, \mu \in \mathbb{R}$, we have

$$(\lambda + \mu)v = \lambda v + \mu v \quad \text{and} \quad \lambda(v + w) = \lambda v + \lambda w.$$

Remark 4.2. (i) Note that the notation $\vec{v}$ with an arrow is reserved for the special case of a vector in $\mathbb{R}^n$ or $\mathbb{C}^n$. Vectors in an abstract vector space are usually denoted without an arrow.

(ii) If $\mathbb{K} = \mathbb{R}$, then V is called a *real vector space*. If $\mathbb{K} = \mathbb{C}$, then V is called a *complex vector space*.

Before we give examples of vector spaces, we first show some basic properties of vector spaces.

Properties 4.3. *(i) The identity element is unique.* (Note that in the vector space axioms we only asked for *existence* of an additive identity element; we did not ask for uniqueness. So one could think that there may be several elements which satisfy (c) in Definition 4.1. However, this is not possible as the following proof shows.)

Proof. Assume there are two neutral elements $\mathbb{0}$ and $\mathbb{0}'$. Then we know that for every v and w in V the following is true:

$$v = \mathbb{0} + v, \qquad w = \mathbb{0}' + w.$$

Now let us take $v = \mathbb{0}'$ and $w = \mathbb{0}$. Then, using commutativity, we obtain

$$\mathbb{0}' = \mathbb{0} + \mathbb{0}' = \mathbb{0}' + \mathbb{0} = \mathbb{0}.$$ □

(ii) For every $v \in V$, its inverse element is unique. (Note that in the vector space axioms we only asked for *existence* of an additive inverse for every element $x \in V$; we did not ask for uniqueness. So one could think that there may be several elements which satisfy (d) in Definition 4.1. However, this is not possible as the following proof shows.)

Proof. Let $v \in V$ and assume that there are elements v', v'' in V such that

$$v + v' = \mathbb{0}, \quad v + v'' = \mathbb{0}.$$

We have to show that $v' = v''$. This follows from

$$v' = v' + \mathbb{0} = v' + (v + v'') = (v' + v) + v'' = \mathbb{0} + v'' = v''.$$ □

(iii) For every $\lambda \in \mathbb{K}$ we have $\lambda\mathbb{0} = \mathbb{0}$.

Proof. Observe that

$$\lambda\mathbb{0} = \lambda(\mathbb{0} + \mathbb{0}) = \lambda\mathbb{0} + \lambda\mathbb{0}.$$

Now let $(\lambda\mathbb{0})'$ be the inverse of $\lambda\mathbb{0}$ and sum it to both sides of the equation. We obtain

$$\begin{aligned} & \lambda\mathbb{0} + (\lambda\mathbb{0})' = (\lambda\mathbb{0} + \lambda\mathbb{0}) + (\lambda\mathbb{0})' \\ \implies \quad & \mathbb{0} = \lambda\mathbb{0} + (\lambda\mathbb{0} + (\lambda\mathbb{0})') \\ \implies \quad & \mathbb{0} = \lambda\mathbb{0} + \mathbb{0} \\ \implies \quad & \mathbb{0} = \lambda\mathbb{0}. \end{aligned}$$ □

(iv) For every $v \in V$ we have that $0v = \mathbb{0}$.

Proof. The proof is similar to the one above. Observe that

$$0v = (0+0)v = 0v + 0v.$$

Now let $(0v)'$ be the inverse of $0v$ and sum it to both sides of the equation. We obtain

$$\begin{aligned} & 0v + (0v)' = (0v + 0v) + (0v)' \\ \Longrightarrow \qquad & \mathbb{0} = 0v + (0v + (0v)') \\ \Longrightarrow \qquad & \mathbb{0} = 0v + \mathbb{0} \\ \Longrightarrow \qquad & \mathbb{0} = 0v. \end{aligned}$$

□

(v) If $\lambda v = \mathbb{0}$, then either $\lambda = 0$ or $v = \mathbb{0}$.

Proof. If $\lambda = 0$, then there is nothing to prove. Now assume that $\lambda \neq 0$. Then v is $\mathbb{0}$ because

$$v = \frac{1}{\lambda}(\lambda v) = \frac{1}{\lambda}\mathbb{0} = \mathbb{0}.$$

□

(vi) For every $v \in V$, its inverse is $(-1)v$.

Proof. Let $v \in V$. Observe that by (v), we have that $0v = \mathbb{0}$. Therefore

$$\mathbb{0} = 0v = (1 + (-1))x = v + (-1)v.$$

Hence $(-1)v$ is an additive inverse of v. By (ii), the inverse of v is unique, therefore $(-1)v$ is the inverse of v. □

Remark 4.4. From now on, we write $-v$ for the additive inverse of a vector. This notation is justified by Property 4.3 (vi).

Examples 4.5. We give some important examples of vector spaces.

- $\mathbb{R}$ is a real vector space. More generally, $\mathbb{R}^n$ is a real vector space. The proof is the same as for $\mathbb{R}^2$ in Chapter 2. Associativity and commutativity are clear. The identity element is the vector whose entries are all equal to zero: $\vec{0} = (0, \dots, 0)^t$. The inverse for a given vector $\vec{x} = (x_1, \dots, x_n)^t$ is $(-x_1, \dots, -x_n)^t$. The distributivity laws are clear, as is the fact that $1\vec{x} = \vec{x}$ for every $\vec{x} \in \mathbb{R}^n$.
- $\mathbb{C}$ is a complex vector space. More generally, $\mathbb{C}^n$ is a real complex space. The proof is as for $\mathbb{R}^n$.
- $\mathbb{C}$ can also be seen as a real vector space.
  Exercise. Check that $\mathbb{C}$ is a real vector space!
- $\mathbb{R}$ is **not** a complex vector space. If it was, then the vectors would be real numbers and the scalars would be complex numbers. But then if we take $1 \in \mathbb{R}$ and $\mathrm{i} \in \mathbb{C}$, then the product $\mathrm{i}1$ must be a vector, that is, a real number, which is not the case.

- $\mathbb{R}$ can be seen as a $\mathbb{Q}$-vector space.
- For every $n, m \in \mathbb{N}$, the space $M(m \times n)$ of all $m \times n$ matrices with real coefficients is a real vector space.

  Proof. Note that in this case, the vectors are matrices. Associativity and commutativity are easy to check. The identity element is the matrix whose entries are all equal to zero. Given a matrix $A = (a_{ij})_{\substack{i=1,\dots,m,\\ j=1,\dots,n}}$ its (additive) inverse is the matrix $-A = (-a_{ij})_{\substack{i=1,\dots,m.\\ j=1,\dots,n}}$ The distributivity laws are clear, as is the fact that $1A = A$ for every $A \in M(m \times n)$. ☐

- For every $n, m \in \mathbb{N}$, the space $M(m \times n, \mathbb{C})$ of all $m \times n$ matrices with complex coefficients, is a complex vector space.

  Proof. As in the case of real matrices. ☐

- Let $C(\mathbb{R})$ be the set of all continuous functions from $\mathbb{R}$ to $\mathbb{R}$. We define the sum of two functions f and g in the usual way as the new function

  $$f + g : \mathbb{R} \to \mathbb{R}, \quad (f+g)(x) = f(x) + g(x).$$

  The product of a function f with a real number λ gives the new function λf defined by

  $$\lambda f : \mathbb{R} \to \mathbb{R}, \quad (\lambda f)(x) = \lambda f(x).$$

  Then $C(\mathbb{R})$ is a vector space with these new operations.

  Proof. It is clear that these operations satisfy associativity, commutativity and distributivity and that $1f = f$ for every function $f \in C(\mathbb{R})$. The additive identity is the zero function (the function which is constant to zero). For a given function f, its (additive) inverse is the function $-f$. ☐

Observe that the sets $M(m \times n)$ and $C(\mathbb{R})$ admit more operations, for example we can multiply functions, or we can multiply matrices or we can calculate $\det A$ for a square matrix. However, all these operations have nothing to do with the question whether they are vector spaces or not. It is important to note that for a vector space we only need the sum of two vectors and the product of a scalar with vector.

We give more examples.

Examples 4.6.

- Consider $\mathbb{R}^2$ but we change the usual sum to the new sum $\oplus$ defined by

  $$\begin{pmatrix} x \\ y \end{pmatrix} \oplus \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} x+a \\ 0 \end{pmatrix}.$$

  With this new sum, $\mathbb{R}^2$ is **not** a vector space. The reason is that there is no additive identity. To see this, assume that we had an additive identity, say $\begin{pmatrix} \alpha \\ \beta \end{pmatrix}$. Then we must have

  $$\begin{pmatrix} \alpha \\ \beta \end{pmatrix} + \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x \\ y \end{pmatrix} \qquad \text{for all} \quad \begin{pmatrix} x \\ y \end{pmatrix} \in \mathbb{R}^2.$$

However, for example,

$$\begin{pmatrix}\alpha\\ \beta\end{pmatrix}+\begin{pmatrix}0\\ 1\end{pmatrix}=\begin{pmatrix}\alpha\\ 0\end{pmatrix}\neq\begin{pmatrix}0\\ 1\end{pmatrix},$$

- Consider $\mathbb{R}^2$ but we change the usual sum to the new sum $\oplus$ defined by

  $$\begin{pmatrix}x\\ y\end{pmatrix}\oplus\begin{pmatrix}a\\ b\end{pmatrix}=\begin{pmatrix}x+b\\ y+b\end{pmatrix}.$$

  With this new sum, $\mathbb{R}^2$ is **not** a vector space. One of the reasons is that the sum is not commutative. For example

  $$\begin{pmatrix}1\\ 0\end{pmatrix}+\begin{pmatrix}0\\ 1\end{pmatrix}=\begin{pmatrix}1+1\\ 0+0\end{pmatrix}=\begin{pmatrix}2\\ 0\end{pmatrix},\qquad\text{but}\qquad\begin{pmatrix}0\\ 1\end{pmatrix}+\begin{pmatrix}1\\ 0\end{pmatrix}=\begin{pmatrix}0+0\\ 1+1\end{pmatrix}=\begin{pmatrix}0\\ 2\end{pmatrix}.$$

  (One could also show that there is no additive identity $\mathbb{0}$ which satisfies $\vec{x}\oplus\mathbb{0}=\vec{x}$ for all $\vec{x}\in\mathbb{R}^2$. You should try to show this.)

- Let $V=\mathbb{R}_+=(0,\infty)$. We make V a real vector space with the following operations: Let $x,y\in V$ and $\lambda\in\mathbb{R}$. We define

  $$x\oplus y=xy\qquad\text{and}\qquad\lambda\odot x=x^\lambda.$$

  Then $(V,\oplus,\odot)$ is a real vector space.

  Proof. Let $u,v,w\in V$ and let $\lambda\in\mathbb{R}$. Then:

  (a) **Associativity:** $(u\oplus v)\oplus w=(uv)\oplus w=(uv)w=u(vw)=u(v\oplus w)=u\oplus(v\oplus w)$.

  (b) **Commutativity:** $v\oplus w=vw=wv=w\oplus v$.

  (c) The **additive identity** of $\oplus$ is 1 because for every $x\in V$ we have that $1\oplus x=1x=x$.

  (d) **Inverse element:** For every $x\in V$, its inverse element is x^{-1} because $x\oplus x^{-1}=xx^{-1}=1$ which is the identity element. (Note that this is in accordance with Properties 4.3 (v) since $(-1)\odot x=x^{-1}$.)

  (e) **Identity element of multiplication by scalar:** For every $x\in V$, we have that $1\odot x=1x=x$.

  (f) **Compatibility:** For every $x\in V$ and $\lambda,\mu\in\mathbb{R}$, we have that

  $$(\lambda\mu)\odot v=v^{\lambda\mu}=(v^\lambda)^\mu=\mu\odot(v^\lambda)=\lambda\odot(\mu\odot v).$$

  (g) **Distributivity laws:** For all $x,y\in V$ and $\lambda,\mu\in\mathbb{R}$, we have

  $$(\lambda+\mu)\odot x=x^{\lambda+\mu}=x^\lambda x^\mu=(\lambda\odot v)(\mu\odot v)=(\lambda\odot v)\oplus(\mu\odot v)$$

  and

  $$\lambda\odot(v\oplus w)=(v\oplus w)^\lambda=(vw)^\lambda=v^\lambda w^\lambda=v^\lambda\oplus w^\lambda=(\lambda\odot v)\oplus(\lambda\odot w).\qquad\square$$

- The example above can be generalised: Let $f : \mathbb{R} \to (a, b)$ be an injective function. Then the interval (a, b) becomes a real vector space with the following operations if we define the sum of two vectors $x, y \in (a, b)$ by

  $$x \oplus y = f(f^{-1}(x) + f^{-1}(y))$$

  and the product of a scalar $\lambda \in \mathbb{R}$ and a vector $x \in (a, b)$ by

  $$\lambda \odot x = f(\lambda f^{-1}(x)).$$

  Note that in the example above we have $(a, b) = (0, \infty)$ and $f = \exp$ (that is: $f(x) = \mathrm{e}^x$).

4.2 Subspaces

In this section, we work mostly with real vector spaces for definiteness sake. However, all the statements are also true for complex vector spaces. We only have to replace everywhere $\mathbb{R}$ by $\mathbb{C}$ and the word *real* by *complex*.

Now we will investigate when a subset of a given vector space is itself a vector space.

Definition 4.7. Let V be a vector space and let $W \subseteq V$ be a subset of V. Then W is called a *subspace* of V if W itself is a vector space with the sum and product with scalars inherited from V. A subspace W is called a *proper subspace* if $W \neq \varnothing$ and $W \neq V$.

First we remark the following basic facts.

Remark 4.8. Let V be a vector space.

- If W is a subspace of V, then $0 \in W$ since W must contain the additive identity.
- If V is a vector space, W is a subspace of V and U is a subspace of W, then U is a subspace of V.
- V always contains the following subspaces: $\{0\}$ and V itself. However, they are not proper subspaces.

  Exercise 4.9. Prove these statements.

Now assume that we are given a vector space V and in it a subset $W \subseteq V$ and we would like to check if W is a vector space. In principle we would have to check all seven vector space axioms from Definition 4.1. However, if W is a subset of V, then we get some of the vector space axioms for free. More precisely, the axioms (a), (b), (e), (f) and (g) hold automatically. For example, to prove (b), we take two elements $w_1, w_2 \in W$. They belong also to V since $W \subseteq V$, and therefore they commute: $w_1 + w_2 = w_2 + w_1$.
We can even show the following proposition:

Proposition 4.10. *Let V be a real vector space and $W \subseteq V$ a subset. Then W is a subspace of V if and only if the following three properties hold:*

- *(i) $W \neq \emptyset$, that is, W is not empty.*
- *(ii) W is* closed under sums*, that is, if we take w_1 and w_2 in W, then their sum $w_1 + w_2$ belongs to W.*
- *(iii) W is* closed under product with scalars*, that is, if we take $w \in W$ and $\lambda \in \mathbb{R}$, then it must follow that $\lambda w \in W$.*

Note that (ii) and (iii) can be resumed in the following:

- (iv) W is *closed under sums and product with scalars*, that is, if we take $w_1, w_2 \in W$ and $\lambda \in \mathbb{R}$, then $\lambda w_1 + w_2 \in W$.

Proof of 4.10. Assume that W is a subspace, then clearly (ii) and (iii) hold. (i) holds because every vector space must contain at least the additive identity $'veczero$.

Now suppose that W is a subset of V such that the properties (i), (ii) and (iii) are satisfied. In order to show that W is a subspace of V, we need to verify the vector space axioms (a) - (f) from Definition 4.1. By assumptions (ii) and (iii) the sum and product with scalars are well defined in W. Moreover, we already convinced ourselves that (a), (b), (e), (f) and (g) hold. Now, for the existence of an additive identity, we take an arbitrary $w \in W$ (such a w exists because W is not empty by assumption (i)). Hence $\mathbb{0} = 0w \in W$ where $\mathbb{0}$ is the additive identity in V. This then is also the additive identity in W. Finally, given $w \in W \subseteq V$, we know from Propertie 4.3 (v) that its additive inverse is $(-1)w$, which, by our assumption (iii), belongs to W. So we have verified that W satisfies all vector space axioms, so it is a vector space. □

The proposition is also true if V is a complex vector space. We only have to replace $\mathbb{R}$ everywhere by $\mathbb{C}$.

In order to verify that a given $W \subseteq V$ is a subspace, one only has to verify (i), (ii) and (iii) from the preceding proposition. In order to verify that W is not empty, one typically checks if it contains $\mathbb{0}$.

The following definition is very important in many applications.

Definition 4.11. Let V be a vector space and $W \subseteq V$ a subset. The W is called an *affine subspace* if there exists an $v_0 \in V$ such that set

$$v_0 + W := \{v_0 + w : w \in W\}$$

is a subspace of V.

Clearly, every subspace is also an affine subspace (take $v_0 = \mathbb{0}$).
Let us see examples of subspaces and affine subspaces.

Examples 4.12. Let V be a vector space. We assume that V is a real vector space, but everything works also for a complex vector space (we only have to replace $\mathbb{R}$ everywhere by $\mathbb{C}$.)

- (i) $\{0\}$ is a subspace of V. It is called the *trivial subspace* of V.
- (ii) V itself is a subspace of V.

(iii) Fix $z \in V$. Then the set $W := \{\lambda z : \lambda \in \mathbb{R}\}$ is a subspace of V.

(iv) More generally, if we fix $z_1, \dots z_k \in V$, then the set $W := \{\lambda_1 z_1 + \cdots \lambda_k z_k : \lambda_1, \dots, \lambda_k \in \mathbb{R}\}$ is a subspace of V.

(v) If we fix v_0 and $z_1, \dots z_k \in V$, then the set $W := \{v_0 + \lambda_1 z_1 + \cdots \lambda_k z_k : \lambda_1, \dots, \lambda_k \in \mathbb{R}\}$ is an affine subspace of V. In general it will not be a subspace.

Exercise. Show that $W := \{v_0 + \lambda_1 z_1 + \cdots \lambda_k z_k : \lambda_1, \dots, \lambda_k \in \mathbb{R}\}$ is an affine subspace of V. Show that it is a subspace if and only if $v_0 \in \operatorname{span}\{z_1, \dots, z_k\}$.

(vi) If W is a subspace of V, then $V \setminus W$ is not a subspace. This can be seen easily if we recall that W must contain $\mathbb{0}$. But then $V \setminus W$ cannot contain $\mathbb{0}$, hence it cannot be a vector space.

Some more examples:

Examples 4.13.

- The set of all solutions of a homogeneous system of linear equations is a vector space.
- The set of all solutions of an inhomogeneous system of linear equations is an affine vector space.
- The set of all solutions of a homogeneous linear differential equation is a vector space.
- The set of all solutions of an inhomogeneous linear differential equation is an affine vector space.

Examples 4.14 (Examples and non-examples of subspaces of $\mathbb{R}^2$).

- $W = \left\{ \begin{pmatrix} \lambda \\ 0 \end{pmatrix} : \lambda \in \mathbb{R} \right\}$ is a subspace of $\mathbb{R}^2$. This is actually a subspace of the form (iii) from Example 4.12 with $z = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$. Note that geometrically W is a line.
- For fixed $x_0, y_0 \in \mathbb{R}$ let $W = \left\{ \lambda \begin{pmatrix} x_0 \\ y_0 \end{pmatrix} : \lambda \in \mathbb{R} \right\}$. Then W is a subspace of $\mathbb{R}^2$. Geometrically, W is a line in $\mathbb{R}^2$ passing through the origin which is parallel to the vector $\begin{pmatrix} x_0 \\ y_0 \end{pmatrix}$.

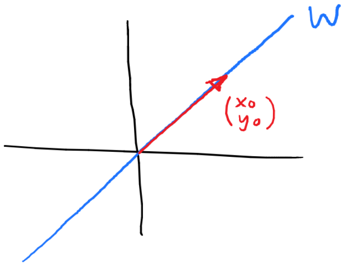

FIGURE 4.1: The subspace W generated by the vector $\binom{x_0}{y_0}$.

- For fixed $a, b, x_0, y_0 \in \mathbb{R}$ let $W = \left\{ \begin{pmatrix} a \\ b \end{pmatrix} + \lambda \begin{pmatrix} x_0 \\ y_0 \end{pmatrix} : \lambda \in \mathbb{R} \right\}$. Then W is an affine subspace. Geometrically, W represents a line in $\mathbb{R}^2$ parallel which passes through the point (a, b) and is parallel to the vector $\begin{pmatrix} x_0 \\ y_0 \end{pmatrix}$. Note that W is a subspace if and only if $\begin{pmatrix} a \\ b \end{pmatrix}$ and $\begin{pmatrix} x_0 \\ y_0 \end{pmatrix}$ are parallel.

FIGURE 4.2: In the figure on the left hand side, W is not a subspace It is only an affine subspace. In the figure on the right hand side, W is a subspace.

- $W = \{\vec{x} \in \mathbb{R}^2 : \vec{x} \geq 3\}$ is not a subspace of $\mathbb{R}^2$ since it does not contain $\vec{0}$.

- $W = \{\vec{x} \in \mathbb{R}^2 : \vec{x} \leq 3\}$ is not a subspace of $\mathbb{R}^2$. For example, take $\vec{z} = \begin{pmatrix} 2 \\ 0 \end{pmatrix}$. Then $\vec{z} \in W$, however $3\vec{z} \notin W$. (or: $\vec{z} + \vec{z} \notin W$)

- $W = \left\{ \begin{pmatrix} x \\ y \end{pmatrix} : x \geq 0 \right\}$. Then W is not a vector space. For example, $\vec{z} = \begin{pmatrix} 2 \\ 0 \end{pmatrix} \in W$, but $(-1)\vec{z} = \begin{pmatrix} -2 \\ 0 \end{pmatrix} \notin W$.

Note that geometrically W is a right half plane in $\mathbb{R}^2$.

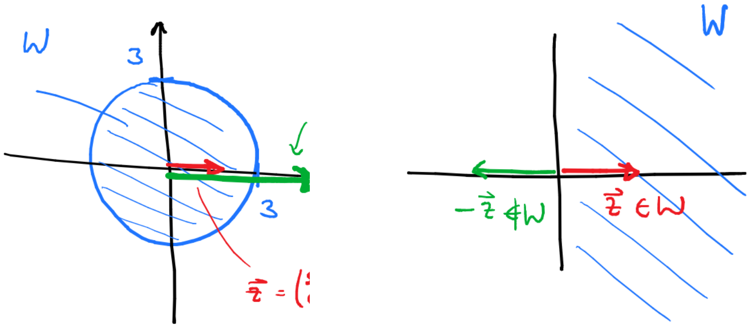

FIGURE 4.3: The sets W in the figures a not subspaces of $\mathbb{R}^2$.

Examples 4.15 (Examples and non-examples of subspaces of $\mathbb{R}^3$).

- For fixed $x_0, y_0, z_0 \in \mathbb{R}$ let $W = \left\{ \lambda \begin{pmatrix} x_0 \\ y_0 \\ z_0 \end{pmatrix} : \lambda \in \mathbb{R} \right\}$. Then W is a subspace of $\mathbb{R}^3$. Geometrically, W is a line in $\mathbb{R}^2$ passing through the origin which is parallel to the vector $\begin{pmatrix} x_0 \\ y_0 \\ z_0 \end{pmatrix}$.
- For fixed $a, b, c \in \mathbb{R}$ the set $W = \left\{ \begin{pmatrix} x \\ y \\ z \end{pmatrix} : ax + by + cz = 0 \right\}$ is a subspace of $\mathbb{R}^3$.

Proof. We use Proposition 4.10 to verify that W is a subspace of $\mathbb{R}^3$. Clearly, $\vec{0} \in W$ since $0a+0b+0c = 0$. Now let $\vec{w}_1 = \begin{pmatrix} x_1 \\ y_1 \\ z_1 \end{pmatrix}$ and $\vec{w}_2 = \begin{pmatrix} x_2 \\ y_2 \\ z_2 \end{pmatrix}$ in W and let $\lambda \in \mathbb{R}$. Then $\vec{w}_1 + \vec{w}_2 \in W$ because

$$a(x_1 + x_2) + b(y_1 + y_2) + c(z_1 + z_2) = (ax_1 + by_1 + cz_1) + (ax_2 + by_2 + cz_2) = 0 + 0 = 0.$$

Also $\lambda \vec{w}_1 \in W$ because

$$a(\lambda x_1) + b(\lambda y_1) + c(\lambda z_1) = \lambda(ax_1 + by_1 + cz_1) = \lambda 0 = 0.$$

Hence W is closed under sum and product with scalars, so it is a subspace of $\mathbb{R}$. □

Remark. If at least one of the numbers $a, b, c \in \mathbb{R}$ is different from zero, then W is a plane in $\mathbb{R}^3$ which passes through the origin and has normal vector $\vec{n} = (a, b, c)^t$. If $a = b = c = 0$, then $W = \mathbb{R}^3$.

- For fixed $a, b, c, d \in \mathbb{R}$ with $d \neq 0$, the set $W = \left\{ \begin{pmatrix} x \\ y \\ z \end{pmatrix} : ax + by + cz = d \right\}$ is **not** a subspace of $\mathbb{R}^3$.

  Proof. Let $\vec{w}_1 = \begin{pmatrix} x_1 \\ y_1 \\ z_1 \end{pmatrix}$ and $\vec{w}_2 = \begin{pmatrix} x_2 \\ y_2 \\ z_2 \end{pmatrix}$ in W. Then $\vec{w}_1 + \vec{w}_2 \notin W$ because

  $$a(x_1 + x_2) + b(y_1 + y_2) + c(z_1 + z_2) = (ax_1 + by_1 + cz_1) + (ax_2 + by_2 + cz_2) = d + d = 2d \neq d.$$

  (We also could have shown that if $\vec{w}_1 \in W$ and $\lambda \in \mathbb{R} \setminus \{1\}$, then $\lambda \vec{w}_1 \notin W$. Show this!) ☐

  Remark. If at least one of the numbers $a, b, c \in \mathbb{R}$ is different from zero, then W is a plane in $\mathbb{R}^3$ which has normal vector $\vec{n} = (a, b, c)^t$ but does not pass through the origin. If $a = b = c = 0$, then $W = \varnothing$.

- $W = \{\vec{x} \in \mathbb{R}^3 : \vec{x} \geq 5\}$ is not a subspace of $\mathbb{R}^3$ since it does not contain $\vec{0}$.

- $W = \{\vec{x} \in \mathbb{R}^3 : \vec{x} \leq 9\}$ is not a subspace of $\mathbb{R}^3$. For example, take $\vec{z} = \begin{pmatrix} 2 \\ 0 \\ 0 \end{pmatrix}$. Then $\vec{z} \in W$, however, for example, $7\vec{z} \notin W$.

- $W = \left\{ \begin{pmatrix} x \\ x^2 \\ x^3 \end{pmatrix} : x \in \mathbb{R} \right\}$. Then W is not a vector space. For example, $\vec{a} = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \in W$, but $2\vec{a} = \begin{pmatrix} 2 \\ 2 \\ 2 \end{pmatrix} \notin W$.

Examples 4.16 (Examples and non-examples of subspaces of $M(m \times n)$. The following sets are examples for subspaces of $M(m \times n)$:

- The set all matrices with $a_{11} = 0$.
- The set all matrices with $a_{11} = 5a_{12}$.
- The set all matrices such that its first row is equal to its last row.

If $m = n$, then also the following sets are subspaces of $M(n \times n)$:

- The set all symmetric matrices.
- The set all antisymmetric matrices.
- The set all diagonal matrices.
- The set all upper triangular matrices.

- The set all lower triangular matrices.

The following sets are **not** subspaces of $M(n \times n)$:

- The set all invertible matrices.
- The set all non-invertible matrices.
- The set all matrices with determinant equal to 1. The set all functions f with $f(7) = 13$.

Examples 4.17 (Examples and non-examples of subspaces of the set all functions from $\mathbb{R}$ to $\mathbb{R}$). Let V be the set of all functions from $\mathbb{R}$ to $\mathbb{R}$. Then V clearly is a real vector space. The following sets are examples for subspaces of V:

- The set all continuous functions.
- The set all differential functions.
- The set all bounded functions.
- The set all polynomials.
- The set all polynomials with degree ≤ 5.
- The set all functions f with $f(7) = 0$.
- The set all even functions.
- The set all odd functions.

The following sets are **not** subspaces of V:

- The set all polynomials with degree 3.
- The set all polynomials with degree ≥ 3.
- The set all functions f with $f(7) = 13$.
- The set all functions f with degree $f(7) \geq 0$.

Exercise. Prove these claims.

Definition 4.18. For $n \in \mathbb{N}_0$ let P_n be the set of all polynomials of degree less or equal to n.

Remark 4.19. P_n is a vector space.

Proof. Clearly, the zero function belongs to P_n (it is the polynomial of degree 0). For polynomials $p, q \in P_n$ and numbers $\lambda \in \mathbb{R}$, we clearly have that $p + q$ and λp are again polynomials of degree at most n, so they belong to P_n. By Proposition 4.10, P_n is a subspace of the space of all real functions, hence it is a vector space. $\square$

4.3 Linear Combinations and linear independence

In this section, we work with real vector spaces for definiteness sake. However, all the statements are also true for complex vector spaces. We only have to replace everywhere $\mathbb{R}$ by $\mathbb{C}$ and the word *real* by *complex*.

We start with a definition.

Definition 4.20. Let V be a real vector space and let $v_1, \dots, v_k \in V$ and $\alpha_1, \dots, \alpha_k \in \mathbb{R}$. Then the vector

$$v = \alpha_1 v_1 + \cdots \alpha_k v_k \tag{4.1}$$

is called a *linear combination of the vectors* $v_1, \dots, v_k \in V$.

Examples 4.21.

- Let $V = \mathbb{R}^3$ and let $\vec{v_1} = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}, \vec{v_2} = \begin{pmatrix} 4 \\ 5 \\ 6 \end{pmatrix}, \vec{a} = \begin{pmatrix} 9 \\ 12 \\ 15 \end{pmatrix}, \vec{b} = \begin{pmatrix} 3 \\ 3 \\ 3 \end{pmatrix}$.

  Then $\vec{a}$ and $\vec{b}$ are linear combinations of $\vec{v_1}$ and $\vec{v_2}$ because $\vec{a} = \vec{v_1} + 2\vec{v_2}$ and $\vec{b} = -\vec{v_1} + \vec{v_2}$.
- Let $V = M(2 \times 2)$ and let $A = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, B = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, R = \begin{pmatrix} 5 & 7 \\ -7 & 5 \end{pmatrix}, S = \begin{pmatrix} 1 & 2 \\ -2 & 3 \end{pmatrix}$.

  Then R is a linear combination of A and B because $R = 5A + 7B$. S is **not** a linear combination of A and B. To see this note that because clearly for every linear combination of A and B

  $$\alpha A + \beta B = \begin{pmatrix} \alpha & \beta \\ -\beta & \alpha \end{pmatrix}$$

  but S is not of this form (S has two different elements on its diagonal).

Definition and Theorem 4.22. *Let V be a real vector space and let $v_1, \dots, v_k \in V$. Then the set of all their possible linear combinations is denoted by*

$$\operatorname{span}\{v_1, \dots, v_k\} := \{\alpha_1 v_1 + \cdots + \alpha_k v_k \ : \ \alpha_1, \dots, \alpha_k \in \mathbb{R}\}.$$

It is a subspace of V and it is called the linear span of the $v_1, \dots, v_k$. *The vectors $v_1, \dots, v_k$ are called* generators *of the subspace* $\operatorname{span}\{v_1, \dots, v_k\}$.

Remark. Other names for "linear span" that are commonly used, are *subspace generated by the* $v_1, \dots, v_k$ or *subspace spanned by the* $v_1, \dots, v_k$. Instead of $\operatorname{span}\{v_1, \dots, v_k\}$ the notation $\operatorname{gen}\{v_1, \dots, v_k\}$ is used frequently. All these names and notations mean exactly the same.

Proof of Theorem 4.22. We have to show that $W := \operatorname{span}\{v_1, \dots, v_k\}$ is a subspace of V. To this end, we use again Proposition 4.10. Clearly, W is not empty since at least $\mathbb{0} \in W$ (we only need to choose all the $\alpha_j = 0$). Now let $u, w \in W$ and $\lambda \in \mathbb{R}$. We have to show that $\lambda u + w \in W$. Since $u, w \in W$, there are real numbers $\alpha_1, \dots, \alpha_k$ and $\beta_1, \dots, \beta_k$ such that $u = \alpha_1 v_1 + \dots, \alpha_k v_k$ and $w = \beta_1 w_1 + \cdots + \beta_k v_k$. Then

$$\begin{aligned} \lambda u + v &= \lambda(\alpha_1 v_1 + \cdots + \alpha_k v_k) + \beta_1 w_1 + \cdots + \beta_k v_k \\ &= \lambda \alpha_1 v_1 + \cdots + \lambda \alpha_k v_k) + \beta_1 w_1 + \cdots + \beta_k v_k \\ &= (\lambda \alpha_1 + \beta_1) v_1 + \cdots + (\lambda \alpha_k + \beta_k) v_k \end{aligned}$$

which belongs to W since it is a linear combination of the $v_1, \dots, v_k$. ☐

Remark.

- $\boxed{\text{The generators of a given subspace are not unique.}}$

  For example, let $A = \left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right), B = \left(\begin{smallmatrix} 0 & -1 \\ 1 & 0 \end{smallmatrix}\right), C = \left(\begin{smallmatrix} 1 & -1 \\ 1 & 1 \end{smallmatrix}\right)$. Then

$$\begin{aligned} \operatorname{span}\{A, B\} &= \{\alpha A + \beta B \,:\, \alpha, \beta \in \mathbb{R}\} &&= \left\{ \begin{pmatrix} \alpha & -\beta \\ \beta & \alpha \end{pmatrix} \,:\, \alpha, \beta \in \mathbb{R} \right\}, \\ \operatorname{span}\{A, B, C\} &= \{\alpha A + \beta B + \gamma C \,:\, \alpha, \beta, \gamma \in \mathbb{R}\} &&= \left\{ \begin{pmatrix} \alpha + \gamma & -(\beta + \gamma) \\ \beta + \gamma & \alpha + \gamma \end{pmatrix} \,:\, \alpha, \beta, \gamma \in \mathbb{R} \right\}, \\ \operatorname{span}\{A, C\} &= \{\alpha A + \gamma C \,:\, \alpha, \gamma \in \mathbb{R}\} &&= \left\{ \begin{pmatrix} \alpha + \gamma & -\gamma \\ \gamma & \alpha + \gamma \end{pmatrix} \,:\, \alpha, \gamma \in \mathbb{R} \right\}. \end{aligned}$$

  We see that $\operatorname{span}\{A, B\} = \operatorname{span}\{A, B, C\} = \operatorname{span}\{A, C\}$ (in all cases it consists of exactly those matrices whose diagonal entries are equal and the off-diagonal entries differ by a minus sign). So we see that neither the generators nor their number is unique.

- $\boxed{\text{If a vector is a linear combination, then the coefficients are not necessarily unique.}}$

  For example, if A, B, C are the matrices above, then $A + B + C = 2A + 2B = 2C$ or $A + 2B + 3C = 4A + 5B = B + 4C$, etc.

Exercise 4.23. (i) Find generators of P_n.
Solution. A set of generators is for example $\{1, X, X^2, \dots, X^{n-1}, X^n\}$ since every vector in P_n is a polynomial of the form $p = \alpha_n X^n + \alpha_{n-1} X^{n-1} + \dots + \alpha_1 X + \alpha_0$, so it is a linear combination of the polynomials X^n, X^{n-1}, …, X, 1. ◇

(ii) Find generators of the set of all antisymmetric 2×2 matrices.
Solution. A set of generators is for example $\{\left(\begin{smallmatrix} 0 & 1 \\ -1 & 0 \end{smallmatrix}\right)\}$ ◇

(iii) Let $V = \mathbb{R}^3$ and let $\vec{v}, \vec{w} \in \mathbb{R}^3 \setminus \{\vec{0}\}$. Describe $\operatorname{span}\{\vec{v}\}$ and $\operatorname{span}\{\vec{v}, \vec{w}\}$ geometrically
Solution.

- $\operatorname{span}\{\vec{v}\}$ is a line which passes through the origin and is parallel to $\vec{v}$.
- $\operatorname{span}\{\vec{v}, \vec{w}\}$ is a plane which passes through the origin and is parallel to $\vec{v}$ and $\vec{w}$ if $\vec{v} \nparallel \vec{w}$. Otherwise, if $\vec{v} \parallel \vec{w}$, then it is a line which passes through the origin and is parallel to $\vec{v}$. ◇

Remark 4.24. Let V be a vector space and let $v_1, \dots, v_n$ and $w_1, \dots, w_m$ be vectors in V. Then the following is equivalent:

(i) $\operatorname{span}\{v_1, \dots, v_n\} = \operatorname{span}\{w_1, \dots, w_m\}$.

(ii) $v_j \in \operatorname{span}\{w_1, \dots, w_m\}$ for every $j = 1, \dots, n$ and $w_j \in \operatorname{span}\{v_1, \dots, v_n\}$ for every $j = 1, \dots, m$.

Proof. (i) $\Longrightarrow$ (ii) is clear.
(i) $\Longrightarrow$ (ii): Note that $v_j \in \operatorname{span}\{w_1, \dots, w_m\}$ for every $j = 1, \dots, n$ implies that every v_j can be written as a linear combination of the $w_1, \dots, w_m$. Then also every linear combination of $v_1, \dots, v_n$ is a linear combination of $w_1, \dots, w_m$. This implies that $\operatorname{span}\{v_1, \dots, v_n\} \subseteq \operatorname{span}\{w_1, \dots, w_m\}$. The converse inclusion $\operatorname{span}\{w_1, \dots, w_m\} \subseteq \operatorname{span}\{v_1, \dots, v_n\}$ can be shown analogously. Both inclusions together show that we must have equality. ☐

No we ask ourselves how many vectors we need at least in order to generate $\mathbb{R}^n$. We now that for example $\mathbb{R}^n = \operatorname{span}\{\vec{e}_1, \dots, \vec{v}_n\}$. So in this case we have n vectors that generate $\mathbb{R}^n$. Could it be that less vetors are sufficient? Clearly, if we take away one of the $\vec{e}_j$, then the remaining system no longer generates $\mathbb{R}^n$ since "one coordinate is missing". However, could we maybe find other vectors so that $n-1$ or less vectors are enough to generate all of $\mathbb{R}^n$? The next Proposition says that this is not possible.

Proposition 4.25. *Let $\vec{v}_1, \dots, \vec{v}_k$ be vectors in $\mathbb{R}^n$. If Let* $\operatorname{span}\{\vec{v}_1, \dots, \vec{v}_k\} = \mathbb{R}^n$*, then $m \geq n$.*

Proof. Let $A = (\vec{v}_1|\dots|\vec{v}_m)$ be the matrix whose columns are the given vectors. We know that there exists an invertible matrix E such that $A' = EA$ is in reduced echelon form (the matrix E is the product of elementary matrices which correspond to the steps in the Gauß-Joradan process to arrive at the reduced echelon form). Now, if $m < n$, then we know that A' must have at least one row which consists of zeros only. If we can find a vector $\vec{w}$ such that it is transformed to $\vec{\mathrm{e}}_n$ under the Gausß-Jordan process, then we would have that $A\vec{x} = \vec{w}$ is inconsistent, which means that $\vec{w} \notin \operatorname{span}\{\vec{v}_1, \dots, \vec{v}_n\}$. How do we find such a vector $\vec{w}$? Well, we only have to start with $\vec{\mathrm{e}}_n$ and "do the Gauß-Jordan porcess backwards". In other words, we define $\vec{w} = E^{-1}\vec{\mathrm{e}}_n$. Now if we apply the Gauß-Jordan process to the augmented matrix $(A|\vec{w})$, we arrive at $(EA|E\vec{w}) = (A'|\vec{\mathrm{e}}_n)$ which we already know is inconsistent.
Therefore, $m < n$ is not possible and we must therefore have that $m \geq n$. ☐

Now we will pay attention to when the coefficients of a linear combination are unique.

Let V be a vector space and fix vectors $v_1, \dots, v_k$ in V. We consider the equation

$$\boxed{\alpha_1 v_1 + \dots + \alpha_k v_k = \mathbb{0}} \tag{4.2}$$

and we ask ourselves how many solutions this equation has for $\alpha_1, \dots, \alpha_k$. In other words, we ask if and in how many ways $\mathbb{0}$ can be written as a linear combination of the $v_1, \dots, v_k$. Clearly, there is always at least one solution, namely $\alpha_1 = \dots = \alpha_k = 0$. This solution is called the *trivial solution*. On the other hand, if we have one non-trivial solution, then we automatically have infinitely many solutions, because if $\alpha_1, \dots, \alpha_k$ is solution, then also $c\alpha_1, \dots, c\alpha_k$ is solution for arbitrary $c \in \mathbb{R}$ since

$$c\alpha_1 v_1 + \dots + c\alpha_k v_k = c(\alpha_1 v_1 + \dots + \alpha_k v_k) = c\mathbb{0} = \mathbb{0}.$$

So we see that only one of the following two cases can occur: (4.2) as exactly one solution (namely the trivial one) or it has infinitely many solutions. Note that this is analogous to the situation of the solutions of homogeneous linear systems: They have either only the trivial solution or they have infinitely many solutions.
The following definition distinguishes between the two cases.

Definition 4.26. In the situation as above, the vectors $v_1, \dots, v_k$ are called *linearly independent* if (4.2) has only one solution. The are called *linearly dependent* if (4.2) has more than one solution.

Before we continue with the theory, we give a few examples.

Examples. (i) The vectors $\vec{v_1} = \binom{1}{2}$ and $\vec{v_2} = \binom{-4}{-8} \in \mathbb{R}^2$ are linearly dependent because $4\vec{v_1} + \vec{v_2} = \vec{0}$.

(ii) The vectors $\vec{v_1} = \left(\begin{smallmatrix}1\\2\end{smallmatrix}\right)$ and $\vec{v_2} = \left(\begin{smallmatrix}5\\0\end{smallmatrix}\right) \in \mathbb{R}^2$ are linearly independent.

Proof. Consider the equation $\alpha\vec{v_1} + \beta\vec{v_2} = \vec{0}$. This equation is equivalent to the following system of linear equations for α and β:

$$\begin{aligned} \alpha + 3\beta &= 0 \\ 2\alpha + 0\beta &= 0. \end{aligned}$$

We can use the Gauß-Jordan process to obtain all solutions. However, in this case we easily see that $\alpha = 0$ (from the second line) and then that $\beta = -\frac{1}{3}\alpha = 0$. Note that we could also have calculated $\det(\left(\begin{smallmatrix}1&3\\2&0\end{smallmatrix}\right) = -6 \neq 0$ to conclude that the homogeneous system above has only the trivial solution. Observe that the columns of the matrix are exactly the given vectors. ☐

(iii) The vectors $\vec{v_1} = \begin{pmatrix}1\\1\\1\end{pmatrix}$ and $\vec{v_2} = \begin{pmatrix}2\\3\\4\end{pmatrix} \in \mathbb{R}^2$ are linearly independent.

Proof. Consider the equation $\alpha\vec{v_1} + \beta\vec{v_2} = \vec{0}$. This equation is equivalent to the following system of linear equations for α and β:

$$\begin{aligned} \alpha + 2\beta &= 0 \\ \alpha + 3\beta &= 0 \\ \alpha + 4\beta &= 0. \end{aligned}$$

If we subtract the first from the second equation, we obtain $\beta = 0$ and then $\alpha = -2\beta = 0$. So again, this system has only the trivial solution and therefore the vectors $\vec{v_1}$ and $\vec{v_2}$ are linearly independent. ☐

(iv) Let $\vec{v_1} = \begin{pmatrix}1\\1\\1\end{pmatrix}$, $\vec{v_2} = \begin{pmatrix}-1\\2\\3\end{pmatrix}$ $\vec{v_3} = \begin{pmatrix}0\\0\\1\end{pmatrix}$ and $\vec{v_4} = \begin{pmatrix}0\\6\\8\end{pmatrix} \in \mathbb{R}^2$ Then

(a) The system $\{\vec{v}_1, \vec{v}_2, \vec{v}_3\}$ is linearly independent.

(b) The system $\{\vec{v}_1, \vec{v}_2, \vec{v}_4\}$ is linearly dependent.

Proof. (a) Consider the equation $\alpha\vec{v_1} + \beta\vec{v_2} + \gamma\vec{v_3} = \vec{0}$. This equation is equivalent to the following system of linear equations for α, β and γ:

$$\begin{aligned} \alpha - 1\beta + 0\gamma &= 0 \\ \alpha + 2\beta + 0\gamma &= 0 \\ \alpha + 3\beta + 1\gamma &= 0. \end{aligned}$$

We use the Gauß-Jordan process to solve the system. Note that the columns of the matrix associated to the above system are exactly the given vectors $\vec{v}_1, \vec{v}_2, \vec{v}_3$.

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

Therefore the unique solution is $\alpha = \beta = \gamma = 0$ and consequently the vectors $\vec{v}_1, \vec{v}_2, \vec{v}_3$ are linearly independent.

Observe that we also could have calculated $\det A = 3 \neq 0$ to conclude that the homogeneous system has only the trivial solution.

(b) Consider the equation $\alpha\vec{v_1} + \beta\vec{v_2} + \delta\vec{v_4} = \vec{0}$. This equation is equivalent to the following system of linear equations for α, β and δ:

$$\begin{aligned} \alpha - 1\beta + 0\delta &= 0 \\ \alpha + 2\beta + 6\delta &= 0 \\ \alpha + 3\beta + 8\delta &= 0. \end{aligned}$$

We use the Gauß-Jordan process to solve the system. Note that the columns of the matrix associated to the above system, are exactly the given vectors.

$$A = \begin{pmatrix} 1 & -1 & 0 \\ 1 & 2 & 6 \\ 1 & 3 & 8 \end{pmatrix} \longrightarrow \begin{pmatrix} 1 & -1 & 0 \\ 0 & 3 & 6 \\ 0 & 4 & 8 \end{pmatrix} \longrightarrow \begin{pmatrix} 1 & -1 & 0 \\ 0 & 1 & 2 \\ 0 & 1 & 2 \end{pmatrix} \longrightarrow \begin{pmatrix} 1 & -1 & 0 \\ 0 & 1 & 2 \\ 0 & 0 & 0 \end{pmatrix}$$
$$\longrightarrow \begin{pmatrix} 1 & 0 & 2 \\ 0 & 1 & 2 \\ 0 & 0 & 0 \end{pmatrix}.$$

Therefore the unique solution is $\alpha = \beta = \gamma = 0$ and consequently the vectors $\vec{v}_1, \vec{v}_2, \vec{v}_3$ are linearly independent. So there are infinitely many solutions. If we take $\delta + t$, then $\alpha = \beta = -2t$. Consequently the vectors $\vec{v}_1, \vec{v}_2, \vec{v}_3$ are linearly dependent, because, for example, $-2\vec{v}_1 - 2\vec{v}_2 + \vec{v}_3 = 0$ (taking $t = 1$).

Observe that we also could have calculated $\det A = 0$ to conclude that the system has infinite solutions. ☐

(v) The matrices $\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$ and $\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$ are linearly independent in $M(2 \times 2)$.

(vi) The matrices $\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ and $\begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}$ are linearly dependent in $M(2 \times 2)$.

After these examples we will proceed with some facts on linear independence. We start with the special case when we have only two vectors.

Proposition 4.27. *Let v_1, v_2 be vectors in a vector space V. Then v_1, v_2 are linearly dependent if and only if one vector is a multiple of the other.*

Proof. Assume that v_1, v_2 are linearly dependent. Then there exist $\alpha_1, \alpha_2 \in \mathbb{R}$ such that $\alpha_1 v_1 + \alpha_2 v_2 = 0$ and at least one of the α_1 and α_2 is different from zero. Let's say that $\alpha_1 \neq 0$. Then we have $v_1 + \frac{\alpha_2}{\alpha_1} v_2 = 0$, hence $v_1 = -\frac{\alpha_2}{\alpha_1} v_2$.
Now assume on the other hand that, e.g., v_1 is a multiple of v_2, that is $v_1 = \lambda v_2$ for some $\lambda \in \mathbb{R}$. Then $v_1 - \lambda v_2 = 0$ which is a nontrivial solution of $\alpha_1 v_1 + \alpha_2 v_2 = 0$ because we can take $\alpha_1 = 1 \neq 0$ (note that λ may be zero). ☐

Proposition 4.28. *Let V be a vector space.*

(i) Every system of vectors which contains $\mathbb{0}$ is linearly dependent.

(ii) *Let $v_1, \dots, v_k \in V$ and assume that there are $\alpha_1, \dots, \alpha_k \in \mathbb{R}$ such that $\alpha_1 v_1 + \cdots + \alpha_k v_k = \mathbb{0}$. If $\alpha_\ell \neq 0$, then v_ℓ is a linear combination of the other v_j.*

(iii) *If the vectors $v_1, \dots, v_k \in V$ are linearly dependent, then for every $w \in V$, the vectors $v_1, \dots, v_k, w$ are linearly dependent.*

(iv) *If the vectors $v_1, \dots, v_k \in V$ are linearly independent, then every subset of them is linearly independent.*

(v) *If $v_1, \dots, v_k$ are vectors in V and w is a linear combination of them, then $v_1, \dots, v_k, w$ are linearly dependent.*

Proof. (i) Let $v_1, \dots, v_k \in V$. Clearly $1\mathbb{0} + 0v_1 + \cdots + 0v_k = \mathbb{0}$ is non-trivial linear combination with gives $\mathbb{0}$. Therefore the system $\{v_1, \dots, v_k, \mathbb{0}\}$ is linearly dependent.

(ii) If $\alpha_\ell \neq 0$, then we can solve for v_ℓ: $v_\ell = -\frac{\alpha_1}{\alpha_\ell} v_1 - \cdots - \frac{\alpha_{\ell-1}}{\alpha_\ell} v_{\ell-1} - \frac{\alpha_{\ell+1}}{\alpha_\ell} v_{\ell+1} - \cdots - \frac{\alpha_k}{\alpha_\ell} v_k$.

(iii) Suppose that the vectors $v_1, \dots, v_k \in V$ are linearly dependent. Then there exist $\alpha_1, \dots, \alpha_k \in \mathbb{R}$ such that at least one of them is different from zero and $\alpha_1 v + 1 + \cdots + \alpha_k v_k = \mathbb{0}$. But then also $\alpha_1 v + 1 + \cdots + \alpha_k v_k + 0w = \mathbb{0}$ which shows that the system $\{v_1, \dots, v_k, w\}$ is linearly dependent.

(iv) If the vectors $v_1, \dots, v_k \in V$ are linearly independent, then there exists a non-trivial linear combination $\alpha_1 v_1 + \cdots + \alpha_k v_k = \mathbb{0}$. But then also $\alpha_1 v_1 + \cdots + \alpha_k v_k + 0w = \mathbb{0}$ is a non-trivial linear combination.

(v) Assume that w is a linear combination of $v_1, \dots, v_k$. Then there exist $\alpha_1, \dots, \alpha_k \in \mathbb{R}$ such that $w = \alpha_1 v_1 + \cdots + \alpha_k v_k$. Therefore we obtain $w - \alpha_1 v_1 - \cdots - \alpha_k v_k = \mathbb{0}$ which is non-trivial linear combination since the coefficient of w is 1.

□

Now we specialise to the case when $V = \mathbb{R}^n$. Let us take vectors $\vec{v}_1, \dots, \vec{v}_k \in \mathbb{R}^n$ and let us write $(\vec{v}_1|\cdots|\vec{v}_k)$ for the $n \times k$ matrix whose columns are the vectors $\vec{v}_1, \dots, \vec{v}_k$.

Lemma 4.29. *With the above notation, the following statements are equivalent:*

(i) *$\vec{v}_1, \dots, \vec{v}_k$ are linearly dependent.*

(ii) *There exist $\alpha_1, \dots, \alpha_k$ not all equal to zero, such that $\alpha_1 \vec{v}_1 + \cdots + \alpha_k \vec{v}_k = 0$.*

(iii) *There exists a vector $\begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_k \end{pmatrix} \neq \vec{0}$ such that $(\vec{v}_1|\cdots|\vec{v}_k) \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_k \end{pmatrix} = \vec{0}$.*

(iv) *The homogeneous system corresponding to the matrix $(\vec{v}_1|\cdots|\vec{v}_k)$ has at least one non-trivial (and therefore infinitely many) solutions.*

Proof. (i) $\Longrightarrow$ (ii) is simply the definition of linear independence. (ii) $\Longrightarrow$ (iii) is only rewriting the vector equation in matrix form. (iv) only says in word what the equation in (iii) means. And finally, (iv) $\Longrightarrow$ (i) because every non trivial solution the inhomogeneous system associated to $(\vec{v}_1|\cdots|\vec{v}_k)$ gives a non-trivial solution of $\alpha_1 \vec{v}_1 + \cdots + \alpha_k \vec{v}_k$. □

Since we know that a homogeneous linear system with more unknowns than equations has infinitely many solutions, we immediately obtain the following corollary.

Corollary 4.30. *Let* $\vec{v}_1, \dots, \vec{v}_k \in \mathbb{R}^n$.

(i) If $k > n$*, then the vectors* $\vec{v}_1, \dots, \vec{v}_k$ *are linearly dependent.*

(ii) If the vectors $\vec{v}_1, \dots, \vec{v}_k$ *are linearly independent, then* $k \le n$.

Observe that (ii) does **not** say that if $k \le n$, then the vectors $\vec{v}_1, \dots, \vec{v}_k$ are linearly independent. It just says that we cannot have a system of more than n vectors which is linearly independent.

Now we specialise further to the case when $k = n$.

Theorem 4.31. *Let* $\vec{v}_1, \dots, \vec{v}_n$ *be vectors in* $\mathbb{R}^n$*. Then the following is equivalent:*

(i) $\vec{v}_1, \dots, \vec{v}_n$ *are linearly independent.*

(ii) *The only solution of* $(\vec{v}_1|\cdots|\vec{v}_n)\begin{pmatrix}\alpha_1\\ \vdots\\ \alpha_n\end{pmatrix} = \vec{0}$ *is the zero vector* $\begin{pmatrix}\alpha_1\\ \vdots\\ \alpha_n\end{pmatrix} = \vec{0}$.

(iii) *The matrix* $(\vec{v}_1|\cdots|\vec{v}_n)$ *is invertible.*

(iv) $\det(\vec{v}_1|\cdots|\vec{v}_n) \neq 0$.

Proof. The prove is analogous to the proof of Lemma 4.29 □

Exercise 4.32. Formulate an analogous theorem for linearly dependent vectors.

Now we can state when a system n vectors in $\mathbb{R}^n$ is generating $\mathbb{R}^n$.

Theorem 4.33. *Let* $\vec{v}_1, \dots, \vec{v}_n$ *be vectors in* $\mathbb{R}^n$*. and let* $A = (\vec{v}_1|\cdots|\vec{v}_n)$ *be the matrix whose columns are the given vectors* $\vec{v}_1, \cdots, \vec{v}_n$*. Then the following is equivalent:*

(i) $\vec{v}_1, \dots, \vec{v}_n$ *are linearly independent.*

(ii) $\mathbb{R}^n = \operatorname{span}\{\vec{v}_1, \dots, \vec{v}_n\}$.

(iii) $\det A \neq 0$.

Proof. (i) $\iff$ (iii) is shown in Theorem 4.31.
(ii) $\iff$ (iii): The vectors $\vec{v}_1, \dots, \vec{v}_n$ generate $\mathbb{R}^n$ if and only if for every $\vec{w} \in \mathbb{R}^n$ there exist numbers $\beta_1, \dots, \beta_n$ such that $\beta_1\vec{v}_1 + \cdots + \beta_n v_n = \vec{w}$. In matrix form that means that $A\begin{pmatrix}\beta_1\\ \vdots\\ \beta_n\end{pmatrix} = \vec{w}$.
By Theorem **??** (in Chapter 3 on existence and uniqueness of solutions of inhomogeneous linear systems) we know that this has a solution for every vector $\vec{w}$ if and only if A is invertible (because if we apply Gauß-Jordan to A, we must get to the identity matrix). □

The proof of the precding theorem basically goes like this: We consider the equation $A\vec{\beta} = \vec{w}$. When are the vectors $\vec{v}_1, \dots, \vec{v}_n$ linearly independent? – They are linearly independent if and only if for $\vec{w} = \vec{0}$ the system has only the trivial solution. This happens if and only if the reduced echelon form of A is the identity matrix. And this happens if and only if $\det A \neq 0$.
When do the vectors $\vec{v}_1, \dots, \vec{v}_n$ generate $\mathbb{R}^n$? – They do, if and only if for every given vector $\vec{w} \in \mathbb{R}^n 0$ the system has at least one solution. This happens if and only if the reduced echelon form of A is the identity matrix. And this happens if and only if $\det A \neq 0$.

Since a square matrix A in invertible if and only if its transpose A^t is invertible, Theorem 4.33 leads immediately to the following corollary.

Corollary 4.34. *For a matrix $A \in M(n \times n)$ the following is equivalent:*

(i) *A is invertible.*

(ii) *The columns of A are linearly independent.*

(iii) *The rows of A are linearly independent.*

We end this section with more examples.

Examples.

- Recall that P_n is the vector space of all polynomials of degree $\leq n$.

In P_3, we consider the vectors $p_1 = X^3 - 1$, $p_2 = X^2 - 1$, $p_3 = X - 1$. These vectors are linearly independent.

Proof. Let $\alpha_1, \alpha_2, \alpha_3 \in \mathbb{R}$ such that $\alpha_1 p_1 + \alpha_2 p_2 + \alpha_3 p_3 = 0$. This means that

$$\begin{aligned} 0 &= \alpha_1(X^3 - 1) + \alpha_2(X^2 - 1) + \alpha_3(X - 1) \\ &= \alpha_1 X^3 + \alpha_2 X^2 + \alpha_3 X - (\alpha_1 + \alpha_2 + \alpha_3). \end{aligned}$$

Comparing coefficients, it follows that $\alpha_1 = 0$, $\alpha_2 = 0$, $\alpha_3 = 0$ and $\alpha_1 + \alpha_2 + \alpha_3 = 0$ which shows that p_1, p_2 and p_3 are linearly independent. ☐

If in addition we take $p_4 = X^3 - X^2$, then the system p_1, p_2, p_3 and p_4 is linearly dependent.

Proof. As before, let $\alpha_1, \alpha_2, \alpha_3, \alpha_4 \in \mathbb{R}$ such that $\alpha_1 p_1 + \alpha_2 p_2 + \alpha_3 p_3 + \alpha_4 p_4 = 0$. This means that

$$\begin{aligned} 0 &= \alpha_1(X^3 - 1) + \alpha_2(X^2 - 1) + \alpha_3(X - 1) + \alpha_4(X^3 - X^2) \\ &= (\alpha_1 + \alpha_4)X^3 + (\alpha_2 - \alpha_4)X^2 + \alpha_3 X - (\alpha_1 + \alpha_2 + \alpha_3). \end{aligned}$$

Comparing coefficients, this is equivalent to $\alpha_1 + \alpha_4 = 0$, $\alpha_2 - \alpha_4 = 0$, $\alpha_3 = 0$ and $\alpha_1 + \alpha_2 + \alpha_3 = 0$. This system of equations has infinitely many solutions. They are given by $\alpha_2 = \alpha_4 = -\alpha_1 \in \mathbb{R}$, $\alpha_3 = 0$. Therefore p_1, p_2, p_3 and p_4 are linearly dependent. ☐

Exercise. Show that p_1, p_2, p_3 and p_5 are linearly independent if $p_5 = X^3 + X^2$.

- In P_2, we consider the vectors $p_1 = X^2 + 2X - 1$, $p_2 = 5X + 2$, $p_3 = 2X^2 - 11X - 8$. These vectors are linearly dependent.

  Proof. Let $\alpha_1, \alpha_2, \alpha_3 \in \mathbb{R}$ such that $\alpha_1 p_1 + \alpha_2 p_2 + \alpha_3 p_3 = 0$. This means that

  $$\begin{aligned} 0 &= \alpha_1(X^2 + 2X - 1) + \alpha_2(5X + 2) + \alpha_3(2X^2 - 11X - 8) \\ &= (\alpha_1 + 2\alpha_3)X^2 + (2\alpha_1 + 5\alpha_2 - 11\alpha_3)X - \alpha_1 + 2\alpha_2 - 8\alpha_3). \end{aligned}$$

  Comparing coefficients, it follows that $\alpha_1 + 2\alpha_3 = 0$, $2\alpha_1 + 5\alpha_2 - 11\alpha_3 = 0$, $-\alpha_1 + 2\alpha_2 - 8\alpha_3 = 0$. We write this in matrix form and apply the Gauß-Jordan:

  $$\begin{pmatrix} 1 & 0 & 2 \\ 2 & 5 & -11 \\ -1 & 2 & -8 \end{pmatrix} \longrightarrow \begin{pmatrix} 1 & 0 & 2 \\ 0 & 5 & -15 \\ 0 & 2 & -6 \end{pmatrix} \longrightarrow \begin{pmatrix} 1 & 0 & 2 \\ 0 & 1 & -3 \\ 0 & 1 & -3 \end{pmatrix} \longrightarrow \begin{pmatrix} 1 & 0 & 2 \\ 0 & 1 & -3 \\ 0 & 0 & 0 \end{pmatrix}$$

  This shows that the system has non-trivial solutions (find them!) and therefore p_1, p_2 and p_3 are linearly dependent. ☐

- In $V = M(2 \times 2)$ consider $A = \left(\begin{smallmatrix} 1 & 2 \\ 2 & 1 \end{smallmatrix}\right), B = \left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right), C = \left(\begin{smallmatrix} 0 & 5 \\ 5 & 0 \end{smallmatrix}\right)$. Then A, B, C are linearly dependent because $A - B - \frac{1}{5}C = 0$.

- In $V = M(2 \times 3)$ consider $A = \left(\begin{smallmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{smallmatrix}\right), B = \left(\begin{smallmatrix} 2 & 2 & 2 \\ 1 & 1 & 1 \end{smallmatrix}\right), C = \left(\begin{smallmatrix} 1 & 2 & 2 \\ 2 & 1 & 1 \end{smallmatrix}\right)$. Then A, B, C are linearly independent.

  Exercise. Prove this!

4.4 Basis and dimension

In this section, we work with real vector spaces for definiteness sake. However, all the statements are also true for complex vector spaces. We only have to replace everywhere $\mathbb{R}$ by $\mathbb{C}$ and the word *real* by *complex*.

Definition 4.35. Let V be a vector space. A *basis* of V is a set of vectors $\{v_1, \dots, v_n\}$ in V which is linearly independent and generates V.

The following remark shows that a basis is a *minimal system of generators of* V and at the same time a *maximal system of linear independent vectors.*

Remark. Let $\{v_1, \dots, v_n\}$ be a basis of V.

(i) Let $w \in V$. Then $\{v_1, \dots, v_n, w\}$ in not a basis of V because this system of vectors is no longer linearly independent by Proposition 4.28 (v)

(ii) If we take away one of the vectors from $\{v_1, \dots, v_n\}$, then it is no longer a basis of V because the new system of vectors no longer generates V. For example, if we take away v_1, then $v_1 \notin \operatorname{span}\{v_2, \dots, v_n\}$ (otherwise $v_1, \dots, v_n$ would be linearly dependent), and therefore $\operatorname{span}\{v_2, \dots, v_n\} \neq \mathbb{R}^n$.

Remark 4.36. Every basis of $\mathbb{R}^n$ has exactly n elements. To see this note that by Corollary 4.30, a basis can have at most n elements because oetherwise it could not be linearly independent. On the other hand, if it had less elements than n elements, then it cannot be a generator of $\mathbb{R}^n$ by Remark 4.25.

Examples 4.37.

- A basis of $\mathbb{R}^3$ is, for example, $\left\{ \begin{pmatrix}1\\0\\0\end{pmatrix}, \begin{pmatrix}0\\1\\0\end{pmatrix}, \begin{pmatrix}0\\0\\1\end{pmatrix} \right\}$. The vectors of this basis are the standard unit vectors. The basis is called the *standard basis* (or *canonical basis*) of $\mathbb{R}^3$.

  Other examples of bases of $\mathbb{R}^3$ are

  $$\left\{ \begin{pmatrix}1\\0\\0\end{pmatrix}, \begin{pmatrix}1\\1\\0\end{pmatrix}, \begin{pmatrix}1\\1\\1\end{pmatrix} \right\}, \quad \left\{ \begin{pmatrix}1\\2\\3\end{pmatrix}, \begin{pmatrix}4\\5\\6\end{pmatrix}, \begin{pmatrix}0\\2\\1\end{pmatrix} \right\}$$

  Exercise. Verify that the systems above are bases of $\mathbb{R}^3$.

  The following systems are **not** bases of $\mathbb{R}^3$

  $$\left\{ \begin{pmatrix}1\\2\\3\end{pmatrix}, \begin{pmatrix}4\\5\\6\end{pmatrix}, \begin{pmatrix}3\\2\\1\end{pmatrix} \right\}, \quad \left\{ \begin{pmatrix}1\\2\\3\end{pmatrix}, \begin{pmatrix}4\\5\\6\end{pmatrix} \right\}, \quad \left\{ \begin{pmatrix}1\\0\\0\end{pmatrix}, \begin{pmatrix}1\\1\\0\end{pmatrix}, \begin{pmatrix}1\\1\\1\end{pmatrix}, \begin{pmatrix}0\\0\\1\end{pmatrix} \right\}.$$

  Exercise. Verify that the systems above are not bases of $\mathbb{R}^3$.

- The *standard basis in* $\mathbb{R}^n$ (or *canonical basis in* $\mathbb{R}^n$) is $\{\vec{e}_1, \dots, \vec{e}_n\}$. Recall that the $\vec{e}_j$ are the standard unit vectors whose jth entry is 1 and all other entries are 0.

  Exercise. Verify that they form a basis of $\mathbb{R}^n$.

- The *standard basis in* P_n (or *canonical basis in* P_n) is $\{1, X, X^2, \dots, X^n\}$.

  Exercise. Verify that they form a basis of $\mathbb{R}^n$.

- Let $p_1 = X,\ p_2 = 2X^2 + 5X - 1, p_3 = 3X^2 + X + 2$. Then the system $\{p_1, p_2, p_3\}$ is a basis of P_2.

  Proof. We have to show that the system in linearly independent and that it generates the space P_2. Let $q = aX^2 + bX + c \in P_2$. We want to see if there are $\alpha_1, \alpha_2, \alpha_3 \in \mathbb{R}$ such that $q = \alpha_1 p_1 + \alpha_2 p_2 + \alpha_3 p_3$. If we write this equation out, we find

  $$\begin{aligned} aX^2 + bX + c &= \alpha_1 X + \alpha_2(2X^2 + 5X - 1) + \alpha_3(3X^2 + X + 2) \\ &= (2\alpha_2 + 3\alpha_3)X^2 + (\alpha_1 + 5\alpha_2 + \alpha_3)X - \alpha_2 + 2\alpha_3. \end{aligned}$$

Comparing coefficients, we obtain the following system of linear equations for the α_j:

$$\left.\begin{aligned} 2\alpha_2 +3\alpha_3 &= a\\ \alpha_1 + \ 5\alpha_2 \ +\alpha_3 &= b\\ -\alpha_2 +2\alpha_3 &= c \end{aligned}\right\} \qquad \text{in matrix form:} \qquad \begin{pmatrix} 0 & 2 & 3\\ 1 & 5 & 1\\ 0 & -1 & 2 \end{pmatrix} \begin{pmatrix} \alpha_1\\ \alpha_2\\ \alpha_3 \end{pmatrix} = \begin{pmatrix} a\\ b\\ c \end{pmatrix}.$$

Now we apply Gauß-Jordan to the augmented matrix:

$$\left(\begin{array}{ccc|c} 0 & 2 & 3 & a\\ 1 & 5 & 1 & b\\ 0 & -1 & 2 & c \end{array}\right) \longrightarrow \left(\begin{array}{ccc|c} 1 & 5 & 1 & b\\ 0 & -1 & 2 & c\\ 0 & 2 & 3 & a \end{array}\right) \longrightarrow \left(\begin{array}{ccc|c} 1 & 0 & 11 & b+5c\\ 0 & 1 & -2 & c\\ 0 & 0 & 7 & a+2c \end{array}\right).$$

So we see that there is exactly one solution for any given q. The existence of such a solution shows that $\{p_1, p_2, p_3\}$ generates P_2. We also see that there for any give $q \in P_2$ there is exactly one way to write it as a linear combination of p_1, p_2, p_3. If we take the special case $q = 0$, this shows that the system is linearly independent. In summary, $\{p_1, p_2, p_3\}$ is a basis of P_2. ☐

- Let $p_1 = X + 1,\ p_2 = X^2 + X, p_3 = X^3 + X^2 .p_4 = X^3 + X^2 + X + 1$. Then the system $\{p_1, p_2, p_3, p_4\}$ is **not** a basis of P_2.

  Exercise. Show this!

- In the spaces $M(m \times n)$, the set of all matrices A_{ij} form a basis, where A_{ij} is the matrix with $a_{ij} = 1$ and all other entries equal to 0. For example, in $M(2\times 3)$ we have the following basis:

$$\begin{pmatrix} 1 & 0 & 0\\ 0 & 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 1 & 0\\ 0 & 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 & 1\\ 0 & 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 & 0\\ 1 & 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 & 0\\ 0 & 1 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 & 0\\ 0 & 0 & 1 \end{pmatrix}.$$

- Let $A = \left(\begin{smallmatrix} 1 & 0\\ 0 & 0 \end{smallmatrix}\right),\ B = \left(\begin{smallmatrix} 1 & 0\\ 1 & 0 \end{smallmatrix}\right),\ C = \left(\begin{smallmatrix} 1 & 0\\ 1 & 1 \end{smallmatrix}\right),\ D = \left(\begin{smallmatrix} 1 & 1\\ 1 & 1 \end{smallmatrix}\right)$. Then $\{A, B, C, D\}$ is a basis of $M(2 \times 2)$.

  Proof. Let $M = \left(\begin{smallmatrix} a & b\\ c & d \end{smallmatrix}\right)$ an arbitrary 2×2 matrix. Consider the equation $M = \alpha_1 A + \alpha_2 B + \alpha_3 C + \alpha_4 D$. This leads to

$$\begin{aligned} \begin{pmatrix} a & b\\ c & d \end{pmatrix} &= \alpha_1 \begin{pmatrix} 1 & 0\\ 0 & 0 \end{pmatrix} + \alpha_2 \begin{pmatrix} 1 & 0\\ 1 & 0 \end{pmatrix} + \alpha_3 \begin{pmatrix} 1 & 0\\ 1 & 1 \end{pmatrix} + \alpha_4 \begin{pmatrix} 1 & 1\\ 1 & 1 \end{pmatrix}\\ &= \begin{pmatrix} \alpha_1 + \alpha_2 + \alpha_3 + \alpha_4 & \alpha_4\\ \alpha_2 + \alpha_3 + \alpha_4 & \alpha_3 + \alpha_4 \end{pmatrix}. \end{aligned}$$

  So we obtain the following set of equations for the α_j:

$$\left.\begin{aligned} \alpha_1 + \alpha_2 + \alpha_3 + \alpha_4 &= a\\ \alpha_4 &= b\\ \alpha_2 + \alpha_3 + \alpha_4 &= c\\ \alpha_3 + \alpha_4 &= d \end{aligned}\right\} \qquad \text{in matrix form:} \qquad \begin{pmatrix} 1 & 1 & 1 & 1\\ 0 & 0 & 0 & 1\\ 0 & 1 & 1 & 1\\ 0 & 0 & 1 & 1 \end{pmatrix} \begin{pmatrix} \alpha_1\\ \alpha_2\\ \alpha_3\\ \alpha_4 \end{pmatrix} = \begin{pmatrix} a\\ b\\ c\\ d \end{pmatrix}.$$

Now we apply Gauß-Jordan to the augmented matrix:

$$\left(\begin{array}{cccc|c} 1 & 1 & 1 & 1 & a \\ 0 & 0 & 0 & 1 & b \\ 0 & 1 & 1 & 1 & c \\ 0 & 0 & 1 & 1 & d \end{array}\right) \longrightarrow \left(\begin{array}{cccc|c} 1 & 1 & 1 & 1 & a \\ 0 & 1 & 1 & 1 & c \\ 0 & 0 & 1 & 1 & d \\ 0 & 0 & 0 & 1 & b \end{array}\right) \longrightarrow \left(\begin{array}{cccc|r} 1 & 1 & 1 & 0 & a-b \\ 0 & 1 & 1 & 0 & c-b \\ 0 & 0 & 1 & 0 & d-b \\ 0 & 0 & 0 & 1 & b \end{array}\right)$$
$$\longrightarrow \left(\begin{array}{cccc|r} 1 & 1 & 0 & 0 & a-d \\ 0 & 1 & 0 & 0 & c-d \\ 0 & 0 & 1 & 0 & d-b \\ 0 & 0 & 0 & 1 & b \end{array}\right) \longrightarrow \left(\begin{array}{cccc|r} 1 & 0 & 0 & 0 & a-c \\ 0 & 1 & 0 & 0 & c-d \\ 0 & 0 & 1 & 0 & d-b \\ 0 & 0 & 0 & 1 & b \end{array}\right).$$

So we see that there is exactly one solution for any given $M \in M(2 \times 2)$. Existence of the solution shows that the matrices A, B, C, D generate $M(2 \times 2)$ and uniqueness shows that they are linearly independent if we choose $M = 0$. ☐

Now we proceed with some theory. The next theorem is very important. It says that if V has a basis which consists of n vectors, then **every** basis consists of exactly n vectors.

Theorem 4.38. *Let V be a vector space and let $\{v_1, \dots, v_n\}$ and $\{w_1, \dots, w_n\}$ be bases of V. Then $n = m$.*

Definition 4.39. The number n (=number of elements of a basis) is called the *dimension of V*. It is denoted by $\dim V$.

Proof of Theorem 4.38. Suppose that $m > n$. We will show that then the vectors $w_1, \dots, w_m$ cannot be linearly independent, hence they cannot be a basis of V. Let us start. Since the vectors $v_1, \dots, v_n$ are a basis of V, every w_j can be written as a linear combination of them. So there exist numbers a_{ij} which

$$\begin{aligned} w_1 &= a_{11}v_1 + a_{12}v_2 + \dots + a_{1n}v_n \\ w_2 &= a_{21}v_1 + a_{22}v_2 + \dots + a_{2n}v_n \\ &\vdots \qquad\qquad\qquad\qquad \vdots \\ w_m &= a_{m1}v_1 + a_{m2}v_2 + \dots + a_{mn}v_n. \end{aligned} \tag{4.3}$$

Now we consider the equation

$$c_1 w_1 + \dots + c_m w_m = \mathbb{0}. \tag{4.4}$$

If the $w_1, \dots, w_m$ were linearly independent, then it should follow that all c_j are 0. We insert (4.3) into (4.4) and obtain

$$\begin{aligned} \mathbb{0} &= c_1(a_{11}v_1 + a_{12}v_2 + \dots + a_{1n}v_n) + c_2(a_{21}v_1 + a_{22}v_2 + \dots + a_{2n}v_n) \\ &\qquad + \dots + c_m(a_{m1}v_1 + a_{m2}v_2 + \dots + a_{mn}v_n) \\ &= (c_1a_{11} + c_2a_{21} + \dots + c_ma_{m1})v_1 + \dots + (c_1a_{1n} + c_2a_{2n} + \dots + c_ma_{mn})v_n. \end{aligned}$$

Since the vectors $v_1, \dots, v_n$ are linearly independent, the expressions in the parentheses must be equal to zero. So we find

$$\begin{array}{c} c_1a_{11} + c_2a_{12} + \cdots + \ c_ma_{1m} = 0 \\ c_1a_{21} + c_2a_{22} + \cdots + \ c_ma_{2m} = 0 \\ \vdots \qquad\qquad\qquad\qquad \vdots \\ c_1a_{1n} + c_2a_{2n} + \cdots + c_ma_{mm} = 0. \end{array} \tag{4.5}$$

This is a homogeneous system of n equations for the m unknowns $c_1, \dots, c_m$. Since $n < m$, we know that it has infinitely many solutions. So the system $\{w_1, \dots, w_m\}$ is not linearly independent and therefore it cannot be a basis of V. Therefore $m > n$ cannto be true and me must have that $n \geq m$.

If we assume that $n > m$, then the same argument as above, with the roles of the v_j and the w_j exchanged, leads to a contradiction and we must have $n \leq m$.
In summary we showed that both m$n \geq m$ and $n \leq m$ must be true. Therefore $m = n$. ☐

Corollary 4.40. *Let V be a vector space.*

- *If the vectors $v_1, \dots, v_k \in V$ are linearly independent, then $k \leq \dim V$.*
- *If the vectors $v_1, \dots, v_m \in V$ generate V, then $m \geq \dim V$.*

Theorem 4.41. *Let V be a vector space with basis $\{v_1, \dots, v_n\}$. Then every $x \in V$ can be written in unique way as linear combination of the vectors $v_1, \dots, v_n$.*

Proof. We have to show existence and uniqueness of numbers $c_1, \dots, c_n$ such that $w = c_1v_1 + \cdots c_nv_n$.

Existence is clear since the set $\{v_1, \dots, v_n\}$ is a set of generators of V (it is even a basis!).
Uniqueness can be shown as follows. Assume that there are numbers $c_1, \dots, c_n$ and $d_1, \dots, d_n$ such that $w = c_1v_1 + \cdots c_nv_n$ and $w = d_1v_1 + \cdots d_nv_n$. Then it follows that

$$\mathbb{0} = w - w = c_1v_1 + \cdots c_nv_n - (d_1v_1 + \cdots d_nv_n) = (c_1 - d_1)v_1 + \cdots (c_n - d_n)v_n.$$

Then all the parentheses have to be zero because the $v_1, \dots, v_n$ are linearly independent. Hence it follows that $c_1 = d_1, \dots, c_n = d_n$, which shows uniqueness. ☐

Definition 4.42. A vector space V is called *finitely generated* if has a basis consisting of finitely many vectors.

For example the spaces $\mathbb{R}^n$, $M(m \times n)$, P_n are finitely generated. The spaces P consisting of all polynomials is not finitely generated. (Can you prove this?[1])
Next we show that every finitely generated vector space has a basis.

---

[1] Assume that P is finitely generated and let $q_1, \dots, q_k$ be a system of generators of P. Note that the q_j are polynomials. We will denote their degrees by $m_j = \deg q_j$ and we set $M = \max\{m_1, \dots, m_k\}$. No matter which coefficients we choose, any linear combination of them will be a polynomial of degree at most M. However, there are elemnts in P which have higer degree, for example X^{m+1}. Therefore $q_1, \dots, q_k$ cannot generate all of P.
Another proof using the conept of dimension will be given in Example 4.49 (f).

Theorem 4.43. *Let V be a vector space and assume that there are vectors $w_1, \dots, w_m \in V$ such that $V = \operatorname{span}\{w_1, \dots, w_m\}$. Then V has a finite basis.*

Proof. Without restriction we may assume that all vectors w_j are different from $\mathbb{0}$. We start with the first vector. If $V = \operatorname{span}\{w_1\}$, then $\{w_1\}$ is a basis of V and $\dim V = 1$. Otherwise we set $V_1 = \operatorname{span}\{w_1\}$ and we note that $V_1 \neq V$. Now we check if $w_2 \in \operatorname{span}\{w_1\}$. If it is, we throw it out because in this case $\operatorname{span}\{w_1\} = \operatorname{span}\{w_1, w_2\}$ so we do not need w_2 to generate V. Next we look if $w_3 \in \operatorname{span}\{w_1\}$. If it is, we throw it out, etc. We proceed like this until we find a vector w_{i_2} in our list which does not belong to $\operatorname{span}\{w_1\}$. Such an i_2 must exist because otherwise we already had that $V_1 = V$. Then we set $V_2 = \operatorname{span}\{w_1, w_{i_2}\}$. If $V_2 = V$, then we are done. Otherwise, we proceed as before. We check if $w_{i_2+1} \in V_2$. If this is the case, then we can throw it out because $\operatorname{span}\{w_1, w_{i_2}\} = \operatorname{span}\{w_1, w_{i_2}, w_{i_2+1}\}$. Then we check w_{i_2+2}, etc., until we find a w_{i_3} such that $w_{i_3} \notin \operatorname{span}\{w_1, w_{i_2}\}$ and we set $V_3 = \operatorname{span}\{w_1, w_{i_2}, w_{i_3}\}$. If $V_3 = V$, then we are done. If not, then we repeat the process. Note that after at most m repetitions, this comes to an end. This shows that we can extract from the system of generators a basis $\{w_1, w_{i_2}, \dots, w_{i_k}\}$ of V. $\square$

The following theorem complements the preceding one.

Theorem 4.44. *Let V be a finitely generated vector space and assume that there are vectors $w_1, \dots, w_m \in V$ which are linearly independent. Then they can be complemented to a basis $\{w_1, \dots, w_m, v_{m+1}, \dots, v_n\}$ of V.*

Proof. Let $n = \dim V$. Note that it follows that $n \geq m$ because we have m linearly independent vectors in V. If $m = n$, then it follows that $w_1, \dots, w_m$ is already a basis of V and we are done.
If $m < n$, then $\operatorname{span}\{w_1, \dots, w_m\} \neq V$ and we choose an arbitrary vector $v_{m+1} \notin \operatorname{span}\{w_1, \dots, w_m\}$ and we define $V_{m+1} := \operatorname{span}\{w_1, \dots, w_m, v_{m+1}\}$. Then $\dim V_{m+1}\} = m + 1$. If $m + 1 = n$, then necessarily $V_{m+1} = V$ and we are done. If $m + 1 < n$, then we choose an arbitrary vector $v_{m+2} \in V \setminus V_{m+1}$ and we let $V_{m+2} := \operatorname{span}\{w_1, \dots, w_m, v_{m+1}, v_{m+2}\}$. If $m+2 = n$, then necessarily $V_{m+2} = V$ and we are done. If not, we repeat the step before. Note that after $n - m$ steps we have found a basis $\{w_1, \dots, w_m, v_{m+1}, \dots, v_n\}$ of V. $\square$

In summary, the two preceding theorems say the following:

- If we have a set of vectors $v_1, \dots v_m$ which generate the vector space V, then it is always possible to extract a subset which is a basis of V (we need to eliminate $m - n$ vectors).
- If we have a set of linearly independent vectors $v_1, \dots v_m$ in a finitely generated vector space V, then it is possible to find vectors $v_{m+1}, \dots, v_n$ such that $\{v_1, \dots, v_n\}$ is a basis of V (we need $\dim V - m$ such vectors).

Example 4.45.

- Let $A = \left(\begin{smallmatrix} 1 & 0 \\ 0 & 0 \end{smallmatrix}\right), B = \left(\begin{smallmatrix} 1 & 1 \\ 1 & 1 \end{smallmatrix}\right) \in M(2 \times 2)$ and suppoose that we want to complete them to a basis of $M(2 \times 2)$ (it is clear that A and B are linearly independent, so this makes sense). Since $\dim(M(2 \times 2)) = 4$, we know that we need 2 more matrices. We take any matrix $C \notin \operatorname{span}\{A, B\}$, for example $C = \left(\begin{smallmatrix} 0 & 1 \\ 0 & 0 \end{smallmatrix}\right)$. Finally we need a matrix $D \notin \operatorname{span}\{A, B, C\}$. We can take for example $D = \left(\begin{smallmatrix} 0 & 0 \\ 1 & 0 \end{smallmatrix}\right)$. So we find that A, B, C, D is a basis of $M(2 \times 2)$.

  Exercise. Check that $D \notin \operatorname{span}\{A, B, C\}$

- Suppose that we are given the vectors $\vec{v}_1 = \left(\begin{smallmatrix}1\\0\\1\end{smallmatrix}\right)$, $\vec{v}_2 = \left(\begin{smallmatrix}4\\0\\4\end{smallmatrix}\right)$, $\vec{v}_3 = \left(\begin{smallmatrix}1\\2\\3\end{smallmatrix}\right)$, $\vec{v}_4 = \left(\begin{smallmatrix}0\\2\\1\end{smallmatrix}\right)$, $\vec{v}_5 = \left(\begin{smallmatrix}0\\0\\2\end{smallmatrix}\right)$, $\vec{v}_6 = \left(\begin{smallmatrix}2\\1\\5\end{smallmatrix}\right)$ and we want to find a subset of them which form a basis of $\mathbb{R}^3$.

  First observe that we need 3 vectors for a basis since $\dim \mathbb{R}^3 = 3$. So we start with the first non-zero vector which is $\vec{v}_1$. We see that $\vec{v}_2 = 4\vec{v}_1$, so we discard it. We keep $\vec{v}_3$ since $\vec{v}_3 \notin \operatorname{span}\{\vec{v}_1\}$. Next, $\vec{v}_4 = \vec{v}_3 - \vec{v}_1$, so $\vec{v}_4 \in \operatorname{span}\{\vec{v}_1, \vec{v}_3\}$ and we discard it. A little calculation shows that $\vec{v}_5 \notin \operatorname{span}\{\vec{v}_1, \vec{v}_3\}$. Hence $\{\vec{v}_1, \vec{v}_3, \vec{v}_5\}$ is a basis of $\mathbb{R}^3$.

  Remark 4.46. We will present a more systematic way to solve exercises of this type in Theorem 5.27 and Remark 5.28.

If we have a vector space V and a subspace $W \subset V$, then we can ask ourselves what the relation between their dimensions is because W itself is a vector space.

Theorem 4.47. *Let V be a finitely generated vector space and let $W \subseteq V$ be a subspace. Then the following is true:*

(i) $\dim W \le \dim V$.

(ii) $\dim W = \dim V$ *if and only if* $W = V$.

Proof. (i) Let $\{w_1, \dots, w_k\}$ be a basis of W. Then these vectors are linearly independent in W, and therefore also in V. By Theorem 4.44 we can find vectors $v_{m+1}, \dots, v_n$ such that $\{w_1, \dots, w_m, v_{m+1}, \dots, v_n\}$ is a basis of V. Note that by construction $m \le n$. We also know that $m = \dim W$ and $n = \dim V$, hence the claim is proved.

(ii) If $V = W$, then clearly $\dim V = \dim W$. To show the converse, we now assume that $\dim V = \dim W$ and we have to show that $V = W$. As before let $\{w_1, \dots, w_k\}$ be a basis of W. Then these vectors are linearly independent in W, and therefore also in V. Since $\dim W = \dim V$, we know that these vectors form a basis of V. Therefore $V = \operatorname{span}\{w_1, \dots, w_m\} = W$.

□

Remark 4.48. Note that (i) is true even when V is not finitely generated. Note however that in general (ii) is **not** true for infinite dimensional vector spaces. In Example 4.49 (f) and (g) we will show that $\dim P = \dim C(\mathbb{R})$ in spite of $P \neq C(\mathbb{R})$. (Recall that P is the set of all polynomials and that $C(\mathbb{R})$ is the set of all continuous functions. So we have $P \subsetneq C(\mathbb{R})$.)

Now we give a few examples of dimensions of spaces.

Examples 4.49. (a) $\dim \mathbb{R}^n = n$, $\dim \mathbb{C}^n = n$.

(b) $\dim M(m \times n) = mn$. This follows because the set of all $m \times n$ matrices A_{ij} which have a 1 in the ith row and jth column and all other entries are equal to zero form a basis of $M(m \times n)$ and there are exactly mn such matrices.

(c) Let $M_{sym}(n \times n)$ be the set of all symmetric $n \times n$ matrices. Then $\dim M_{sym}(n \times n) = \frac{n(n+1)}{2}$. To see this, let A_{ij} be the $n \times n$ matrix with $a_{ij} = a_{ji} = 1$ and all other entries equal to 0. Observe that $A_{ij} = A_{ji}$. It is not hard to see that the set of all A_{ij} with $i \le j$ form a basis of

$M_{sym}(n \times n)$. The dimension of $M_{sym}(n \times n)$ is the number of different matrices of this type. So how many of them are there? If we fix $j = 1$, then only $i = 1$ is possible. If we fix $j = 2$, then $i = 1, 2$ is possible, etc. until for $j = n$ the allowed values for i are $1, 2, \dots, n$. In total we have $1 + 2 + \dots + n = \frac{n(n+1)}{2}$ possibilities. For example, in the case $n = 2$, the matrices are

$$A_{11} = \left(\begin{smallmatrix} 1 & 0 \\ 0 & 0 \end{smallmatrix}\right),\ A_{12} = \left(\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right),\ A_{12} = \left(\begin{smallmatrix} 0 & 0 \\ 0 & 1 \end{smallmatrix}\right).$$

In the case $n = 3$, the matrices are

$$A_{11} = \left(\begin{smallmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{smallmatrix}\right),\ A_{12} = \left(\begin{smallmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{smallmatrix}\right),\ A_{22} = \left(\begin{smallmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{smallmatrix}\right),\ A_{13} = \left(\begin{smallmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{smallmatrix}\right),\ A_{23} = \left(\begin{smallmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{smallmatrix}\right),\ A_{33} = \left(\begin{smallmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{smallmatrix}\right).$$

Exercise. Convince yourself that the A_{ij} form a basis of $M_{sym}(n \times n)$.

(d) Let $M_{asym}(n \times n)$ be the set of all antisymmetric $n \times n$ matrices. Then $\dim M_{asym}(n \times n) = \frac{n(n-1)}{2}$. To see this, for $i \neq j$ let A_{ij} be the $n \times n$ matrix with $a_{ij} = -a_{ji} = 1$ and all other entries equal to 0 form a basis of $M_{sym}(n \times n)$. It is not hard to see that the set of all A_{ij} with $i < j$ form a basis of $M_{asym}(n \times n)$. How many of these matrices are there? If we fix $j = 2$, then only $i = 1$ is possible. If we fix $j = 3$, then $i = 1, 2$ is possible, etc. until for $j = n$ the allowed values for i are $1, 2, \dots, n - 1$. In total we have $1 + 2 + \dots + (n - 1) = \frac{n(n-1)}{2}$ possibilities. For example, in the case $n = 2$, the matrices are

$$A_{12} = \left(\begin{smallmatrix} 0 & 1 \\ -1 & 0 \end{smallmatrix}\right).$$

In the case $n = 3$, the matrices are

$$A_{12} = \begin{pmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix},\ A_{13} = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ -1 & 0 & 0 \end{pmatrix},\ A_{23} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & -1 & 0 \end{pmatrix}.$$

Exercise. Convince yourself that the A_{ij} form a basis of $M_{asym}(n \times n)$.

Remark. Observe that $\dim M_{sym}(n \times n) + \dim M_{asym}(n \times n) = n^2 = \dim M(n \times n)$. This is no coincidence. Observe that every $n \times n$ matrix M can be written as

$$M = \tfrac{1}{2}(M + M^t) + \tfrac{1}{2}(M - M^t)$$

and that $\frac{1}{2}(M + M^t) \in M_{sym}(n \times n)$ and $\frac{1}{2}(M - M^t) \in M_{asym}(n \times n)$. Therefore $M(n \times n)$ is the *direct sum* of $M_{sym}(n \times n)$ and $M_{asym}(n \times n)$. We will talk about direct sums later.

(e) $\dim P_n = n + 1$ since $\{1, X, \dots, X^n\}$ is a basis of P_n and consists of $n + 1$ vectors.

(f) $\dim P = \infty$. Recall that P is the space of all polynomials.

Proof. We know that for every $n \in \mathbb{N}$, the space P_n is a subspace of P. Therefore for every $n \in \mathbb{N}$, we must have that $n + 1 = \dim P_n \leq \dim P$. This is possible only if $\dim P = \infty$. ☐

(g) $\dim C(\mathbb{R}) = \infty$. Recall that $C(\mathbb{R})$ is the space of all continuous functions.

Proof. Since P is a subspace of $C(\mathbb{R})$, it follows that $\dim P \leq \dim(C(\mathbb{R}))$, hence $\dim(C(\mathbb{R})) = \infty$. ☐

Now we use the concept of dimension to classify all subspaces of $\mathbb{R}^2$ and $\mathbb{R}^3$. We already know that for examples lines and planes which pass through the origin are subspaces of $\mathbb{R}^3$. Now we can show that there are no other proper subspaces.

Subspaces of $\mathbb{R}^2$. Let U be a subspace of $\mathbb{R}^2$. Then U must have a dimension. So we have the following cases:

- $\dim U = 0$. In this case $U = \{\vec{0}\}$ is the trivial subspace.
- $\dim U = 1$. Then U is of the form $U = \operatorname{span}\{\vec{v}_1\}$ with some vector $\vec{v}_1 \in \mathbb{R}^2 \setminus \{\vec{0}\}$. Then U is a line parallel to $\vec{v}_1$ passing through the origin.
- $\dim U = 2$. In this case $\dim U = \dim \mathbb{R}^2$. Hence it follows that $U = \mathbb{R}^2$ by Theorem 4.47 (ii).
- $\dim U \geq 3$ is not possible.

In conclusion, the only subspaces of $\mathbb{R}^2$ are $\{\vec{0}\}$, lines passing through the origin and $\mathbb{R}^2$ itself.

Subspaces of $\mathbb{R}^3$. Let U be a subspace of $\mathbb{R}^3$. Then U must have a dimension. So we have the following cases:

- $\dim U = 0$. In this case $U = \{\vec{0}\}$ is the trivial subspace.
- $\dim U = 1$. Then U is of the form $U = \operatorname{span}\{\vec{v}_1\}$ with some vector $\vec{v}_1 \in \mathbb{R}^3 \setminus \{\vec{0}\}$. Then U is a line parallel to $\vec{v}_1$ passing through the origin.
- $\dim U = 2$. Then U is of the form $U = \operatorname{span}\{\vec{v}_1, \vec{v}_2\}$ with linearly independent vectors $\vec{v}_1, \vec{v}_2 \in \mathbb{R}^3$. Hence U is a plane parallel to the vectors $\vec{v}_1$ and $\vec{v}_2$ which passes through the origin.
- $\dim U = 3$. In this case $\dim U = \dim \mathbb{R}^3$. Hence it follows that $U = \mathbb{R}^3$ by Theorem 4.47 (ii).
- $\dim U \geq 4$ is not possible.

In conclusion, the only subspaces of $\mathbb{R}^3$ are $\{\vec{0}\}$, lines passing through the origin, planes passing through the origin and $\mathbb{R}^3$ itself.

Chapter 5

Linear transformations and change of bases

In the first section of this chapter we will define linear maps between vector spaces and discuss their properties. These are fuctions which "behave well" with respect to the vector space structure. For example, $m \times n$ matrices can be viewed as linear maps from $\mathbb{R}^m$ to $\mathbb{R}^n$. We will prove the so-called *dimension formula* for linear maps. In Section 5.2 we will study the special case of matrices. One of the main results will be the dimension formula (5.7). In Section 5.4 we will see that, after choice of a basis, every linear map between finite dimensional vector spaces, can be represented as a matrix. This will allow us to carry over results on matrices to the case of linear transformations. In partiuclar the dimension formula (**??**) holds.

5.1 Linear maps

Definition 5.1. Let U, V be vector spaces. A function $A : U \to V$ is called a *linear map* (or *linear function* or *linear operator*) if for all $x, y \in U$ and $\lambda \in \mathbb{K}$ the following is true:

$$A(x + y) = Ax + Ay, \quad A(\lambda x) = \lambda Ax. \tag{5.1}$$

Remark. Note that very often one writes Ax instead of $A(x)$ when A is a linear function.

Remark 5.2. (i) Clearly, (5.1) is equivalent to

$$A(x + \lambda y) = Ax + \lambda Ay \qquad \text{for all } x, y \in U \text{ and } \lambda \in \mathbb{K}.$$

(ii) It follows immediately from the definition that

$$A(\lambda_1 v_1 + \cdots + \lambda_k v_k) = \lambda_1 A v_1 + \cdots + \lambda_k A v_k$$

for all $v_1, \dots, v_k \in V$ and $\lambda_1, \dots, \lambda_k \in \mathbb{K}$.

(iii) The condition (5.1) says that **a linear map respects the vector space structures of its domain and its target space**.

Examples 5.3 (Linear maps). (i) Every matrix $A \in M(m \times n)$ can be identified with a linear map $\mathbb{R}^n \to \mathbb{R}^m$.

(ii) Differentiation is a linear map, for example

(a) $T : C^1(\mathbb{R}) \to C(\mathbb{R}),\ Tf = f'$, where $C^1(\mathbb{R})$ is the space of continuously differentiable functions.

Proof. First of all note that $f' \in C(\mathbb{R})$ if $f \in C^1(\mathbb{R})$, so the map T is well-defined. Now want to see that it is linear. So we take $f, g \in C^1(\mathbb{R})$ and $\lambda \in \mathbb{R}$. We find

$$T(\lambda f + g) = (\lambda f + g)' = (\lambda f)' + g' = \lambda f' + g' = \lambda Tf + Tg. \qquad \square$$

(b) $T : P_n \to P_{n-1},\ Tf = f'$.

(iii) Integration is a linear map. For example:

$$I : C([0,1]) \to C([0,1]),\ f \mapsto If \quad \text{where } (If)(x) = \int_0^x f(t)\ \mathrm{d}t.$$

Proof. Clearly I is well-defined since the integral of a continuous function is again continuous. In order to show that I is linear, we fix $f, g \in C(\mathbb{R})$ and $\lambda \in \mathbb{R}$. We find for every $x \in \mathbb{R}$:

$$\begin{aligned}(I(\lambda f + g)(t) &= \int_0^x (\lambda f + g)(t)\ \mathrm{d}t = \int_0^x \lambda f(t) + g(t)\ \mathrm{d}t = \lambda \int_0^t f(t)\ \mathrm{d}t + \int_0^x g(t)\ \mathrm{d}t \\ &= \lambda(If)(x) + (Ig)(x).\end{aligned}$$

Since this is true for every x, it follows that $I(\lambda f + g) = \lambda(If) + (Ig)$. $\square$

Lemma 5.4. *If A is a linear map, then $A\mathbb{0} = \mathbb{0}$.*

Proof. $\mathbb{0} = A\mathbb{0} - A\mathbb{0} = A(\mathbb{0} - \mathbb{0}) = A\mathbb{0}$. $\square$

Definition 5.5. Let $A : U \to V$ be a linear map.

(i) A is called *injective* (or *one-to-one*) if

$$x, y \in U,\ x \neq y \quad \Longrightarrow \quad Ax \neq Ay.$$

(ii) A is called *surjective* if for all $v \in V$ exists at least one $x \in U$ such that $Ax = v$.

(iii) A is called *bijective* if it is injective and surjective.

(iv) The *kernel of A* (or *null space of A*, *espacio nulo de A*) is

$$\ker(A) := \{x \in U : Ax = 0\}.$$

Sometimes the notations $N(A)$ or N_A instead of $\ker(A)$ are used.

(v) The *image of A* (or *range of A*, *imagen de A*) is

$$\operatorname{Im}(A) := \{v \in V : y = Ax \text{ for some } y \in U\}.$$

Sometimes the notations $\operatorname{Rg}(A)$ or $\operatorname{R}(A)$ instead of $\operatorname{Im}(A)$ are used.

Remark 5.6. (i) Observe that $\ker(A)$ is a subset of U, $\operatorname{Im}(A)$ is a subset of V. In Proposition 5.9 we will show that they are even subspaces.

(ii) It follows immediately from the definition that A is surjective if and only if $\operatorname{Im}(A) = V$.

(iii) Clearly, A is injective if and only if for all $x, y \in U$ the following is true:

$$Ax = Ay \quad \Longrightarrow \quad x = y.$$

(iv) If A is a linear injective map, then its inverse $A^{-1} : \operatorname{Im}(A) \to U$ exists and is linear too.

The following lemma is very useful.

Lemma 5.7. *A linear map A is injective if and only if* $\ker(A) = \{\mathbb{0}\}$.

Proof. By Lemma 5.4, we always have $\mathbb{0} \in \ker(A)$. Assume that A is injective, then $\ker(A)$ cannot contain any other element, hence $\ker(A) = \{\mathbb{0}\}$.
Now assume that $\ker(A) = \{\mathbb{0}\}$ and let $x, y \in U$ with $Ax = Ay$. By Remark 5.6 it is sufficient to show that $x = y$. By assumption, $\mathbb{0} = Ax - Ay = A(x - y)$, hence $x - y \in \ker(A) = \{\mathbb{0}\}$. Therefore $x - y = \mathbb{0}$, which means that $x = y$. ☐

Examples 5.8. (i) Let $A \in M(m \times n)$ with $m < n$. Then A cannot be injective.

(ii) Let $T : C^1(\mathbb{R}) \to C(\mathbb{R}), Tf = f'$ the operator of differentiation from Example 5.3. Then it is easy to see that the kernel of T consists exactly of the constant functions and that T is surjective.

Proposition 5.9. *Let $A : U \to V$ be a linear map. Then*

(i) $\ker(A)$ *is a subspace of U.*

(ii) $\operatorname{Im}(A)$ *is a subspace of V.*

Proof. (i) Let $x, y \in \ker(A)$ and $\lambda \in \mathbb{K}$. Then

$$A(x + \lambda y) = Ax + \lambda Ay = 0 + \lambda 0 = 0,$$

hence $x + \lambda y \in \ker(A)$.

(ii) Let $v, w \in \operatorname{Im}(A)$ and $\lambda \in \mathbb{K}$. Then there exist Let $x, y \in U$ such that $Ax = v$ and $Ay = y$. Then $v + \lambda w = Ax + \lambda Ay = A(x + \lambda y) \in \operatorname{Im}(A)$. hence $v + \lambda w \in \operatorname{Im}(A)$. ☐

Since we now know that $\ker(A)$ and $\operatorname{Im}(A)$ are subspaces, the following definition makes sense.

Definition 5.10. Let $A : U \to V$ be a linear map. We define

$$\dim(\ker(A)) = \textit{nullity of } A, \qquad \dim(\operatorname{Im}(A)) = \textit{rank of } A.$$

Sometimes the notations $\nu(A) = \dim(\ker(A))$ and $\rho(A) = \dim(\operatorname{Im}(A))$ are used.

Let us pause for a moment and see an example.

Example. Let $T : P_4 \to P_4$ be defined by $Tp = p'$.

- $\boxed{\operatorname{Im}(T) = \{q \in P_3 : \deg q \leq 2\}}$ We know that differentiation lowers the degree of a polynomial by 1. Hence $\operatorname{Im}(T) \subseteq \{q \in P_3 : \deg q \leq 2\}$. On the other hand, we know that every polynomial of degree ≤ 2 is the derivative of a polynomial of degree ≤ 3. So the claim follows.
- $\boxed{\ker(T) = \{q \in P_3 : \deg q = 0\}}$ Recall that $\ker(T) = \{p \in P_3 : Tp = 0\}$. So the kernel of T are exactly those polynomials whose first derivative is 0. These are exactly the constant polynomials, i.e., the polynomials of degree 0.

Proposition 5.11. *Let U, V be $\mathbb{K}$-vector spaces, $A : U \to V$ a linear map. Let $x_1, \dots, x_k \in U$ and set $y_1 := Ax_1, \dots, y_k := Ax_k$. Then the following is true.*

(i) If the $x_1, \dots, x_k$ are linearly dependent, then $y_1, \dots, y_k$ are linearly dependent too.

(ii) If the $y_1, \dots, y_k$ are linearly independent, then $x_1, \dots, x_k$ are linearly independent too.

(iii) Suppose additionally that A invertible. Then $x_1, \dots, x_k$ are linearly independent if and only if $y_1, \dots, y_k$ are linearly independent.

Remark. In general the implication "If $x_1, \dots, x_k$ are linearly independent, then $y_1, \dots, y_k$ are linearly independent." is *false*. Can you give an example?

Proof of Proposition 5.11. (i) Assume that $x_1, \dots, x_k$ are linearly dependent. Then there exist $\lambda_1, \dots, \lambda_k \in \mathbb{K}$ such that $\lambda_1 x_1 + \dots + \lambda_k x_k = 0$ and at least one $\lambda_j \neq 0$. But then

$$\begin{aligned} \mathbb{0} = A\mathbb{0} &= A(\lambda_1 x_1 + \dots + \lambda_k x_k) = \lambda_1 Ax_1 + \dots + \lambda_k Ax_k \\ &= \lambda_1 y_1 + \dots + \lambda_k y_k, \end{aligned}$$

hence $y_1, \dots, y_k$ are linearly dependent.

(ii) follows directly from (i).

(iii) Suppose that $y_1, \dots, y_k$ are linearly independent. Then so are the $x_1, \dots, x_k$ by (i). Now suppose that $x_1, \dots, x_k$ are linearly independent. Note that A is invertible, so A^{-1} exists and in invertible too. Therefore we can apply (i) to A^{-1} in order to conclude that the system $y_1, \dots, y_k$ is linearly independent. (Note that $x_j = A^{-1} y_j$.) ☐

Exercise 5.12. Assume that $A : U \to V$ is an injective linear map and suppose that $\{u_1, \dots, u_\ell\}$ is a set of are linearly independent vectors in U. Show that $\{Au_1, \dots, Au_\ell\}$ is a set of are linearly independent vectors in V.

The following lemma is very useful and it is used in the proof of Theorem 5.14.

Lemma 5.13. *(i) If $T : U \to V$ is a bijective linear transformation, then* $\dim U = \dim V$*.*

(ii) If $T : U \to V$ is an injective linear transformation, then $\dim U = \dim \operatorname{Im}(T)$*.*

Proof. (i) Let $k = \dim U$ and $n = \dim V$. Choose a basis $\{w_1, \dots, w_k\}$ of U and set $v_1 := Tw_1, \dots, v_k := Tw_k$. Then the vectors $v_1, \dots, v_k$ are linearly independent in V by Proposition 5.11 (iii). Therefore $\dim V \geq k = \dim U$. Now choose a basis $z_1, \dots, z_n$ of V and set $u_1 := T^{-1}z_1, \dots, u_\ell := T^{-1}z_n$. Then, again by Proposition 5.11 (iii), the vectors $u_1, \dots, u_n$ are linearly independent in U and it follows that $\dim W \geq n = \dim V$.

In summary, both $\dim V \geq \dim U$ and $\dim U \geq \dim V$ must be true. This is possible only if $\dim V = \dim U$.

(ii) Assume that T is injective. Then the map $T : U \to \operatorname{Im} T$ is bijective (it is injective by assumption and surjective by construction). Therefore, by (i), it follows that $\dim U = \dim(\operatorname{Im} T)$. □

Theorem 5.14. *Let U, V be finite-dimensional $\mathbb{K}$-vector spaces and let $A : U \to V$ a linear map. Moreover, let $E : U \to U$, $F : V \to V$ be linear bijective maps. Then the following is true:*

(i) $\operatorname{Im}(A) = \operatorname{Im}(AE)$, *in particular* $\dim(\operatorname{Im}(A)) = \dim(\operatorname{Im}(AE))$.

(ii) $\ker(AE) = E^{-1}(\ker(A))$ *and* $\dim(\ker(A)) = \dim(\ker(AE))$.

(iii) $\ker(A) = \ker(FA)$, *in particular* $\dim(\ker(A)) = \dim(\ker(FA))$.

(iv) $\operatorname{Im}(FA) = F(\operatorname{Im}(A))$ *and* $\dim(\operatorname{Im}(A)) = \dim(\operatorname{Im}(FA))$.

In summary we have

$$\boxed{\begin{array}{ll} \ker(FA) = \ker(A), & \ker(AE) = E^{-1}(\ker(A)), \\ \operatorname{Im}(FA) = F(\operatorname{Im}(A)), & \operatorname{Im}(AE) = \operatorname{Im}(A). \end{array}} \tag{5.2}$$

and

$$\boxed{\begin{array}{l} \dim\ker(A) = \dim\ker(FA) = \dim\ker(AE) = \dim\ker(FAE), \\ \dim\operatorname{Im}(A) = \dim\operatorname{Im}(FA) = \dim\operatorname{Im}(AE) = \dim\operatorname{Im}(FAE). \end{array}} \tag{5.3}$$

Remark 5.15. In general, $\ker(A) = \ker(AE)$ and $\ker(A) = \ker(FA)$ is false. Take for example $U = V = \mathbb{R}^2$, $A = \left(\begin{smallmatrix} 1 & 0 \\ 0 & 0 \end{smallmatrix}\right)$ and $E = F = \left(\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right)$. Then clearly the hypotheses of the theorem are satisfied and

$$\ker(A) = \operatorname{gen}\left\{\begin{pmatrix} 0 \\ 1 \end{pmatrix}\right\}, \qquad \operatorname{Im}(A) = \operatorname{gen}\left\{\begin{pmatrix} 1 \\ 0 \end{pmatrix}\right\},$$

but

$$\ker(AE) = \operatorname{gen}\left\{\begin{pmatrix} 1 \\ 0 \end{pmatrix}\right\}, \qquad \operatorname{Im}(FA) = \operatorname{gen}\left\{\begin{pmatrix} 0 \\ 1 \end{pmatrix}\right\}.$$

Remark 5.16. The theorem is also true for infinite dimensional vector spaces, but the proofs of (ii) and (iv) must be changed a little bit.

Proof of Theorem 5.14. (i) Let $v \in V$. If $v \in \operatorname{Im}(A)$, then there exists $x \in U$ such that $Ax = v$. Set $y = E^{-1}x$. Then $v = Ax = AEE^{-1}x = AEy \in \operatorname{Im}(AE)$. On the other hand, if $v \in \operatorname{Im}(AE)$, then there exists $y \in U$ such that $AEy = v$. Set $x = E$. Then $v = AEy = Ax \in \operatorname{Im}(A)$.

(ii) To show $\ker(AE) = E^{-1}\ker(A)$ observe that

$$\ker(AE) = \{x \in U : Ex \in \ker(A)\} = \{E^{-1}u : u \in \ker(A)\} = E^{-1}(\ker(A)).$$

The claim on the dimensions follows from Lemma 5.13 with E^{-1} as T and $\ker(A)$ as W.

(iii) Let $x \in U$. Then $x \in \ker(FA)$ if and only if $FAx = 0$. Since F is injective, we know that $\ker(F) = \{0\}$, hence it follows that $Ax = 0$. But this is equivalent to $x \in \ker(A)$.

(iv) To show $\operatorname{Im}(FA) = F\operatorname{Im}(A)$ observe that

$$\begin{aligned}\operatorname{Im}(FA) &= \{y \in V : y = FAx \text{ for some } x \in U\} = \{Fv : v \in \operatorname{Im}(A)\} \\ &= F(\operatorname{Im}(A)),\end{aligned}$$

The claim on the dimensions follows from Lemma 5.13 with F as T and $\operatorname{Im}(A)$ as W. ☐

5.2 Matrices as linear maps

Let $\in M(m \times n)$. We already know that we can view A as a linear map from $\mathbb{R}^n$ to $\mathbb{R}^m$. Hence $\ker(A)$ and $\operatorname{Im}(A)$ and the terms *injectivity* and *surjectivity* are defined.

If we view the matrix A at the same time as a linear system of equations, then we obtain the following.

Remark 5.17.

(i) $\ker(A) =$ all solutions of the homogeneous system $A\vec{x} = \vec{0}$.

(ii) A is injective
$\iff$ $\ker(A) = \{\vec{0}\}$
$\iff$ the homogenous system $A\vec{x} = \vec{0}$ has only the trivial solution $\vec{x} = \vec{0}$.

(iii) $\operatorname{Im}(A) =$ all vectors $\vec{b}$ such that the system $A\vec{x} = \vec{b}$ has a solution.

(iv) A is surjective
$\iff$ $\operatorname{Im}(A) = \mathbb{R}^m$
$\iff$ for every $\vec{b} \in \mathbb{R}^m$, the system $A\vec{x} = \vec{b}$ has at least one solution.

Definition 5.18. Let $A \in M(m \times n)$ and let $\vec{c}_1, \dots, \vec{c}_n$ be the columns of A and $\vec{r}_1, \dots, \vec{r}_m$ be the rows of A. We define

(i) $C_A := \operatorname{gen}\{\vec{c}_1, \dots, \vec{c}_m\} =:$ *column space of* A.

(ii) $R_A := \operatorname{gen}\{\vec{r}_1, \dots, \vec{r}_n\} =:$ *row space of* A,

Observe that $\vec{c}_1, \dots, \vec{c}_n \in \mathbb{R}^m$ and $\vec{r}_1, \dots, \vec{r}_m \in \mathbb{R}^n$.
It follows immediately from the definition above that

$$R_A = C_{A^t} \quad \text{and} \quad C_A = R_{A^t}. \tag{5.4}$$

Proposition 5.19. $C_A = \operatorname{Im}(A)$, $R_A = \operatorname{Im}(A^t)$.

Proof. Let $\vec{y} \in \mathbb{R}^m$. Then:

$$\begin{aligned} \vec{y} \in \operatorname{Im}(A) \quad & \iff \text{exists } \vec{x} \in \mathbb{R}^n \text{ such that } \vec{y} = A\vec{x} = (\vec{c}_1 | \dots | \vec{c}_n) \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \\ & \qquad\qquad\qquad\qquad\qquad\qquad = x_1\vec{c}_1 + \dots x_n\vec{c}_n \\ & \iff \vec{y} \in \operatorname{gen}\{\vec{c}_1, \dots, \vec{c}_n\} = C_A. \end{aligned}$$

This shows $C_A = \operatorname{Im}(A)$. From this it follows that $R_A = C_{A^t} = \operatorname{Im}(A^t)$. □

The next theorem follows easily from the general theory in Section 5.1. We will give another proof at the end of this section.

Proposition 5.20. *Let $A \in M(m \times n)$, $E \in M(n \times n)$, $F \in M(m \times m)$ and assume that E and F are invertible. Then*

(i) $C_A = C_{AE}$.

(ii) $R_A = R_{FA}$.

Proof. (i) Note that $C_A = \operatorname{Im}(A) = \operatorname{Im}(AE) = C_{AE}$, where in the first and third equality we used Proposition 5.19, and in the second equality we used Theorem 5.14.

(ii) Recall that, if F is invertible, then F^t is invertible too. With (5.4) and what we already proved in (i), we obtain $R_{FA} = C_{(FA)^t} = C_{A^tF^t} = C_{A^t} = R_A$. □

This proposition implies immediately the following proposition.

Proposition 5.21. *Let $A, B \in M(m \times n)$.*

(i) *If A and B are row equivalent, then*

$$\dim(\ker(A)) = \dim(\ker(B)), \quad \dim(\operatorname{Im}(A)) = \dim(\operatorname{Im}(B)), \quad \operatorname{Im}(A^t) = \operatorname{Im}(B^t), \quad R_A = R_B.$$

(ii) *If A and B are column equivalent, then*

$$\dim(\ker(A)) = \dim(\ker(B)), \quad \dim(\operatorname{Im}(A)) = \dim(\operatorname{Im}(B)), \quad \operatorname{Im}(A) = \operatorname{Im}(B), \quad C_A = C_B.$$

Proof. We will only prove (i). The claim (ii) can be proved similar (or can be deduced easily from (i) by applying (i) to the transposed matrices). If A and B are row equivalent, then there are elementary matrices $F_1, \dots, F_k \in M(m \times m)$ such that $A = F_1 \dots F_k B$. Note that all F_j are invertible. Let $F := F_1 \dots F_k$. Then F is invertible and $A = FB$. Hence all the claims in (i) follow from Theorem 5.14 and Proposition 5.20. ☐

The proposition above is very useful to calculate the kernel of a matrix A: Let A' be the reduced row-echelon form of A. Then the proposition can be applied to A and A' (for B), and we find that $\ker(A) = \ker(A')$. We know this actually since the first chapter of this course. This says nothing else then the solutions of a homogenous system do not change if we apply row transformations.
We will calculate the kernel and range of a matrix later in Examples 5.29 and 5.30.
Now we will prove to technical lemmas.

Lemma 5.22. *Let $A \in M(m \times n)$. Then there exist elementary matrices $E_1, \dots, E_k \in M(n \times n)$ and $F_1, \dots, F_\ell \in M(m \times m)$ such that*

$$F_1 \cdots F_\ell A E_1 \cdots E_k = A''$$

where A'' is of the form

$$A'' = \begin{array}{c} \begin{array}{cc} \overbrace{\qquad\quad}^{r} & \overbrace{\qquad}^{n-r} \end{array} \\ \left(\begin{array}{c:c} \begin{matrix} 1 & & \\ & \ddots & \\ & & 1 \end{matrix} & 0 \\ \hdashline \\ 0 & 0 \end{array}\right) \end{array} \begin{array}{l} \left.\vphantom{\begin{matrix}1\\ \ddots\\ 1\end{matrix}}\right\} r \\ \left.\vphantom{\begin{matrix}0\\0\end{matrix}}\right\} m-r \end{array} \tag{5.5}$$

Proof. Let A' be the reduced row-echelon form of A. Then there exist $F_1, \dots, F_\ell \in M(m \times m)$ such that $F_1 \cdots F_\ell A = A'$ and A' is of the form

$$A' = \left(\begin{array}{ccccccccccc} 1 & * & \cdots & * & 0 & * & \cdots & * & 0 & * & \cdots \\ & & & & 1 & * & \cdots & * & 0 & * & \cdots \\ & & & & & & & & 1 & * & \cdots \\ & & & & & & & & & & \\ & & & & & & & & & & \end{array}\right). \tag{5.6}$$

Now clearly we can find "allowed" column transformations such that A' is transformed into the form A''. If we observe that applying row transformations is equivalent to multiply A' from the right by elementary matrices. ☐

Lemma 5.23. *Let A'' be as in (5.5). Then*

(i) $\dim(\ker(A)) = m - r =$ *number of zero rows of A'',*

(ii) $\dim(\operatorname{Im}(A)) = r =$ *number of pivots A'',*

(iii) $\dim(C_{A''}) = \dim(R_{A''}) = r$.

Proof. All assertions are clear if we note that

$$\ker(A'') = \text{gen}\{\vec{e}_{r+1}, \dots, \vec{e}_n\}, \quad \text{Im}(A'') = \text{gen}\{\vec{e}_1, \dots, \vec{e}_r\},$$

where the $\vec{e}_j$ are the standard unit vectors (that is, their jth component is 1 and all other components are 0). ☐

Proposition 5.24. *Let $A \in M(m \times n)$ and let A' be its reduced row-echelon form. Then*

$$\dim(\text{Im}(A)) = \text{ number of pivots of } A'.$$

Proof. Let $F_1, \dots, F_\ell$, $E_1, \dots, E_k$ and A'' be as in (5.22) and set $F := F_1 \cdots F_\ell$ and $E := E_1 \cdots E_k$. It follows that $A' = FA$ and $A'' = FAE$. Clearly, the number of pivots of A' and A'' coincide. Therefore, with the help of Theorem 5.14 we obtain

$$\begin{aligned} \dim(\text{Im}(A)) &= \dim(\text{Im}(FAE)) \\ &= \text{ number of pivots of } A'' \\ &= \text{ number of pivots of } A'. \end{aligned}$$ ☐

Proposition 5.25. *Let $A \in M(m \times n)$. Then*

$$\dim(\text{Im}(A)) = \dim C_A = \dim R_A.$$

That means: (rank of row space) = (rank of column space).

Proof. Since $C_A = \text{Im}(A)$ by Proposition 5.19, the first equality is clear.
Now let $F_1, \dots, F_\ell$, $E_1, \dots, E_k$ and A', A'' be as in Lemma 5.22 and set $F := F_1 \cdots F_\ell$ and $E := E_1 \cdots E_k$. Then

$$\begin{aligned} \dim(R_A) &= \dim(R_{FAE}) = \dim(R_{A''}) = r = \dim(C_{A''}) = \dim(C_{FAE}) \\ &= \dim(C_A). \end{aligned}$$ ☐

As an immediate consequence we obtain

Theorem 5.26. *Let $A \in M(m \times n)$. Then*

$$\boxed{\dim(\ker(A)) + \dim(\text{Im}(A)) = n.} \tag{5.7}$$

Proof. With the notation a above, we obtain

$$\begin{aligned} \dim(\ker(A)) &= \dim(\ker(A'')) = n - r, \\ \dim(\text{Im}(A)) &= \dim(\text{Im}(A'')) = r \end{aligned}$$

and the desired formula follows. ☐

We will give a different proof of a more general version in theorem in Theorem 5.34.
For the calculation of a basis of Im(A), the following theorem is useful.

Theorem 5.27. *Let $A \in M(m \times n)$ and let A' be its reduced row-echelon form with columns $\vec{c}_1, \dots, \vec{c}_n$ and $\vec{c}_1{}', \dots, \vec{c}_n{}'$ respectively. Assume that the pivot columns of A' are the columns $j_1 < \dots < j_k$. Then $\dim(\operatorname{Im}(A)) = k$ and a basis of $\operatorname{Im}(A)$ is given by the columns $\vec{c}_{j_1}, \dots, \vec{c}_{j_k}$ of A.*

Proof. Let E be an invertible matrix such that $A = EA'$. By assumption on the pivot columns of A', we know that $\dim(\operatorname{Im}(A')) = k$ and that a basis of $\operatorname{Im}(A')$ is given by the columns $\vec{c}_{j_1}{}', \dots, \vec{c}_{j_k}{}'$. By Theorem 5.14, it follows that $\dim(\operatorname{Im}(A)) = \dim(\operatorname{Im}(A')) = k$. Now observe that by definition of E we have that $E\vec{c}_\ell{}' = \vec{c}_\ell$ for every $\ell = 1, \dots, n$ and in particular this is true for the pivot columns of A'. Moreover, since E in invertible and the vectors $\vec{c}_{j_1}{}', \dots, \vec{c}_{j_k}{}'$ are linearly independent, it follows from Theorem 5.11 that the vectors $\vec{c}_{j_1}, \dots, \vec{c}_{j_k}$ are linearly independent. Clearly they belong to $\operatorname{Im}(A)$, so we have $\operatorname{gen}\{\vec{c}_{j_1}, \dots, \vec{c}_{j_k}\} \subseteq \operatorname{Im}(A)$. Since both spaces have the same dimension, they must be equal. ☐

Remark 5.28. The theorem above can be used to determine a basis of a subspace given in the form $U = \operatorname{gen}\{\vec{v}_1, \dots, \vec{v}_k\} \subseteq \mathbb{R}^m$ as follows: Define the matrix $A = (\vec{v}_1 | \dots | \vec{v}_k)$. Then clearly $U = \operatorname{Im} A$ and we can apply Theorem 5.27 to find a basis of U.

Example 5.29. Find $\ker(A)$, $\operatorname{Im}(A)$, $\dim(\ker(A))$, $\dim(\operatorname{Im}(A))$ and R_A for

$$A = \begin{pmatrix} 1 & 1 & 5 & 1 \\ 3 & 2 & 13 & 1 \\ 0 & 2 & 4 & -1 \\ 4 & 5 & 22 & 1 \end{pmatrix}.$$

Solution. First, let us row-reduce the matrix A:

$$A = \begin{pmatrix} 1 & 1 & 5 & 1 \\ 3 & 2 & 13 & 1 \\ 0 & 2 & 4 & -1 \\ 4 & 5 & 22 & 1 \end{pmatrix} \xrightarrow{\substack{Q_{21}(-1) \\ Q_{41}(-4)}} \begin{pmatrix} 1 & 1 & 5 & 1 \\ 0 & -1 & -2 & -2 \\ 0 & 2 & 4 & -1 \\ 0 & 1 & 2 & -3 \end{pmatrix} \xrightarrow{\substack{Q_{32}(2) \\ Q_{42}(1)}} \begin{pmatrix} 1 & 1 & 5 & 1 \\ 0 & -1 & -2 & -2 \\ 0 & 0 & 0 & -5 \\ 0 & 0 & 0 & -5 \end{pmatrix}$$

$$\xrightarrow{\substack{S_2(-1) \\ Q_{43}(-1)}} \begin{pmatrix} 1 & 1 & 5 & 1 \\ 0 & 1 & 2 & 2 \\ 0 & 0 & 0 & 5 \\ 0 & 0 & 0 & 0 \end{pmatrix} \xrightarrow{\substack{S_4(1/5) \\ Q_{12}(-1)}} \begin{pmatrix} 1 & 0 & 3 & -1 \\ 0 & 1 & 2 & 2 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix} \xrightarrow{\substack{Q_{14}(1) \\ Q_{24}(-2)}} \begin{pmatrix} 1 & 0 & 3 & 0 \\ 0 & 1 & 2 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix} =: A'.$$

Now it follows immediately that $\dim R_A = \dim C_A = 3$ and

$$\begin{aligned} \dim(\operatorname{Im}(A)) &= \#\text{non-zero rows of } A' = 3, \\ \dim(\ker(A)) &= 4 - \dim(\operatorname{Im}(A)) = 1 \end{aligned}$$

(or: $\dim(\operatorname{Im}(A)) = \#\text{pivot columns } A' = 3$, or: $\dim(\operatorname{Im}(A)) = \dim(R_A) = 3$ or: $\dim(\ker(A)) = \#\text{non-pivot columns } A' = 1$).

Kernel of A: We know that $\ker(A) = \ker(A')$ by Theorem 5.14 or Proposition 5.21. From the explicit form of A', it is clear that $A\vec{x} = 0$ if and only if $x_4 = 0$, x_3 arbitrary, $x_2 = -2x_3$ and

$x_1 = -3x_3$. Therefore

$$\ker(A) = \ker(A') = \left\{ \begin{pmatrix} -3x_3 \\ -2x_3 \\ x_3 \\ 0 \end{pmatrix} : x_3 \in \mathbb{R} \right\} = \operatorname{gen} \left\{ \begin{pmatrix} -3 \\ -2 \\ 1 \\ 0 \end{pmatrix} \right\}.$$

Image of A: The pivot columns of A' are the columns 1, 2 and 4. Therefore, by Theorem 5.27 a basis of $\operatorname{Im}(A)$ are the columns 1, 2 and 4 of A:

$$\operatorname{Im}(A) = \operatorname{gen} \left\{ \begin{pmatrix} 1 \\ 3 \\ 0 \\ 4 \end{pmatrix}, \begin{pmatrix} 1 \\ 2 \\ 2 \\ 5 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \\ -1 \\ 1 \end{pmatrix} \right\}.$$

Example 5.30. Find a basis of $\operatorname{gen}\{p_1, p_2, p_3, p_4\} \subseteq P_3$ and its dimension for

$$\begin{aligned} p_1 &= x^3 - x^2 + 2x + 2, & p_2 &= x^3 + 2x^2 + 8x + 13, \\ p_3 &= 3x^3 - 6x^2 - 5, & p_3 &= 5x^3 + 4x^2 + 26x - 9. \end{aligned}$$

Solution. First we identify P_3 with $\mathbb{R}^4$ by $ax^3 + bx^2 + cx + d \mathrel{\widehat{=}} (a, b, c, d)^t$. The polynomials p_1, p_2, p_3, p_4 correspond to the vectors

$$\vec{v}_1 = \begin{pmatrix} 1 \\ -1 \\ 2 \\ 2 \end{pmatrix}, \quad \vec{v}_2 = \begin{pmatrix} 1 \\ 2 \\ 8 \\ 13 \end{pmatrix}, \quad \vec{v}_3 = \begin{pmatrix} 3 \\ -6 \\ 0 \\ -5 \end{pmatrix}, \quad \vec{v}_4 = \begin{pmatrix} 5 \\ 4 \\ 26 \\ -9 \end{pmatrix}.$$

Now we use Remark 5.28 to find a basis of $\operatorname{gen}\{v_1, v_2, v_3, v_4\}$. To this end we consider the A whose columns are the vectors $\vec{v}_1, \dots, \vec{v}_4$:

$$A = \begin{pmatrix} 1 & 1 & 3 & 5 \\ -1 & 2 & -6 & 4 \\ 2 & 8 & 0 & 26 \\ 2 & 13 & -5 & -9 \end{pmatrix}.$$

Clearly, $\operatorname{gen}\{v_1, v_2, v_3, v_4\} = \operatorname{Im}(A)$, so it suffices to find a basis of $\operatorname{Im}(A)$. Applying row transformation to A, we obtain

$$A = \begin{pmatrix} 1 & 1 & 3 & 5 \\ -1 & 2 & -6 & 4 \\ 2 & 8 & 0 & 26 \\ 2 & 13 & -5 & -9 \end{pmatrix} \longrightarrow \quad \cdots \quad \longrightarrow \begin{pmatrix} 1 & 0 & 4 & 5 \\ 0 & 1 & 2 & 3 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} = A'.$$

The pivot columns of A' are the first and the second column, hence by Theorem 5.27, a basis of $\operatorname{Im}(A)$ are its first and second columns, i.e. the vectors $\vec{v}_1$ and $\vec{v}_2$.
It follows that $\{p_1, p_2\}$ is a basis of $\operatorname{gen}\{p_1, p_2, p_3, p_4\} \subseteq P_3$ and consequently $\dim(\operatorname{gen}\{p_1, p_2, p_3, p_4\}) = 2$.

Remark 5.31. Let us use the abbreviation $\pi = \operatorname{gen}\{p_1, p_2, p_3, p_4\}$. The calculation above actually shows that any two vectors of p_1, p_2, p_3, p_4 form a basis of π. To see this, observe that clearly any two of them are linearly independent, hence the dimension of their generated space is 2. On the other hand, this generated space is a subspace of π which has the same dimension 2. Therefore they must be equal.

Remark 5.32. If we wanted to complete p_1, p_2 to a basis of P_3, we have (at least) the two following options:

(i) Find two linearly independent vectors which are orthogonal to $\vec{v}_1$ an $\vec{v}_2$. This leads to a homogenous system of two equations for four unknowns, namely

$$\begin{aligned} x_1 - x_2 + 2x_3 + 2x_4 &= 0, \\ x_1 + 2x_2 - 6x_3 + 4x_4 &= 0 \end{aligned}$$

or, in matrix notation, $P\vec{x} = 0$ where P is the 2×4 matrix whose rows are p_1 and p_2. Since clearly $\operatorname{Im}(P) \subseteq \mathbb{R}^2$, it follows that $\dim(\operatorname{Im}(P)) \leq 2$ and therefore $\dim(\ker(P)) \geq 4 - 2 = 2$.

(ii) Another way to find $q_3, q_4 \in P_3$ such that p_1, p_2, q_3, q_4 forms a basis of P_3 is to use the reduction process that was employed to find A'. Assume that E is an invertible matrix such that $A = EA'$. Such an E can be found by keeping track of the row operations that transform A into A'. Let $\vec{e}_j$ be the standard unit vectors of $\mathbb{R}^4$. Then we already know that $\vec{v}_1 = E\vec{e}_1$ and $\vec{v}_2 = E\vec{e}_2$. If we set $\vec{w}_3 = E\vec{e}_3$ and $\vec{w}_4 = E\vec{e}_4$, then $\vec{v}_1, \vec{v}_2, \vec{w}_3, \vec{w}_4$ form a basis of $\mathbb{R}^4$. This is because $\vec{e}_1, \dots, \vec{e}_4$ are linearly independent and E is injective. Hence $E\vec{e}_1, \dots, E\vec{e}_4$ are linearly independent too (by Proposition 5.11).

Sometimes useful is the following theorem.

Theorem 5.33. *Let $A \in M(m \times n)$. Then* $\ker(A) = (R_A)^\perp$.

Proof. Let $\vec{r}_1, \dots, \vec{r}_n$ be the rows of A. Since $R_A = \operatorname{gen}\{\vec{r}_1, \dots, \vec{r}_n\}$, it suffices to show that $\vec{x} \in \ker(A)$ if and only if $\vec{x} \perp \vec{r}_j$ for all $j = 1, \dots, m$.
By definition $\vec{x} \in \ker(A)$ if and only if

$$\vec{0} = A\vec{x} = \left(\begin{array}{c} \vec{r}_1 \\ \hline \vdots \\ \hline \vec{r}_m \end{array}\right) \begin{pmatrix} x_1 \\ \vdots \\ x_m \end{pmatrix} = \begin{pmatrix} \langle \vec{r}_1 \,, \vec{x} \rangle \\ \vdots \\ \langle \vec{r}_m \,, \vec{x} \rangle \end{pmatrix}$$

This is the case if and only if $\langle \vec{r}_j \,, \vec{x} \rangle$ for all $j = 1, \dots, m$, that is, if and only if $\vec{x} \perp \vec{\vec{r}}_j$ for all $j = 1, \dots, m$. ($\langle \cdot \,, \cdot \rangle$ denotes the inner product on $\mathbb{R}^n$.) ☐

Alternative proof of Theorem 5.33. Observe that $R_A = C_{A^t} = \operatorname{Im}(A^t)$. So we have to show that $\ker(A) = (\operatorname{Im}(A^t))^\perp$. Recall that $\langle Ax \,, y \rangle = \langle x \,, A^t y \rangle$. Therefore

$$\begin{aligned} x \in \ker(A) &\iff Ax = 0 \iff Ax \perp \mathbb{R}^m \\ &\iff \langle Ax \,, y \rangle = 0 \text{ for all } y \in \mathbb{R}^m \\ &\iff \langle x \,, A^t y \rangle = 0 \text{ for all } y \in \mathbb{R}^m \iff x \in (\operatorname{Im}(A))^t. \end{aligned}$$ ☐

Finally we want to give an alternative (coordinate free!) proof of Theorem 5.26.

Theorem 5.34. *Let U, V be vector spaces, $T : U \to V$ a linear map and set $n = \dim U$. Then*

$$\boxed{\dim(\ker(T)) + \dim(\operatorname{Im}(T)) = n.} \tag{5.7}$$

Proof. Let $k = \dim(\ker(T))$ and let $\{u_1, \dots, u_k\}$ be a basis of $\ker(T)$. We complete it to a basis $\{u_1, \dots, u_k, w_{k+1}, \dots, w_n\}$ to a basis of U. We set $W := \operatorname{span}\{w_{k+1}, \dots, w_n\}$ and we consider $\widetilde{T} = T|_W$ the restriction of T to W.
$\widetilde{T}$ *is injective* because $\widetilde{T}x = 0$ for some $x \in W$ if and only if $x \in \ker(T) \cap W = \{0\}$, hence $\ker(\widetilde{T}) = \{0\}$. Therefore we know that $\widetilde{T}w_{k+1}, \dots, \widetilde{T}w_n$ are linearly independent by Exercise 5.12. On the other hand, $\operatorname{Im}(\widetilde{T}) = \operatorname{span}\{\widetilde{T}w_{k+1}, \dots, \widetilde{T}w_n\}$, therefore $\dim(\operatorname{Im}\widetilde{T}) = n - k$. It follows that

$$n = \dim(\ker T) + \dim(\operatorname{Im}\widetilde{T}). \tag{5.8}$$

To complete the proof, it suffices to show that $\operatorname{Im}\widetilde{T} = \operatorname{Im}T$. First note that $\operatorname{Im}\widetilde{T} \subseteq \operatorname{Im}T$ since $\widetilde{T}$ is a restriction of T. On the other hand, let $v \in \operatorname{Im}(T)$. Then there exists an $x \in U$ with $Tx = v$. Now we write x as a linear combination of the basis $\{u_1, \dots, u_k, w_{k+1}, \dots, w_n\}$: $x = \alpha_1 u_1 + \dots + \alpha_k u_k + \alpha_{k+1} w_{k+1} + \dots + \alpha_n w_n$. Therefore

$$\begin{aligned} v = Tx &= T(\alpha_1 u_1 + \dots + \alpha_k u_k + \alpha_{k+1} w_{k+1} + \dots + \alpha_n w_n) \\ &= T(\alpha_1 u_1 + \dots + \alpha_k u_k) + T(\alpha_{k+1} w_{k+1} + \dots + \alpha_n w_n) \\ &= T(\alpha_{k+1} w_{k+1} + \dots + \alpha_n w_n) \\ &= \widetilde{T}(\alpha_{k+1} w_{k+1} + \dots + \alpha_n w_n) \in \operatorname{Im}(\widetilde{T}). \end{aligned}$$

Here we used that $\alpha_1 u_1 + \dots + \alpha_k u_k \in \ker(T)$ so that $T(\alpha_1 u_1 + \dots + \alpha_k u_k) = \mathbb{O}$ and that $\alpha_{k+1} w_{k+1} + \dots + \alpha_n w_n \in W$, therefore $T(\alpha_{k+1} w_{k+1} + \dots + \alpha_n w_n) = \widetilde{T}(\alpha_{k+1} w_{k+1} + \dots + \alpha_n w_n)$. So we have showed that $\operatorname{Im}\widetilde{T} = \operatorname{Im}T$, in particular their dimensions are equal and the claim follows from (5.8). ☐

5.3 Change of bases

Usually we represent vectors in $\mathbb{R}^n$ as column of numbers, for example

$$\vec{v} = \begin{pmatrix} 3 \\ 2 \\ -1 \end{pmatrix}, \qquad \text{or more generally,} \qquad \vec{w} = \begin{pmatrix} x \\ y \\ z. \end{pmatrix}, \tag{5.9}$$

Such columns of numbers are usually interpreted as the Cartesian coordinates of the tip of the vector if its initial point is in the origin. So for example, we can visualise $\vec{v}$ as a vector which we obtain when we move 3 units along the x-axis, 4 units along the y-axis and -1 unit along the z-axis. If we set $\vec{e}_1$, $\vec{e}_2$, $\vec{e}_3$ the unit vectors which are parallel to the x-, y- and z-axis, respectively, then we can write $\vec{v}$ as a weighted sum of them:

$$\vec{v} = \begin{pmatrix} 3 \\ 2 \\ -1 \end{pmatrix} = 3\vec{e}_1 + 2\vec{e}_2 - \vec{e}_3. \tag{5.10}$$

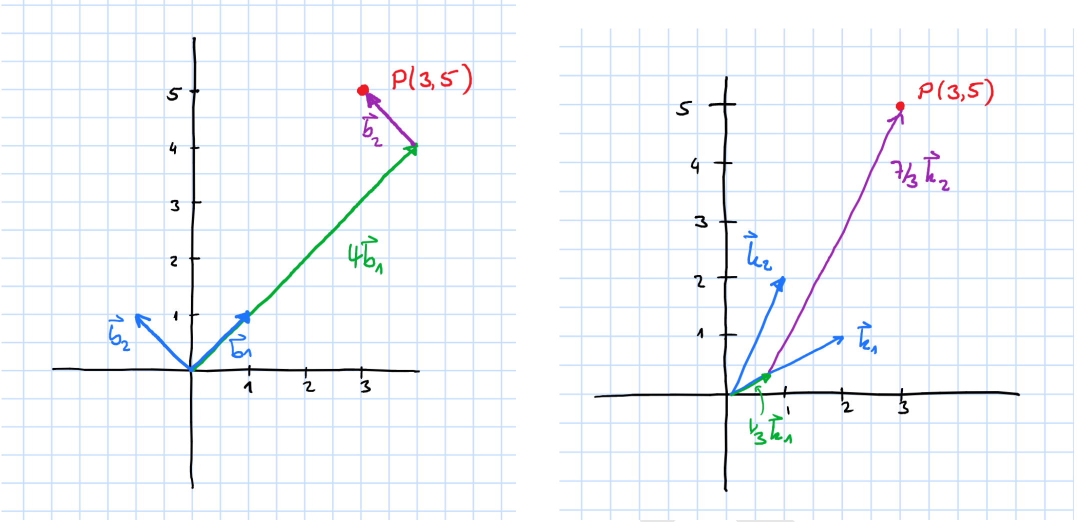

FIGURE 5.1: The pictures shows the point $(3,5)$ in "bishop" and "knight" coordinates. The vectors for the bishop are $\vec{b}_1 = \binom{1}{1}$, $\vec{b}_2 = \binom{-1}{1}$ and $\vec{x}_B = \binom{3}{1}$. The vectors for the knight are $\vec{k}_1 = \binom{2}{1}$, $\vec{k}_2 = \binom{1}{2}_B$ and $\vec{x}_K = \begin{pmatrix} \frac{1}{3} \\ \frac{7}{3} \end{pmatrix}_K$.

So the column of numbers which we use to describe $\vec{v}$ in (5.9) can be seen as a convenient way to abbreviate the sum in (5.10).

Sometimes it makes sense to describe a certain vector not by its Cartesian coordinates. For instance, think of an infinitely big chess field (this is $\mathbb{R}^2$). Then the rock is moving a along the Cartesian axis while the bishop moves a along the diagonals, that is along $\vec{b}_1 = \binom{1}{1}$, $\vec{b}_2 = \binom{-1}{1}$ and the knight moves in directions parallel to $\vec{k}_1 = \binom{2}{1}$, $\vec{k}_2 = \binom{1}{2}$. We suppose that in our imaginary chess game the rock, the bishop and the knight may move in arbitrary multiples of their directions. Suppose all three of them are situated in the origin of the field and we want to move them to the field $(3,5)$. For the rock, this is very easy. It only has to move 3 steps to the right and then 5 steps up. He would denote his movement as $\vec{v}_r = \binom{3}{5}$. The bishop cannot do this. He can move only along the diagonals. So what does he have to do? We has to move 4 steps in direction of $\vec{b}_1$ and 1 step indirection $\vec{b}_2$. So he would denote his movement with respect to his bishop coordinate system as $\vec{v}_B = \binom{4}{2}_B$. Finally the knight has to move $\frac{1}{3}$ steps in direction $\vec{k}_1$ and $\frac{7}{3}$ steps in direction $\vec{k}_2$ to reach the point $(3,5)$. So he would denote his movement with respect to his knight coordinate system as $\vec{v}_K = \binom{1/3}{7/3}_K$.

Exercise. Check that $\vec{v}_B = \binom{4}{2}_b = 4\vec{b}_1 + 2\vec{b}_2 = \binom{3}{5}$ and that $\vec{v}_K = \binom{1/3}{7/3}_k = 1/3\vec{k}_1 + 7/3\vec{k}_2 = \binom{3}{5}$.

So the three vectors $\vec{v}$, $\vec{v}_B$ and $\vec{v}_K$ look very differently but they describe the same vector if we remember that the have to be interpreted as linear combinations of the vectors that describe their movements.

What we just did was to perform a change of bases in $\mathbb{R}^2$: Instead of describing a point in the plane in Cartesian coordinates, we used “bishop”- and “knight”-coordinates.

We can also go in the other direction and transform from “bishop”- or “knight”-coordinates to Cartesian coordinates. Assume that we know that the bishop moves 3 steps in his direction $\vec{b}_1$ and -2 steps in his direction $\vec{b}_2$, where does he end up? In his coordinate system, he is displaced by the vector $\vec{u}_B = \left(\begin{smallmatrix} 3 \\ -2 \end{smallmatrix}\right)$. In Cartesian coordinates this would be $\vec{u}_B = 3\vec{b}_1 - 2\vec{b}_2 = \left(\begin{smallmatrix} 3 \\ 3 \end{smallmatrix}\right) + \left(\begin{smallmatrix} -2 \\ -2 \end{smallmatrix}\right) = \left(\begin{smallmatrix} 1 \\ 1 \end{smallmatrix}\right)$.

If we move the knight 2 steps in his direction $\vec{k}_1$ and 3 step in his direction $\vec{k}_2$, that is, we move him along $\vec{w}_K = \left(\begin{smallmatrix} 2 \\ 3 \end{smallmatrix}\right)$ according to his coordinate system. In Cartesian coordinates this would be $\vec{w}_K = 4\vec{b}_1 + 3\vec{b}_2 = \left(\begin{smallmatrix} 8 \\ 4 \end{smallmatrix}\right) + \left(\begin{smallmatrix} 3 \\ 6 \end{smallmatrix}\right) = \left(\begin{smallmatrix} 11 \\ 10 \end{smallmatrix}\right)$.

Can the bishop and the knight reach every point in the plane? If so, in how many ways? The answer is yes, and they can do so in exactly on way. The reason is that for the bishop and for the knight, their set of direction vectors each form a basis of $\mathbb{R}^2$ (verify this!).

Let us formalise what we just did. Assume we are given an ordered basis $B = \{\vec{b}_1, \dots, \vec{b}_n\}$ of $\mathbb{R}^n$. If we write

$$\vec{x}_B = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}_B \tag{5.11}$$

then we interprete it a vector which is expressed with respect to the basis B. If there in no index attached to the vector, then we interprete it as an vector with respect to the canonical basis $\vec{e}_1, \dots, \vec{e}_n$ of $\mathbb{R}^n$. Now we want to find a way to calculate the Cartesian coordinates (that is, those with respect to the canonical basis) if we are given a vector in B-coordinates and the other way around.

It will turn out that the following matrix will be very useful: $A_{B\to can} = (\vec{v}_1 | \dots | \vec{v}_n)$ = matrix whose columns are the vectors of the basis B. We will explain the index “$B \to can$” in a moment.

- Suppose we are given a vector as in (5.11). How do we obtain its Cartesian coordinates?

This is quite straightforward. We only need to remember what the notation $(\vdots)_B$ means. We will denote by $\vec{x}_B$ the representation of the vector with respect to the basis B and by $\vec{x}$ its representation with respect to the standard basis of $\mathbb{R}^n$.

$$\vec{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}_B = x_1\vec{b}_1 + x_2\vec{b}_2 + \cdots + x_n\vec{b}_n = (\vec{b}_1|\vec{b}_2|\cdots|\vec{b}_n) \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = A_{B\to can} \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = A_{B\to can}\vec{x}_B,$$

that is

$$\vec{x} = A_{B\to can}\vec{x}_B, \tag{5.12}$$

The last vector (the one with the $y_1, \dots, y_n$ in it) describes the same vector as $\vec{x}_B$, but it does so with respect to the standard basis of $\mathbb{R}^n$). The matrix is called the *transition matrix from the basis B to the canonical basis* (which explains the subscript “$B \to can$”). The matrix is also called the *change-of-coordinates matrix*

- Suppose we are given a vector $\vec{x}$ in Cartesian coordinates. How do we calculate its coordinates $\vec{x}_B$ with respect to the basis B?

We only need to remember the relation between $\vec{x}$ and $\vec{x}_B$ which according to (5.12) is

$$\vec{x} = A_{B\to can}\vec{x}_B.$$

In this case, we know the entries of the vector $\vec{x}_B$. So we only need to invert the matrix $A_{B\to can}$ in order to obtain the entries of $\vec{x}_B$:

$$\vec{x}_B = A^{-1}_{B\to can}\vec{x}.$$

This requires of course to know that $A_{B\to can}$ invertible. But this is guaranteed by Theorem 4.33 since we know that its columns are linearly independent. So it follows that the transitions matrix from the canonical basis to the basis B is given by

$$A_{can\to B} = A^{-1}_{B\to can}.$$

Note that we could do this also "by hand": We are given $\vec{x} = (y_1, \dots, y_n)$ and we want to find the entries $x_1, \dots, x_n$ of the vector $\vec{x}_B$ which describes the same vector. That is, we need numbers $x_1, \dots, x_n$ such that

$$\vec{x} = x_1\vec{b}_1 + \dots + \vec{b}_n x_n.$$

If we know the vectors $\vec{b}_1, \dots, \vec{b}_n$, then we can write this as a $n \times n$ system of linear equations and then solve it for $x_1, \dots, x_n$ which of course in reality is the same as inverting the matrix $A_{B\to can}$.

Now assume that we have two ordered bases $B = \{\vec{b}_1, \dots, \vec{b}_n\}$ and $C = \{\vec{c}_1, \dots, \vec{c}_n\}$ of $\mathbb{R}^n$ and we are given a vector $\vec{x}_B$ with respect to the basis B. How can we calculate its representation $\vec{x}_C$ with respect to the basis C? The easiest way is to use the canonical basis of $\mathbb{R}^n$ as an auxiliary basis. So we first calculate the given vector $\vec{x}_B$ with respect to the canonical basis, we call this vector $\vec{x}$. Then we go from $\vec{x}$ to $\vec{x}_C$. According to the formulas above, this is

$$\vec{x}_C = \vec{A}_{can\to C}\vec{x} = A_{can\to C}A_{B\to can}\vec{x}_B$$

Hence the transition matrix from the basis B to the basis C is

$$A_{B\to C} = A_{can\to C}A_{B\to can}.$$

Example 5.35. Let us go back to our example of our imaginary chess board. We have the "bishop basis" $B = \{\vec{b}_1, \vec{b}_2\}$ where $\vec{b}_1 = \left(\begin{smallmatrix}1\\1\end{smallmatrix}\right)$, $\vec{b}_2 = \left(\begin{smallmatrix}-1\\1\end{smallmatrix}\right)$ and the "knight basis" $K = \{\vec{k}_1, \vec{k}_2\}$ $\vec{k}_1 = \left(\begin{smallmatrix}2\\1\end{smallmatrix}\right)$, $\vec{k}_2 = \left(\begin{smallmatrix}1\\2\end{smallmatrix}\right)$. Then the transition matrices to the canonical basis are

$$A_{B\to can} = \begin{pmatrix}1 & -1\\ 1 & 1\end{pmatrix}, \quad A_{K\to can} = \begin{pmatrix}2 & 1\\ 1 & 2\end{pmatrix},$$

their inverses are

$$A_{can\to B} = \frac{1}{2}\begin{pmatrix}1 & 1\\ -1 & 1\end{pmatrix}, \quad A_{can\to K} = \frac{1}{3}\begin{pmatrix}2 & -1\\ -1 & 2\end{pmatrix}$$

and the transition matrices from C to K and from K to C are

$$A_{B\to K} = \frac{1}{3}\begin{pmatrix}3 & -3\\ 1 & 1\end{pmatrix}, \quad A_{K\to C} = \frac{1}{2}\begin{pmatrix}1 & 3\\ -1 & 3\end{pmatrix}.$$

- Given a vector $\vec{x}_B = \left(\begin{smallmatrix}2\\7\end{smallmatrix}\right)_B$ in bishop coordinates, how does it look like in knight coordinates?

Solution. $\vec{x}_K = A_{B\to K}\vec{x}_B = \frac{1}{3}\left(\begin{smallmatrix} 3 & -3 \\ 1 & 1 \end{smallmatrix}\right)\left(\begin{smallmatrix} 2 \\ 7 \end{smallmatrix}\right) = \left(\begin{smallmatrix} -5 \\ 3 \end{smallmatrix}\right)$. ◇

- Given a vector $\vec{y}_K = \left(\begin{smallmatrix} 5 \\ 1 \end{smallmatrix}\right)_K$ in knight coordinates, how does it look like in bishop coordinates?

  Solution. $\vec{y}_B = A_{K\to B}\vec{y}_K = \frac{1}{2}\left(\begin{smallmatrix} 1 & 3 \\ -1 & 3 \end{smallmatrix}\right)\left(\begin{smallmatrix} 5 \\ 1 \end{smallmatrix}\right) = \left(\begin{smallmatrix} 3 \\ -1 \end{smallmatrix}\right)$. ◇

- Given a vector $\vec{z} = \left(\begin{smallmatrix} 1 \\ 3 \end{smallmatrix}\right)$ in standard coordinates, how does it look like in bishop coordinates?

  Solution. $\vec{z}_B = A_{can\to B}\vec{z} = \frac{1}{2}\left(\begin{smallmatrix} 1 & 1 \\ -1 & 1 \end{smallmatrix}\right)\left(\begin{smallmatrix} 1 \\ 3 \end{smallmatrix}\right) = \left(\begin{smallmatrix} 2 \\ 1 \end{smallmatrix}\right)$. ◇

Example 5.36. Recall the example where we had a shop that sold different types of packages of food. Package type A contains 4 sausages and 3 potatoes and package type B contains 1 sausage and 2 potatoes and we wanted to know how many packages of each type we had to buy if we want to have 7 sausages and 9 potatoes. This can be viewed a as a change-of-bases problem. If we view all every point in the xy-plane as representing a configuration (sausage, potato), then what we wanted to do is to write a given sausage-potato vector as a (package A)-(package B)-vector.

In the rest of this section we will apply these ideas to introduce coordinates in abstract (finitely generated) vector spaces V given a basis. This allows us to identify in a certain sense V with an appropriate $\mathbb{R}^n$ or $\mathbb{C}^n$.
Assume we are given a real vector space V with an ordered basis $B = \{v_1, \dots, v_n\}$. (Everything works the same if V is a complex vector space; we only need to replace $\mathbb{R}$ by $\mathbb{C}$ and the word "real" by "complex" everywhere). Given a vector $w \in V$, we know that there are uniquely determined real numbers $\alpha_1, \dots, \alpha_n$ such that

$$w = \alpha_1 v_1 + \cdots + \alpha_n v_n.$$

So, if we are given w, we can find the numbers $\alpha_1, \dots, \alpha_n$. On the other hand, if we are given the numbers $\alpha_1, \dots, \alpha_n$, we can easily reconstruct the vector w (just replace in the right hand side of the above equation). Therefore it makes sense to write

$$w = \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{pmatrix}_B$$

where again the index B reminds us that the column of numbers has to be understood at the coefficients with respect to the basis B. In this way, we identify V with $\mathbb{R}^n$ since every column vector gives on vector w in V and every vector w gives one column vector in $\mathbb{R}^n$. Note that if we start with some w in V, calculate its coordinates in $\mathbb{R}^n$ and then go back to V, we end up again with the original vector w.

Example 5.37. In P_2, consider the bases $B = \{p_1, p_2, p_3\}$, $C = \{q_1, q_2, q_3\}$, $D = \{r_1, r_2, r_3\}$ where

$$p_1 = 1,\ p_2 = X,\ p_3 = X^2, \qquad q_1 = X^2,\ q_2 = X,\ q_3 = 1, \qquad r_1 = X^2 + 2X,\ r_2 = 5X + 2,\ r_3 = 1.$$

We want to write the polynomial $t(X) = aX^2 + bX + c$ with respect to the given basis.

- Basis B: Clearly, $t = cp_1 + bp_2 + ap_3$, therefore
$$t = \begin{pmatrix} c \\ b \\ a \end{pmatrix}_B .$$
- Basis C: Clearly, $t = aq_1 + bq_2 + cq_3$, therefore
$$t = \begin{pmatrix} a \\ b \\ c \end{pmatrix}_C .$$
- Basis D: This requires some calculations. Recall that we need numbers $\alpha, \beta, \gamma \in \mathbb{R}$ such that
$$t = \begin{pmatrix} \alpha \\ \beta \\ \gamma \end{pmatrix}_D = \alpha r_1 + r_2\beta + r_3\gamma.$$
This leads to the following equation
$$aX^2 + bX + c = \alpha(X^2 + 2X) + \beta(5X + 2) + \gamma = \alpha X^2 + (2\alpha + 5\beta)X + 2\beta + \gamma.$$
Comparing coefficients we obtain
$$\left.\begin{aligned} \alpha \quad\quad &= a \\ 2\alpha + 5\beta \quad &= b \\ 2\beta + \gamma &= c. \end{aligned}\right\} \quad \text{in matrix form:} \quad \begin{pmatrix} 1 & 0 & 0 \\ 2 & 5 & 0 \\ 0 & 2 & 1 \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \\ \gamma \end{pmatrix} = \begin{pmatrix} a \\ b \\ c \end{pmatrix}. \tag{5.13}$$
Note that the columns of the matrix appearing on the right hand side are exactly are exactly the vector representations with respect to the basis C and the column vector $\left(\begin{smallmatrix} a \\ b \\ c \end{smallmatrix}\right)$ is exactly the vector representation of t with respect to the basis C! The solution of the system is
$$\alpha = a, \quad \beta = -\tfrac{1}{5}a + \tfrac{1}{5}b, \quad \gamma = \tfrac{4}{5}a - \tfrac{2}{5}b + c,$$
therefore
$$t = \begin{pmatrix} a \\ -\frac{1}{5}a + \frac{1}{5}b \\ \frac{4}{5}a - \frac{2}{5}b + c \end{pmatrix}_D .$$

We could have found the solution also by doing a detour through $\mathbb{R}^3$ as follows: We identify the vectors q_1, q_2, q_3 with the canonical basis vectors e_1, $\vec{e}_2, \vec{e}_3$ of $\mathbb{R}^3$. Then the vectors r_1, r_2, r_3 and t correspond to
$$\vec{r_1}' = \begin{pmatrix} 1 \\ 2 \\ 0 \end{pmatrix}, \quad \vec{r_2}' = \begin{pmatrix} 0 \\ 5 \\ 2 \end{pmatrix}, \quad \vec{r_3}' = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}, \quad \vec{t}' = \begin{pmatrix} a \\ b \\ c \end{pmatrix}.$$

Let $R = \{\vec{r}_1', \vec{r}_2', \vec{r}_3'\}$. In order to find the coordinates of $\vec{t}'$ with respect to the basis $\vec{r}_1', \vec{r}_2', \vec{r}_3'$, we note that

$$\vec{t}' = A_{R\to can}\vec{t}_R'$$

where $A_{R\to can}$ is the transition matrix from the basis R to the canonical basis of $\mathbb{R}$ whose columns consist of the vectors $\vec{r}_1', \vec{r}_2', \vec{r}_3'$. So we see that this is exactly the same equation as the one in (5.13).

We give a final example in a space of matrices.

Example 5.38. Consider the matrices

$$R = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}, \quad S = \begin{pmatrix} 1 & 0 \\ 0 & 3 \end{pmatrix}, \quad T = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad Z = \begin{pmatrix} 2 & 3 \\ 3 & 0 \end{pmatrix}.$$

(i) Show that $\mathcal{B} = \{R, S, T\}$ is a basis of $M_{sym}(2\times 2)$ (the space of all symmetric 2×2 matrices).

(ii) Write Z in terms of the basis $\mathcal{B}$.

Solution. (i) Clearly, $R, S, T \in M_{sym}(2\times 2)$. Since we already now that $\dim M_{sym}(2\times 2) = 3$, it suffices to show that R, S, T are linearly independent. So let us consider the equation

$$0 = \alpha R + \beta S + \gamma T = \begin{pmatrix} \alpha+\beta & \alpha+\gamma \\ \alpha+\gamma & \alpha+3\beta \end{pmatrix}.$$

We obtain the system of equations

$$\left.\begin{aligned} \alpha + \ \ \beta \quad\ &= 0 \\ \alpha + \quad\ + \gamma &= 0 \\ \alpha + 3\beta + \quad &= 0 \end{aligned}\right\} \qquad \text{in matrix form:} \quad \underbrace{\begin{pmatrix} 1\ 1\ 0 \\ 1\ 0\ 1 \\ 1\ 3\ 0 \end{pmatrix}}_{=A} \begin{pmatrix} \alpha \\ \beta \\ \gamma \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}. \tag{5.14}$$

Doing some calculations, if follows that $\alpha = \beta = \gamma = 0$. Hence we showed that R, S, T are linearly independent and therefore they are a basis of $M_{sym}(2\times 2)$.

(ii) In order to write Z in terms of the basis $\mathcal{B}$, we need to find $\alpha, \beta, \gamma \in \mathbb{R}$ such that

$$Z = \alpha R + \beta S + \gamma T = \begin{pmatrix} \alpha+\beta & \alpha+\gamma \\ \alpha+\gamma & \alpha+3\beta \end{pmatrix}.$$

We obtain the system of equations

$$\left.\begin{aligned} \alpha + \ \ \beta \quad\ &= 2 \\ \alpha + \quad\ + \gamma &= 3 \\ \alpha + 3\beta + \quad &= 0 \end{aligned}\right\} \qquad \text{in matrix form:} \quad \underbrace{\begin{pmatrix} 1\ 1\ 0 \\ 1\ 0\ 1 \\ 1\ 3\ 0 \end{pmatrix}}_{=A} \begin{pmatrix} \alpha \\ \beta \\ \gamma \end{pmatrix} = \begin{pmatrix} 2 \\ 3 \\ 0 \end{pmatrix}. \tag{5.15}$$

Therefore

$$\begin{pmatrix} \alpha \\ \beta \\ \gamma \end{pmatrix} = A^{-1}\begin{pmatrix} 2 \\ 3 \\ 0 \end{pmatrix} = \frac{1}{2}\begin{pmatrix} 3 & 0 & -1 \\ -1 & 0 & 1 \\ -3 & 2 & 1 \end{pmatrix}\begin{pmatrix} 2 \\ 3 \\ 0 \end{pmatrix} = \begin{pmatrix} 3 \\ -1 \\ 0 \end{pmatrix},$$

therefore $Z = 3R - S = \begin{pmatrix} 3 \\ -1 \\ 0 \end{pmatrix}_{\mathcal{B}}$. ◇

Now we give an alternative solution (which is essentially the same as the above) doing a detour through $\mathbb{R}^3$. Let $\mathcal{C} = \{A_1,\, A_2,\, A_3\}$ where $A_1 = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$, $A_2 = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}$, $A_3 = \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}$. This is clearly a basis of $M_{sym}(2 \times 2)$. We identify it with the standard basis $\vec{e}_1, \vec{e}_2, \vec{e}_3$ of $\mathbb{R}^3$. Then the vectors R, S, T in this basis look like

$$R' = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \qquad S' = \begin{pmatrix} 1 \\ 0 \\ 3 \end{pmatrix}, \qquad T' = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \qquad \text{and} \qquad Z' = \begin{pmatrix} 2 \\ 3 \\ 0 \end{pmatrix}.$$

(i) In order to show that R, S, T are linearly independent, we only have to show that the vectors R', S' and T' are linearly independent in $\mathbb{R}$. To this end, we consider the matrix A whose columns are these vectors. Note that his is the same matrix that appeared in (5.15). It is easy to show that this matrix is invertible (we already calculated its inverse!). Therefore the vectors R', S', T' are linearly independent in $\mathbb{R}^3$, hence R, S, T are linearly independent in $M_{sym}(2 \times 2)$.

(ii) Now in order to find the representation of Z in terms of the basis $\mathcal{B}$, we only need to find the representation of Z' in terms of the basis $\mathcal{B}' = \{R', S', T'\}$. This is done as follows:

$$Z'_{\mathcal{B}'} = A_{can \to \mathcal{B}'} Z' = A^{-1} Z' = \begin{pmatrix} 2 \\ 3 \\ 0 \end{pmatrix}.$$

5.4 Matrix representation of linear maps

Let U, V be $\mathbb{K}$-vector spaces and let $T : U \to V$ be a linear map. Recall that T satisfies

$$T(\lambda_1 x_1 + \cdots + \lambda_k x_k) = \lambda_1 T(x_1) + \cdots + \lambda_k T(x_k)$$

for all $x_1, \ldots, x_k \in U$ and $\lambda_1, \ldots, \lambda_k \in \mathbb{K}$. This shows that in order to know T, it is in reality enough to know how a T acts on a basis of U. Suppose that we are given a basis $\mathcal{B} = \{u_1, \ldots, u_n\} \in U$ and let $w \in U$ arbitrary. Then there exist uniquely determined $\lambda_1, \ldots, \lambda_k \in \mathbb{K}$ such that $w = \lambda_1 u_1 + \cdots + \lambda_n u_n$. Hence

$$Tw = T(\lambda_1 u_1 + \cdots + \lambda_n u_n) = \lambda_1 T u_1 + \cdots + \lambda_n T u_n. \tag{5.16}$$

So Tw is a linear combination of the vectors $Tu_1, \ldots, Tu_n \in V$, and the coefficients are exactly the $\lambda_1, \ldots, \lambda_n$.

Suppose we are given a basis $\mathcal{C} = \{v_1, \ldots, v_k\}$ of V. Then we know that for every $j = 1, \ldots, n$, the vector Tu_j is a linear combination of the basis vectors $v_1, \ldots, v_m$ of V. Therefore there exist uniquely determined numbers $a_{ij} \in K$ $(i = 1, \ldots, m,\ j = 1, \ldots n)$ such that $Tu_j = a_{j1} v_1 + \cdots + a_{jm} v_m$, that is

$$\begin{aligned} Tu_1 &= a_{11} v_1 + a_{21} v_2 + \cdots + a_{m1} v_m, \\ Tu_2 &= a_{12} v_1 + a_{22} v_2 + \cdots + a_{m2} v_m, \\ \vdots \quad & \quad \vdots \qquad\quad \vdots \qquad\qquad\quad \vdots \\ Tu_n &= a_{1n} v_1 + a_{2n} v_2 + \cdots + a_{mn} v_m. \end{aligned} \tag{5.17}$$

Let us define the matrix A_T and the vector $\vec{\lambda}$ by

$$A_T = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix} \in M(m \times n), \qquad \vec{\lambda} = \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_n \end{pmatrix} \in \mathbb{R}^n.$$

Recall that A_T represents a linear map from $\mathbb{R}^n$ to $\mathbb{R}^m$.

Now let us come back to the calculation of Tw and its connection with the matrix A_T. From (5.16) and (5.17) we obtain

$$\begin{aligned} Tw &= \lambda_1 Tu_1 + \lambda_2 Tu_2 + \cdots + \lambda_n Tu_n \\ &= \lambda_1(a_{11}v_1 + a_{21}v_2 + \cdots + a_{m1}v_m) \\ &\quad + \lambda_2(a_{12}v_1 + a_{22}v_2 + \cdots + a_{m2}v_m) \\ &\quad + \qquad \cdots \\ &\quad + \lambda_n(a_{1n}v_1 + a_{2n}v_2 + \cdots + a_{mn}v_m) \\ &= (a_{11}\lambda_1 + a_{12}\lambda_2 + \cdots + a_{1n}\lambda_n)v_1 \\ &\quad + (a_{21}\lambda_1 + a_{22}\lambda_2 + \cdots + a_{2n}\lambda_n)v_2 \\ &\quad + \cdots \\ &\quad + (a_{m1}\lambda_1 + a_{m2}\lambda_2 + \cdots + a_{mn}\lambda_n)v_m. \end{aligned}$$

The calculation shows that for every k the coefficient of v_k is the kth component of the vector $A_T\vec{\lambda}$! Now we can go one step further. Recall the choice of the basis $\mathcal{B}$ of U and the basis $\mathcal{C}$ of V lets us write w and Tw as a column vectors:

$$w = \vec{w}_{\mathcal{B}} \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_1 \end{pmatrix}_{\mathcal{B}}, \qquad Tw = \begin{pmatrix} a_{11}\lambda_1 + a_{12}\lambda_2 + \cdots + a_{1n}\lambda_n \\ a_{21}\lambda_1 + a_{22}\lambda_2 + \cdots + a_{2n}\lambda_n \\ \vdots \\ a_{m1}\lambda_1 + a_{m2}\lambda_2 + \cdots + a_{mn}\lambda_n \end{pmatrix}_{\mathcal{C}}.$$

This shows that

$$(Tw)_{\mathcal{C}} = A_T\vec{w}_{\mathcal{B}}.$$

For obvious reasons, the matrix A_T is called the *matrix representation of T with respect to the bases $\mathcal{B}$ and $\mathcal{C}$.*

So every linear transformation $T : U \to V$ can be represented as a matrix $A_T \in M(m \times n)$. On the other hand, every a matrix $A(m \times n)$ induces a linear transformation $T_A : U \to V$.

Very important remark. This identification of $m \times n$-matrices with linear maps $U \to V$ depends on the choice of the basis! See Example 5.40.

Let us summarise what we have found so far.

Theorem 5.39. *Let U, V be vector spaces and let $\mathcal{B} = \{u_1, \dots, u_n\}$ be a basis of U and $\mathcal{C} = \{v_1, \dots, v_n\}$ be a basis of V. Then the following is true:*

(i) *Every linear map* $T : U \to V$ *can be represented as a matrix* $A_T \in M(m \times n)$ *such that*

$$(Tw)_{\mathcal{C}} = A_T \vec{w}_B$$

where $(Tw)_{\mathcal{C}}$ *is the representation of* $Tw \in V$ *with respect to the basis* $\mathcal{C}$ *and* $\vec{w}_{\mathcal{B}}$ *is the representation of* $w \in U$ *with respect to the basis* $\mathcal{B}$*. The entries* a_{ij} *of* A_T *can be calculated as in* (5.17)*.*

(ii) *Every matrix* $A = (a_{ij})_{\substack{i=1,\dots,m\\ j=1,\dots,n}} \in M(mt \times n$ *induces a linear transformation* $T : U \to V$ *defined by*

$$T(u_j) = a_{1j}v_1 + \dots a_{mj}v_m, \qquad j = 1, \dots, n.$$

(iii) $T = T_{A_T}$ *and* $A = A_{T_A}$*. , That means: If we start with a linear map* $T : U \to V$*, calculate its matrix representation* A_T *and then the linear map* $T_{A_T} : U \to V$ *induced by* A_T*, then we get back our original map* T*. If on the other hand we start with a matrix* $A \in M(m \times n)$*, calculate the linear map* $T_A : U \to V$ *induced by* A *and then calculate its matrix representation* A_{T_A}*, then we get back our original matrix* A*.*

Proof. We already show (i) and (ii) in the text before the theorem. To see **??**, let us start with a linear transformation $T : U \to V$ and let $A_T = (a_{ij})$ be the matrix representation of T with respect to the bases $\mathcal{B}$ and $\mathcal{C}$. For T_{A_T}, the linear map induced by A_T, it follows that

$$TA_Tu_j = a_{1j}v_1 + \dots a_{mj}v_m = Tu_j, \qquad j = 1, \dots, n$$

Since this is true for all basis vectors and both T and T_{A_T} are linear, they must be equal.
If on the other hand we are given a matrix $A = (a_{ij})_{\substack{i=1,\dots,m\\ j=1,\dots,n}} \in M(mt \times n$ then we have that the linear transformation T_A induced by A acts on the basis vectors $u_1, \dots, u_n$ as follows:

$$T_Au_j = TA_Tu_j = a_{1j}v_1 + \dots a_{mj}v_m.$$

But then, by definition of the matrix representation A_{T_A} of T_A, it follows that $A_{T_A} = A$. ☐

Let us see this “identifications” of matrices with linear transformations a bit more formally. By choosing a basis $\mathcal{B} = \{u_1, \dots, u_n\}$ in U and thereby identifying U with $\mathbb{R}^n$, we are in reality defining a linear bijection

$$\Psi : U \to \mathbb{R}^n, \qquad \Psi(\lambda u_1 + \dots + \lambda_n u_n) = \begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_n \end{pmatrix}.$$

Recall that we denoted the vector on the right hand side by $\vec{w}_{\mathcal{B}}$.
The same happens if we choose a basis $\mathcal{C} = \{v_1, \dots, v_m\}$ of V. We obtain a linear bijection

$$\Phi : V \to \mathbb{R}^m, \qquad \Phi(\mu v_1 + \dots + \mu_m v_m) = \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_m \end{pmatrix}.$$

With these linear maps, we find that

$$A_T = \Phi \circ T \circ \Psi^{-1} \qquad \text{and} \qquad T_A = \Phi^{-1} \circ A \circ \Psi.$$

The maps Ψ and Φ "translate" the spaces U and V in $\mathbb{R}^n$ and $\mathbb{R}^m$ where the chosen bases serve as "dictionary". Thereby they "translate" linear maps $U : U \to V$ to matrices $A \in M(m \times n)$ and vice versa. In a diagram this looks likes this:

$$\begin{tikzcd} U \arrow[r, "T"] \arrow[d, "\Psi"'] & V \arrow[d, "\Phi"] \\ \mathbb{R}^n \arrow[r, "A_T"] & \mathbb{R}^m \end{tikzcd}$$

So in order to go from U to V, we can take the detour through $\mathbb{R}^n$ and $\mathbb{R}^m$. One say that the diagram above *commutes*. That means that it does not matter which path we take to go from on corner of the diagram to another one as long as we move in the directions of the arrows. Note that in this case we are even allowed to go in the opposite directions of the arrows representing Ψ and Φ because they are bijections.

What is the use of a matrix representation of a linear map? Sometimes calculations are easier in the world of matrices. For example, we know how to calculate the range and the kernel of a matrix. Therefore:

- If we want to calculate $\operatorname{Im} T$, we only need to calculate $\operatorname{Im} A_T$ and then use Φ to "translate back" to the range of T. In formula: $\operatorname{Im} T = \operatorname{Im}(\Phi A_T) = \Phi(\operatorname{Im} A_T)$.
- If we want to calculate $\ker T$, we only need to calculate $\ker A_T$ and then use Ψ to "translate back" to the kernel of T. In formula: $\ker T = \ker(A_T \Psi) = \Psi^{-1}(\ker A_T)$.
- If $\dim U = \dim V$, i.e., if $n = m$, then T is invertible if and only if A_T is invertible. This is the case if and only if $\det A_T \neq 0$.

In particular, we obtain the following formula for a linear transformation $T : U \to V$:

$$\boxed{\dim U = \dim(\ker T) + \dim(\operatorname{Im} T)} \tag{5.18}$$

Let us see some examples.

Example 5.40. We consider the operator of differentiation

$$T : P_3 \to P_3, \quad Tp = p'.$$

Note that in this case the vector spaces U and V are both equal to P_3.

(i) Represent T with respect to the basis $\mathcal{B} = \{p_1, p_2, p_3, p_4\}$ and find its kernel where $p_1 = 1, p_2 = X, p_3 = X^2, p_4 = X^3$.

Solution. We only need to evaluate T in the elements of the basis and the write the result again as linear combination of the basis. Since in this case, the bases are "easy", the calculations are fairly easy:

$$Tp_1 = 0, \quad Tp_2 = 1 = p_1, \quad Tp_3 = 2X = 2p_2, \quad Tp_4 = 3X^2 = 3p_3.$$

Therefore the matrix representation of T is

$$A_T^{\mathcal{B}} = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

The kernel of A_T is clearly $\operatorname{span}\{\vec{e}_1\}$, hence $\ker T = \operatorname{span}\{p_1\} = \operatorname{span}\{1\}$. ◇

(ii) Represent T with respect to the basis $\mathcal{C} = \{q_1, q_2, q_3, q_4\}$ and find its kernel where $q_1 = X^3, q_2 = X^2, q_3 = X, q_4 = 1$.

Solution. Again we only need to evaluate T in the elements of the basis and the write the result again as linear combination of the basis.

$$Tq_1 = 3X^2 = 3q_2, \quad Tq_2 = 2X = 2q_3, \quad Tq_3 = X = q_4, \quad Tq_4 = 0.$$

Therefore the matrix representation of T is

$$A_T^{\mathcal{C}} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 3 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}.$$

The kernel of A_T is clearly $\operatorname{span}\{\vec{e}_4\}$, hence $\ker T = \operatorname{span}\{q_4\} = \operatorname{span}\{1\}$. ◇

(iii) Represent T with respect to the basis $\mathcal{B}$ in the domain of T (in the "left" P_3) and the basis $\mathcal{C}$ in the target space (in the "right" P_3).

Solution. We calculate

$$Tp_1 = 0, \quad Tp_2 = 1 = q_4, \quad Tp_3 = 2X = 2q_3, \quad Tp_4 = 3X^2 = 3q_2.$$

Therefore the matrix representation of T is

$$A_T^{\mathcal{B},\mathcal{C}} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 3 \\ 0 & 0 & 2 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}.$$

The kernel of A_T is clearly $\operatorname{span}\{\vec{e}_1\}$, hence $\ker T = \operatorname{span}\{p_1\} = \operatorname{span}\{1\}$. ◇

(iv) Represent T with respect to the basis $\mathcal{D} = \{r_1, r_2, r_3, r_4\}$ and find its kernel where

$$r_1 = X^3 + X, \quad r_2 = 2X^2 + X^2 + 2X, \quad r_3 = 3X^3 + X^2 + 4X + 1, \quad r_4 = 4X^3 + X^2 + 4X + 1.$$

Solution 1. Again we only need to evaluate T in the elements of the basis and the write the result again as linear combination of the basis. This time the calculations are a bit more tedious.

$$\begin{aligned} Tr_1 &= 3X^2+1 &&= -\,8r_1+2r_2+r_4, \\ Tr_2 &= 6X^2+2X+2 &&= -\,14r_1+4r_2+r_3, \\ Tr_3 &= 9X^2+2X+4 &&= -\,24r_1+5r_2+2r_3+2r_4, \\ Tr_4 &= 12X^2+2X+4 &&= \quad 30r_1+8r_2+2r_3+2r_4. \end{aligned}$$

Therefore the matrix representation of T is

$$A_T^{\mathcal{D}} = \begin{pmatrix} \text{‘}-8 & -14 & -24 & -30 \\ 2 & 4 & 5 & 8 \\ 0 & 2 & 2 & 2 \\ 1 & 0 & 2 & 2 \end{pmatrix}.$$

In order to calculate the kernel of A_T, we apply the Gauß-Jordan process and obtain

$$A_T^{\mathcal{D}} = \begin{pmatrix} -8 & -14 & -24 & -30 \\ 2 & 4 & 5 & 8 \\ 0 & 2 & 2 & 2 \\ 1 & 0 & 2 & 2 \end{pmatrix} \longrightarrow \quad \cdots \longrightarrow \begin{pmatrix} 1 & 0 & 0 & 2 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

The kernel of A_T is clearly $\operatorname{span}\{-2\vec{e}_1 - \vec{\mathrm{e}}_2 + \vec{\mathrm{e}}_4\}$, hence $\ker T = \operatorname{span}\{-2r_1 - r_2 + r_4\} = \operatorname{span}\{1\}$. ◇

Solution 2. We already have the matrix representation $A_T^{\mathcal{C}}$ and we can use it to calculate $A_T^{\mathcal{D}}$. To this end define the vectors

$$\vec{\rho}_1 = \begin{pmatrix} 1 \\ 0 \\ 1 \\ 0 \end{pmatrix}, \ \vec{\rho}_2 = \begin{pmatrix} 2 \\ 1 \\ 2 \\ 0 \end{pmatrix}, \ \vec{\rho}_3 = \begin{pmatrix} 3 \\ 1 \\ 4 \\ 1 \end{pmatrix}, \ \vec{\rho}_4 = \begin{pmatrix} 4 \\ 1 \\ 4 \\ 1 \end{pmatrix}.$$

Note that these vectors are the representations of our basis vectors $r_1, \dots, r_4$ in the basis $\mathcal{C}$. The change-of-bases matrix from $\mathcal{C}$ to $\mathcal{D}$ and its inverse are, in coordinates,

$$S_{\mathcal{D}\to\mathcal{C}} = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 0 & 1 & 1 & 1 \\ 1 & 2 & 4 & 4 \\ 0 & 0 & 1 & 1 \end{pmatrix}, \qquad S_{\mathcal{C}\to\mathcal{D}} = S_{\mathcal{D}\to\mathcal{C}}^{-1} = \begin{pmatrix} 0 & -2 & 1 & -2 \\ 0 & 1 & 0 & -1 \\ -1 & 0 & 1 & 0 \\ 1 & 0 & -1 & 1 \end{pmatrix}.$$

It follows that

$$\begin{aligned} A_T^{\mathcal{D}} &= S_{\mathcal{C}\to\mathcal{D}} A_T^{\mathcal{C}} S_{\mathcal{D}\to\mathcal{C}} \\ &= \begin{pmatrix} 0 & -2 & 1 & -2 \\ 0 & 1 & 0 & -1 \\ -1 & 0 & 1 & 0 \\ 1 & 0 & -1 & 1 \end{pmatrix} \begin{pmatrix} 0 & 0 & 0 & 0 \\ 3 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} 1 & 2 & 3 & 4 \\ 0 & 1 & 1 & 1 \\ 1 & 2 & 4 & 4 \\ 0 & 0 & 1 & 1 \end{pmatrix} = \begin{pmatrix} -8 & -14 & -24 & -30 \\ 2 & 4 & 5 & 8 \\ 0 & 2 & 2 & 2 \\ 1 & 0 & 2 & 2 \end{pmatrix}. \end{aligned}$$

Let us see how this looks in diagrams. We define the two bijections of P_3 with $\mathbb{R}^4$ which are given by choosing the bases $\mathcal{C}$ and $\mathcal{D}$ by $\Psi_{\mathcal{C}}$ and $\Psi_{\mathcal{D}}$

$$\begin{aligned} &\Psi_{\mathcal{C}} : P_3 \to \mathbb{R}^4, \qquad && \Psi_{\mathcal{C}}(q_1) = \vec{e}_1,\ \Psi_{\mathcal{C}}(q_2) = \vec{e}_2,\ \Psi_{\mathcal{C}}(q_3) = \vec{e}_3,\ \Psi_{\mathcal{C}}(q_4) = \vec{e}_4, \\ &\Psi_{\mathcal{D}} : P_3 \to \mathbb{R}^4, \qquad && \Psi_{\mathcal{D}}(r_1) = \vec{e}_1,\ \Psi_{\mathcal{D}}(r_2) = \vec{e}_2,\ \Psi_{\mathcal{D}}(r_3) = \vec{e}_3,\ \Psi_{\mathcal{D}}(r_4) = \vec{e}_4. \end{aligned}$$

Then we have the following diagrams:

$$\begin{tikzcd}
P_3 \arrow[r, "T"] \arrow[d, "\Psi_{\mathcal{C}}"'] & P_3 \arrow[d, "\Psi_{\mathcal{C}}"] \\
\mathbb{R}^4 \arrow[r, red, "A_T^{\mathcal{C}}"] & \mathbb{R}^4
\end{tikzcd}
\qquad\qquad
\begin{tikzcd}
P_3 \arrow[r, "T"] \arrow[d, "\Psi_{\mathcal{D}}"'] & P_3 \arrow[d, "\Psi_{\mathcal{D}}"] \\
\mathbb{R}^4 \arrow[r, blue, dashed, "A_T^{\mathcal{D}}"] & \mathbb{R}^4
\end{tikzcd}$$

We already know everything in the diagram on the left and we want to calculate $A_T^{\mathcal{D}}$ in the diagram on the right. We can put the diagrams together as follows:

$$\begin{tikzcd}
P_3 \arrow[rrr, "T"] \arrow[d, "\Psi_{\mathcal{D}}"'] \arrow[dr, "\Psi_{\mathcal{C}}"] & & & P_3 \arrow[dl, "\Psi_{\mathcal{C}}"'] \arrow[d, "\Psi_{\mathcal{D}}"] \\
\mathbb{R}^4 \arrow[r, "S_{\mathcal{D}\to\mathcal{C}}"] \arrow[rrr, blue, dashed, bend right, "A_T^{\mathcal{D}}"] & \mathbb{R}^4 \arrow[r, red, "A_T^{\mathcal{C}}"] & \mathbb{R}^4 \arrow[r, "S_{\mathcal{C}\to\mathcal{D}}"] & \mathbb{R}^4
\end{tikzcd}$$

We can also see that the change-of-basis maps $S_{\mathcal{D}\to\mathcal{C}}$ and $S_{\mathcal{C}\to\mathcal{D}}$ are

$$S_{\mathcal{D}\to\mathcal{C}} = \Psi_{\mathcal{C}} \circ \Psi_{\mathcal{D}}^{-1}, \qquad S_{\mathcal{C}\to\mathcal{D}} = \Psi_{\mathcal{D}} \circ \Psi_{\mathcal{C}}^{-1}.$$

For $A_T^{\mathcal{D}}$ we obtain

$$A_T^{\mathcal{D}} = \Psi_{\mathcal{D}} \circ T \circ \Psi_{\mathcal{D}}^{-1} = S_{\mathcal{D}\to\mathcal{C}} \circ A_T^{\mathcal{C}} \circ S_{\mathcal{C}\to\mathcal{D}}.$$

Another way to draw the diagram above is

$$\begin{tikzcd}
P_3 \arrow[rrr, "T"] \arrow[dd, "\Psi_{\mathcal{D}}"'] \arrow[dr, "\Psi_{\mathcal{C}}"] & & & P_3 \arrow[dl, "\Psi_{\mathcal{C}}"'] \arrow[dd, "\Psi_{\mathcal{D}}"] \\
& \mathbb{R}^4 \arrow[r, red, "A_T^{\mathcal{C}}"] & \mathbb{R}^4 \arrow[dr, "S_{\mathcal{C}\to\mathcal{D}}"] & \\
\mathbb{R}^4 \arrow[ur, "S_{\mathcal{D}\to\mathcal{C}}"] \arrow[rrr, blue, dashed, "A_T^{\mathcal{D}}"] & & & \mathbb{R}^4
\end{tikzcd}$$

◇

Note that the matrices $A_T^{\mathcal{B}}$, $A_T^{\mathcal{C}}$, $A_T^{\mathcal{D}}$ and $A_T^{\mathcal{B},\mathcal{C}}$ all look different but they describe the same linear transformation. The reason why they look different is that in each case we used different bases to describe them.

Example 5.41. The next example is not very applied but it serves to practice a bit more. We consider the operator given

$$T : M(2\times 2) \to P_2, \quad T(\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)) = (a+c)X^2 + (a-b)X + a - b + d.$$

Show that T is a linear transformation and represent T with respect to the bases $\mathcal{B} = \{B_1, B_2, B_3, B_4\}$ of $M(2\times 2)$ and $\mathcal{C} = \{p_1, p_2, p_3\}$ of P_2 where

$$B_1 = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \quad B_2 = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \quad B_3 = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, \quad B_4 = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix},$$

and

$$p_1 = 1, \quad p_2 = X, \quad p_3 = X^2.$$

Find based for $\ker T$ and $\operatorname{Im} T$ and their dimensions.

Solution. First we verify that T is indeed a linear map. To this end, we take matrices $A_1 = \left(\begin{smallmatrix} a_1 & b_1 \\ c_1 & d_1 \end{smallmatrix}\right)$ and $A_2 = \left(\begin{smallmatrix} a_2 & b_2 \\ c_2 & d_2 \end{smallmatrix}\right)$ and $\lambda \in \mathbb{R}$. Then

$$\begin{aligned} T(\lambda A_1 + A_2) &= T\left(\lambda\begin{pmatrix} a_1 & b_1 \\ c_1 & d_1 \end{pmatrix} + \begin{pmatrix} a_2 & b_2 \\ c_2 & d_2 \end{pmatrix}\right) = T\left(\lambda\begin{pmatrix} \lambda a_1 + a_2 & \lambda b_1 + b_2 \\ \lambda c_1 + c_2 & \lambda d_1 + d_2 \end{pmatrix}\right) \\ &= (\lambda a_1 + a_2 + \lambda c_1 + c_2)X^2 + (\lambda a_1 + a_2 - \lambda b_1 - b_2)X + \lambda a_1 + a_2 - (\lambda b_1 + b_2) + \lambda d_1 + d_2 \\ &= \lambda(a_1 + c_1)X^2 + (a_1 - b_1)X + a_1 - b_1 + d_1) \\ &\quad + \left[(a_2 + c_2)X^2 + (a_2 - b_2)X + a_2 - b_2 + d_2)\right] \\ &= \lambda T(A_1) + T(A_2). \end{aligned}$$

This shows that T is a linear transformation.
Now we calculate its matrix representation with respect to the given bases.

$$\begin{aligned} TB_1 &= X^2 + X + 1 = p_1 + p_2 + p_3, \\ TB_2 &= -X = -p_2, \\ TB_3 &= X^2 = p_3, \\ TB_4 &= 1 = p_1. \end{aligned}$$

Therefore the matrix representation of T is

$$A_T = \begin{pmatrix} 1 & 0 & 0 & 1 \\ 1 & -1 & 0 & 0 \\ 1 & 0 & 1 & 0 \end{pmatrix}$$

In order to determine the kernel and range of A_T, we apply the Gauß-Jordan process:

$$A_T = \begin{pmatrix} 1 & 0 & 0 & 1 \\ 1 & -1 & 0 & 0 \\ 1 & 0 & 1 & 0 \end{pmatrix} \longrightarrow \begin{pmatrix} 1 & 0 & 0 & 1 \\ 0 & -1 & 0 & -1 \\ 0 & 0 & 1 & -1 \end{pmatrix} \longrightarrow \begin{pmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & -1 \end{pmatrix}.$$

So the range of A_T is $\mathbb{R}^3$ and its kernel is $\ker A_T = \operatorname{span}\{\vec{\mathrm{e}}_1 + \vec{\mathrm{e}}_2 - \vec{\mathrm{e}}_3 - \vec{\mathrm{e}}_3\}$. Therefore $\operatorname{Im} T = P_2$ and $\ker T = \operatorname{span}\{B_1 + B_2 - B_3 - B_4\}$. For their dimensions we find $\dim(\operatorname{Im} T) = 3$ and $\dim(\ker T) = 1$. ◇

Example 5.42 (Reflection in $\mathbb{R}^2$). In $\mathbb{R}^2$, consider the line $L : 3x - 2y = 0$. Let $R : \mathbb{R}^2 \to \mathbb{R}^2$ which takes a vector in $\mathbb{R}^2$ and reflects it on the line L. Find the matrix representation of R with respect to the standard basis of $\mathbb{R}^2$.
Observation. Note that L is the line which passes through the origin and is parallel to the vector $\vec{v} = \left(\begin{smallmatrix} 2 \\ 3 \end{smallmatrix}\right)$.

Solution 1 (use coordinates adapted to the problem). Clearly, there are two directions which are special in this problem: the direction parallel and the direction orthogonal to the line. So a basis which is adapted to the exercise, is $\mathcal{B} = \{\vec{v}, \vec{w}\}$ where $\vec{v} = \left(\begin{smallmatrix} 2 \\ 3 \end{smallmatrix}\right)$ and $\vec{w} = \left(\begin{smallmatrix} -3 \\ 2 \end{smallmatrix}\right)$. Clearly, $R\vec{v} = \vec{v}$ and $R\vec{w} = -\vec{w}$. Therefore the matrix representation of R with respect to the basis $\mathcal{B}$ is

$$A_R^{\mathcal{B}} = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

In order to obtain the representation A_R with respect to the standard basis, we only need to perform a change of basis. Recall that change-of-bases matrices are given by

$$S_{\mathcal{B}\to can} = (\vec{v}|\vec{w}) = \begin{pmatrix} 2 & -3 \\ 3 & 2 \end{pmatrix}, \qquad S_{can\to\mathcal{B}} = S_{\mathcal{B}\to can}^{-1} = \frac{1}{13}\begin{pmatrix} 2 & 3 \\ -3 & 2 \end{pmatrix}.$$

Therefore

$$A_R = S_{\mathcal{B}\to can} A_R^{\mathcal{B}} S_{can\to\mathcal{B}} = \frac{1}{13}\begin{pmatrix} 2 & -3 \\ 3 & 2 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}\begin{pmatrix} 2 & 3 \\ -3 & 2 \end{pmatrix} = \frac{1}{13}\begin{pmatrix} -5 & 12 \\ 12 & 5 \end{pmatrix}.$$

◇

Solution 2 (reduce the problem to a known reflection). The problem would be easy if we were asked to calculate the matrix representation of the reflection on the x-axis. This would simply be $A_0 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$. Now we can proceed as follows: First we rotate $\mathbb{R}^2$ about the origin such that the line L is parallel to the x-axis, then we reflect on the x-axis and then we rotate back. The result is the same as reflecting on L. Assume that Rot is the rotation matrix. Then

$$A_T = \operatorname{Rot}^{-1} \circ A_0 \circ \operatorname{Rot}. \tag{5.19}$$

How can we calculate Rot? We know that $\operatorname{Rot}\vec{v} = \vec{\mathrm{e}}_1$ and that $\operatorname{Rot}\vec{w} = \vec{\mathrm{e}}_2$. It follows that $\operatorname{Rot}^{-1} = (\vec{v}|\vec{w}) = \left(\begin{smallmatrix} 2 & 3 \\ -3 & 2 \end{smallmatrix}\right)$. Note that up to a numerical factor, this is $S_{\mathcal{B}\to can}$. We can calculate easily that $\operatorname{Rot} = (\operatorname{Rot}^{-1})^{-1} = \frac{1}{13}\left(\begin{smallmatrix} 2 & -3 \\ 3 & 2 \end{smallmatrix}\right)$. If we insert this in (5.19), we find again $A_R = \left(\begin{smallmatrix} -5 & 12 \\ 12 & 5 \end{smallmatrix}\right)$. ◇

Solution 3 (straight forward calculation). Lastly, we can form a system of linear equations in order to find A_T. We write $A_R = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$ with unknown numbers a, b, c, d. Again, we use that we know that $A_T\vec{v} = \vec{v}$ and $A_T\vec{w} = -\vec{w}$. This gives the following equations:

$$\begin{pmatrix} 2 \\ 3 \end{pmatrix} = \vec{v} = A_T\vec{v} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}\begin{pmatrix} 2 \\ 3 \end{pmatrix} = \begin{pmatrix} 2a + 3b \\ 2c + 3d \end{pmatrix},$$
$$\begin{pmatrix} -3 \\ 2 \end{pmatrix} = \vec{w} = -A_T\vec{w} = -\begin{pmatrix} a & b \\ c & d \end{pmatrix}\begin{pmatrix} -3 \\ 2 \end{pmatrix} = \begin{pmatrix} 3a - 2b \\ 3c - 2d \end{pmatrix}$$

which gives the system

$$2a + 3b = 2, \quad 2c + 3d = 3, \quad 3a - 2b = -3, \quad 3c - 2d = 2,$$

Its unique solution is $a = -\frac{5}{13}$, $b = c = \frac{12}{13}$, $d = \frac{5}{13}$, hence $A_R = \left(\begin{smallmatrix} -5 & 12 \\ 12 & 5 \end{smallmatrix}\right)$. ◇

Example 5.43 (Reflection and orthogonal projection in $\mathbb{R}^3$). In $\mathbb{R}^3$, consider the plane $E : x - 2y + 3z = 0$. Let $R : \mathbb{R}^3 \to \mathbb{R}^3$ which takes a vector in $\mathbb{R}^3$ and reflects it on the plane E and let $P : \mathbb{R}^3 \to \mathbb{R}^3$ be the orthogonal projection onto E. Find the matrix representation of R with respect to the standard basis of $\mathbb{R}^E$.

Observation. Note that E is the line which passes through the origin and is orthogonal to the vector $\vec{n} = \left(\begin{smallmatrix} 1 \\ -2 \\ 3 \end{smallmatrix}\right)$. Moreover, if we set $\vec{v} = \left(\begin{smallmatrix} 2 \\ 1 \\ 0 \end{smallmatrix}\right)$ and $\vec{w} = \left(\begin{smallmatrix} 0 \\ 3 \\ 2 \end{smallmatrix}\right)$, then it is easy to see that $\{\vec{v}, \vec{w}\}$ is a basis of E.

Solution 1 (use coordinates adapted to the problem). Clearly, a basis which is adapted to the exercise, is $\mathcal{B} = \{\vec{n},\ \vec{v},\ \vec{w}\}$ because for these vectors we have $R\vec{v} = \vec{v}$, $R\vec{w} = \vec{w}$ and $P\vec{v} = \vec{v}$, $P\vec{w} = \vec{w}$ and $P\vec{n} = \vec{0}$. Therefore the matrix representation of R with respect to the basis $\mathcal{B}$ is

$$A_R^{\mathcal{B}} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{pmatrix}$$

and the one of P is

$$A_R^{\mathcal{B}} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

In order to obtain the representations A_R and A_P with respect to the standard basis, we only need to perform a change of basis. Recall that change-of-bases matrices are given by

$$S_{\mathcal{B}\to can} = (\vec{v}|\vec{w}|\vec{n}) = \begin{pmatrix} 2 & 0 & 1 \\ 1 & 3 & -2 \\ 0 & 2 & 3 \end{pmatrix}, \qquad S_{can\to\mathcal{B}} = S_{\mathcal{B}\to can}^{-1} = \frac{1}{28}\begin{pmatrix} 13 & 2 & -3 \\ -3 & 6 & 5 \\ 2 & -4 & 6 \end{pmatrix}.$$

Therefore

$$\begin{aligned} A_R = S_{\mathcal{B}\to can} A_R^{\mathcal{B}} S_{can\to\mathcal{B}} &= \frac{1}{28}\begin{pmatrix} 2 & 0 & 1 \\ 1 & 3 & -2 \\ 0 & 2 & 3 \end{pmatrix}\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{pmatrix}\begin{pmatrix} 13 & 2 & -3 \\ -3 & 6 & 5 \\ 2 & -4 & 6 \end{pmatrix} \\ &= \frac{1}{7}\begin{pmatrix} 6 & 2 & -3 \\ 2 & 3 & 6 \\ -3 & 6 & -2 \end{pmatrix} \end{aligned}$$

and

$$\begin{aligned} A_P = S_{\mathcal{B}\to can} A_P^{\mathcal{B}} S_{can\to\mathcal{B}} &= \frac{1}{28}\begin{pmatrix} 2 & 0 & 2 \\ 1 & 3 & -1 \\ 0 & 2 & 3 \end{pmatrix}\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}\begin{pmatrix} 13 & 2 & -3 \\ -3 & 6 & 5 \\ 2 & -4 & 6 \end{pmatrix} \\ &= \frac{1}{14}\begin{pmatrix} 13 & 2 & -3 \\ 2 & 10 & 6 \\ -3 & 6 & 5 \end{pmatrix} \end{aligned}$$

◇

Solution 2 (reduce the problem to a known reflection). The problem would be easy if we were asked to calculate the matrix representation of the reflection on the *xy*-plane. This would simply be $A_0 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{pmatrix}$. Now we can proceed as follows: First we rotate $\mathbb{R}^3$ about the origin such that the plane E is parallel to the *xy*-axis, then we reflect on the *xy*-plane and then we rotate back. The result is the same as reflecting on the plane E. We leave the details to the reader. An analogous procedure works for the orthogonal projection. ◇

Solution 3 (straight forward calculation). Lastly, we can form a system of linear equations in order to find A_R. We write $A_R = \left(\begin{smallmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{smallmatrix}\right)$ with unknowns a_{ij}. Again, we use that we know that $A_R\vec{v} = \vec{v}$, $A_R\vec{w} = \vec{w}$ and $A_R\vec{n} = -\vec{n}$. This gives a system of 9 linear equations for the nine unknowns a_{ij} which can be solved. ◇

Remark 5.44. Yet another solution is the following. Let Q be the orthogonal projection onto $\vec{n}$. We already know how to calculate its representing matrix:

$$Q\vec{x} = \frac{\langle \vec{x}\,, \vec{n}\rangle}{\|\vec{n}\|^2}\vec{n} = \frac{x - 2y + 3z}{14}\vec{n} = \frac{1}{14}\begin{pmatrix} 1 & -2 & 3 \\ -2 & 4 & -6 \\ 3 & -6 & 9 \end{pmatrix}\begin{pmatrix} x \\ y \\ z \end{pmatrix}.$$

Hence $A_Q = \frac{1}{14}\left(\begin{smallmatrix} 1 & -2 & 3 \\ -2 & 4 & -6 \\ 3 & -6 & 9 \end{smallmatrix}\right)$. Geometrically, it is clear that $P = \operatorname{id} - Q$ and $R = \operatorname{id} - 2Q$. Hence it follows that

$$A_P = \operatorname{id} - A_Q = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} - \frac{1}{14}\begin{pmatrix} 1 & -2 & 3 \\ -2 & 4 & -6 \\ 3 & -6 & 9 \end{pmatrix} = \frac{1}{14}\begin{pmatrix} 13 & 2 & -3 \\ 2 & 10 & 6 \\ -3 & 6 & 5 \end{pmatrix}$$

and

$$A_R = \operatorname{id} - 2A_Q = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} - \frac{1}{7}\begin{pmatrix} 1 & -2 & 3 \\ -2 & 4 & -6 \\ 3 & -6 & 9 \end{pmatrix} = \frac{1}{7}\begin{pmatrix} 6 & 2 & -3 \\ 2 & 3 & 6 \\ -3 & 6 & -2 \end{pmatrix}.$$

Index