Linear Algebra for the Sciences

Thomas Kappeler, Riccardo Montalto

Contents

Chapter 1

Systems of linear equations

At the core of linear algebra is the solving of systems of linear equations. A linear system (S) with n unknowns and m equations has the form

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = b_1 & (Eq_1) \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n = b_2 & (Eq_2) \\ \vdots & \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n = b_m & (Eq_m) \end{cases}$$

The numbers $a_{11}, \ldots, a_{1n}$, $a_{21}, \ldots, a_{2n}$, $\ldots$, $a_{m1}, \ldots, a_{mn}$ are called the coefficients of the system. The system is said to be real if all coefficients a_{ij}, $i = 1, \ldots, n$, $j = 1, \ldots, m$ as well as the numbers $b_1, \ldots, b_m$ are real. In this chapter we will only consider real systems of linear equations, but the techniques we develop to solve them also apply to complex systems, i.e., systems where a_{ij} and b_i are complex numbers which will only be introduced in a later chapter. Given the real system (S), a *solution* of (S) is a set of n real numbers $x_1, \ldots, x_n$ so that equations $(Eq_1) - (Eq_m)$ are satisfied simultaneously. The basic questions with regard to linear systems of the form (S) are the following ones:

(1) Does (S) have a solution? (Existence)

(2) Does (S) have at most 1 solution? (Uniqueness)

Or formulated in more general terms:

(3) What are the properties of the set of solutions

$$L := \Big\{(x_1, \ldots, x_n) : x_1 \in \mathbb{R}, \ldots, x_n \in \mathbb{R}; \quad (x_1, \ldots, x_n) \quad \text{satisfies} \quad (S_1) - (S_m)\Big\}.$$

Here and in the sequel, $\mathbb{R}$ denotes the set of real numbers. Note that in this terminology questions (1) and (2) can be reformulated as follows:

(1') Is L a nonempty set?

(2') Does L have at most one element?

In practical applications, systems of linear equations can be very large. One therefore needs theoretical concepts and numerical algorithms to investigate respectively solve such systems.

1.1 Linear systems with two equations and two unknowns

As an introduction we consider in this section real systems with two equations and two unknowns:

$$\begin{cases} a_{11}x_1 + a_{12}x_2 = b_1 \\ a_{21}x_1 + a_{22}x_2 = b_2 \,. \end{cases}$$

In such a case, we usually choose a simpler notation: $x_1 \rightsquigarrow x$, $x_2 \rightsquigarrow y$

$$\begin{cases} ax + by = e \\ cx + dy = f \,. \end{cases}$$

where a, b, c, d, e, f are in $\mathbb{R}$. To get some experience, let us first treat the simpler case of one linear equation with one unknown

$$ax = b \tag{1.1.1}$$

where $a, b \in \mathbb{R}$. The solvability of this equation depends on the values of a and b:
Case $a \neq 0$: equation (1) has precisely one solution

$$x = \frac{b}{a} \in \mathbb{R} \quad \text{or} \quad L = \left\{ \frac{b}{a} \right\}.$$

Case $a = 0$ *and* $b \neq 0$: equation (1) has no solution, i.e., $L = \emptyset$.
Case $a = 0$ *and* $b = 0$: any real number $x \in \mathbb{R}$ is a solution of (1), i.e., $L = \mathbb{R}$.

Next let us treat the case of one equation and two unknowns:

$$ax + by = c\,, \quad a, b, c \in \mathbb{R} \tag{1.1.2}$$

As above we denote the set of solutions by L,

$$L = \left\{ (x, y) \in \mathbb{R}^2 : ax + by = c \right\}.$$

Case $a = 0, b = 0, c = 0$: $L = \mathbb{R}^2$
Case $a = 0, b = 0, c \neq 0$: $L = \emptyset$
Case $(a, b) \neq (0, 0)$: the set L is a straight line in $\mathbb{R}^2$. In case $b \neq 0$, this straight line is given by

$$y = -\frac{a}{b}x + \frac{c}{b}$$

and

$$L = \left\{ \left(x, -\frac{a}{b}x + \frac{c}{b} \right) : x \in \mathbb{R} \right\}.$$

In case $b = 0$ and $a \neq 0$,

$$L = \left\{ \left(\frac{c}{a}, y \right) : y \in \mathbb{R} \right\}.$$

Now let us go back to the linear system of two equations with two unknowns,

$$ax + by = e \tag{1.1.3}$$

$$cx + dy = f \tag{1.1.4}$$

The set of solutions L of (1.1.3), (1.1.4) is then given by the *intersection* of the set of solutions of (1.1.3) with the set of solutions of (1.1.3), $L = L_1 \cap L_2$, where

$$L_1 = \Big\{(x, y) \in \mathbb{R}^2 : ax + by = e\Big\}\,, \quad L_2 = \Big\{(x, y) \in \mathbb{R}^2 : cx + dy = f\Big\}\,.$$

In case $(a, b) \neq (0, 0)$ and $(c, d) \neq (0, 0)$, L_1 and L_2 are lines in $\mathbb{R}^2$ and L is the intersection of them. Thus L can be a one point set (L_1 and L_2 are not parallel), or a line ($L_1 = L_2$) or the emptyset (L_1 and L_2 are parallel but do not coincide).

Let us now describe an algorithm how to determine the set of solutions of (1.1.3), (1.1.4) in a systematic way. You know this algorithm already from high school. To simplify the algorithm we assume that

$$a \neq 0\,. \tag{1.1.5}$$

STEP 1: eliminate x from equation (1.1.4) by replacing (1.1.4) by (1.1.4) $-\frac{c}{a}$(1.1.3) It means that the left hand side of (1.1.4) is replaced by

$$cx + dy - \frac{c}{a}\Big(ax + by\Big) \quad (\text{use that } a \neq 0!)$$

whereas the right hand side of (1.1.4) is replaced by

$$f - \frac{c}{a}e\,.$$

The new system of equations then reads as follows

$$ax + by = e \tag{1.1.6}$$

$$\Big(d - \frac{c}{a}b\Big)y = f - \frac{c}{a}e\,. \tag{1.1.7}$$

It is straightforward to see that under the assumption (1.1.5), the set solutions of (1.1.3), (1.1.4) coincides with the set of solutions of (1.1.6), (1.1.7). In such a case we say that the two systems are equivalent.

STEP 2: The system (1.1.6), (1.1.7) is solved by first solving (1.1.7) for y and then use (1.1.6) to determine x:
Case $d - \frac{c}{a}b \neq 0$*:* then (1.1.7) has the unique solution

$$y = \frac{f - \frac{c}{a}e}{d - \frac{c}{a}b} = \frac{af - ce}{ad - bc}$$

and when substituted into (1.1.6) one obtains

$$ax = e - b\Big(\frac{af - ce}{ad - bc}\Big)$$

or

$$x = \frac{de - bf}{ad - bc}\,.$$

Hence the set of solutions L consists of one element

$$L = \Big\{ \Big(\frac{de - bf}{ad - bc}, \frac{af - ce}{ad - bc} \Big) \Big\} \tag{1.1.8}$$

Case $d - \frac{c}{a}b = 0\,,\ f - \frac{c}{a}e \neq 0$: equation (1.1.7) has no solutions and hence $L = \emptyset$.
Case $d - \frac{c}{a}b = 0\,,\ f - \frac{c}{a}e = 0$: then any $y \in \mathbb{R}$ is a solution of (1.1.7) and solutions of (1.1.6) are given by $x = \frac{e}{a} - \frac{b}{a}y$. Hence the set of solutions L is given by

$$L = \Big\{ \Big(\frac{e}{a} - \frac{b}{a}y, y \Big) : y \in \mathbb{R} \Big\}.$$

Motivated by formula (1.1.8) for the solutions of (1.1.6), (1.1.7) in the case where $d - \frac{c}{a}b \neq 0$ we make the following definitions:

Definition 1.1.1. (*i*) *A real* 2×2 *matrix (plural: matrices) is an array* A *of real numbers of the form*

$$A := \begin{pmatrix} a & b \\ c & d \end{pmatrix}\,, \quad a, b, c, d \in \mathbb{R}\,;$$

(*ii*) *the determinant of a* 2×2 *matrix* A *is defined as*

$$\det(A) := ad - bc\,.$$

The notion of the determinant can be used to characterize the solvability of the system (1.1.6), (1.1.7) and to obtain formulas for its solutions. We state without proof the following

Theorem 1.1.1. (*i*) *The system of linear equations* (1.1.6), (1.1.7) *has a unique solution if and only if*

$$\det \begin{pmatrix} a & b \\ c & d \end{pmatrix} \neq 0\,.$$

(*ii*) *If* $\det \begin{pmatrix} a & b \\ c & d \end{pmatrix} \neq 0$, *then the unique solution of* (1.1.6), (1.1.7) *is given by the following formulas (Cramer's rule)*

$$x = \frac{\det \begin{pmatrix} e & b \\ f & d \end{pmatrix}}{\det \begin{pmatrix} a & b \\ c & d \end{pmatrix}}\,, \qquad y = \frac{\det \begin{pmatrix} a & e \\ c & f \end{pmatrix}}{\det \begin{pmatrix} a & b \\ c & d \end{pmatrix}}$$

PROBLEM: Analyze the following system of linear equations

$$\begin{cases} 2y + 4x = 3 \\ -x + y = 5 \end{cases}$$

and find its set of solutions.
ANSWER: Let A denote the coefficient matrix,

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

Then $\det(A) = 6 \neq 0$. Hence according to Theorem 1.1.1, the system of equations has a unique solution given by

$$x = \frac{\det \begin{pmatrix} 3 & 2 \\ 5 & 1 \end{pmatrix}}{6} = -\frac{7}{6}, \qquad y = \frac{\det \begin{pmatrix} 4 & 3 \\ -1 & 5 \end{pmatrix}}{6} = \frac{23}{6}$$

1.2 Gaussian elimination

Gaussian elimination is an algorithm to determine the set of solutions of an arbitrary system of linear equations with m equations and n unknowns, denoted by $x_1, \ldots, x_n$,

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

where $a_{ij} \in \mathbb{R}$, $1 \leq i \leq m$ (index for the equations) and $1 \leq j \leq n$ (index for the unknowns). A compact way of writing the above system of equations is achieved using the symbol $\sum$ for the sum,

$$\sum_{j=1}^{n} a_{ij}x_j = b_i \quad \text{for any} \quad 1 \leq i \leq m\,. \tag{1.2.1}$$

Of course we could also use different letters than i and j. We denote by L the set of solutions of (1.2.1),

$$L := \left\{ (x_1, \ldots, x_n) \in \mathbb{R}^n : \sum_{j=1}^{n} a_{ij}x_j = b_i \quad \text{for any} \quad 1 \leq i \leq m \right\}.$$

Note that

$$L = \bigcap_{i=1}^{m} L_i\,, \quad L_i := \left\{ (x_1, \ldots, x_n) \in \mathbb{R}^n : \sum_{j=1}^{n} a_{ij}x_j = b_i \right\}.$$

Definition 1.2.1. *We say that two systems of linear equations*

$$\sum_{j=1}^{n} a_{ij}x_j = b_i \quad \textit{for} \quad \textit{any} \quad 1 \leq i \leq m$$

and

$$\sum_{j=1}^{n} c_{kj}x_j = d_k \quad \textit{for} \quad \textit{any} \quad 1 \leq k \leq p$$

are equivalent if their sets of solutions coincide.

An example of two equivalent systems with three unknowns is the following one:

$$\begin{cases} 4x_1 + 3x_2 + 2x_3 = 1 \\ x_1 + x_2 + x_3 = 4 \end{cases}$$

and

$$\begin{cases} 4x_1 + 3x_2 + 2x_3 = 1 \\ x_1 + x_2 + x_3 = 4 \\ 2x_1 + 2x_2 + 2x_3 = 8 \end{cases}$$

since the latter equation is obtained from the equation $x_1 + x_2 + x_3 = 4$ by multiplying left and right hand side by the factor 2.

The idea of Gaussian elimination is to replace a given system of linear equations in a systematic way by an equivalent one which is easy to solve. In Subsection 1.1 we have demonstrated this method in the case of two equations ($m = 2$) and two unknowns ($n = 2$). Gaussian elimination uses the following basic operations, referred to as row operations, which leave the set of solutions of a given system of linear equations invariant:

(R1) Exchange of two equations (rows) of a system of linear equations.

EXAMPLE:

$$\begin{cases} 5x_2 + 15x_3 = 10 \\ 4x_1 + 3x_2 + x_3 = 1 \end{cases} \quad \rightsquigarrow \quad \begin{cases} 4x_1 + 3x_2 + x_3 = 1 \\ 5x_2 + 15x_3 = 10 \end{cases}$$

It means that the equations get listed in a different order.

(R2) Multiplication of an equation (row) by a real number $\alpha \neq 0$.

EXAMPLE:

$$\begin{cases} 4x_1 + 3x_2 + x_3 = 1 \\ 5x_2 + 15x_3 = 10 \end{cases} \quad \rightsquigarrow \quad \begin{cases} 4x_1 + 3x_2 + x_3 = 1 \\ x_2 + 3x_3 = 2 \end{cases}$$

We have multiplied the left and right hand side of the second equation by the factor $1/5$.

(R3) An equation (row) gets replaced by the equation obtained by adding to it the multiple of another equation. More formally, this can be expressed as follows: the kth equation $\sum_{j=1}^{n} a_{kj}x_j = b_k$ is replaced by the equation

$$\sum_{j=1}^{n} a_{kj}x_j + \alpha \sum_{j=1}^{n} a_{\ell j}x_j = b_k + \alpha b_\ell$$

for some where $1 \leq \ell \leq m$ with $\ell \neq k$ – or more explicitly,

$$(a_{k1} + \alpha a_{\ell 1})x_1 + \cdots + (a_{kn} + \alpha a_{\ell n})x_n = b_k + \alpha b_\ell \, .$$

EXAMPLE:

$$\begin{cases} x_1 + x_2 = 5 & (Eq1) \\ 4x_1 + 2x_2 = 3 & (Eq2) \end{cases} \quad \overset{(Eq2)\rightsquigarrow(Eq2)-4(Eq1)}{\rightsquigarrow} \quad \begin{cases} x_1 + x_2 = 5 \\ -2x_2 = -17 \end{cases}$$

It is not difficult to verify that these basic row operations lead to equivalent linear systems. We state without proof the following

Theorem 1.2.1. *The basic row operations lead to equivalent systems of linear equations.*

We now show how these basic row operations can be used to solve a system (S) of linear equations of the form

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

where a_{ij} $(1 \leq i \leq m \, , \, 1 \leq j \leq n)$ and b_i $(1 \leq i \leq m)$ are real numbers. To (S) we associate its coefficient matrix

$$A = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix} .$$

A is an array of real numbers with m rows and n columns. Such an array of real numbers is called an $m \times n$ matrix and written in a compact form as

$$A = (a_{ij})_{\substack{1 \leq i \leq m \\ 1 \leq j \leq n}}$$

The augmented coefficient matrix of (S) is the following array of real numbers

$$\begin{pmatrix} a_{11} & \cdots & a_{1n} & b_1 \\ \vdots & & \vdots & \vdots \\ a_{m1} & \cdots & a_{mn} & b_m \end{pmatrix}$$

written in a compact form as $(A \,|\, b)$.

Definition 1.2.2. $(A\,|\,b)$ *is said to be in* row echelon form *(Zeilenstufenform) if it is of the form*

$$\left(\begin{array}{ccccc|c} \underline{\star} & \cdot & \cdot & \cdot & & \\ 0 & \underline{\star} & \cdot & \vdots & & \\ 0 & 0 & \underline{\star} & \vdots & & \\ \vdots & & & & & b \\ 0 & \cdots & & \underline{\star} & & \\ 0 & \cdots & & 0 & \underline{\star} & \\ 0 & \cdots & & & 0 & \\ 0 & \cdots & & & 0 & \end{array}\right)$$

where $\star$ stands for nonzero elements of A.

Note that below the echelon (Stufe), all the coefficients of A vanish. We point out that it is possible that the coefficient matrix A does not have any zero columns (Spalten). Furthermore, if it has zero rows (Zeilen), then these rows have to be at the bottom of A. Examples of augmented coefficient matrices in row echelon form are

$$\left(\begin{array}{cc|c} 0 & 1 & 1 \\ 0 & 0 & 0 \end{array}\right), \quad \left(\begin{array}{cc|c} 1 & 0 & 0 \\ 0 & 0 & 1 \end{array}\right), \quad \left(\begin{array}{ccc|c} 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{array}\right)$$

$$\left(\begin{array}{ccc|c} 4 & 0 & 0 & 1 \\ 0 & 3 & 0 & 5 \end{array}\right), \qquad \left(\begin{array}{ccc|c} 4 & 5 & 2 & 1 \\ 0 & 0 & 3 & 5 \end{array}\right)$$

whereas the following ones are *not* in this form

$$\left(\begin{array}{cc|c} 1 & 0 & 1 \\ 1 & 1 & 1 \end{array}\right), \quad \left(\begin{array}{cc|c} 0 & 1 & 1 \\ 1 & 1 & 1 \end{array}\right), \quad \left(\begin{array}{cc|c} 0 & 0 & 1 \\ 0 & 1 & 1 \end{array}\right)$$

If the augmented coefficient matrix of a linear system is in row echelon form, it can easily be solved. To illustrate this let us look at few examples.

Examples:

(1)
$$\begin{cases} 2x_1 + x_2 = 2 \\ 3x_2 = 6 \end{cases} \qquad \rightsquigarrow \qquad \left(\begin{array}{cc|c} 2 & 1 & 2 \\ 0 & 3 & 6 \end{array}\right)$$

is in row echelon form. Solve it by first determining x_2 by the second equation and then solving the first equation for x_1 by substituting the obtained value of x_2

$$3x_2 = 6 \quad \rightsquigarrow \quad x_2 = 2$$

$$2x_1 = 2 - x_2 \quad \rightsquigarrow \quad 2x_1 = 0 \quad \rightsquigarrow \quad x_1 = 0$$

hence the set of solutions is given by $L = \{(0, 2)\}$.

(2)
$$\begin{cases} 2x_1 + x_2 + x_3 = 2 \\ 3x_3 = 6 \end{cases} \qquad \rightsquigarrow \qquad \left(\begin{array}{ccc|c} 2 & 1 & 1 & 2 \\ 0 & 0 & 3 & 6 \end{array}\right)$$
is in row echelon form. Solve the second equation for x_3 and then the first equation for x_1:
$$3x_3 = 6 \quad \rightsquigarrow \quad x_3 = 2\,; \quad x_2 \text{ is a free variable;}$$
$$2x_1 = 2 - x_2 - x_3 \quad \rightsquigarrow \quad x_1 = -\frac{1}{2}x_2$$
and hence
$$L = \left\{(-\frac{1}{2}x_2, x_2, 2) : x_2 \in \mathbb{R}\right\},$$
which is a straight line in $\mathbb{R}^3$ trough $(0, 0, 2)$ in direction $(-1, 2, 0)$.

(3) The augmented coefficient matrix $\left(\begin{array}{cc|c} 2 & 1 & 1 \\ 0 & 3 & 6 \\ 0 & 0 & 3 \end{array}\right)$ is in row echelon form. The corresponding system of linear equations can be solved as follows: since
$$0 \cdot x_1 + 0 \cdot x_2 = 3$$
has no solutions, one concludes that $L = \emptyset$.

(4) The augmented coefficient matrix $\left(\begin{array}{cc|c} 2 & 1 & 1 \\ 0 & 3 & 6 \\ 0 & 0 & 0 \end{array}\right)$ is in row echelon form. The corresponding system of linear equations can be solved as follows:
$$0 \cdot x_1 + 0 \cdot x_2 = 0$$
is satisfied for any $x_1, x_2 \in \mathbb{R}$;
$$3x_2 = 6 \quad \rightsquigarrow \quad x_2 = 2$$
$$2x_1 = 1 - x_2 \quad \rightsquigarrow \quad x_1 = -\frac{1}{2}\,.$$
Hence $L = \{(-1, 2)\}$.

(5) The augmented coefficient matrix $\left(\begin{array}{ccc|c} 0 & 1 & 2 & 1 \\ 0 & 0 & 1 & 6 \end{array}\right)$ is in row echelon form.
$$x_3 = 6\,, \quad x_2 = 1 - 2x_3 = -11\,, \quad x_1 \text{ free variable}\,,$$
hence
$$L = \left\{(x_1, -11, 6) : x_1 \in \mathbb{R}\right\}$$
which is a straight line in $\mathbb{R}^3$ trough the point $(0, -11, 6)$ in direction $(1, 0, 0)$.

(6) The augmented coefficient matrix $\left(\begin{array}{cccc|c} 1 & 2 & 0 & 1 & 0 \\ 0 & 0 & 0 & 3 & 6 \end{array}\right)$ is in row echelon form.

$$3x_4 = 6 \quad \rightsquigarrow \quad x_4 = 2\,; \quad x_3 \text{ and } x_4 \text{ are a free variables}\,;$$
$$x_1 = -2x_2 - x_4 = -2x_2 - 2\,,$$

hence

$$L = \Big\{(-2x_2 - 2, x_2, x_3, 2) : x_2, x_3 \in \mathbb{R}\Big\}$$

which is a plane in $\mathbb{R}^4$ containing the point $(-2, 0, 0, 2)$ and spanned by the vectors $(0, 0, 1, 0)$ and $(-2, 1, 0, 0)$.

(7) The augmented coefficient matrix $\left(\begin{array}{cccc|c} 1 & 2 & 1 & 0 & 0 \\ 0 & 0 & 3 & 0 & 6 \end{array}\right)$ is in row echelon form.

$$x_4 \text{ is a free variable}\,; \quad 3x_3 = 6 \quad \rightsquigarrow \quad x_3 = 2$$
$$x_1 = -2x_2 - x_3 = -2x_2 - 2\,.$$

Hence

$$L = \Big\{(-2x_2 - 2, x_2, 2, x_4) : x_2, x_4 \in \mathbb{R}\Big\}$$

which is a plane in $\mathbb{R}^4$ containing the point $(-2, 0, 2, 0)$ and spanned by the vectors $(0, 0, 0, 1)$ and $(-2, 1, 0, 0)$.

Gaussian elimination is a method of transforming a given augmented coefficient matrix with the help of basic row operations into row echelon form. How can this be achieved? We explain the procedure with a few examples. It is convenient to introduce for the three basic row operations the following notations:

- $R_{i \leftrightarrow k}$: exchange rows i and k
- $R_k \quad \rightsquigarrow \quad \alpha R_k$: replace k-th row R_k by αR_k, $\alpha \neq 0$.
- $R_k \quad \rightsquigarrow \quad R_k + \alpha R_\ell$: replace k-th row by adding to it αR_ℓ where $\ell \neq k$ and $\alpha \in \mathbb{R}$.

Examples

(1) The augmented coefficient matrix $\left(\begin{array}{cc|c} 0 & 3 & 6 \\ 2 & 1 & 2 \end{array}\right)$ is not in row echelon form. Apply $R_{1 \leftrightarrow 2}$ to get

$$\left(\begin{array}{cc|c} 2 & 1 & 2 \\ 0 & 3 & 6 \end{array}\right)$$

(2) The augmented coefficient matrix $\left(\begin{array}{cc|c} 2 & 1 & 1 \\ 4 & 3 & 0 \end{array}\right)$ is not in row echelon form. The first row is ok; in the second row we have to eliminate 4; hence

$$R_2 \quad \rightsquigarrow \quad R_2 - 2R_1$$

yielding

$$\left(\begin{array}{cc|c} 2 & 1 & 1 \\ 0 & 1 & -2 \end{array}\right)$$

(3) The augmented coefficient matrix $\left(\begin{array}{ccc|c} 1 & 1 & 1 & 1 \\ 2 & 1 & 1 & 0 \\ 4 & 1 & 2 & 0 \end{array}\right)$ is not in row echelon form. The first row is ok; in second and third row we have to eliminate 2 respectively 4. Hence $R_2 \rightsquigarrow R_2 - 2R_1$ and $R_3 \rightsquigarrow R_3 - 4R_1$

$$\left(\begin{array}{ccc|c} 1 & 1 & 1 & 1 \\ 0 & -1 & -1 & -2 \\ 0 & -3 & -2 & -4 \end{array}\right)$$

Now R_1 and R_2 are ok, but we need to eliminate -3 from the last row. Hence $R_3 \rightsquigarrow R_3 - 3R_2$, yielding

$$\left(\begin{array}{ccc|c} 1 & 1 & 1 & 1 \\ 0 & -1 & -1 & -2 \\ 0 & 0 & 1 & 2 \end{array}\right)$$

which is in row echelon form.

(4) The augmented coefficient matrix

$$\left(\begin{array}{ccccc|c} 1 & 1 & 1 & 1 & 1 & 1 \\ -1 & -1 & 0 & 0 & 1 & -1 \\ -2 & -2 & 0 & 0 & 3 & 1 \\ 0 & 0 & 1 & 1 & 3 & -1 \\ 1 & 1 & 2 & 2 & 4 & 1 \end{array}\right)$$

is not in row echelon form. R_1 is ok, but we need to eliminate the first coefficients from subsequent rows:

$$R_2 \rightsquigarrow R_2 + R_1\,, \qquad R_3 \rightsquigarrow R_3 + 2R_1\,, \qquad R_5 \rightsquigarrow R_5 - R_1\,,$$

yielding

$$\left(\begin{array}{ccccc|c} 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 2 & 0 \\ 0 & 0 & 2 & 2 & 5 & 3 \\ 0 & 0 & 1 & 1 & 3 & -1 \\ 0 & 0 & 1 & 1 & 3 & 0 \end{array}\right)$$

Rows R_1, R_2 are ok, but we need to eliminate the third coefficients in the rows R_3, R_4, R_5.

$$R_3 \rightsquigarrow R_3 - 2R_2\,, \qquad R_4 \rightsquigarrow R_4 - R_2\,, \qquad R_5 \rightsquigarrow R_5 - R_2\,,$$

yielding

$$\left(\begin{array}{ccccc|c} 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 2 & 0 \\ 0 & 0 & 0 & 0 & 1 & 3 \\ 0 & 0 & 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{array}\right)$$

Now R_1, R_2, R_3 are ok, but we need to eliminate the last coefficients in R_4 and R_5, i.e.,

$$R_4 \rightsquigarrow R_4 - R_3\,, \qquad R_5 \rightsquigarrow R_5 - R_3\,,$$

leading to

$$\left(\begin{array}{ccccc|c} 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 2 & 0 \\ 0 & 0 & 0 & 0 & 1 & 3 \\ 0 & 0 & 0 & 0 & 0 & -4 \\ 0 & 0 & 0 & 0 & 0 & -3 \end{array}\right)$$

which is in row echelon form.

(5) The augmented coefficient matrix $\left(\begin{array}{ccc|c} 1 & 1 & 1 & 0 \\ -1 & -1 & 0 & 0 \\ -2 & 1 & 0 & 1 \end{array}\right)$ is not in row echelon form.

R_1 is ok, but we need to eliminate the first coefficients in R_2, R_3, i.e.,

$$R_2 \rightsquigarrow R_2 + R_1\,, \qquad R_3 \rightsquigarrow R_3 + 2R_1\,,$$

yielding

$$\left(\begin{array}{ccc|c} 1 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 3 & 2 & 1 \end{array}\right)$$

To bring the lattest augmented coefficient matrix in row echelon form we need to exchange the second and the third row, $R_{2\leftrightarrow 3}$, leading to

$$\left(\begin{array}{ccc|c} 1 & 1 & 1 & 0 \\ 0 & 3 & 2 & 1 \\ 0 & 0 & 1 & 0 \end{array}\right)$$

which is in row echelon form.

We now want to go one step further and describe the set of solutions of a system of linear equations in more detail. We begin by making some preliminary considerations. Consider the system (S)

$$\begin{cases} x_1 + 4x_2 = b_1 \\ 5x_1 + 2x_2 = b_2\,. \end{cases}$$

The corresponding augmented coefficient matrix is given by

$$(A\,|\,b) = \left(\begin{array}{cc|c} 1 & 4 & b_1 \\ 5 & 2 & b_2 \end{array}\right)\,.$$

Let us compare it with the system $(\widetilde{S})$, obtained by exchanging the two columns of A. Introducing as new unknowns y_1, y_2 this system reads

$$\begin{cases} 4y_1 + y_2 = b_1 \\ 2y_1 + 5y_2 = b_2 \end{cases}$$

and the corresponding augmented coefficient matrix is given by $(\widetilde{A}\,|\,b)$ where

$$\widetilde{A} = \begin{pmatrix} 4 & 1 \\ 2 & 5 \end{pmatrix}.$$

Denote by L and $\widetilde{L}$ the set of solutions of (S) respectively $(\widetilde{S})$. It is easy to see that the map $(x_1, x_2) \mapsto (y_1, y_2) := (x_2, x_1)$ induces a bijection between L and $\widetilde{L}$. It means that any solution (x_1, x_2) of (S) leads to the solution $y_1 := x_2$, $y_2 := x_1$ of $(\widetilde{S})$ and conversely, any solution (y_1, y_2) of $(\widetilde{S})$ leads to a solution $x_1 := y_2$, $x_2 := y_1$ of (S), or said differently, by renumerating the unknowns x_1, x_2, we can read off the set of solutions of $(\widetilde{S})$ from the one of (S). This procedure can be used to bring an augmented coefficient matrix $(A\,|\,b)$ in row echelon form into an even simpler form: assume that A has m rows and n columns. By renumerating the unknowns, which corresponds to a permutation of the columns of A, $(A\,|\,b)$ can be brought into the echelon form $(\widetilde{A}\,|\,b)$,

$$\widetilde{A} = \left(\begin{array}{ccccccc|c} \underline{\widetilde{a}_{11}} & & & & & & & \\ 0 & |\underline{\widetilde{a}_{22}} & & & & & & \\ 0 & 0 & |\underline{\widetilde{a}_{33}} & & & & & \\ \vdots & & & & & & & \\ 0 & \cdots & & & |\underline{\widetilde{a}_{kk}} & & & b \\ 0 & \cdots & & & & 0 & | & \\ \vdots & & & & & & & \\ 0 & \cdots & & & & 0 & | & \end{array}\right)$$

where k is an integer with $0 \leq k \leq \min(m, n)$ and $\widetilde{a}_{11} \neq 0$, $\widetilde{a}_{22} \neq 0, \ldots, \widetilde{a}_{kk} \neq 0$. If $k = 0$, then $\widetilde{A}$ is the matrix whose entries are all zero. Note that now all echelons have height *and* length equal to 'one'. Using the row operations $(R2)$ and $(R3)$, $(\widetilde{A}\,|\,b)$ can be simplified. First we apply $(R2)$ to the rows R_i, $1 \leq i \leq k$, $(R_i) \rightsquigarrow \frac{1}{\widetilde{a}_{ii}}(R_i)$ yielding

$$\widetilde{\widetilde{A}} = \left(\begin{array}{ccccccc|c} \underline{1} & \widetilde{\widetilde{a}}_{12} & \widetilde{\widetilde{a}}_{13} & & \widetilde{\widetilde{a}}_{1k} & & & \\ 0 & |\underline{1} & \widetilde{\widetilde{a}}_{23} & & \vdots & & & \\ 0 & 0 & |\underline{1} & & \vdots & & & \\ \vdots & & & & & & & \\ 0 & \cdots & & & \widetilde{\widetilde{a}}_{(k-1)k} & & & \widetilde{\widetilde{b}} \\ 0 & \cdots & & & |\underline{1} & 0 & | & \\ \vdots & & & & & & & \\ 0 & \cdots & & & & 0 & | & \end{array}\right)$$

and then we apply $(R3)$ to remove all coefficients $\widetilde{a}_{ij}$ with $1 < i < j \leq k$ to obtain

$$\widehat{A} = \left(\begin{array}{cccccccc|c} \underline{1} & 0 & \cdots & 0 & 0 & \widehat{a}_{1(k+1)} & \cdots & \widehat{a}_{1n} & | & \\ 0 & |\underline{1} & \cdots & 0 & 0 & \widehat{a}_{2(k+1)} & \cdots & \widehat{a}_{2n} & | & \\ \vdots & & & & & \vdots & & \vdots & | & \\ 0 & \cdots & & \overline{0} & |\underline{1} & \widehat{a}_{k(k+1)} & \cdots & \widehat{a}_{kn} & | & \widehat{b} \\ 0 & \cdots & & & & 0 & \cdots & 0 & | & \\ \vdots & & & & & & & & & \\ 0 & \cdots & & & & 0 & \cdots & 0 & | & \end{array}\right)$$

referred to as being in refined echelon form. The system of linear equations corresponding to this latter augmented coefficient matrix $(\widehat{A}\,|\,b)$ is then the following one:

$$\begin{cases} y_1 + \sum_{j=k+1}^{n} \widehat{a}_{1j} y_j = \widehat{b}_1 \\ \vdots \\ y_k + \sum_{j=k+1}^{n} \widehat{a}_{kj} y_j = \widehat{b}_k \\ \sum_{j=1}^{n} 0 \cdot y_j = \widehat{b}_i \quad \forall\ k+1 \leq i \leq m\,. \end{cases}$$

Note that the unknowns are denoted by $y_1, \ldots, y_n$ since the original unknowns $x_1, \ldots, x_n$ might have been permuted, that $0 \leq k \leq \min(m, n)$, and that the set of solutions of the latter system is given by $\widetilde{L}$, introduced above. The sets $\widetilde{L}$ and L can now easily be determined. We have to distinguish between different cases:

- CASE 1: $k < m$ and there exists i with $k+1 \leq i \leq m$ so that $\widehat{b}_i \neq 0$. Then $\widetilde{L} = \emptyset$ and hence $L = \emptyset$.
- CASE 2: either $[k = m]$ or $[k < m$ and $\widehat{b}_{k+1} = 0\,, \ldots\,, \widehat{b}_m = 0]$. Then the system above reduces to the system of equations

$$y_i + \sum_{j=k+1}^{n} \widehat{a}_{ij} y_j = \widehat{b}_i\,, \quad \text{for any} \quad 1 \leq i \leq k\,. \tag{1.2.2}$$

- CASE 2A: if in addition $k = n$, then the system (1.2.2) reads $y_i = \widehat{b}_i$ for any $1 \leq i \leq n$. It means that $\widetilde{L} = \{(\widehat{b}_1, \ldots, \widehat{b}_n)\}$ and therefore the system with augmented coefficient matrix $(A\,|\,b)$ we started with, has a unique solution.
- CASE 2B: if in addition $k < n$, then the system (1.2.2) reads

$$y_i = \widehat{b}_i - \sum_{j=k+1}^{n} \widehat{a}_{ij} y_j \quad \text{for any} \quad 1 \leq i \leq k$$

and the unknowns $y_{k+1}, \ldots, y_n$ are free variables, also referred to as parameters and denoted by $t_{k+1}, \ldots, t_n$. The set of solutions $\widetilde{L}$ is then given by

$$\Big\{\Big(\widehat{b}_1 - \sum_{j=k+1}^{n} \widehat{a}_{1j} t_j\,, \ldots\,, \widehat{b}_k - \sum_{j=k+1}^{n} \widehat{a}_{kj} t_j,\ t_{k+1}\,, \ldots\,, t_n\Big) : \quad t_{k+1}, \ldots, t_n \in \mathbb{R} \text{ arbitrary}\Big\}\,.$$

Hence the system (1.2.2) and therefore also the original system with augmented coefficient matrix $(A \,|\, b)$ has infinitely many solutions. The map $\mathbb{R}^{n-k} \to \mathbb{R}^n$, given by

$$\begin{pmatrix} t_{k+1} \\ \vdots \\ t_n \end{pmatrix} \mapsto \begin{pmatrix} \widehat{b}_1 \\ \vdots \\ \widehat{b}_k \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix} + t_{k+1} \begin{pmatrix} -\widehat{a}_{1(k+1)} \\ \vdots \\ -\widehat{a}_{k(k+1)} \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} + t_{k+2} \begin{pmatrix} -\widehat{a}_{1(k+2)} \\ \vdots \\ -\widehat{a}_{k(k+2)} \\ 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix} + \cdots + t_n \begin{pmatrix} -\widehat{a}_{1n} \\ \vdots \\ -\widehat{a}_{kn} \\ 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix}$$

is a parameter representation of $\widetilde{L}$.

Let us now illustrate the discussed procedure with a few examples:

(1) Consider the case of one equation and two unknowns,

$$a_{11}x_1 + a_{21}x_2 = b_1\,, \quad a_{11} \neq 0\,.$$

Then

$$x_1 + \widehat{a}_{21}x_2 = \widehat{b}_1\,, \quad \widehat{a}_{21} = \frac{a_{21}}{a_{11}}\,, \quad \widehat{b}_1 = \frac{b_1}{a_{11}}$$

and we are in the CASE 2B with $k = 1, m = 1, n = 2$. The set of solutions $L = \widetilde{L}$ (no renumeration of unknowns were necessary) has the following parameter representation

$$\mathbb{R} \to \mathbb{R}^2\,, \quad t_2 \to \begin{pmatrix} \widehat{b}_1 \\ 0 \end{pmatrix} + t_2 \begin{pmatrix} -\widehat{a}_{21} \\ 1 \end{pmatrix}\,.$$

It is a straight line in $\mathbb{R}^2$, passing through the point $(\widehat{b}_1, 0)$ and having the direction $(-\widehat{a}_{21}, 1)$.

(2) Consider the case of one equation and three unknowns,

$$a_{11}x_1 + a_{12}x_2 + a_{13}x_3 = b_1, \quad a_{11} \neq 0\,.$$

We divide the equation by a_{11} and obtain

$$x_1 + \widehat{a}_{12}x_2 + \widehat{a}_{13}x_3 = \widehat{b}_1$$

where

$$\widehat{a}_{12} = \frac{a_{12}}{a_{11}}\,, \quad \widehat{a}_{13} = \frac{a_{13}}{a_{11}}\,, \quad \widehat{b}_1 = \frac{b_1}{a_{11}}$$

and we are again in the case 2-B with $k = 1, m = 1, n = 3$. The set of solutions $L = \widetilde{L}$ has the following parameter representation

$$\mathbb{R}^2 \to \mathbb{R}^3\,, \quad \begin{pmatrix} t_2 \\ t_3 \end{pmatrix} \mapsto \begin{pmatrix} \widehat{b}_1 \\ 0 \\ 0 \end{pmatrix} + t_2 \begin{pmatrix} -\widehat{a}_{12} \\ 1 \\ 0 \end{pmatrix} + t_3 \begin{pmatrix} -\widehat{a}_{13} \\ 0 \\ 1 \end{pmatrix}\,.$$

It is a plane in $\mathbb{R}^3$ passing through the point $(\widehat{b}_1, 0, 0)$ and spanned by the vectors $(-\widehat{a}_{12}, 1, 0)$ and $(-\widehat{a}_{13}, 0, 1)$.

(3) Consider the following system

$$x_1 + x_2 + x_3 = 4\,, \qquad x_1 - x_2 - 2x_3 = 0\,. \tag{1.2.3}$$

The corresponding augmented coefficient matrix $(A|b)$ is given by

$$\left(\begin{array}{ccc|c} 1 & 1 & 1 & 4 \\ 1 & -1 & -2 & 0 \end{array}\right).$$

Apply row operation $(R3)$ and replace R_2 by $R_2 - R_1$ to get

$$\left(\begin{array}{ccc|c} 1 & 1 & 1 & 4 \\ 0 & -2 & -3 & -4 \end{array}\right).$$

Apply row operation $(R2)$, $R_2 \rightsquigarrow -\frac{1}{2}R_2$ to get

$$\left(\begin{array}{ccc|c} 1 & 1 & 1 & 4 \\ 0 & 1 & \frac{3}{2} & 2 \end{array}\right)$$

and finally apply again row operation $(R3)$, $R_1 \rightsquigarrow R_1 - R_2$, yielding

$$\left(\begin{array}{ccc|c} 1 & 0 & -\frac{1}{2} & 2 \\ 0 & 1 & \frac{3}{2} & 2 \end{array}\right),$$

which is in refined echelon form. We are in the CASE 2B with $k = 2$, $m = 2$, $n = 3$. Since we have not permuted the unknowns, $\tilde{L} = L$ and a parameter representation of L is given by

$$\mathbb{R} \to \mathbb{R}^3\,, \quad t_3 \mapsto \begin{pmatrix} 2 \\ 2 \\ 0 \end{pmatrix} + t_3 \begin{pmatrix} 1/2 \\ -3/2 \\ 1 \end{pmatrix}$$

which is a straight line in $\mathbb{R}^3$, passing through the point $(2, 2, 0)$ and having the direction $(1/2, -3/2, 1)$.

(4) Consider

$$\begin{cases} x_1 + 2x_2 - x_3 = 1 \\ 2x_1 + x_2 + x_3 = 0 \\ 3x_1 + 0 \cdot x_2 + 3x_3 = -1\,. \end{cases}$$

Apply row operation $(R3)$ by replacing $R_2 \rightsquigarrow R_2 - 2R_1$, $R_3 \rightsquigarrow R_3 - 3R_1$, yielding

$$\left(\begin{array}{ccc|c} 1 & 2 & -1 & 1 \\ 0 & -3 & 3 & -2 \\ 0 & -6 & 6 & -4 \end{array}\right)$$

Apply $(R3)$ once were, $R_3 \rightsquigarrow R_3 - 2R_2$, leading to the following augmented coefficient matrix in echelon form

$$\left(\begin{array}{ccc|c} 1 & 2 & -1 & 1 \\ 0 & -3 & 3 & -2 \\ 0 & 0 & 0 & 0 \end{array}\right)$$

Now apply $(R2)$, $R_2 \rightsquigarrow -\frac{1}{3}R_2$ to get

$$\left(\begin{array}{ccc|c} 1 & 2 & -1 & 1 \\ 0 & 1 & -1 & \frac{2}{3} \\ 0 & 0 & 0 & 0 \end{array}\right)$$

and finally we apply $(R3)$ once more, $R_1 \rightsquigarrow R_1 - 2R_2$, to get the following augmented coefficient matrix in refined echelon form

$$\left(\begin{array}{ccc|c} 1 & 0 & 1 & -\frac{1}{3} \\ 0 & 1 & -1 & \frac{2}{3} \\ 0 & 0 & 0 & 0 \end{array}\right)$$

We are in CASE 2B with $k = 2$, $m = 3$, $n = 3$ and $\widetilde{L} = L$. Hence a parameter representation of L is given by

$$\mathbb{R} \to \mathbb{R}^3\,, \quad t_3 \mapsto \begin{pmatrix} -1/3 \\ 2/3 \\ 0 \end{pmatrix} + t_3 \begin{pmatrix} -1 \\ 1 \\ 1 \end{pmatrix}$$

which is a straight line in $\mathbb{R}^3$ passing through the point $(-1/3, 2/3, 0)$ and having the direction $(-1, 1, 1)$.

From our analysis one can deduce the following

Theorem 1.2.2. *If (S) is a system of m linear equations and n unknowns with $m < n$, then its set of solutions is either empty or infinite.*

Remark 1.2.3. To see that Theorem 1.2.2 holds, one argues as follows: bring the augmented coefficient matrix in refined row echelon form. Then $k \leq \min(m, n) = m < n$ since by assumption $m < n$. Hence CASE 2A cannot occur and we are either in CASE 1 ($L = \emptyset$) or in CASE 2B (L infinite). □

An important class of linear systems is the one where the number of equations is the same as the number of unknowns, $m = n$.

Definition 1.2.3. *A matrix A is called quadratic if the number of its rows equals the number of its columns.*

Definition 1.2.4. *We say that a $n \times n$ matrix $A = (a_{ij})_{1\le i,j\le n}$ is a diagonal matrix if $A = \mathrm{diag}A$, where $\mathrm{diag}A = (d_{ij})_{1\le i,j\le n}$ is the $n \times n$ matrix with*

$$d_{ii} = a_{ii} \quad \forall\, 1 \le i \le n, \qquad d_{ij} = 0 \quad \forall\, i \ne j.$$

It is called the $n \times n$ identity matrix if $a_{ii} = 1$ for any $1 \le i \le n$ and $a_{ij} = 0$ for $i \ne j$. We denote it by Id_n or $\mathrm{Id}_{n\times n}$.

Going through the procedure described above for transforming the augmented coefficient matrix $(A\,|\,b)$ of a given system (S) of n linear equations with n unknowns into refined echelon form, one sees that (S) has a unique solution if and only if it is possible to bring $(A\,|\,b)$ *without* renumbering the unknowns into the form $(\mathrm{Id}_n\,|\,\widehat{b})$, yielding the solution $x_1 = \widehat{b}_1, \ldots, x_n = \widehat{b}_n$.

Definition 1.2.5. *We say that a $n \times n$ matrix A is regular if it can be transformed by the row operations $(R1)$–$(R3)$ to the identity matrix Id_n. Otherwise A is called singular.*

Theorem 1.2.4. *Assume that (S) is a system of n linear equations and n unknowns with augmented coefficients matrix $(A\,|\,b)$, i.e.,*

$$\sum_{j=1}^{n} a_{ij}x_j = b_i\,, \quad 1 \le i \le n\,.$$

Then the following holds:
(i) If A is regular, then for any $b' \in \mathbb{R}^n$,

$$\sum_{j=1}^{n} a_{ij}x_j = b'_i\,, \quad 1 \le i \le n\,.$$

has a unique solution.
(ii) If A is singular, then for any $b' \in \mathbb{R}^n$, the system

$$\sum_{j=1}^{n} a_{ij}x_j = b'_i\,, \quad 1 \le i \le n\,.$$

has either no solution at all or infinitely many.

Example: Assume that A is a singular $n \times n$ matrix. Then the system with augmented coefficient matrix $(A\,|\,0)$ has infinitely many solutions.

Corollary 1.2.5. *Assume that A is a $n \times n$ matrix, $A = (a_{ij})_{1\le i,j\le n}$. If there exists $b \in \mathbb{R}^n$ so that*

$$\sum_{j=1}^{n} a_{ij}x_j = b_i\,, \quad 1 \le i \le n\,,$$

has a unique solution, then for any $b' \in \mathbb{R}^n$,

$$\sum_{j=1}^{n} a_{ij}x_j = b'_i\,, \quad 1 \le i \le n\,.$$

has a unique solution.

Definition 1.2.6. *A system of the form*

$$\sum_{j=1}^{n} a_{ij}x_j = b_i\,, \quad 1 \leq i \leq m$$

is called a homogeneous system of linear equations if $b = (b_1, \ldots, b_m) = (0, \ldots, 0)$ *and otherwise inhomogeneous. Given a inhomogeneous system*

$$\sum_{j=1}^{n} a_{ij}x_j = b_i\,, \quad 1 \leq i \leq m\,,$$

the system

$$\sum_{j=1}^{n} a_{ij}x_j = 0\,, \quad 1 \leq i \leq m$$

is referred to as the corresponding homogeneous system.

Note that a homogeneous system of linear equations has always the zero solution. Theorem 1.2.2 then leads to the following corollary.

Corollary 1.2.6. *A homogeneous system of* m *linear equations with* n *unknowns and* $m < n$ *has infinitely many solutions.*

We summarize the results for a system of n linear equations with n unknowns as follows.

Corollary 1.2.7. *Assume that we are given a system* (S)

$$\sum_{j=1}^{n} a_{ij}x_j = b_i\,, \quad 1 \leq i \leq n\,,$$

where $b = (b_1, \ldots, b_n) \in \mathbb{R}^n$*. Then the following statements are equivalent:*
(i) If (S) *has a unique solution.*
(ii) The homogeneous system corresponding to (S) *has only the zero solution.*
(iii) $A = (a_{ij})_{1 \leq i,j \leq n}$ *is regular.*

Finally, we already remark at this point that the set of solutions L of the linear system $\sum_{j=1}^{n} a_{ij}x_j = b\,, \quad 1 \leq i \leq m$, and the set of solutions L_h of the corresponding homogeneous system $\sum_{j=1}^{n} a_{ij}x_j = 0\,, \quad 1 \leq i \leq m$, have the following properties which can be verified in a straightforward way:

- (P1) for any $x, \widetilde{x} \in L_h$ and any $\lambda \in \mathbb{R}$, one has $x + \widetilde{x} \in L_h$, and $\lambda x \in L_h$.
- (P2) for any $x, \widetilde{x} \in L$, one has $x - \widetilde{x} \in L_h$.
- (P3) for any $x \in L$ and $\widetilde{x} \in L_h$, one has $x + \widetilde{x} \in L$.

We will come back to these properties after having introduced the notion of vector spaces.

Chapter 2

Matrix calculus and related topics

The aim of this chapter to introduce the notion of a matrix and to discuss the elementary properties of matrices as well as related topics.

2.1 Matrix calculus

The aim of this section is to discuss the basics of the matrix calculus. We denote by $\mathrm{Mat}_{m\times n}(\mathbb{R})$ or by $\mathbb{R}^{m\times n}$ the set of all real $m \times n$ matrices

$$A = (a_{ij})_{\substack{1\le i\le m\\ 1\le j\le n}} = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix}$$

Definition 2.1.1 (Sum, multiplication by scalars). *(i) For any $A = (a_{ij})_{\substack{1\le i\le m\\ 1\le j\le n}}$ and $B = (b_{ij})_{\substack{1\le i\le m\\ 1\le j\le n}}$ in $\mathbb{R}^{m\times n}$ we denote by $A + B$ the $m \times n$ matrix given by*

$$(a_{ij} + b_{ij})_{\substack{1\le i\le m\\ 1\le j\le n}} \in \mathbb{R}^{m\times n}\,.$$

(ii) For any $A = (a_{ij})_{\substack{1\le i\le m\\ 1\le j\le n}}$ in $\mathbb{R}^{m\times n}$ and $\lambda \in \mathbb{R}$ we denote by λA the $m \times n$ matrix

$$(\lambda a_{ij})_{\substack{1\le i\le m\\ 1\le j\le n}} \in \mathbb{R}^{m\times n}\,.$$

Note that for the sum of two matrices A and B to be defined, they have to have the same number of rows and the same number of columns. So , e.g., the two matrices

$$\begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} 6 & 7 \end{pmatrix}$$

cannot be added. The definition of the multiplication of matrices is more complicated. It is motivated by the interpretation of a $m \times n$ matrix as a linear map from $\mathbb{R}^m$ to $\mathbb{R}^n$ which we will discuss in the subsequent chapter in detail. We will see that the multiplication of matrices corresponds to the composition of the corresponding linear maps. At this point however it is only important to know that the definition of the multiplication of matrices is very well motivated.

Definition 2.1.2. *Assume that* $A = (a_{i\ell})_{\substack{1\le i\le m\\ 1\le \ell\le n}} \in \mathbb{R}^{m\times n}$ *and* $B = (b_{\ell j})_{\substack{1\le \ell\le n\\ 1\le j\le k}} \in \mathbb{R}^{n\times k}$. *Then the product of* A *and* B, *denoted by* $A\cdot B$ *or* AB *for short, is the* $m\times k$ *matrix* $C = (c_{ij})_{\substack{1\le i\le m\\ 1\le j\le k}}$ *with coefficients given by*

$$c_{ij} := \sum_{\ell=1}^{n} a_{i\ell} b_{\ell j}\,.$$

Note that in the computation of the coefficient c_{ij} only

$$\text{the i-th row of A, } (a_{i1},\dots,a_{in})\,, \qquad \text{and the j-th column of B, } \begin{pmatrix} b_{1j}\\ \vdots\\ b_{nj}\end{pmatrix}$$

are involved.

Examples

(1) Let $A = \begin{pmatrix}1 & 2\\ 3 & 4\end{pmatrix}, B = \begin{pmatrix}1\\ 2\end{pmatrix}$. Then $A\in\mathbb{R}^{2\times 2}$, $B\in\mathbb{R}^{2\times 1}$ and AB is well defined.

$$AB = \begin{pmatrix}1\cdot 1+2\cdot 2\\ 3\cdot 1+4\cdot 2\end{pmatrix} = \begin{pmatrix}5\\ 11\end{pmatrix}\in\mathbb{R}^{2\times 1}$$

(2) Let $A = \begin{pmatrix}1 & 2\\ 3 & 4\end{pmatrix}, B = \begin{pmatrix}1 & 0\\ 3 & 1\end{pmatrix}$. Then $A, B\in\mathbb{R}^{2\times 2}$ and

$$AB = \begin{pmatrix}1\cdot 1+2\cdot 3 & 1\cdot 0+2\cdot 0\\ 3\cdot 1+4\cdot 3 & 3\cdot 0+4\cdot 1\end{pmatrix} = \begin{pmatrix}7 & 2\\ 15 & 4\end{pmatrix}\in\mathbb{R}^{2\times 2}\,.$$

(3) Let $A = \begin{pmatrix}1 & 2\\ 3 & 4\end{pmatrix}, B = \begin{pmatrix}1 & 0 & 1\\ 3 & 1 & 0\end{pmatrix}$. Then $A\in\mathbb{R}^{2\times 2}$, $B\in\mathbb{R}^{2\times 3}$ and

$$AB = \begin{pmatrix}1\cdot 1+2\cdot 3 & 1\cdot 0+2\cdot 1 & 1\cdot 1+2\cdot 0\\ 3\cdot 1+4\cdot 3 & 3\cdot 0+4\cdot 1 & 3\cdot 1+4\cdot 0\end{pmatrix} = \begin{pmatrix}7 & 2 & 1\\ 15 & 4 & 3\end{pmatrix}\in\mathbb{R}^{2\times 3}$$

(4) Let $A = (1\ \ 2), B = \begin{pmatrix}3\\ 4\end{pmatrix}$. Then $A\in\mathbb{R}^{1\times 2}$, $B\in\mathbb{R}^{2\times 1}$ and

$$AB = 3+8 = 11\in\mathbb{R}^{1\times 1}(\simeq\mathbb{R})$$

and

$$BA = \begin{pmatrix}3\cdot 1 & 3\cdot 2\\ 4\cdot 1 & 4\cdot 2\end{pmatrix} = \begin{pmatrix}3 & 6\\ 4 & 8\end{pmatrix}\in\mathbb{R}^{2\times 2}\,.$$

The following theorem states elementary properties of the matrix multiplication.

Theorem 2.1.1. *The following holds:*
(*i*) *matrix multiplication is associative: for any* $A \in \mathbb{R}^{m\times n}$, $B \in \mathbb{R}^{n\times k}$, $C \in \mathbb{R}^{k\times \ell}$,

$$(AB)C = A(BC)\,;$$

(*ii*) *matrix multiplication is distributive:*
(*ii*1) *for any* $A, B \in \mathbb{R}^{m\times n}$, $C \in \mathbb{R}^{n\times k}$

$$(A+B)C = AC + BC\,;$$

(*ii*2) *for any* $A \in \mathbb{R}^{m\times n}$, $B, C \in \mathbb{R}^{n\times k}$

$$A(B+C) = AB + AC\,;$$

(*iii*) *operation with the identity matrix: for any* $A \in \mathbb{R}^{m\times n}$,

$$A \cdot \mathrm{Id}_n = A\,, \quad \mathrm{Id}_m \cdot A = A\,.$$

To get acquainted with matrix mutiplication let us verify item (i) of Theorem 2.1.1 in the case $m = 2, n = 2, k = 2$. For any $A, B, C \in \mathbb{R}^{2\times 2}$, the identity $A(BC) = (AB)C$ is verified as follows: let $D := BC$, $E := AB$. We claim that $AD = EC$. Indeed

$$(AD)_{ij} = \sum_{k=1}^{2} a_{ik} d_{kj} = \sum_{k=1}^{2} a_{ik} \sum_{\ell=1}^{2} b_{k\ell} c_{\ell j} = \sum_{k=1}^{2}\sum_{\ell=1}^{2} a_{ik} b_{k\ell} c_{\ell j}$$

and

$$(EC)_{ij} = \sum_{\ell=1}^{2} e_{i\ell} c_{\ell j} = \sum_{\ell=1}^{2} \Big(\sum_{k=1}^{2} a_{ik} b_{k\ell} \Big) c_{\ell j} = \sum_{\ell=1}^{2}\sum_{k=1}^{2} a_{ik} b_{k\ell} c_{\ell j}\,.$$

Matrix multiplication is not commutative, i.e., in general, for $A, B \in \mathbb{R}^{n\times n}$, one has $AB \neq BA$. As an example consider $A = \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}$ and $B = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}$. Then $AB \neq BA$ since

$$AB = \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 2 & 2 \end{pmatrix}$$

whereas

$$BA = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 1 & 2 \end{pmatrix}$$

As a consequence, in general $(AB)^2 \neq A^2B^2$.

However matrices in certain special classes commute with each other. The set of diagonal $n \times n$ matrices is such a class. Indeed if $A, B \in \mathbb{R}^{n\times n}$ are both diagonal matrices, $A = \mathrm{diag}(A)$, $B = \mathrm{diag}(B)$, then AB is also a diagonal matrix, $AB = \mathrm{diag}(AB)$, with

$$(AB)_{ii} = a_{ii} b_{ii}\,, \qquad 1 \le i \le n\,,$$

implying that $AB = BA$.

We say that two $n \times n$ matrices A, B commute if $AB = BA$. They anticommute if $AB = -BA$.

Definition 2.1.3. *A matrix $A \in \mathbb{R}^{n\times n}$ is called* invertible *if there exists $B \in \mathbb{R}^{n\times n}$ so that $AB = BA = \mathrm{Id}_n$.*

One can easily verify that for any given $A \in \mathbb{R}^{n\times n}$ there exists at most one matrix $B \in \mathbb{R}^{n\times n}$ so that $AB = BA = \mathrm{Id}_n$. Indeed assume that $C \in \mathbb{R}^{n\times n}$ satisfies $AC = CA = \mathrm{Id}_n$. Then

$$B = B \cdot \mathrm{Id}_n = B(AC) = (BA)C = \mathrm{Id}_n \cdot C = C\,.$$

Hence if $A \in \mathbb{R}^{n\times n}$ is invertible, there exists a unique matrix $B \in \mathbb{R}^{n\times n}$ so that $AB = BA = \mathrm{Id}_n$. This matrix is denoted by A^{-1} and is called the *inverse* of A. Note that the notion of the inverse is only defined for quadratic matrices.

Examples in $\mathbb{R}^{2\times 2}$:

(1) The null matrix $\begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$ is not invertible since for any $B \in \mathbb{R}^{2\times 2}$

$$\begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} B = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} b_1 & b_2 \\ b_3 & b_4 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} \neq \mathrm{Id}_2\,.$$

(2) Similarly, $\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$ is not invertible since for any $B \in \mathbb{R}^{2\times 2}$

$$\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} B = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} b_1 & b_2 \\ b_3 & b_4 \end{pmatrix} = \begin{pmatrix} b_3 & b_4 \\ 0 & 0 \end{pmatrix} \neq \mathrm{Id}_2\,.$$

(3) The identity matrix Id_2 is invertible and $\mathrm{Id}_2^{-1} = \mathrm{Id}_2$.

(4) The diagonal matrix $A = \begin{pmatrix} 3 & 0 \\ 0 & 4 \end{pmatrix}$ is invertible and

$$A^{-1} = \begin{pmatrix} \frac{1}{3} & 0 \\ 0 & \frac{1}{4} \end{pmatrix}\,.$$

(5) The matrix $A = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}$ is invertible. One can easily verify that

$$A^{-1} = -\frac{1}{2}\begin{pmatrix} 4 & -2 \\ -3 & 1 \end{pmatrix}\,.$$

How can the inverse of an invertible 2×2 matrix $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ be computed? In the case where $\det(A) \neq 0$, it turns out that A is invertible and its inverse A^{-1} is given by

$$A^{-1} = \frac{1}{\det(A)}\begin{pmatrix} d & -b \\ -c & a \end{pmatrix} = \begin{pmatrix} \frac{d}{\det(A)} & \frac{-b}{\det(A)} \\ \frac{-c}{\det(A)} & \frac{a}{\det(A)} \end{pmatrix}\,.$$

Indeed, one has

$$A^{-1}A = \frac{1}{\det(A)} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix} \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \frac{1}{\det(A)} \begin{pmatrix} da - bc & db - bd \\ -ca + ac & -cb + ad \end{pmatrix} = \mathrm{Id}_2 \,.$$

Similarly one verifies that $AA^{-1} = \mathrm{Id}_2$.

Before we describe a procedure to find the inverse of an invertible $n \times n$ matrix, let us state some general results on invertible $n \times n$ matrices. First let us introduce

$$\mathrm{GL}_{\mathbb{R}}(n) := \Big\{ A \in \mathbb{R}^{n \times n} : A \text{ is } \text{ invertible} \Big\} \tag{2.1.1}$$

where we remark that GL stands for 'general linear'.

Theorem 2.1.2. *The following holds:*
(i) For any $A, B \in \mathrm{GL}_{\mathbb{R}}(n)$, one has $AB \in \mathrm{GL}_{\mathbb{R}}(n)$ and

$$(AB)^{-1} = B^{-1}A^{-1} \,.$$

(ii) For any $A \in \mathrm{GL}_{\mathbb{R}}(n)$, $A^{-1} \in \mathrm{GL}_{\mathbb{R}}(n)$ and

$$(A^{-1})^{-1} = A \,.$$

To get acquainted with the notion of the inverse of an invertible matrix, let us verify the above statement *(i)*: first note that if $A, B \in \mathrm{GL}_{\mathbb{R}}(n)$, then A^{-1}, B^{-1} are well defined and so is $B^{-1}A^{-1}$. To see that $B^{-1}A^{-1}$ is the inverse of AB we compute

$$(B^{-1}A^{-1})(AB) = B^{-1}(A^{-1}A)B = B^{-1} \cdot \mathrm{Id}_n \cdot B = B^{-1}B = \mathrm{Id}_n$$

and similarly

$$(AB)B^{-1}A^{-1} = A(BB^{-1})A^{-1} = A \cdot \mathrm{Id}_n \cdot A^{-1} = AA^{-1} = \mathrm{Id}_n \,.$$

Hence by definition of the inverse we have that AB is invertible and $(AB)^{-1}$ is given by $B^{-1}A^{-1}$. To see that *(ii)* holds we argue similarly. Note that in general, $A^{-1}B^{-1}$ is not the inverse of AB, but of BA. Hence in case A and B do not commute, neither do A^{-1} and B^{-1}.

The important questions with regard to the invertibility of a quadratic matrix A are the following ones:

(1) How can we decide if A is invertible?

(2) In case A is invertible, how can we find its inverse?

It turns out that the two questions are closely related and can be answered by the following procedure: consider the system (S) of n linear equations with n unknowns

$$\begin{cases} a_{11}x_1 + \cdots + a_{1n}x_n = b_1 \\ \vdots \\ a_{n1}x_1 + \cdots + a_{nn}x_n = b_n \end{cases}$$

where $A = (a_{ij})_{1\leq i,j\leq n}$ and $b = (b_1, \ldots, b_n) \in \mathbb{R}^n$. We want to write this system in matrix notation. For this purpose we consider b as a $n \times 1$ matrix and similarly, we do so for x,

$$b = \begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix}, \quad x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}.$$

Since A is a $n \times n$ matrix, the matrix multiplication of A and x is well defined and $Ax \in \mathbb{R}^{n\times 1}$. Note that

$$(Ax)_j = \sum_{k=1}^{n} a_{jk}x_k = a_{j1}x_1 + \cdots + a_{jn}x_n \,.$$

Hence the above linear system (S), when written in matrix notation, takes the form

$$Ax = b \,.$$

Now let us assume that A is invertible. Then the matrix multiplication of A^{-1} and Ax is well defined and

$$A^{-1}(Ax) = (A^{-1}A)x = \mathrm{Id}_n x = x \,.$$

Hence multiplying left and right hand side of $Ax = b$ with A^{-1}, we get

$$x = A^{-1}b \,.$$

If we choose

$$b = e^{(1)} = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix},$$

then $x = A^{-1}e^{(1)}$ is the first column of A^{-1}. More generally, if

$$b = e^{(j)} = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

where 1 is the j-th component of $b = e^{(j)}$ and $1 \leq j \leq n$, then $A^{-1}e^{(j)}$ is the j-th column of A^{-1}. Summarizing, we have seen that we can decide if A is invertible and if so, determine A^{-1}, by solving the following systems of linear equations

$$Ax = e^{(1)}\,,\; Ax = e^{(2)}\,, \cdots, Ax = e^{(n)}\,.$$

In case A is regular, the solutions $x^{(1)}, \cdots, x^{(n)}$ of the latter equations are uniquely determined and are the columns of A^{-1}. In that case, we are thus led to apply the following procedure for determining the inverse of A: form the following version of the augmented coefficient matrix

$$\left(\begin{array}{ccc|cccc} a_{11} & \cdots & a_{1n} & 1 & 0 & \cdots & 0 \\ \vdots & & \vdots & 0 & 1 & \cdots & 0 \\ \vdots & & \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{n1} & \cdots & a_{nn} & 0 & 0 & \cdots & 1 \end{array}\right)$$

Use Gaussian elimination to determine the solutions $x^{(1)}, \cdots, x^{(n)}$. As an immediate consequence, one gets the following

Theorem 2.1.3. *The quadratic matrix A is invertible if and only if A is regular. In case A is invertible, the linear system $Ax = b$ has the solution $x = A^{-1}b$.*

We recall that a $n \times n$ matrix A is regular if it can be transformed into the identity matrix Id_n by the row operations $(R1) - (R3)$. In such a case, the above version of the augmented coefficient matrix

$$\left(\begin{array}{ccc|cccc} a_{11} & \cdots & a_{1n} & 1 & 0 & \cdots & 0 \\ \vdots & & \vdots & 0 & 1 & \cdots & 0 \\ \vdots & & \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{n1} & \cdots & a_{nn} & 0 & 0 & \cdots & 1 \end{array}\right)$$

gets transformed into

$$\left(\begin{array}{ccc|ccc} 1 & \cdots & 0 & b_{11} & \cdots & b_{1n} \\ \vdots & & \vdots & \vdots & & \vdots \\ 0 & \cdots & 1 & b_{n1} & \cdots & b_{nn} \end{array}\right)$$

and A^{-1} is given by the matrix $(b_{ij})_{1 \leq i,j \leq n}$. Let us now illustrate the procedure with a few examples: for each of the matrices A below, decide if A is invertible and if so, determine its inverse.

(1) For $A = \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix}$ consider the augmented coefficient matrix

$$\left(\begin{array}{cc|cc} 1 & 1 & 1 & 0 \\ -1 & 1 & 0 & 1 \end{array}\right)$$

and apply Gaussian elimination

(i) $R_2 \rightsquigarrow R_2 + R_1$:

$$\left(\begin{array}{cc|cc} 1 & 1 & 1 & 0 \\ 0 & 2 & 1 & 1 \end{array}\right)$$

(ii) $R_2 \rightsquigarrow \frac{1}{2}R_2$:

$$\left(\begin{array}{cc|cc} 1 & 1 & 1 & 0 \\ 0 & 1 & \frac{1}{2} & \frac{1}{2} \end{array}\right)$$

(iii) $R_1 \rightsquigarrow R_1 - R_2$:

$$\left(\begin{array}{cc|cc} 1 & 0 & \frac{1}{2} & -\frac{1}{2} \\ 0 & 1 & \frac{1}{2} & \frac{1}{2} \end{array}\right)$$

thus

$$A^{-1} = \begin{pmatrix} 1/2 & -1/2 \\ 1/2 & 1/2 \end{pmatrix} = \frac{1}{2}\begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}.$$

Note that the result coincides with the one obtained by the formula

$$\frac{1}{\det(A)}\begin{pmatrix} d & -b \\ -c & a \end{pmatrix}.$$

(2) For $A = \begin{pmatrix} 1 & -2 & 2 \\ 1 & 1 & -1 \\ 2 & 3 & 1 \end{pmatrix}$ consider the augmented coefficients matrix

$$\left(\begin{array}{ccc|ccc} 1 & -2 & 2 & 1 & 0 & 0 \\ 1 & 1 & -1 & 0 & 1 & 0 \\ 2 & 3 & 1 & 0 & 0 & 1 \end{array}\right)$$

and apply Gaussian elimination:

(i) $R_1 \rightsquigarrow R_2 - R_1$, $R_3 \rightsquigarrow R_3 - 2R_1$:

$$\left(\begin{array}{ccc|ccc} 1 & -2 & 2 & 1 & 0 & 0 \\ 0 & 3 & -3 & -1 & 1 & 0 \\ 0 & 7 & -3 & -2 & 0 & 1 \end{array}\right)$$

(ii) $R_3 \rightsquigarrow R_3 - \frac{7}{3}R_2$:

$$\left(\begin{array}{ccc|ccc} 1 & -2 & 2 & 1 & 0 & 0 \\ 0 & 3 & -3 & -1 & 1 & 0 \\ 0 & 0 & 4 & \frac{1}{3} & -\frac{7}{3} & 1 \end{array}\right)$$

(iii) $R_3 \rightsquigarrow \frac{1}{4}R_3$, $R_2 \rightsquigarrow \frac{1}{3}R_2$:

$$\left(\begin{array}{ccc|ccc} 1 & -2 & 2 & 1 & 0 & 0 \\ 0 & 1 & -1 & -\frac{1}{3} & \frac{1}{3} & 0 \\ 0 & 0 & 1 & \frac{1}{12} & -\frac{7}{12} & \frac{1}{4} \end{array}\right)$$

(iv) $(R_1) \rightsquigarrow R_1 - 2R_3$, $R_2 \rightsquigarrow R_2 + R_3$:

$$\left(\begin{array}{ccc|ccc} 1 & -2 & 0 & 1-\frac{2}{12} & \frac{14}{12} & -\frac{2}{4} \\ 0 & 1 & 0 & -\frac{3}{12} & -\frac{3}{12} & \frac{1}{4} \\ 0 & 0 & 1 & \frac{1}{12} & -\frac{7}{12} & \frac{1}{4} \end{array}\right)$$

(v) $R_1 \rightsquigarrow R_1 + 2R_2$:

$$\left(\begin{array}{ccc|ccc} 1 & 0 & 0 & \frac{1}{2}-\frac{1}{6} & \frac{7}{6}-\frac{1}{2} & -\frac{1}{2}+\frac{1}{2} \\ 0 & 1 & 0 & -\frac{1}{4} & -\frac{1}{4} & \frac{1}{4} \\ 0 & 0 & 1 & \frac{1}{12} & -\frac{7}{12} & \frac{1}{4} \end{array}\right)$$

and

$$A^{-1} = \begin{pmatrix} \frac{1}{2}-\frac{1}{6} & \frac{7}{6}-\frac{1}{2} & -\frac{1}{2}+\frac{1}{2} \\ -\frac{1}{4} & -\frac{1}{4} & \frac{1}{4} \\ \frac{1}{12} & -\frac{7}{12} & \frac{1}{4} \end{pmatrix}$$

(3) For $A = \begin{pmatrix} 2 & 3 \\ -4 & -6 \end{pmatrix}$ consider the augmented coefficients matrix

$$\left(\begin{array}{cc|cc} 2 & 3 & 1 & 0 \\ -4 & -6 & 0 & 1 \end{array}\right)$$

and apply Gaussian elimination:

(i) $R_2 \rightsquigarrow R_2 + 2R_1$:

$$\left(\begin{array}{cc|cc} 2 & 3 & 1 & 0 \\ 0 & 0 & 2 & 1 \end{array}\right)$$

It follows that A is not regular and hence according to Theorem 2.1.3, A is not invertible.

We finish this section by introducing the notion of the transpose of a matrix.

Definition 2.1.4. *Given $A \in \mathbb{R}^{m\times n}$, we denote by A^T the $n \times m$ matrix for which the i-th row is given by the j-th column of A. More formally,*

$$A^T = (b_{ij})_{\substack{1\le i\le n \\ 1\le j\le m}}, \quad \textit{with} \quad b_{ij} := a_{ji}$$

Examples:

(1) $\quad A = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} \in \mathbb{R}^{2\times 2} \quad \rightsquigarrow \quad A^T = \begin{pmatrix} 1 & 3 \\ 2 & 4 \end{pmatrix} \in \mathbb{R}^{2\times 2}$

(2) $\quad A = \begin{pmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{pmatrix} \in \mathbb{R}^{3\times 2} \quad \rightsquigarrow \quad A^T = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix} \in \mathbb{R}^{2\times 3}$

(3) $A = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} \in \mathbb{R}^{3\times 1} \quad \rightsquigarrow \quad A^T = (1\ 2\ 3) \in \mathbb{R}^{1\times 3}$

Definition 2.1.5. *A quadratic matrix $A = (a_{ij})_{1\le i,j\le n} \in \mathbb{R}^{n\times n}$ is said to be symmetric if $A = A^T$ or, written coefficient wise*

$$a_{ij} = a_{ji} \qquad \forall 1 \le i, j \le n\,.$$

Examples:

(1) $A = \begin{pmatrix} 1 & 2 \\ 2 & 3 \end{pmatrix}$ is symmetric.

(2) $A = \begin{pmatrix} 1 & 2 \\ 1 & 3 \end{pmatrix}$ is not symmetric.

(3) If $A \in \mathbb{R}^{n\times n}$ is a diagonal matrix, then A is symmetric. (Recall that A is a diagonal matrix if $A = \mathrm{diag}(A)$.)

Theorem 2.1.4. *(i) For any $A \in \mathbb{R}^{m\times n}$, $B \in \mathbb{R}^{n\times k}$,*

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

(ii) For any $A \in \mathrm{GL}_{\mathbb{R}}(n)$, also $A^T \in \mathrm{GL}_{\mathbb{R}}(n)$ and

$$(A^T)^{-1} = (A^{-1})^T\,.$$

To get more acquainted with the notion of the transpose of a matrix, let us verify the statements above: given $A \in \mathbb{R}^{m\times n}$, $B \in \mathbb{R}^{n\times k}$, one has $AB \in \mathbb{R}^{m\times k}$ and for any $1 \le i \le m$, $1 \le j \le k$,

$$(AB)_{ij} = \sum_{\ell=1}^{n} a_{i\ell} b_{\ell j} = \sum_{\ell=1}^{n} (A^T)_{\ell i} (B^T)_{j\ell} = \sum_{\ell=1}^{n} (B^T)_{j\ell} (A^T)_{\ell i} = (B^T A^T)_{ji}$$

and on the other hand, by the definition of the transpose matrix, $((AB)^T)_{ji} = (AB)_{ij}$. Combining the two identities yields $(AB)^T = B^T A^T$. To see that for any $A \in \mathrm{GL}_{\mathbb{R}}(n)$ also $A^T \in \mathrm{GL}_{\mathbb{R}}(n)$, the candidate for the inverse of A^T is the matrix $(A^{-1})^T$. Indeed one has

$$(A^T)(A^{-1})^T = (A^{-1}A)^T = \mathrm{Id}_n^T = \mathrm{Id}_n$$

since $\mathrm{Id}_n = \mathrm{Id}_n^T$, since Id_n is diagonal. Similarly

$$(A^{-1})^T A^T \overset{(i)}{=} (AA^{-1})^T = \mathrm{Id}_n^T = \mathrm{Id}_n\,.$$

2.2 Linear dependence, bases, coordinates

In this section we introduce the important notions of linear (in)dependence of vectors in $\mathbb{R}^k$, of a basis of $\mathbb{R}^k$ and of coordinates of a vector in $\mathbb{R}^k$ with respect to a basis. Later in this course we will discuss these notions in the more general framework of vector spaces of which $\mathbb{R}^k$ is an example. Elements of $\mathbb{R}^k$ are called vectors and written as $a = (a_1, \dots, a_k)$ where a_j, $1 \le j \le k$ are called the components of a. Alternatively, we view a as a $k \times 1$ matrix

$$a = \begin{pmatrix} a_1 \\ \vdots \\ a_k \end{pmatrix} .$$

Definition 2.2.1. *Assume that $a^{(1)}, \cdots, a^{(n)}$ are vectors in $\mathbb{R}^k$. A vector $b \in \mathbb{R}^k$ is said to be a linear combination of the vectors $a^{(1)}, \cdots, a^{(n)}$ if there exist real numbers $\alpha_1, \cdots, \alpha_n$ so that*

$$b = \alpha_1 a^{(1)} + \cdots + \alpha_n a^{(n)} \qquad or \qquad b = \sum_{j=1}^{n} \alpha_j a^{(j)}.$$

Example: Consider the vectors $a^{(1)} = (1, 2)$, $a^{(2)} = (2, 1) \in \mathbb{R}^2$. Then $b = (b_1, b_2) = (1, 5)$ is a linear combination of $a^{(1)}$ and $a^{(2)}$ since $b = 3a^{(1)} - a^{(2)}$: indeed

$$b_1 = 1\,, \quad (3a^{(1)} - a^{(2)})_1 = 3 \cdot 1 - 1 \cdot 2 = 1$$

and

$$b_2 = 5\,, \quad (3a^{(1)} - a^{(2)})_2 = 3 \cdot 2 - 1 \cdot 1 = 5.$$

Definition 2.2.2. *Assume that $a^{(1)}, \cdots, a^{(n)}$ are vectors in $\mathbb{R}^k$. They are said to be* linearly dependent *if there exists $1 \le i \le n$ such that $a^{(i)}$ is a linear combination of $a^{(j)}$, $j \ne i$, i.e., if there exists $\alpha_j \in \mathbb{R}$, $j \ne i$, so that*

$$a^{(i)} = \sum_{\substack{j \ne i \\ 1 \le j \le n}} \alpha_j a^{(j)} .$$

The vectors $a^{(1)}, \cdots, a^{(n)}$ are said to be linearly independent *if they are not linearly dependent.*

Example: We claim that the two vectors $a^{(1)} = (1, 2)$, $a^{(2)} = (2, 1)$ are linearly independent in $\mathbb{R}^2$. To see it, we assume that they are linearly dependent and then show that this is not possible. If $a^{(1)}$ and $a^{(2)}$ are linearly independent, then either there exists $\alpha_1 \in \mathbb{R}$ so that

$$a^{(2)} = \alpha_1 a^{(1)} \quad \rightsquigarrow \quad 2 = \alpha_1 \cdot 1, \quad 1 = \alpha_1 \cdot 2 \quad \rightsquigarrow \quad \alpha_1 = 2, \quad \alpha_1 = 1/2$$

or there exists $\alpha_2 \in \mathbb{R}$ with

$$a^{(1)} = \alpha_2 a^{(2)} \quad \rightsquigarrow \quad 1 = 2 \cdot \alpha_2, \quad 2 = \alpha_2 \cdot 1 \quad \rightsquigarrow \quad \alpha_2 = 1/2\,, \quad \alpha_2 = 2.$$

In both cases we arrive at a contraddiction and hence $a^{(1)}, a^{(2)}$ cannot be linearly dependent. Thus by definition they are linearly independent.

EXAMPLES: Decide if the following vectors are linearly dependent:

(1) $a^{(1)} = (1, 2),\ a^{(2)} = (3, -5)$ in $\mathbb{R}^2$

(2) $a^{(1)} = (1, 2, -3),\ a^{(2)} = (4, 5, -6)$ in $\mathbb{R}^3$

(3) $a^{(1)} = (2, 4, -8),\ a^{(2)} = (3, 6, -12)$ in $\mathbb{R}^3$

(4) $a^{(1)} = (1, -3),\ a^{(2)} = (-2, 6)$ in $\mathbb{R}^2$.

Important questions with regard to the notions of linear combination and linear independence are the following ones:

- How can we decide whether given vectors $a^{(1)}, \dots, a^{(n)}$ in $\mathbb{R}^k$ are linearly independent?
- Given vectors $b, a^{(1)}, \dots, a^{(n)}$ in $\mathbb{R}^k$ how can we decide whether b is a linear combination of $a^{(1)}, \dots, a^{(n)}$ and if so, how can we find $\alpha_1, \dots, \alpha_n$ in $\mathbb{R}$ so that

$$b = \sum_{j=1}^{n} \alpha_j a^{(j)} .$$

  Are the numbers $\alpha_1, \dots, \alpha_n$ uniquely determined?

It turns out that these questions are closely related with each other and can be rephrased in terms of systems of linear equations: assume that $b, a^{(1)}, \dots, a^{(n)}$ are vectors in $\mathbb{R}^k$. We view them as $k \times 1$ matrices and write

$$b = \begin{pmatrix} b_1 \\ \vdots \\ b_k \end{pmatrix}, \quad a^{(1)} = \begin{pmatrix} a_1^{(1)} \\ \vdots \\ a_k^{(1)} \end{pmatrix} = \begin{pmatrix} a_{11} \\ \vdots \\ a_{1k} \end{pmatrix}, \ \dots \ , a^{(n)} = \begin{pmatrix} a_1^{(n)} \\ \vdots \\ a_k^{(n)} \end{pmatrix} = \begin{pmatrix} a_{1n} \\ \vdots \\ a_{kn} \end{pmatrix} .$$

Denote by A the $k \times n$ matrix whose columns are given by $a^{(1)}, \dots, a^{(n)}$,

$$A = \begin{pmatrix} a_{11} & \dots & a_{1n} \\ \vdots & & \vdots \\ a_{k1} & \dots & a_{kn} \end{pmatrix} = \begin{pmatrix} a_1^{(1)} & \dots & a_1^{(n)} \\ \vdots & & \vdots \\ a_k^{(1)} & \dots & a_k^{(n)} \end{pmatrix} .$$

Recall that b is a linear combination ov the vectors $a^{(1)}, \dots, a^{(n)}$ if there exist real numbers $\alpha_1, \dots, \alpha_n$ so that

$$b = \sum_{j=1}^{n} \alpha_j a^{(j)} .$$

Written componentwise, it means that

$$\begin{cases} b_1 = \alpha_1 a_1^{(1)} + \cdots + \alpha_n a_1^{(n)} = \sum_{j=1}^n a_{1j}\alpha_j \\ \vdots \\ b_k = \alpha_1 a_k^{(1)} + \cdots + \alpha_n a_k^{(n)} = \sum_{j=1}^n a_{kj}\alpha_j \end{cases}$$

or in matrix notation $b = A\alpha$, where

$$\alpha = \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{pmatrix} \in \mathbb{R}^{n\times 1}\,.$$

Note that $A \in \mathbb{R}^{k\times n}$ and hence the matrix multiplication of A by α is well defined. Hence we can express the question if b is a linear combination of $a^{(1)}, \ldots, a^{(n)}$ in terms of the matrix calculus as follows:

• b is a *linear combination* of $a^{(1)}, \ldots, a^{(n)}$ if and only if the linear system $Ax = b$ has a solution $x \in \mathbb{R}^{n\times 1}$.

• $a^{(1)}, \ldots, a^{(n)}$ are *linearly dependent* if and only if the linear homogeneous system $Ax = 0$ has a solution $x \in \mathbb{R}^{n\times 1}$, $x \neq 0$.

• $a^{(1)}, \ldots, a^{(n)}$ are *linearly independent* if and only if the linear homogeneous system $Ax = 0$ has only the trivial solution $x = 0 \in \mathbb{R}^{n\times 1}$.

In the important case where $k = n$, we thus have, in view of the definition of a regular matrix, the following

Theorem 2.2.1. *Assume that $a^{(1)}, \ldots, a^{(n)}$ are vectors in $\mathbb{R}^n$. Then the following two statements are equivalent:*
(i) $a^{(1)}, \ldots, a^{(n)}$ are linearly independent;
(ii) the $n \times n$ matrix

$$A = \begin{pmatrix} a_1^{(1)} & \ldots & a_1^{(n)} \\ \vdots & & \vdots \\ a_n^{(1)} & \ldots & a_n^{(n)} \end{pmatrix}$$

is regular.

EXAMPLE: Let $b = (5, 5, -1, -2), a^{(1)} = (1, 1, 0, 1), a^{(2)} = (0, 2, -1, -1), a^{(3)} = (-2, -1, 0, 1)$. Show that b is alinear combination of $a^{(1)}, a^{(2)}, a^{(3)}$.

SOLUTION: it suffices to study the linear system $Ax = b$ where $x = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}$ and A is the 4×3 matrix

$$A = \begin{pmatrix} a_1^{(1)} & a_1^{(2)} & a_1^{(3)} \\ \vdots & & \vdots \\ a_4^{(1)} & a_4^{(2)} & a_4^{(3)} \end{pmatrix} = \begin{pmatrix} 1 & 0 & -2 \\ 1 & 2 & -1 \\ 0 & -1 & 0 \\ 1 & -1 & 1 \end{pmatrix}$$

Consider the augmented coefficient matrix

$$\left(\begin{array}{ccc|c} 1 & 0 & -2 & 5 \\ 1 & 2 & -1 & 5 \\ 0 & -1 & 0 & -1 \\ 1 & -1 & 1 & -2 \end{array}\right)$$

- $R_2 \rightsquigarrow R_2 - R_1$, $R_4 \rightsquigarrow R_4 - R_1$

$$\rightsquigarrow \left(\begin{array}{ccc|c} 1 & 0 & -2 & 5 \\ 0 & 2 & 1 & 0 \\ 0 & -1 & 0 & -1 \\ 0 & -1 & 3 & -7 \end{array}\right)$$

- $R_{2 \rightsquigarrow 3}$

$$\rightsquigarrow \left(\begin{array}{ccc|c} 1 & 0 & -2 & 5 \\ 0 & -1 & 0 & -1 \\ 0 & 2 & 1 & 0 \\ 0 & -1 & 3 & -7 \end{array}\right)$$

- $R_3 \rightsquigarrow R_3 + 2R_2$, $R_4 \rightsquigarrow R_4 - R_2$

$$\rightsquigarrow \left(\begin{array}{ccc|c} 1 & 0 & -2 & 5 \\ 0 & -1 & 0 & -1 \\ 0 & 0 & 1 & -2 \\ 0 & 0 & 3 & -6 \end{array}\right)$$

- $R_4 \rightsquigarrow R_4 - 3R_3$

$$\rightsquigarrow \left(\begin{array}{ccc|c} 1 & 0 & -2 & 5 \\ 0 & -1 & 0 & -1 \\ 0 & 0 & 1 & -2 \\ 0 & 0 & 0 & 0 \end{array}\right)$$

Therefore, $x_3 = -2$, $x_2 = 1$, $x_1 = 5 + 2x_3 = 1$ hence

$$b = 1 \cdot a^{(1)} + 1 \cdot a^{(2)} + (-2) \cdot a^{(3)} = a^{(1)} + a^{(2)} - 2a^{(3)} \, .$$

Definition 2.2.3. *Vectors $a^{(1)}, \ldots, a^{(n)}$ in $\mathbb{R}^k$ are called a basis of $\mathbb{R}^k$ if any vector $b \in \mathbb{R}^k$ can be represented in a unique way as linear combination of $a^{(1)}, \ldots, a^{(n)}$, i.e. if for every $b \in \mathbb{R}^k$ there exist unique real numbers $x_1, \ldots, x_n$ such that*

$$b = \sum_{j=1}^{n} x_j a^{(j)}.$$

The numbers $x_1, \ldots, x_n$ are called the coordinates of b with respect to the basis $a^{(1)}, \ldots, a^{(n)}$.

Theorem 2.2.2. *(i) Any basis of $\mathbb{R}^k$ consists of k vectors.*
(ii) If the vectors $a^{(1)}, \dots, a^{(k)}$ in $\mathbb{R}^k$ are linearly independent, then they form a basis in $\mathbb{R}^k$.
(iii) If $a^{(1)}, \dots, a^{(n)}$ are vectors in $\mathbb{R}^k$ with $1 \le n < k$, then there exist vectors $a^{(n+1)}, \dots, a^{(k)}$ in $\mathbb{R}^k$ so that $a^{(1)}, \dots, a^{(n)}, a^{(n+1)}, \dots, a^{(k)}$ form a basis of $\mathbb{R}^k$. In words, a collection of linearly independent vectors of $\mathbb{R}^k$ can always be completed to a basis of $\mathbb{R}^k$.

EXAMPLES:

- The vectors $e^{(1)} = (1, 0, \dots, 0), e^{(2)} = (0, 1, 0, \dots, 0), \dots, e^{(n)} = (0, \dots, 0, 1)$ form a basis of $\mathbb{R}^n$, referred to as the standard basis of $\mathbb{R}^n$. Indeed any vector $b = (b_1, \dots, b_n) \in \mathbb{R}^n$ can be written uniquely as

$$b = \sum_{j=1}^{n} b_j e^{(j)} .$$

  Note that $b_1, \dots, b_n$ are the coordinates of b with respect to the standard basis $e^{(1)}, \dots, e^{(n)}$ of $\mathbb{R}^n$.

- $a^{(1)} = (1, 1), a^{(2)} = (2, 1)$ is a basis of $\mathbb{R}^2$: according to Theorem 2.2.2, it suffices to verify that $a^{(1)}, a^{(2)}$ are linearly independent in $\mathbb{R}^2$ and according to Theorem 2.2.1, this is the case iff

$$A = \begin{pmatrix} a_1^{(1)} & a_1^{(2)} \\ a_2^{(1)} & a_2^{(2)} \end{pmatrix} = \begin{pmatrix} 1 & 2 \\ 1 & 1 \end{pmatrix}$$

  is regular. Since $\det(A) = 1 - 2 = -1 \neq 0$, A is indeed regular.

- Find the coordinates of the vector $b = (1, 3) \in \mathbb{R}^2$ with respect to the basis $a^{(1)} = (1, 1), a^{(2)} = (2, 1)$ of $\mathbb{R}^2$. We need to solve the linear system $Ax = b$, whose augmented coefficient matrix is given by

$$\left(\begin{array}{cc|c} 1 & 2 & 1 \\ 1 & 1 & 3 \end{array}\right) \overset{R_2 \rightsquigarrow R_2 - R_1}{\rightsquigarrow} \left(\begin{array}{cc|c} 1 & 2 & 1 \\ 0 & -1 & 2 \end{array}\right)$$

  and hence $x_2 = -2$, $x_1 = 1 - 2x_2 = 5$. The coordinates of b withe respect to $a^{(1)}$ and $a^{(2)}$ are thus 5 and -2,

$$b = 5a^{(1)} - 2a^{(2)} , \quad x_1 = 5 , \, x_2 = -2 .$$

  Note that the coordinates of b with respect to the standard basis of $\mathbb{R}^2$ are 1 and 3. An important question is how the coefficients $5, -2$ and $1, 3$ are related with each other.

Let us consider the latter question in a more general context. Assume that

$$[a] := [a^{(1)}, \dots, a^{(n)}] \quad \text{and} \quad b := [b^{(1)}, \dots, b^{(n)}]$$

are bases of $\mathbb{R}^n$ and consider a vector $u \in \mathbb{R}^n$. Then

$$u = \sum_{j=1}^{n} \alpha_j a^{(j)}\,, \qquad u = \sum_{j=1}^{n} \beta_j b^{(j)}$$

where $\alpha_1, \ldots, \alpha_n$ are the coordinates of u with respect to $[a]$ and $\beta_1, \ldots, \beta_n$ the ones of u with respect to $[b]$. We would like to have a method of computing β_j, $1 \le j \le n$ from α_j, $1 \le j \le n$. To this end we have to express the vectors $a^{(j)}$ as linear combination of the vectors $b^{(1)}, \ldots, b^{(n)}$:

$$a^{(j)} = \sum_{i=1}^{n} t_{ij} b^{(i)}\,, \quad T := \begin{pmatrix} t_{11} & \ldots & t_{1n} \\ \vdots & & \vdots \\ t_{n1} & \ldots & t_{nn} \end{pmatrix}$$

Then

$$\sum_{i=1}^{n} \beta_i b^{(i)} = u = \sum_{j=1}^{n} \alpha_j a^{(j)} = \sum_{j=1}^{n} \alpha_j \sum_{i=1}^{n} t_{ij} b^{(i)}$$

or

$$\sum_{i=1}^{n} \beta_i b^{(i)} = \sum_{i=1}^{n} \Big(\sum_{j=1}^{n} \alpha_j t_{ij} \Big) b^{(i)}\,.$$

Since the coordinates of u with respect to the basis $[b]$ are uniquely determined one conludes that $\beta_i = \sum_{j=1}^{n} t_{ij}\alpha_j$ for any $1 \le i \le n$. In matrix notation, we thus obtain the relation

$$\beta = T\alpha\,, \quad \alpha = \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{pmatrix}\,, \quad \beta = \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_n \end{pmatrix}\,. \tag{2.2.1}$$

It turns out that a convenient notation for T is the following one

$$T = \mathrm{Id}_{[a]\to[b]}\,.$$

The j-th column of T is the vector of coordinates of $a^{(j)}$ with respect to the basis $[b]$. Similarly, we express $b^{(j)}$ as a linear combination of $a^{(1)}, \ldots, a^{(n)}$,

$$b^{(j)} = \sum_{i=1}^{n} s_{ij} a^{(j)}\,, \quad S = \begin{pmatrix} s_{11} & \ldots & s_{1n} \\ \vdots & & \vdots \\ s_{n1} & \ldots & s_{nn} \end{pmatrix}\,.$$

We then obtain

$$\alpha = S\beta\,, \quad S = \mathrm{id}_{[b]\to[a]}\,. \tag{2.2.2}$$

Theorem 2.2.3. *Assume that $[a]$ and $[b]$ are bases of $\mathbb{R}^n$. Then*
(i) S and T are regular $n \times n$ matrices, hence invertible.
(ii) $T = S^{-1}$.

Note that item (ii) can be deduced by (2.2.1), (2.2.2) and item (i): indeed

$$\beta = T\alpha \overset{(2.2.2)}{\rightsquigarrow} \beta = TS\beta$$

$$\alpha = S\beta \overset{(2.2.1)}{\rightsquigarrow} \alpha = ST\alpha\,.$$

From item (i), it follows that $TS = \mathrm{Id}_n = ST$, i.e., $T = S^{-1}$.

EXAMPLE: Consider the standard basis $[e] = [e^{(1)}, e^{(2)}, e^{(3)}]$ and the basis $[b] = [b^{(1)}, b^{(2)}, b^{(3)}]$ of $\mathbb{R}^3$ given by

$$b^{(1)} = (1,1,1)\,, \quad b^{(2)} = (0,1,2)\,, \quad b^{(3)} = (2,1,-1)\,.$$

Let us compute $S = \mathrm{Id}_{[b]\to[e]}$. The first column of S is the vector of coordinates of $b^{(1)}$ with respect to the standard basis $[e]$,

$$b^{(1)} = 1\cdot e^{(1)} + 1\cdot e^{(2)} + 1\cdot e^{(3)}$$

and similarly,

$$b^{(2)} = 0\cdot e^{(1)} + 1\cdot e^{(2)} + 2\cdot e^{(3)}$$

and

$$b^{(3)} = 2\cdot e^{(1)} + 1\cdot e^{(2)} - 1\cdot e^{(3)}\,.$$

Hence

$$S = \begin{pmatrix} 1 & 0 & 2 \\ 1 & 1 & 1 \\ 1 & 2 & -1 \end{pmatrix}$$

With Gaussian elimination, we then compute $T = \mathrm{Id}_{[e]\to[b]} = S^{-1}$,

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

The coordinates $\beta_1, \beta_2, \beta_3$ of $u = (1,2,3) \in \mathbb{R}^3$ with respect to the basis $[b]$, $u = \sum_{j=1}^3 \beta_j b^{(j)}$, can then be computed as follows:

- compute the coordinates of u with respect to the basis $[e]$: $\alpha_1 = 1, \alpha_2 = 2, \alpha_3 = 3$.
- $$\beta = \mathrm{Id}_{[e]\to[b]}\,\alpha = T\alpha = \begin{pmatrix} 3 & -4 & 2 \\ -2 & 3 & -1 \\ -1 & 2 & -1 \end{pmatrix} \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}\,.$$

  Hence $u = 1\cdot b^{(1)} + 1\cdot b^{(2)} + 0\cdot b^{(3)}$.
- How to remember the formula ?

$$\beta = \mathrm{Id}_{[e]\to[b]}\,\alpha$$

$$\begin{cases} \beta = \text{'new' coordinates}, & \alpha = \text{'old' coordinates} \\ [b] = \text{'new' basis}, & [e] = \text{'old' basis} \end{cases}$$

We will come back on this topic in Chapter 3, when we discuss the notion of a linear map.

2.3 Determinants

We already introduced the notion of the determinant of a 2×2 matrix $A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$,

$$\det(A) = a_{11}a_{22} - a_{12}a_{21}$$

and established important properties:

- A is regular if and only if $\det(A) \neq 0$
  (simple criterion to decide if a 2×2 matrix is regular or not)
- Cramer's rule for solving $Ax = b$, $b \in \mathbb{R}^2$. Write

$$a^{(1)} = \begin{pmatrix} a_{11} \\ a_{21} \end{pmatrix}, \quad a^{(2)} = \begin{pmatrix} a_{12} \\ a_{22} \end{pmatrix}, \quad b = \begin{pmatrix} b_1 \\ b_2 \end{pmatrix}.$$

  Then

$$x_1 = \frac{\det(b \;\; a^{(2)})}{\det(a^{(1)} \;\; a^{(2)})}, \quad x_2 = \frac{\det(a^{(1)} \;\; b)}{\det(a^{(1)} \;\; a^{(2)})}.$$

We now want to address the question if the notion of determinant can be extended to $n \times n$ matrices so that corresponding properties hold as for 2×2 matrices. The answer is yes! Among the many equivalent ways of defining the determinant of $n \times n$ matrices, we choose a recursive definition, which defines the determinant of a $n \times n$ matrix in terms of determinants of certain $(n-1) \times (n-1)$ matrices. First we need to introduce some more notations. For $A \in \mathbb{R}^{n \times n}$ and $1 \leq i, j \leq n$, we denote by $A^{(i,j)} \in \mathbb{R}^{(n-1)\times(n-1)}$ the $(n-1) \times (n-1)$ matrix, obtained from A by deleting the i-th row and the j-th column.
EXAMPLE: For

$$A = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix} \in \mathbb{R}^{3\times 3}$$

one has

$$A^{(1,1)} = \begin{pmatrix} 5 & 6 \\ 8 & 9 \end{pmatrix}, \quad A^{(1,3)} = \begin{pmatrix} 4 & 5 \\ 7 & 8 \end{pmatrix}, \quad A^{(1,2)} = \begin{pmatrix} 4 & 6 \\ 7 & 9 \end{pmatrix}$$

$$A^{(2,1)} = \begin{pmatrix} 2 & 3 \\ 8 & 9 \end{pmatrix}, \quad A^{(2,2)} = \begin{pmatrix} 1 & 3 \\ 7 & 9 \end{pmatrix}, \quad A^{(2,3)} = \begin{pmatrix} 1 & 2 \\ 7 & 8 \end{pmatrix}$$

To motivate the inductive definition of the determinant of a $n \times n$ matrix, we first consider the cases $n = 1$, $n = 2$. One has

$$n = 1: \qquad A = (a_{11}) \in \mathbb{R}^{1\times 1} \quad \rightsquigarrow \quad \det(A) = a_{11}$$

$$n = 2: \qquad A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \quad \rightsquigarrow \quad \det(A) = a_{11}a_{22} - a_{12}a_{21}$$

which can be written as

$$\begin{aligned}\det(A) &= a_{11}\cdot\det(A^{(1,1)}) - a_{12}\cdot\det(A^{(1,2)})\\ &= (-1)^{1+1}a_{11}\cdot\det(A^{(1,1)}) + (-1)^{1+2}a_{12}\cdot\det(A^{(1,2)})\\ &= \sum_{j=1}^{2}(-1)^{1+j}a_{1j}\cdot\det(A^{(1,j)})\,.\end{aligned}$$

Definition 2.3.1. *For any $A\in\mathbb{R}^{n\times n}$ with $n\geq 3$,*

$$\det(A) = \sum_{j=1}^{n}(-1)^{1+j}a_{1j}\det(A^{(1,j)})\,. \tag{2.3.1}$$

Since $A^{(1,j)}$ is a $(n-1)\times(n-1)$ matrix for any $1\leq j\leq n$, this is indeed a recursive definition. We refer to (2.3.1) as the expansion of $\det(A)$ with respect to the first row.

Example:

$$A = \begin{pmatrix}1 & 2 & 3\\ 4 & 2 & 1\\ 1 & 0 & 1\end{pmatrix}\in\mathbb{R}^{3\times 3}\,.$$

Then

$$\det(A) = (-1)^{1+1}1\cdot\det(A^{(1,1)}) + (-1)^{1+2}2\cdot\det(A^{(1,2)}) + (-1)^{1+3}3\cdot\det(A^{(1,3)})\,.$$

Since

$$A^{(1,1)} = \begin{pmatrix}2 & 1\\ 0 & 1\end{pmatrix}\,,\quad A^{(1,2)} = \begin{pmatrix}4 & 1\\ 1 & 1\end{pmatrix}\,,\quad A^{(1,3)} = \begin{pmatrix}4 & 2\\ 1 & 0\end{pmatrix}\,,$$

one gets

$$\det(A) = (2\cdot 1 - 0) - 2(4\cdot 1 - 1\cdot 1) + 3(4\cdot 0 - 2\cdot 1) = 2 - 6 - 6 = -10\,.$$

Let us state some elementary important properties of the determinant.

Theorem 2.3.1. *For any $A\in\mathbb{R}^{n\times n}$ and $1\leq k\leq n$ the following holds:*
(i)

$$\det(A) = \sum_{j=1}^{n}(-1)^{k+j}a_{kj}\det(A^{(k,j)})$$

(expansion of $\det(A)$ *with respect to the k-th row)*
(ii)

$$\det(A) = \sum_{j=1}^{n}(-1)^{j+k}a_{jk}\det(A^{(j,k)})$$

(expansion of $\det(A)$ *with respect to the k-th column)*
(iii) $\det(A) = \det(A^T)$.

Note that item (iii) of the latter theorem follows from item (ii), since $(A^T)_{1j} = a_{j1}$ and $(A^T)^{(1,j)} = A^{(j,1)}$.

To state the next theorem let us introduce some more notations. For $A \in \mathbb{R}^{n\times n}$, denote by $a^{(1)}, \dots, a^{(n)}$ its columns and by $A^{(1)}, \dots, A^{(n)}$ its rows. We then have

$$A = (a^{(1)} \cdots a^{(n)}), \qquad A = \begin{pmatrix} A^{(1)} \\ \vdots \\ A^{(n)} \end{pmatrix}.$$

Theorem 2.3.2. *For any $A \in \mathbb{R}^{n\times n}$, the following identities hold:*
(i) *for any $1 \le j < i \le n$,*

$$\det(a^{(1)} \cdots a^{(j)} \cdots a^{(i)} \cdots a^{(n)}) = -\det(a^{(1)} \cdots a^{(i)} \cdots a^{(j)} \cdots a^{(n)})$$

and

$$\det \begin{pmatrix} A^{(1)} \\ \vdots \\ A^{(j)} \\ \vdots \\ A^{(i)} \\ \vdots \\ A^{(n)} \end{pmatrix} = -\det \begin{pmatrix} A^{(1)} \\ \vdots \\ A^{(i)} \\ \vdots \\ A^{(j)} \\ \vdots \\ A^{(n)} \end{pmatrix}.$$

(ii) *For any $\lambda \in \mathbb{R}$ and $1 \le i \le n$*

$$\det(a^{(1)} \cdots \lambda a^{(i)} \cdots a^{(n)}) = \lambda \cdot \det(a^{(1)} \cdots a^{(i)} \cdots a^{(n)})$$

and for any $b \in \mathbb{R}^n$

$$\det(a^{(1)} \cdots (a^{(i)} + b) \cdots a^{(n)}) = \det(a^{(1)} \cdots a^{(i)} \cdots a^{(n)}) + \det(a^{(1)} \cdots b \cdots a^{(n)})$$

and an analogous statement holds for the rows of A.
(iii) *for any $1 \le i, j \le n$ and $i \ne j$, $\lambda \in \mathbb{R}$,*

$$\det(a^{(1)} \cdots (a^{(i)} + \lambda a^{(j)}) \cdots a^{(n)}) = \det(a^{(1)} \cdots a^{i} \cdots a^{(n)}).$$

An analogous statement is valid for the rows of A.
(iv) *If A is upper triangular, namely $a_{ij} = 0 \quad \forall i > j$, then*

$$\det(A) = a_{11}a_{22} \cdots a_{nn} = \prod_{j=1}^{n} a_{jj}.$$

Expressed in words, item (ii) says that $\det(A)$ is linear with respect to its i-th column, $1 \le i \le n$ (cf Chapter 3 for the notion of linear maps). Similarly, $\det(A)$ is linear with respect to its i-th row, $1 \le i \le n$.

In view of the rules for computing $\det(A)$, stated in Theorem 2.3.2, one can compute $\det(A)$ with the help of Gaussian elimination.

EXAMPLE: Let

$$A = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 2 & 1 \\ 1 & 0 & 1 \end{pmatrix} \in \mathbb{R}^{3\times 3}\,.$$

Then by Gaussian elimination,

- $R_2 \rightsquigarrow R_2 - 4R_1$, $R_3 \rightsquigarrow R_3 - R_1$

$$\begin{pmatrix} 1 & 2 & 3 \\ 0 & -6 & -11 \\ 0 & -2 & -2 \end{pmatrix}$$

- $R_3 \rightsquigarrow R_3 - R_1$

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

hence

$$\det \begin{pmatrix} 1 & 2 & 3 \\ 4 & 2 & 1 \\ 1 & 0 & 1 \end{pmatrix} = \det \begin{pmatrix} 1 & 2 & 3 \\ 0 & -6 & -11 \\ 0 & 0 & \frac{5}{3} \end{pmatrix} = -10\,.$$

We finish this section by stating the following important properties of determinants.

Theorem 2.3.3. *For any $A \in R^{n\times n}$, A is regular if and only if* $\det(A) \neq 0$.

Theorem 2.3.4. *For any $A, B \in \mathbb{R}^{n\times n}$, the following holds:*
(i) $\det(AB) = \det(A)\det(B)$.
(ii) If A is invertible, then $\det(A) \neq 0$ *and*

$$\det(A^{-1}) = \frac{1}{\det(A)}\,.$$

Theorem 2.3.5 (Cramer's rule). *Assume that $A \in \mathbb{R}^{n\times n}$ is regular. Then for any $b \in \mathbb{R}^{n\times 1}$, the unique solution of $Ax = b$ is given by $x = A^{-1}b$ and for any $1 \leq j \leq n$, the j-th coefficient x_j of x is given by*

$$x_j = \frac{\det(A_{a^{(j)}\rightsquigarrow b})}{\det(A)}\,,$$

where $A_{a^{(j)}\rightsquigarrow b}$ is the $n \times n$ matrix, obtained from A by replacing the j-th column $a^{(j)}$ by b.

Chapter 3

Complex numbers and complex systems of linear equations

So far we have worked with real numbers and used that they are ordered and can be added and multiplied, tacitly assuming that addition and multiplication satisfy the classical computational rules, i.e., that this operations are commutative, associative, ... It turns out that for many reasons, it is necessary to consider an extension of the set $\mathbb{R}$ of real numbers. These more general numbers are referred to as complex numbers and the set of them is denoted by $\mathbb{C}$. One important feature of complex numbers is that for any equation of the form

$$x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0 = 0$$

with $n \geq 1$ and $a_{n-1}, \ldots, a_0$ arbitrary real numbers, there exists at least one solution in $\mathbb{C}$. In particular, the equation

$$x^2 + 1 = 0\,,$$

admits two solutions in $\mathbb{C}$, which are denoted by i and $-$i.
Before we introduce the complex numbers in a more formal way, let us put them in a more general context, hopefully making it easier to work with them. The natural numbers

$$1, 2, 3, \ldots$$

appear in the process of counting and have been studied for thousands of years. It is standard to denote the set of these numbers by $\mathbb{N}$. A first extension of $\mathbb{N}$ is necessary if one wants to solve an equation of the form

$$x + a = b, \quad a, b \in \mathbb{N} \quad \text{with} \quad a \geq b\,.$$

Note that such equations frequently come up in the business of accounting. Since $a \geq b$, this equation has no solution in $\mathbb{N}$ and it is necessary to introduce negative numbers and zero. The set

$$\ldots, -2, -1, 0, 1, 2, \ldots$$

is denoted by $\mathbb{Z}$ and its elements are referred to as entire numbers. For each $a \in \mathbb{N}$, $x + a = a$ has the unique solution $0 \in \mathbb{Z}$. More generally, for any $a, b \in \mathbb{Z}$, $x + a = b$ has

the unique solution

$$x = b + (-a) \in \mathbb{Z}\,.$$

To solve equations of the form

$$ax = b\,, \quad a, b \in \mathbb{Z}\,, \quad a \neq 0$$

one needs to extend $\mathbb{Z}$ and introduce the set of rational numbers

$$\mathbb{Q} := \left\{ \frac{p}{q} : p \in \mathbb{Z}\,,\ q \in \mathbb{N};\ p, q \text{ relatively prime} \right\}.$$

Finally, we mention that the equation

$$x^2 = 2$$

has no solution in $\mathbb{Q}$. (To see this, argue by assuming that it does and show that this leads to a contraddiction.) Note that x can be interpreted as the length of the hypotenuse of a rectangular triangle whose smaller sides have both length 1. Considerations of this type led to the extension of $\mathbb{Q}$ to the set $\mathbb{R}$ of real numbers. Elements in $\mathbb{R}$, which are not in $\mathbb{Q}$, are referred to as irrational numbers. (Irrational numbers can be further distinguished. E.g. numbers such as π and e are called trascendental numbers.)

3.1 Complex numbers and their calculus

A complex number is an element $(a, b) \in \mathbb{R}^2$ which we conveniently write as $z = a + \mathrm{i}b$ where the letter i stands for *imaginary*. The real numbers a and b are referred to as real and imaginary parts of z and denoted as follows

$$a = \mathrm{Re}(z)\,, \qquad b = \mathrm{Im}(z)\,.$$

We write $z = 0$ if $\mathrm{Re}(z) = 0$ and $\mathrm{Im}(z) = 0$. If $z = \mathrm{i}b$, $b \in \mathbb{R}$, then z is called a purely imaginary number. The set of all the complex numbers is denoted by $\mathbb{C}$. Keep in mind that numbers are only useful if we can add and multiply them with each other and if the operations of addition and multiplications have nice computational rules.
Addition: Complex numbers are added as vectors in $\mathbb{R}^2$, i.e., since for $(a, b), (a', b') \in \mathbb{R}^2$ one has

$$(a, b) + (a', b') = (a + a', b + b')$$

one has that for $z = a + \mathrm{i}b$, $z' = a' + \mathrm{i}b'$,

$$z + z' = (a + \mathrm{i}b) + (a' + \mathrm{i}b') := (a + a') + \mathrm{i}(b + b')\,.$$

Note that for $z' = 0$ one has

$$z + 0 = (a + 0) + \mathrm{i}(b + 0) = z\,.$$

Examples: For $z = 2 + 4\mathrm{i}$, $z' = 3 + \mathrm{i}$ one has

$$z + z' = (2 + 3) + \mathrm{i}(4 + 1) = 5 + 5\mathrm{i}\,.$$

Multiplication: Up till now we never multiplied vectors. The key to define multiplication between complex numbers is to interpret i as a solution of

$$x^2 + 1 = 0\,,$$

i.e.,

$$i^2 = -1\,.$$

Hence, we define for $z = a + \mathrm{i}b$, $z' = a' + \mathrm{i}b'$

$$zz' = (aa' - bb') + \mathrm{i}(ab' + ba')\,.$$

For reasons of clarity we sometimes write $z \cdot z'$ for zz'. We can compute zz' as follows

$$zz' = (a + \mathrm{i}b)(a' + \mathrm{i}b') = aa' + \mathrm{i}ba' + \mathrm{i}ab' + \mathrm{i}^2bb'\,.$$

Using that $\mathrm{i}^2 = -1$ and collecting terms containing i and those which do not, one gets indeed

$$zz' = (aa' - bb') + \mathrm{i}(ab' + ba')\,.$$

EXAMPLES: Again consider $z = 2 + 4\mathrm{i}$, $z' = 3 + \mathrm{i}$, then

$$zz' = (2 \cdot 3 - 4 \cdot 1) + \mathrm{i}(2 \cdot 1 + 4 \cdot 3) = 2 + \mathrm{i}14\,.$$

We have the following special cases:

- If $z = a \in \mathbb{R}$, then
$$zz' = az' = (aa') + \mathrm{i}(ab')$$
corresponds to the scalar multiplication of the vector $(a', b') \in \mathbb{R}^2$ by the real number $a \in \mathbb{R}$. In particular $1 \cdot z' = z'$.

- If $z = \mathrm{i}b$, $b \in \mathbb{R}$, then
$$zz' = \mathrm{i}b(a' + \mathrm{i}b') = b(-b' + ia') = -bb' + \mathrm{i}ba'\,.$$
Geometrically, this corresponds in $\mathbb{R}^2$ to a rotation by $\pi/2$ composed with a scalar multiplication by b.

- If $z = 0$, then $0 \cdot z' = 0$.

One can verify that the standard computational rules are satisfied: the operations of addition and multiplication are associative and commutative and the distributive laws hold.

Absolute value, polar representation: The absolute value of a complex number $z = a + \mathrm{i}b$ is defined as the length of the vector $(a, b) \in \mathbb{R}^2$ and denoted by $|z|$,

$$|z| = \sqrt{a^2 + b^2}\,.$$

In particular $|z| = 0$, if and only if $z = 0$. It leads to the polar representation of a complex number $z \neq 0$. Indeed

$$(a, b) = \sqrt{a^2 + b^2}(\cos\varphi, \sin\varphi) = (\sqrt{a^2 + b^2}\cos\varphi, \sqrt{a^2 + b^2}\sin\varphi)\,, \tag{3.1.1}$$

where φ is the oriented angle (determined modula 2π) between the x–axis and the vector (a, b). Hence, we have by the definition of multiplication that

$$z = |z|\cos\varphi + \mathrm{i}|z|\sin\varphi = |z|\big(\cos\varphi + \mathrm{i}\sin\varphi\big)\,.$$

To shorten notation we introduce

$$e(\varphi) := \cos\varphi + \mathrm{i}\sin\varphi$$

yielding the polar representation

$$z = |z|e(\varphi)\,.$$

Note that $|e(\varphi)| = 1$. For reason of clarity we sometimes also write $z = |z| \cdot e(\varphi)$. Note that the angle φ is only determined modulo 2π, i.e., φ might be replaced by $\varphi + 2\pi k$ for an arbitrary integer $k \in \mathbb{Z}$.
EXAMPLES:

- Polar representation of $z = -2$

$$|z| = 2 \quad \rightsquigarrow \quad z = 2(\cos\pi + \mathrm{i}\sin\pi) = 2 \cdot e(\pi)\,.$$

- Polar representation of $z = 1 + \mathrm{i}$

$$|z| = \sqrt{1+1} = \sqrt{2} \quad \rightsquigarrow \quad z = \sqrt{2}(\cos(\pi/4) + \mathrm{i}\sin(\pi/4)) = \sqrt{2} \cdot e(\pi/4)$$

- Polar representation of $z = 1 - \mathrm{i}$

$$|z| = \sqrt{2} \quad \rightsquigarrow \quad z = \sqrt{2} \cdot e(-\pi/4) = \sqrt{2} \cdot e(7\pi/4)\,.$$

The polar representation is particularly useful for the multiplication and the division of non zero complex numbers: let z, z' be nonzero complex numbers with polaar representation $z = |z|e(\varphi), z' = |z'|e(\varphi')$. Then

$$\begin{aligned} zz' &= |z||z'|(\cos\varphi + \mathrm{i}\sin\varphi)(\cos\varphi' + \mathrm{i}\sin\varphi') \\ &= |z||z'|\Big[(\cos\varphi\cos\varphi' - \sin\varphi\sin\varphi') + \mathrm{i}(\cos\varphi\sin\varphi' + \sin\varphi\cos\varphi')\Big]\,. \end{aligned}$$

Since, by the trigonometric addition theorems

$$\cos\varphi\cos\varphi' - \sin\varphi\sin\varphi' = \cos(\varphi + \varphi')\,, \quad \cos\varphi\sin\varphi' + \sin\varphi\cos\varphi' = \sin(\varphi + \varphi')$$

one obtains that

$$zz' = |z||z'|e(\varphi + \varphi')\,.$$

In particula one has $|zz'| = |z||z'|$. Given $z \neq 0$, the formula above can be used to determine the inverse of z, denoted by $\frac{1}{z}$. It is the complex number z' characterized by

$$1 = zz' = |z||z'|e(\varphi + \varphi')\,,$$

hence

$$|z'| = \frac{1}{|z|} \quad \text{and} \quad \varphi' = -\varphi \quad \text{modulo} \quad 2\pi\,,$$

i.e.,

$$\frac{1}{z} = \frac{1}{|z|}e(-\varphi)\,.$$

Combining the formula for zz' and the one of the inverse of a complex number one sees that for nonzero complex numbers z, z' with polar representations $z = |z|e(\varphi)$, $z' = |z'|e(\varphi')$, one has

$$z \cdot \frac{1}{z'} = \frac{|z|}{|z'|}e(\varphi - \varphi')\,.$$

Conjugation: The complex conjugate of a complex number $z = a + \mathrm{i}b$ is defined to be $a - \mathrm{i}b$ and denoted by $\overline{z}$. Note that $\overline{0} = 0$ and that for any $z \neq 0$ with polar representation $z = |z|e(\varphi)$, one has

$$\overline{z} = |z|e(-\varphi)\,.$$

Geometrically, the map $z \mapsto \overline{z}$ corresponds in $\mathbb{R}^2$ to the reflection across the x–axis, $(a, b) \mapsto (a, -b)$. Note that for any $z \in \mathbb{C}$,

$$z\overline{z} = (a + \mathrm{i}b)(a - \mathrm{i}b) = a^2 + b^2 = |z|^2\,.$$

This can be used to compute the quotient z'/z of two complex numbers z' and $z \neq 0$. Indeed let $z = a + \mathrm{i}b$ and $z' = a' + \mathrm{i}b'$. Then we multiply nominator and denominator of z'/z by $\bar{z}$ to obtain

$$\begin{aligned}\frac{z'}{z} = \frac{z'\overline{z}}{z\overline{z}} &= \frac{(a' + \mathrm{i}b')(a - \mathrm{i}b)}{a^2 + b^2} \\ &= \frac{a'a + b'b}{a^2 + b^2} + \mathrm{i}\frac{b'a - a'b}{a^2 + b^2}\end{aligned}$$

Example: Compute real and imaginary part of the quotient $\frac{2+3\mathrm{i}}{1-\mathrm{i}}$: since $\overline{1 - \mathrm{i}} = 1 + \mathrm{i}$,

$$\frac{2 + 3\mathrm{i}}{1 - \mathrm{i}} \cdot \frac{1 + \mathrm{i}}{1 + \mathrm{i}} = \frac{(2 + 3\mathrm{i})(1 + \mathrm{i})}{1 + 1} = \frac{2 - 3 + \mathrm{i}(3 + 2)}{2} = -\frac{1}{2} + \mathrm{i}\frac{5}{2}\,.$$

The computations in polar coordinates would be more complicated as the polar representation of $2 + 3\mathrm{i}$ does not come with a 'nice' angle.
The following identities can be easily verified:

$$\overline{z + z'} = \overline{z} + \overline{z'}\,, \quad \overline{zz'} = \overline{z} \cdot \overline{z'}\,.$$

Furthermore, we point out that $z = \overline{z'}$ if and only if z is real, i.e., $\mathrm{Im}(z) = 0$.

Powers and roots: The powers z^n, $n = 1, 2 \ldots$ of a complex number $z \neq 0$ can easily be computed by using the polar representation of z, $z = |z|e(\varphi)$. For $n = 2$ one gets

$$z^2 = |z|e(\varphi)|z|e(\varphi) = |z|^2 e(2\varphi)$$

and then inductively for n arbitrary,

$$z^n = |z|^n e(n\varphi) = |z|^n \Big(\cos(n\varphi) + \mathrm{i}\sin(n\varphi) \Big) .$$

EXAMPLE: Compute real and imaginary part of $(1+\mathrm{i})^{12}$. Note that $|1+\mathrm{i}| = \sqrt{2}$ and thus

$$1 + \mathrm{i} = \sqrt{2}e(\pi/4) .$$

Hence

$$(1+\mathrm{i})^{12} = 2^{12/2} e(12\pi/4) = 64e(3\pi) = -64 .$$

Hence

$$\mathrm{Re}((1+\mathrm{i})^{12}) = -64 \qquad \text{and} \qquad \mathrm{Im}((1+\mathrm{i})^{12}) = 0 .$$

The same procedure works to compute

$$z^{-n} = \Big(\frac{1}{z}\Big)^n = \frac{1}{|z|^n} e(-n\varphi) .$$

EXAMPLE: Compute real and imaginary part of $(1+\mathrm{i})^{-8}$. Since $1 + \mathrm{i} = \sqrt{2}e(\pi/4)$, one has that

$$\frac{1}{1+\mathrm{i}} = \frac{1}{\sqrt{2}} e(-\pi/4)$$

and therefore

$$(1+\mathrm{i})^{-8} = \frac{1}{2^{8/2}} e(-8\pi/4) = \frac{1}{16} ,$$

yielding

$$\mathrm{Re}((1+\mathrm{i})^{-8}) = \frac{1}{16} \quad \text{and} \quad \mathrm{Im}((1+\mathrm{i})^{-8}) = 0 .$$

Next we want to find all the solutions of the equation

$$\zeta^n = z ,$$

where z is a given complex number. A solution of the equation $\zeta^n = z$ is called a n-th root of z. If $z = 0$, then the only solution is $\zeta = 0$. If $z \neq 0$, consider the polar representation $z = |z|e(\varphi)$. A solution ζ of $\zeta^n = z$ then satisfies $\zeta \neq 0$ and hence has also a polar representation $\zeta = \rho e(\psi)$. Hence we want to solve

$$\rho^n e(n\psi) = |z|e(\varphi) .$$

Clearly one has

$$\rho = |z|^{\frac{1}{n}}$$

and

$$n\psi = \varphi \qquad \text{modulo} \qquad 2\pi\,.$$

There are n solutions of the latter equation and they are given by

$$\psi_0 = \frac{\varphi}{n}\,, \quad \psi_1 = \frac{\varphi}{n} + \frac{2\pi}{n}\,, \quad \psi_2 = \frac{\varphi}{n} + 2\frac{2\pi}{n}\,, \ldots, \psi_{n-1} = \frac{\varphi}{n} + (n-1)\frac{2\pi}{n}\,.$$

Hence if $z \neq 0$, $\zeta^n = z$ has n solutions given by

$$\zeta_0 = |z|^{\frac{1}{n}} e\Big(\frac{\varphi}{n}\Big)\,,\ \zeta_1 = |z|^{\frac{1}{n}} e\Big(\frac{\varphi}{n} + \frac{2\pi}{n}\Big)\,, \ldots\,,\ \zeta_{n-1} = |z|^{\frac{1}{n}} e\Big(\frac{\varphi}{n} + (n-1)\frac{2\pi}{n}\Big)\,.$$

In the case $z = 1$, $\zeta_0, \ldots, \zeta_{n-1}$ are called the n-th roots of unity. Note that in this case one can choose the polar representation $z = e(0)$ and $\zeta_0 = 1$.
EXAMPLE:

- Compute the 6-th roots of the unity, $\zeta^6 = 1$. Then

$$\zeta_0 = 1\,, \quad \zeta_1 = e\Big(\frac{2\pi}{6}\Big) = e\Big(\frac{\pi}{3}\Big)\,, \quad \zeta_2 = e\Big(\frac{2\pi}{3}\Big)\,,$$

$$\zeta_3 = e\Big(\frac{3\pi}{3}\Big) = -1\,, \quad \zeta_4 = e\Big(\frac{4\pi}{3}\Big)\,, \quad \zeta_5 = e\Big(\frac{5\pi}{3}\Big)\,.$$

- Compute the solutions of $\zeta^3 = 2(\sqrt{3} + \mathrm{i})$. Note that $2(\sqrt{3} + \mathrm{i}) = 4e(\pi/6)$. Thus

$$\zeta_0 = 4^{\frac{1}{3}} e\Big(\frac{1}{3}\frac{\pi}{6}\Big) = 4^{\frac{1}{3}} e\Big(\frac{\pi}{18}\Big)\,,$$

$$\zeta_1 = 4^{\frac{1}{3}} e\Big(\frac{\pi}{18} + \frac{2\pi}{3}\Big)\,, \quad \zeta_2 = 4^{\frac{1}{3}} e\Big(\frac{\pi}{18} + 2\cdot\frac{2\pi}{3}\Big)\,.$$

Remark: The expression $e(\varphi) = \cos\varphi + \mathrm{i}\sin\varphi$ can be computed via the complex exponential function,

$$e^z = \sum_{n=0}^{\infty} \frac{1}{n!} z^n = 1 + z + \frac{1}{2!} z^2 + \ldots$$

Euler showed that

$$e^{\mathrm{i}x} = \cos x + \mathrm{i}\sin x \qquad \textbf{(Euler's formula)}$$

The complex exponential function has the same properties of the real exponential function e^x; in particular one has

$$e^{z+z'} = e^z e^{z'}\,,$$

implying that for any $z = a + \mathrm{i}b$

$$e^z = e^{a+\mathrm{i}b} = e^a e^{\mathrm{i}b} = e^a(\cos b + \mathrm{i}\sin b)\,.$$

It means that

$$|e^z| = e^a\,,$$

and that b, modulo 2π is the argument of e^z. By Euler's formula

$$\sin x = \frac{e^{\mathrm{i}x} - e^{-\mathrm{i}x}}{2\mathrm{i}}\,, \quad \cos x = \frac{e^{\mathrm{i}x} + e^{-\mathrm{i}x}}{2}$$

and the trigonometric addition theorems are an easy consequence:

$$e^{\mathrm{i}(\varphi+\psi)} = \cos(\varphi+\psi) + \mathrm{i}\sin(\varphi+\psi)$$

and

$$\begin{aligned} e^{\mathrm{i}(\varphi+\psi)} = e^{\mathrm{i}\varphi}e^{\mathrm{i}\psi} &= \Big(\cos\varphi + \mathrm{i}\sin\varphi\Big)\Big(\cos\psi + \mathrm{i}\sin\varphi\Big) \\ &= \cos\varphi\cos\psi - \sin\varphi\sin\psi \\ &\quad + \mathrm{i}\Big(\sin\varphi\cos\psi + \cos\varphi\sin\psi\Big)\,. \end{aligned}$$

Comparing the two expressions leads to

$$\cos(\varphi+\psi) = \cos\varphi\cos\psi - \sin\varphi\sin\psi\,, \quad \sin(\varphi+\psi) = \sin\varphi\cos\psi + \sin\psi\cos\varphi\,.$$

3.2 The fundamental theorem of algebra

An important class of functions is the class of polynomials. Later in this course, polynomials will play a prominent role for computing the eigenvalues of a matrix. Here, we consider polynomials in one complex variable z,

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

or, written in compact notation,

$$p(z) = \sum_{k=0}^{n} a_k z^k\,.$$

It can be viewed as a function

$$\mathbb{C} \to \mathbb{C}\,, \quad z \mapsto p(z)\,.$$

The coefficients $a_0, a_1, \ldots, a_n$ are assumed to be complex numbers. If $a_n \neq 0$, p is said to be a polynomial of degree n, where n is assumed to be a non negative integer. Note that a polynomial p of degree 0 is a constant $p = a_0 \neq 0$. In case all the coefficients are real, p is said to be a polynomial with real coefficients. Of special interest are the zeros (Nullstellen) of p, also referred to as roots (Wurzeln). By definition, a complex zero of a polynomial p of degree $n \geq 1$ is a complex number w so that $p(z)$ can be factorized in the following way :

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

where q is a polynomial of degree $n-1$.

Theorem 3.2.1 (Fundamental theorem of algebra). *Any polynomial* $p(z) = \sum_{k=0}^{n} a_k z^k$ *of degree* $n \geq 1$ *and with complex coefficients* $a_0, a_1, \ldots, a_n$ *has at least one complex root.*

It means that for such a polynomial there exists a complex number z_1 so that $p(z)$ can be written as a product

$$p(z) = (z - z_1)p_1(z)$$

where $p_1(z)$ is a polynomial of degree $n-1$ (≥ 0). We can apply Theorem 3.2.1 inductively to get the following

Theorem 3.2.2. *Any polynomial* $p(z) = \sum_{k=0}^{n} a_k z^k$ *of degree* $n \geq 1$ *and with complex coefficients* $a_0, a_1, \ldots, a_n$ *can be written as a product of* n *linear factors*

$$p(z) = a_n(z - z_1) \cdots (z - z_n)$$

or, written in a compact form,

$$a_n \prod_{k=1}^{n} (z - z_k)\,.$$

Note that the complex numbers $z_1, \ldots, z_n$ do not need to be different from each other. Hence, alternatively we can write

$$p(z) = a_n(z - \zeta_1)^{m_1} \ldots (z - \zeta_\ell)^{m_\ell}$$

where $\zeta_1, \ldots, \zeta_\ell$ are the different complex numbers among $z_1, \ldots, z_n$ and $m_i \in \mathbb{Z}_{\geq 1}$ denotes the multiplicity of the root ζ_i. One has $\sum_{i=1}^{\ell} m_i = n$. We remark that the results corresponding to Theorems 3.2.1, 3.2.2 do not hold for polynomials p with *real* coefficients. More precisely , it is *not* true that a polynomial p of abitrary degree $n \geq 2$ with real coefficients has at least one real zero. As an example we mention the polynomial $p(x) = x^2 + 1$ of degree two, with real coefficients

$$a_2 = 1\,, \quad a_1 = 0\,, \quad a_0 = 1\,.$$

However, as it should be, p has two complex roots, $z_1 = \mathrm{i}, z_2 = -\mathrm{i}$ and p factors over the complex numbers

$$p(z) = (z - \mathrm{i})(z + \mathrm{i})\,.$$

This observation is one of the main reasons why we need to introduce the complex numbers for the sequel of the course: Eigenvalues of a real $n \times n$ matrix ($n \geq 2$), might be complex numbers.

To finish, let us consider polynomials of degree two with complex coefficients. In this case we are fortunate to have a formula for its roots. Consider $p(z) = az^2 + bz + c$ with $a, b, c \in \mathbb{C}$, $a \neq 0$. To find the roots of p, we proceed as the Babylonians thousands of years before us: by completing the square (quadratische Ergänzung) one has

$$p(z) = a\Big(z^2 + \frac{b}{a}z + \frac{b^2}{4a^2}\Big) + c - \frac{b^2}{4a} = a\Big(z + \frac{b}{2a}\Big)^2 + c - \frac{b^2}{4a}\,.$$

To find z_1, z_2 with $p(z_i) = 0$, one needs to solve

$$a\Big(z + \frac{b}{2a}\Big)^2 = \frac{b^2}{4a} - c$$

or

$$\Big(z + \frac{b}{2a}\Big)^2 = \frac{b^2}{4a^2} - \frac{c}{a}\,.$$

We need to distinguish two cases:
CASE 1 : $\frac{b^2}{4a^2} - \frac{c}{a} = 0$. Then

$$\Big(z + \frac{b}{2a}\Big)^2 = 0$$

and we get

$$z_1 = z_2 = -\frac{b}{2a}.$$

It then follows that

$$p(z) = a\Big(z + \frac{b}{2a}\Big)^2.$$

CASE 2: $\frac{b^2}{4a^2} - \frac{c}{a} \neq 0$. In this case, the equation

$$w^2 = \frac{b^2}{4a^2} - \frac{c}{a} = \frac{1}{4a^2}(b^2 - 4ac)$$

has two dinstinct complex solutions w_1 and w_2, where $w_2 = -w_1$ and $w_1 \neq 0$ is given by

$$w_1 = \frac{1}{2a}\sqrt{b^2 - 4ac}$$

with a specific choice of the sign of the square root. Hence the two roots of p are

$$z_1 = -\frac{b}{2a} + w_1\,, \quad z_2 = -\frac{b}{2a} - w_1\,,$$

and we have

$$p(z) = a\Big(z - \Big(-\frac{b}{2a} + w_1\Big)\Big)\Big(z - \Big(-\frac{b}{2a} - w_1\Big)\Big)\,.$$

Now let us treat the special case where $p(z) = az^2+bz+c$ has real coefficients $a, b, c \in \mathbb{R}$ with $a \neq 0$. We argue in a similar way and write by completing the square

$$p(z) = a\Big(z + \frac{b}{2a}\Big)^2 + c - \frac{b^2}{4a}\,.$$

CASE 1(REAL), $\frac{b^2}{4a^2} - \frac{c}{a} = 0$. Then

$$\Big(z + \frac{b}{2a}\Big)^2 = 0$$

and we conclude that

$$z_1 = z_2 = -\frac{b}{2a}$$

is real. The polynomial $p(x) = ax^2 + bx + c$ can be written as product of real linear factors,

$$p(x) = a\Big(x + \frac{b}{2a}\Big)^2 .$$

CASE 2 (REAL) : $\frac{b^2}{4a^2} - \frac{c}{a} \neq 0$. We distinguish between two cases:
CASE 2 A: $\frac{b^2}{4a^2} - \frac{c}{a} > 0$. Then

$$w_1 = \sqrt{b^2 - 4ac}\,, \quad w_2 = -\sqrt{b^2 - 4ac}$$

are real numbers and so are the roots

$$z_1 = -\frac{b}{2a} + \frac{1}{2a}\sqrt{b^2 - 4ac}\,, \quad z_2 = -\frac{b}{2a} - \frac{1}{2a}\sqrt{b^2 - 4ac}$$

and $p(x)$ can be written as the product of two real linear factors

$$p(x) = a\Big(x - \Big(-\frac{b}{2a} + \frac{1}{2a}\sqrt{b^2 - 4ac}\Big)\Big)\Big(x - \Big(-\frac{b}{2a} - \frac{1}{2a}\sqrt{b^2 - 4ac}\Big)\Big)\,.$$

CASE 2B $b^2 - 4ac < 0$. Then

$$w_1 = \mathrm{i}\sqrt{4ac - b^2}\,, \quad w_2 = -\mathrm{i}\sqrt{4ac - b^2}$$

are purely imaginary numbers and the roots z_1, z_2 are now the complex numbers

$$z_1 = -\frac{b}{2a} + \mathrm{i}\frac{1}{2a}\sqrt{4ac - b^2}\,, \quad z_2 = -\frac{b}{2a} - \mathrm{i}\frac{1}{2a}\sqrt{4ac - b^2}\,.$$

Note that z_2 is the complex conjugate of z_1, $z_2 = \overline{z}_1$.

For polynomials of degree three, there are formulas for the three roots (formulas of Cardano) but they are more complicated. For polynomials of degree $n \geq 5$ it can be proved that the roots can no longer be written as an expression of roots of complex numbers.

For polynomials with real coefficients it can be shown that for any complex root w, also its complex coordinates $\bar{w}$ is a root and the two roots have the same multiplicity. To see that if w is a root of p, then $p(\bar{w}) = 0$, take the complex conjugate of $p(w) = 0$,

$$0 = \overline{a_n w^n + \ldots + a_0} = \overline{a}_n \overline{w}^n + \ldots + \overline{a}_0\,.$$

Since the coefficients are assumed to be real one has $\overline{a}_n = a_n, \ldots, \overline{a}_0 = a_0$ and hence

$$0 = a_n \overline{w}^n + \ldots + a_0 = p(\overline{w})\,,$$

i.e., $\overline{w}$ is indeed a root of p.
EXAMPLE 1. Find the representation of $p(z) = z^4 + z^3 - z - 1$ as a product of linear factors: by inspection, one sees that $z_1 = 1$, $z_2 = -1$ are roots of p. We get the representation

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

The polynomial $q(z) = z^2 + z + 1$ has the roots

$$z_3 = -\frac{1}{2} + \mathrm{i}\frac{\sqrt{3}}{2}\,, \qquad z_4 = \overline{z}_3 = -\frac{1}{2} - \mathrm{i}\frac{\sqrt{3}}{2}\,.$$

EXAMPLE 2. Find the representation of $p(z) = z^6 - 1$ as a product of linear factors: by inspection, one sees that

$$z^6 - 1 = (z^3 - 1)(z^3 + 1)$$

and

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

Hence $z_1 = 1$, $z_2 = -1$ are two real roots. As above the roots of $z^2 + z + 1$ are

$$z_3 = -\frac{1}{2} + \mathrm{i}\frac{\sqrt{3}}{2}\,, \qquad z_4 = \overline{z}_3 = -\frac{1}{2} - \mathrm{i}\frac{\sqrt{3}}{2}$$

whereas the roots of $z^2 - z + 1$ are

$$z_5 = \frac{1}{2} + \mathrm{i}\frac{\sqrt{3}}{2}\,, \qquad z_6 = \frac{1}{2} - \mathrm{i}\frac{\sqrt{3}}{2}\,.$$

3.3 Systems of linear equations with complex coefficients

The results on systems of linear equations with real coefficients and matrices $A \in \mathbb{R}^{m\times n}$ with their calculus, discussed in Chapter 1, extend in a very natural way to the corresponding objects with complex coefficients:

Definition 3.3.1. *An element $A \in \mathbb{C}^{m\times n}$ is a $m \times n$ matrix with complex coefficients,*

$$A = (a_{ij})_{\substack{1\le i\le m\\ 1\le j\le n}}\,, \quad a_{ij} \in \mathbb{C}\,,$$

also referred to as a complex $m \times n$ matrix. A system of m linear equations with complex coefficients for the n unknowns $z_1, \ldots, z_n$ is a system of the form

$$\sum_{j=1}^{n} a_{ij} z_j = b_i\,, \quad 1 \le i \le m$$

where $a_{ij} \in \mathbb{C}$ $(1 \le i \le m\,,\, 1 \le j \le n)$ and $b_i \in \mathbb{C}$ $(1 \le i \le m)$. It is referred to as complex linear system. In matrix notation, it is given by $Az = b$ where

$$z = \begin{pmatrix} z_1 \\ \vdots \\ z_n \end{pmatrix} \in \mathbb{C}^{n\times 1}\,, \quad b = \begin{pmatrix} b_1 \\ \vdots \\ b_m \end{pmatrix} \in \mathbb{C}^{m\times 1}\,, \quad A = (a_{ij})_{\substack{1\le i\le m\\ 1\le j\le n}} \in \mathbb{C}^{m\times n}$$

and Az denotes matrix multiplication of the $m \times n$ matrix A with the $n \times 1$ matrix z, resulting in the $m \times 1$ matrix Az with coefficients

$$\sum_{j=1}^{n} a_{ij} z_j\,, \quad 1 \le i \le m\,.$$

As in the case of real linear systems, complex linear systems can be solved by Gauss elimination. We say that a $n \times n$ matrix A is regular if A can be transformed into the $n \times n$ identity matrix by the row operations $(R1), (R2)$, and $(R3)$, suitably defined in the context of complex numbers.

Definition 3.3.2.

$$\mathrm{GL}_{\mathbb{C}}(n) := \left\{ A \in \mathbb{C}^{n\times n} : A \text{ is regular} \right\}.$$

One introduces the notion of the determinant of a complex $n \times n$ matrix as in the case of real $n \times n$ matrices and can show as in that case that for any $A \in \mathbb{C}^{n\times n}$, A is regular if and only if $\det(A) \neq 0$.

Definition 3.3.3. *We say that $b \in \mathbb{C}^m$ is a $\mathbb{C}$-linear combination of vectors $a^{(1)}, \ldots, a^{(n)}$ in $\mathbb{C}^m$ if there exist $\alpha_1, \ldots, \alpha_n \in \mathbb{C}$ such that*

$$b = \sum_{j=1}^{n} \alpha_j b^{(j)}.$$

Definition 3.3.4. *Vectors $a^{(1)}, \ldots, a^{(n)} \in \mathbb{C}^m$ are said to be $\mathbb{C}$-linearly independent if for any $\alpha_1, \ldots, \alpha_n \in \mathbb{C}$ with*

$$\sum_{j=1}^{n} \alpha_j a^{(j)} = 0$$

it follows that $\alpha_1 = 0, \ldots, \alpha_n = 0$. Otherwise $a^{(1)}, \ldots, a^{(n)}$ are said to be $\mathbb{C}$-linearly dependent.

Definition 3.3.5. *The vectors $a^{(1)}, \ldots, a^{(n)}$ form a basis of $\mathbb{C}^m$ if the following holds:*

(1) $a^{(1)}, \ldots, a^{(n)}$ are linearly independent and

(2) every element $b \in \mathbb{C}^m$ can be represented as a $\mathbb{C}$-linear combination of $a^{(1)}, \ldots, a^{(n)}$.

In such a case one has $n = m$ and every element in $\mathbb{C}^m$ can be written in a unique way as $\mathbb{C}$ linear combination of $a^{(1)}, \ldots, a^{(m)}$.
EXAMPLE: The vectors $e^{(1)} = (1, 0)$, $e^{(2)} = (0, 1)$ in $\mathbb{R}^2$ are $\mathbb{R}$-linearly independent in $\mathbb{R}^2$. Viewed as complex numbers, $e^{(1)} \sim 1$, $e^{(2)} \sim \mathrm{i}$, they are $\mathbb{C}$-linearly dependent since for the complex number $\alpha_1 = \mathrm{i}$ one has $e^{(2)} = \alpha_1 e^{(1)}$.

Chapter 4

Vector spaces and linear maps

In this chapter we introduce the notion of a vector space (over $\mathbb{R}$ or $\mathbb{C}$) and of maps between vector spaces which respect the structure of vector spaces. Such maps are referred to as linear maps.

4.1 Vector spaces and their linear subspaces

First let us consider the space $\mathbb{R}^n$ consisting of elements $a = (a_1, \ldots, a_n)$ with $a_1, \ldots, a_n \in \mathbb{R}$. These elements are also referred to as vectors. Vectors in $\mathbb{R}^n$ can be added and multiplied by a real number:
Vector addition: For $a = (a_1, \ldots, a_n)$ and $b = (b_1, \ldots, b_n)$ in $\mathbb{R}^n$, the sum of a and b is defined as

$$a + b := (a_1 + b_1, \ldots, a_n + b_n)\,.$$

Scalar multiplication: For any $\lambda \in \mathbb{R}$, and $a = (a_1, \ldots, a_n) \in \mathbb{R}^n$,

$$\lambda a \equiv \lambda \cdot a = (\lambda a_1, \ldots, \lambda a_n)\,.$$

Recall that in $\mathbb{R}, \mathbb{R}^2$, and $\mathbb{R}^3$, vectors can be also multiplied:
(i) in $\mathbb{R}$: multiplication of real numbers:

$$a, b \in \mathbb{R} \quad \rightsquigarrow \quad a \cdot b$$

(ii) in $\mathbb{R}^2$: multiplication of complex numbers:

$$a = (a_1, a_2)\,,\ b = (b_1, b_2) \quad \rightsquigarrow \quad (a_1 b_1 - a_2 b_2, a_2 b_1 + a_1 b_2)$$

(iii) in $\mathbb{R}^3$: vector product of vectors in $\mathbb{R}^3$: for any $a = (a_1, a_2, a_3)$, $b = (b_1, b_2, b_3)$,

$$a \times b := (a_2 b_3 - a_3 b_2, -a_1 b_3 + a_3 b_1, a_1 b_2 - a_2 b_1)\,.$$

Note that the vector product is anticommutative, i.e., $a \times b = -b \times a$.
In $\mathbb{R}^4$ one can define the multiplication of quaternions which however is no longer associative. In $\mathbb{R}^n$ with $n \geq 5$, no multiplication rules exist which turned out to be useful.

Similarly we can consider the space $\mathbb{C}^n$ consisting of elements $a = (a_1, \ldots, a_n)$ with $a_1, \ldots, a_n \in \mathbb{C}$. These elements are also referred to as vectors or complex vectors. Complex vectors a, b can be added,

$$a + b = (a_1 + b_1, \ldots, a_n + b_n)$$

and multiplied by a complex scalar $\lambda \in \mathbb{C}$,

$$\lambda a = \lambda \cdot a = (\lambda a_1, \ldots, \lambda a_n)\,.$$

In the sequel, $\mathbb{K}$ denotes either $\mathbb{R}$ (field of real numbers) or $\mathbb{C}$ (field of complex numbers).

Definition 4.1.1. *A $\mathbb{K}$–vector space is a non empty set V, endowed with two operations $+$ (addition) and $\cdot$ (multiplication with a scalar $\lambda \in \mathbb{K}$),*

$$+ : V \times V \to V\,,\ (a, b) \mapsto a + b\,, \quad \text{and} \quad \cdot : \mathbb{K} \times V \to V\,,\ (\lambda, a) \mapsto \lambda \cdot a$$

with the following properties:
$(VS1)$ $(V, +)$ is an abelian group:
(i) addition is associative: $(a + b) + c = a + (b + c)$ for any $a, b, c \in V$
(ii) addition is commutative: $a + b = b + a$ for any $a, b \in V$
(iii) there is an element $0 \in V$ with the property that $0 + a = a + 0 = a$ for any $a \in V$. (The element is uniquely determined by these properties and referred to as zero or zero element of V. This justifies the notation 0.)
(iv) For any element $a \in V$ there exists an element $b \in V$ such that $a + b = 0$. (The element b is referred to as the inverse of a and denoted by $-a$. Since it is uniquely determined, this notation is justified.)
$(VS2)$ The two distributive laws hold, i.e., for any $\lambda, \mu \in \mathbb{K}$, $a, b \in V$
(v) $(\lambda + \mu) \cdot a = \lambda \cdot a + \mu \cdot a$
(vi) $\lambda \cdot (a + b) = \lambda \cdot a + \lambda \cdot b$
$(VS3)$ Scalar multiplication is associative, i.e., for any $\lambda, \mu \in \mathbb{K}$ and $a \in V$,
(vii) $\lambda \cdot (\mu \cdot a) = (\lambda\mu) \cdot a$
$(VS4)$ For any $a \in V$: $1 \cdot a = a$.

Elements of a $\mathbb{K}$–vector space are often referred to as vectors.

Remark 4.1.1. Various useful identities can be derived by $(VS1) - (VS4)$:
(i) For any $a \in V$, $0 \cdot a = 0$. To verify this identity, note that $0 \cdot a = (0+0) \cdot a = 0 \cdot a + 0 \cdot a$. Add on both sides of the latter equality $-(0 \cdot a)$ to conclude that

$$0 \overset{(VS1)}{=} 0 \cdot a + (-0 \cdot a) = 0 \cdot a + 0 \cdot a + (-0 \cdot a) \overset{(VS1)}{=} 0 \cdot a$$

(ii) For any $a \in V$, $(-1) \cdot a = -a$. To verify this identity, note that

$$a + (-1) \cdot a = 1 \cdot a + (-1) \cdot a = (1 + (-1)) \cdot a = 0 \cdot a \overset{(i)}{=} 0\,.$$ □

Examples

(1) For any $n \in \mathbb{Z}_{\geq 0}$, $(\mathbb{R}^n, +, \cdot)$ with addition $+$ and scalar multiplication $\cdot$ described as above is a $\mathbb{R}$-vector space. Similarly $(\mathbb{C}^n, +, \cdot)$ is a $\mathbb{C}$-vector space.

(2) For any $m, n \in \mathbb{Z}_{\geq 1}$, let $V = \mathbb{R}^{m\times n}$ the space of $m \times n$ matrices, and define addition of matrices and multiplication of a matrix by a scalar $\lambda \in \mathbb{R}$ as introduced earlier,

$$+ : V \times V \to V\,,\ (A, B) \mapsto A + B\,, \quad \text{and} \quad \cdot : \mathbb{R} \times V \to V\,,\ (\lambda, A) \mapsto \lambda \cdot A\,.$$

Then $(V, +, \cdot)$ ia a $\mathbb{R}$–vector space.

(3) Let $M[0, 1] := \{f : [0, 1] \to \mathbb{R}\}$ and define

$$+ : M[0, 1] \times M[0, 1] \to M[0, 1]\,,\ (f, g) \mapsto f + g$$

where $f + g : [0, 1] \to \mathbb{R}\,,\ t \mapsto (f + g)(t) = f(t) + g(t)$, and

$$\cdot : \mathbb{R} \times M[0, 1] \to M[0, 1]\,,\ (\lambda, f) \mapsto \lambda \cdot f\,,$$

where $\lambda \cdot f : [0, 1] \to \mathbb{R}\,,\ t \mapsto (\lambda f)(t) := \lambda \cdot f(t)$. Then $(M[0, 1], +, \cdot)$ is a $\mathbb{R}$–vector space.

(4) Let $\mathcal{P}_n$ denote the set of all polynomials with real coefficients of degree at most n with $n \in \mathbb{Z}_{\geq 1}$. An element $p \in \mathcal{P}_n$ can be viewed as a function

$$p : \mathbb{R} \to \mathbb{R}\,, \quad t \mapsto a_n t^n + a_{n-1} t^{n-1} + \cdots + a_1 t + a_0\,,$$

where $a_n, \ldots, a_0$ are the coefficients of p and are assumed to be real numbers. Define addition and scalar multiplication as in the Example (3). Then $(\mathcal{P}_n, +, \cdot)$ is a $\mathbb{R}$–vector space.

Definition 4.1.2. *If $V \equiv (V, +, \cdot)$ is a $\mathbb{K}$–vector space and $W \subseteq V$ a nonempty subset of V, then W is said to be a $\mathbb{K}$–subspace (or subspace for short) of V if the following holds:*
$(SS1)$ W closed with respect to addition: $a + b \in W\,,\ \forall a, b \in W\,.$
$(SS2)$ W closed with repsect to scalar multiplication: $\lambda a \in W\,,\ \forall a \in W\,, \forall \lambda \in \mathbb{K}\,.$

Note that W with $+$ and $\cdot$ of V forms a $\mathbb{K}$–vector space.
Examples:

(1) Assume that $A \in \mathbb{R}^{m\times n}$ and let $L := \{x \in \mathbb{R}^{n\times 1} : Ax = 0\}$. Then L is a $\mathbb{R}$–subspace of $\mathbb{R}^{n\times 1}$ $(\simeq \mathbb{R}^n)$. To verify this assertion we have to show that L is non empty and $(SS1), (SS2)$ hold. Indeed, $0 \in \mathbb{R}^{n\times 1}$ is an element of L since $A0 = 0$. Furthermore, for any $x, y \in L$, $\lambda \in \mathbb{R}$, one has

$$A(x + y) = Ax + Ay = 0 + 0 = 0 \quad \text{and} \quad A(\lambda x) = \lambda Ax = \lambda \cdot 0 = 0\,.$$

(2) Let $M(\mathbb{R}) := \{f : \mathbb{R} \to \mathbb{R}\}$ and define addition $+$ and scalar multiplication $\cdot$ as for $M[0, 1]$ above. Then $(M(\mathbb{R}), +, \cdot)$ is a $\mathbb{R}$–vector space and $\mathcal{P}_n$ is a subspace of $(M(\mathbb{R}), +, \cdot)$.

Definition 4.1.3. *Assume that V is a $\mathbb{K}$–vector space and $v^{(1)}, \ldots, v^{(n)}$ are elements of V. Then $[v] = [v^{(1)}, \ldots, v^{(n)}]$ is said to be a basis of V if the following holds:*
$(B1)$ Any $a \in V$ can be represented as a $\mathbb{K}$–linear combination of $v^{(1)}, \ldots, v^{(n)}$, i.e. there exist $\lambda_1, \ldots, \lambda_n \in \mathbb{K}$ so that $a = \sum_{j=1}^{n} \lambda_j v^{(j)}$.
$(B2)$ The vectors $v^{(1)}, \ldots, v^{(n)}$ are $\mathbb{K}$–linearly independent.

Remark 4.1.2. Equivalently, $[v]$ is a basis if the following holds:
(B) Any $a \in V$ can be written as a $\mathbb{K}$–linear combination in a *unique* way, $a = \sum_{j=1}^{n} \lambda_j v^{(j)}$. The numbers $\lambda_1, \ldots, \lambda_n \in \mathbb{K}$ are called the coordinates of a with respect to the basis $[v]$. □

We now discuss a few elementary, but important properties concerning the notion of a basis of a vector space.

Theorem 4.1.3. *Assume that V is a $\mathbb{K}$–vector space and $[v] = [v^{(1)}, \ldots, v^{(n)}]$ is a basis of V. Then every basis of V has precisely n elements.*

Definition 4.1.4. *(i) Assume that V is a $\mathbb{K}$–vector space with basis $[v] = [v^{(1)}, \ldots, v^{(n)}]$. Then*

$$\dim(V) = n$$

is referred to as the dimension of V (as a $\mathbb{K}$–vector space).
(ii) A $\mathbb{K}$–vector space is said to be finite dimensional if it admits a basis with finitely many elements.
(iii) If $V = \{0\}$, then $\dim(V) = 0$.

Examples:
(i) $V = \mathbb{R}^n$ is a $\mathbb{R}$–vector basis of dimension n. Standard basis:

$$e^{(1)} = (1, 0, \ldots, 0)\,,\ e^{(2)} = (0, 1, 0, \ldots, 0)\,,\ \ldots\,,\ e^{(n)} = (0, \ldots, 0, 1)\,.$$

(ii) The space $\mathcal{P}_n = \left\{p(z) = \sum_{k=0}^{n} a_k z^k : a_0, \ldots, a_n \in \mathbb{C}\right\}$ is a $\mathbb{C}$–vector space of dimension $n+1$. A basis $p^{(0)}, \ldots, p^{(n)}$ is given by

$$p^{(0)}(z) = 1\,,\ p^{(1)}(z) = z\,,\ \ldots\,,\ p^{(n)}(z) = z^n\,.$$

Theorem 4.1.4. *Assume that V is a $\mathbb{K}$–vector space of dimension n. Then:*
(i) Every subspace W of V is a finite dimensional $\mathbb{K}$–vector space and $\dim(W) \leq \dim(V)$.
(ii) If the vectors $v^{(1)}, \ldots, v^{(m)}$ span all of V, $V = \left\{\sum_{j=1}^{n} \lambda_j v^{(j)} : \lambda_1, \ldots, \lambda_m \in \mathbb{K}\right\}$, then the following holds: $(ii1)$ $m \geq n$, $(ii2)$ $v^{(1)}, \ldots, v^{(m)}$ is a basis of V if and only if $m = n$, and $(ii3)$ if $m > n$, then one can choose n elements among $v^{(1)}, \ldots, v^{(m)}$ which form a basis of V.
(iii) If the vectors $v^{(1)}, \ldots, v^{(m)}$ in V are $\mathbb{K}$–linearly independent then the following holds: $(iii1)$ $m \leq n$, $(iii2)$ $v^{(1)}, \ldots, v^{(m)}$ is a basis of V if and only if $m = n$, and $(iii3)$ if $m < n$, then there exist elements $w^{(1)}, \ldots, w^{(m-n)}$ so that $v^{(1)}, \ldots, v^{m}, w^{(1)}, \ldots, w^{(n-m)}$ form a basis of V.

As an application of the notion of a subspace of a vector space and its dimension, we introduce the notion of the rank of a matrix. For any $A \in \mathbb{K}^{m\times n}$, denote by $a^{(1)}, \ldots, a^{(n)}$ its columns and define the set

$$C_A := \Big\{ \sum_{j=1}^{n} \lambda_j a^{(j)} : \lambda_1, \ldots, \lambda_n \in \mathbb{K} \Big\} .$$

Then C_A is a $\mathbb{K}$–subspace of $\mathbb{K}^{m\times 1} \simeq \mathbb{K}^m$ and we define

Definition 4.1.5. *The rank of A, denoted by* $\mathrm{rank}(A)$ *is defined as*

$$\mathrm{rank}(A) := \dim(C_A) .$$

Note that $\mathrm{rank}(A) \leq \min\{m, n\}$, since $C_A \subseteq \mathbb{K}^{m\times 1}$.

Theorem 4.1.5. *For any $A \in \mathbb{K}^{m\times n}$,* $\mathrm{rank}(A^T) = \mathrm{rank}(A)$.

Remark 4.1.6. For any $A \in \mathbb{K}^{m\times n}$, denote by $A^{(1)}, \ldots, A^{(m)}$ the rows of A and introduce

$$R_A := \Big\{ \sum_{k=1}^{m} \lambda_k A^{(k)} : \lambda_1, \ldots, \lambda_m \in \mathbb{K} \Big\} \subseteq \mathbb{K}^{1\times n} .$$

Then one has $\dim(R_A) = \mathrm{rank}(A^T)$. □

Example: Determine the rank of the matrix $A \in \mathbb{R}^{3\times 4}$,

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

The vector space generated by the columns of A is

$$C_A := \Big\{ \sum_{j=1}^{4} \lambda_j a^{(j)} : \lambda_1, \ldots, \lambda_4 \in \mathbb{R} \Big\} .$$

where

$$a^{(1)} = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} , \quad a^{(2)} = \begin{pmatrix} 0 \\ 2 \\ 1 \end{pmatrix} , \quad a^{(3)} = \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix} , \quad a^{(4)} = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} .$$

We want to construct a basis of C_A, using Theorem 4.1.4. We first verify that $a^{(1)}, \ldots, a^{(4)}$ span all of $\mathbb{R}^{3\times 1} \simeq \mathbb{R}^3$. It means that for any given $b \in \mathbb{R}^3$, we have to find $x_1, x_2, x_3 \in \mathbb{R}$ so that

$$b = \sum_{j=1}^{3} x_j a^{(j)} .$$

In matrix notation, $Ax = b$. We solve for x by Gauss elimination. The augmented coefficient matrix is given by

$$\left(\begin{array}{cccc|c} 1 & 0 & 1 & 1 & b_1 \\ 1 & 2 & -1 & 2 & b_2 \\ 1 & 1 & 0 & 3 & b_3 \end{array}\right)$$

- $R_2 \rightsquigarrow R_2 - R_1$, $R_3 \rightsquigarrow R_3 - R_1$

$$\left(\begin{array}{cccc|c} 1 & 0 & 1 & 1 & b_1 \\ 0 & 2 & -2 & 1 & b_2 - b_1 \\ 0 & 1 & -1 & 2 & b_3 - b_1 \end{array}\right)$$

- $R_3 \rightsquigarrow R_3 - \frac{1}{2}R_2$

$$\left(\begin{array}{cccc|c} 1 & 0 & 1 & 1 & b_1 \\ 0 & 2 & -2 & 1 & b_2 - b_1 \\ 0 & 0 & 0 & \frac{3}{2} & b_3 - \frac{1}{2}b_1 - \frac{1}{2}b_2 \end{array}\right)$$

It then follows that for any $b \in \mathbb{R}^3$, the latter system has a solution. Then by Theorem 4.1.4, we know that $a^{(1)}, \ldots, a^{(4)}$ are linearly dependent. This can be also verified directly: consider the homogeneous system

$$\sum_{j=1}^{4} x_j v^{(j)} = 0\,, \quad (i.e., \quad b = 0)\,.$$

Then according to the above calculations,

$$x_4 = 0\,; \quad 2x_2 = 2x_3 - x_4\,, \quad \rightsquigarrow \quad x_2 = x_3\,; \quad x_1 = -x_3 - x_4 = -x_3\,.$$

With $x_3 = 1$ we get $x_2 = 1$, $x_1 = -1$ and hence

$$-a^{(1)} + a^{(2)} + a^{(3)} = 0\,.$$

We conclude that $a^{(2)}, a^{(3)}, a^{(4)}$ span $\mathbb{R}^3$. By Theorem 4.1.4 it then follows that $a^{(2)}, a^{(3)}, a^{(4)}$ is a basis of $\mathbb{R}^3$. As a consequence, $\operatorname{rank}(A) = 3$.
Application: Determine a basis of the $\mathbb{K}$–subspace $L \subset \mathbb{K}^{n\times 1}$,

$$L = \left\{x \in \mathbb{K}^{n\times 1} : Ax = 0\right\},$$

of solutions of the homogeneous system $Ax = 0$, where $A \in \mathbb{K}^{m\times n}$. To this end recall that in Chapter 1 we have seen that by possibly renumerating the variables $x_1, \ldots, x_n$, denoted by $y_1, \ldots, y_n$, the system $Ax = 0$ can be broguht by the row operations $(R1) - (R3)$ into the form $By = 0$ with

$$B = \begin{pmatrix} \underline{1} & 0 & \cdots & 0 & 0 & b_{1(k+1)} & \cdots & b_{1n} \\ 0 & \underline{1} & \cdots & 0 & 0 & b_{2(k+1)} & \cdots & b_{2n} \\ \vdots & & & & & \vdots & & \vdots \\ 0 & \cdots & & \overline{0} & \underline{1} & b_{k(k+1)} & \cdots & b_{kn} \\ 0 & \cdots & & & & 0 & \cdots & 0 \\ \vdots & & & & & & & \\ 0 & \cdots & & & & 0 & \cdots & 0 \end{pmatrix}$$

The solutions of $By = 0$ are given by

$$y_1 = -\sum_{j=k+1}^{n} b_{1j}t_j \, , \, \ldots \, , \, y_k = -\sum_{j=k+1}^{n} b_{kj}t_j$$

and $y_{k+1} := t_{k+1} \, , \, \ldots \, , \, y_n := t_n$.

Theorem 4.1.7. *The space of solutions*

$$\widetilde{L} = \left\{ y \in \mathbb{K}^{n\times 1} : By = 0 \right\}$$

is a subspace of $\mathbb{K}^{n\times 1}$ *with basis*

$$\widetilde{v}^{(1)} = \begin{pmatrix} -b_{1(k+1)} \\ \vdots \\ -b_{k(k+1)} \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \, , \, \ldots\ldots \, , \, \widetilde{v}^{(n-k)} = \begin{pmatrix} -b_{1n} \\ \vdots \\ -b_{kn} \\ 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix} .$$

By appropriately renumerating the unknowns $y_1, \ldots, y_n$*, one obtains a basis* $v^{(1)}, \ldots, v^{(n-k)}$ *of* $L = \left\{ x \in \mathbb{K}^{n\times 1} : Ax = 0 \right\}$.

A corresponding result holds for an inhomogeneous system of linear equations:

Theorem 4.1.8. *Let* $A \in \mathbb{K}^{m\times n}$ *and* $b \in \mathbb{K}^{m\times 1}$ *and assume that* $Ax = b$ *has at least one solution. Then the space of solutions*

$$L = \left\{ x \in \mathbb{K}^{n\times 1} : Ax = b \right\}$$

is an affine space,

$$L = x^{part} + L_{hom}$$

where $L_{hom} = \left\{ x \in \mathbb{K}^{n\times 1} : Ax = 0 \right\}$ *and* x^{part} *is an arbitrary solution of* $Ax = b$ *(referred to as particular solution).*

Remark 4.1.9. Note that $x^{part} + L_{hom} \subseteq L$, since for any $x \in L_{hom}$,

$$A(x^{part} + x) = A(x^{part}) + Ax = b + 0 = b.$$

Similarly, for any solution $y \in L$ one has

$$A(y - x^{part}) = Ay - Ax^{part} = b - b = 0$$

and hence $x := y - x^{part} \in L_{hom}$ or $y = x^{part} + x \in x^{part} + L_{hom}$. □

Let us discuss some further issues in form of exercises.

EXERCISE 1. Consider the following vectors in $\mathbb{R}^3$,

$$v^{(1)} = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \quad v^{(2)} = \begin{pmatrix} 0 \\ 2 \\ 1 \end{pmatrix}, \quad v^{(3)} = \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix}, \quad v^{(4)} = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}$$

(i) Verify that $v^{(1)}, \ldots, v^{(4)}$ span all of $\mathbb{R}^3$. It means that every element $b \in \mathbb{R}^3$ can be represented as a linear combination of $v^{(1)}, \ldots, v^{(4)}$, i.e., $b = \sum_{j=1}^4 \lambda_j v^{(j)}$, where $\lambda_1, \ldots, \lambda_4 \in \mathbb{R}$. This can be done by showing that for any $b \in \mathbb{R}^3$, the linear system $\sum_{j=1}^4 x_j v^{(j)} = b$ has at least one solution. We form the corresponding augmented coefficient matrix

$$\left(\begin{array}{cccc|c} 1 & 0 & 1 & 1 & b_1 \\ 1 & 2 & -1 & 2 & b_2 \\ 1 & 1 & 0 & 3 & b_3 \end{array}\right)$$

and use Gaussian elimination to get

$$R_2 \rightsquigarrow R_2 - R_1\,, \quad R_3 \rightsquigarrow R_3 - R_1\,, \qquad \left(\begin{array}{cccc|c} 1 & 0 & 1 & 1 & b_1 \\ 0 & 2 & -2 & 1 & b_2 - b_1 \\ 0 & 1 & -1 & 2 & b_3 - b_1 \end{array}\right)$$

$$R_3 \rightsquigarrow R_3 - \frac{1}{2} R_2\,, \qquad \left(\begin{array}{cccc|c} 1 & 0 & 1 & 1 & b_1 \\ 0 & 2 & -2 & 1 & b_2 - b_1 \\ 0 & 0 & 0 & \frac{3}{2} & b_3 - \frac{1}{2}b_1 - \frac{1}{2}b_2 \end{array}\right).$$

We conclude that $b = \sum_{j=1}^4 \lambda_j v^{(j)}$ with $\frac{3}{2}\lambda_4 = b_3 - \frac{1}{2}b_1 - \frac{1}{2}b_2$, $\lambda_3 = 0$, $2\lambda_2 = -\lambda_4 + b_2 - b_1$, and $\lambda_1 = -\lambda_4 + b_1$.

(ii) Can we form a basis of $\mathbb{R}^3$ by choosing three of the four vectors $v^{(1)}, \ldots, v^{(4)}$. How to proceed? Note that $\dim(\mathbb{R}^3) = 3$, and hence $v^{(1)}, \ldots, v^{(4)}$ are linearly dependent. We want to find a vector among $v^{(1)}, \ldots, v^{(4)}$ which can be expressed as a linear combination of the remaining three vectors. It means to find a non trivial solution of the homogeneous system $\sum_{j=1}^4 x_j v^{(j)} = 0$. According to the above scheme with $b = 0$ we get

$$\left(\begin{array}{cccc|c} 1 & 0 & 1 & 1 & 0 \\ 0 & 2 & -2 & 1 & 0 \\ 0 & 0 & 0 & \frac{3}{2} & 0 \end{array}\right).$$

Then $x_4 = 0$, $x_3 = 1$, $x_2 = 1$, and $x_1 = -x_3 = -1$, yielding

$$-v^{(1)} + v^{(2)} + v^{(3)} = 0\,, \quad \text{or} \quad v^{(1)} = v^{(2)} + v^{(3)}\,.$$

Since $v^{(1)}, \ldots, v^{(4)}$ span $\mathbb{R}^3$ and $\dim(\mathbb{R}^3) = 3$, $v^{(2)}, v^{(3)}, v^{(4)}$ form a basis of $\mathbb{R}^3$.

EXERCISE 2. Determine the dimension of the space of solutions of the homogeneous system $Ax = 0$, where $A \in \mathbb{R}^{3\times 5}$ is given by

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

First note that $L_{hom} = \{x \in \mathbb{R}^{5\times 1} : Ax = 0\}$ is a subspace of $\mathbb{R}^{5\times 1}(\simeq \mathbb{R}^5)$, since for any $x, x' \in L_{hom}$ and any $\lambda \in \mathbb{R}$

$$A(x + x') = Ax + Ax' = 0\,, \quad A(\lambda x) = \lambda Ax = 0\,.$$

To determine the dimension of L_{hom}, we have to find a basis and the dimension of L_{hom} is then given by the number of the vectors of this basis. To this end use Gaussian elimination to bring the coefficient matrix A into the refined row echelon form

$$R_2 \rightsquigarrow R_2 - 2R_1\,, \quad R_3 \rightsquigarrow R_3 - 2R_1\,, \qquad \left(\begin{array}{ccccc|c} 1 & 1 & 1 & 1 & 1 & 0 \\ 0 & -1 & -3 & 1 & -2 & 0 \\ 0 & -1 & 0 & -1 & 0 & 0 \end{array}\right)$$

$$R_3 \rightsquigarrow R_3 - R_2\,, \qquad \left(\begin{array}{ccccc|c} 1 & 1 & 1 & 1 & 1 & 0 \\ 0 & -1 & -3 & 1 & -2 & 0 \\ 0 & 0 & 3 & -2 & 2 & 0 \end{array}\right)$$

$$R_2 \rightsquigarrow R_2 + R_3\,, \quad R_1 \rightsquigarrow R_1 - \frac{1}{3}R_3\,, \qquad \left(\begin{array}{ccccc|c} 1 & 1 & 0 & \frac{5}{3} & \frac{1}{3} & 0 \\ 0 & -1 & 0 & -1 & 0 & 0 \\ 0 & 0 & 3 & -2 & 2 & 0 \end{array}\right)$$

$$R_1 \rightsquigarrow R_1 + R_2\,, \qquad \left(\begin{array}{ccccc|c} 1 & 0 & 0 & \frac{2}{3} & \frac{1}{3} & 0 \\ 0 & -1 & 0 & -1 & 0 & 0 \\ 0 & 0 & 3 & -2 & 2 & 0 \end{array}\right)$$

$$R_2 \rightsquigarrow -R_2\,, \quad R_3 \rightsquigarrow \frac{1}{3}R_3\,, \qquad \left(\begin{array}{ccccc|c} 1 & 0 & 0 & \frac{2}{3} & \frac{1}{3} & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & -\frac{2}{3} & \frac{2}{3} & 0 \end{array}\right)\,.$$

Hence x_4, x_5 are free variables and any solution of the system $Ax = 0$ has the form

$$x = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{pmatrix} = \begin{pmatrix} -\frac{2}{3}x_4 - \frac{1}{3}x_5 \\ -x_4 \\ \frac{2}{3}x_4 - \frac{2}{3}x_5 \\ x_4 \\ x_5 \end{pmatrix} = x_4 \begin{pmatrix} -\frac{2}{3} \\ -1 \\ \frac{2}{3} \\ 1 \\ 0 \end{pmatrix} + x_5 \begin{pmatrix} -\frac{1}{3} \\ 0 \\ -\frac{2}{3} \\ 0 \\ 1 \end{pmatrix}\,.$$

It means that x is a linear conbination of

$$v^{(1)} = \begin{pmatrix} -\frac{2}{3} \\ -1 \\ \frac{2}{3} \\ 1 \\ 0 \end{pmatrix}\,, \qquad v^{(2)} = \begin{pmatrix} -\frac{1}{3} \\ 0 \\ -\frac{2}{3} \\ 0 \\ 1 \end{pmatrix}$$

Since x_4, x_5 are free variables, $v^{(1)}, v^{(2)}$ are linearly independent, therefore they are a basis of L_{hom}.

EXERCISE 3. Given $A \in \mathbb{R}^{m\times n}$ and $b \in \mathbb{R}^{m\times 1}$, represent the set L of solutions of the linear system $Ay = b$,

$$L = \big\{y \in \mathbb{R}^{n\times 1} : Ay = b\big\},$$

using the space of solutions of the homogeneous linear system

$$L_{hom} = \big\{x \in \mathbb{R}^{n\times 1} : Ax = 0\big\}\,.$$

This can be achieved in two steps:
STEP 1. Determine a solution y of $Ay = b$. We denote such a solution by y^{part} and refer to it as a particular solution of $Ay = b$.
STEP 2. Determine L_{hom} using Gaussian elimination. The set L is then given by

$$L = y^{part} + L_{hom} = \big\{y^{part} + x : x \in L_{hom}\big\}\,.$$

To verify this identity we show that $L \subseteq y^{part} + L_{hom}$ (claim 1) and $y^{part} + L_{hom} \subseteq L$ (claim 2). To verify claim 1, let $y \in L$. Then $Ay = b$. Since $Ay^{part} = b$, it follows that

$$A(y - y^{part}) = Ay - Ay^{part} = b - b = 0\,,$$

hence $y - y^{part} \in L_{hom}$ and

$$y = y^{part} + (y - y^{part}) \in y^{part} + L_{hom}\,.$$

To verify claim 2, let $x \in L_{hom}$. Then $Ax = 0$ and one has

$$A(y^{part} + x) = Ay^{part} + Ax = b + 0 = b\,,$$

hence $y^{part} + x \in L$. Note that for $b = 0$, $L = L_{hom}$ is a subspace of $\mathbb{R}^{n\times 1}$. In contrast, for $b \neq 0$, $y^{part} \notin L_{hom}$ and $y^{part} + L_{hom}$ is not a subspace of $\mathbb{R}^{n\times 1}$. Such subsets are translates of subspaces and referred to as affine subspaces.

4.2 Linear maps

Assume that V, W are $\mathbb{R}$–vector spaces.

Definition 4.2.1. *A map $f : V \to W$ is said to be $\mathbb{R}$–linear (or linear for short) if it is compatible with the vector space structures of V and W. It means that*

(L1) $f(u + v) = f(u) + f(v) \quad \forall u, v \in V$

(L2) $f(\lambda u) = \lambda f(u) \quad \forall \lambda \in \mathbb{R},\ u \in V$

Remark 4.2.1. Similarly, if V, W are $\mathbb{C}$-vector spaces, a map $f : V \to W$ is said to be $\mathbb{C}$-linear if for any $u, v \in V$, $\lambda \in \mathbb{C}$ the identities $(L1), (L2)$ hold. □

From $(L1)$–$(L2)$ it follows that for any $v^{(1)}, \ldots, v^{(n)} \in V$, $\lambda_1, \ldots, \lambda_n \in \mathbb{R}$,

$$f\Big(\sum_{j=1}^{n} \lambda_j v^{(j)}\Big) = \sum_{j=1}^{n} \lambda_j f(v^{(j)})\,.$$

Lemma 4.2.2. *If $f : V \to W$ is a linear map, then $f(0_V) = 0_W$ where 0_V denotes the zero element of V and 0_W the one of W.*

Note that the above lemma can be verified in a straightforward way: using that $0_V = 0_V + 0_V$, it follows that $f(0_V) = f(0_V) + f(0_V) = 2f(0_V)$ implying that

$$0_W = f(0_V) - f(0_V) = 2f(0_V) - f(0_V) = f(0_V)\,.$$

Examples:
(*i*) Let $V = W = \mathbb{R}$ and $a \in \mathbb{R}$. Then $f : \mathbb{R} \to \mathbb{R}\,, x \mapsto ax$ is a linear map. Indeed, $(L1), (L2)$ are clearly satisfied.
(*ii*) Let $V = \mathbb{R}^{n\times 1}$, $W = \mathbb{R}^{m\times 1}$. Then for any $A \in \mathbb{R}^{m\times n}$

$$f_A : \mathbb{R}^{n\times 1} \to \mathbb{R}^{m\times 1}\,, \quad x \mapsto Ax = \begin{pmatrix} \sum_{j=1}^n a_{1j}x_j \\ \vdots \\ \sum_{j=1}^n a_{mj}x_j \end{pmatrix}$$

is linear since $(L1), (L2)$ are clearly satisfied. The map f_A is called the linear map associated to the matrix A.
(*iii*) Let $f : \mathbb{R}^2 \to \mathbb{R}^2$ be the map which maps an arbitrary point $x \in \mathbb{R}^2$ to the point obtained by reflecting x over the x_1-axis. The map f is given by $f : (x_1, x_2) \to (x_1, -x_2)\,.$ Identifying $\mathbb{R}^2$ and $\mathbb{R}^{2\times 1}$ one sees that f can be represented as

$$f(x_1, x_2) = (x_1, -x_2) = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$

hence by the previous example, f is linear.
(*iv*) Let $R(\varphi)$ denote the map $\mathbb{R}^2 \to \mathbb{R}^2$, defined by rotating a given vector $x \in \mathbb{R}^2$ counterclockwise by the angle φ (modulo 2π). Note that

$$(1, 0) \mapsto (\cos\varphi, \sin\varphi)\,, \quad (0, 1) \mapsto (-\sin\varphi, \cos\varphi)\,.$$

Identifying again $\mathbb{R}^2$ and $\mathbb{R}^{2\times 1}$ one can verify in a straightforward way that

$$f \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = A \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}\,, \quad A := \begin{pmatrix} \cos\varphi & -\sin\varphi \\ \sin\varphi & \cos\varphi \end{pmatrix}\,,$$

hence by item (*ii*), f is linear.
(*v*) Let $f : \mathbb{R} \to \mathbb{R}$, $x \mapsto x^2$. Then f is not linear since $f(2) = 2^2 = 4$, but $f(1) + f(1) = 2$, implying that $f(1 + 1) \neq 2f(1)$, which violates $(L1)$.
(*vi*) Let $f : \mathbb{R} \to \mathbb{R}$, $x \mapsto 1 + 2x$. Then f is not linear since $f(0) = 1$, $f(1) = 3$, and hence $f(1 + 0) \neq f(1) + f(0)$, violating $(L1)$.

Theorem 4.2.3. *For any linear map $f : \mathbb{R}^n \to \mathbb{R}^m$, there exists a matrix $A \in \mathbb{R}^{m\times n}$ so that $f(x) = Ax$, for any $x \in \mathbb{R}^n$. (Here we again identify $\mathbb{R}^{n\times 1}$ with $\mathbb{R}^n$.) The matrix A is referred to as the matrix representation of f with respect to the standard bases $[e_{\mathbb{R}^n}] = [e^{(1)}_{\mathbb{R}^n}, \ldots, e^{(n)}_{\mathbb{R}^n}]$ of $\mathbb{R}^n$ and $[e_{\mathbb{R}^m}] = [e^{(1)}_{\mathbb{R}^m}, \ldots, e^{(m)}_{\mathbb{R}^m}]$ of $\mathbb{R}^m$.*

The following notation for A turns out to be very convenient:

$$A = f_{[e_{\mathbb{R}^n}]\mapsto[e_{\mathbb{R}^m}]}\,.$$

Remark 4.2.4. In the sequel, we will often drop the subscript $\mathbb{R}^n$ in $e^{(j)}_{\mathbb{R}^n}$ and simply write $e^{(j)}$. □

To get more familiar with the notion of a linear map, let us determine the matrix A. Given $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$, we write $x = \sum_{j=1}^n x_j e^{(j)}_{\mathbb{R}^n}$ and get $f(x) = f\Big(\sum_{j=1}^n x_j e^{(j)}_{\mathbb{R}^n}\Big) = \sum_{j=1}^n x_j f(e^{(j)}_{\mathbb{R}^n})$. With $f(e^{(j)}_{\mathbb{R}^n}) = (a_{1j}, \ldots, a_{mj}) \in \mathbb{R}^m$, one has

$$f(x) = \begin{pmatrix} \sum_{j=1}^n a_{1j}x_j \\ \vdots \\ \sum_{j=1}^n a_{mj}x_j \end{pmatrix} = Ax\,,$$

where A is the $m \times n$ matrix with columns given by

$$a^{(1)} := \begin{pmatrix} a_{11} \\ \vdots \\ a_{m1} \end{pmatrix}\,, \ldots\ldots\,, a^{(n)} := \begin{pmatrix} a_{1n} \\ \vdots \\ a_{mn} \end{pmatrix}\,.$$

More generally, assume that V is a $\mathbb{R}$–vector space of dimension n with basis $[v] = [v^{(1)}, \ldots, v^{(n)}]$, W a $\mathbb{R}$–vector space of dimension m with basis $[w] = [w^{(1)}, \ldots, w^{(m)}]$, and $f : V \to W$ a linear map. Then the image $f(v^{(j)})$ of the vector $v^{(j)}$ can be written uniquely as a linear conbination of the vectors of the basis $[w]$, namely

$$f(v^{(j)}) = \sum_{i=1}^m a_{ij} w^{(i)}\,.$$

Definition 4.2.2. *The matrix $A = (a_{ij})_{\substack{1\le i\le m\\ 1\le j\le n}}$ is called the matrix representation of f with respect to the basis $[v]$ of V and $[w]$ of W and is denoted by $f_{[v]\to[w]}$.*

Theorem 4.2.5. *Assume that $f : V \to W$ is a linear map, $x = \sum_{j=1}^n x_j v^{(j)}$, and $f(x) = \sum_{i=1}^m y_i w^{(i)}$. Then $y = (y_1, \ldots, y_m)$ is related to $x = (x_1, \ldots, x_n)$ by*

$$y = Ax\,.$$

In words: the coordinates of $f(x)$ with respect to the basis $[w]$ can be computed from the coordinates of x with respect to the basis $[v]$ with the $m \times n$ matrix $f_{[v]\to[w]}$, i.e., $y = f_{[v]\to[w]}x$.

To get more familiar with linear maps and their matrix representations, let us verify the statement of the above theorem: since $f(v^{(j)}) = \sum_{i=1}^m a_{ij} w^{(i)}$ one has

$$f\Big(\sum_{j=1}^n x_j v^{(j)}\Big) = \sum_{j=1}^n x_j f(v^{(j)}) = \sum_{j=1}^n x_j \sum_{i=1}^m a_{ij} w^{(i)} = \sum_{i=1}^m \Big(\sum_{j=1}^n a_{ij}x_j\Big) w^{(i)}\,.$$

So $\sum_{j=1}^n a_{ij}x_j$ is the i-th coordinate of the vector $f\Big(\sum_{j=1}^n x_j v^{(j)}\Big)$ with respect to the basis $[w]$.

We now discuss the relation between the composition of linear maps and the definition of matrix multiplication, mentioned already in Chapter 1. Assume that V is a $\mathbb{R}$–vector space of dimension n with basis $[v] = [v^{(1)}, \dots, v^{(n)}]$, W a $\mathbb{R}$–vector space of dimension m with basis $[w] = [w^{(1)}, \dots, w^{(m)}]$ and U a $\mathbb{R}$–vector space of dimension k with basis $[u] = [u^{(1)}, \dots, u^{(k)}]$. Furthermore, assume that

$$f : V \to W\,, \qquad g : W \to U$$

are linear maps with matrix representations $A = f_{[v]\to[w]}$ and $B = g_{[w]\to[u]}$. The following theorem says how to compute the matrix representation $(g \circ f)_{[v]\to[u]}$ of the linear map $g \circ f : V \to U$,

$$g \circ f : V \xrightarrow{f} W \xrightarrow{g} U\,.$$

Theorem 4.2.6.

$$(g \circ f)_{[v]\to[u]} = g_{[w]\to[u]} \cdot f_{[v]\to[w]}\,.$$

To get more familiar with the composition of linear maps, let us verify the statement of the above theorem. With $x = \sum_{j=1}^n x_j v^{(j)}$ and $f(v^{(j)}) = \sum_{i=1}^m a_{ij} w^{(i)}$ one gets

$$f(x) = \sum_{i=1}^m \Big(\sum_{j=1}^n a_{ij}x_j\Big) w^{(i)}\,.$$

Writing $g(w^{(i)}) = \sum_{\ell=1}^k b_{\ell i} u^{(\ell)}$, one then concludes that

$$\begin{aligned} g(f(x)) &= \sum_{i=1}^m \Big(\sum_{j=1}^n a_{ij}x_j\Big) g(w^{(i)}) = \sum_{i=1}^m \Big(\sum_{j=1}^n a_{ij}x_j\Big) \sum_{\ell=1}^k b_{\ell i} u^{(\ell)} \\ &= \sum_{\ell=1}^k \Big(\sum_{j=1}^n \Big(\sum_{i=1}^m b_{\ell i} a_{ij}\Big) x_j\Big) u^{(\ell)}\,. \end{aligned}$$

But $\sum_{i=1}^m b_{\ell i} a_{ij} = (BA)_{\ell j}$ and hence

$$g(f(x)) = \sum_{\ell=1}^k z_\ell u^{(\ell)}\,, \quad z_\ell = \sum_{j=1}^n (BA)_{\ell j} x_j = (BAx)_\ell\,, \quad 1 \le \ell \le k\,.$$

Special case: $V = W$. In this case one often chooses the same basis $[v] = [v^{(1)}, \dots, v^{(n)}]$ in the domain and the target of a linear map $f : V \to W$ and $f_{[v]\to[v]}$ is said to be the matrix representation of f with respect to $[v]$. The $n \times n$ matrix $f_{[v]\to[v]}$ has columns $a^{(j)}$, $1 \le j \le n$, whose coefficients are the coordinates of $f(v^{(j)})$ with respect to the basis $[v]$,

$$f(v^{(j)}) = \sum_{i=1}^n a_{ij} v^{(i)}\,.$$

EXAMPLES:
(i) Let us consider the rotation $R(\varphi) : \mathbb{R}^2 \to \mathbb{R}^2$, introduced earlier, and denote by $[e] = [e^{(1)}, e^{(2)}]$ the standard basis of $\mathbb{R}^2$. The two columns $a^{(1)}$, $a^{(2)}$ of $R(\varphi)_{[e]\to[e]}$ are then computed as follows:

$$R(\varphi)e^{(1)} = \begin{pmatrix} \cos\varphi \\ \sin\varphi \end{pmatrix} = \cos\varphi\, e^{(1)} + \sin\varphi\, e^{(2)}\,,$$

implying that $a^{(1)} = \begin{pmatrix} \cos\varphi \\ \sin\varphi \end{pmatrix}$. Similarly

$$R(\varphi)e^{(2)} = \begin{pmatrix} -\sin\varphi \\ \cos\varphi \end{pmatrix} = -\sin\varphi\, e^{(1)} + \cos\varphi\, e^{(2)}\,,$$

yielding $a^{(2)} = \begin{pmatrix} -\sin\varphi \\ \cos\varphi \end{pmatrix}$. Altogether one obtains

$$R(\varphi)_{[e]\to[e]} = \begin{pmatrix} \cos\varphi & -\sin\varphi \\ \sin\varphi & \cos\varphi \end{pmatrix}$$

(ii) Let $w^{(1)} = (1, 0)$, $w^{(2)} = (1, 1)$. To compute $R(\varphi)_{[w]\to[w]}$ we proceed as follows : write

$$R(\varphi)w^{(1)} = \begin{pmatrix} \cos\varphi \\ \sin\varphi \end{pmatrix} = a_{11}w^{(1)} + a_{21}w^{(2)}$$

and determine a_{11}, a_{21} by solving the linear system

$$\begin{cases} a_{11} + a_{21} = \cos\varphi \\ a_{21} = \sin\varphi \end{cases} .$$

Hence $a_{21} = \sin\varphi$ and $a_{11} = \cos\varphi - \sin\varphi$. Similarly, to find the coordinate a_{21}, a_{22} of

$$R(\varphi)w^{(2)} = \begin{pmatrix} \cos\varphi - \sin\varphi \\ \sin\varphi + \cos\varphi \end{pmatrix}$$

with respect to the basis $[w]$ we need to solve

$$\begin{cases} a_{12} + a_{22} = \cos\varphi - \sin\varphi \\ a_{22} = \sin\varphi + \cos\varphi \end{cases}$$

yielding $a_{22} = \sin\varphi + \cos\varphi$ and $a_{12} = -2\sin\varphi$. Altogether, one then obtains

$$R(\varphi)_{[w]\to[w]} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} = \begin{pmatrix} \cos\varphi - \sin\varphi & -2\sin\varphi \\ \sin\varphi & \sin\varphi + \cos\varphi \end{pmatrix} .$$

With the introduced notations we can also express the matrix representation of the identity map with respect to bases $[v]$ and $[w]$ of a given vector space : assume that V is a $\mathbb{R}$–vector space of dimension n and that

$$[v] = [v^{(1)}, \ldots, v^{(n)}]\,, \quad [w] = [w^{(1)}, \ldots, w^{(n)}]$$

are two bases of V. The matrix representation of the identity map $Id : V \to V$ with respect to the bases $[v]$ and $[w]$ is then given by

$$\mathrm{Id}_{[v]\to[w]}\,.$$

The meaning of this matrix is the following one: given any vector u in V, write u as a linear combination with respect to the two bases $[v]$ and $[w]$, $u = \sum_{j=1}^n x_j v^{(j)}$ and $u = \sum_{j=1}^n y_j w^{(j)}$. The coordinate vectors $x = (x_j)_{1\le j\le n}$ and $y = (y_j)_{1\le j\le n}$ are then related by the formula $y = \mathrm{Id}_{[v]\to[w]}x$ and the matrix $\mathrm{Id}_{[v]\to[w]}$ is referred to as the matrix of the change of basis $[v]$ to the basis $[w]$.

The j-th column of the above matrix is given by the coordinates of $v^{(j)}$ with respect to the basis $[w]$, $v^{(j)} = \sum_{i=1}^n s_{ij} w^{(i)}$. Note that $\mathrm{Id}_{[v]\to[v]} = \mathrm{Id}_n$, where Id_n is the standard $n \times n$ identity matrix in $\mathbb{R}^{n\times n}$ and $\mathrm{Id}_{[v]\to[w]} \cdot \mathrm{Id}_{[w]\to[v]} = \mathrm{Id}_{[v]\to[v]} = \mathrm{Id}_n$, yielding

$$\mathrm{Id}_{[w]\to[v]} = \left(\mathrm{Id}_{[v]\to[w]}\right)^{-1}.$$

Theorem 4.2.7. *Assume that V is a $\mathbb{R}$–vector space of dimension n, $[v]$, $[w]$ are bases of V and $f : V \to V$ is a linear map. Then*

$$f_{[w]\to[w]} = \mathrm{Id}_{[w]\to[v]} \cdot f_{[v]\to[v]} \cdot \mathrm{Id}_{[v]\to[w]} = (\mathrm{Id}_{[v]\to[w]})^{-1} \cdot f_{[v]\to[v]} \cdot \mathrm{Id}_{[v]\to[w]}\,.$$

Examples: Assume that $V = W = \mathbb{R}^2$ and let $[v] = [v^{(1)}, v^{(2)}]$ be the basis with $v^{(1)} = (1, 1)$, $v^{(2)} = (1, -1)$. As usual, $[e] = [e^{(1)}, e^{(2)}]$ denotes the standard basis.
(i) Compute $\mathrm{Id}_{[v]\to[e]}$: write $v^{(1)} = 1 \cdot e^{(1)} + 1 \cdot e^{(2)}$ and $v^{(2)} = 1 \cdot e^{(1)} + (-1) \cdot e^{(2)}$. Hence

$$\mathrm{Id}_{[v]\to[e]} = (a^{(1)}\ a^{(2)}), \quad \text{with} \quad a^{(1)} = \begin{pmatrix}1\\1\end{pmatrix}, \quad a^{(2)} = \begin{pmatrix}1\\-1\end{pmatrix}.$$

(ii) Compute $\mathrm{Id}_{[e]\to[v]}$: it suffices to compute $(\mathrm{Id}_{[v]\to[e]})^{-1}$. It is given by $\begin{pmatrix}\frac{1}{2} & \frac{1}{2}\\ \frac{1}{2} & -\frac{1}{2}\end{pmatrix}$.
(iii) Consider the counterclockwise rotation $R(\varphi)$ in $\mathbb{R}^2$ with angle φ (modulo 2π).
$(iii1)$ Compute $R(\varphi)_{[e]\to[e]}$: we have

$$R(\varphi)e^{(1)} = \cos\varphi\, e^{(1)} + \sin\varphi\, e^{(2)}, \quad R(\varphi)e^{(2)} = -\sin\varphi\, e^{(1)} + \cos\varphi\, e^{(2)}$$

or

$$R(\varphi)_{[e]\to[e]} = \begin{pmatrix}\cos\varphi & -\sin\varphi\\ \sin\varphi & \cos\varphi\end{pmatrix}$$

$(iii2)$ Compute $R(\varphi)_{[v]\to[e]}$: we have

$$R(\varphi)_{[v]\to[e]} = R(\varphi)_{[e]\to[e]} \cdot \mathrm{Id}_{[v]\to[e]}\,.$$

Combining (i) and $(iii1)$ we get

$$R(\varphi)_{[v]\to[e]} = \begin{pmatrix}\cos\varphi & -\sin\varphi\\ \sin\varphi & \cos\varphi\end{pmatrix}\begin{pmatrix}1 & 1\\ 1 & -1\end{pmatrix} = \begin{pmatrix}\cos\varphi - \sin\varphi & \cos\varphi + \sin\varphi\\ \sin\varphi + \cos\varphi & \sin\varphi - \cos\varphi\end{pmatrix}$$

(*iii*3) Compute $R(\varphi)_{[v]\to[v]}$: we have

$$\begin{aligned} R(\varphi)_{[v]\to[v]} &= \mathrm{Id}_{[e]\to[v]} \cdot R(\varphi)_{[e]\to[e]} \mathrm{Id}_{[v]\to[e]} \\ &= \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} \end{pmatrix} \begin{pmatrix} \cos\varphi - \sin\varphi & \cos\varphi + \sin\varphi \\ \sin\varphi + \cos\varphi & \sin\varphi - \cos\varphi \end{pmatrix} \\ &= \begin{pmatrix} \cos\varphi & \sin\varphi \\ -\sin\varphi & \cos\varphi \end{pmatrix} \end{aligned}$$

Note that one has to distinguish between linear maps and their representations. Linear maps are independent of a choice of bases and a matrix can be the representation of many different linear maps, depending on the choice of bases made. In the case where f is a linear map of a vector space into itself, this fact led to the following

Definition 4.2.3. *Two matrices $A, B \in \mathbb{R}^{n\times n}$ are said to be similar if there exists a matrix $S \in \mathrm{GL}_{\mathbb{R}}(n)$ with $B = S^{-1}AS$.*

Note that the matrix S can be interpreted as $\mathrm{Id}_{[v]\to[e]}$, with $[v] = [v^{(1)}, \ldots, v^{(n)}]$ given by the columns of S, implying that B is the matrix representation of the linear map $f_A : \mathbb{R}^n \to \mathbb{R}^n$, $x \mapsto Ax$ with respect to the basis $[v]$.

Examples: (*i*) For $A = \lambda \mathrm{Id}_n$, $\lambda \in \mathbb{R}$, the set of all matrices similar to A is determined as follows: let $S \in \mathrm{GL}_{\mathbb{R}}(n)$. Then

$$S^{-1}AS = S^{-1}\lambda \mathrm{Id}_n S = \lambda S^{-1}S = \lambda \mathrm{Id}_n = A\,.$$

In words: the matrix representation of $f_{\lambda \mathrm{Id}_n} : \mathbb{R}^n \to \mathbb{R}^n\,, x \mapsto \lambda x$, with respect to any basis of $\mathbb{R}^n$ is the matrix $\lambda \mathrm{Id}_n$.
(*ii*) Determine if

$$A = \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix} \qquad \text{and} \qquad B = \begin{pmatrix} 3 & 2 \\ -1 & 0 \end{pmatrix}$$

are similar or not. It turns out that the regular matrix $S = \begin{pmatrix} 1 & 2 \\ -1 & -1 \end{pmatrix}$ satisfies $B = S^{-1}AS$. We will see later how S can be found in a systematic way.

Definition 4.2.4. *Let V be a $\mathbb{R}$–vector space of dimension n and $f : V \to V$ linear. Then f is said to be diagonalizable if there exists a basis $[v] = [v^{(1)}, \ldots, v^{(n)}]$ of V so that the matrix representation $f_{[v]\to[v]}$ of f with respect to the basis $[v]$ is diagonal.*

Note that in the example above, the linear map

$$f_B : \mathbb{R}^2 \to \mathbb{R}^2\,, \qquad x \mapsto Bx$$

is diagonalizable: with $v^{(1)}, v^{(2)}$ given by the columns of

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

one has $S^{-1} = \mathrm{Id}_{[v]\to[e]}$ and since $B = (f_B)_{[e]\to[e]}$ it follows that

$$\mathrm{Id}_{[e]\to[v]} B \mathrm{Id}_{[v]\to[e]} = SBS^{-1} = A = (f_B)_{[v]\to[v]} .$$

In the next chapter, we will characterize a whole class of linear maps $f : \mathbb{R}^n \to \mathbb{R}^n$ which are diagonalizable.

4.3 Inner products on $\mathbb{R}$–vector spaces

Often in applications we are given a vector space with additional geometric structures, allowing e.g. to measure the length of a vector or the angle between two nonzero vectors. In this section, we introduce such an additional geometric structure, called inner product, for $\mathbb{R}$–vector spaces. In a later section we will consider the corresponding notion for $\mathbb{C}$–vector spaces.

Definition 4.3.1. *A map*

$$\langle \cdot, \cdot \rangle : V \times V \to \mathbb{R}$$

is said to be an inner product (or scalar product) for the $\mathbb{R}$–vector space V if the following conditions are satisfied:

(IP1) $\langle v, w \rangle = \langle w, v \rangle , \qquad \forall v, w \in V$

(IP2) $\forall v, w, u \in V$, $\forall \lambda \in \mathbb{R}$

$$\langle v + w, u \rangle = \langle v, u \rangle + \langle w, u \rangle , \qquad \langle \lambda v, w \rangle = \lambda \langle v, w \rangle$$

(IP3) $\langle v, v \rangle \geq 0$, $\forall v \in V$ *and* $\langle v, v \rangle = 0$ *if and only if* $v = 0$.

In words: $\langle \cdot, \cdot \rangle$ is a symmetric, positive definite, real valued, bilinear map.
In case $\dim(V) < \infty$, *the pair $(V, \langle \cdot, \cdot \rangle)$ is said to be a finite dimensional Hilbert space.*

If a vector space is given an inner product $\langle \cdot, \cdot \rangle$ one can define the notion of the length (or norm) of a vector as follows:

$$\|v\| := (\langle v, v \rangle)^{\frac{1}{2}} .$$

$\|v\|$ is referred to as the norm of v induced by $\langle \cdot, \cdot \rangle$ and the following holds:

$$\|v\| \geq 0 , \quad \forall v \in V \quad \text{and} \quad \|v\| = 0 \quad \text{iff} \quad v = 0$$

and

$$\|\lambda v\| = |\lambda| \|v\| \qquad \forall \lambda \in \mathbb{R} , \quad \forall v \in V .$$

In this section, we will always assume that V is equipped with an inner product $\langle \cdot, \cdot \rangle$. The following fundamental inequality holds:

Lemma 4.3.1 (Cauchy-Schwartz inequality)**.**

$$|\langle v, w\rangle| \leq \|v\|\|w\|\,, \qquad \forall v, w \in V\,.$$

Why is this inequality true? Obviously, it holds if $v = 0$ or $w = 0$, so let us assume that $v \neq 0$ and $w \neq 0$. Let us consider the vector $v + tw$, where $t \in \mathbb{R}$ is arbitrary. Then

$$0 \leq \|v + tw\|^2 = \langle v + tw, v + tw\rangle = \langle v, v\rangle + 2t\langle v, w\rangle + t^2\langle w, w\rangle\,.$$

Let us find the minimum of $\|v + tw\|^2$ when viewed as a function of t. To this end we want to determine the values of t with

$$\frac{d}{dt}\Big(\langle v, v\rangle + 2t\langle v, w\rangle + t^2\langle w, w\rangle\Big) = 0$$

leading to the following equation $2t\|w\|^2 + 2\langle v, w\rangle = 0$. Hence such a t is uniquely determined

$$t_0 = -\frac{\langle v, w\rangle}{\|w\|^2}$$

and we conclude that $0 \leq \langle v, v\rangle + 2t_0\langle v, w\rangle + t_0^2\langle w, w\rangle$ yielding the Cauchy-Schwarz inequality

$$0 \leq \|v\|^2\|w\|^2 - (\langle v, w\rangle)^2 \quad \text{or} \quad |\langle v, w\rangle| \leq \|v\|^2\|w\|^2\,.$$

Note that for $v, w \in V$ with $\|v\| = \|w\| = 1$ the Cauchy-Schwarz inequality yields

$$|\langle v, w\rangle| \leq 1 \quad \text{or} \qquad -1 \leq \langle v, w\rangle \leq 1\,.$$

Hence there exists a unique angle $0 \leq \varphi \leq \pi$ with

$$\langle v, w\rangle = \cos\varphi\,.$$

Definition 4.3.2. *For any vectors $v, w \in V \setminus \{0\}$, the uniquely determined real number $0 \leq \varphi \leq \pi$ with*

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

is said to be the non–oriented angle between v and w. (The non–oriented angle does not allow to determine whether $\frac{v}{\|v\|}$ needs to be rotated clockwise or counterclockwise to be mapped to $\frac{w}{\|w\|}$.)

As a consequence, one has for any $v, w \in V \setminus \{0\}$,

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

Definition 4.3.3. *The vectors $v, w \in V$ are orthogonal to each other if $\langle v, w\rangle = 0$.*

Note that if $v = 0$ or $w = 0$, one always has $\langle v, w\rangle = 0$ whereas for $v, w \in V \setminus \{0\}$

$$0 = \langle v, w\rangle = \|v\|\|w\|\cos\varphi$$

implies $\varphi = \pi/2$. As an easy application of the notion of an inner product one obtains the following well known result:

Theorem 4.3.2 (Theorem of Pythagoras)**.** *For any $v, w \in V$ which are orthogonal to each others one has*

$$\|v+w\|^2 = \|v\|^2 + \|w\|^2 .$$

This identity follows immediately from the assumption $\langle v, w\rangle = 0$ and the identity

$$\|v+w\|^2 = \langle v+w, v+w\rangle = \langle w, w\rangle + 2\langle v, w\rangle + \langle v, v\rangle .$$

Theorem 4.3.3 (Law of cosines)**.** *For any $v, w \in V \setminus \{0\}$ one has*

$$\|v-w\|^2 = \|v\|^2 + \|w\|^2 - 2\|v\|\|w\|\cos\varphi$$

where φ is the non–oriented angle between v and w.

The above theorems are indeed generalizations of the corresponding well known results in the Euclidean plane $\mathbb{R}^2$. This is explained with the following :
EXAMPLE 1. Let $V = \mathbb{R}^2$ and define

$$\langle\cdot,\cdot\rangle_{Euclid} : \mathbb{R}^2 \times \mathbb{R}^2 \to \mathbb{R} , \quad (x, y) \mapsto x_1y_1 + x_2y_2$$

where $x = (x_1, x_2)$, $y = (y_1, y_2)$. One easily verifies that $\langle\cdot,\cdot\rangle_{Euclid}$ is an inner product for $\mathbb{R}^2$. The corresponding norm $\|x\|$ is given by

$$\|x\| := \sqrt{\langle x, x\rangle} = \sqrt{x_1^2 + x_2^2}$$

and the angle φ between vectors $x, y \in \mathbb{R}^2$ of norm 1 can be computed as follows. Using polar coordinates one has

$$x = (\cos\theta, \sin\theta) , \qquad y = (\cos\psi, \sin\psi) , \qquad 0 \le \theta, \psi < 2\pi .$$

Then the angle $0 \le \varphi \le \pi$ defined by

$$\langle x, y\rangle = \|x\|\|y\|\cos\varphi = \cos\varphi$$

is determined as follows: by the addition formula for the cosine,

$$\cos\varphi = \langle x, y\rangle = \cos\theta\cos\psi + \sin\theta\sin\psi = \cos(\theta - \psi)$$

yielding $\varphi = \theta - \psi$ modulo π. (Recall that $0 \le \varphi \le \pi$. Since $\theta - \psi$ need not to be in the interval $[0, \pi]$, the angles φ and $\theta - \psi$ can only be equal modulo an integer multiple of π.) Hence the non–oriented angle φ between $x, y \in \mathbb{R}^2 \setminus \{0\}$, defined by the Euclidean inner product, coincides with the classical notion of non–oriented angle in $\mathbb{R}^2$. In the sequel we will write simply $\langle\cdot,\cdot\rangle$ for $\langle\cdot,\cdot\rangle_{Euclid}$ in order to simplify notations.

APPLICATION: Compute the area F of a triangle ABC where $A = (0, 0)$ is the origin, $B = x = (x_1, x_2)$, $C = y = (y_1, y_2)$, and the oriented angle ψ between x and y is assumed to satisfy $0 \le \psi \le \pi$. Recall that

$$F = \frac{1}{2}\text{length of } \overline{AB} \cdot \text{height} .$$

where the height is given by $\|y\| \sin\psi$ and the length of $\overline{AB}$ is $\|x\|$. Since $0 < \sin\psi \leq 1$ one then has

$$F = \frac{1}{2}\|x\|\|y\| \sin\psi = \frac{1}{2}\|x\|\|y\|\sqrt{1-\cos^2\psi}\,, \qquad \cos\psi = \frac{\langle x,y\rangle}{\|x\|\|y\|}\,.$$

Thus

$$F = \frac{1}{2}\sqrt{\|x\|^2\|y\|^2 - (\langle x,y\rangle)^2}\,.$$

By an elementary computation

$$\|x\|^2\|y\|^2 - (\langle x,y\rangle)^2 = (x_1y_2 - x_2y_1)^2$$

or

$$\sqrt{\|x\|^2\|y\|^2 - (\langle x,y\rangle)^2} = |x_1y_2 - x_2y_1| = \left|\det\begin{pmatrix} x_1 & y_1 \\ x_2 & y_2 \end{pmatrix}\right|.$$

The absolute value of the determinant can thus be interpreted as the area of the parallelogram spanned by x and y.

EXAMPLE 2. Assume that $V = \mathbb{R}^n$. Then the map

$$\mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}\,, \qquad (x,y) \mapsto \sum_{i=1}^n x_iy_i$$

is an inner product on $\mathbb{R}^n$, referred to as Euclidean inner product. The corresponding norm of a vector $x \in \mathbb{R}^n$ is given by

$$\|x\| = \Big(\sum_{i=1}^n x_i^2\Big)^{\frac{1}{2}}$$

and the non–oriented angle $0 \leq \varphi \leq \pi$ between two vectors $x, y \in \mathbb{R}^n \setminus \{0\}$ is defined by

$$\cos\varphi = \frac{1}{\|x\|\|y\|}\sum_{i=1}^n x_iy_i\,.$$

There exist many inner products on $\mathbb{R}^n$, but usually we denote them by the same symbol. Examples on $\mathbb{R}^2$: for $x, y \in \mathbb{R}^2$

$$\langle x,y\rangle := 2x_1y_1 + 5x_2y_2 \qquad \text{or} \qquad \langle x,y\rangle := 3x_1y_1 + x_1y_2 + x_2y_1 + 3x_2y_2\,.$$

We will see later how inner products such as the two examples above can be found. An important property of inner product is the following inequality.

Lemma 4.3.4 (Triangle inequality)**.** *For any $v, w \in V$*

$$\|v+w\| \leq \|v\| + \|w\|\,.$$

The inequality is an immediate application of the Cauchy-Schwarz inequality,

$$\|v+w\|^2 = \|v\|^2 + \|w\|^2 + 2\langle v,w\rangle \leq \|v\|^2 + \|w\|^2 + 2\|v\|\|w\| = (\|v\| + \|w\|)^2\,.$$

4.4 Isometries and orthogonal matrices

Assume that V is a $\mathbb{R}$–vector space equipped with an inner product $\langle \cdot, \cdot \rangle$.

Definition 4.4.1. *A linear map $f : V \to V$ is said to be isometric (or an isometry) with respect to $\langle \cdot, \cdot \rangle$ if*

$$\langle f(v), f(w) \rangle = \langle v, w \rangle \,, \qquad \forall v, w \in V \,.$$

In words: f preserves the length of vectors and the non–orieneted angle between nonzero vectors.

Definition 4.4.2. *A matrix $A \in \mathbb{R}^{n \times n}$ is said to be orthogonal if*

$$\langle Ax, Ay \rangle = \langle x, y \rangle \,, \qquad \forall x, y \in \mathbb{R}^n$$

where $\langle \cdot, \cdot \rangle$ denotes the Euclidean inner product in $\mathbb{R}^n$,

$$\langle x, y \rangle := \sum_{j=1}^{n} x_j y_j \,, \qquad \forall x, y \in \mathbb{R}^n \,.$$

Lemma 4.4.1. *For any $A \in \mathbb{R}^{n \times n}$, A is orthogonal if and only if $A^T A = \mathrm{Id}_n$.*

Note that if A is orthogonal one has for any vectors of the standard basis $[e] = [e^{(1)}, \ldots, e^{(n)}]$

$$\langle Ae^{(i)}, Ae^{(j)} \rangle = \langle e^{(i)}, e^{(j)} \rangle = \delta_{ij} \,.$$

Since for any $1 \leq j \leq n$, the j-th column of A is $a^{(j)} = Ae^{(j)}$, one has

$$\langle a^{(i)}, a^{(j)} \rangle = \langle Ae^{(i)}, Ae^{(j)} \rangle = \sum_{k=1}^{n} a_{ki} a_{kj} = (A^T A)_{ij}$$

and thus indeed $A^T A = \mathrm{Id}_n$. Note that Lemma 4.4.1 implies that any orthogonal $n \times n$ matrix A is invertible since

$$1 = \det(\mathrm{Id}_n) = \det(A^T A) = \det(A^T)\det(A) = (\det(A))^2 \,.$$

The following theorem states important features of orthogonal matrices.

Theorem 4.4.2. *For any $A \in \mathbb{R}^{n \times n}$:*
(i) If A is orthogonal, then A is regular and $A^{-1} = A^T$.
(ii) If A is orthogonal so is A^{-1} (and A^T).
(iii) Id_n is orthogonal.
(iv) If $A, B \in \mathbb{R}^{n \times n}$ are orthogonal, so is AB.
(v) If A is orthogonal then $\det(A) \in \{-1, 1\}$.

The statements of Theorem 4.4.2 can be verified in a straightforward way. We introduce the following notation

$$O(n) := \left\{ A \in \mathbb{R}^{n \times n} : A \text{ is orthogonal} \right\}, \quad SO(n) := \left\{ A \in O(n) : \det(A) = 1 \right\}.$$

(From Theorem 4.4.2 it follows that $O(n)$ and $SO(n)$ are groups.)

Definition 4.4.3. *Assume that V is a $\mathbb{R}$–vector space of dimension n, equipped with an inner product $\langle\cdot,\cdot\rangle$. A basis $[v] = [v^{(1)}, \dots, v^{(n)}]$ is said to be orthonormal if*

$$\langle v^{(i)}, v^{(j)}\rangle = \delta_{ij}\,, \qquad \forall 1 \le i, j \le n\,.$$

In words: the vectors $v^{(i)}$ have length 1 and are pairwise orthogonal.

Examples (i) Let $V = \mathbb{R}^n$ and let $\langle\cdot,\cdot\rangle$ be the Euclidean inner product. Then the standard basis $[e] = [e^{(1)}, \dots, e^{(n)}]$ is an orthonormal basis.
(ii) Let $V = \mathbb{R}^2$ and let $\langle\cdot,\cdot\rangle$ be the Euclidean inner product. Define

$$v^{(1)} = (\cos\varphi, \sin\varphi)\,, \quad v^{(2)} = (-\sin\varphi, \cos\varphi)\,.$$

Then $[v] = [v^{(1)}, v^{(2)}]$ is an orthonormal basis of $\mathbb{R}^2$.
(iii) Assume that $V = \mathbb{R}^n$ and $\langle\cdot,\cdot\rangle$ is the Euclidean inner product of $\mathbb{R}^n$. The columns $a^{(1)}, \dots, a^{(n)}$ of an arbitrary orthogonal matrix A form an orthonormal basis of $\mathbb{R}^n$. (To verify this statement use Lemma 4.4.1.)

Applications:(i) Let $v^{(1)}, \dots, v^{(n)}$ be an orthonormal basis of a $\mathbb{R}$–vector space V with respect to the inner product $\langle\cdot,\cdot\rangle$. If $f : V \to V$ is an isometry, then $f(v^{(1)}), \dots, f(v^{(n)})$ is also an orthonormal basis of V. (Straightforward to verify.)
(ii) Orthonormal basis are particularly convenient to compute the corresponding coordinates of a vector. Assume that $v^{(1)}, \dots, v^{(n)}$ is an orthonormal basis of a $\mathbb{R}$–vector space V and $w \in V$. Then

$$w = \sum_{j=1}^{n} \langle w, v^{(j)}\rangle v^{(j)}\,.$$

So in this case, there is no need to solve a system of linear equations to determine the coordinates $x_1, \dots, x_n$ of w. They are given by $x_1 = \langle w, v^{(1)}\rangle, \dots, x_n = \langle w, v^{(n)}\rangle$.
There is a close connection between orthogonal matrices $A \in O(n)$ and isometries of $\mathbb{R}$-Hilbert spaces of dimension n.

Theorem 4.4.3. *Assume that $V \equiv (V, \langle\cdot,\cdot\rangle)$ is a $\mathbb{R}$-Hilbert space of dimension n, $[v]$ an orthonormal basis of V and $f : V \to V$ an isometry. Then $f_{[v]\to[v]}$ is an orthogonal matrix.*

Finally we introduce the notion of orthogonal complement.

Definition 4.4.4. *Assume that $M \subseteq V$ is a subset of the $\mathbb{R}$-Hilbert space $V \equiv (V, \langle\cdot,\cdot\rangle)$. Then*

$$M^{\perp} := \big\{ w \in V : \langle w, v\rangle = 0\,, \quad \forall v \in M \big\}$$

is referred to as the orthogonal complement of M.

Theorem 4.4.4. *Assume that $(V, \langle\cdot,\cdot\rangle)$ is a $\mathbb{R}$-Hilbert space of dimension n and $M \subseteq V$. Then the following holds:*
(i) *$M^{\perp}$ is a subspace of V. In particular, $\{0\}^{\perp} = V$ and $V^{\perp} = \{0\}$.*
(ii) *Assume that $M \subseteq V$ is a subspace of V and $v^{(1)}, \dots, v^{(m)}$ is an orthonormal basis of M. Then there are vectors $v^{(m+1)}, \dots, v^{(n)} \in V$ so that $[v^{(1)}, \dots, v^{(n)}]$ is an orthonormal basis of V and $[v^{(m+1)}, \dots, v^{(n)}]$ is an orthonormal basis of $M^{\perp}$.*
(iii) *If $M \subseteq V$ is a subspace of V then* $\dim(V) = \dim(M) + \dim(M^{\perp})$.

4.5 Vector product in $\mathbb{R}^3$

As already discussed in Chapter 1, one can define a multiplication of vectors in $\mathbb{R}^3$ which turns out to be useful.

Definition 4.5.1. *The vector product in $\mathbb{R}^3$ is the map $\mathbb{R}^3 \times \mathbb{R}^3 \to \mathbb{R}^3$, $(v, w) \mapsto v \times w$ where*

$$v \times w := (v_2 w_3 - v_3 w_2, -v_1 w_3 + v_3 w_1, v_1 w_2 - v_2 w_1)\,.$$

EXAMPLE: For $v = (0, 1, 0)$, $w = (1, 0, 0)$, one has $v \times w = (0, 0, -1)$.

Theorem 4.5.1. *The vector product in $\mathbb{R}^3$, with $\mathbb{R}^3$ being equipped with the Euclidean inner product $\langle \cdot, \cdot \rangle$, has the following properties:*
(i) antisymmetric: $v \times w = -w \times v$, $\forall v, w \in \mathbb{R}^3$
(ii) bilinear: for any $u, v, w \in \mathbb{R}^3$, $\alpha, \beta \in \mathbb{R}$

$$(\alpha v + \beta w) \times u = \alpha(v \times u) + \beta(w \times u)\,, \quad u \times (\alpha v + \beta w) = \alpha(u \times v) + \beta(u \times w)$$

(iii) $\langle v \times w, v \rangle = 0$, $\langle v \times w, w \rangle = 0$, $\forall v, w \in \mathbb{R}^3$
(iv) If $v \times w = 0$, then v and w are linearly dependent. In more detail either $v = 0$ or $w = 0$, or $v = \alpha w$ for some $\alpha \in \mathbb{R}$.
(v) $\|v \times w\| = \|v\|\|w\| \sin\varphi$ where $0 \leq \varphi \leq \pi$ is given by $\langle v, w \rangle = \|v\|\|w\| \cos\varphi$.
(vi) The vectors $v, w, v \times w$ are positively oriented, meaning that $\det(A) \geq 0$ where A is the 3×3 matrix whose columns are given by $v, w, v \times w$ (right thumb rule).

APPLICATION Assume that $v^{(1)}, v^{(2)}$ are vectors in $\mathbb{R}^3$ with $\|v^{(1)}\| = \|v^{(2)}\| = 1$ and $\langle v^{(1)}, v^{(2)} \rangle = 0$. Then $[v^{(1)}, v^{(2)}, v^{(1)} \times v^{(2)}]$ is an orthonormal basis of $\mathbb{R}^3$. Indeed, with the assumptions made, $\langle v^{(1)}, v^{(2)} \rangle = \cos\varphi$ with $\varphi = \pi/2$, implying that

$$\|v^{(1)} \times v^{(2)}\| = \|v^{(1)}\|\|v^{(2)}\| \sin\varphi = 1\,.$$

By Theorem 4.5.1-(*iii*), $[v^{(1)}, v^{(2)}, v^{(1)} \times v^{(2)}]$ is an orthonormal basis.
EXAMPLE: Let $v^{(1)} = \frac{1}{\sqrt{6}}(2, 1, -1)$. Then $\|v^{(1)}\| = 1$. We choose a vector $v^{(2)}$ of norm 1 which is orthogonal to $v^{(1)}$, $v^{(2)} = \frac{1}{\sqrt{5}}(1, -2, 0)$. Then $[v^{(1)}, v^{(2)}, v^{(3)}]$, $v^{(3)} = v^{(1)} \times v^{(2)}$ is an orthonormal basis of $\mathbb{R}^3$. One computes $v^{(3)} = \frac{1}{\sqrt{30}}(-2, -1, -5)$.

4.6 Inner products on $\mathbb{C}$–vector spaces

Assume that V is a $\mathbb{C}$–vector space.

Definition 4.6.1. *The map*

$$\langle \cdot, \cdot \rangle : V \times V \to \mathbb{C}$$

is said to be an inner product if the following is satisfied
$(IP1)_{\mathbb{C}}$ (Hermitian) $\langle v, w \rangle = \overline{\langle w, v \rangle}$ $\forall v, w \in V$.
$(IP2)_{\mathbb{C}}$ (linear in first argument) $\forall u, w, u \in V$, $\forall \alpha, \beta \in \mathbb{C}$,

$$\langle \alpha v + \beta w, u \rangle = \alpha \langle v, u \rangle + \beta \langle w, u \rangle$$

$(IP3)_{\mathbb{C}}$ (positive definite) $\forall v \in V$, $\langle v, v \rangle \geq 0$ and $\langle v, v \rangle = 0$ if and only if $v = 0$.

Again one can define the norm of a vector $v \in V$, induced by $\langle \cdot, \cdot \rangle$, $\|v\| := \sqrt{\langle v, v \rangle}$. It has the following properties

$$\|v\| \geq 0\,, \quad \|\alpha v\| = |\alpha| \|v\| \ \forall \alpha \in \mathbb{R}\,, \quad \|v + w\| \leq \|v\| + \|w\|\,.$$

The Cauchy-Schwarz inequality continues to hold, $|\langle v, w \rangle| \leq \|v\| \|w\|$. However note that in this case, $\langle v, w \rangle \in \mathbb{C}$ and hence we cannot define an angle.

EXAMPLE Euclidean inner product in $\mathbb{C}^n$:

$$\langle v, w \rangle := \sum_{j=1}^{n} v_j \overline{w}_j\,, \qquad \forall v, w \in \mathbb{C}^n\,.$$

It is straightforward to verify that $(IP1)_{\mathbb{C}} - (IP3)_{\mathbb{C}}$ are satisfied.

In case V is a $\mathbb{C}$–vector space of dimension n, equipped with an inner product $\langle \cdot, \cdot \rangle$, $(V, \langle \cdot, \cdot \rangle)$ is referred to as a n-dimensional $\mathbb{C}$–Hilbert space. A basis $[v] = [v^{(1)}, \ldots, v^{(n)}]$ of the $\mathbb{C}$-vector space V of dimension n with inner product $\langle \cdot, \cdot \rangle$ is said to be an orthonormal basis if $\langle v^{(j)}, v^{(j)} \rangle = 1$, for any $1 \leq j \leq n$ and $\langle v^{(i)}, v^{(j)} \rangle = 0$ for $j \neq i$.

A matrix $A \in \mathbb{C}^{n \times n}$ is said to be unitary if $\langle Av, Aw \rangle = \langle v, w \rangle$ for any $v, w \in \mathbb{C}^n$. Here $\langle \cdot, \cdot \rangle$ denotes the Euclidean inner product on $\mathbb{C}^n$. Similarly, a linear map $f : V \to V$ on a n-dimensional $\mathbb{C}$-Hilbert space $V \equiv (V, \langle \cdot, \cdot \rangle)$ is said to be an isometry if $\langle f(v), f(w) \rangle = \langle v, w \rangle$ for all $v, w \in V$. Note that $A \in \mathbb{C}^{n \times n}$ is unitary if and only if $f_A : \mathbb{C}^n \to \mathbb{C}^n$, $v \mapsto Av$ is an isometry. Furthermore A is unitary if and only if $\bar{A}^T A = \mathrm{Id}_n$. Here $\bar{A}^T$ is the conjugate transpose (or Hermitian transpose) of A,

$$(\bar{A}^T)_{ij} = \bar{a}_{ji} \qquad \text{where} \qquad A = (a_{ij})_{1 \leq i,j \leq n}\,.$$

Theorem 4.6.1. *Assume that $[v] = [v^{(1)}, \ldots, v^{(n)}]$ is an orthonormal basis of $(V, \langle \cdot, \cdot \rangle)$ and $f : V \to V$ a linear map. Then f is an isometry if and only if $f_{[v] \to [v]}$ is unitary.*

Chapter 5

Eigenvalues and eigenvectors

In this chapter we introduce the important notion of eigenvalues and eigenvectors of a linear map $f : V \to V$ on a vector space V of finite dimension. Since the case where V is a $\mathbb{C}$–vector space is somewhat simpler, we first treat this case.

5.1 Eigenvalues and eigenvectors of $\mathbb{C}$–linear maps on $\mathbb{C}$–vector spaces

Assume that V is a $\mathbb{C}$–vector space of dimension n and f $\mathbb{C}$-linear map.

Definition 5.1.1. *A complex number $\lambda \in \mathbb{C}$ is said to be an eigenvalue of f if there exists $v \in V \setminus \{0\}$ such that*

$$f(v) = \lambda v\,.$$

The vector v is called an eigenvector of f for the eigenvalue λ. Geometrically, it means that in the direction v, f is a dilation by the complex number λ.

Example Let $V = \mathbb{C}^n$ and let $f : \mathbb{C}^n \to \mathbb{C}^n$ be the linear map $f(v) = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)v$ where $\lambda_1, \ldots, \lambda_n$ are given complex numbers. Then $f(e^{(j)}) = \lambda_j e^{(j)}$ for any $1 \leq j \leq n$, where $e^{(1)}, \ldots, e^{(n)}$ is the standard basis of $\mathbb{C}^n$. Since for any $1 \leq j \leq n$, $e^{(j)} \neq 0$, $e^{(j)}$ is an eigenvector of f for the eigenvalue λ_j.

How to compute the eigenvalues of the linear map $f : V \to V$? Choose a basis $[v] = [v^{(1)}, \ldots, v^{(n)}]$ of V and consider the matrix representation $f_{[v]\to[v]} \in \mathbb{C}^{n\times n}$ of f with respect to $[v]$. Assume that $v \in V \setminus \{0\}$ is an eigenvector of f for the eigenvalue λ. Then v can be uniquely represented as a linear combination of $v^{(1)}, \ldots, v^{(n)}$,

$$v = \sum_{j=1}^{n} x_j v^{(j)}\,.$$

Then $x = (x_1, \ldots, x_n) \in \mathbb{C}^n \setminus \{0\}$ and

$$f_{[v]\to[v]}x = \lambda x \qquad \text{or} \qquad \big(f_{[v]\to[v]} - \lambda \mathrm{Id}_n\big)x = 0\,.$$

It means that $f_{[v]\to[v]} - \lambda \mathrm{Id}_n$ is not regular and hence

$$\det\big(f_{[v]\to[v]} - \lambda \mathrm{Id}_n\big) = 0\,.$$

We have therefore obtained the following

Theorem 5.1.1. *Assume that V is a $\mathbb{C}$–vector space of dimension n with basis $[v] = [v^{(1)}, \dots, v^{(n)}]$ and $f : V \to V$ is $\mathbb{C}$ linear. Then $\lambda \in \mathbb{C}$ is an eigenvalue of f if and only if $\det\big(f_{[v]\to[v]} - \lambda \mathrm{Id}_n\big) = 0$.*

Let us now investigate the function

$$\mathbb{C} \to \mathbb{C}\,, \qquad z \mapsto \det\big(f_{[v]\to[v]} - z\mathrm{Id}_n\big)\,.$$

Consider first the case where

$$f_{[v]\to[v]} = (a_{ij})_{1\le i,j\le n}$$

is a diagonal matrix, i.e., $a_{ij} = 0$ for $i \neq j$. Then

$$f_{[v]\to[v]} - z\mathrm{Id}_n = \mathrm{diag}\big(a_{11} - z, \cdots, a_{nn} - z\big)$$

and hence $\det\big(f_{[v]\to[v]} - z\mathrm{Id}_n\big)$ is given by

$$(a_{11} - z)\cdots(a_{nn} - z) = (-1)^n z^n + (-1)^{n-1}(a_{11} + \cdots + a_{nn})z^{n-1} + \ldots + \det(f_{[v]\to[v]})\,,$$

which is a polynomial of degree n in z. Actually this holds in general. As an illustration, consider the case when $n = 2$,

$$\det\big(f_{[v]\to[v]} - z\mathrm{Id}_2\big) = \det\begin{pmatrix} a_{11} - z & a_{12} \\ a_{21} & a_{22} - z \end{pmatrix} = z^2 - (a_{11} + a_{22})z + \det(A)\,.$$

More generally, one has

Theorem 5.1.2. *Assume that V is a $\mathbb{C}$–vector space of dimension n with basis $[v] = [v^{(1)}, \dots, v^{(n)}]$ and $f : V \to V$ is $\mathbb{C}$-linear. Then the following holds:*
(i) $\det\big(f_{[v]\to[v]} - z\mathrm{Id}_n\big)$ is a polynomial of degree n in z,

$$\chi(z) = p_n z^n + \cdots + p_1 z + p_0 \quad \textit{with} \quad p_n := (-1)^n\,,\ p_0 := \det(f_{[v]\to[v]})\,.$$

(ii) Every root (zero) of χ is an eigenvalue of f.

Remark 5.1.3. The polynomial $\chi(z)$ is independent of the choice of the basis $[v]$. This can be verified as follows: Assume that $[w] = [w^{(1)}, \dots, w^{(n)}]$ is another basis of V. Since

$$f_{[w]\to[w]} = \mathrm{Id}_{[v]\to[w]} f_{[v]\to[v]} \mathrm{Id}_{[w]\to[v]}\,, \qquad \mathrm{Id}_{[v]\to[w]} = \big(\mathrm{Id}_{[w]\to[v]}\big)^{-1}$$

it follows that

$$\begin{aligned}
\det\big(f_{[w]\to[w]} - z\mathrm{Id}_n\big) &= \det\big((\mathrm{Id}_{[w]\to[v]})^{-1} f_{[v]\to[w]} \mathrm{Id}_{[w]\to[v]} - z(\mathrm{Id}_{[w]\to[v]})^{-1}\mathrm{Id}_{[w]\to[v]}\big) \\
&= \det\big((\mathrm{Id}_{[w]\to[v]})^{-1}\big(f_{[v]\to[w]} - z\mathrm{Id}_n\big)\mathrm{Id}_{[w]\to[v]}\big) \\
&= \det\big((\mathrm{Id}_{[w]\to[v]})^{-1}\big)\det\big(f_{[v]\to[w]} - z\mathrm{Id}_n\big)\det\big(\mathrm{Id}_{[w]\to[v]}\big) \\
&= \det\big(f_{[v]\to[w]} - z\mathrm{Id}_n\big)
\end{aligned}$$

where we have used that the determinant is multiplicative and hence it particular

$$\det\big((\mathrm{Id}_{[w]\to[v]})^{-1}\big) = \big(\det\big(\mathrm{Id}_{[w]\to[v]}\big)\big)^{-1}.$$

□

Since $\chi(z)$ depends only on f we denote it also by $\chi_f(z)$. It is referred to as the characteristic polynomial of f. How many eigenvalues does f have? Or equivalently, how may zeros does χ_f have? Since χ_f is a polynomial of degree n it has n complex zeros when counted with their multiplicities. The multiplicity of a zero of χ_f is also referred to as the *algebraic multiplicity* of the corresponding eigenvalue of f.

EXAMPLES:
(i) Assume that V is a $\mathbb{C}$–vector space of dimension n and $f : V \to V$ is a linear map with characteristic polynomial $\chi_f(z) = (1-z)^2(3-z)^5(4-z)^3$. The zeros of χ_f are $z_1 = 1$, $z_2 = 3$ and $z_3 = 4$. The multiplicities of z_1, z_2, z_3 are

$$m_{z_1} = 2\,, \quad m_{z_2} = 5\,, \quad m_{z_3} = 3\,.$$

Note that $m_{z_1} + m_{z_2} + m_{z_3} = 10 = \dim(V)$.
(ii) Assume that V is a $\mathbb{C}$–vector space of dimension n and f is the identity map, $f = \mathrm{Id}$. Choose an arbitrary basis $[v]$ of V. Then $f_{[v]\to[v]} = \mathrm{Id}_n$ and hence

$$\det\big(f_{[v]\to[v]} - z\mathrm{Id}_n\big) = (1-z)^n\,.$$

Hence χ_f has the root $z_1 = 1$ and its multiplicity m_{z_1} equals n.

We summarize our discussion as follows:

Theorem 5.1.4. *Assume that V is a $\mathbb{C}$–vector space of dimension n and $f : V \to V$ is linear. Then f has precisely n eigenvalues, when counted with their multiplicities.*

TERMINOLOGY: The collection of the eigenvalues of f, listed with their algebraic multiplicities, is called the spectrum of f and denoted by $\mathrm{spec}(f)$. For any eigenvalue λ of f, we denote by m_λ its algebraic multiplicity.
EXAMPLE: (i) For $\chi_f(z) = (1-z)^2(3-z)^5(4-z)^3$, the spectrum of f is given by

$$\lambda_1 = \lambda_2 = 1\,, \quad \lambda_3 = \lambda_4 = \lambda_5 = \lambda_6 = \lambda_7 = 3\,, \quad \lambda_8 = \lambda_9 = \lambda_{10} = 4$$

Definition 5.1.2. *Given any eigenvalue λ of a linear map $f : V \to V$, the set*

$$E_\lambda = E_\lambda(f) := \big\{v \in V : f(v) = \lambda v\big\}$$

is said to be the eigenspace of f for the eigenvalue λ. Note that

$$E_\lambda = \{0\} \cup \big\{v \in V : v \text{ eigenvector of } f \text{ for the eigenvalue } \lambda\big\}\,.$$

Lemma 5.1.5. *For any eigenvalue λ of f, $E_\lambda(f)$ is a $\mathbb{C}$–subspace of V.*

Indeed if $v, w \in E_\lambda(f)$, then $f(v) = \lambda v$, $f(w) = \lambda w$ and hence

$$f(v+w) = f(v) + f(w) = \lambda v + \lambda w = \lambda(v+w)\,.$$

Furthermore, for any $\alpha \in \mathbb{C}$, $f(\alpha v) = \alpha f(v) = \alpha\lambda v = \lambda(\alpha v)$.

Definition 5.1.3. *For an eigenvalue λ of a linear map $f : V \to V$ on a $\mathbb{C}$–vector space V of dimension n, $\dim(E_\lambda(f))$ is said to be the geometric multiplicity of λ.*

Theorem 5.1.6. *Assume that V is a $\mathbb{C}$–vector space of dimension n and $\lambda \in \mathbb{C}$ is an eigenvalue of a linear map $f : V \to V$. Then the following holds:*
(i) $\dim(E_\lambda(f)) \leq m_\lambda$ *and* $m_\lambda \leq n$.
(ii) If $\dim(E_\lambda(f)) = m_\lambda$ *for any eigenvalue λ of f, then V admits a basis consisting of eigenvectors of f.*

Examples: We consider examples of linear maps $f : \mathbb{C}^2 \to \mathbb{C}^2$ of the form f_A with $A \in \mathbb{C}^{2\times 2}$. Recall that $(f_A)_{[e]\to[e]} = A$. Hence

$$\chi_{f_A}(z) = \det(A - \lambda \mathrm{Id}_2)\,.$$

We refer to $\chi_{f_A}(z)$ also as the characteristic polynomial of A and denote it by $\chi_A(z)$. We want to determine the eigenvalues of f_A and the corresponding eigenspaces.
(i)

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

Step 1: Determine the eigenvalues of f_A

$$\det \begin{pmatrix} 1-z & 2 \\ 3 & 4-z \end{pmatrix} = (1-z)(4-z) - 6 = z^2 - 5z - 2$$

The roots are $z_\pm = \frac{5}{2} \pm \frac{1}{2}\sqrt{33}$ or $\lambda_1 = \frac{5}{2} + \frac{1}{2}\sqrt{33}$, $\lambda_2 = \frac{5}{2} - \frac{1}{2}\sqrt{33}$.
Step 2: Determine the eigenspaces for λ_1 and λ_2.
$E_{\lambda_1}(f) \equiv E_{\lambda_1}(A)$: we need to solve the linear system

$$\begin{pmatrix} 1-\lambda_1 & 2 \\ 3 & 4-\lambda_1 \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

Consider the equation

$$(1-\lambda_1)v_1 + 2v_2 = 0\,.$$

Since $1 - \lambda_1 \neq 0$, we may choose $v_2 = 1$ and get $v_1 = \frac{2}{\lambda_1 - 1}$. One then verifies that

$$v^{(1)} = \big(\frac{-2}{1-\lambda_1}, 1\big)$$

is an eigenvector for λ_1. Since $m_{\lambda_1} = 1$, one has $\dim(E_{\lambda_1}(A)) = 1$.

$E_{\lambda_2}(f) \equiv E_{\lambda_2}(A)$: we need to solve the linear system

$$\begin{pmatrix} 1-\lambda_2 & 2 \\ 3 & 4-\lambda_2 \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

Consider the equation $(1-\lambda_2)+2v_2 = 0$. Since $1-\lambda_2 \neq 0$ we again choose $v_2 = 1$. One verifies that

$$v^{(2)} = \Big(\frac{2}{\lambda_2 - 1}, 1\Big)$$

is an eigenvector corresponding to the eigenvalue λ_2. Since $m_{\lambda_2} = 1$ one concludes that $\dim(E_{\lambda_2}(A)) = 1$. The vectors $v^{(1)}, v^{(2)}$ form a basis $[v]$ of $\mathbb{C}^2$. (Since they are eigenvectors of f_A corresponding to dinstinct eigenvalues, they are linearly independent.) One has

$$(f_A)_{[v]\to[v]} = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix} .$$

Since $A = (f_A)_{[e]\to[e]}$ it follows that A is diagonalizable

$$A = \mathrm{Id}_{[v]\to[e]} (f_A)_{[v]\to[v]} \mathrm{Id}_{[e]\to[v]} ,$$

where $\mathrm{Id}_{[v]\to[e]} = (v^{(1)}\, v^{(2)})$ and $\mathrm{Id}_{[e]\to[v]} = \big(\mathrm{Id}_{[v]\to[e]}\big)^{-1}$.
(ii)

$$A = \det \begin{pmatrix} 1 & 1 \\ 0 & 2 \end{pmatrix}$$

Step 1: Determine the eigenvalues of f_A.

$$\det \begin{pmatrix} 1-z & 1 \\ 0 & 2-z \end{pmatrix} = (1-z)(2-z) = 0 ,$$

thus $\lambda_1 = 1$, $\lambda_2 = 2$.
Step 2: Determine the eigenspaces for λ_1, λ_2.
$E_{\lambda_1}(f_A) \equiv E_{\lambda_1}(A)$. We need to solve the following linear system

$$\begin{pmatrix} 1-1 & 1 \\ 0 & 2-1 \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} .$$

Consider the equation $0 \cdot v_1 + v_2 = 0$. It has the solution $v_2 = 0$, $v_1 = 1$ and one verifies that $v^{(1)} = (1, 0)$ is an eigenvector corresponding to the eigenvalue λ_1. Since $m_{\lambda_1} = 1$ one has $\dim(E_{\lambda_1}(A)) = 1$.
$E_{\lambda_2}(f) \equiv E_{\lambda_2}(A)$: We need to solve

$$\begin{pmatrix} 1-2 & 1 \\ 0 & 2-2 \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} .$$

Consider the equation $-v_1 + v_2 = 0$. It has the solution $v_1 = 1$, $v_2 = 1$ and one verifies that $v^{(1)} = (1, 1)$ is an eigenvector corresponding to the eigenvalue λ_2. Since $m_{\lambda_2} = 1$ one has $\dim(E_{\lambda_2}(A)) = 1$. Hence $[v] = [v^{(1)}, v^{(2)}]$ is a basis of $\mathbb{C}^2$ and

$$(f_A)_{[v]\to[v]} = \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix} .$$

(*iii*)

$$A = \begin{pmatrix} 1 & \mathrm{i} \\ 0 & 2 \end{pmatrix}$$

Step 1: determine the eigenvalues of A.

$$\chi_A(z) = \det\Big(A - z\mathrm{Id}_2\Big) = \det\begin{pmatrix} 1-z & \mathrm{i} \\ 0 & 2-z \end{pmatrix} = (1-z)(2-z)$$

implying that $\lambda_1 = 1$, $\lambda_2 = 2$ with $m_{\lambda_1} = 1$, $m_{\lambda_2} = 2$.
Step 2. Determine the eigenspaces. We need to solve the linear system

$$E_{\lambda_1}(A): \qquad \begin{pmatrix} 1-1 & \mathrm{i} \\ 0 & 2-1 \end{pmatrix}\begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

The corresponding augmented coefficient matrix is

$$\left(\begin{array}{cc|c} 0 & \mathrm{i} & 0 \\ 0 & 1 & 0 \end{array}\right) \rightsquigarrow \left(\begin{array}{cc|c} 0 & \mathrm{i} & 0 \\ 0 & 0 & 0 \end{array}\right)$$

hence

$$E_{\lambda_1}(A) = \big\{\alpha(1,0) : \alpha \in \mathbb{C}\big\}.$$

$$E_{\lambda_2}(A): \qquad \begin{pmatrix} 1-2 & \mathrm{i} \\ 0 & 0 \end{pmatrix}\begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

thus

$$E_{\lambda_2}(A) = \big\{\alpha(\mathrm{i},1) : \alpha \in \mathbb{C}\big\}.$$

Note that the eigenvectors $v^{(1)} = (1,0)$ and $v^{(2)} = (\mathrm{i},1)$ are linearly independent in $\mathbb{C}^2$ and thus indeed form a basis of $\mathbb{C}^2$. Clearly

$$f_{[v]\to[v]} = \mathrm{diag}(1,2),$$

$$\mathrm{Id}_{[v]\to[e]} = \begin{pmatrix} 1 & \mathrm{i} \\ 0 & 1 \end{pmatrix}$$

$$\mathrm{Id}_{[e]\to[v]} = \Big(\mathrm{Id}_{[v]\to[e]}\Big)^{-1} = \begin{pmatrix} 1 & -\mathrm{i} \\ 0 & 1 \end{pmatrix}.$$

Thus

$$f_{[e]\to[e]} = \begin{pmatrix} 1 & \mathrm{i} \\ 0 & 1 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}\begin{pmatrix} 1 & -\mathrm{i} \\ 0 & 1 \end{pmatrix}$$

it means that A is similar to $\mathrm{diag}(1,2)$ and hence diagonalizable.

(*iv*)

$$A = \begin{pmatrix} 1 & \mathrm{i} \\ 0 & 1 \end{pmatrix} \in \mathbb{C}^{2\times 2}$$

STEP 1. Determine the eigenvalues of A

$$\chi_A(z) = \det(A - z\mathrm{Id}_2) = \det \begin{pmatrix} 1-z & \mathrm{i} \\ 0 & 1-z \end{pmatrix} = (1-z)^2 .$$

Hence $\lambda_1 = \lambda_2 = 1$ and $m_{\lambda_1} = 2$.
STEP 2. Determine the eigenspace $E_{\lambda_1}(A)$. Solve

$$\begin{pmatrix} 0 & \mathrm{i} \\ 0 & 0 \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} .$$

The solutions are $(v_1, 0)$ with $v_1 \in \mathbb{C}$ arbitrary, hence

$$E_{\lambda_1}(A) = \big\{\alpha(1,0) : \alpha \in \mathbb{C}\big\}$$

is of dimension 1, $n_{\lambda_1} = 1 < 2 = m_{\lambda_1}$. As a consequence there does not exist a basis of $\mathbb{C}^2$ of eigenvectors of A (and A cannot be diagonalized).

Let us now discuss special type of matrices

Upper triangular matrices. A matrix A is upper triangular if it is of the form

$$A = (a_{ij})_{1 \le i,j \le n} , \qquad a_{ij} = 0 \quad \text{for} \quad i < j .$$

Then the coefficients $a_{11}, \dots, a_{nn}$ of the diagonal of A are the eigenvalues of A, counted with their algebraic multiplicity. Indeed

$$\det(A - z\mathrm{Id}_n) = (a_{11} - z) \dots (a_{nn} - z) .$$

The eigenvalues of A can then be directly read from the diagonal of A. Similarly, the eigenvalues of a lower triangular matrix $A \in \mathbb{C}^{n\times n}$ can be read off from the diagonal.

Invertible matrices. Assume that $A \in \mathbb{C}^{n\times n}$ is invertible. Then the eigenvalues $\lambda_1, \dots, \lambda_n$ of A (listed with their algebraic multiplicities) are in $\mathbb{C} \setminus \{0\}$ and $\lambda_1^{-1}, \dots, \lambda_n^{-1}$ are the eigenvalues of A^{-1}. If $v^{(j)}$ is an eigenvector of A for λ_j, then $A^{-1}v^{(j)} = \lambda_j^{-1}v^{(j)}$, i.e. $v^{(j)}$ is also an eigenvector of A^{-1} for λ_j^{-1}.

TRANSPOSE MATRIX: given a matrix $A \in \mathbb{C}^{n\times n}$, the eigenvalues of its transpose A^T, coincide with the eigenvalues of A. Indeed

$$\det(A^T - z\mathrm{Id}_n) = \det([A - z\mathrm{Id}_n]^T) = \det(A - z\mathrm{Id}_n) ,$$

i.e. A and A^T have the same characterisitc polynomial.

Definition 5.1.4. *A $\mathbb{C}$-linear map $f : V \to V$ on a $\mathbb{C}$ vector space of dimension n is said to be diagonalizable if there exists a basis $[v] = [v^{(1)}, \dots, v^{(n)}]$ of eigenvectors of f. In such a case one has that*

$$f_{[v]\to[v]} = \mathrm{diag}(\lambda_1, \dots, \lambda_n)$$

where for any $1 \le i \le n$, $\lambda_i \in \mathbb{C}$ is the eigenvalue of f corresponding to the eigenvector $v^{(j)}$.

Theorem 5.1.7. *Let $A \in \mathbb{C}^{n\times n}$. Then $f_A : \mathbb{C}^n \to \mathbb{C}^n$, $x \mapsto Ax$ has a basis of eigenvectors if and only if there is a regular matrix $S \in \mathbb{C}^{n\times n}$ such that $S^{-1}AS$ is a diagonal matrix. The elements of the diagonal of $S^{-1}AS$ are the eigenvalues of f_A.*

Remark 5.1.8. If $[v] = [v^{(1)}, \dots, v^{(n)}]$ is a basis of $\mathbb{C}^n$ consisting of eigenvectors of $f_A : \mathbb{C}^n \to \mathbb{C}^n$, then $S = \mathrm{Id}_{[v]\to[e]}$ is regular and $S^{-1}AS$ is a diagonal matrix. Conversely, let $S \in \mathbb{C}^{n\times n}$ be regular and $S^{-1}AS$ be a diagonal matrix $B = \mathrm{diag}(\lambda_1, \dots, \lambda_n)$. Denote by $s^{(1)}, \dots, s^{(n)}$ the columns of S. Then $s^{(1)}, \dots, s^{(n)}$ is a basis of $\mathbb{C}^n$. Since $AS = SB$ it follows that for any $1 \le j \le n$

$$\lambda_j s^{(j)} = S(\lambda_j e^{(j)}) = SB(e^{(j)}) = AS(e^{(j)}) = As^{(j)}\,.$$

Hence $s^{(j)}$ is an eigenvector of f_A corresponding to the eigenvalue λ_j. □

Theorem 5.1.9. *Assume that $f : V \to V$ is a $\mathbb{C}$-linear map on the $\mathbb{C}$-vector space V of dimension n. If f has n dinstinct eigenvalues $\lambda_1, \dots, \lambda_n \in \mathbb{C}$, then f is diagonalizable.*

Remark 5.1.10. If the eigenvalues of f $\lambda_1, \dots, \lambda_n \in \mathbb{C}$ are dinstinct, then for any $1 \le j \le n$

$$1 \le n_{\lambda_j} = \dim(E_{\lambda_j}(f)) \le m_{\lambda_j}\,, \quad \sum_{j=1}^{n} m_{\lambda_j} = n\,.$$

It follows that $\dim(E_{\lambda_j}(f)) = 1$ for any $1 \le j \le n$. One can show that for any choice of $v^{(j)} \in E_{\lambda_j}(f) \setminus \{0\}$, $v^{(1)}, \dots, v^{(n)}$ form a basis of V. □

Examples: (i) let

$$A = \begin{pmatrix} 2 & 1 & 0 \\ 0 & 1 & -1 \\ 0 & 2 & 4 \end{pmatrix} \in \mathbb{C}^{3\times 3}\,.$$

Compute the eigenvalues and the eigenvectors of f_A.

Step 1. Determine the eigenvalues of f_A.

$$\det(A - z\mathrm{Id}_3) = (2-z)\Big((1-z)(4-z)+2\Big)\,.$$

Then $\lambda_1 = 2$ and λ_2, λ_3 are zeros of

$$0 = (1-z)(4-z)+2 = z^2 - 5z + 6\,,$$

i.e. $\lambda_2 = 2$ and $\lambda_3 = 3$. note that

$$m_{\lambda_1} = 2\,, \qquad m_{\lambda_3} = 1\,.$$

Step 2: determine the eigenspaces.

$$E_{\lambda_3}(A): \qquad \begin{pmatrix} -1 & 1 & 0 \\ 0 & -2 & -1 \\ 0 & 2 & 1 \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

Gaussian elimination leadsto ($R_3 \rightsquigarrow R_2 + R_3$)

$$\left(\begin{array}{ccc|c} -1 & 1 & 0 & 0 \\ 0 & -2 & -1 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right) .$$

Thus the solutions are given by the vectors $\alpha(-1,-1,2)$, $\alpha \in \mathbb{C}$

$$E_{\lambda_3}(A) = \left\{\alpha(-1,-1,2) : \alpha \in \mathbb{C}\right\} .$$

$$E_{\lambda_1}(A): \qquad \begin{pmatrix} 0 & 1 & 0 \\ 0 & -1 & -1 \\ 0 & 2 & 2 \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

Gaussian elimination ($R_3 \rightsquigarrow R_3 + 2R_2$)

$$\left(\begin{array}{ccc|c} 0 & 1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right)$$

hence $v = \alpha(1,0,0)$, $\alpha \in \mathbb{C}$ are all solutions and

$$E_{\lambda_1}(A) = \left\{\alpha(1,0,0) : \alpha \in \mathbb{C}\right\} .$$

Since

$$n_{\lambda_1} = \dim(E_{\lambda_1}(A)) < m_{\lambda_1} = 2$$

the matrix A cannot be diagonalized.

(*ii*)

$$A = \begin{pmatrix} \cos\varphi & -\sin\varphi \\ \sin\varphi & \cos\varphi \end{pmatrix} , \qquad \varphi \neq 0 \mod \pi ,$$

i.e. $\sin\varphi \neq 0$.

Step 1. Determine the eigenvalues of A.

$$\det \begin{pmatrix} \cos\varphi - z & -\sin\varphi \\ \sin\varphi & \cos\varphi - z \end{pmatrix} = (\cos\varphi - z)^2 + \sin^2\varphi = z^2 - 2z\cos\varphi + 1 .$$

The roots are given by

$$\lambda_1 = \cos\varphi + \mathrm{i}\sin\varphi = e^{\mathrm{i}\varphi} , \qquad \lambda_2 = \cos\varphi - \mathrm{i}\sin\varphi = e^{-\mathrm{i}\varphi} .$$

Clearly $m_{\lambda_1} = m_{\lambda_2} = 1$.

Step 2. Determine the eigenspaces. We need to solve

$$\begin{pmatrix} \mathrm{i}\sin\varphi & -\sin\varphi \\ \sin\varphi & \mathrm{i}\sin\varphi \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} .$$

Since by assumption $\sin\varphi \neq 0$, we can divide by $\sin\varphi$ the augmented coefficient matrix

$$\begin{pmatrix} -\mathrm{i} & -1 & | & 0 \\ 1 & -\mathrm{i} & | & 0 \end{pmatrix}$$

and thus

$$E_{\lambda_1}(A) = \left\{\alpha(\mathrm{i}, 1) : \alpha \in \mathbb{C}\right\}.$$

$E_{\lambda_2}(A)$: we need to solve

$$\begin{pmatrix} -\mathrm{i}\sin\varphi & -\sin\varphi \\ \sin\varphi & -\mathrm{i}\sin\varphi \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

the augmented coefficient matrix, after dividing by $\sin\varphi \neq 0$ is given by

$$\begin{pmatrix} \mathrm{i} & -1 & | & 0 \\ 1 & \mathrm{i} & | & 0 \end{pmatrix}$$

and thus

$$E_{\lambda_2}(A) = \left\{\alpha(-\mathrm{i}, 1) : \alpha \in \mathbb{C}\right\}.$$

Then $v^{(1)} = (\mathrm{i}, 1)$, $v^{(2)} = (1, \mathrm{i})$ form a basis of eigenvectors of $\mathbb{C}^2$ with

$$(f_A)_{[v]\to[v]} = \mathrm{diag}(e^{\mathrm{i}\varphi}, e^{-\mathrm{i}\varphi})$$

and

$$\begin{aligned} A &= \mathrm{Id}_{[v]\to[e]}\mathrm{diag}(e^{\mathrm{i}\varphi}, e^{-\mathrm{i}\varphi})(\mathrm{Id}_{[v]\to[e]})^{-1} \\ &= \begin{pmatrix} \mathrm{i} & 1 \\ 1 & \mathrm{i} \end{pmatrix} \begin{pmatrix} e^{\mathrm{i}\varphi} & 0 \\ 0 & e^{-\mathrm{i}\varphi} \end{pmatrix} \begin{pmatrix} \mathrm{i} & 1 \\ 1 & \mathrm{i} \end{pmatrix}^{-1} \end{aligned}$$

Terminology: the spectrum of a $n \times n$ matrix $A \in \mathbb{C}^{n\times n}$ and the eigenvalues of A are defined to be the spectrum respectively the eigenvalues of f_A. As already mentioned above, A and f_A are often not distinguished if the context permits.

Definition 5.1.5. *An eigenvalue λ of $A \in \mathbb{C}^{n\times n}$ is called simple if $m_\lambda = 1$. The spectrum of A is called simple if any eigenvalue of A is simple.*

Finally let us discuss two important classes of matrices. Recall that a matrix $A \in \mathbb{C}^{n\times n}$ is called unitary if $A\overline{A}^T = \mathrm{Id}_n$. Equivalently these are matrices with the property

$$\langle Av, Aw\rangle = \langle v, w\rangle, \qquad \forall v, w \in \mathbb{C}^n$$

where

$$\langle v, w\rangle = \sum_{j=1}^{n} v_j\overline{w}_j$$

denotes the Euclidean inner product on $\mathbb{C}^n$.

Theorem 5.1.11. *For any unitary matrix $A \in \mathbb{C}^{n\times n}$, the following holds:*
(i) Every eigenvalue λ of A satisfies $|\lambda| = 1$.
(ii) If $\lambda, \mu \in \mathbb{C}$ are eigenvalues of A with $\lambda \neq \mu$, then for any eigenvectors v, w of A for λ respectively μ, v and w are orthogonal, i.e. $\langle v, w\rangle = 0$.
(iii) There exists an orthonormal basis of eigenvectors of A, i.e. A is diagonalizable.

Remark 5.1.12. let us briefly discuss items (i) and (ii). Concerning item (i), assume that $\lambda \in \mathbb{C}$ is an eigenvalue of A with eigenvector $v \in \mathbb{C}^n \setminus \{0\}$. Then

$$\langle v, v\rangle = \langle Av, Aw\rangle = \langle \lambda v, \lambda v\rangle = \lambda\overline{\lambda}\langle v, v\rangle\,.$$

Since $v \neq 0$ one has that $\langle v, v\rangle \neq 0$ and thus $|\lambda|^2 = 1$. Regarding (ii), assume that $v, w \in \mathbb{C}^n \setminus \{0\}$ are eigenvectors of A for the eigenvalues λ respectively μ, with $\lambda \neq \mu$. Then

$$\lambda\langle v, w\rangle = \langle \lambda v, w\rangle = \langle Av, w\rangle = \langle v, \overline{A}^T w\rangle\,.$$

Since $\overline{A}^T = A^{-1}$ and $A^{-1}w = \mu^{-1}w$ with $\mu^{-1} = \bar{\mu}$ it then follows that

$$\lambda\langle v, w\rangle = \mu\langle v, w\rangle$$

or $(\lambda - \mu)\langle v, w\rangle = 0$. By assumption $\lambda - \mu \neq 0$ and hence $\langle v, w\rangle = 0$. □

The second class of matrices we want to discuss are the Hermitian matrices.

Definition 5.1.6. *A matrix $A \in \mathbb{C}^{n\times n}$ is called Hermitian if $\overline{A}^T = A$ of equivalently,*

$$\langle Av, Aw\rangle = \langle v, w\rangle\,, \quad \forall v, w \in \mathbb{C}^n\,.$$

$\langle \cdot, \cdot\rangle$ denotes the Euclidean inner product on $\mathbb{C}^n$.

Theorem 5.1.13. *For any Hermitian matrix $A \in \mathbb{C}^{n\times n}$, the following holds:*
(i) Every eigenvalue of A is real.
(ii) Any two eigenvectors $v, w \in \mathbb{C}^n$ of A corresponding to eigenvalues λ, μ of A with $\lambda \neq \mu$ are orthogonal, i.e. $\langle v, w\rangle = 0$.
(iii) There exists an orthonormal basis $[v] = [v^{(1)}, \ldots, v^{(n)}]$ of $\mathbb{C}^n$ consisting of eigenvectors of A so that $S^{-1}AS$ is diagonal and S is the unitary matrix whise columns are $v^{(1)}, \ldots, v^{(n)}$.

Remark 5.1.14. Let us briefly discuss items (i) and (ii). Concerning (i), let λ be an eigenvalue of A and $v \in \mathbb{C}^n$, $\langle v, v\rangle \neq 0$ a corresponding eigenvector, $Av = \lambda v$. Since

$$\lambda\langle v, v\rangle = \langle \lambda v, v\rangle = \langle Av, v\rangle = \langle v, Av\rangle = \langle v, \lambda v\rangle = \overline{\lambda}\langle v, v\rangle$$

it then follows that $\lambda = \overline{\lambda}$.
Regarding item (ii), assume that λ and μ are eigenvalues of A with $\lambda \neq \mu$ and $v, w \in \mathbb{C}^n \setminus \{0\}$ the corresponding eigenvectors. Then

$$\lambda\langle v, w\rangle = \langle \lambda v, w\rangle = \langle Av, w\rangle = \langle v, Aw\rangle = \langle v, \mu w\rangle = \overline{\mu}\langle v, w\rangle\,.$$

Since μ is real and by assumption $\lambda \neq \mu$ it then follows that $\langle v, w\rangle = 0$. □

5.2 Eigenvalues and eigenvectors of $\mathbb{R}$-linear maps on $\mathbb{R}$-vector spaces

The aim of this section is to discuss issues in spectral theory, arising when one considers $\mathbb{R}$-linear maps on $\mathbb{R}$-vector spaces V. To simplify the exposition we limit ourselves to the case where $V = \mathbb{R}^n$ and consider linear maps $f_A : \mathbb{R}^n \to \mathbb{R}^n$, $x \to Ax$ associated to the matrix $A \in \mathbb{R}^{n\times n}$. Of course we can extend the linear map f_A to a map on $\mathbb{C}^n$, $f_A : \mathbb{C}^n \to \mathbb{C}^n$, $x \mapsto Ax$ and then apply the results obtained in subsection 4.1. But A having real coefficients has interesting implications and arises new questions which we now would like to discuss in some detail. For $A \in \mathbb{R}^{n\times n}$, consider the characteristic polynomial

$$\chi_A(z) = \det\Big(A - z\mathrm{Id}_n\Big)\,.$$

It is not difficult to see that the polynomial $\chi_A(z)$ has real coefficients. However, note that in general this does not imply that the roots of χ_A are real numbers.

Lemma 5.2.1. *Assume that $A \in \mathbb{R}^{n\times n}$ and $v \in \mathbb{C}^n \setminus \{0\}$ is an eigenvector of A for the eigenvalue $\lambda \in \mathbb{C}$. Then $\overline{\lambda}$ is an eigenvalue of A as well and $\overline{v}$ is a corresponding eigenvector.*

Indeed, since $\overline{A} = A$ and $\overline{Av} = \overline{A}\overline{v}$, one has

$$A\overline{v} = \overline{Av} = \overline{\lambda v} = \overline{\lambda}\overline{v}\,.$$

Remark 5.2.2. Actually one can verify that for any eigenvalue $\lambda \in \mathbb{C}$ of $A \in \mathbb{R}^{n\times n}$,

$$m_\lambda = m_{\overline{\lambda}} \qquad n_\lambda = n_{\overline{\lambda}}\,,$$

i.e., λ and $\overline{\lambda}$ have the same algebraic and geometric multiplicities. □

Lemma 5.2.1 can be applied as follows: assume that $\lambda \in \mathbb{C}$ is an eigenvalue of A with eigenvector $v = (v_j)_{1\le j\le n} \in \mathbb{C}^n$. Then

$$Av = \lambda v\,, \qquad A\overline{v} = \overline{\lambda}\overline{v}\,.$$

Note that $v + \overline{v} = (2\mathrm{Re}(v_j))_{1\le j\le n} \in \mathbb{R}^n$ and $\mathrm{i}(\overline{v} - v) = (2\mathrm{Im}(v_j))_{1\le j\le n} \in \mathbb{R}^n$. Hence with $\lambda_1 = \mathrm{Re}(\lambda)$, $\lambda_2 = \mathrm{Im}(\lambda)$, one has

$$A(v + \overline{v}) = (\lambda_1 + \mathrm{i}\lambda_2)v + (\lambda_1 - \mathrm{i}\lambda_2)\overline{v} = \lambda_1(v + \overline{v}) - \mathrm{i}\lambda_2(\overline{v} - v)$$

and similarly,

$$A(\mathrm{i}(\overline{v} - v)) = \mathrm{i}A\overline{v} - \mathrm{i}Av = \mathrm{i}(\lambda_1 - \mathrm{i}\lambda_2)\overline{v} - \mathrm{i}(\lambda_1 + \mathrm{i}\lambda_2)v = \lambda_1\mathrm{i}(\overline{v} - v) + \lambda_2(v + \overline{v})\,.$$

In the case where $v + \overline{v}$ and $\mathrm{i}(\overline{v} - v)$ are linearly independent vectors in $\mathbb{R}^n$, we consider the two dimensional subspace W generated by

$$v^{(1)} := v + \overline{v}\,, \qquad v^{(2)} := \mathrm{i}(\overline{v} - v)\,.$$

Then $Aw \in W$ for any $w \in W$ and

$$\Big((f_A)_{|W}\Big)_{[v^{(1)},v^{(2)}]\to[v^{(1)},v^{(2)}]} = \begin{pmatrix} \mathrm{Re}(\lambda) & \mathrm{Im}(\lambda) \\ -\mathrm{Im}(\lambda) & \mathrm{Re}(\lambda) \end{pmatrix} .$$

We have the following

Lemma 5.2.3. *Assume that $A \in \mathbb{R}^{n\times n}$ and $v \in \mathbb{C}^n$ is an eigenvector of A with eigenvalue $\lambda \in \mathbb{C} \setminus \mathbb{R}$. Then the following holds:*
(i) $v^{(1)} = v + \overline{v}$ and $v^{(2)} = \mathrm{i}(\overline{v} - v)$ are linearly independent vectors in $\mathbb{R}^n$.
(ii) The two dimensional subspace W, spanned by $v^{(1)}$ and $v^{(2)}$, is invariant under f_A, i.e., for any $\alpha, \beta \in \mathbb{R}$, $A\big(\alpha v^{(1)} + \beta v^{(2)}\big) \in W$.
(iii) The matrix representation of the restriction $(f_A)_{|V} : W \to W$ with respect to the basis $[v^{(1)}, v^{(2)}]$ is given by

$$\Big((f_A)_{|W}\Big)_{[v^{(1)},v^{(2)}]\to[v^{(1)},v^{(2)}]} = \begin{pmatrix} \mathrm{Re}(\lambda) & \mathrm{Im}(\lambda) \\ -\mathrm{Im}(\lambda) & \mathrm{Re}(\lambda) \end{pmatrix} = |\lambda| \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}$$

where $|\lambda|e^{\mathrm{i}\theta}$ is the polar representation of the complex number λ. Since $\lambda \notin \mathbb{R}$ we have $\sin\theta \neq 0$.

To finish this section, we consider two classes of matrices in $\mathbb{R}^{n\times n}$.
Orthogonal matrices: A matrix $A \in \mathbb{R}^{n\times n}$ is said to be orthogonal if $A^T A = \mathrm{Id}_n$. Equivalently, it means that

$$\langle Ax, Ay\rangle = \langle x, y\rangle \qquad \forall x, y \in \mathbb{R}^n ,$$

where $\langle\cdot,\cdot\rangle$ denotes the Euclidean scalar product on $\mathbb{R}^n$, $\langle x, y\rangle = \sum_{k=1}^n x_k y_k$.

Theorem 5.2.4. *For any orthogonal matrix $A \in \mathbb{R}^{n\times n}$, the following holds:*
(i) Any eigenvalue λ of A satisfies $|\lambda| = 1$.
(ii) If $v, w \in \mathbb{C}^n$ are eigenvectors for dinstinct eigenvalues λ, μ of A, then $\sum_{k=1}^n v_k \overline{w}_k = 0$.
(iii) $\mathbb{C}^n$ admits a basis of eigenvectors of A.

Remark 5.2.5. Since any orthogonal matrix is unitary Theorem 5.1.11 applies. □

EXAMPLE: The 2×2 matrix $R(\varphi) = \begin{pmatrix} \cos\varphi & \sin\varphi \\ -\sin\varphi & \cos\varphi \end{pmatrix}$, describing the rotation in $\mathbb{R}^2$ by the angle φ, is orthogonal. Its eigenvalues are $\lambda_1 = e^{\mathrm{i}\varphi}$, $\lambda_2 = e^{-\mathrm{i}\varphi}$.

Symmetric matrices: A matrix $A \in \mathbb{R}^{n\times n}$ is said to be symmetric if $A^T = A$. Equivalently, it means that

$$\langle Ax, y\rangle = \langle x, Ay\rangle , \qquad \forall x, y \in \mathbb{R}^n .$$

Theorem 5.2.6. *For any symmetric matrix $A \in \mathbb{R}^{n\times n}$, the following holds:*
(i) Every eigenvalue of A is real and admits an eigenvector x in $\mathbb{R}^n$, $Ax = \lambda x$.
(ii) If λ, μ are eigenvalues of A with $\lambda \neq \mu$, then $\langle x, y\rangle = 0$ for any eigenvector $x \in \mathbb{R}^n$ $[y \in \mathbb{R}^n]$ of A for the eigenvalue λ $[\mu]$.
(iii) $\mathbb{R}^n$ admits an orthonormal basis of eigenvectors of A.

Remark 5.2.7. Since a symmetric matrix $A \in \mathbb{R}^{n\times n}$ is Hermitian, it follows by Theorem 5.1.13 that every eigenvalue of A is real. Assume that $v \in \mathbb{C}^n \setminus \{0\}$ is an eigenvector for the eigenvalue λ of A, $Av = \lambda v$. Then $A\bar{v} = \lambda\bar{v}$ and hence

$$A\big(v+\bar{v}\big) = \lambda(v+\bar{v})\,, \qquad A\big(\mathrm{i}(\bar{v}-v)\big) = \lambda\mathrm{i}(\bar{v}-v)\,.$$

Since $0 \neq 2v = (v+\bar{v}) + \mathrm{i}(\mathrm{i}(\bar{v}-v))$ and $v \neq 0$, either $v+\bar{v} \neq 0$ or $\mathrm{i}(\bar{v}-v) \neq 0$, implying item (i). To see that item (ii) holds, one argues as in the case of Theorem 5.1.13. □

In the case $A \in \mathbb{R}^{n\times n}$ is symmetric one can define the geometric multiplicity n_λ of an eigenvalue λ in the following alternative way: consider

$$E_{\lambda,\mathbb{R}^n}(A) := \Big\{x \in \mathbb{R}^n : Ax = \lambda x\Big\} = E_\lambda(A) \cap \mathbb{R}^n\,.$$

It is straightforward to verify that $E_{\lambda,\mathbb{R}^n}(A)$ is a subspace of $\mathbb{R}^n$. By Theorem 5.2.6 we know that it is a non trivial subspace of $\mathbb{R}^n$. It can be proved that

$$\dim\big(E_{\lambda,\mathbb{R}^n}(A)\big) = \dim\big(E_\lambda(A)\big)$$

and hence the geometric multiplicity n_λ of λ is given by $\dim\big(E_{\lambda,\mathbb{R}^n}(A)\big)$.

Theorem 5.2.8. *Any symmetric matrix $A \in \mathbb{R}^{n\times n}$ can be diagonalized in the following sense: there exists an orthonormal basis $[v] = [v^{(1)}, \ldots, v^{(n)}]$ of $\mathbb{R}^n$, consisting of eigenvectors of A, $Av^{(j)} = \lambda_j v^{(j)}$, $1 \leq j \leq n$, so that*

$$(f_A)_{[v]\to[v]} = S^T A S$$

where S is the orthogonal matrix in $\mathbb{R}^{n\times n}$, given by $\mathrm{Id}_{[v]\to[e]}$. Hence the jth column of S is given by $v^{(j)}$. Furthermore, for any eigenvalue λ of A, the algebraic multiplicity m_λ coincides with the geometric multiplicity n_λ and $n_\lambda = \dim(E_{\lambda,\mathbb{R}^n}(A))$.

5.3 Quadratic forms on $\mathbb{R}^n$

In this section we discuss the notion of quadratic forms on $\mathbb{R}^n$ and discuss applications to geometry.

Definition 5.3.1. *We say that a function $Q : \mathbb{R}^n \to \mathbb{R}$ is a quadratic form on $\mathbb{R}^n$ if it is given by*

$$Q(x) = \sum_{1\leq i,j\leq n} a_{ij}x_i x_j\,, \qquad a_{ij} \in \mathbb{R}\,.$$

It means that Q is a polynomial homogeneous of degree 2 in the variables $x_1, \ldots, x_n$ with real coefficients a_{ij}, $1 \leq i, j \leq n$.

Since

$$Q(x) = \sum_{1\leq i,j\leq n} \frac{a_{ij}+a_{ji}}{2} x_i x_j = \sum_{1\leq i,j\leq n} \frac{1}{2}(A+A^T)_{ij}x_i x_j$$

we can assume without loss of generality that A is symmetric, i.e., $A = A^T$. We say that $Q = Q_A$ is the quadratic form associated to the symmetric matrix A. One can represent the quadratic form Q_A with the help of the Euclidean inner product on $\mathbb{R}^n$,

$$Q_A(x) = \langle Ax, x\rangle\,, \qquad \forall x \in \mathbb{R}^n\,.$$

EXAMPLE: Assume that $Q(x) = 3x_1^2 + 5x_1x_2 + 8x_2^2$. With

$$A = \begin{pmatrix} 3 & 5/2 \\ 5/2 & 8 \end{pmatrix} \in \mathbb{R}^{2\times 2}$$

one gets $Q(x) = Q_A(x) = \langle Ax, x\rangle$ for all $x \in \mathbb{R}^2$.

Quadratic forms can be classified as follows:

Definition 5.3.2. *A quadratic form $Q : \mathbb{R}^n \to \mathbb{R}$ is said to be positive definite [positive semi-definite] if $Q(x) > 0$ [$Q(x) \geq 0$] for any $x \in \mathbb{R}^n \setminus \{0\}$. It is said to be negative definite [negative semi-definite] if $Q(x) < 0$ [$Q(x) \leq 0$] for any $x \in \mathbb{R}^n \setminus \{0\}$. Finally, Q is said to be indefinite if there exists $x, y \in \mathbb{R}^n$ such that $Q(x) > 0$ and $Q(y) < 0$.*

Note that for a positive definite quadratic form Q, $x = 0$ is a global strict minimum of Q. If Q is indefinite, then $x = 0$ is a saddle point. To decide the type a given quadratic form Q_A, it is useful to introduce the following calssification of symmetric $n \times n$ matrices:

Definition 5.3.3. *A symmetric $n \times n$ matrix $A \in \mathbb{R}^{n\times n}$ is said to be positive definite [positive semi-definite] if any eigenvalue λ of A satisfies $\lambda > 0$ [$\lambda \geq 0$]. Similarly A is said to be negative definite [negative semi-definite] if any eigenvalue of A satisfies $\lambda < 0$ [$\lambda \leq 0$]. Finally, A is said to be indefinite if there exist eigenvalues λ, μ such that $\lambda < 0 < \mu$.*

It turns out that the classifications of quadratic forms and symmetric matrices are closely related. For any quadratic form Q_A with $A \in \mathbb{R}^{n\times n}$ symmetric, it follows from Theorem 5.2.8 that there exists an orthogonal matrix $S \in \mathbb{R}^{n\times n}$ and real numbers $\lambda_1, \ldots, \lambda_n$ such that $A = SBS^T$ where $B = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$. Here $\lambda_1, \ldots, \lambda_n$ are the eigenvalues of A and the jth column of S is an eigenvector of A for λ_j. As a consequence

$$Q_A(x) = \langle Ax, x\rangle = \langle SBS^Tx, x\rangle = \langle BS^Tx, S^Tx\rangle\,.$$

With $y = S^Tx$, $x \in \mathbb{R}^n$ one then can write

$$Q_A(x) = \sum_{j=1}^{n} \lambda_j y_j^2, \qquad y = (y_1, \ldots, y_n) = S^Tx\,.$$

It yields the following relationship between the classification of quadratic forms Q_A and symmetric matrices A.

- Q_A is positive definite [positive semi-definite] if and only if A is positive definite [positive semi-definite]

- Q_A is negative definite [negative semi-definite] if and only if A is negative definite [negative semi-definite]
- Q_A is indefinite if and only if A is indefinite.

To decide the type of a given symmetric $n \times n$ matrix A, it is not necessary to compute the spectrum of A. The following result characterizes positive definite and positive semi-definite symmetric matrices.

Theorem 5.3.1. *Assume that $A \in \mathbb{R}^{n\times n}$ is symmetric and denote by $A^{(k)}$ the $k\times k$ matrix $A^{(k)} = (a_{ij})_{1\le i,j\le k}$. Then the following holds:*
(i) A is positive definite if and only if $\det(A^{(k)}) > 0$ for any $1 \le k \le n$.
(ii) A is positive semi-definite if and only if $\det(A^{(k)}) \ge 0$ for any $1 \le k \le n$.

EXAMPLE: Consider the symmetric 3×3 matrix

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

Then $A^{(1)} = 1$, $A^{(2)} = \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}$ and $A^{(3)} = A$. One computes $\det A^{(1)} = 1 > 0$, $\det A^{(2)} = 1 > 0$, $\det A^{(3)} = 1 > 0$.

We finish this section with an application to geometry. Consider a function $f : \mathbb{R}^2 \to \mathbb{R}$ of the form

$$f(x_1, x_2) = ax_1^2 + bx_1x_2 + cx_2^2 + dx_1 + ex_2 + k \tag{5.3.1}$$

with real coefficients a, b, c, d, e, and k. It means that f is a polynomial of degree 2 in the variables x_1, x_2. The coefficients of the polynomial f are assumed to be real. Then we refer to

$$Q_A(x) = \langle x, Ax\rangle, \qquad A = \begin{pmatrix} a & b/2 \\ b/2 & c \end{pmatrix} \in \mathbb{R}^{2\times 2}$$

as the quadratic form associated to f.

Definition 5.3.4. *A conic section is a subset of the form*

$$K_f = \big\{(x_1, x_2) \in \mathbb{R}^2 : f(x_1, x_2) = 0\big\},$$

where f is a polynomial of the form (5.3.1).

EXAMPLES
(i) For $f(x_1, x_2) = (x_1 + x_2)^2 + 1$, $K_f = \emptyset$.
(ii) For $f(x_1, x_2) = (x_1+x_2)^2 - 1$, $K_f = \big\{(x_1, x_2) \in \mathbb{R}^2 : x_1+x_2 = 1 \quad \text{or} \quad x_1+x_2 = -1\big\}$, i.e., K_f is the union of two straight lines.
(iii) For $f(x_1, x_2) = (x_1+x_2)^2$, $K_f = \big\{(x_1, x_2) \in \mathbb{R}^2 : x_1+x_2 = 0\big\}$, i.e., K_f is one straight line.

(*iv*) For $f(x_1, x_2) = x_1^2 + x_2^2$, $K_f = \{(0,0)\}$.
(*v*) For $f(x_1, x_2) = \frac{1}{4}x_1^2 + x_2^2 - 1$, $K_f = \left\{(x_1, x_2) \in \mathbb{R}^2 : \frac{1}{4}x_1^2 + x_2^2 = 1\right\}$, i.e., K_f is an ellipse centered at $(0,0)$ with half axes of length 2 and 1.
(*vi*) For $f(x_1, x_2) = x_1^2 - x_2$, $K_f = \left\{(x_1, x_2) \in \mathbb{R}^2 : x_2 = x_1^2\right\}$ is a parabola with vertex $(0,0)$.
(*vii*) For $f(x_1, x_2) = x_1 x_2 - 1$, $K_f = \left\{(x_1, x_2) \in \mathbb{R}^2 : x_2 = 1/x_1\right\}$ is a hyperbola with two branches.

Remark 5.3.2. A conic section K_f is said to be *degenerate* if K_f is empty or a point set or a straight line or a union of two straight lines. Otherwise it is called nondegenerate. □

Theorem 5.3.3. *By means of translations and / or rotations in $\mathbb{R}^2$, any nondegenerate conic section can be brought into canonical form, i.e., in one of the following curves:*
(*i*) *Ellipse with center* $(0,0)$,

$$\frac{x_1^2}{a_1^2} + \frac{x_2^2}{a_2^2} = 1\,, \qquad a_1, a_2 \in \mathbb{R}_{>0}\,.$$

The associated symmetric matrix A is given by $A = \mathrm{diag}(\frac{1}{a_1^2}, \frac{1}{a_2^2})$ and $\det A > 0$.
(*ii*) *Hyperbola, centered at* $(0,0)$, *with two branches,*

$$\frac{x_1^2}{a_1^2} - \frac{x_2^2}{a_2^2} = 1\,, \qquad a_1, a_2 \in \mathbb{R}_{>0}\,.$$

The associated symmetric matrix A is given by $A = \mathrm{diag}(\frac{1}{a_1^2}, -\frac{1}{a_2^2})$ and $\det A < 0$.
(*iii*) *Parabola with vertex* $(0,0)$,

$$x_1^2 = a x_2 \qquad \text{or} \qquad x_2^2 = a x_1\,, \qquad a \in \mathbb{R} \setminus \{0\}\,.$$

The associated symmetric matrix A is given by $A = \mathrm{diag}(1, 0)$ and $\det A = 0$.

Examples: (*i*) Consider the polynomial

$$f(x_1, x_2) = 9x_1^2 - 18x_1 + 4x_2^2 + 16x_2 - 11\,.$$

Bring K_f in canonical form, if possible. First we complete squares in the following expressions

$$9x_1^2 - 18x_1 = 9(x_1^2 - 2x_2) = 9((x_1 - 1)^2 - 1) = 9(x_1 - 1)^2 - 9$$
$$4x_2^2 + 16x_2 = 4(x_2^2 + 4x_2) = 4((x_2 + 2)^2 - 4) = 4(x_2 + 2)^2 - 16$$

to get

$$f(x_1, x_2) = 9(x_1 - 1)^2 + 4(x_2 + 2)^2 - 36\,.$$

Then $f(x_1, x_2) = 0$ if and only if

$$\frac{9(x_1 - 1)^2}{36} + \frac{4(x_2 + 2)^2}{36} = 1$$

or

$$\frac{(x_1-1)^2}{2^2}+\frac{(x_2+2)^2}{3^2}=1\,.$$

Hence by the translation $x\mapsto y:=x-(1,-2)$ one gets

$$\frac{y_1^2}{a_1^2}+\frac{y_2^2}{a_2^2}=1\,,\qquad a_1=2\,,\,a_2=3\,,$$

i.e., K_f is an ellipse with center $(1,-2)$.

(ii) Consider

$$f(x_1,x_2)=3x_1^2+2x_1x_2+3x_2^2-8\,.$$

Decide whether K_f is nondegenerate and if so, bring it into canonical form. First note that

$$f(x)=\langle x,Ax\rangle-8\,,\qquad A=\begin{pmatrix}3&1\\1&3\end{pmatrix}$$

The eigenvalues of A can be computed to be $\lambda_1=2$, $\lambda_2=4$. Hence $\det A=2\cdot 4=8>0$. This shows that K_f is an ellipse. To bring it in canonical form, we have to diagonalize A. One verifies that

$$v^{(1)}=\frac{1}{\sqrt{2}}(1,-1)\,,\qquad v^{(2)}=\frac{1}{\sqrt{2}}(1,1)$$

are eigenvectors of A corresponding to the eigenvalues λ_1 and λ_2. Then $S=\mathrm{Id}_{[v]\to[e]}$ is the orthogonal 2×2 matrix

$$S=\frac{1}{\sqrt{2}}\begin{pmatrix}1&1\\-1&1\end{pmatrix}$$

and $A=S\mathrm{diag}(2,4)S^T$. Hence

$$\langle x,Ax\rangle=\langle S^Tx,\begin{pmatrix}2&0\\0&4\end{pmatrix}S^Tx\rangle\,,\qquad\forall x\in\mathbb{R}^2\,.$$

Thus we have shown that $K_f=\left\{x=Sy:\frac{y_1^2}{4}+\frac{y_2^2}{2}=1\right\}$. When, expressed in the y coordinates, K_f is an ellipse with axes of length 2 and $\sqrt{2}$.

One can investigate the zero sets of polynomials also in higher dimensions. In the case $n=3$, one considers polynomials $f:\mathbb{R}^3\to\mathbb{R}$ of the form

$$f(x)=\langle x,Ax\rangle+\langle b,x\rangle+c$$

where $A\in\mathbb{R}^{3\times 3}$ is symmetric, $b\in\mathbb{R}^3$ and $c\in\mathbb{R}$. The zero set $K_f=\left\{x\in\mathbb{R}^3:f(x)=0\right\}$ is called a quadric surface or a quadric. Types of non degenerate quadrics in $\mathbb{R}^3$ are the following surfaces:

- *Ellipsoid:* for any $a_1,a_2,a_3>0$,

$$a_1^2x_1^2+a_2^2x_2^2+a_3^2x_3^2=1\,.$$

- *Hyperboloids:* for any $a_1, a_2, a_3 > 0$,

$$\text{hyperboloid with one sheet :} \qquad -a_1^2 x_1^2 + a_2^2 x_2^2 + a_3^2 x_3^2 = 1\,;$$

$$\text{hyperboloid with two sheets :} \qquad -a_1^2 x_1^2 - a_2^2 x_2^2 + a_3^2 x_3^2 = 1\,.$$

- *Paraboloids:* for any $a_3 \neq 0$ and $a_1, a_2 > 0$,

$$\text{elliptic paraboloid :} \qquad a_3 x_3 = a_1^2 x_1^2 + a_2^2 x_2^2\,;$$

$$\text{hyperbolic paraboloid :} \qquad a_3 x_3 = -a_1^2 x_1^2 + a_2^2 x_2^2\,.$$

Chapter 6

Differential equations

The aim of this chapter is to present a brief introduction to the theory of ordinary differential equations. The main focus is on systems of linear differential equations of first order in $\mathbb{R}^n$ with constant coefficients. They can be solved by the means of linear algebra. Hence this chapter is an application of what we have learnt so far to a topic in the field of analysis.

6.1 Introduction

Mathematical models describing the dynamics of systems, considered in the sciences, are often expressed in terms of differential equations, relating the quantities describing the essential features of the system. Such models are introduced and analyzed with the aim of predicting how the systems evolve in time. Prominent examples are models of mechanical systems, such as the motion of a particle of mass m in the space $\mathbb{R}^3$, models for radioactive decay, population models etc.

Motion of a particle in $\mathbb{R}^3$. By Newton's law, the motion of a particle in $\mathbb{R}^3$ is described by

$$y : \mathbb{R} \to \mathbb{R}^3\,, \quad t \mapsto y(t)$$

with t (independent variable) denoting time and $y(t)$ (dependent variable) the position of the particle at time t, determined by

$$my''(t) = F$$

together with the initial conditions $y(0) = y^{(0)}$, $y'(0) = y^{(1)}$. Here $y'(t) = \frac{d}{dt}y(t)$ is the velocity and $y''(t) = \frac{d^2}{d^2t}y(t)$ is the acceleration of the particle at time t, whereas F is the vector resulting from all the forces acting on the particle. Furthermore $y^{(0)}$ denotes the position and $y^{(1)}$ the velocity of the particle at time $t = 0$. The equation $my''(t) = F$ is a differential equation in case the force F only depends on t, on the position $y(t)$ at time t, on the velocity $y'(t)$ at time t, ... , but not on the values $y(s)$ for $s \neq t$.

Radioactive decay: A substance, such as radium, decays by a stochastic process. It is assumed that the probability P of the decay of an atom of the substance in an infinitesimal

time interval $[t, t+\Delta t]$ is proportional to Δt, i.e., there exists a constant λ, depending on the substance considered, so that

$$P\Big(\text{decay of atom in } [t, t+\Delta t]\Big) = \lambda \Delta t\,.$$

Denote by $N(t)$ the number of atoms of the substance at time t. It is then expected that

$$N(t+\Delta t) - N(t) = -N(t) \cdot P\Big(\text{decay of atom in } [t, t+\Delta t]\Big) = -N(t)\lambda \Delta t$$

and hence that the total mass of the substance $x(t) = mN(t)$ obeys the law

$$x(t+\Delta t) - x(t) = -x(t)\lambda \Delta t\,.$$

Here m is the mass of a single atom. Taking the limit $\Delta t \to 0$, one is led to the equation

$$x'(t) = -\lambda x(t)\,.$$

More generally, if one is given three substances X, Y, Z, where X decays to Y and Y decays to Z, then the total masses $x(t)$, $y(t)$, $z(t)$ of the substances X, Y, Z at time t satisfy a system of equations of the form

$$\begin{cases} x'(t) = -\lambda x(t) \\ y'(t) = \lambda x(t) - \mu y(t) \\ z'(t) = \mu y(t)\,. \end{cases}$$

Population models: Denote by $p(t)$ the number of individuals of a given species at time t. Often it is assumed that the growth rate

$$\frac{p'(t)}{p(t)}$$

depends on the population $p(t)$, i.e., that it is modeled by a function $r(p)$, yielding the equation

$$\frac{p'(t)}{p(t)} = r(p(t))\,.$$

In case the function r is a constant $r \equiv \alpha$, one speaks of exponential growth

$$p'(t) = \alpha \cdot p(t)\,.$$

In applications, it is often observed that the growth rate becomes negative if the population exceeds a certain threshold p_0. In such a case one frequently chooses for r an affine function

$$r(p) = \beta \cdot (p_0 - p)$$

leading to the equation

$$p'(t) = \alpha \cdot (p_0 - p(t)) \cdot p(t) = \alpha p_0 p(t) - \alpha p(t)^2\,.$$

(The equation is also referred to as logistic equation.)

Assume that

$$f : \mathbb{R} \times \mathbb{R} \times \mathbb{R} \to \mathbb{R}\,, \qquad (t, x_0, x_1) \mapsto f(t, x_0, x_1)$$

is a sufficiently regular function for the purposes considered. We then say that

$$f(t, x(t), x'(t)) = 0\,, \qquad t \in \mathbb{R} \tag{6.1.1}$$

is a ordinary differential equation (ODE) of first order and that a continuously differentiable function $x : \mathbb{R} \to \mathbb{R}$ satisfying $f(t, x(t), x'(t)) = 0$ for any $t \in \mathbb{R}$ (or alternatively for any t in some open nonempty interval) is a solution of (6.1.1). More generally, if

$$F : \mathbb{R} \times \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^m$$

is a sufficiently regular vector valued map, we say that

$$F(t, y(t), y'(t)) = 0\,, \qquad t \in \mathbb{R} \tag{6.1.2}$$

is a system of ordinary differential equations of first order and that the continuously differentiable vector valued function

$$y : \mathbb{R} \to \mathbb{R}^n$$

satisfying $F(t, y(t), y'(t)) = 0$ for any $t \in \mathbb{R}$ is a solution of (6.1.2). We say that (6.1.2) is in explicit form if it can be written in the form

$$y'(t) = G(t, y(t))\,. \tag{6.1.3}$$

Note that in this case one has $m = n$ and G is a map $G : \mathbb{R} \times \mathbb{R}^n \to \mathbb{R}^n$. More generally a system of ODEs of n-th order in explicit form is an equation of the form

$$y^{(n)}(t) = G(t, y(t), y'(t), \ldots, y^{(n-1)}(t)) \tag{6.1.4}$$

where $y^{(j)}(t)$ denotes the j-th derivative of $y : \mathbb{R} \to \mathbb{R}^n$ and

$$G : \mathbb{R} \times \mathbb{R}^n \times \cdots \times \mathbb{R}^n \to \mathbb{R}^n$$

is sufficiently regular for the purposes considered. In the sequel, we will only consider ODEs of the form (6.1.4). We remark that systems of order n can always be converted into systems of first order, albeit of higher dimension. See the discussion in Section 6.3 concerning second order ODEs. So, we may restrict our attention to equations of the form

$$y'(t) = G(t, y(t)) \tag{6.1.5}$$

where $G : \mathbb{R} \times \mathbb{R}^n \to \mathbb{R}^n$.

The main questions concerning the equation (6.1.5) are the existence and the uniqueness of solutions and their properties. We remark that only in very rare cases, the solution

can be represented in terms of an explicit formula. Hence investigations of qualitative properties of solutions play an important role. In particular one is interested to know whether solutions exist for all time or whether some of them blow up in finite time. Of special interest is their asymptotic behaviour as $t \to +\infty$ or as t approaches the blow up time. In addition, one wants to find out if there are special solutions such as stationary solutions or periodic solutions and investigate their stability.

Associated to (6.1.5) is the so called initial value problem (IVP)

$$\begin{cases} y'(t) = G(t, y(t)) \\ y(0) = y^{(0)} \end{cases} \tag{6.1.6}$$

where $y^{(0)} \in \mathbb{R}^n$ is a given vector. There are general theorems saying that under appropriate conditions of the map G, (6.1.6) has a unique solution at least in some time interval containing $t = 0$. An important class of equations of the form (6.1.5) are the so called linear ODEs. We say that (6.1.5) is a linear ODE if G is of the form

$$G(t, y) = A(t)y + f(t)$$

or written componentwise

$$G_j(t, y) = \sum_{k=1}^{n} a_{jk}(t) y_k + f_j(t)\,, \qquad 1 \le j \le n \tag{6.1.7}$$

where $A : \mathbb{R} \to \mathbb{R}^{n\times n}$, $t \mapsto A(t) = (a_{jk}(t))_{1\le j,k\le n}$ is a matrix valued map. Note that the linearity refers to the (dependent) variable y but not to the (independent) variable t. The real valued functions $t \mapsto a_{jk}(t)$ are referred to as the coefficients of (6.1.7). In case $f = 0$, (6.1.7) is said to be a homogeneous ODE, otherwise a inhomogeneous one. We say that a linear ODE of the form (6.1.7) has constant coefficients if all the coefficients $a_{jk}(t)$, $1 \le j, k \le n$, are independent of t. Note that f need not to be constant in t. Linear ODEs of the form (6.1.7) with constant coefficients can be solved explicitly. Such ODEs will be discussed in Section 6.2 whereas ODEs of second order (with constant coefficients) will be studied in Section 6.3. We point out that the theory of ODEs is a large field within analysis and that in this chapter we only discuss a tiny, albeit important, part of it. We also mention that a field closely related to the field of ordinary differential equations is the field of partial differential equations including equations such as the wave equation, the heat equation, the transport equation, the Schrödinger equation, the Maxwell equations, the Einstein equations, ... In contrast to ODEs, besides time, there are other independent variables such as space variables or more generally, variables in a phase space ... . The field of partial differential equations is huge and currently a very active research area.

6.2 Systems of linear ODEs of first order with constant coefficients

In this section we treat systems of linear differential equations of first order with n unknowns $y = (y_1, \dots, y_n)$,

$$Py'(t) + Qy(t) = g(t) \tag{6.2.1}$$

where P, Q are matrices in $\mathbb{R}^{n\times n}$ with constant coefficients and $g : \mathbb{R} \to \mathbb{R}^n$ is a continuous function. In applications, the variable t has often the meaning of time and $y'(t)$ denotes the derivative of $y(t)$ with respect to t. We are looking for solutions $y : t \mapsto y(t)$ which are continuously differentiable. In the sequel we will always assume that P is invertible. Hence we might multiply left and right hand side of (6.2.1) by P^{-1} yielding

$$y'(t) = Ay(t) + f(t)$$

where $A \in \mathbb{R}^{n\times n}$ has constant coefficients and $f : \mathbb{R} \to \mathbb{R}^n$ is assumed to be continuous. We first treat the case in which f identically vanishes. The system

$$y'(t) = Ay(t) \tag{6.2.2}$$

is referred to as a homogeneous system of linear ODEs with constant coefficients. Fundamental questions are whether (6.2.2) has a solution and if a solution is unique. Written componentwise, the system reads

$$\begin{cases} y_1'(t) = \sum_{j=1}^n a_{1j} y_j(t) \\ \vdots \\ y_n'(t) = \sum_{j=1}^n a_{nj} y_j(t) \end{cases}$$

where $A = (a_{ij})_{1\le i,j\le n}$.

Let us first consider the special case $n = 1$. Writing a for $A = (a_{11})$, equation (6.2.2) becomes $y'(t) = ay(t)$. We claim that $y(t) = ce^{at}$, $c \in \mathbb{R}$ arbitrary, is a solution. Indeed by substituting $y(t) = ce^{at}$ into the equation $y'(t) = ay(t)$ and using that $\frac{d}{dt}(e^{at}) = ae^{at}$ one sees that this is the case.

How can one find a solution of $y'(t) = ay(t)$? A possible way is to use the *method of separation of variables*. It consists in transforming the equation in such a way that the left hand side is an expression in y and its derivative only whereas the right hand side does not involve y and its derivative at all. In the case at hand we argue formally. Divide $y' = ay$ by y to get

$$\frac{y'}{y} = a\,.$$

Since, formally, $y'/y = \frac{d}{dt}\log(y(t))$, one concludes that

$$\int_0^t \frac{d}{dt}\log(y(t))\,dt = at\,.$$

Using that $\log(y(t)) - \log(y(0)) = \log(y(t)/y(0))$, one gets $y(t) = y(0)e^{at}$. Hence for any initial value $y_0 \in \mathbb{R}$, $y_0 e^{at}$ is a solution of the initial value problem

$$\begin{cases} y'(t) = ay(t) \\ y(0) = y_0 \end{cases} \qquad (IVP)$$

It turns out that $y_0 e^{at}$ is the unique solution of (IVP). Indeed assume that $z : \mathbb{R} \to \mathbb{R}$ is a continuously differentiable function such that (IVP) holds. Then consider $w(t) = e^{-at}z(t)$. By the product rule for differentiation

$$w'(t) = -ae^{-at}z(t) + e^{-at}z'(t)\,.$$

Since $z'(t) = az(t)$, it follows that $w'(t) = 0$ for all $t \in \mathbb{R}$. Hence $w(t)$ is constant in time. Since $w(0) = z(0) = y_0$, one has $y_0 = e^{-at}z(t)$ implying that $z(t) = y_0 e^{-at}$. In summary we have seen that in the case $n = 1$, the equation (6.2.2) admits a one parameter family of solutions and the initial value problem (IVP) has, for any initial value $y_0 \in \mathbb{R}$, a unique solution. In the general case $n \geq 2$, do similar results hold? First we would like to investigate if the exponential e^a can be defined when a is replaced by an arbitrary $n \times n$ matrix A. Recall that in Chapter 2, we defined the exponential e^z of a complex number by the power series expansion

$$e^z = \sum_{n=0}^{\infty} \frac{z^n}{n!} = 1 + z + \frac{z^2}{2!} + \dots$$

If we replace z by a $n \times n$ matrix A, then $A^2 = AA$, and inductively, for any $n \geq 1$, $A^{n+1} = AA^n$ is well defined. It can be shown that the series $\sum_{n=0}^{\infty} \frac{A^n}{n!}$ converges and defines a $n \times n$ matrix which is denoted by e^A,

$$e^A = \sum_{n=0}^{\infty} \frac{A^n}{n!}\,. \qquad (6.2.3)$$

Examples:
(i)

$$A = \mathrm{diag}(-1, 2)\,.$$

e^A can be computed as follows

$$A^2 = \mathrm{diag}((-1)^2, 2^2)\,, \quad A^3 = \mathrm{diag}((-1)^3, 2^3)\,, \dots$$

hence

$$\begin{aligned} e^A &= \sum_{n=0}^{\infty} \frac{A^n}{n!} = \sum_{n=0}^{\infty} \frac{\mathrm{diag}((-1)^n, 2^n)}{n!} \\ &= \begin{pmatrix} \sum_{n=0}^{\infty} \frac{(-1)^n}{n!} & 0 \\ 0 & \sum_{n=0}^{\infty} \frac{2^n}{n!} \end{pmatrix} = \begin{pmatrix} e^{-1} & 0 \\ 0 & e^2 \end{pmatrix}. \end{aligned}$$

(*ii*) Let $S \in \mathbb{R}^{2\times 2}$ be invertible and consider $B = S^{-1}AS$, $A = \mathrm{diag}(-1, 2)$. Then e^B can be computed as follows

$$B^2 = (S^{-1}AS)(S^{-1}AS) = S^{-1}A^2S\,,$$

and, inductively,

$$B^{n+1} = (S^{-1}AS)B^n = (S^{-1}AS)S^{-1}A^nS = S^{-1}A^{n+1}S\,,$$

hence

$$e^B = \sum_{n=0}^{\infty}\frac{B^n}{n!} = \sum_{n=0}^{\infty}\frac{S^{-1}A^nS}{n!} = S^{-1}\Big(\sum_{n=0}^{\infty}\frac{A^n}{n!}\Big)S = S^{-1}e^AS\,.$$

By item (i) it then follows that

$$e^B = S^{-1}\mathrm{diag}(e^{-1}, e^2)S\,.$$

Theorem 6.2.1. *For any $A \in \mathbb{R}^{n\times n}$, the map $\mathbb{R} \to \mathbb{R}^{n\times n}$, $t \mapsto e^{tA}$ is continuously differentiable and satisfies*

$$\begin{cases} \frac{d}{dt}e^{tA} = Ae^{tA} \\ e^{tA}{}_{|t=0} = \mathrm{Id}_n \end{cases} \tag{6.2.4}$$

Remark 6.2.2. (*i*) Let us comment on why Theorem 6.2.1 holds. Clearly $e^{0A} = \mathrm{Id}_n$. Furthermore, by differnetiating term by term, one gets at least formally,

$$\begin{aligned}\frac{d}{dt}(e^{tA}) &= \frac{d}{dt}\Big(\mathrm{Id}_n + tA + \frac{t^2A^2}{2!} + \dots\Big) \\ &= A + \frac{2tA^2}{2!} + \frac{3t^2A^3}{3!}\dots \\ &= A\Big(\mathrm{Id} + tA + \frac{t^2A^2}{2!} + \dots\Big) = Ae^{tA}\,.\end{aligned}$$

(*ii*) One can show that for any $t, s \in \mathbb{R}$, $e^{(t+s)A} = e^{tA}e^{sA}$. Indeed for any given $s \in \mathbb{R}$, let $E(t) = e^{tA}e^{sA}$. Then $t \mapsto E(t)$ is continuously differentiable and satisfies

$$\begin{cases} E'(t) = AE(t) \\ E(0) = e^{sA}\,. \end{cases}$$

On the other hand the latter theorem implies that

$$\frac{d}{dt}e^{(t+s)A} = Ae^{(t+s)A}\,, \qquad e^{(t+s)A}{}_{|t=0} = e^{sA}\,.$$

One can show that the solution of (6.2.4) is unique and hence $E(t) = e^{(t+s)A}$ for any $t \in \mathbb{R}$. Since $s \in \mathbb{R}$ is arbitrary, the claimed identity follows.
(*iii*) By item (*i*) applied for $s = -t$, one has

$$e^{tA}e^{-tA} = e^{(t-t)A} = \mathrm{Id}_n\,,$$

meaning that for any $t \in \mathbb{R}$, e^{tA} is invertible with inverse e^{-tA}. □

Theorem 6.2.3. *For any $A \in \mathbb{R}^{n\times n}$ and $y^{(0)} \in \mathbb{R}^n$, the initial value problem*

$$y'(t) = Ay(t)\,, \quad y(0) = y^{(0)}$$

has a unique solution. It is given by $y(t) = e^{tA}y^{(0)}$. Furthermore, the general solution of $y' = Ay$ is given by $e^{tA}v$, where $v = (v_1, \dots, v_n) \in \mathbb{R}^n$. In particular, if $t \mapsto u(t)$ and $t \mapsto v(t)$ are solutions of $y' = Ay$, so is $t \mapsto au(t) + bv(t)$ for any $a, b \in \mathbb{R}$.

Remark 6.2.4. (i) We often write y_0 instead of $y^{(0)}$ for the initial value.
(ii) By Theorem 6.2.1, $y(t) = e^{tA}y^{(0)}$ satisfies $y(0) = e^{0A}y^{(0)} = y^{(0)}$ and

$$\frac{d}{dt}e^{At}y^{(0)} = Ae^{tA}y^{(0)} = Ay(t)\,.$$

To see that this is the only solution, assume that $z : \mathbb{R} \to \mathbb{R}^n$ is another solution, i.e, $z'(t) = Az(t)$ and $z(0) = y^{(0)}$. Define $w(t) = e^{-tA}z(t)$ and note that

$$w(0) = e^{-0A}z(0) = \mathrm{Id}_n y^{(0)} = y^{(0)}$$

and

$$\begin{aligned} w'(t) &= \frac{d}{dt}(e^{-tA})z(t) + e^{-tA}z'(t) \\ &= -Ae^{-tA}z(t) + e^{-tA}Az(t) \\ &= -Ae^{-tA}z(t) + Ae^{-tA}z(t) = 0\,, \end{aligned}$$

where we have used that $e^{-tA}A = Ae^{-tA}$ in view of the definition of $e^{-tA} = \sum_{n=0}^{\infty} \frac{(-t)^n}{n!}A^n$. It then follows that $w(t)$ is a vector independent of t, i.e., $w(t) = w(0) = y^{(0)}$, implying that for any $t \in \mathbb{R}$

$$e^{tA}y^{(0)} = e^{tA}w(t) = e^{tA}e^{-tA}z(t) = z(t)\,.$$

□

Examples (i) Find the general solution of

$$\begin{cases} y_1' = 3y_1 + 4y_2 \\ y_2' = 3y_1 + 2y_2 \end{cases}$$

In matrix notation, the system reads $y' = Ay$ where

$$A = \begin{pmatrix} 3 & 4 \\ 3 & 2 \end{pmatrix} \in \mathbb{R}^{2\times 2}\,.$$

The general solution is given by $y(t) = e^{tA}v$, $v = (v_1, v_2) \in \mathbb{R}^2$. To determine the solution $y(t)$ in a more explicit way, we analyze e^{tA} as follows. The eigenvalues of A can be determined to be $\lambda_1 = -1$, $\lambda_2 = 6$ with corresponding eigenvectors

$$v^{(1)} = (1, -1)\,, \qquad v^{(2)} = (4, 3)\,.$$

Note that $A^k v^{(1)} = \lambda_1^k v^{(1)}$ for any $k \geq 0$ and hence

$$e^{tA} v^{(1)} = \sum_{k=0}^{\infty} \frac{t^k}{k!} A^k v^{(1)} = \sum_{k=0}^{\infty} \frac{t^k}{k!} \lambda_1^k v^{(1)} = e^{t\lambda_1} v^{(1)}$$

and similarly $e^{tA} v^{(2)} = e^{t\lambda_2} v^{(2)}$. Since $v^{(1)}$ and $v^{(2)}$ are linearly independent, any vector $v \in \mathbb{R}^2$ can be uniquely represented as a linear combination

$$v = a_1 v^{(1)} + a_2 v^{(2)} .$$

Hence

$$\begin{aligned} e^{tA} v &= e^{tA} \Big(a_1 v^{(1)} + a_2 v^{(2)} \Big) = a_1 e^{tA} v^{(1)} + a_2 e^{tA} v^{(2)} \\ &= a_1 e^{t\lambda_1} v^{(1)} + a_2 e^{t\lambda_2} v^{(2)} \\ &= \Big(a_1 e^{-t} + 4 a_2 e^{6t} \, , \, -a_1 e^{-t} + 3 a_2 e^{6t} \Big) . \end{aligned}$$

To solve the initial value problem with $y^{(0)} = (6, 1)$ we need to solve the linear system

$$\begin{cases} a_1 + 4a_2 = 6 \\ -a_1 + 3a_2 = 1 \end{cases}$$

One computes $a_1 = 2$ and $a_2 = 1$, hence

$$y(t) = \Big(2e^{-t} + 4e^{6t} \, , \, -2e^{-t} + 3e^{6t} \Big) .$$

The asymptotics of the solution $y(t)$ for $t \to \pm\infty$ can be described as follows:

$$y(t) \simeq (2e^{-t}, -2e^{-t}) \qquad t \to -\infty ,$$

$$y(t) \simeq (4e^{6t}, 3e^{6t}) \qquad t \to +\infty .$$

(ii) Find the general solution of

$$\begin{cases} y_1' = y_1 + y_2 \\ y_2' = -2y_1 + 3y_2 \end{cases}$$

In matrix notation, the system reads $y' = Ay$ where

$$A = \begin{pmatrix} 1 & 1 \\ -2 & 3 \end{pmatrix} \in \mathbb{R}^{2\times 2} .$$

The general solution is given by $y(t) = e^{tA} v$, $v = (v_1, v_2) \in \mathbb{R}^2$. To determine the solutions $y(t)$ in a more explicit way, we analyze e^{tA} as follows. The eigenvalues of A can be computed to be $\lambda_1 = 2 + \mathrm{i}$, $\lambda_2 = \bar{\lambda}_1 = 2 - \mathrm{i}$ with corresponding eigenvectors

$$v^{(1)} = (1, 1 + \mathrm{i}) \in \mathbb{C}^2 , \qquad v^{(2)} = \bar{v}^{(1)} = (1, 1 - \mathrm{i}) \in \mathbb{C}^2$$

which form a basis of $\mathbb{C}^2$. The general complex solution of $y' = Ay$ is then given by

$$y(t) = a_1 e^{\lambda_1 t} v^{(1)} + a_2 e^{\lambda_2 t} v^{(2)}\,, \qquad a_1, a_2 \in \mathbb{C}\,.$$

How can we obtain the general real solution? Recall that by Euler's formula, $e^{\alpha+\mathrm{i}\beta} = e^{\alpha}\Big(\cos\beta + \mathrm{i}\sin\beta\Big)$, hence

$$e^{\lambda_1 t} = e^{2t}\Big(\cos t + \mathrm{i}\sin t\Big)$$

implying that

$$e^{\lambda_1 t} v^{(1)} = \Big(e^{2t}\cos t + \mathrm{i}e^{2t}\sin t\,,\ e^{2t}(\cos t - \sin t) + \mathrm{i}e^{2t}(\cos t + \sin t)\Big)\,.$$

Since $e^{\lambda_1 t} v^{(2)} = \overline{e^{\lambda_2 t} v^{(1)}}$ one gets real solutions by considering

$$\frac{1}{2}\Big(e^{\lambda_1 t} v^{(1)} + e^{\lambda_2 t} v^{(2)}\Big)\,, \quad \frac{1}{2\mathrm{i}}\Big(e^{\lambda_1 t} v^{(1)} - e^{\lambda_2 t} v^{(2)}\Big)\,.$$

They can be computed to be

$$\Big(e^{2t}\cos t\,,\, e^{2t}(\cos t - \sin t)\Big)\,, \quad \Big(e^{2t}\sin t\,,\, e^{2t}(\cos t + \sin t)\Big).$$

Hence the general real solution is given by

$$a_1\Big(e^{2t}\cos t\,,\, e^{2t}(\cos t - \sin t)\Big) + a_2\Big(e^{2t}\sin t\,,\, e^{2t}(\cos t + \sin t)\Big)\,, \quad a_1, a_2\ \in \mathbb{R}\,.$$

(iii) Find the general solution of

$$\begin{cases} y_1' = y_1 + y_2 \\ y_2' = y_2\,. \end{cases}$$

In matrix notation, the system reads

$$y' = Ay\,, \qquad A := \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}\,.$$

The general solution is given by $e^{tA}v$, where $v = (v_1, v_2)$ is an arbitrary vector in $\mathbb{R}^2$. To determine the solutions in a more explicit way we analyze e^{tA} further. Note that in this case A is not diagonalizable. To compute e^{tA} we use that for $S, T \in \mathbb{R}^{n\times n}$ with $ST = TS$, one has $e^{S+T} = e^S e^T$ (cf Exercises). Since

$$A = \mathrm{Id}_2 + T\,, \quad T = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}\,, \quad \mathrm{Id}_2 T = T\mathrm{Id}_2\,,$$

it follows that $e^{tA} = e^{t\mathrm{Id}_2} e^{tT}$. Clearly $e^{t\mathrm{Id}_2} = e^t \mathrm{Id}_2$ and

$$e^{tT} = \sum_{k=0}^{\infty} \frac{t^k}{k!} T^k = \mathrm{Id}_2 + tT\,,$$

since $T^2 = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$. Altogether $e^{tA} = e^t \begin{pmatrix} 1 & t \\ 0 & 1 \end{pmatrix}$ implying that the general real solution is given by

$$y_1(t) = v_1 e^t + v_2 t e^t\,, \quad y_2(t) = v_2 e^t$$

where $v_1, v_2 \in \mathbb{R}$ are arbitrary constants.

Let us now turn to the system

$$y'(t) = Ay(t) + f(t)\,, \tag{6.2.5}$$

referred to as inhomogeneous system of linear ODEs of first order with constant coefficients.

Theorem 6.2.5. *Assume that $A \in \mathbb{R}^{n\times n}$, $f : \mathbb{R} \to \mathbb{R}^n$ continuous and $y_p : \mathbb{R} \to \mathbb{R}^n$ a given solution of (6.2.5), $y_p'(t) = Ay_p(t) + f(t)$. Then the following holds:*
(i) For any solution $u(t)$ of the homogeneous system, $u'(t) = Au(t)$, $y_p(t) + u(t)$ is a solution of (6.2.5).
(ii) For any solution $y(t)$ of (6.2.5), there exists a solution $u(t)$ of the homogeneous system $u'(t) = Au(t)$, such that $y(t) = y_p(t) + u(t)$.

Remark 6.2.6. (*i*) By substituting $y_p(t) + u(t)$ into (6.2.5), one gets

$$\frac{d}{dt}\Big(y_p(t) + u(t)\Big) = y_p'(t) + u'(t) = Ay_p(t) + f(t) + Au(t) = A\big(y_p(t) + u(t)\big) + f(t)$$

and hence $y_p(t) + u(t)$ is a solution of (6.2.5).
(*ii*) Let $u(t) := y(t) - y_p(t)$. Then

$$u'(t) = y'(t) - y_p'(t) = Ay(t) + f(t) - Ay_p(t) - f(t) = A(y(t) - y_p(t)) = Au(t)\,.$$

□

As a consequence of the theorem above, the general solution of (6.2.5) can be found as follows:

1. Find a particular solution of (6.2.5).

2. Find the general solution of the homogeneous system $u' = Au$.

In special cases, this method allows to find the general solution in a very efficient way – see our discussion later in this section. First we want to present a method, referred to as the *method of variation of the constants*, which allows to construct the general solution of (6.2.5). The method is always applicable, but for some classes of equations there are easier ways to construct the general solution. The starting point for the method of variation of constants is the ansatz

$$y(t) = e^{tA} w(t)\,.$$

In case $w(t)$ is a constant vector $v \in \mathbb{R}^n$, $y(t)$ is a solution of the homogeneous system. Hence the ansatz consists in replacing the constant v by a t-dependent unknown function $w(t)$, explaining the name of the method. Substituting the ansatz into the inhomogeneous equation one gets

$$\begin{aligned} y'(t) &= \frac{d}{dt}\big(e^{tA}\big)w(t) + e^{tA}w'(t) \\ &= Ay(t) + e^{tA}w'(t)\,. \end{aligned}$$

In case $y(t)$ is a solution of (6.2.5), it then follows that

$$Ay(t) + f(t) = Ay(t) + e^{tA}w'(t) \quad \text{or} \quad f(t) = e^{tA}w'(t)\,.$$

Multiplying both sides of the equation by the matrix e^{-tA} one gets

$$w'(t) = e^{-tA}f(t)\,.$$

Integrating in t, one obtains $w(t) = w(0) + \int_0^t e^{-sA}f(s)\,ds$. Altogether we found

$$y(t) = e^{tA}w(t) = e^{tA}w(0) + e^{tA}\int_0^t e^{-sA}f(s)\,ds\,.$$

Hence

$$y(t) = e^{tA}v + \int_0^t e^{(t-s)A}f(s)\,ds \tag{6.2.6}$$

with $v \in \mathbb{R}^n$ arbitrary, is the general solution of (6.2.5). Note that $e^{tA}v$ is the general solution of the homogeneous system $u' = Au$, whereas $\int_0^t e^{(t-s)A}f(s)\,ds$ is a particular solution y_p of (6.2.5) with $y_p(0) = 0$.
Special case: Assume that in the equation

$$y'(t) = Ay(t) + f(t)\,,$$

f is a solution of the homogeneous equation $u'(t) = Au(t)$, i.e., $f(t) = e^{tA}a$, $a \in \mathbb{R}^n$. Substituting f into the integral $\int_0^t e^{(t-s)A}f(s)\,ds$ leads to

$$\int_0^t e^{(t-s)A}f(s)\,ds = \int_0^t e^{tA}e^{-sA}e^{sA}a\,ds = te^{tA}a\,.$$

Alternatively, one verifies in a straightforward way that whenever $f(t)$ is a solution of the homogeneous system, then $y_p(t) = tf(t)$ is a particular solution of $y'(t) = Ay(t) + f(t)$.

Application: Formula (6.2.6) can be used to solve the initial value problem (IVP)

$$\begin{cases} y' = Ay + f \\ y(0) = y^{(0)}\,. \end{cases}$$

The solution is given by

$$y(t) = e^{tA}y^{(0)} + \int_0^t e^{(t-s)A} f(s)\, ds\,.$$

Example: Find the general solution of

$$\begin{cases} y_1' = -y_2 \\ y_2' = y_1 + t \end{cases}\,.$$

In matrix notation, the system reads $y' = Ay + (0,t)$ where $A = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$. Fisrt we compute e^{tA}. The eigenvalues of A are $\lambda_1 = \mathrm{i}$ and $\lambda_2 = -\mathrm{i}$ and the corresponding eigenvectors are

$$v^{(1)} = (\mathrm{i}, 1)\,, \qquad v^{(2)} = (-\mathrm{i}, 1)\,.$$

Hence

$$e^{tA}v^{(1)} = e^{\mathrm{i}t}v^{(1)}\,, \qquad e^{tA}v^{(2)} = e^{-\mathrm{i}t}v^{(2)}\,.$$

Using that $e^{\pm \mathrm{i}t} = \cos t \pm \mathrm{i}\sin t$ and

$$\frac{v^{(1)} + v^{(2)}}{2} = (0,1)\,, \qquad \frac{\mathrm{i}(v^{(2)} - v^{(1)})}{2} = (1,0)$$

it follows that

$$e^{tA}\begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} -\sin t \\ \cos t \end{pmatrix}\,, \qquad e^{tA}\begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} \cos t \\ \sin t \end{pmatrix}$$

implying that

$$(e^{tA})_{[e]\to[e]} = \begin{pmatrix} \cos t & -\sin t \\ \sin t & \cos t \end{pmatrix} = (\cos t)\mathrm{Id}_2 + (\sin t)A\,.$$

The general solution then reads

$$e^{tA}v + \int_0^t e^{(t-s)A}\begin{pmatrix} 0 \\ s \end{pmatrix}\, ds\,.$$

One has

$$e^{(t-s)A}\begin{pmatrix} 0 \\ s \end{pmatrix} = \cos(t-s)\begin{pmatrix} 0 \\ s \end{pmatrix} - \sin(t-s)\begin{pmatrix} -s \\ 0 \end{pmatrix}$$

and

$$\int_0^t -s\sin(t-s)\, ds = \Big[-s\cos(t-s) - \sin(t-s)\Big]_0^t = -t + \sin t$$

$$\int_0^t s\cos(t-s)\, ds = \Big[-s\sin(t-s) + \cos(t-s)\Big]_0^t = 1 - \cos t\,.$$

Altogether we have

$$\int_0^t e^{(t-s)A}\begin{pmatrix}0\\ s\end{pmatrix}\,ds=\begin{pmatrix}-t+\sin t\\ 1-\cos t\end{pmatrix}$$

and the general solution reads

$$v_1\begin{pmatrix}\cos t\\ \sin t\end{pmatrix}+v_2\begin{pmatrix}-\sin t\\ \cos t\end{pmatrix}+\begin{pmatrix}-t+\sin t\\ 1-\cos t\end{pmatrix}$$

As an aside we remark that there is the following alternative way of computing the exponential e^{tA}. Note that

$$A=\begin{pmatrix}0&-1\\ 1&0\end{pmatrix}\,,\quad A^2=\begin{pmatrix}-1&0\\ 0&-1\end{pmatrix}\,,\quad A^3=-A\,,\quad A^4=\mathrm{Id}_2\,,\ \ldots$$

and thus

$$e^{tA}=\sum_{k=0}^{\infty}\frac{t^k}{k!}A^k=\sum_{k=0}^{\infty}\frac{t^{2k}}{(2k)!}A^{2k}+\sum_{k=0}^{\infty}\frac{t^{2k+1}}{(2k+1)!}A^{2k+1}=\sum_{k=0}^{\infty}\frac{t^{2k}}{(2k)!}\mathrm{Id}_2+\sum_{k=0}^{\infty}\frac{t^{2k+1}}{(2k+1)!}A$$

implying that $e^{tA}=(\cos t)\mathrm{Id}_2+(\sin t)A\,.$

The example above shows that the computation of $\int_0^t e^{(t-s)A}f(s)\,ds$ can be quite involved. For special classes of functions $f:\mathbb{R}\to\mathbb{R}^n$ it is easier to get a particular solution by making an ansatz. We now discuss three classes of functions to illustrate this method.

Polynomials: Assume that each component of f is a polynomial in t of degree at most L. It means that

$$f(t)=\sum_{j=0}^{L}t^j f^{(j)}\,,\quad f^{(0)},\ldots,f^{(L)}\in\mathbb{R}^n\,.$$

We restrict to the case where the $n\times n$ matrix A is invertible and make an ansatz for a particular solution $y_p(t)$ by assuming that each component of $y_p(t)$ is a polynomial in t of degree at most L, i.e., we assume that

$$y_p(t)=\sum_{j=0}^{L}t^j w^{(j)}\,,\quad w^{(0)},\ldots,w^{(L)}\in\mathbb{R}^n\,.$$

Since $y_p'(t)=\sum_{j=0}^{L-1}(j+1)t^j w^{(j+1)}$, one obtains, upon substitution of $y_p(t)$ into the equation $y_p'(t)=Ay_p(t)+f(t)$,

$$\sum_{j=0}^{L-1}(j+1)t^j w^{(j+1)}=\sum_{j=0}^{L}t^j Aw^{(j)}+\sum_{j=0}^{L}t^j f^{(j)}.$$

By comparison of coefficients, one gets

$$(j+1)w^{(j+1)}=Aw^{(j)}+f^{(j)}\quad 0\le j\le L-1\,,\quad\text{and}\quad 0=Aw^{(L)}+f^{(L)}\,.$$

Since A is assumed to be invertible, we can solve this linear system recursively. Solving the equation $0 = Aw^{(L)} + f^{(L)}$ yields $w^{(L)} = -A^{-1}f^{(L)}$, which then allows to solve the remaining equations by setting for $j = L-1, \dots, j = 0$,

$$Aw^{(j)} = -f^{(j)} + (j+1)w^{(j+1)} \quad \text{or} \quad w^{(j)} = A^{-1}((j+1)w^{(j+1)} - f^{(j)})$$

To illustrate how to determine $q_1, \dots, q_n$, let us go back to the example treated above

$$y'(t) = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} y(t) + \begin{pmatrix} 0 \\ t \end{pmatrix} .$$

where $f(t) = (0, t)$. Note that the matrix $\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$ is invertible and hence we make the ansatz $y_p(t) = (q_1(t), q_2(t))$,

$$q_1(t) = a_0 + a_1 t\,, \qquad q_2(t) = b_0 + b_1 t\,.$$

Upon substitution of the ansatz in the above equation and using that $y_p'(t) = (a_1, b_1)$, one gets

$$\begin{pmatrix} a_1 \\ b_1 \end{pmatrix} = \begin{pmatrix} -b_0 - b_1 t \\ a_0 + a_1 t + t \end{pmatrix} \qquad \forall t \in \mathbb{R}\,.$$

Comparing coefficients yields

$$a_1 = -b_0\,, \quad 0 = -b_1 \quad \text{and} \quad b_1 = a_0\,, \quad 0 = a_1 + 1\,,$$

hence $a_1 = -1$, $b_1 = 0$, $a_0 = 0$, and $b_0 = 1$, yielding $y_p(t) = (-t, 1)$.

To see that in the case where the components of f are polynomials of degree at most L, the ansatz for $y_p(t)$, consisting in choosing the components of y_p to be polynomials of degree at most L, not always works, consider the following scalar valued ODE

$$y'(t) = f(t) \quad \text{with} \quad f(t) = 4 + 6t\,.$$

The general solution can be found by integration, $y_p(t) = c_1 + 4t + 3t^2$. Clearly, for no choice of c_1, y_p will be a polynomial of degree at most one. Not that in this example the 1×1 matrix A is zero.

Trigonometric polynomials: Assume that each component of $f : \mathbb{R} \to \mathbb{R}^n$ is a trigonometric polynomial. It means that f can be written in the form

$$f(t) = f^{(0)} + \sum_{j=1}^{J} \cos(\xi_j t) f^{(2j-1)} + \sin(\xi_j t) f^{(2j)}\,, \quad \text{with} \quad f^{(0)}\,, \dots, f^{(2J)} \in \mathbb{R}^n.$$

Here $\xi_1, \dots, \xi_J$ are in $\mathbb{R} \setminus \{0\}$. We restrict to the case where the $n \times n$ matrix A is invertible and $\pm i\xi_j$ are not eigenvalues of A. We then make the ansatz for a particular solution $y_p(t)$ of the form

$$y_p(t) = w^{(0)} + \sum_{j=1}^{J} \cos(\xi_j t) w^{(2j-1)} + \sin(\xi_j t) w^{(2j)}\,, \quad \text{with} \quad w^{(0)}\,, \dots, w^{(2J)} \in \mathbb{R}^n.$$

Upon substitution of the ansatz into the equation $y' = Ay + f$, the vectors $w^{(j)}$ can then be determined by comparison of coefficients. Let us illustrate the method with an example. Let

$$y'(t) = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} y(t) + \begin{pmatrix} \cos 2t \\ 1 \end{pmatrix} .$$

Clearly the 2×2 matrix $\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$ is invertible and its eigenvalues are i and $-i$. Following the considerations above, we make the ansatz $y_p(t) = (q_1(t), q_2(t))$ where

$$q_1(t) = a_0 + a_1 \cos 2t + a_2 \sin 2t \,, \quad q_2(t) = b_0 + b_1 \cos 2t + b_2 \sin 2t \,.$$

Since

$$y_p'(t) = \begin{pmatrix} -2a_1 \sin 2t + 2a_2 \cos 2t \\ -2b_1 \sin 2t + 2b_2 \cos 2t \end{pmatrix} \quad \text{and} \quad Ay_p(t) = \begin{pmatrix} -b_0 - b_1 \cos 2t - b_2 \sin 2t \\ a_0 + a_1 \cos 2t + a_2 \sin 2t \end{pmatrix}$$

one is led to the following equation

$$\begin{pmatrix} -2a_1 \sin 2t + 2a_2 \cos 2t \\ -2b_1 \sin 2t + 2b_2 \cos 2t \end{pmatrix} = \begin{pmatrix} -b_0 - b_1 \cos 2t - b_2 \sin 2t + \cos 2t \\ a_0 + a_1 \cos 2t + a_2 \sin 2t + 1 \end{pmatrix}$$

Comparison of coefficients leads to the linear system

$$b_0 = 0 \,, \quad -2a_1 = -b_2 \,, \quad 2a_2 = -b_1 + 1 \,, \qquad 0 = a_0 + 1 \,, \quad -2b_1 = a_2 \,, \quad 2b_2 = a_1$$

having the solution

$$a_0 = -1 \,, \quad b_0 = 0 \,, \quad a_1 = 0 \,, \quad b_2 = 0 \,, \quad a_2 = \frac{2}{3} \,, \quad b_1 = -\frac{1}{3} \,.$$

Altogether, we get

$$y_p(t) = (-1 + \frac{2}{3} \sin 2t, -\frac{1}{3} \cos 2t) \,.$$

Exponentials: Assume that each component of $f : \mathbb{R} \to \mathbb{R}^n$ is a linear combination of exponential functions, i.e., each component is an element in

$$V := \Big\{ \sum_{j=1}^{J} a_j e^{\xi_j t} : a_1, \ldots, a_J \in \mathbb{R} \Big\}$$

and $\xi_1, \ldots, \xi_J$ are dinstinct real numbers. Then V is a $\mathbb{R}$–vector space of dimension J with the property that for any $g \in V$, its derivative g' is again in V. In the case where no exponent ξ_j is an eigenvalue of A, we make the ansatz for a particular solution $y_p(t)$, by assuming that each component of $y_p(t)$ is an element in V. The particular solution $y_p(t)$ is then again computed by comparison of coefficients.

6.3 Linear ODEs of higher order with constant coefficients

The main goal of this section is to discuss linear ODEs of second order with constant coefficients. They come up in many applications and therefore are particularly important. Specifically, we consider ODEs of the form

$$y'' = b_1 y' + b_0 y + f \tag{6.3.1}$$

where $b_1, b_0 \in \mathbb{R}$ are constants and $f : \mathbb{R} \to \mathbb{R}$ is a continuous function. We want to discuss two methods for solving such equations.

Method 1: Convert the equation (6.3.1) into a system of ODEs of first order with constant coefficients in the following way: let $x(t) = (x_0(t), x_1(t)) \in \mathbb{R}^2$ be given by

$$x_0(t) := y(t)\,, \qquad x_1(t) := y'(t)\,.$$

Then

$$\begin{aligned} x'(t) &= \begin{pmatrix} x_0'(t) \\ x_1'(t) \end{pmatrix} = \begin{pmatrix} y'(t) \\ y''(t) \end{pmatrix} = \begin{pmatrix} x_1(t) \\ b_1 y'(t) + b_0 y(t) + f(t) \end{pmatrix} \\ &= \begin{pmatrix} x_1(t) \\ b_1 x_1(t) + b_0 x_0(t) + f(t) \end{pmatrix} \end{aligned}$$

or, in matrix notation,

$$x'(t) = Ax(t) + \begin{pmatrix} 0 \\ f(t) \end{pmatrix}\,, \qquad A := \begin{pmatrix} 0 & 1 \\ b_0 & b_1 \end{pmatrix}\,. \tag{6.3.2}$$

The equations (6.3.1) and (6.3.2) are equivalent in the sense that they have the same set of solutions: a solution y of (6.3.1) gives rise to a solution $x(t) = (y(t), y'(t))$ of (6.3.2) and conversely, a solution $x(t) = (x_0(t), x_1(t))$ of (6.3.2) yields a solution $y(t) = x_0(t)$ of (6.3.1). More generally, an ODE of the form

$$y^{(n)}(t) = b_{n-1} y^{(n-1)}(t) + \cdots + b_1 y'(t) + b_0 \tag{6.3.3}$$

can be converted into a system of ODEs of first order by setting

$$x_0(t) := y(t)\,,\ \ x_1(t) := y'(t)\,,\ \ \ldots\ , x_{n-1}(t) := y^{(n-1)}(t)\,.$$

Then one gets $x_0'(t) = x_1(t)\,, \ldots, x_{n-2}'(t) = x_{n-1}(t)$ and

$$x_{n-1}'(t) = y^{(n)}(t) = b_{n-1} y^{(n-1)}(t) + \cdots + b_1 y'(t) + b_0 y + f(t)\,.$$

In matrix notation,

$$x'(t) = Ax(t) + (0, \ldots, 0, f(t))\,, \tag{6.3.4}$$

where

$$A := \begin{pmatrix} 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \dots & 1 \\ b_0 & b_1 & b_2 & \dots & b_{n-1} \end{pmatrix} \in \mathbb{R}^{n\times n}\,. \tag{6.3.5}$$

Again the equations (6.3.3) and (6.3.4) are equivalent in the sense that they have the same set of solutions.

Example: Consider

$$y''(t) = -k^2 y(t)\,, \qquad k \in \mathbb{R}\,, \quad k > 0\,. \tag{6.3.6}$$

The equation is a model for the vibrations of a string without damping. Converting this equation into a 2×2 system leads to the following first order ODE,

$$x'(t) = Ax(t)\,, \qquad x(t) = \begin{pmatrix} x_0(t) \\ x_1(t) \end{pmatrix}\,, \quad A = \begin{pmatrix} 0 & 1 \\ -k^2 & 0 \end{pmatrix}\,.$$

Then the general solution is given by $x(t) = e^{tA}c$, $c = (c_1, c_2) \in \mathbb{R}^2$. In order to determine $y(t) = x_0(t)$ in a more explicit form we need to analyze e^{tA} further. The eigenvalues of A are $\lambda_1 = \mathrm{i}k$, $\lambda_2 = -\mathrm{i}k$ and corresponding eigenvectors are

$$v^{(1)} = (1,\, \mathrm{i}k)\,, \qquad v^{(2)} = (1,\, -\mathrm{i}k)\,.$$

Note that $v^{(1)}, v^{(2)}$ form a basis of $\mathbb{C}^2$. The general complex solution is then given by

$$a_1 e^{\mathrm{i}kt} v^{(1)} + a_2 e^{-\mathrm{i}kt} v^{(2)}\,, \qquad a_1, a_2 \in \mathbb{C}\,. \tag{6.3.7}$$

The general real solution is then given by an arbitrary linear combination of real and imaginary part of the general complex solution of (6.3.7),

$$c_1 \begin{pmatrix} \cos kt \\ -k \sin kt \end{pmatrix} + c_2 \begin{pmatrix} \sin kt \\ k \cos kt \end{pmatrix}\,, \quad c_1, c_2 \in \mathbb{R}\,.$$

Method 2: This is a method to find the general solution of a homogeneous ODE of order 2, $y'' = b_1 y' + b_0 y$ or more generally of order n, $y^{(n)} = b_{n-1} y^{(n-1)} + \cdots + b_0 y$. The ansatz is $y(t) = e^{\lambda t}$ with $\lambda \in \mathbb{C}$ to be determined. Let us illustrate this method with the equation (6.3.6), $y'' = -k^2 y$. Substituting the ansatz $y(t) = e^{\lambda t}$ into the equation one gets

$$\lambda^2 e^{\lambda t} = y''(t) = -k^2 y(t) = -k^2 e^{\lambda t}\,,$$

hence

$$\lambda^2 = -k^2\,, \qquad \text{or} \qquad \lambda_1 = \mathrm{i}k\,, \quad \lambda_2 = -\mathrm{i}k\,.$$

The general complex solution reads

$$a_1 e^{\mathrm{i}kt} + a_2 e^{-\mathrm{i}kt}\,, \qquad a_1, a_2 \in \mathbb{C}\,,$$

whereas the general real solution is given by

$$c_1 \cos kt + c_2 \sin kt\,, \qquad c_1, c_2 \in \mathbb{R}\,.$$

The advantage of this method is that the given equation has not to be converted into a system of first order. However we will see that in some situations, the ansatz has to be modified to get all solutions in this way.

EXAMPLE: (i) Consider the homogeneous ODE of second order

$$y'' + 3y' + 2y = 0\,. \tag{6.3.8}$$

With the ansatz $y(t) = e^{\lambda t}$ we get $y'(t) = \lambda e^{\lambda t}$, $y''(t) = \lambda^2 e^{\lambda t}$. Hence substituting $y(t) = e^{\lambda t}$ into (6.3.8) one is led to

$$\lambda^2 + 3\lambda + 2 = 0\,.$$

The roots are given by

$$\lambda_1 = -1\,, \qquad \lambda_2 = -2\,.$$

The general real solution of (6.3.8) is therefore

$$y(t) = c_1 e^{-t} + c_2 e^{-2t}\,, \qquad c_1, c_2 \in \mathbb{R}\,.$$

(ii) Consider the inhomogeneous ODE of second order

$$y'' + 3y' + 2y = 8\,. \tag{6.3.9}$$

To find the general solution, we could apply method 1 and convert the equation into a system of first order ODEs. As an alternative, we could also use method 2. First, note that for the homogeneous ODE

$$u'' + 3u' + 2u = 0 \tag{6.3.10}$$

the superposition principle holds, i.e., for any solutions u, v of (6.3.10), and any $\alpha, \beta \in \mathbb{R}$, $\alpha u(t) + \beta v(t)$ is also a solution of (6.3.10). Arguing as for systems of first order, it is straightforward to verify that the general solution of (6.3.9) is given by

$$\text{general solution of (6.3.10)} \; + y_p$$

where y_p is a particular solution of (6.3.9). To find a particular solution, we can try to make an appropriate ansatz. Note that the inhomogeneous term is the constant function $f = 8$ which is a polynomial of degree 0. Thus we try the ansatz $y_p = c$. Substituting into the equation (6.3.9) one gets

$$0 + 3 \cdot 0 + 2c = 8 \quad \text{implying} \quad c = 4\,.$$

Altogether we conclude that

$$c_1 e^{-t} + c_2 e^{-2t} + 4\,, \qquad c_1, c_2 \in \mathbb{R}$$

is the general solution of (6.3.9). As for systems of ODEs of first order, one can study the initial value problem for the equation (6.3.1) or more generally for the equation (6.3.3)

$$(IVP) \quad \begin{cases} y^{(n)} = b_{n-1} y^{(n-1)} + \cdots + b_1 y' + b_0 y + f \\ y(0) = a_0\,,\ \ldots\ , y^{(n-1)}(0) = a_{n-1} \end{cases}$$

where $a_0, \ldots, a_{n-1}$ are arbitrary real numbers. Note that there are n initial conditions for an equation of order n. This corresponds to a vector $a \in \mathbb{R}^n$ of initial values for the initial value problem of a $n \times n$ system of first order ODEs. As in that case, the above (IVP) has a unique solution. Let us illustrate how to find this solution with the following example.

EXAMPLE: (i) Consider the initial value problem

$$y'' = -k^2 y\,, \qquad y(0) = 1\,, \quad y'(0) = 0\,,$$

where we assume again that $k > 0$. We have seen that the general complex solution is given by

$$y(t) = a_1 e^{\mathrm{i}kt} + a_2 e^{-\mathrm{i}kt}\,, \qquad a_1, a_2 \in \mathbb{C}\,.$$

We need to determine $a_1, a_2 \in \mathbb{C}$ so that $y(0) = 1$, $y'(0) = 0$, i.e.,

$$1 = y(0) = a_1 + a_2\,, \qquad 0 = y'(0) = \mathrm{i}k a_1 - \mathrm{i}k a_2\,.$$

This is a linear 2×2 system with the two unknowns a_1 and a_2. We get $a_1 = a_2$ and $a_1 = 1/2$. Hence

$$y(t) = \frac{1}{2} e^{\mathrm{i}kt} + \frac{1}{2} e^{-\mathrm{i}kt} = \cos kt$$

is the solution of the initial value problem. Note that it is automatically real valued.

(ii) Consider the initial value problem

$$y'' + 3y' + 2y = 8\,, \quad y(0) = 1\,, \quad y'(0) = 0\,.$$

We have seen that the general solution of $y'' + 3y' + 2y = 8$ is given by

$$y(t) = c_1 e^{-t} + c_2 e^{-2t} + 4\,, \qquad c_1, c_2 \in \mathbb{R}\,.$$

We need to determine c_1, c_2 in such a way that $y(0) = 1$, $y'(0) = 0$, i.e.,

$$1 = y(0) = c_1 + c_2 + 4\,, \qquad 0 = y'(0) = -c_1 - 2c_2$$

yielding $c_1 = -6$, $c_2 = 3$ and hence

$$y(t) = -6e^{-t} + 3e^{-2t} + 4\,.$$

EXAMPLE: Since second order ODEs of the form

$$y'' + ay' + by = 0\,, \qquad a, b \in \mathbb{R}$$

frequently come up in applications, we finish this section with a detailed discussion on this equation. We concentrate on the case where $a > 0$ and $b = \omega^2$ with $\omega > 0$, which is a model for the vibrations of a string with damping.

Let us use method 2. Making the ansatz $y(t) = e^{\lambda t}$ and taking into account that $y'(t) = \lambda e^{\lambda t}$, $y''(t) = \lambda^2 e^{\lambda t}$, one is led to the equation

$$(\lambda^2 + a\lambda + b)e^{\lambda t} = 0\,, \qquad \forall t \in \mathbb{R}\,.$$

Hence $\lambda^2 + a\lambda + b = 0$ or

$$\lambda_\pm = -\frac{a}{2} \pm \frac{1}{2}\sqrt{a^2 - 4\omega^2}\,.$$

Case 1: $a^2 - 4\omega^2 > 0$ or $0 < 2\omega < a$. This is the case of strong damping, leading to solutions with no oscillations. Indeed, one has $\lambda_- < \lambda_+ < 0$ and the general real solution is given by

$$y(t) = c_+ e^{\lambda_+ t} + c_- e^{\lambda_- t}\,, \qquad c_+, c_- \in \mathbb{R}\,.$$

Case 2: $a^2 - 4\omega^2 < 0$ or $0 < a < 2\omega$. This is the case of weak damping, leading to solutions with oscillations. Then $\lambda_\pm = -\frac{a}{2} \pm \mathrm{i}\gamma$ where $\gamma := \frac{1}{2}\sqrt{4\omega^2 - a^2}$. Then

$$e^{\lambda_\pm t} = e^{-\frac{a}{2}t} e^{\pm \mathrm{i}\gamma} = e^{-\frac{a}{2}t}\Big(\cos\gamma t + \mathrm{i}\sin\gamma t\Big)$$

and the general real solution has is given by

$$c_1 e^{-\frac{a}{2}t}\cos\gamma t + c_2 e^{-\frac{a}{2}t}\sin\gamma t\,, \qquad c_1, c_2 \in \mathbb{R}\,, \qquad \gamma = \frac{1}{2}\sqrt{4\omega^2 - a^2}\,.$$

Note that as $t \to \infty$, the solutions tend to 0. When compared with the case without damping, i.e., the case where $a = 0$, the frequency γ of the oscillation is reduced by it.

Case 3: $a^2 - 4\omega^2 = 0$ or $a = 2\omega$. Then $\lambda_+ = \lambda_- = -\frac{a}{2}$. It can be verified that besides $e^{-\frac{a}{2}t}$, also $te^{-\frac{a}{2}t}$ is a solution and hence the general real solution is given by

$$c_1 e^{-\frac{a}{2}t} + c_2 t e^{-\frac{a}{2}t}\,, \qquad c_1, c_2 \in \mathbb{R}\,.$$

There are no oscillations in this case, but as $t \to \infty$ the decay is weaker than in the case 1. Recall that by converting $y'' = -by' - ay$ to a 2×2 system we get $x' = Ax$, with

$$A = \begin{pmatrix} 0 & 1 \\ -b & -a \end{pmatrix}\,, \quad b = \frac{a^2}{4}\,.$$

Note that A cannot be diagonalized as otherwise it would have to be the matrix $-\frac{a}{2}\mathrm{Id}_2$.

Let us finish this section by discussing the inhomogeneous ODE of second order

$$y'' + ay' + by = f \tag{6.3.11}$$

where $f : \mathbb{R} \to \mathbb{R}$ is continuous and $a, b \in \mathbb{R}$. As mentioned above, the general solution of (6.3.11) can be found by converting the equation to a 2×2 system of first order. As an alternative we can first determine the general solution of the corresponding homogeneous equation

$$u'' + au' + bu = 0$$

by making the ansatz $u(t) = e^{\lambda t}$. The set of solutions of (6.3.11) is then given by

$$\{y_p + u : u \text{ satisfies } \ u'' + au' + bu = 0\}$$

where y_p is a particular solution of (6.3.11). Similarly as in the case of systems of differential equations of first order with constant coefficients, for certain classes of functions f it is quite easy to find a particular solution.

Special case: f solution of homogeneous equation. Assume that f satisfies $f''+af'+bf = 0$. If $a^2 = 4b$, we have seen that the general solution of $u'' + au' + bu = 0$ is given by $c_1 e^{-\frac{a}{2}t} + c_2 t e^{-\frac{a}{2}t}$, $c_1, c_2 \in \mathbb{R}$. By assumption, $f(t) = f_1 e^{-\frac{a}{2}t} + f_2 t e^{-\frac{a}{2}t}$ where $f_1, f_2 \in \mathbb{R}$ are determined by $f(0) = f_1$, $f'(0) = (-a/2)f_1 + f_2$. Making the ansatz

$$y_p(t) = (\alpha_1 t + \alpha_2 t^2) t e^{-\frac{a}{2}t}$$

and substituting it into (6.3.11) one finds by comparison of coefficients $\alpha_1 = \frac{f_1}{2}$, $\alpha_2 = \frac{f_2}{6}$, yielding

$$y_p(t) = \Big(\frac{f_1}{2} t^2 + \frac{f_2}{6} t^3\Big) e^{-\frac{a}{2}t}\,.$$

If $a^2 \neq 4b$, we have seen that the general complex solution of $u'' + au' + bu = 0$ is given by $c_1 e^{\lambda_1 t} + c_2 e^{\lambda_2 t}$, where $c_1, c_2 \in \mathbb{C}$ and

$$\lambda_1 = -\frac{a}{2} + \frac{1}{2}\sqrt[+]{a^2 - 4b}\,, \qquad \lambda_2 = -\frac{a}{2} - \frac{1}{2}\sqrt[+]{a^2 - 4b}\,,$$

(with $\sqrt[+]{a^2 - 4b}$ defined as $\mathrm{i}\sqrt[+]{4b - a^2}$ in the case $a^2 < 4b$). By assumption, $f(t) = f_1 e^{\lambda_1 t} + f_2 e^{\lambda_2 t}$ where $f_1, f_2 \in \mathbb{C}$ are uniquely determined by $f(0) = f_1 + f_2$, $f'(0) = \lambda_1 f_1 + \lambda_2 f_2$. Note that f_1, f_2 might be complex numbers even if f is real valued. However, in such a case $f_2 = \overline{f}_1$ and $\lambda_2 = \overline{\lambda}_1$. Making the ansatz

$$y_p(t) = \alpha_1 t e^{\lambda_1 t} + \alpha_2 t e^{\lambda_2 t}$$

and substituting it into (6.3.11) one finds by comparison of coefficients

$$\alpha_1 = \frac{f_1}{2\lambda_1 + a}\,, \qquad \alpha_2 = \frac{f_2}{2\lambda_2 + a}\,.$$

Note that $2\lambda_1 + a = \sqrt[+]{a^2 - 4b}$ and $2\lambda_2 + a = -\sqrt[+]{a^2 - 4b}$ do not vanish by assumption, so that α_1, α_2 are well defined. Hence

$$y_p(t) = \frac{f_1}{2\lambda_1 + a} t e^{\lambda_1 t} + \frac{f_2}{2\lambda_2 + a} t e^{\lambda_2 t}$$

is a particular solution of (6.3.11).

Polynomials: In case f ia a polynomial of degree L, $f(t) = f_0 + f_1 t + \ldots + f_L t^L$, we make the ansatz

$$y_p(t) = \alpha_0 + \alpha_1 t + \ldots + \alpha_{L+2} t^{L+2} .$$

Substituting the ansatz into the equation (6.3.11), one gets

$$\sum_{j=2}^{L+2} \alpha_j j(j-1) t^{j-2} + a \sum_{j=1}^{L+2} \alpha_j j t^{j-1} + b \sum_{j=0}^{L+2} \alpha_j t^j = \sum_{j=0}^{L} f_j t^j .$$

The coefficients $\alpha_0, \ldots, \alpha_L$ are then determined by comparison of coefficients:

$$b\alpha_{L+2} = 0 , \quad (L+2)a\alpha_{L+2} + b\alpha_{L+1} = 0 ,$$

$$(j+2)(j+1)\alpha_{j+2} + (j+1)a\alpha_{j+1} + b\alpha_j = f_j , \qquad 0 \le j \le L .$$

To see that a particular solution might not be given by a polynomial of degree L, consider

$$y'' = 2 + 3t^2 .$$

A particular solution can be easily found by integration,

$$y_p(t) = t^2 + \frac{1}{4} t^4 .$$

As another example, consider

$$y'' + 2y' = t .$$

Substituting the ansatz $y_p(t) = \alpha_0 + \alpha_1 t + \alpha_2 t^2 + \alpha_3 t^3$, one gets

$$6\alpha_3 t + 2\alpha_2 + 6\alpha_3 t^2 + 4\alpha_2 t + 2\alpha_1 = t$$

and comparison of coefficients yields

$$\alpha_3 = 0 , \quad 6\alpha_3 + 4\alpha_2 = 1 , \quad 2\alpha_2 + 2\alpha_1 = 0 .$$

Hence $\alpha_3 = 0$, $\alpha_2 = 1/4$, and $\alpha_1 = -1/4$, whereas α_0 can be chosen arbitrarily. A particular solution is hence given by

$$y_p(t) = -\frac{1}{4} t + \frac{1}{4} t^2 .$$

Trigonometric polynomials: In case f is a trigonometric polynomial

$$f(t) = f_0 + \sum_{j=1}^{L} f_{2j-1} \cos(\xi_j t) + \sum_{j=1}^{L} f_{2j} \sin(\xi_j t)$$

with ξ_j real, pairwise different, and $\mathrm{i}\xi_j \notin \{0, \lambda_1, \lambda_2\}$, we make the ansatz

$$y_p(t) = \alpha_0 + \sum_{j=1}^{L} \Big(\alpha_{2j-1} \cos(\xi_j t) + \alpha_{2j} \sin(\xi_j t) \Big) .$$

Here λ_1, λ_2 are the zeros of $\lambda^2 + a\lambda + b = 0$. Substituting the ansatz into the equation, the coefficients $\alpha_0, \ldots, \alpha_{2L}$ are then determined by comparison of coefficients. As an example consider

$$y'' + 2y' + y = const .$$

Note that $\lambda_1 = \lambda_2 = -1$ and $\xi_1 = 1$, hence $\mathrm{i}\xi_1 \notin \{0, -1\}$. Substituting the ansatz

$$y_p(t) = \alpha_0 + \alpha_1 \cos t + \alpha_2 \sin t$$

into the equation and using that $y_p'(t) = -\alpha_1 \sin t + \alpha_2 \cos t$ and $y_p''(t) = -\alpha_1 \cos t - \alpha_2 \sin t$, one gets

$$-\alpha_1 \cos t - \alpha_2 \sin t + 2(-\alpha_1 \sin t + \alpha_2 \cos t) + (\alpha_0 + \alpha_1 \cos t + \alpha_2 \sin t) = \cos t .$$

Hence by comparison of coefficients,

$$-\alpha_1 + 2\alpha_2 + \alpha_1 = 1 , \qquad -\alpha_2 - 2\alpha_1 + \alpha_2 = 0 , \quad \alpha_0 = 0 ,$$

yielding $\alpha_2 = \frac{1}{2}$, $\alpha_1 = 0$, and $\alpha_0 = 0$, i.e.,

$$y_p(t) = \frac{1}{2} \sin t . \tag{6.3.12}$$

As a second example we consider

$$y''(t) + \omega^2 y(t) = A \sin(\omega t) , \qquad \omega > 0 . \tag{6.3.13}$$

Since $\lambda_1 = \mathrm{i}\omega$, $\lambda_2 = -\mathrm{i}\omega$, and $\xi_1 = \omega$, it follows that $\mathrm{i}\xi_1 \in \{0, \lambda_1, \lambda_2\}$. Hence the ansatz $\alpha_1 \cos(\omega t) + \alpha_2 \sin(\omega t)$ for a particular solution of (6.3.13) has to be modified. Note that $c_1 e^{\lambda_1 t} + c_2 e^{\lambda_2 t}$, $c_1, c_2 \in \mathbb{C}$, is the general complex solution of $u'' + \omega^2 u = 0$. It then follows that $\cos(\omega t)$, $\sin(\omega t)$ are a basis of the $\mathbb{R}$-vector space V of the real solutions of the homogeneous equation $u'' + \omega^2 u = 0$,

$$V = \big\{ c_1 \cos(\omega t) + c_2 \sin(\omega t) : c_1, c_2 \in \mathbb{R} \big\} .$$

According to our considerations above, we make the following ansatz for a particular solution

$$y_p(t) = \alpha_1 t \cos(\omega t) + \alpha_2 t \sin(\omega t) .$$

Substituting it into (6.3.13) one gets with

$$\begin{aligned} y_p'(t) &= \alpha_1 \cos(\omega t) + \alpha_2 \sin(\omega t) + t \Big(\alpha_1 \cos(\omega t) + \alpha_2 \sin(\omega t) \Big)' \\ y_p''(t) &= 2 \Big(\alpha_1 \cos(\omega t) + \alpha_2 \sin(\omega t) \Big)' + t \Big(\alpha_1 \cos(\omega t) + \alpha_2 \sin(\omega t) \Big)'' \end{aligned}$$

$$2\Big(\alpha_1\cos(\omega t)+\alpha_2\sin(\omega t)\Big)' = A\sin(\omega t)$$

implying that $\alpha_2 = 0$ and $-2\alpha_1\omega = A$. Hence

$$y_p(t) = -\frac{A}{2\omega}t\cos(\omega t)$$

is a particular solution of (6.3.13). Note that this solution is oscillatory with unbounded amplitude. The general real solution of (6.3.13) is given by

$$c_1\cos(\omega t)+c_2\sin(\omega t)-\frac{A}{2\omega}t\cos(\omega t)\,,\quad c_1,c_2\in\mathbb{R}\,. \tag{6.3.14}$$

It can be used to solve the corresponding initial value problem. To illustrate how this can be done let us find the unique solution of (6.3.13) satisfying the initial conditions

$$y(0)=0\,,\quad y'(0)=1\,. \tag{6.3.15}$$

It means that in the formula (6.3.14), the constants c_1, c_2 have to be determined in such a way that (6.3.15) holds. Therefore, we need to solve the following linear system:

$$c_1 = 0\,,\qquad \omega c_2 - \frac{A}{2\omega} = 1\,.$$

The solution $y(t)$ is then given by

$$y(t) = \frac{1}{\omega}\Big(1+\frac{A}{2\omega}\Big)\sin(\omega t) - \frac{A}{2\omega}t\cos(\omega t)\,.$$

Exponentials: In case f is a linear combination of exponentials

$$f(t) = \sum_{j=1}^{L} f_j e^{\xi_j t}\,,\qquad f_1,\dots,f_L\in\mathbb{R}$$

where $\xi_1,\dots,\xi_L$ are in $\mathbb{R}\setminus\{\lambda_1,\lambda_2\}$ and pairwise different and λ_1,λ_2 are the zeros of $\lambda^2+a\lambda+b=0$. We make the ansatz $y_p(t)=\sum_{j=1}^{L}\alpha_j e^{\xi_j t}$ for a particular solution of $y''+ay'+by=f$. Since

$$y_p'(t) = \sum_{j=1}^{L}\xi_j\alpha_j e^{\xi_j t}\,,\qquad y_p''(t) = \sum_{j=1}^{L}\xi_j^2\alpha_j e^{\xi_j t}\,,$$

one obtains, upon substution, that

$$\sum_{j=1}^{L}\Big(\alpha_j\xi_j^2 + a\alpha_j\xi_j + b\alpha_j\Big)e^{\xi_j t} = \sum_{j=1}^{L} f_j e^{\xi_j t}\,.$$

Hence

$$\alpha_j = \frac{f_j}{\xi_j^2+a\xi_j+b}\,,\qquad 1\le j\le L$$

and

$$y_p(t) = \sum_{j=1}^{L} \frac{f_j}{\xi_j^2 + a\xi_j + b} e^{\xi_j t}$$

is a particular solution of the equation. Note that $\xi_j^2 + a\xi_j + b \neq 0$ for any $1 \leq j \leq L$, since by assumption, $\xi_j \neq \lambda_1, \lambda_2$. As an example consider

$$y'' - 5y' + 4y = 1 + e^t \,.$$

Note that $\lambda^2 - 5\lambda + 4 = 0$ has the solutions $\lambda_1 = 4$, $\lambda_2 = 1$. Hence the general solution of the homogeneous equation $u'' - 5u' + 4u = 0$ is given by $c_1 e^{4t} + c_2 e^t$ with $c_1, c_2 \in \mathbb{R}$. On the other hand, $f(t) = e^{\xi_1 t} + e^{\xi_2 t}$ with $\xi_1 = 0$ and $\xi_2 = \lambda_2$. We look for a particular solution of the form

$$y_p(t) = y_p^{(1)}(t) + y_p^{(2)}(t)$$

where $y_p^{(1)}$ is a particular solution of $y'' - 5y' + 4y = 1$ and $y_p^{(2)}$ is a particular solution of $y'' - 5y' + 4y = e^t$. For $y_p^{(1)}$ we make the ansatz $y_p^{(1)}(t) = \alpha \in \mathbb{R}$. Upon substitution into the equation $y'' - 5y' + 4y = 1$ we get $y_p^{(1)}(t) = 1/4$. Since e^t is a solution of the homogeneous problem $y'' - 5y' + 4y = 0$, we make the ansatz $y_p^{(2)}(t) = \alpha t e^t$. Substituting it into $y'' - 5y' + 4y = e^t$ we obtain, with $y' = \alpha e^t + \alpha t e^t$ and $y'' = 2\alpha e^t + \alpha t e^t$,

$$-3\alpha e^t + \alpha t\Big(e^t - 5e^t + 4e^t\Big) = e^t$$

yielding $\alpha = -1/3$ and hence $y_p^{(2)} = -\frac{1}{3} t e^t$. Altogether

$$y_p(t) = \frac{1}{4} - \frac{1}{3} t e^t$$

is a solution of $y'' - 5y' + 4y = 1 + e^t$ and the general real solution is given by

$$c_1 e^{4t} + c_2 e^t + \frac{1}{4} - \frac{1}{3} t e^t \,, \quad c_1, c_2 \in \mathbb{R} \,.$$

The above formula can be used to solve the initial value problem

$$\begin{cases} y'' - 5y' + 4y = 1 + e^t \\ y(0) = \frac{1}{4} \,, \qquad y'(0) = \frac{8}{3} \,. \end{cases}$$

Indeed the constants $c_1, c_2 \in \mathbb{R}$ can be determined by solving the following linear system

$$c_1 + c_2 + \frac{1}{4} = \frac{1}{4} \,, \qquad 4c_1 + c_2 - \frac{1}{3} = \frac{8}{3} \,,$$

yielding $c_1 + c_2 = 0$ and $4c_1 + c_2 = 3$. It implies that $c_1 = 1$, $c_2 = -1$ and hence the unique solution of the initial value problem is

$$e^{4t} - e^t + \frac{1}{4} - \frac{1}{3} t e^t \,.$$